On time-dependent Gibbs-type random probability measures

Federico Camerlenghi[*]   Igor Prünster[†]   Matteo Ruggiero[‡]

Abstract

We review some recent constructions of time-dependent random probability measures of Gibbs type which exhibit diffusive behaviour. The characterization of the dynamics of the type frequencies, and those of the type heterogeneity in the underlying population, also known as alpha diversity, allows for a qualitative classification of these models according to whether the heterogeneity is driven by state-dependent quantities and whether the frequencies dynamics are directly affected by the overall heterogeneity.

Key Words: Scaling limits, Moran model, Dependent nonparametric model, Diffusion process, Bayesian Nonparametrics, Generalized Polya urn.

1. Introduction

Gibbs-type priors, introduced in Gnedin and Pitman (2005), represent a large class of discrete random probability measures whose laws act as nonparametric priors for Bayesian inference. They can be characterized in terms of the predictive distributions on the observables they induce (see Lijoi, Mena and Prünster (2007a,b)). These are of the form

$$\mathbb{P}\left(X_{n+1} \in \cdot \mid X_1, \ldots, X_n\right) = \frac{V_{n+1,k+1}}{V_{n,k}} P^*(\cdot) + \frac{V_{n+1,k}}{V_{n,k}} \sum_{i=1}^{k} (n_i - \alpha)\, \delta_{X_i^*}(\cdot). \quad (1)$$

with $\alpha < 1$ and $(X_1^*, \ldots, X_k^*)$ denoting the $k \leq n$ distinct values observed in the sample $(X_1, \ldots, X_n)$, each X_i^* having multiplicity n_i. Moreover, the set of non–negative weights $\{V_{n,k} : n \geq 1, 1 \leq k \leq n\}$ satisfies the forward recursive equation

$$V_{n,k} = (n - \alpha k)V_{n+1,k} + V_{n+1,k+1} \quad (2)$$

for any $k = 1, \ldots, n$ and $n \geq 1$, with $V_{1,1} = 1$. Assuming P^* is a non–atomic probability measure on some space $\mathbb{X}$, a natural interpretation for (1) is that $V_{n+1,k+1}/V_{n,k}$ is the probability of observing a new distinct value not included in the sample $X_1, \ldots, X_n$, whereas the second summand describes the probability of observing a replicate of an already seen value. Relevant special cases of the above predictive structure include the Pitman–Yor process (Pitman and Yor, 1997), for which

$$\frac{V_{n+1,k+1}}{V_{n,k}} = \frac{\theta + \alpha k}{\theta + n}, \qquad \frac{V_{n+1,k}}{V_{n,k}} = \frac{1}{\theta + n}; \quad (3)$$

the Dirichlet process (Ferguson, 1973), obtained by setting $\alpha = 0$ in (3); the normalized generalized gamma processes (Lijoi, Mena and Prünster, 2007a), which arises by setting

$$\frac{V_{n+1,k+1}}{V_{n,k}} = \frac{\alpha}{n} \frac{\sum_{i=0}^{n} \binom{n}{i}(-1)^i \beta^{i/\alpha} \Gamma(k_n + 1 - i/\alpha, \beta)}{\sum_{i=0}^{n-1} \binom{n-1}{i}(-1)^i \beta^{i/\alpha} \Gamma(k_n - i/\alpha, \beta)},$$

$$\frac{V_{n+1,k}}{V_{n,k}} = \frac{1}{n} \frac{\sum_{i=0}^{n} \binom{n}{i}(-1)^i \beta^{i/\alpha} \Gamma(k_n - i/\alpha, \beta)}{\sum_{i=0}^{n-1} \binom{n-1}{i}(-1)^i \beta^{i/\alpha} \Gamma(k_n - i/\alpha, \beta)},$$

$$(4)$$

[*]Department of Decision Sciences, Bocconi University

[†]Department of Decision Sciences and BIDSA, Bocconi University

[‡]University of Torino and Collegio Carlo Alberto

for $\beta > 0$ and $\Gamma(a, b)$ being the incomplete gamma function; and the normalized inverse-Gaussian process (Lijoi, Mena and Prünster, 2005), recovered by setting $\alpha = 1/2$ in (4). Applications of Gibbs-type priors include mixture modeling (Ishwaran and James, 2001; Lijoi, Mena and Prünster, 2007a), linguistics and information retrieval (Teh and Jordan, 2010), species sampling (Lijoi, Mena and Prünster, 2007b), Lijoi, Mena and Prünster (2007c), Navarrete, Quintana and Müller (2008) and survival analysis Jara et al. (2010), among others.

Here we discuss some recent extensions of Gibbs-type random probability measures to time-dependent models. A recent strand of literature on Bayesian nonparametric dependent processes in this framework has been initiated by the proposal of the *dependent Dirichlet process* in MacEachern (1999, 2000), a family of random probability measures indexed by a covariate, whereby the observations are exchangeable only conditionally on the covariate value. See Hjort, Holmes, Müller and Walker (2010) and Müller and Mitra (2013) for a recent account on the state of the art on this line of research. In the Dirichlet case, a considerable research production has been favored by the availability of the so-called stick-breaking representation, due to Sethuraman (1994), which has proven to be a valuable and versatile instrument from a practical point of view. Stick-breaking representations are also available for the Pitman–Yor process (Pitman, 1995), for the normalized generalized gamma process with a completely explicit representation for the normalized inverse-Gaussian process (Favaro, Lijoi and Prünster, 2012), and for homogeneous normalized random measures with independent increments (Favaro et al., 2016).

In the following we review some recent results on certain classes of temporally dependent Gibbs-type random probability measures. Such constructions are based on first principles, without relying on stick-breaking representations by means of which one can impose virtually any dynamic feature, but will rather be defined in terms of interaction of the underlying vector of observables. This is somehow similar in spirit to the recent random measure approaches to the definition and study of dependent nonparametric priors in Lijoi, Nipoti and Prünster (2014); Camerlenghi et al. (2016). In order to obtain well-defined scaling limits, in the sense explained below, for the dynamics of the frequencies, one must condition on a latent environment for dynamic random measures of Gibbs-type which are not in the Dirichlet or Pitman–Yor classes. Such constructions have a somewhat direct interpretation in population genetics, but can also be of interest for Bayesian nonparametric inference in a framework of partial exchangeability.

2. Dynamic models for Gibbs-type priors

2.1 A general updating scheme

We concentrate on the case of Gibbs-type random probability measures with parameter $0 \leq \alpha < 1$, assumed to hold throughout. A simple and general recipe for generating discrete time dynamics for an arbitrary, but fixed, number of observations $X_1, \ldots, X_n$ can be obtained by exploiting the predictive scheme (1). Given the state $X_1, \ldots, X_n$ at time t, remove a uniformly chosen element, X_i say, and sample a replacement X_i' from

$$\mathbb{P}(X_i' \in \cdot \mid X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \tag{5}$$

$$= \frac{V_{n,k_{(-i)}+1}}{V_{n-1,k_{(-i)}}} P^*(\cdot) + \frac{V_{n,k_{(-i)}}}{V_{n-1,k_{(-i)}}} \sum_{j=1}^{k_{(-i)}} (n_j - \alpha) \, \delta_{X_j^*}(\cdot).$$

Here $k_{(-i)}$ is the number of distinct observations in the sample after removing X_i. By exchangeability, it can be easily seen that the resulting vector $X_1, \ldots, X_{i-1}, X_i', X_{i+1},$

$\ldots, X_n$ at time $t + 1$ has the same marginal distribution as the starting one. The iteration of the above scheme generates a discrete time Markov chain on $\mathbb{X}^n$, denoted as $X^{(n)}(\cdot) := \{X^{(n)}(m) : m \in \mathbb{N}\}$, whose marginals states are samples from a Gibbs-type random probability measure. This can also be seen as obtained by performing a random scan Gibbs sampler on $X_1, \ldots, X_n$, with (5) taking the role of the full conditional distribution, which immediately implies stationarity and reversibility of the Markov chain. This strategy has been exploited in Ruggiero and Walker (2009a,b); Favaro, Ruggiero and Walker (2009) for constructing infinite-dimensional dynamic processes associated with Dirichlet and Poisson–Dirichlet related random probability measures. For example in Ruggiero and Walker (2009b), the chain $X^{(n)}(\cdot)$ is embedded in continuous time by superimposing it to a Poisson process of intensity λ_n, which governs the n-dependent exponentially distributed time laps between successive updates of the chain.

An interesting question is which requirements are needed in order to yield asymptotically continuous trajectories for the dynamics of the frequencies of types induced by the above scheme applied to a certain Gibbs-type random random probability measure. These are of extreme interest in population genetics, where it is fundamental to approximate intractable dynamic objects with diffusions (see, e.g., Etheridge, 2009). In Bayesian nonparametric inference such objects can be seen as describing a dependent family of random probability measures with continuous trajectories which are useful in a framework of partial exchangeability with data collected at (non necessarily equally spaced) discrete time points; see, e.g., Mena, Ruggiero and Walker (2011); Mena and Ruggiero (2016).

Let $p_i(t)$ be the $[0, 1]$-valued process which describes the evolution of the frequency of the i-th observed distinct type, induced asymptotically, as $n \to \infty$, by the scheme outlined above. The task is then to derive infinitesimal dynamics such that the time-dependent vector of frequencies $p(t) = (p_1(t), p_2(t), \ldots)$ has coordinatewise continuous trajectories and the Gibbs-type law is preserved at the margin, for any t. Being able to characterize such an object allows, by the simple association of *iid* locations $Y_i \sim P^*$ to the frequencies, to provide a time-dependent extension of a discrete random probability measure to a dynamic model with continuous sample-paths. How to characterize such behaviour depends on the analytical tractability of the weights $V_{n,k}$ and poses some challenging technical issues. In the following we summarize results contained in Ethier and Kurtz (1981); Petrov (2009); Ruggiero, Walker and Favaro (2013); Ruggiero (2014); Ruggiero and Sordello (2016) which can be interpreted under this framework.

2.2 Clustering dynamics

Before presenting the results on the frequencies dynamics, we discuss some aspects of what can be called the *clustering dynamics*, or the *dynamic alpha diversity* associated to the systems considered.

A structurally relevant quantity for Gibbs-type random probability measures is the asymptotic behaviour of the number of distinct types observed in the sample generated by (1) as the sample size n increases. Let this quantity be K_n, and define the alpha diversity of the model as the random variable S on $(0, \infty)$ such that

$$\lim_{n \to \infty} \frac{K_n}{n^\alpha} = S \qquad \text{a.s.;} \qquad (6)$$

cf. Pitman (2006), Definition 3.10. This implicitly characterizes the heterogeneity in the population, if we interpret the observed random variables as species. A dynamic extension of (6) turns out to be crucial for characterizing the asymptotic dynamics of the frequencies induced via (5), since it carries up-to-date information on the species heterogeneity in the population and characterizes the environment in which the population evolves. In the
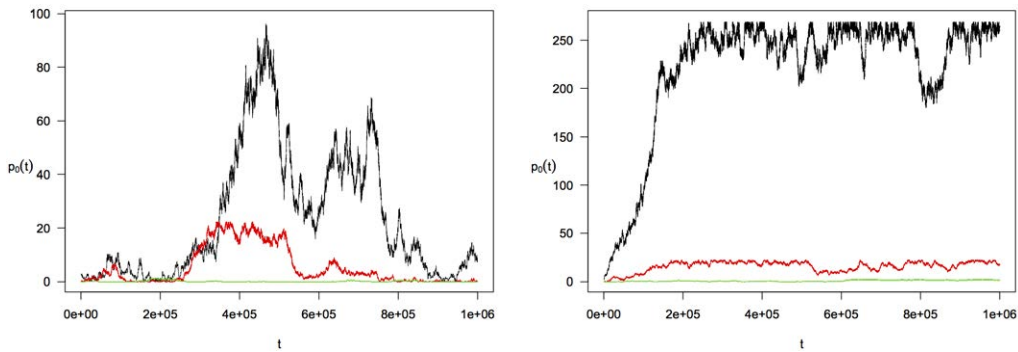
Figure 1: Some approximate sample paths of (7) with $n = 500$ and for values $\theta = .1$ (left) and $\theta = 1$ (right), and $\alpha = 0.1$ (black), 0.5 (red), 0.9 (green). As this is an approximation with finite n, for some parameter values the trajectory can touch the boundary $n^{1-\alpha}$.

Dirichlet case ($\alpha = 0$), the meaningful equivalent of (6) is given by the fact that $K_n / \log n$ converges to a constant (Korwar and Hollander, 1973).

In the dynamic case, one can derive an inhomogeneous Markov chain $K_n(m)$ on the integers for the number of distinct types, and study the asymptotic behaviour of this in terms of scaling limits, or from a different viewpoint in terms of weak convergence of the induced probability measures on the space of sample paths. In the Dirichlet case it is easy to show that the dynamic version $K_n(m) / \log n$ converges to a constant process. A consequence is that the asymptotic result for the frequencies dynamics is parametrized by this constant limit but the dynamics can be described independently of the current heterogeneity in the population, which is somehow degenerate. In order to be consistent with the notation of the next section, we will denote by p_0 the appropriately rescaled limit of K_n, and by $p_0(t)$ its dynamic counterpart.

Consider the Pitman–Yor case and rescale the space of the Markov chain $K_n(m)$ by n^α and time by $n^{1+\alpha}$ yielding the continuous–time process $\{K_n(\lfloor n^{1+\alpha}t \rfloor)/n^\alpha\}_{t \geq 0}$, where $\lfloor \cdot \rfloor$ denotes the floor function. One may prove that the distributional limit of such a process is a well defined diffusion on $\mathbb{R}^+$, which is the solution of

$$\mathrm{d}p_0(t) = \theta \mathrm{d}t + \sqrt{2\alpha p_0(t)}\mathrm{d}B(t), \qquad p_0(t) \geq 0, \tag{7}$$

where $B(t)$ is a standard Brownian motion. The interpretation of a trajectory of such processes is that in the periods of time when $p_0(t)$ is relatively high, there is a relatively high number of small clusters in the population, and viceversa. Note however that the height of the path depends on the parameter values, as for example α close to 1 will produce paths near the origin even when K_n is close to n. Figure 1 shows some approximate sample paths of (7), for different values of α and θ.

A similar space-time transformation for $K_n(m)$ yields a different scaling limit in the normalized generalized gamma case. Specifically, we have

$$\mathrm{d}p_0(t) = (\beta/p_0^{1/\alpha})\mathrm{d}t + \sqrt{2\alpha p_0(t)}\mathrm{d}B(t), \qquad p_0(t) \geq 0, \tag{8}$$

with $\beta > 0$ as in (4). This has been recently shown in Ruggiero and Sordello (2016) (and supersedes Theorem 4.1 in Ruggiero, Walker and Favaro, 2013).

Note that the volatility is lower in correspondence of low values of the process and vice versa, as implied by the diffusion coefficient in (7) and (8). Furthermore, the origin acts as an entrance boundary, meaning that the process never reaches zero (except if it is made to
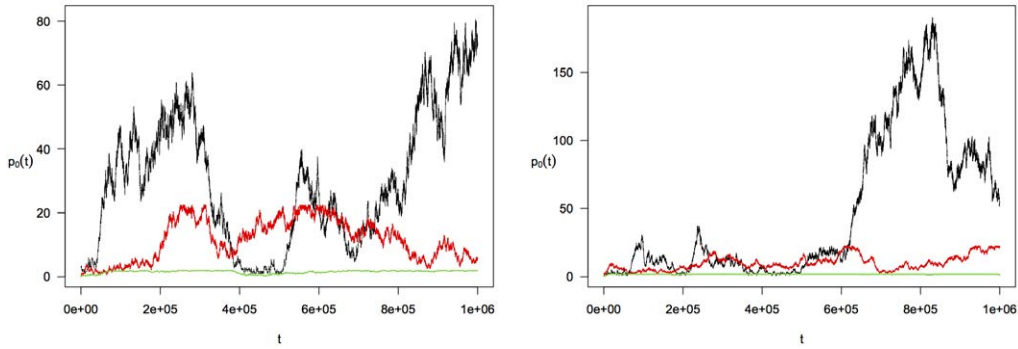
Figure 2: Some approximate sample paths of (8) with $n = 500$ and for values $\beta = 1$ (left) and $\beta = 10$ (right), and $\alpha = 0.1$ (green), 0.5 (red), 0.9 (black).

start from there). This is due to the fact that when $p_0(t)$ approaches zero, the volatility also drops to zero, and the positive drift pushes the process back towards the interior of the state space. Figure 2 shows some approximate sample path examples of (8) for different values of α and β.

2.3 Frequencies dynamics

Let now $p(\cdot) = \{p_i(t) : i \in \mathbb{Z}_+, t \geq 0\}$ denote the infinite vector of dynamic asymptotic frequencies obtained according to the scheme of Section 2.1 under appropriate transformations. We will now denote by $(a_{i,j}(p))_{i,j \geq 0}$ the infinitesimal covariances between the ith and jth coordinates of p, and by $(b_i(p))_{i \geq 0}$ the associated infinitesimal means of the ith coordinate. Index by $h = 1, 2, 3$ the Dirichlet related model, the Pitman–Yor related model and the normalized inverse-Gaussian related model, respectively, that is the dynamic models derived as described in Section 2.1 when using the relative urn scheme (5) for updating the vector. Then the above quantities equal

$$
a_{ij}^{(h)}(p) := \begin{cases} 2\alpha p_0, & i = j = 0, \\ p_i(\delta_{ij} - p_j), & i, j \geq 1, \\ 0, & \text{else,} \end{cases} \qquad b_i^{(h)}(p) := -[b_0^{(h)}(p)p_i + \alpha], \quad i \geq 1.
$$

(9)

for all h, where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, and

$$
b_0^{(h)}(p) = \begin{cases} 0, & h = 1, \\ \theta, & h = 2, \\ \beta/p_0^{1/\alpha}, & h = 3. \end{cases}
$$

The three models presented above are based on decreasingly ordered Dirichlet, Pitman–Yor and normalized inverse-Gaussian sample frequencies. It is important to emphasize that the identification of the frequencies dynamics in the ordered case is also sufficient for characterizing the temporal model associated to each of the three mentioned random probability measures. This is due to the invariance under size–biased permutation of Gibbs-type priors.

The above coefficients characterize an infinite dimensional stochastic differential equation that describes the dynamics of infinitely-many frequencies, indexed by $i \geq 1$, and a positive component, indexed by $i = 0$. The frequencies are the relative abundances of

the different species in a large population which marginally is of the specified Gibbs type, whereas the 0th coordinate is the dynamic alpha diversity described in Section 2.2 for each case. The state space of the joint processes is $[0, \infty) \times \overline{\nabla}_\infty$, where $[0, \infty)$ is the state space of the dynamic heterogeneity and the vector of frequencies takes values in

$$\overline{\nabla}_\infty = \left\{ p = (p_1, p_2, \ldots) : p_1 \geq p_2 \geq \ldots \geq 0, \ \sum_{i=1}^\infty p_i \leq 1 \right\};$$

(vectors summing to less than one are included for technical reasons omitted here; see, e.g., Costantini et al., 2016). In the first model ($h = 1$) the dynamic heterogeneity is trivial since $\alpha = 0$; cf. the Dirichlet case in Section 2.2. In the second model ($h = 2$), the diversity process provides insight into the model properties in terms of how the population heterogeneity evolves, but acts independently of the frequencies. In the third model ($h = 3$), all components must be considered jointly as the frequencies dynamics depend on the current value of the diversity process.

For what concerns the coefficient describing the frequencies dynamics, the first model identifies the infinitesimal dynamics which preserve the marginal distributions of the Dirichlet weights, as follows:

- the expected increment or drift of the ith component $p_i(t)$ equals $-\theta p_i$,
- the variance of the ith component equals $p_i(1 - p_i)$,
- the covariance between ith and jth component equals $-p_i p_j$.

The covariance structure is reminiscent of a multinomial sampling scheme, which is indeed at the base of Wright–Fisher type constructions of such models from finitely-many components with finitely-many species. See Ethier and Kurtz (1981). The interpretation of the parameter θ in the drift is not very dissimilar from that of the Dirichlet process: in this setting it regulates how often (in time, instead of along the sampling sequence) new species or types appear in the population. In mathematical biology θ is interpreted as overall mutation rate, that is the rate at which mutation events occur in the population, and the negative sign of the drift reflects the fact that mutations of individuals of type i to some other type decrease the frequency p_i.

The second model, associated to the Pitman–Yor process, is analogous to the Dirichlet process case except for the drift, which is $-(\theta p_i + \alpha)$. Here the additional parameter α introduces a further negative effect on the expected increment of each frequency. This is related to the reinforcement effect of α in the generalized Pólya urn scheme for Gibbs-type models. See De Blasi et al. (2015) for details. In mathematical biology the interpretation of α in this model is to a certain extent still unclear, but some recent insight into this problem has been provided by Costantini et al. (2016).

A distinctive feature of the third model, associated to normalized inverse-Gaussian processes with parameter $(\beta, 1/2)$, is the dependence of the ith component's drift on the component p_0, since it coincides with $-[(\beta/p_0^2)p_i + 1/2]$. Although it appears as structurally analogous to the previous cases, here the role of the alpha diversity is crucial, since $p_0(t)$ is itself a diffusion which drives the time-dependent overall mutation rate $\theta(t) := \beta/p_0^2(t)$.

2.4 Concluding remarks

The summarized results expose a structural difference between the first two and the third model. In the Dirichlet and Pitman–Yor related models, the (time-dependent) distributional properties of the distinct number of components do not directly affect the mutation rate of the population, which is instead the case in the normalized inverse-Gaussian related model. In the latter case the chances single individuals have to mutate vary over time, creating a randomly changing environment to which the population evolution is subject to. This

suggests the following qualitative classification of the three models, according to the role of the dynamic heterogeneity:

- Dirichlet related diffusion: heterogeneity driven by constant quantities, frequencies dynamics not directly affected by the heterogeneity;

- Pitman–Yor related diffusion: heterogeneity driven by state-dependent quantities, frequencies dynamics not directly affected by the heterogeneity;

- normalized inverse-Gaussian related diffusion: heterogeneity driven by state-dependent quantities, frequencies dynamics directly affected by the heterogeneity.

The normalized generalized gamma case has not been included in the above classification nor in (9), as to the present date only the dynamics for the alpha diversity have been characterized. Given that normalized inverse-Gaussian models are included in the normalized generalized gamma class, it is however reasonable to conjecture that the structure of the problem is also the same in the dynamic case, which would yield dynamics for p_0 as in (8) and for the frequencies as in (9), analogously to the case $h = 3$, but for general $\alpha \in (0, 1)$.

Acknowledgments

REFERENCES

CAMERLENGHI, F., LIJOI, A., ORBANZ and PRÜNSTER, I. (2016). Distribution theory for hierarchical processes. *Submitted.*

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.

DE BLASI, P., FAVARO, S., LIJOI, A., MENA, R.H., RUGGIERO, M. and PRÜNSTER, I. (2015). Are Gibbs–type priors the most natural generalization of the Dirichlet process? *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 212–229.

ETHERIDGE, A.M. (2009). *Some mathematical models from population genetics.* École d'été de Probabilités de Saint-Flour XXXIX. Lecture Notes in Math. **2012**. Springer.

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

FAVARO, S., LIJOI, A., NAVA, C., NIPOTI, B., PRÜNSTER, I. and TEH, Y.W. (2016). On the stick-breaking representation for homogeneous NRMIs. *Bayesian Analysis* **11**, 697-724.

FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2012). On the stick-breaking representation of normalized inverse Gaussian priors. *Biometrika* **99**,663–674.

FAVARO, S., RUGGIERO, M. AND WALKER, S.G. (2009). On a Gibbs sampler based random process in Bayesian nonparametrics. *Electron. J. Statist.* **3**, 1556–1566.

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

GNEDIN, A. and PITMAN, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **325**, 83–102.

HJORT, N.L., HOLMES, C.C., MÜLLER, P. and WALKER, S.G., EDS. (2003). *Bayesian nonparametrics.* Cambridge Univ. Press, Cambridge.

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

JARA, A., LESAFFRE, E., DE IORIO, M. and QUINTANA, F.A. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Ann. Appl. Statist.* **4**, 2126–2149.

KORWAR, R.M. AND HOLLANDER, M. (1973). Contribution to the theory of Dirichlet processes. *Ann. Probab.* **1**, 705–711.

LIJOI, A., MENA, R.H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-gaussian priors. *J. Amer. Statist. Assoc.* **472**, 1278–1291.

LIJOI, A., MENA, R.H. and PRÜNSTER, I. (2007a). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Statist. Soc. Ser. B* **69**, 715–740.

LIJOI, A., MENA, R.H. AND PRÜNSTER, I. (2007b). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94**, 769–786.

LIJOI, A., MENA, R.H. AND PRÜNSTER, I. (2007c). A Bayesian nonparametric method for prediction in EST analysis. *BMC Bioinformatics* **8**, article 339.

LIJOI, A., NIPOTI, B. AND PRÜNSTER, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260–1291.

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

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

MÜLLER, P. and MITRA, R. (2013). Bayesian nonparametric inference - why and how. *Bayesian Anal.* **8**, 269–302.

NAVARRETE, C., QUINTANA, F.A. and MÜLLER, P.. (2008). Some issues on nonparametric Bayesian modeling using species sampling models. *Statist. Modell.* **41**, 3–21.

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 Relat. Fields* **102**, 145–158.

PITMAN, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Springer, Berlin.

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

RUGGIERO, M. (2014). Species dynamics in the two-parameter Poisson–Dirichlet diffusion model. *J. Appl. Probab.* **51**, 174–190.

RUGGIERO, M. AND SORDELLO, M. (2016). Clustering dynamics in a class of normalized generalized gamma dependent priors. *Ann. Inst. Statist. Math.*, to appear.

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.

TEH, Y.W. and JORDAN, M.I. (2010). Bayesian nonparametrics in machine learning. In Bayesian Nonparametrics (N. L. Hjort, C. C. Holmes, P. Mller and S. G. Walker, eds.). Cambridge Univ. Press, Cambridge.