On GEM diffusive mixtures

Luis Gutiérrez Inostroza* Ramsés H. Mena† Matteo Ruggiero‡

Abstract

We present and discuss the class of dependent mixture models induced by diffusive measure-valued GEM processes. These extend the family of diffusive Dirichlet mixtures to the multiparameter case, including dependent Pitman–Yor mixtures as special cases, and have applications to Bayesian nonparametric inference for dynamic density estimation with discretely collected data.

Key Words: Bayesian nonparametrics, dependent process, mixture model, diffusion process, measure-valued process

1. Introduction

A current major research frontier in Bayesian nonparametric poses as the main task to extend mixtures of type $\tilde{f}_P(y) = \int K(y|\, x)P(\mathrm{d}x)$, where $K(\cdot|y)$ is a kernel density, and P is a discrete random probability measure, to yield a dependent mixture of the type

$$\tilde{f}_{P_z}(y) = \int K(y|\, x)P_z(\mathrm{d}x), \tag{1}$$

where the random probability measure is now indexed by a covariate z. This entails accommodating forms of dependence more general than exchangeability. One such case is that of partial exchangeability, whereby observations are assumed to be exchangeable conditionally on the values of z, but not overall exchangeable. This line of research was initiated by MacEachern (1999, 2000), who proposed the so-called dependent Dirichlet process or more generally a class of dependent processes given by the collection of random measures

$$\left\{ P_z = \sum_{i=1}^{\infty} W_i(z)\delta_{X_i(z)},\ z \in \mathbb{Z} \right\}$$

where the dependence is induced by means of $z \in \mathbb{Z}$ through the random weights W_i and/or the random atoms X_i, for some appropriate space $\mathbb{Z}$. A non exhaustive list of papers in this area includes De Iorio et al. (2004); Gelfand et al. (2005); Dunson (2006); Griffin and Steel (2006); Caron et al. (2008); Dunson, Pillai and Park (2007); Dunson and Park (2008); Caron et al. (2007); Rodriguez and Ter Horst (2008); Fuentes-García (1973); Mena et al. (2011); Caron et al. (2016); Gutierrez et al. (2016); Mena and Ruggiero (2016). See also Hjort et al. (2010) and references therein. For the case of Dirichlet processes (Ferguson, 1973), the time evolution has often been built into the process by exploiting its stick-breaking representation (Sethuraman, 1994). For the Dirichlet case this is obtained by specifying the weights in

$$P = \sum_{i=1}^{\infty} W_i\delta_{X_i} \tag{2}$$

as follows

$$W_1 = V_1, \qquad W_i = V_i \prod_{j=1}^{i-1}(1 - V_j), \qquad V_i \overset{iid}{\sim} \text{Beta}(1, \theta), \tag{3}$$

*Pontificia Universidad Católica de Chile
†Universidad Nacional Autónoma de México
‡University of Torino and Collegio Carlo Alberto

and associating to each of them a distinct location $X_i \in \mathbb{X}$, where $X_i \overset{\text{iid}}{\sim} G^*$ and G^* is nonatomic. The stick-breaking construction has demonstrated to lend itself quite easily to the implementation in applied problems with the aid of Markov Chain Monte Carlo techniques, most notably via exploitation of the slice sampler (Damien, Wakefield and Walker, 1999; Walker, 2007) and the retrospective sampler (Papaspiliopoulos and Roberts, 2008). See also Ishwaran and James (2001).

In this note, we discuss some details on an extension of diffusive Dirichlet mixtures, only briefly mentioned in Mena and Ruggiero (2016), to GEM diffusive mixtures. A diffusive Dirichlet process is a collection of random probability measures

$$P_t = \sum_{i=1}^{\infty} W_i(t)\delta_{X_i}, \quad t \geq 0, \qquad X_i \overset{iid}{\sim} G^* \tag{4}$$

where

$$W_1(t) = V_1(t), \qquad W_i(t) = V_i(t)\prod_{j<i}(1 - V_j(t)), \qquad V_i(\cdot) \overset{iid}{\sim} \text{WF}(1, \theta). \tag{5}$$

Here $V_i(\cdot) \overset{\text{iid}}{\sim} \text{WF}(1, \theta)$ indicates that for each $i \geq 1$, $V_i(\cdot) = \{V_i(t), t \geq 0\}$ is defined to be one-dimensional Wright–Fisher diffusion with parameters $(1, \theta)$, i.e. the unique solution in $[0, 1]$ of the stochastic differential equation

$$\mathrm{d}V(t) = \frac{1}{2}[1 - (1 + \theta)V(t)]\mathrm{d}t + \sqrt{V(t)(1 - V(t))}\mathrm{d}B(t), \qquad V(t) \in [0, 1], \tag{6}$$

where $B(t)$ denotes a one-dimensional standard Brownian motion. Such diffusion is stationary and reversible with respect to a Beta$(1, \theta)$ distribution, and this in turn implies that (4) is a diffusion process, taking values in the space of discrete probability measures on $\mathbb{X}$, with marginal laws coinciding with that of a Dirichlet process. A diffusive Dirichlet mixture is then defined through (1) by letting P_z be as in (4)-(6), with z being the time index t.

Measure-valued diffusions are well known objects in population genetics, usually describing the temporal evolution of a large population of individuals or alleles, whose random reproduction mechanism can be subject to mutation and possibly selection. See Ethier and Kurtz (1993); Dawson (1993, 2010); Feng (2010) for reviews, and Ruggiero and Walker (2009a,b); Prünster and Ruggiero (2013); Ruggiero et al. (2013); Papaspiliopoulos et al. (2016) for some connections with Bayesian nonparametrics.

The proofs of the results presented in this note are easily adapted from Mena and Ruggiero (2016), and therefore omitted.

2. GEM diffusive mixtures

Let $\mathbf{a} = (a_1, a_2, \ldots)$ and $\mathbf{b} = (b_1, b_2, \ldots)$ be vectors of positive parameters. The GEM distribution with parameters $(\mathbf{a}, \mathbf{b})$ is the law of the vector $W = (W_1, W_2, \ldots)$, where

$$W_1 = V_1, \qquad W_i = V_i\prod_{j=1}^{i-1}(1 - V_j), \qquad V_i \overset{ind}{\sim} \text{Beta}(a_i, b_i).$$

Here $(\mathbf{a}, \mathbf{b})$ are chosen so that the W_i's sum up to one; see Ishwaran and James (2001). When $a_i = 1$ and $b_i = \theta > 0$ for all $i \geq 1$, we recover the law GEM(θ) of the weights of a Dirichlet process as in (3), and the law of the vector obtained by sorting in decreasing order

the components of W is the Poisson–Dirichlet distribution with parameter θ (Kingman, 1975). See Ethier (1981); Ethier and Kurtz (1981) for related diffusion models. When $a_i = 1-\sigma$ for all $i \geq 1$ and $\mathbf{b} = (\theta+\sigma, \theta+2\sigma, \ldots)$, we obtain the GEM$(\theta, \sigma)$ distribution, whose decreasingly ordered components have the two-parameter Poisson–Dirichlet distribution (Perman, Pitman and Yor, 1992; Pitman, 1995; Pitman and Yor, 1997). See Petrov (2009); Ruggiero and Walker (2009b); Feng and Sun (2010); Costantini et al. (2016) for related diffusion models.

Let now $(V_1(t), V_2(t), \ldots)$ be a collection of infinitely many independent Wright–Fisher diffusions with parameters (a_i, b_i), where $V_i(t)$ solves

$$\mathrm{d}V_i(t) = \frac{1}{2}[a_i(1-V(t)) - b_i V(t)]\mathrm{d}t + \sqrt{V(t)(1-V(t))}\mathrm{d}B(t), \qquad V(t) \in [0,1]. \quad (7)$$

A GEM(**a**,**b**) diffusion is defined as the vector process $W(t) = (W_1(t), W_2(t), \ldots)$ where each component $W_i(t)$ is constructed by means of the stick-breaking procedure

$$W_1(t) = V_1(t), \qquad W_i(t) = V_i(t)\prod_{j<i}(1-V_j(t)), \qquad V_i(\cdot) \overset{ind}{\sim} \mathrm{WF}(a_i, b_i). \quad (8)$$

Feng and Wang (2007) characterised this class of processes, which are well defined infinite-dimensional diffusions, whose sample paths are continuous functions from $[0,\infty)$ to the infinite simplex

$$\Delta_\infty = \left\{ \mathbf{w} : (w_1, w_2, \ldots) : w_i \geq 0, \ \sum_{i\geq 1} w_i = 1 \right\}.$$

When $a_i = 1$ and $b_i = \theta > 0$ for all $i \geq 1$, it is immediate that the GEM diffusions reduce to (5)-(6).

Denote by B_{a_i,b_i} a Beta(a_i, b_i) distribution and, for $\mathbf{a} = (a_1, a_2, \ldots)$ and $\mathbf{b} = (b_1, b_2, \ldots)$, let

$$B_{\mathbf{a},\mathbf{b}} = \prod_{i=1}^{\infty} B_{a_i,b_i} \qquad\qquad \tilde{B}_{\mathbf{a},\mathbf{b}} = B_{\mathbf{a},\mathbf{b}} \circ \phi^{-1},$$

where ϕ is the function

$$\begin{aligned} &\phi : [0,1]^{\mathbb{N}} \to \Delta_\infty \\ &\mathbf{v} = (v_1, v_2, \ldots) \mapsto \phi_1(\mathbf{v}) = v_1, \ \phi_i(\mathbf{v}) = v_i(1-v_1)\cdots(1-v_{i-1}), \ i \geq 1. \end{aligned} \quad (9)$$

From Feng and Wang (2007), Theorem 2.1, it follows that the GEM$(\mathbf{a}, \mathbf{b})$ diffusion has reversible measure $\tilde{B}_{\mathbf{a},\mathbf{b}}$. In particular, this implies that at stationarity the vector $(W_1(t), W_2(t), \ldots)$ has the GEM$(\mathbf{a}, \mathbf{b})$ distribution.

Letting the dependence in a random probability measure of type (4) be induced by the above construction, or, from a different viewpoint, attaching *iid* atoms to each component of the GEM diffusion, leads easily to the definition of measure-valued GEM processes.

Definition 2.1. *Let* $\boldsymbol{a} = (a_1, a_2, \ldots)$ *and* $\boldsymbol{b} = (b_1, b_2, \ldots)$, *where* $a_i, b_i > 0$ *for all* $i \geq 1$. *A measure-valued GEM diffusion is a dependent random probability measure*

$$P_t = \sum_{i\geq 1} W_i(t)\delta_{X_i}, \qquad\qquad t \geq 0,$$

with weights given by the GEM$(\mathbf{a}, \mathbf{b})$ *diffusion* $W(t)$ *defined as in (8) and atoms* $X_i \overset{iid}{\sim} G^*$, *where* G^* *is a non atomic probability measure on* $\mathbb{X}$. □

We will denote by MV-GEM($\mathbf{a}, \mathbf{b}$) the measure-valued process $P = \{P_t, t \geq 0\}$ of Definition 2.1. The following proposition shows that a measure-valued GEM process has nice path properties. In particular it is Feller, hence strong Markov, and has trajectories given by continuous functions from $[0, \infty)$ to the set $\mathscr{P}(\mathbb{X})$ of discrete probability measures on $\mathbb{X}$.

Proposition 2.2. *Let P_t be a measure-valued* GEM($\mathbf{a}, \mathbf{b}$) *process as in Definition 2.1. Then P_t is a Feller process with almost surely continuous sample paths from $[0, \infty)$ to $\mathscr{P}(\mathbb{X})$.*

We now turn to stationarity. The following result identifies the reversible measure of the measure-valued GEM process. For any fixed sequence $\mathbf{x} = (x_i)_{i \geq 1}$, denote by $\varphi_x : \Delta_\infty \times \mathbb{X}^\infty \to \mathscr{P}(\mathbb{X})$ the transformation

$$\varphi_{\mathbf{x}}(\phi(\mathbf{v})) = \sum_{i=1}^{\infty} w_i \delta_{x_i}, \tag{10}$$

where $\phi(\mathbf{v})$ is as in (9).

Proposition 2.3. *Let P be a* MV-GEM($\mathbf{a}, \mathbf{b}$) *process. Then P is reversible with respect to $\mathcal{B}_{\boldsymbol{a},\boldsymbol{b}} = \tilde{B}_{\boldsymbol{a},\boldsymbol{b}} \circ \varphi_{\mathbf{x}}^{-1}$.*

From Proposition 2.3 it follows immediately that a measure-valued GEM($\mathbf{a}, \mathbf{b}$) process has invariant measure $\mathcal{B}_{\mathbf{a},\mathbf{b}}$. Again, two immediate special cases implied by this stationarity result are familiar: when $a_i = 1, b_i = \theta > 0$ for all $i \geq 1$, P_t is marginally a Dirichlet process; when $a_i = 1 - \sigma$ for all $i \geq 1$ and $\mathbf{b} = (\theta + \sigma, \theta + 2\sigma, \ldots)$, then P_t is marginally a Pitman–Yor process.

Measure-valued GEM processes are therefore amenable to be used at the top level of Bayesian nonparametric mixtures in a time dependent framework, to define a class of diffusive dependent GEM mixtures. This entails considering the hierarchical model

$$\begin{aligned} y|x &\sim K(y|x) \\ x|t, P_t &\sim P_t \\ P = \{P_t\}_{t \geq 0} &\sim \text{MV-GEM}(\mathbf{a}, \mathbf{b}) \end{aligned} \tag{11}$$

or equivalently (1) with a MV-GEM($\mathbf{a}, \mathbf{b}$) process in place of P_z. This formulation provides additional flexibility with respect to the diffusive Dirichlet mixture of Mena and Ruggiero (2016). Note that by restricting (11) to a single t yields the class of mixture models considered in Ishwaran and James (2001).

3. Posterior computation

Mena and Ruggiero (2016) presented a Gibbs sampler algorithm that can be used for posterior inference under the mixture model (11) when P is a diffusive Dirichlet process. Here we highlight the differences with respect to the Dirichlet case, and refer the reader to Section 4 and Appendix B of the cited source for further details.

We assume for simplicity that univariate data points $\mathbf{y}^{(n)} = (y_{t_1}, \ldots, y_{t_n})$ are observed at times $0 \leq t_1 < \ldots < t_n$, where time intervals need not be equally spaced. The target of inference is the data generating time-varying distribution (1), such that $y_{t_i} \sim \tilde{f}_{P_{t_i}}$ where we model $P = \{P_t\}_{t \geq 0} \sim$ MV-GEM($\mathbf{a}, \mathbf{b}$). To this end, let $V(t) = (V_1(t), V_2(t), \ldots)$ be as in (8). The above non-parametric mixture model can be written as

$$f_{P_t}(y \mid V(t), X) = \sum_{j \geq 1} W_j(t)\, K(y \mid X_j). \tag{12}$$

Following Walker (2007) this random density can be augmented to

$$f_{P_t}(y, u, s \mid V(t), X) = \mathbb{I}(u < W_s(t))\, K(y \mid X_s) \tag{13}$$

or to a variation of it aimed at a more efficient MCMC, given by

$$f_{P_t}(y, u, s \mid V(t), X) = \psi_s^{-1}\mathbb{I}(u < \psi_s)\, W_s(t)\, K(y \mid X_s) \tag{14}$$

where $s \mapsto \psi_s$ is a $\mathbb{N}$-valued function with known inverse ψ^*, e.g. $e^{-\eta s}$, for $0 \leq \eta \leq 1$. The latent variable s indexes the specific kernel K which better represents the observation y and $u \sim \mathsf{U}(0, \psi_s)$. Notice that when $\psi_s = W_s$ we can recover (13) from (14). See Kalli et al. (2011) for more details on (14). The conditional augmented likelihood is given by

$$\mathcal{L}\left(\mathbf{y}^{(n)}, \mathbf{u}^{(n)}, \mathbf{s}^{(n)} \mid V(t), X\right) = \prod_{i=1}^{n} \psi_{s_i}^{-1}\mathbb{I}(u_i < \psi_{s_i}) \left[V_{s_i}(t_i) \prod_{k<s_i} (1 - V_k(t_i))\right] K(y_i \mid Xs_i)$$

where $\mathbf{y}^{(n)} := (y_1, \ldots, y_n)$, $\mathbf{u}^{(n)} := (u_1, \ldots, u_n)$, $\mathbf{s}^{(n)} := (s_1, \ldots, s_n)$, $y_i := y_{t_i}$, $u_i := u_{t_i}$ and $s_i := s_{t_i}$. The random truncation induced by the slice sampler in Walker (2007) simplifies the posterior learning process, thus now we have to learn about the first m Wright–Fisher processes and locations. The value of m is an integer given

$$m := \max(\lfloor \psi^*(u_{t_1}) \rfloor, \lfloor \psi^*(u_{t_2}) \rfloor, \ldots, \lfloor \psi^*(u_{t_n}) \rfloor),$$

with $\lfloor A \rfloor$ denoting the integer part of A. Under the above setting, at each iteration of the algorithm, we have to update a matrix V of dimension $m \times n$ and a vector x of length m. Each row of V denoted by $v_j^{(n)}$ is a stochastic process and the m processes are mutually independent. The assumption of independence also holds for the m locations, denoted here by x. Thus, the prior distribution on V and x can be defined as

$$\mathcal{L}(\mathsf{V}) = \prod_{j=1}^{m} \mathcal{L}(v_j^{(n)}), \text{ and } \mathcal{L}(\mathsf{x}) = \prod_{j=1}^{m} \mathcal{L}(x_j)$$

where

$$\mathcal{L}(v_j^{(n)}) = \pi_v(v_j(t_1)) \prod_{i=2}^{n} \mathsf{p}_v(v_j(t_i) \mid v_j(t_{i-1})),$$

with $\pi_v := \text{Beta}(a_1, b_1)$ and p_v is the transition density of the Wright–Fisher diffusion process. An expression for such a transition density can be found in Ethier and Griffiths (1993), and is given by

$$\mathsf{p}_v(v_j(t_j) \mid v_j(t_{j-1})) = \sum_{k=0}^{\infty} q_k^{\theta}(t) \sum_{l=0}^{k} \text{Beta}(v_j(t_j) \mid a + l, b + k - l)\text{Bin}(l \mid k, v_j(t_{j-1})) \tag{15}$$

where $\text{Beta}(\cdot \mid a, b)$ is the Beta density with parameters a, b, $\text{Bin}(\cdot \mid m, q)$ is the Binomial probability mass function with m trials and success probability q and $q_k^{\theta}(t)$ are the transition probabilities of a death process with an entrance boundary of infinity, and death rates $k(k + \theta - 1)/2$, $k \geq 1$. See Griffiths and Spanò (2010). With the above elements we have that the posterior distribution for V and x is given by

$$\mathcal{L}(\mathsf{V}, \mathsf{x} \mid \mathbf{y}^{(n)}, \mathbf{u}^{(n)}, \mathbf{s}^{(n)}) \propto \mathcal{L}(\mathbf{y}^{(n)}, \mathbf{u}^{(n)}, \mathbf{s}^{(n)} \mid \mathsf{V}, \mathsf{x})\mathcal{L}(\mathsf{V})\mathcal{L}(\mathsf{x}) \tag{16}$$

The posterior distribution remains the unchanged prior for all v_l when $l > m$.

3.1 Updating the locations

As locations are not time dependent we have that,

$$\mathcal{L}(x_j \mid \cdots) \propto \mathcal{L}(x_j) \prod_{\{i:s_i=j\}} K(y_{t_i} \mid x_j) \tag{17}$$

which is the usual way of update the parameters in a mixture model.

3.2 Updating of the weights processes

For each process $j = 1, \ldots, m$ and $i \neq 1, n$ the weight process is updated component–wise by

$$\begin{aligned} \mathcal{L}(v_j(t_i) \mid \cdots) &\propto \mathsf{p}_v(v_j(t_{i+1}) \mid v_j(t_i))\, \mathsf{p}_v(v_j(t_i) \mid v_j(t_{i-1})) \\ &\times v_j(t_i)^{\mathbb{I}(s_i=j)}\, (1 - v_j(t_i))^{\mathbb{I}(s_i>j)} \end{aligned} \tag{18}$$

and

$$\begin{aligned} \mathcal{L}(v_j(t_1) \mid \cdots) &\propto \mathsf{p}_v(v_j(t_2) \mid v_j(t_1))\, \mathsf{p}_v(v_j(t_1)) \\ &\times v_j(t_1)^{\mathbb{I}(s_1=j)}\, (1 - v_j(t_1))^{\mathbb{I}(s_1>j)} \end{aligned} \tag{19}$$

$$\begin{aligned} \mathcal{L}(v_j(t_n) \mid \cdots) &\propto \mathsf{p}_v(v_j(t_n) \mid v_j(t_{n-1})) \\ &\times v_j(t_n)^{\mathbb{I}(s_n=j)}\, (1 - v_j(t_n))^{\mathbb{I}(s_n>j)} \end{aligned} \tag{20}$$

3.3 Updating of the slice and membership variables

The slice variable u is updated from a uniform distribution given by

$$\mathcal{L}(u_{t_i} \mid \cdots) = \mathsf{U}(0, \psi_{s_{t_i}})$$

and the membership variable s from a discrete distribution given by

$$\mathcal{L}(s_{t_i} \mid \cdots) \propto \psi_{s_{t_i}}^{-1}\, w_{s_{t_i}}(t_i)\, K(y_{t_i} \mid \mathsf{x}_{s_{t_i}}(t_i))\, \mathbb{I}(s_{t_i} \in \left\{k : \psi_{s_{t_i}} > u_{t_i}\right\})$$

Notice that since $\left\{k : \psi_{s_{t_i}} > u_{t_i}\right\}$ is a finite set, this latter distribution is easy to sample, i.e. from $s_{t_i} = 1, \ldots, \lfloor \psi^*(u_{t_i}) \rfloor$.

The resulting Monte Carlo sample can be used for computing various estimates of interest, such as, e.g., the estimated posterior density $\hat{f}_{P_t}$, posterior mean functionals $\eta_t = \int y \hat{f}_{P_t}(y) \mathrm{d}y$ and their moments, or for constructing appropriate credible intervals.

Ackowledgements

The first author is partially supported by Fondecyt grant 11140013. The second author is supported by CONACyT project 241195. The third author is supported by the European Research Council (ERC) through StG "N-BNP" 306406.

REFERENCES

CARON, F., DAVY. M. AND DOUCET, A. (2007) Generalized Polya urn for time-varying Dirichlet process mixtures. *Proc. 23rd Conf. on Uncertainty in Artificial Intelligence*, Vancouver.

CARON, F., DAVY, M., DOUCET, A., DUFLOS, E. AND VANHEEGHE, P. (2008). Bayesian inference for linear dynamic models with Dirichlet process mixtures. *IEEE Trans. Sig. Proc.* **56**, 71–84.

CARON, F., NEISWANGER, W., WOOD, F., DOUCET, A. and DAVY, M. (2016). Generalized Pólya urn for time-varying Pitman–Yor processes. *J. Mach. Learn. Res.*, in press.

COSTANTINI, C, DE BLASI, P., ETHIER, S.N., RUGGIERO, M. and SPANÒ, D.(2016). Wright–Fisher construction of the two-parameter Poisson–Dirichlet diffusion. arXiv:1601.06064.

DAMIEN, P., WAKEFIELD, J.C. AND WALKER, S.G. (1999). Gibbs sampling for Bayesian nonconjugate and hierarchical models using auxiliary variables. *J. Roy. Statist. Soc. Ser. B* **61**, 331–344.

DAWSON, D.A. (1993). *Measure-valued Markov processes*. Ecole d'Eté de Probabilités de Saint Flour XXI. Lecture Notes in Mathematics **1541**. Springer, Berlin.

DAWSON, D.A. (2010). *Introductory lectures on stochastic population systems*. Technical Report Series **451**, Laboratory for Research in Statistics and Probability, Carleton University.

DE IORIO, M., MÜLLER, P., ROSNER, G. AND MACEACHERN, S.N.(2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.* **99**, 205–215.

DUNSON, D.B. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics* 7, 551–568.

DUNSON, D.B. AND PARK, J-H. (2008). Kernel stick breaking processes. *Biometrika* **95**, 307–323.

DUNSON, D.B., PILLAI, N. AND PARK, J-H. (2007). Bayesian density regression. *J. Roy. Statist. Soc. Ser. B* **69**, 163–183.

Ethier, S.N. (1981). A class of infinite-dimensional diffusions occurring in population genetics. *Indiana Univ. Math. J.* **30**, 925–935.

ETHIER, S.N. AND KURTZ, T.G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.* **13**, 429–452.

ETHIER, S.N. AND KURTZ, T.G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31**, 345–386.

ETHIER, S.N. AND GRIFFITHS, R. C. (1993). The transition function of a Fleming–Viot process. *Ann. Probab.* **21**, 1571–1590.

FENG, S. (2010). *The Poisson–Dirichlet distribution and related topics*. Springer, Heidelberg.

FENG, S. AND SUN, W. (2010). Some diffusion processes associated two-parameter Poisson–Dirichlet distribution and Dirichlet process. *Probab. Theory Relat. Fields* **148**, 501–525.

FENG, S. AND WANG, F.Y. (2007). A class of infinite-dimensional diffusion processes with connection to population genetics. *J. Appl. Probab.* **44**, 938–949.

FERGUSON, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

FUENTES-GARCÍA, R., MENA, R.H. and WALKER, S.G. (2009). A nonparametric dependent process for Bayesian regression. *Statist. Probab. Lett.* **79**, 1112–1119.

GELFAND, A., KOTTAS, A. AND MACEACHERN, S. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Amer. Statist. Assoc.* **100**, 1021–1035.

GRIFFIN, J.E. AND STEEL, M.F.J. (2006). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 179–194.

GRIFFITHS, R.C. AND SPAN, D. (2010). Diffusion processes and coalescent trees. In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*. Ed. N. H. Bingham and C. M. Goldie. London Mathematical Society Lecture Notes Series, Cambridge University Press 2010.

GUTIERREZ, L., MENA, R.H. and RUGGIERO, M. (2016). A time dependent Bayesian nonparametric model for air quality analysis. *Comput. Statist. Data Anal.* **95**, 161–175.

HJORT, N.L., HOLMES, C.C. MÜLLER, P., WALKER, S.G. (Eds.) (2010). *Bayesian Nonparametrics*. Cambridge University Press, Cambridge.

ISHWARAN, H. AND JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Ass.* **96**, 161–173.

KINGMAN, J.F.C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B* **37**, 1–22.

KALLI, M., GRIFFIN, J.E. AND WALKER, S.G. (2011). Slice sampling mixture models. *Stat. Comput.* **21**, 93–105.

MACEACHERN, S.N. (1999). Dependent Nonparametric Processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statist. Assoc., Alexandria, VA.

MACEACHERN, S.N. (2000). Dependent Dirichlet processes. *Tech. Rep.*, Ohio State University.

MENA, R.H. and RUGGIERO, M. (2016). Dynamic density estimation with diffusive Dirichlet mixtures. *Bernoulli* **22**, 901–926.

MENA, R.H., RUGGIERO, M. AND WALKER, S.G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. *J. Statist. Plann. Inf.* **141**, 3217–3230.

PAPASPILIOPOULOS, O. AND ROBERTS, G.O. (2008). Retrospective mcmc for dirichlet process hierarchical models. *Biometrika* **95**, 169–186.

PAPASPILIOPOULOS, O., RUGGIERO, M. and SPANÒ, D. (2016). Conjugacy properties of time-evolving Dirichlet and gamma random measures. *Electron. J. Statist.*, in press.

PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92**, 21–39.

PETROV, L. (2009). Two-parameter family of diffusion processes in the Kingman simplex. *Funct. Anal. Appl.* **43**, 279–296.

PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory and Related Fields* **102**, 145–158.

PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.

PRÜNSTER, I. and RUGGIERO, M. (2013). A bayesian nonparametric approach to modeling market share dynamics. *Bernoulli* **19**, 64–92.

RODRIGUEZ, A. AND TER HORST, E. (2008). Bayesian dynamic density estimation. *Bayes. Anal.* 3, 339366.

RUGGIERO, M. AND WALKER, S.G. (2009a). Bayesian nonparametric construction of the Fleming–Viot process with fertility selection. *Statist. Sinica* **19**, 707–720.

RUGGIERO, M. AND WALKER, S.G. (2009b). Countable representation for infinite-dimensional diffusions derived from the two-parameter Poisson–Dirichlet process. *Electron. Comm. Probab.* **14**, 501–517.

RUGGIERO, M., WALKER, S.G. AND FAVARO, S. (2013). Alpha-diversity processes and normalized inverse-Gaussian diffusions. *Ann. Appl. Probab.* **23**, 386–425.

SETHURAMAN, J. (1994). A constructive definition of the Dirichlet process prior. *Statist. Sinica* **2**, 639–650.

WALKER, S. (2007). Sampling the dirichlet mixture model with slices. *Comm. Statist. Sim. Comput.* **36**, 45–54.