STA 293, Fall 2001
ISDS, Duke University

Instructor: Athanasios Kottas

The Dirichlet process

Consider a space Θ and a σ-field $\mathcal{B}$ of subsets of Θ. Following Ferguson (1973), a random probability measure, or equivalently a random distribution function, G on $(\Theta, \mathcal{B})$ follows a Dirichlet process $DP(\nu G_0)$ if, for any finite measurable partition, B_1,...,B_r of Θ, the distribution of the random vector $(G(B_1),...,G(B_r))$ is $Dirichlet(\nu G_0(B_1),...,\nu G_0(B_r))$, where $G(B_i)$ and $G_0(B_i)$ denote the probability of set B_i under G and G_0, respectively. Hence the Dirichlet process is characterized by two parameters, G_0 a specified distribution on $(\Theta, \mathcal{B})$ and ν a positive scalar parameter. In fact, since for any $B \in \mathcal{B}$, $E(G(B)) = G_0(B)$ and $Var(G(B)) = \{G_0(B)(1 - G_0(B))\}/(\nu + 1)$, G_0 is viewed as the center of the process while ν can be interpreted as a precision parameter; the larger ν is the *closer* we expect a realization from the process to be to G_0. νG_0 is referred to as the base measure of the process.

The standard criticism of the Dirichlet process is that it places all of its mass on the subset of discrete distributions on Θ (Ferguson, 1973, Blackwell, 1973). This property becomes evident if we consider the constructive definition of the Dirichlet process provided by Sethuraman and Tiwari (1982) and Sethuraman (1994). Specifically, let $\{z_s,\ s = 1,2,...\}$ and $\{\theta_j,\ j = 1,2,...\}$ be independent sequences of independent identically distributed (i.i.d.) random variables such that $z_s \sim Beta(1,\nu)$ and $\theta_j \sim G_0$. Then if we define $\omega_j = z_j \prod_{s=1}^{j-1}(1 - z_s)$, $j = 1,2,...$, whence $\sum_{j=1}^{\infty} \omega_j = 1$, a realization G from $DP(\nu G_0)$ is almost surely of the form

$$G = \sum_{j=1}^{\infty} \omega_j \delta_{\theta_j}, \tag{1}$$

where δ_a denotes the measure giving mass 1 to the point a. Another limitation of the Dirichlet process stems from the fact that it assigns negative correlation between $G(B_i)$ and $G(B_j)$ for any disjoint pair of B_i, $B_j \in \mathcal{B}$, an immediate consequence of a property of the Dirichlet distribution. This feature might be counter-intuitive in certain applications.

Prior to posterior updating using Dirichlet process priors is attractively straightforward. In particular, Ferguson (1973) proved that if $\theta = \{\theta_i,\ i = 1,...,n\}$ is an i.i.d. sample from G and a priori $G \sim DP(\nu G_0)$, then the posterior distribution of G given the data θ is again a Dirichlet process $DP(\nu^* G_0^*)$ with $\nu^* = \nu + n$ and $G_0^* = (\nu + n)^{-1}(\nu G_0 + \sum_{i=1}^{n} \delta_{\theta_i})$. Note that as ν tends to 0 (corresponding to a noninformative prior specification for G) the Bayes estimate for G, under integrated squared error loss, converges to the empirical distribution function of the sample which is the classical nonparametric estimator and also forms the basis for the Bayesian bootstrap (Rubin, 1981).

Further clarification for the Dirichlet process has been provided by the work of various authors on its theoretical properties and characterizations. Some of the related references are Blackwell and MacQueen (1973), Fabius (1973), Korwar and Hollander (1973), James and Mosimann (1980), Hannum, Hollander and Langberg (1981), Doss and Sellke (1982), Sethuraman and Tiwari (1982) and Lo (1983, 1991). The mean functional $\mu(G) = \int \theta G(d\theta)$, with $G \sim DP(\nu G_0)$, has received special attention. This is an almost surely finite random variable provided G_0 has finite first moment. Its distribution has been studied by Hannum, Hollander and Langberg (1981), Yamato (1984), Cifarelli and Regazzini (1990) and Diaconis and Kemperman (1996).

Regarding inference based on Dirichlet process priors, Ferguson (1973), apart from estimation for the unknown distribution function, presented a few other applications including estimation of the mean, variance and quantiles of the distribution. He also considered hypothesis testing involving quantiles and estimation of $P(X < Y)$ assigning independent Dirichlet process priors to the distribution functions of X and Y. The Mann-Whitney statistic (see, e.g., Randles and Wolfe, 1979) arises naturally in the latter case. Susarla and van Ryzin (1976, 1978) and Blum and Susarla (1977) extended the results of Ferguson on estimation of the distribution function (equivalently the survival function) based on right censored data. The Kaplan-Meier estimator is a limit of the resulting Bayes estimate under integrated squared error loss, again, when the precision parameter tends to 0. Treatments of the same problem but under a dependent censoring mechanism have been carried out by Phadia and Susarla (1983) and Tsai (1986). The case of grouped data was handled by Johnson and Christensen (1986). Incorporation of covariate information through the accelerated failure time model was considered by Christensen and Johnson (1988), employing

a semi-Bayesian approach for censored data, and Johnson and Christensen (1989) using a fully Bayesian approach in the absence of censoring. The use of Gibbs sampling (Gelfand and Smith, 1990) to provide full inference from doubly censored data was illustrated in Kuo and Smith (1992). The Dirichlet process has also found wide applicability as a prior for the tolerance distribution, or potency curve, in Bayesian bioassay. We refer to Ramsey (1972), Antoniak (1974), Bhattacharya (1981), Disch (1981), Ammann (1984) and Kuo (1983, 1988) for point estimates and various approximations to the associated posteriors, and Gelfand and Kuo (1991), Kottas, Branco and Gelfand (2000) and Mukhopadhyay (2000) for richer inference through the use of MCMC methods. For other Bayesian analyses with Dirichlet process priors see Campbell and Hollander (1978) for rank order estimation, Breth (1978, 1979) for construction of confidence bands for the distribution function and interval estimates for the associated mean and quantiles, Johnson, Susarla and van Ryzin (1979) in estimation for distribution functions of a branching process, Lo (1981) for an application to shock models and wear processes, Binder (1982) and Lo (1986) with regard to sampling from finite populations, Dalal and Phadia (1983) for estimation of a measure of dependence for bivariate distributions and Tamura (1988) in the context of statistical auditing. An extensive review of the work on the Dirichlet process, including additional references up to 1990, can be found in Ferguson, Phadia and Tiwari (1992).

Dalal (1979a) introduced the Dirichlet invariant process, an extension of the Dirichlet process, and used it to infer about the location parameter of a symmetric distribution (Dalal, 1979b). Diaconis and Freedman (1986a, b) were concerned with the consistency of the Bayes estimate of this parameter proving that it can be inconsistent for certain prior choices. See also Freedman and Diaconis (1983) for related work including discussion for Dirichlet process priors. Other variants of the Dirichlet process can be found in Doss (1985a, b), Newton, Czado and Chappell (1996), including applications to median estimation and binary regression, respectively, and Muliere and Tardella (1998) who defined the ϵ-Dirichlet process an approximation to the Dirichlet process suggested by its almost sure representation given in (1). Finally, the recent work of MacEachern (2000) on dependent Dirichlet processes holds promise since it can provide flexible modeling for a collection of dependent random distributions with direct applications in regression problems.