

AMS 241: Bayesian Nonparametric Methods (Winter 2009)
Instructor: Athanasios Kottas

References for Dirichlet processes

Consider a sample space \mathcal{X} and a σ -field \mathcal{B} of subsets of \mathcal{X} . The Dirichlet process (Ferguson, 1973, 1974) is a random probability measure Q on $(\mathcal{X}, \mathcal{B})$, which therefore, equivalently, defines a random distribution function G on \mathcal{X} . It is characterized by two parameters, Q_0 a specified probability measure on $(\mathcal{X}, \mathcal{B})$ (equivalently, G_0 a specified distribution function on \mathcal{X}) and α a positive scalar parameter. The stochastic process $Q = \{Q(\omega, B) : \omega \in \Omega, B \in \mathcal{B}\}$, with sample paths $Q_\omega \equiv Q(\omega, B)$, $B \in \mathcal{B}$, that are probability measures on $(\mathcal{X}, \mathcal{B})$, is a Dirichlet process if, for any finite measurable partition B_1, \dots, B_k of \mathcal{X} , the distribution of the random vector $(Q(B_1), \dots, Q(B_k))$ is $\text{Dirichlet}(\alpha Q_0(B_1), \dots, \alpha Q_0(B_k))$, 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 $B \in \mathcal{B}$,

$$E(Q(B)) = Q_0(B)$$

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 α can be interpreted as a precision parameter. For example, with $\mathcal{X} = R$ and $B = (-\infty, x]$, $x \in R$, $Q(B) = G(x)$ 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 α 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 \mathcal{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 $\{z_r, r = 1, 2, \dots\}$ and $\{\theta_\ell, \ell = 1, 2, \dots\}$ 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 $\omega_1 = z_1$, $\omega_\ell = z_\ell \prod_{r=1}^{\ell-1} (1 - z_r)$, $\ell = 2, 3, \dots$ (note that $\sum_{\ell=1}^{\infty} \omega_\ell = 1$), a realization G from $\text{DP}(\alpha, G_0)$ is, almost surely, of the form

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

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 θ_ℓ that are i.i.d. draws from G_0 and weights ω_ℓ that are generated using a “stick-breaking” procedure based on the i.i.d. $\text{Beta}(1, \alpha)$ draws z_r .

Prior to posterior updating using Dirichlet process priors is straightforward. In particular, Ferguson (1973) proved that if data = $\{y_i, i = 1, \dots, n\}$ 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 α 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

(1981) for an application to shock models and wear processes, Binder (1982) and Lo (1986) with regard to sampling from finite populations, 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).

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

- Ammann, L.P. (1984), "Bayesian Nonparametric Inference for Quantal Response Data," *The Annals of Statistics*, 12, 636-645.
- Antoniak, C.E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.
- Bhattacharya, P.K. (1981), "Posterior Distribution of a Dirichlet Process from Quantal Response Data," *The Annals of Statistics*, 9, 803-811.
- Binder, D.A. (1982), "Non-parametric Bayesian Models for Samples from Finite Populations," *Journal of the Royal Statistical Society, Ser. B*, 44, 388-393.
- Blackwell, D. (1973), "Discreteness of Ferguson Selections," *The Annals of Statistics*, 1, 356-358.
- Blackwell, D., and MacQueen, J.B. (1973), "Ferguson Distributions via Pólya Urn Schemes," *The Annals of Statistics*, 1, 353-355.
- Blum, J., and Susarla, V. (1977), "On the Posterior Distribution of a Dirichlet Process Given Randomly Right Censored Observations," *Stochastic Processes and their Applications*, 5, 207-211.
- Breth, M. (1978), "Bayesian Confidence Bands for a Distribution Function," *The Annals of Statistics*, 6, 649-657.
- Breth, M. (1979), "Nonparametric Bayesian Interval Estimation," *Biometrika*, 66, 641-644.
- Campbell, G., and Hollander, M. (1978), "Rank Order Estimation With the Dirichlet Prior," *The Annals of Statistics*, 6, 142-153.
- Christensen, R., and Johnson, W. (1988), "Modeling Accelerated Failure Time With a Dirichlet Process," *Biometrika*, 75, 693-704.
- Cifarelli, D.M., and Regazzini, E. (1990), "Distribution Functions of Means of a Dirichlet Process," *The Annals of Statistics*, 18, 429-442.

- Dalal, S.R. (1979a), "Dirichlet Invariant Processes and Applications to Nonparametric Estimation of Symmetric Distribution Functions," *Stochastic Processes and their Applications*, 9, 99-107.
- Dalal, S.R. (1979b), "Nonparametric and Robust Bayes Estimation of Location," in *Optimizing Methods in Statistics*, ed. J.S. Rustagi, New York: Academic Press, pp. 141-166.
- Diaconis, P., and Freedman, D. (1986a), "On the Consistency of Bayes Estimates" (with discussion), *The Annals of Statistics*, 14, 1-67.
- Diaconis, P., and Freedman, D. (1986b), "On Inconsistent Bayes Estimates of Location," *The Annals of Statistics*, 14, 68-87.
- Diaconis, P., and Kemperman, J. (1996), "Some New Tools for Dirichlet Priors," in *Bayesian Statistics 5, Proceedings of the Fifth Valencia International Meeting*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford: Oxford Clarendon Press, pp. 97-106.
- Disch, D. (1981), "Bayesian Nonparametric Inference for Effective Doses in a Quantal-Response Experiment," *Biometrics*, 37, 713-722.
- Doss, H. (1985a), "Bayesian Nonparametric Estimation of the Median; part I: Computation of the Estimates," *The Annals of Statistics*, 13, 1432-1444.
- Doss, H. (1985b), "Bayesian Nonparametric Estimation of the Median; part II: Asymptotic Properties of the Estimates," *The Annals of Statistics*, 13, 1445-1464.
- Doss, H., and Sellke, T. (1982), "The Tails of Probabilities Chosen from a Dirichlet Prior," *The Annals of Statistics*, 10, 1302-1305.
- Ferguson, T.S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209-230.
- Ferguson, T.S. (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615-629.
- Ferguson, T.S., Phadia, E.G., and Tiwari, R.C. (1992), "Bayesian Nonparametric Inference," in *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, pp. 127-150.
- Freedman, D., and Diaconis, P. (1983), "On Inconsistent Bayes Estimates in the Discrete Case," *The Annals of Statistics*, 11, 1109-1118.
- Gelfand, A.E., and Kuo, L. (1991), "Nonparametric Bayesian Bioassay Including Ordered Polytomous Response," *Biometrika*, 78, 657-666.
- Ghosh, J.K., and Ramamoorthi, R.V. (2003), *Bayesian Nonparametrics*, New York: Springer.
- Hannum, R.C., Hollander, M., and Langberg, N.A. (1981), "Distributional Results for Random Functionals of a Dirichlet Process," *The Annals of Probability*, 9, 665-670.
- James, I.R., and Mosimann, J.E. (1980), "A New Characterization of the Dirichlet Distribution Through Neutrality," *The Annals of Statistics*, 8, 183-189.
- Johnson, R.A., Susarla, V., and van Ryzin, J. (1979), "Bayesian Non-parametric Estimation for Age-dependent Branching Processes," *Stochastic Processes and their Applications*, 9, 307-318.

- Johnson, W., and Christensen, R. (1986), "Bayesian Nonparametric Survival Analysis for Grouped Data," *The Canadian Journal of Statistics*, 14, 307-314.
- Johnson, W., and Christensen, R. (1989), "Nonparametric Bayesian Analysis of the Accelerated Failure Time Model," *Statistics and Probability Letters*, 8, 179-184.
- Korwar, R.M., and Hollander, M. (1973), "Contributions to the Theory of Dirichlet Processes," *The Annals of Probability*, 1, 705-711.
- Kottas, A., Branco, M.D., and Gelfand, A.E. (2002), "A Nonparametric Bayesian Modeling Approach for Cytogenetic Dosimetry," *Biometrics*, 58, 593-600.
- Kuo, L. (1983), "Bayesian Bioassay Design," *The Annals of Statistics*, 11, 886-895.
- Kuo, L. (1988), "Linear Bayes Estimators of the Potency Curve in Bioassay," *Biometrika*, 75, 91-96.
- Kuo, L., and Smith, A.F.M. (1992), "Bayesian Computations in Survival Models via the Gibbs Sampler," in *Survival Analysis: State of the Art*, eds. J.P. Klein and P.K. Goel, Dordrecht: Kluwer Academic Publishers, pp. 11-24.
- Lo, A.Y. (1981), "Bayesian Nonparametric Statistical Inference for Shock Models and Wear Processes," *Scandinavian Journal of Statistics*, 8, 237-242.
- Lo, A.Y. (1983), "Weak Convergence for Dirichlet Processes," *Sankhyā*, Ser. A, 45, 105-111.
- Lo, A.Y. (1986), "Bayesian Statistical Inference for Sampling a Finite Population," *The Annals of Statistics*, 14, 1226-1233.
- Lo, A.Y. (1991), "A Characterization of the Dirichlet Process," *Statistics and Probability Letters*, 12, 185-187.
- Mukhopadhyay, S. (2000), "Bayesian Nonparametric Inference on the Dose Level With Specified Response Rate," *Biometrics*, 56, 220-226.
- Newton, M.A., Czado, C., and Chappell, R. (1995), "Bayesian Inference for Semiparametric Binary Regression," *Journal of the American Statistical Association*, 91, 132-141.
- Phadia, E.G., and Susarla, V. (1983), "Nonparametric Bayesian Estimation of a Survival Curve With Dependent Censoring Mechanism," *Annals of the Institute of Statistical Mathematics*, 35, 389-400.
- Ramsey, F.L. (1972), "A Bayesian Approach to Bioassay," *Biometrics*, 28, 841-858.
- Regazzini, E., Guglielmi, A., and Di Nunno, G. (2002), "Theory and numerical analysis for exact distributions of functionals of a Dirichlet process," *The Annals of Statistics*, 30, 1376-1411.
- Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639-650.
- Sethuraman, J., and Tiwari, R.C. (1982), "Convergence of Dirichlet Measures and the Interpretation of their Parameter," in *Statistical Decision Theory and Related Topics III*, eds. S. Gupta and J.O. Berger, New York: Springer-Verlag, 2, pp. 305-315.

- Susarla, V., and Van Ryzin, J. (1976), “Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations,” *Journal of the American Statistical Association*, 71, 897-902.
- Susarla, V., and Van Ryzin, J. (1978), “Large Sample Theory for a Bayesian Nonparametric Survival Curve Estimator Based on Censored Samples,” *The Annals of Statistics*, 6, 755-768. Addendum (1980), *The Annals of Statistics*, 8, 693.
- Tamura, H. (1988), “Estimation of Rare Errors Using Expert Judgement,” *Biometrika*, 75, 1-9.
- Tsai, W.-Y. (1986), “Estimation of Survival Curves from Dependent Censorship Models via a Generalized Self-consistent Property With Nonparametric Bayesian Estimation Application,” *The Annals of Statistics*, 14, 238-249.
- Yamato, H. (1984), “Characteristic Functions of Means of Distributions Chosen from a Dirichlet Process,” *The Annals of Probability*, 12, 262-267.