Ann Inst Stat Math (2018) 70:83–98
https://doi.org/10.1007/s10463-016-0583-8

Clustering dynamics in a class of normalised generalised gamma dependent priors

Matteo Ruggiero[1,2] · **Matteo Sordello**[3]

Received: 27 November 2015 / Revised: 2 August 2016 / Published online: 20 October 2016
© The Institute of Statistical Mathematics, Tokyo 2016

Abstract Normalised generalised gamma processes are random probability measures that induce nonparametric prior distributions widely used in Bayesian statistics, particularly for mixture modelling. We construct a class of dependent normalised generalised gamma priors induced by a stationary population model of Moran type, which exploits a generalised Pólya urn scheme associated with the prior. We study the asymptotic scaling for the dynamics of the number of clusters in the sample, which in turn provides a dynamic measure of diversity in the underlying population. The limit is formalised to be a positive non-stationary diffusion process which falls outside well-known families, with unbounded drift and an entrance boundary at the origin. We also introduce a new class of stationary positive diffusions, whose invariant measures are explicit and have power law tails, which approximate weakly the scaling limit.

Keywords Alpha diversity · Bayesian nonparametrics · Dependent process · Diffusion process · Generalised Pólya urn · Moran model · Scaling limit

The first author is supported by the European Research Council (ERC) through StG "N-BNP" 306406.

✉ Matteo Ruggiero
matteo.ruggiero@unito.it

Matteo Sordello
sordello@wharton.upenn.edu

1 University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy

2 Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, TO, Italy

3 The Wharton School, University of Pennsylvania, 3730 Walnut St, Philadelphia, PA 19104, USA

1 Introduction

A key aspect in Bayesian nonparametric inference, both from a methodological and a computational point of view, is the clustering of the observations. Regardless of whether these represent real quantities of interest or latent features used in intermediate levels of hierarchies, an often important inferential issue is the estimation of the number of components underlying the mixture. A Bayesian nonparametric mixture model typically takes the form

$$f(y) = \int f(y \mid x) P(\mathrm{d}x) \tag{1}$$

where $f(\cdot|x)$ is a density function for every value of x, and the latent quantity x is modelled through a random probability measure P. When P is a Dirichlet process (Ferguson 1973), (1) is the Dirichlet process mixture model introduced by Lo (1984), which is to date the most popular Bayesian nonparametric approach. The mixture (1) can be equivalently expressed in hierarchical form by writing

$$\begin{aligned} &P \sim \mathscr{P} \\ &X_i \mid P \overset{iid}{\sim} P \\ &Y_i \mid X_i \overset{ind}{\sim} f(\cdot \mid X_i), \end{aligned} \tag{2}$$

where $\mathscr{P}$ is the prior induced by the random probability measure P on the set of distributions over the latent space. If $\mathscr{P}$ selects almost surely discrete probability measures, as is the case when P is a Dirichlet process (Blackwell 1973), then the latent variables X_i will feature ties and can be used to cluster the observations Y_i according to the kernel $f(\cdot|X_i)$ from which they are generated. For this reason, the number of distinct values $K_n \leq n$ in the exchangeable sequence $X_1, \ldots, X_n$ is sometimes loosely referred to as the number of clusters.

The implications on inference of the clustering structures induced by the Dirichlet process, which behave as logarithmic functions of the number of observations (Korwar and Hollander 1973), have long been the object of extensive investigation. Recent advances in the field have drawn attention to different clustering behaviours, such as those induced by Pitman–Yor processes (Pitman 1995; Pitman and Yor 1997), normalised inverse-Gaussian processes (Lijoi et al. 2005) and normalised generalised gamma processes (Lijoi et al. 2007). Despite the increased generality, these priors stand out for their tractability among the various generalisations of the Dirichlet process, and contrast with the latter by inducing clustering structures which behave as power functions of the number of observation. See the study by Lijoi and Prünster (2010).

Another recent, significant line of research in Bayesian nonparametrics aims at extending nonparametric priors to accommodate forms of dependence more general than exchangeability. MacEachern (1999, 2000) proposed a class of so-called *dependent processes* for modelling partially exchangeable sequences, where observations are exchangeable conditional on a given set of covariates, but not overall exchangeable. These are modelled through a collection of random probability measures with series representation of the form

$$P_z = \sum_{i \geq 1} p_{i,z} \delta_{x_{i,z}},$$

where the weights $p_{i,z}$ and/or the atoms $x_{i,z}$ depend on some covariate z, which can be multidimensional or possibly represent time. See the study by Hjort et al. (2010) for a review and for recent developments. In particular, the discrete nature of these dependent priors and their wide applicability to concrete problems call for new understanding of how the clustering structure depends on the covariate, which is in turn induced by the type of dependence used for defining $p_{i,z}$ and $x_{i,z}$.

In this paper, we construct a new class of temporally dependent priors which are induced by a normalised generalised gamma population model, and investigate the scaling limit for the dynamics of the number of groups or clusters. Other classes of dependent normalised random measures have been constructed recently by Griffin et al. (2013), Lijoi et al. (2014) and Griffin and Leisen (2016). Here, by taking a different approach, the construction embeds normalised generalised gamma priors in a temporal environment. We define a dynamic population model of Moran type (see Section. 3 for details), which can also be seen as the iteration of Gibbs sampling steps at the level of the latent variables in the hierarchy. We study the rescaling of the induced number of groups in the population and identify the limit to be a positive, nonstationary diffusion process which seems to fall outside well-known classes. As the use of stationary components in dependent, the hierarchical model is often desirable and we formulate a weak approximation of the scaling limit by introducing a new family of stationary and positive diffusions whose invariant measure is explicit and exhibits a power law right tail.

Scaling limits of the clustering dynamics for other classes of dependent models connected with Bayesian nonparametrics have been studied in Ruggiero et al. (2013); Ruggiero (2014) for the normalised inverse Gaussian and the two-parameter Poisson–Dirichlet case, respectively. See also Ruggiero and Walker (2009a, b); Mena et al. (2011); Mena and Ruggiero (2016); Papaspiliopoulos et al. (2016) for different dependent models connected with diffusions processes.

2 Preliminaries on normalised generalised gamma priors

Generalised gamma processes, introduced by Brix (1999), are completely random measures with generalised gamma mean intensity, that is Levy processes with positive jumps and Levy measure on $[0, \infty)$ given by

$$\lambda(\mathrm{d}t) = \frac{e^{-\tau t} t^{-(1+\alpha)}}{\Gamma(1-\alpha)} \mathrm{d}t,$$

with $\alpha \in (0, 1)$ and $\tau \geq 0$. Lijoi et al. (2007) exploited this construction for proposing a prior distribution for Bayesian nonparametric mixture modelling. This is obtained by normalising the jumps of a generalised gamma process via

$$p_i = \frac{J_i}{\sum_{k=1}^{\infty} J_k},$$

where J_i are the jump sizes and $\sum_{k=1}^{\infty} J_k < \infty$ almost surely. The resulting random weights allow to define a discrete random probability measure by writing

$$P = \sum\nolimits_{i \geq 1} p_i \delta_{Z_i},$$

where $Z_i \overset{iid}{\sim} P_0$ and P_0 is a nonatomic probability measure on a Polish space $\mathbb{X}$. The resulting normalised generalised gamma random probability measure induces a prior distribution on the space of discrete laws on $\mathbb{X}$, denoted here $\mathrm{GG}(\beta, \alpha)$ for $\beta = \tau^{\alpha}/\alpha$. This can then be used at the top level of the hierarchy for Bayesian nonparametric modelling, replacing $\mathscr{P}$ with $\mathrm{GG}(\beta, \alpha)$ in (2).

Denote by K_n the number of distinct values $X_1^*, \ldots, X_{K_n}^*$ observed in a sample $X_1, \ldots, X_n$ with $X_i \mid P \overset{iid}{\sim} P$. When $P \sim \mathrm{GG}(\beta, \alpha)$, Lijoi et al. (2007) showed that

$$\mathbb{P}(K_n = k) = \frac{e^{\beta} \mathcal{G}(n, k, \alpha)}{\alpha \Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\alpha} \Gamma(k - i/\alpha; \beta), \tag{3}$$

where $\Gamma(a; x)$ is the incomplete gamma function, $\mathcal{G}(n, k, \alpha)$ is the generalised factorial coefficient (see Charalambides (2005))

$$\mathcal{G}(n, k, \alpha) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-j\alpha)_n$$

and $(a)_n = a(a+1) \cdots (a+n-1)$ is the increasing factorial. Furthermore, K_n grows as n^{α} and

$$\lim_{n \to \infty} \frac{K_n}{n^{\alpha}} = S \qquad \text{a.s.}, \tag{4}$$

where S is a random variable on $(0, \infty)$ with density

$$g_{\beta,\alpha}(s) = \exp\{\beta - (\beta/s)^{1/\alpha}\} \alpha^{-1} s^{-1-1/\alpha} f_{\alpha}(s^{-1/\alpha}),$$

and where f_{α} is the density of a positive stable random variable of index α. Since S summarises the asymptotic diversity in terms of number of groups which grows as a power function of α, the partition associated with generalised gamma priors is said to have α-diversity S. Cf. Pitman (2006), Definition 3.10.

Normalised generalised gamma priors belong to the larger class of Gibbs-type priors (Gnedin and Pitman 2006; De Blasi et al. 2015). These can be characterised, among other ways, in terms of the marginal law of the observations, which is given by a generalised Pólya urn scheme. More specifically, conditionally on $K_n = k_n$, the predictive distribution for the observations associated with Gibbs-type models is given by the following generalised Pólya urn scheme:

$$X_{n+1} | X_1, \ldots, X_n \sim g_0(n, k_n) P_0(\cdot) + g_1(n, k_n) \sum_{j=1}^{k_n} (n_j - \alpha) \delta_{X_j^*}(\cdot). \tag{5}$$

Here, P_0 is as above, and the weights $g_0(n, k_n)$, $g_1(n, k_n)$, possibly dependent on other fixed parameters that characterise the specific model, satisfy

$$g_0(n, k_n) + (n - \alpha k_n) g_1(n, k_n) = 1$$

for all $n \geq 1$ and $1 \leq k_n \leq n$. The interpretation of (5) is that $g_0(n, k_n)$ is the probability of sampling a previously unobserved value, and $g_1(n, k_n)(n_j - \alpha)$ is the probability of duplicating the distinct value X_j^*, thus enlarging the associated group by one unit. Lijoi et al. (2007) showed that in the normalised generalised gamma case, we have

$$\begin{aligned} g_0(n, k_n) &= \frac{\alpha}{n} \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i \beta^{i/\alpha} \Gamma(k_n + 1 - i/\alpha, \beta)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\alpha} \Gamma(k_n - i/\alpha, \beta)}, \\ g_1(n, k_n) &= \frac{1}{n} \frac{\sum_{i=0}^{n} \binom{n}{i} (-1)^i \beta^{i/\alpha} \Gamma(k_n - i/\alpha, \beta)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\alpha} \Gamma(k_n - i/\alpha, \beta)}. \end{aligned} \tag{6}$$

for $\alpha \in (0, 1)$ and β as above. Furthermore, the generalised gamma model is the only normalised completely random measure that is also of Gibbs type. See Proposition 2 in Lijoi et al. (2008). The Pitman–Yor process is also a member of the Gibbs family, in which case these quantities simplify to

$$g_0(n, k_n) = \frac{\theta + \alpha k_n}{\theta + n}, \qquad g_1(n, k_n) = \frac{1}{\theta + n} \tag{7}$$

for either $\alpha < 0$ and $\theta = |\alpha| m$, $m \in \mathbb{N}$, or

$$\alpha \in [0, 1), \qquad \theta > -\alpha. \tag{8}$$

In this paper, we aim at studying a dynamic version of the α-diversity asymptotic result (4) for the generalised gamma model, after appropriately extending the distribution (3) of the number of groups to a temporal framework, through the definition of a population model based on (5), (6). To this end, we will make use of a recent result by Arbel et al. (2016), who extend a result contained in the study by Ruggiero et al. (2013). In particular, they derive the second-order approximation of (6) to be

$$\begin{aligned} g_0(n, k_n) &= \frac{\alpha k_n}{n} + \frac{\beta}{k_n^{1/\alpha}} + o(n^{-1}), \\ g_1(n, k_n) &= \frac{1}{n} - \frac{\beta}{n k_n^{1/\alpha}} + o(n^{-2}), \end{aligned} \tag{9}$$

for $n \to \infty$, which allows to avoid, in view of an asymptotic study, a cumbersome computation with alternating sums and incomplete Gamma functions.

3 A generalised gamma population model and its group dynamics

Ruggiero and Walker (2009b) proposed a discrete construction for a class of two-parameter Poisson–Dirichlet diffusion models, introduced in Petrov (2009), based on

the generalised Pólya urn scheme (7). Here, we extend such approach for defining a stationary generalised gamma population model and derive the scaling limit for the dynamics of the number of groups (or species) in the population.

Fix n, and let $X^{(n)} = (X_1, \ldots, X_n)$ be a sample from a generalised gamma model, with $X_1 \sim P_0$ and $X_i | X_1, \ldots, X_{i-1}$ as in (5) for $i = 2, \ldots, n$. We update $X^{(n)}$ at discrete times by substituting a uniformly chosen coordinate of the vector with a replacement from its conditional distribution given the remaining observations. Given the exchangeability of the sample, and assuming we replace X_i, the new element has distribution

$$X_i' | X_{(-i)}^{(n)} \sim g_0(n-1, k_{n,i}) P_0(\cdot) + g_1(n-1, k_{n,i}) \sum_{j=1}^{k_{n,i}} (n_{j,i} - \alpha) \delta_{X_j^*}(\cdot) \tag{10}$$

where $X_{(-i)}^{(n)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ is the remaining sample after removing X_i, $k_{n,i}$ is the number of distinct values in $X_{(-i)}^{(n)}$ and $n_{j,i}$ is the cardinality of the jth cluster after removing X_i. Thus, X_i' is of a new type with probability $g_0(n-1, k_{n,i})$ or a copy of an existing type with probability $(n - 1 - \alpha k_{n,i}) g_1(n-1, k_{n,i})$. In terms of the population model, copying an existing type is interpreted as a birth, whereby the offspring takes the parent type in a haploid population. New types are interpreted as births with mutation, where the mutant type does not depend on the parent type and is drawn from a pool of infinitely many alleles. Removals are interpreted as deaths, which here keep the population size constant. The resulting dynamics are those of a Moran model, which, together with Wright–Fisher models, are among the oldest approaches to mathematical population genetics. See the studies by Etheridge (2009) for background, and Feng (2010) for Moran and Wright–Fisher models applied to infinitely many-allele dynamics, with some connections to Bayesian nonparametrics. See also the work of Costantini et al. (2016) for a recent Wright–Fisher construction of the two-parameter Poisson-Dirichlet diffusion.

Denote the Markov chain resulting from the above-described replacements by $X^{(n)}(\cdot) = \{X^{(n)}(m), m \in \mathbb{N}\}$, and define $K_n(\cdot) = \{K_n(m), m \in \mathbb{N}\}$ to be the process that tracks the number of distinct types in $X^{(n)}$. Note that the dynamics of the Moran chain $X^{(n)}$ are equivalent to running a random scan Gibbs sampler (Smith and Roberts 1993) on the joint distribution of a generalised gamma sample of size n. This implies the following.

Proposition 1 *Let $X^{(n)}(\cdot) = \{X^{(n)}(m), m \in \mathbb{N}\}$ be the Markov chain described above, with transitions determined by replacing a randomly chosen coordinate X_i with a sample from* (10)*. Then $X^{(n)}(\cdot)$ is stationary.*

Proof It follows by adapting the proof of Proposition 4.1 by Ruggiero and Walker (2009b), which does not depend on the specific form of the urn weights, or equivalently by the stationarity of the Markov chain generated by a Gibbs sampler on $X^{(n)}$, given that (10) are the full conditional distributions of the coordinates. □

The stationary distribution of $X^{(n)}(\cdot)$ is clearly the joint law of an n-sized sample from (10). The previous result suggests that the present construction can be natu-

rally embedded in broader Monte Carlo strategies where the distinct values of such observations represent the latent clusters for the data points.

The transition probabilities of $K_n(m)$, denoted

$$p_n(k, k') = \mathbb{P}(K_n(m+1) = k' | K_n(m) = k)$$

can be easily derived from the dynamics of $X^{(n)}$. Denote by $M_{1,n}$ the number of types appearing only once in $X^{(n)}$. Then, the probability of a transition $k \mapsto k+1$ is given by the probability $1 - M_{1,n}/n$ of not removing a group of size 1, times the probability $g_0(n-1, k)$ of sampling a new type as a replacement. Similarly, the probability of a transition $k \mapsto k-1$ is given by the probability $M_{1,n}/n$ of removing a singleton, times the probability $(n-1-\alpha(k-1))g_1(n-1, k-1)$ of duplicating an existing type as a replacement. Such transitions are not Markov, since $M_{1,n}$ carries more information than K_n. Following a similar approach to that by Ruggiero (2014), we can exploit the approximation $M_{1,n} \approx \alpha K_n$, deduced from Corollary 1 by Lijoi et al. (2007), to define

$$p_n(k, k') = \begin{cases} \left(1 - \dfrac{\alpha k}{n}\right) g_0(n-1, k), & k' = k+1, \\ \dfrac{\alpha k}{n}(n-1-\alpha(k-1))g_1(n-1, k-1), & k' = k-1, \\ 1 - p(k, k+1) - p(k, k-1), & k' = k, \\ 0, \quad \text{else}; \end{cases} \tag{11}$$

(note that there is a misprint in eq. (8) of the paper by Ruggiero (2014), which should be as in (11) with g_0, g_1 as in (7), i.e. dropping the small o terms, with obvious modifications to the subsequent proof). Using now (9), we can approximate (11) with

$$p_n(k, k') = \begin{cases} \left(1 - \dfrac{\alpha k}{n}\right)\left(\dfrac{\alpha k}{n-1} + \dfrac{\beta}{k^{1/\alpha}} + o(n^{-1})\right), & k' = k+1, \\ \dfrac{\alpha k}{n}(n-1-\alpha(k-1))\left(\dfrac{1}{n-1} - \dfrac{\beta}{(n-1)(k-1)^{1/\alpha}} + o(n^{-2})\right), & \\ & k' = k-1, \\ 1 - p(k, k+1) - p(k, k-1), & k' = k, \end{cases} \tag{12}$$

and 0 otherwise. Due to the approximation of $g_0(n, k_n)$ and $g_1(n, k_n)$, non-admissible values can arise for certain choices of parameters when k_n is close to the boundary; hence, the probabilities of K_n stepping up or down are intended as $\min(p_n(k, k+1), 1)$ and $\max(p_n(k, k-1), 0)$, respectively. Completed by the boundary conditions $p_n(1, 0) = p_n(n, n+1) = 0$, $K_n(m)$ with transitions (12) is clearly recurrent on $\{1, \ldots, n\}$.

Define now $S(\cdot) = \{S(t), t \geq 0\}$ as the solution of the stochastic differential equation

$$\mathrm{d}S(t) = \frac{\beta}{S(t)^{1/\alpha}}\mathrm{d}t + \sqrt{2\alpha S(t)}\mathrm{d}B(t), \tag{13}$$

where $B(t)$ is a standard Brownian motion. To the best of our knowledge, (13) does not seem to belong to any well-known class of diffusions. We will first show that $S(t)$ above is a well-defined diffusion process on $[0, \infty)$, it has an entrance boundary at 0 and a natural boundary at ∞, and it is non-stationary. An entrance boundary at the origin means that 0 can be the starting point of the process which instantly enters $(0, \infty)$ and never touches the origin again. A natural boundary at ∞ is instead attractive, but never reached. Then, we will show that $S(\cdot)$ is the scaling limit, as $n \to \infty$, of the above-defined sequence of Markov chains after an appropriate space-time transformation.

Proposition 2 *Let $S(\cdot)$ be the solution to (13). Then $S(\cdot)$ is a Feller process, has an entrance boundary at 0 and a natural boundary at ∞, and it is non-stationary.*

Proof Classical Feller theory leads to studying the boundary behaviour of the process by investigating some functionals of the drift and diffusion coefficients that characterise the process. Here, we highlight the relevant quantities and refer to Karlin and Taylor (1981), Section 15.6, for further details (see also Etheridge, 2009, Section 3). Define the scale function

$$Z(x) = \int^x z(y)\mathrm{d}y, \qquad z(x) = \exp\left\{-\int^x \frac{2\mu(y)}{\sigma^2(y)}\mathrm{d}y\right\}$$

and the speed measure

$$M(x) = \int^x m(y)\mathrm{d}y, \qquad m(x) = \frac{1}{\sigma^2(x)z(x)}.$$

A standard calculation leads to find

$$z(x) = e^{\beta x^{-\frac{1}{\alpha}}}, \qquad m(x) = \frac{1}{2\alpha x}e^{-\beta x^{-\frac{1}{\alpha}}}. \tag{14}$$

Letting $Z(a) = \lim_{x\to a} Z(x)$ and similarly for M, for α, β as in (6) it is easy to see that $Z(0) = Z(+\infty) = M(+\infty) = \infty$ and $M(0) < \infty$. Moreover, from

$$\Sigma(x) = \int^x \left(\int_t^x z(y)\mathrm{d}y\right) m(t)\mathrm{d}t, \qquad N(x) = \int^x \left(\int_t^x m(y)\mathrm{d}y\right) z(t)\mathrm{d}t$$

we deduce $\Sigma(0) = \Sigma(+\infty) = N(+\infty) = \infty$ and $N(0) < \infty$. The second assertion now follows from Karlin and Taylor (1981), Section 15.6.

Let now $\hat{C}([0, \infty))$ be the Banach space of continuous functions on $[0, \infty)$ vanishing at infinity. Let also

$$Af(s) = (\beta/s^{1/\alpha})f'(s) + \alpha s f''(s) \tag{15}$$

be the infinitesimal operator corresponding to (13) and define

$$\mathcal{D} = \{f \in C([0, \infty)) \cap C^2((0, \infty)) : Af \in C([0, \infty))\}.$$

Corollary 8.1.1 in Ethier and Kurtz (1986), together with the second assertion, implies that $\{(f, Af) : f \in \mathcal{D} \cap \hat{C}([0, \infty))\}$ generates a Feller semigroup on $\hat{C}([0, \infty))$, which is the first statement.

The proof is completed by the fact that a stationary distribution must take the form

$$\psi(x) = m(x)[C_1 Z(x) + C_2] \tag{16}$$

and the above arguments imply that both constants must vanish. □

Given the boundary properties shown in Proposition 2, it follows that, without loss of generality, we can start $S(\cdot)$ from $(0, \infty)$ and take the latter as the state space of the processes.

The following Theorem, which extends Proposition 3 by Lijoi et al. (2007), shows that (13) is the scaling limit of the sequence of Markov chains $\{K_n\}_{n\geq 1}$ with transitions (12), in the sense that, as $n \to \infty$, the sequence of appropriately transformed chains converges in distribution to $S(\cdot)$. To this end, denote by $Z_n \xrightarrow{d} Z$ convergence in distribution, let $D_A(B)$ be the Skorokhod space of right-continuous functions from A to B with left limits, and $C_A(B)$ its subspace of continuous functions endowed with the topology of uniform convergence. Let also $\lfloor\cdot\rfloor$ be the floor function.

Theorem 1 *Let $K_n(\cdot) = \{K_n(m), m \in \mathbb{N}\}$ be the Markov chain on $\mathbb{N}$ with transition probabilities as in (12), where $\alpha \in (0, 1)$ and $\beta > 0$, and define $\tilde{K}_n(\cdot) = \{\tilde{K}_n(t), t \geq 0\}$ by*

$$\tilde{K}_n(t) = \frac{K_n(\lfloor n^{1+\alpha} t\rfloor)}{n^\alpha}.$$

Let also $S(\cdot)$ be as in (13). If $\tilde{K}_n(0) \xrightarrow{d} S(0)$, then

$$\tilde{K}_n(\cdot) \xrightarrow{d} S(\cdot) \qquad \text{in } C_{[0,\infty)}([0, \infty))$$

as $n \to \infty$.

Proof Let U_n be the semigroup operator induced by (12). Writing n and k in place of $n - 1$ and $k - 1$ for brevity given their asymptotic equivalence, we have

$$\begin{aligned} U_n f(k) &= E[f(K_n(m+1))|K_n(m) = k] \\ &= f(k+1)p_n(k, k+1) + f(k-1)p_n(k, k-1) + f(k)p_n(k, k). \end{aligned}$$

Consider now the spatially rescaled variable and let I denote the identity operator, leading to

$$\begin{aligned} (U_n - I)f\left(\frac{k_n}{n^\alpha}\right) &= E\left[f\left(\frac{K_n(m+1)}{n^\alpha}\right) - f\left(\frac{K_n(m)}{n^\alpha}\right)\Big|K_n(m) = k_n\right] \\ &= \left[f\left(\frac{k_n+1}{n^\alpha}\right) - f\left(\frac{k_n}{n^\alpha}\right)\right] \end{aligned}$$

$$\times \left[\left(1 - \frac{\alpha k_n}{n}\right) \left(\frac{\alpha k_n}{n} + \frac{\beta}{k_n^{1/\alpha}} + o(n^{-1}) \right) \right]$$
$$+ \left[f\left(\frac{k_n - 1}{n^\alpha}\right) - f\left(\frac{k_n}{n^\alpha}\right) \right]$$
$$\times \left[\frac{\alpha k_n}{n}(n - \alpha k_n) \left(\frac{1}{n} - \frac{\beta}{n k_n^{1/\alpha}} + o(n^{-2}) \right) \right].$$

A second-order Taylor expansion, together with some standard computation, yields

$$(U_n - I) f(s_n) = n^{-1-\alpha} \frac{\beta}{s_n^{1/\alpha}} f'(s_n) + n^{-1-\alpha} \alpha s_n f''(s_n) + o(n^{-1-\alpha})$$

where $s_n = k_n / n^\alpha$. Since $s_n \to s$ from (4), it follows that

$$n^{1+\alpha}(U_n - I) f(s) \to A f(s)$$

uniformly on $(0, \infty)$, for $f \in \mathcal{D}$, with A as in (15). Theorem 1.6.5 by Ethier and Kurtz (1986) now implies that

$$U_n(t/\varepsilon_n) f(s) \to U(t) f(s), \qquad \text{as } n \to \infty, \qquad \forall f(s) \in \hat{C}((0, \infty)),$$

where $\varepsilon_n = n^{-1-\alpha}$, where U is the Feller semigroup operator corresponding to A. Then Theorem 4.2.6 of Ethier and Kurtz (1986) in turn implies that

$$\frac{K_n(\lfloor n^{1+\alpha} t \rfloor)}{n^\alpha} \xrightarrow{d} S(t)$$

holds in $D_{[0,\infty)}((0, \infty))$, provided the weak convergence of the initial distributions holds on $(0, \infty)$. Since the $S(t)$ has null probability of touching the origin for all $t > 0$, if the convergence of the initial distributions holds on $[0, \infty)$, then the weak convergence holds in $D_{[0,\infty)}([0, \infty))$. The full statement now follows from the fact that convergence in distribution on $D_{[0,\infty)}([0, \infty))$ to an object that belongs to $C_{[0,\infty)}([0, \infty))$ with probability one implies convergence in distribution on $C_{[0,\infty)}([0, \infty))$. □

The above Theorem states that the sequence of laws induced by the Markov chains $\tilde{K}_n(t)$ on the cadlag space of sample paths $D_{[0,\infty)}([0, \infty))$ converges weakly to the law induced by $S(\cdot)$ onto $C_{[0,\infty)}([0, \infty))$. By analogy with (4), the scaling limit $S(\cdot)$ in (13) can be interpreted as a dynamic measure of diversity in the generalised gamma population model constructed at the beginning of the present section. Figure 1 shows some examples of sample paths of (13) for different values of α.

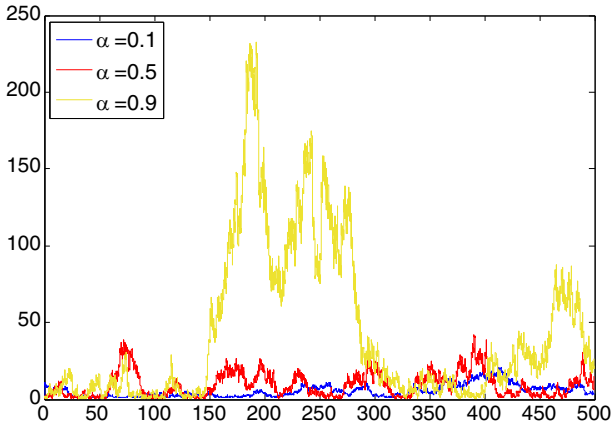

Fig. 1 Some paths of (13) for different values of α

4 Stationary approximations to the scaling limit

Bayesian nonparametric inference in the presence of temporally structured data usually tries to use stationary processes as building blocks of a broader model, as one typically has in mind a certain distributional structure for the marginal states and wants to make the latter depend on some covariate, such as time. Note that this approach is not particularly restrictive, as one can still model non-stationary time series starting from stationary elements used for the construction in different hierarchical levels, in a similar spirit to hidden Markov models (Mena and Ruggiero 2016). It would then be desirable to have a stationary process describing the dynamics of the number of clusters. As this is not the case for the dynamics associated to generalised gamma clustering, as shown in Proposition 2, we devise a weak approximation to the diffusion in Theorem 1 such that any term of the approximating sequence is a stationary diffusion. This will provide stationary dynamics for the number of groups which are as close as desired to those induced by a generalised gamma population, with an explicit invariant measure.

It is instructive to construct each term of the sequence of stationary diffusions from a continuous-time Markov chain, which highlights the underlying dynamics and allows a comparison with the results of the previous section. For any $\gamma > 0$, consider a continuous-time Markov chain $\{K_{n,\gamma}(t), t \geq 0\}$ on $\mathbb{N}$ with transition rates

$$\lambda_1 = \frac{\alpha k^{1+\gamma}}{n^{1+\alpha\gamma}} + \frac{\beta}{k^{1/\alpha}} + o(n^{-1}) = \frac{\alpha s^{1+\gamma}}{n^{1-\alpha}} + \frac{\beta/s^{1/\alpha}}{n} + o(n^{-1})$$

from k to $k+1$ and

$$\lambda_2 = \frac{\alpha k^{1+\gamma}}{n^{1+\alpha\gamma}} + o(n^{-1}) = \frac{\alpha s^{1+\gamma}}{n^{1-\alpha}} + o(n^{-1})$$

from k to $k-1$. Here, we are still assuming that $k_n/n^\alpha \to s_n$. The following result mimics Theorem 1 and identifies the scaling limit of the sequence of Markov chains.

Proposition 3 *Let $\{K_{n,\gamma}(t), t \geq 0\}$ be the above-defined continuous-time Markov chain with rates λ_1 and λ_2, and let $\{\tilde{K}_{n,\gamma}(t), t \geq 0\}$ be defined as*

$$\tilde{K}_{n,\gamma}(t) = \frac{K_{n,\gamma}(n^{1+\alpha}t)}{n^{\alpha}}.$$

Let $\{S_\gamma(t), t \geq 0\}$ be the diffusion process on $[0, \infty)$ driven by the stochastic differential equation

$$\mathrm{d}S_\gamma(t) = \frac{\beta}{S_\gamma(t)^{1/\alpha}}\mathrm{d}t + \sqrt{2\alpha(S_\gamma(t))^{1+\gamma}}\mathrm{d}B(t). \tag{17}$$

If $\tilde{K}_{n,\gamma}(0) \xrightarrow{d} S_\gamma(0)$ then

$$\tilde{K}_{n,\gamma}(\cdot) \xrightarrow{d} S_\gamma(\cdot) \quad \text{in} \quad C_{[0,\infty)}([0, \infty))$$

as $n \to \infty$.

Proof The proof proceeds along the same lines of that of Theorem 1. In particular, the well definedness of the diffusion follows by the same argument for

$$A_\gamma f(s) = \frac{\beta}{s^{1/\alpha}} f'(s) + \alpha s^{1+\gamma} f''(s).$$

with $f \in \mathcal{D}_\gamma([0, \infty))$ and

$$\mathcal{D}_\gamma([0, \infty)) = \{f \in C([0, \infty)) \cap C^2((0, \infty)) : A_\gamma f \in C([0, \infty))\}.$$

Letting now $U_{n,\gamma}$ be the semigroup corresponding to the Markov chain $K_{n,\gamma}$, we have

$$\begin{aligned}(U_{n,\gamma} - I)f\left(\frac{k_n}{n^\alpha}\right) &= \frac{1}{n^\alpha} f'(s_n)(\lambda_1 - \lambda_2) + \frac{1}{2n^{2\alpha}} f''(s_n)(\lambda_1 + \lambda_2) + o(n^{-1-\alpha}) \\ &= n^{-1-\alpha} f'(s_n)\frac{\beta}{s_n^{1/\alpha}} + n^{-1-\alpha} f''(s_n)\alpha s_n^{1+\gamma} + o(n^{-1-\alpha}),\end{aligned}$$

and the rest of the proof follows similarly. □

We conclude by showing that any process in the class $\{S_\gamma(\cdot)\}_{\gamma>0}$ is stationary; we identify the invariant measure and prove that for any sequence $\gamma_\ell \to 0$, the associated sequence of diffusions $\{S_{\gamma_\ell}(\cdot)\}_{\gamma_\ell}$ converges in distribution to $S(\cdot)$ in Theorem 1, as $\ell \to \infty$. For notational simplicity, we write $S_\gamma(\cdot)$ in place of $S_{\gamma_\ell}(\cdot)$.

Proposition 4 *Let $S(\cdot)$ and $S_\gamma(\cdot)$ be as in (13) and (17), respectively. For any $\gamma > 0$, $S_\gamma(\cdot)$ is stationary with invariant measure*

$$\psi_\gamma(x) \propto \frac{1}{2\alpha x^{1+\gamma}} \exp\left\{-\frac{\beta}{1+\alpha\gamma} x^{-\frac{1+\alpha\gamma}{\alpha}}\right\}, \qquad x > 0, \tag{18}$$

with normalising constant $C = 2\beta(\frac{1+\alpha\gamma}{\beta})^{\frac{1}{1+\alpha\gamma}} / \Gamma(\frac{\alpha\gamma}{1+\alpha\gamma})$*, and has an entrance boundary at* 0 *and a natural boundary at* ∞*. Moreover, as* $\gamma \to 0$*,* $S_\gamma(\cdot)$ *converges in distribution to* $S(\cdot)$ *on* $C_{[0,\infty)}([0,\infty))$*, provided the initial distributions converge.*

Proof Denote by $\mu_\gamma(\cdot)$ and $\sigma_\gamma(\cdot)$ the drift and diffusion coefficients in (17). Then

$$z_\gamma(x) = \exp\left\{-\int^x \frac{2\mu_\gamma(y)}{\sigma_\gamma^2(y)}\mathrm{d}y\right\} = \exp\left\{\frac{\beta}{1+\alpha\gamma}x^{-\frac{1+\alpha\gamma}{\alpha}}\right\}$$

and

$$m_\gamma(x) = \frac{1}{\sigma_\gamma^2(x)z_\gamma(x)} = \frac{1}{2\alpha x^{1+\gamma}}\exp\left\{-\frac{\beta}{1+\alpha\gamma}x^{-\frac{1+\alpha\gamma}{\alpha}}\right\}.$$

The function $z_\gamma(x)$ behaves essentially as $z(x) = z_0(x)$, so $Z_\gamma(0) = Z_\gamma(+\infty) = \infty$. The function $m_\gamma(x)$, instead, behaves like $m(x) = m_0(x)$ in a neighbourhood of $x = 0$, but goes to 0 as $1/x^{1+\gamma}$ for $x \to \infty$, so $M_\gamma(0) < \infty$ and $M_\gamma(+\infty) < \infty$. We immediately have $\Sigma_\gamma(0) = \Sigma_\gamma(+\infty) = \infty$. Moreover, $N_\gamma(0) < \infty$ and $N_\gamma(+\infty) = \infty$. The boundary classification then again follows from Karlin and Taylor (1981), Section 15.6. From (16), we now find that $Z_\gamma(x) \equiv \infty$ implies $C_1 = 0$, whence $\psi_\gamma(x) \propto m_\gamma(x)$.

Note now that the infinitesimal generators of $S_\gamma(\cdot)$ and $S(\cdot)$ satisfy

$$\begin{aligned}\left|A_\gamma f(s) - Af(s)\right| &= \left|\frac{\beta}{s^{1/\alpha}}f'(s) + \alpha s^{1+\gamma}f''(s) - \left(\frac{\beta}{s^{1/\alpha}}f'(s) + \alpha s f''(s)\right)\right| \\ &= \alpha|(s^{1+\gamma} - s)f''(s)| \to 0\end{aligned}$$

uniformly on $[0,\infty)$, for $f \in \mathcal{D}_0 = \mathcal{D} \cap C^2([0,\infty))$, as $\gamma \to 0$. Now, $\mathcal{D}_0$ can be easily shown to be a core for A (cf. Ethier and Kurtz (1986), Section 1.3), i.e. its closure is such that $\overline{\mathcal{D}_0} = \mathcal{D}$ and $\overline{A|_{\mathcal{D}_0}} = A$ (here $\mathcal{D}_0$ and $\mathcal{D}$ differ for functions with one or two infinite derivatives at 0). Theorems 1.6.1 and 4.2.5 by Ethier and Kurtz (1986) then yield

$$\lim_{\gamma\to 0} U_\gamma(t)f = U(t)f \qquad \forall f \in \hat{C}([0,\infty))$$

and

$$S_\gamma(\cdot) \xrightarrow{d} S(\cdot) \qquad \text{as } \gamma \to 0$$

on $D_{[0,\infty)}[0,\infty)$, provided the initial distributions converge. The rest of the argument is now analogous to the proof of Theorem 1. □

Figure 2 shows the qualitative difference among sample paths of $S_\gamma(\cdot)$ for decreasing values of γ. Figure 3 shows the convergence of the unnormalised stationary measures of $S_\gamma(\cdot)$ to the speed measure $m(x)$ of $S(\cdot)$ (blue curve), which does not

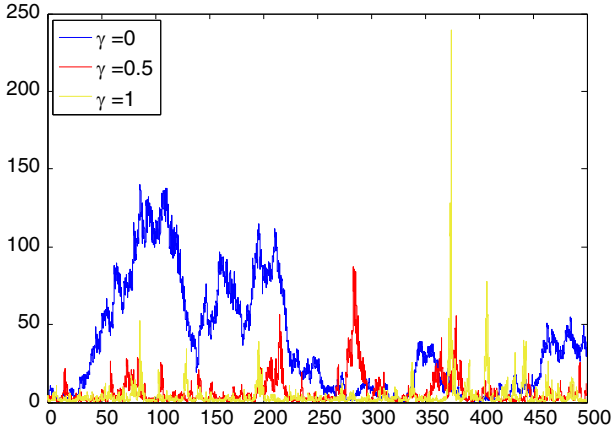

Fig. 2 Some paths of (17) for different values of γ

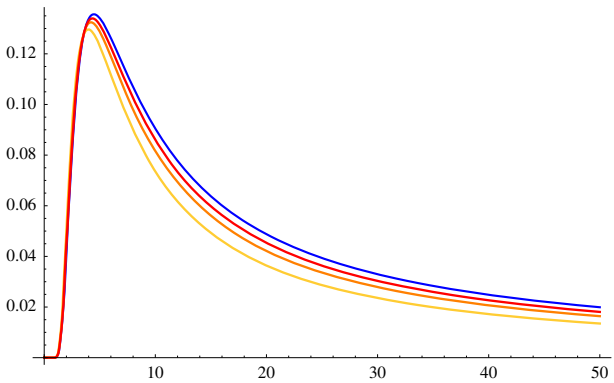

Fig. 3 Convergence of unnormalised stationary densities (16) for $\gamma = 0.1, 0.05, 0.025$ (*yellow to red*) to the speed measure (14) of (13) (*blue*), which does not integrate

integrate, for decreasing values of γ (bottom to top); cf. (14) and (18). Here, the stationary distribution of $S_\gamma(\cdot)$ has right tail decaying as $x^{-1-\gamma}$.

Acknowledgements The authors are grateful to two anonymous referees for helpful comments and to Pierpaolo De Blasi and Bertrand Lods for useful suggestions. This work was conducted while the second author was affiliated to the University of Torino and Collegio Carlo Alberto, Italy.

References

Arbel, J., Favaro, S., Nipoti, B., Teh, Y.W. (2016). Bayesian nonparametric inference for discovery probabilities: credible intervals and large sample asymptotics. *Statistica Sinica* (**in press**).

Blackwell, D. (1973). Discreteness of Ferguson selections. *The Annals of Statistics*, *2*, 356–358.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, *31*, 929–953.

Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*. Hoboken, NJ: Wiley.

Costantini, C., De Blasi, P., Ethier, S. N., Ruggiero, M., 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., Prünster, I., Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Transactions Pattern Analysis and Machine Intelligence*, *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. Berlin, Heidelberg: Springer.

Ethier, S. N., Kurtz, T. G. (1986). *Markov processes*: *characterization and convergence*. New York: Wiley.

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

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, *1*, 209–230.

Gnedin, A., Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *Journal of Mathematical Sciences*, *138*, 5674–5685.

Griffin, J. E., Kolossiatis, M., Steel, M. F. J. (2013). Comparing distributions by using dependent normalized random-measure mixtures. Journal of the Royal Statistical Society. *Series B*, *75*, 499–529.

Griffin, J. E., Leisen, F. (2016). Compound random measures and their use in Bayesian nonparametrics . *Journal of the Royal Statistical Society. Series B* (**in press**).

Hjort, N. L., Holmes, C. C., Müller, P., Walker, S. G. (Eds.). (2003). *Bayesian nonparametrics*. Cambridge: Cambridge University Press.

Karlin, S., Taylor, H. M. (1981). *A second course in stochastic processes*. New York: Academic Press.

Korwar, R. M., Hollander, M. (1973). Contribution to the theory of Dirichlet processes. *The Annals of Probability*, *1*, 705–711.

Lijoi, A., Prünster, I. (2010). Models beyond the Dirichlet process. In N. L. Hjort, C. C. Holmes, P. Müller, S. G.Walker (Eds.), Bayesian nonparametrics (pp. 80–136). Cambridge: Cambridge University Press.

Lijoi, A., Mena, R. H., Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-gaussian priors. *Journal of the American Statistical Association*, *472*, 1278–1291.

Lijoi, A., Mena, R. H., Prünster, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. Journal of the Royal Statistical Society. *Series B*, *69*, 715–740.

Lijoi, A., Prünster, I., Walker, S. G. (2008). Investigating nonparametric priors with Gibbs structure. Statistica Sinica, 18, 1653–1668.

Lijoi, A., Nipoti, B., Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli*, *20*, 1260–1291.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, *12*, 351–357.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association.

MacEachern, S.N. (2000). Dependent Dirichlet processes. *Technical Report*, Department of Statistics, Ohio State University, Columbus.

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

Mena, R. H., Ruggiero, M., Walker, S. G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modelling. *Journal of Statistical Planning and Inference*, *141*, 3217–3230.

Papaspiliopoulos, O., Ruggiero, M., Spanò, D. (2016). Conjugacy properties of time-evolving Dirichlet and gamma random measures. *Electronic Journal of Statistics* (**to appear**).

Petrov, L. (2009). Two-parameter family of diffusion processes in the Kingman simplex. *Functional Analysis and its Applications*, *43*, 279–296.

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

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In T. S. Ferguson, L. S. Shapley, J. B. MacQueen (Eds.), *Statistics, Probability and Game Theory*. *Institute of Mathematical Statistics Lecture Notes-Monograph Series*, *30*. Hayward, CA: Institute of Mathematical Statistics.

Pitman, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Mathematics, 1875*. Berlin: Springer.

Pitman, J., Yor, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, *25*, 855–900.

Ruggiero, M. (2014). Species dynamics in the two-parameter Poisson–Dirichlet diffusion model. *Journal of Applied Probability*, *51*, 174–190.

Ruggiero, M., Walker, S. G. (2009a). Bayesian nonparametric construction of the Fleming–Viot process with fertility selection. *Statistica Sinica*, *19*, 707–720.

Ruggiero, M., Walker, S. G. (2009b). Countable representation for infinite-dimensional diffusions derived from the two-parameter Poisson–Dirichlet process. *Electronic Communications in Probability*, *14*, 501–517.

Ruggiero, M., Walker, S. G., Favaro, S. (2013). Alpha-diversity processes and normalized inverse-Gaussian diffusions. *The Annals of Applied Probability*, *23*, 386–425.

Smith, A. F. M., Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B*, *55*, 3–23.