Consider a sample space $X$ and a $\sigma$-field $B$ of subsets of $X$. The Dirichlet process (Ferguson, 1973, 1974) is a random probability measure $Q$ on $(X, B)$, which therefore, equivalently, defines a random distribution function $G$ on $X$. It is characterized by two parameters, $Q_0$ a specified probability measure on $(X, B)$ (equivalently, $G_0$ a specified distribution function on $X$) and $\alpha$ a positive scalar parameter. The stochastic process $Q = \{Q(\omega, B) : \omega \in \Omega, B \in B\}$, with sample paths $Q_\omega \equiv Q(\omega, B)$, $B \in B$, that are probability measures on $(X, B)$, is a Dirichlet process if, for any finite measurable partition $B_1, \ldots, B_k$ of $X$, the distribution of the random vector $(Q(B_1), \ldots, Q(B_k))$ is $\text{Dirichlet}(\alpha Q_0(B_1), \ldots, \alpha Q_0(B_k))$, where $Q(B_i)$ (a random variable) and $Q_0(B_i)$ (a constant) denote the probability of set $B_i$ under $Q$ and $Q_0$, respectively. In fact, since for any $B \in B$, $E(Q(B)) = Q_0(B)$ and $\text{Var}(Q(B)) = \frac{Q_0(B)(1 - Q_0(B))}{\alpha + 1}$, $Q_0$ is viewed as the center of the process, whereas $\alpha$ can be interpreted as a precision parameter. For example, with $X = \mathbb{R}$ and $B = (-\infty, x], x \in \mathbb{R}, Q(B) = G(x)$ has a $\text{Beta}(\alpha G_0(x), \alpha(1 - G_0(x)))$ distribution, hence $E(G(x)) = G_0(x)$ and $\text{Var}(G(x)) = (G_0(x)(1 - G_0(x)))/{(\alpha + 1)}$. Thus the larger $\alpha$ is, the closer we expect a realization $G$ from the process to be to $G_0$. We will write $\text{DP}(\alpha, G_0)$ to denote the Dirichlet process. Moreover, $G$ will denote, depending on the context, the probability measure (distribution) or the associated distribution function generated by the DP.

The standard criticism of the Dirichlet process is that it, almost surely, generates discrete distributions on $X$ (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 $\{z_r, r = 1, 2, \ldots\}$ and $\{\theta_\ell, \ell = 1, 2, \ldots\}$ be independent sequences of independent identically distributed (i.i.d.) random variables such that $z_r \sim \text{Beta}(1, \alpha)$ and $\theta_\ell \sim G_0$. Then if we define $\omega_1 = z_1$, $\omega_\ell = z_\ell \prod_{r=1}^{\ell-1}(1 - z_r)$, $\ell = 2, 3, \ldots$ (note that $\sum_{\ell=1}^{\infty} \omega_\ell = 1$), a realization $G$ from $\text{DP}(\alpha, G_0)$ is, almost surely, of the form

$$G(\cdot) = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\theta_\ell}(\cdot),$$

(1)

where $\delta_a(\cdot)$ denotes a point mass at $a$. Hence the Dirichlet process yields, with probability 1, distributions that can be represented as countable mixtures of point masses, with locations $\theta_\ell$ that are i.i.d. draws from $G_0$ and weights $\omega_\ell$ that are generated using a “stick-breaking” procedure based on the i.i.d. $\text{Beta}(1, \alpha)$ draws $z_r$. 

AMS 241 — Winter 2009

Athanasios Kottas
Prior to posterior updating using Dirichlet process priors is straightforward. In particular, Ferguson (1973) proved that if data = \{y_i, i = 1,...,n\} is an i.i.d. sample from \(G\) and a priori \(G \sim \text{DP}(\alpha, G_0)\), then \(G \mid \text{data} \sim \text{DP}(\tilde{\alpha}, \tilde{G}_0)\) with \(\tilde{\alpha} = \alpha + n\) and

\[
\tilde{G}_0(t) = \frac{\alpha}{\alpha + n} G_0(t) + \frac{1}{\alpha + n} \sum_{i=1}^{n} 1[y_i, \infty)(t).
\]

Note that as \(\alpha\) tends to 0, \(E(G(t) \mid \text{data}) = \tilde{G}_0(t)\), which is 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.

Several authors have studied the theoretical properties of the Dirichlet process. Some of the related references are Antoniak (1974), Blackwell and MacQueen (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 \(\mu(G) = \int tG(dt)\), with \(G \sim \text{DP}(\alpha, G_0)\), has received special attention. (It can be shown that if \(G_0\) has finite first moment, then \(\mu(G)\) is an almost surely finite random variable.) Its distribution has been studied by Hannum, Hollander and Langberg (1981), Yamato (1984), Cifarelli and Regazzini (1990), Diaconis and Kemperman (1996), and Regazzini, Guglielmi and Di Nunno (2002). Consistency of estimators arising from nonparametric priors (and, in particular, Dirichlet process priors) has been studied by Diaconis and Freedman (1986a,b) and Freedman and Diaconis (1983). See also Ghosh and Ramamoorthi (2003) for a careful treatment of posterior consistency under nonparametric priors, including several more recent references.

Regarding inference based on Dirichlet process priors, Ferguson (1973), in addition to estimation for the unknown distribution function, presented details on 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 \(Y\). 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 arises as a limiting case of the 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 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 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), Mukhopadhyay (2000), and Kottas, Branco and Gelfand (2002) for full 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

Regarding extensions of the Dirichlet process, Dalal (1979a) introduced the Dirichlet invariant process and used it to infer about the location parameter of a symmetric distribution (Dalal, 1979b). Other variants of the Dirichlet process can be found in Doss (1985a,b) and Newton, Czado and Chappell (1996), including applications to median estimation and binary regression, respectively. Refer to the first set of course notes for more recent work involving several extensions of DP prior structures.

References


