A Birth-Death Process for Feature Allocation

Konstantina Palla [1] **David Knowles** [2] **Zoubin Ghahramani** [3 4]

Abstract

We propose a Bayesian nonparametric prior over feature allocations for sequential data, the *birth-death feature allocation process* (BDFP). The BDFP models the evolution of the feature allocation of a set of N objects across a covariate (e.g. time) by creating and deleting features. A BDFP is exchangeable, projective, stationary and reversible, and its equilibrium distribution is given by the Indian buffet process (IBP). We show that the Beta process on an extended space is the de Finetti mixing distribution underlying the BDFP. Finally, we present the finite approximation of the BDFP, the Beta Event Process (BEP), that permits simplified inference. The utility of the BDFP as a prior is demonstrated on real world dynamic genomics and social network data.

1. Introduction - Problem Statement

We are interested in time series settings where we observe data $\{Y_t \in \mathcal{Y} : t = 1, \ldots, L\}$. We consider problems where the observations are explained by a latent structure which assigns objects to features and this feature allocation changes over time. For instance, consider the topics covered by a number of newspapers over time; some topics "die" while new ones are "born". The topic coverage of each paper is its latent feature allocation which could be modelled with an Indian buffet process (Griffiths & Ghahramani, 2011, IBP). While static feature allocation models are well studied, these are not able to handle the time series nature of many datasets. We propose a process that extends the IBP by allowing the feature allocation to evolve over the covariate as a result of "birth" and "death" of features.

2. Related Work

We target problems where the data depends on a covariate, such as time or space, and is explained by a latent structure, in particular a (multi-membership) clustering of the data points. The observations are result of the underlying partitioning and its evolution over the covariate. Typical models fall in two main categories: clustering and feature allocation. The former allow each data point to belong to one and only one class (cluster), while the latter let each data point belong to multiple groups (features). Bayesian nonparametric approaches are primarily based on the Chinese restaurant process (CRP, Aldous, 1983) or the Indian buffet process (IBP, Griffiths & Ghahramani, 2005) corresponding to the two categories. In particular, a sample from a CRP is an assignment of data points to disjoint classes (a clustering), while a sample from an IBP is an allocation of the data points to (possibly) overlapping classes (a feature allocation). Dependent nonparametric processes extend distributions over partitions to distributions over collections of partitions indexed by locations in some covariate space, such as $\mathbb{R}^+$ (e.g. continuous time), $\mathbb{Z}$ (e.g. discrete time), or $\mathbb{R}^d$ (e.g. geographical location). Teh et al. (2013) define such a process based on the duality between Kingman's coalescent (Kingman, 1982) and the Dirichlet diffusion tree (Neal, 2003). In the resulting "Fragmentation-Coagulation" process (FCP) a partitioning of the data points evolves over the covariate undergoing fragmentation and coagulation events while maintaining CRP marginals. More recently, Palla et al. (2013) derived a dependent partition-valued process (DPVP) on an arbitrary covariate space which, like the FCP, is exchangeable and has CRP distributed marginals. In the setting of feature allocations, Williamson et al. (2010) propose a nonparametric process, the dependent IBP (dIBP), with IBP distributed marginals and in which the feature allocations are coupled over the covariate space using a Gaussian process (GP, Rasmussen & Williams, 2006). In a similar vein, Van Gael et al. (iFHMM, 2009) define the Markov Indian Buffet process (mIBP), a probability distribution over a potentially infinite number of binary Markov chains evolving in discrete time. They use the mIBP to extend the factorial hidden Markov model (FHMM, Ghahramani & Jordan, 1997) to the infinite FHMM (iFHMM).

In this paper, we address the problem of dependence for

[1] University of Oxford, Oxford, UK [2] Stanford University, California, USA [3] University of Cambridge, Cambridge, UK [4] Uber AI Labs, SF, California, USA. Correspondence to: Konstantina Palla <konstantina.palla@gmail.com>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017. Copyright 2017 by the author(s).

binary latent feature models. We propose a process that extends the IBP by allowing features to be "born" and "die" at times learnt by the model, while maintaining the essential mathematical properties of the IBP. The process is a Markov Jump process (MJP) where the events are the birth or the death of a feature. The idea is closely related to the FCP where the events are either a fragmentation of a cluster or a coagulation of two clusters. The partitions at each location in the FCP are marginally a sample from a Chinese restaurant process, while the feature allocations in the BDFP are marginally samples from an IBP. Compared to the dIBP, both processes model feature allocations evolving over the covariate. However, while in the dIBP the assignment of data points to a feature might change over the covariate, in our process, it remains the same until the feature dies. In the case of the iFHMM, the authors model the dependence of a feature allocation on a discrete time variable as opposed to our process where continuous covariate space is assumed. Moreover, in the iFHMM, the marginal distribution of a feature allocation is analogous but not equal to an IBP. We call the proposed process the birth-death feature allocation process (BDFP). The BDFP is exchangeable, projective, stationary and reversible, and its equilibrium distribution is given by the Indian buffet process.

3. Feature Allocations and the Indian Buffet Process

Consider a dataset with N data points indexed by integers $[N] := \{1, 2, \ldots, N\}$ (allowing $N \to \infty$). Each datapoint n is associated with a binary vector $\mathbf{Z}_n$ of length K that defines its feature allocation; $Z_{nk} = 1$ if datapoint n has feature k and $Z_{nk} = 0$ otherwise. The potential total number of features K may be infinite. The binary matrix $\mathbf{Z}_{[N]} = [\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_N^T]^T$ specifies a random feature allocation of $[N]$, while $\mathcal{Z}_N$ denotes the space of all feature allocations of $[N]$, i.e. $\mathbf{Z}_{[N]} \in \mathcal{Z}_N$. We define m_k as the number of datapoints that possess feature k, $K_+ = \sum_{h=1}^{2^N-1} K_h$ as the number of features for which $m_k > 0$ and K_h as the multiplicity of feature h, that is the number of times the same binary column h appears in $\mathbf{Z}_{[N]}$. Under the IBP (Griffiths & Ghahramani, 2011), the probability of a matrix $\mathbf{Z}_{[N]}$ is

$$g([\mathbf{Z}_{[N]}];\alpha) = \frac{\alpha^{K_+}}{\prod_{h=1}^{\mathcal{H}} K_{h!}} \exp(-\alpha H_N) \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)}{N!} \tag{1}$$

where $\alpha > 0$ is the concentration parameter, $H_N = \sum_{j=1}^{N} \frac{1}{j}$ is the Nth harmonic number and $\mathcal{H} \leq 2^N - 1$ is the number of distinct nonzero features in the allocation.

Thibaux & Jordan (2007) showed one can construct the Indian buffet process from a Beta-Bernoulli process using the following two stage sampling process for $n = 1, \ldots, N$:

$$B|c, \mu_0 \sim \mathrm{BP}(c, \mu_0) \quad \mathbf{Z}_n|B \sim \mathrm{BeP}(B) \tag{2}$$

where $B = \sum_{k=1}^{\infty} \omega_k \delta_{\theta_k}$ and $Z = \sum_{k=1}^{\infty} f_k \delta_{\theta_k}$. First a draw B is sampled from the Beta process $\mathrm{BP}(c\mu_0)$ (Hjort, 1990) with μ_0 as the base distribution. B is a set of pairs (ω_k, θ_k) sampled from a Poisson process on the product space $[0, 1] \times \Theta$ with Lévy intensity $\nu(\mathrm{d}\omega, \mathrm{d}\theta) = c\omega^{-1}(1-\omega)^{c-1}\mathrm{d}\omega\mu_0(\mathrm{d}\theta)$. Then, B is used as the atomic hazard measure for a Bernoulli process $\mathrm{BeP}(B)$. Each $\mathbf{Z}_n$ is a draw from the Bernoulli process and constitutes a collection of atoms of unit mass on Θ. Then, $\mathbf{Z}_n$ is a binary vector containing the $\{f_k\}_{k=1}^{\infty}$ values resulting from tossing a countably infinite sequence of (conditionally independent) coins with success probabilities ω_k, i.e. $f_k|\omega_k \sim \mathrm{Bernoulli}(\omega_k)$. This construction allows the use of de Finetti's theorem (de Finetti, 1931) that lets the joint distribution of the rows to be written as

$$P(\mathbf{Z}_1, \ldots, \mathbf{Z}_N) = \int \Big[\prod_{n=1}^{N} P(\mathbf{Z}_n|B)\Big] \mathrm{d}P(B) \tag{3}$$

where B is the random measure that renders the variables $\mathbf{Z}_n$ conditionally independent. Equation (3) shows the exchangeability of the rows of $\mathbf{Z}_n$, since they can be described as a mixture of Bernoulli processes.

4. Birth-Death Process for Feature Allocation

We consider a continuous-time Markov process $(Z(t))_{t\geq 0}$ in which each $Z(t)$ is a random feature allocation taking values in the discrete space $\mathcal{Z}_N$. The state space is countably infinite; it is determined by all the possible feature allocations defined by N datapoints and K features, where $K \to \infty$. The Markov process $(Z(t))$ evolves over time jumping to different states (feature allocations). Let $\{t_1, \ldots, t_J \in \mathbb{R} : J \in \mathbb{N}\}$ denote the times when the chain jumps such that $t_j = \inf\{\tau \geq t_{j-1} : Z(\tau) \neq Z(t_{j-1})\}$ and $Z(t_j) \in \mathcal{Z}_N$. These jumps are a result of a *birth* or a *death* of a feature. The process $(Z(t))$ can only jump to neighbouring states, i.e. if the chain is currently at state $Z(t_j) = s$, then at time t_{j+1} it transitions to $Z(t_{j+1}) = s'$ where a new feature is created or an existing feature is deleted after a birth or a death event respectively. Let $\mathcal{Z}_N^s \subset \mathcal{Z}_N$ be the discrete space of neighboring states to state s. The process is time homogeneous with transition probabilities $\mathrm{P}(Z(t+y) = s'|Z(y) = s) = \mathrm{P}(Z(t) = s'|Z(0) = s) = \mathrm{p}_{ss'}(t)$ for all t, y, where $s, s' \in \mathcal{Z}_N$. At time t_{j+1} the process jumps to the next state $Z(t_{j+1}) = s'$ with rate determined by the current state $Z(t) = s$ and the corresponding event, i.e birth or death. More specifically,

- **Birth:** Suppose $s \in \mathcal{Z}_N$ is a feature allocation with K_s nonzero features and $s' \in \mathcal{Z}_N^s$ is another feature allocation that differs from s in having one additional feature of size $|a|$ so that $K_s' = K_s + 1$. We choose the transition rate from s to s' as

$$q_{ss'} = R\frac{(|a|-1)!(N-|a|)!}{N!} \tag{4}$$

where $R > 0$ is a parameter governing the birth rate. The new feature a is a binary column of length N. There are $\binom{N}{|a|}$ binary formulations for this feature and $2^N - 1 = \sum_{n=1}^{N}\binom{N}{n}$ for all possible feature births and thus, the total birth rate from s is $\sum_{n=1}^{N}\begin{pmatrix}N\\n\end{pmatrix} R\frac{(n-1)!(N-n)!}{N!} = R\sum_{n=1}^{N}\frac{1}{n} = R \cdot H_N$ where $H_N = \sum_{n=1}^{N} 1/n$ is the N-th harmonic number and $n = |a|$.

- **Death:** The rate of transitioning from s' to s is

$$q_{s's} = \frac{Rr}{\alpha} \tag{5}$$

where $D = \frac{R}{a}$ is a parameter governing the death rate and r is the multiplicity of the feature in s' that dies. The multiplicity r is the combinatorial factor that accounts for all the possible ways of obtaining the same equivalence class as defined in Griffiths & Ghahramani (2011) . There are $K_{s'}$ features (including repetitions of the same feature) in s' that might "die", thus the total death rate from s' is $\frac{RK_{s'}}{\alpha}$.

The total rate of transition out of state $s \in \mathcal{Z}_N$ is the sum of the total birth and death rates, $q_s = RH_N + \frac{RK_s}{\alpha} = R\Big(H_N + \frac{K_s}{\alpha}\Big)$. We call $(Z(t))_{t>=0}$ a **birth-death feature allocation process** with birth rate R and death rate $\frac{R}{\alpha}$ and write $\mathbf{BDFP}(\alpha, R)$.

Theorem 1. *The Markov process $(Z(t))_{t\geq 0}$ is irreducible and has stationary distribution* $\mathrm{IBP}(\alpha)$*. Furthermore, it is reversible.*

Proof. A continuous time Markov chain is irreducible if it is possible to eventually get from every state to every other state with positive probability. It is reversible if detailed balance holds, i.e. there is a probability distribution π on $\mathcal{Z}_N$ such that $\pi_s q_{ss'} = \pi_{s'} q_{s's}$ for all $s, s' \in \mathcal{Z}_N$. Then π is also the invariant (equilibrium) distribution of the Markov chain. The chain in BDFP is irreducible, because for any $T > 0$ and any two distinct feature allocations $\gamma, \rho \in \mathcal{Z}_N$, there is a positive probability that if it starts at $\gamma \in \mathcal{Z}_N$, it will end at $\rho \in \mathcal{Z}_N$. Reversibility and the equilibrium distribution can be demonstrated by detailed balance. Suppose γ, ρ are feature allocations such that $\gamma, \rho \in \mathcal{Z}_N$ and ρ differs from γ in that it has one additional feature a of size $|a|$. The number of (nonzero) features in ρ is $K_\rho = K_\gamma + 1$. Then,

$$\begin{aligned}
g(\gamma;\alpha)q_{\gamma\rho} &= \frac{\alpha^{K_\gamma}}{\Pi_{h=1}^{\mathcal{H}_\gamma} K_h!}\exp\left(-\alpha H_N\right)\prod_{k=1}^{K_\gamma}\frac{(N-m_k)!(m_k-1)!}{N!}R\frac{(|a|-1)!(N-|a|)!}{N!} \\
&\overset{m_{K_\gamma+1}=|\alpha|}{=} \frac{\alpha^{K_\gamma+1}}{\alpha\Pi_{h=1}^{\mathcal{H}_\gamma} K_h!}\exp\left(-\alpha H_N\right)\prod_{k=1}^{K_\gamma+1}\frac{(N-m_k)!(m_k-1)!}{N!}R \\
&= \frac{\alpha^{K_\rho}}{r_\alpha\Pi_{h=1}^{\mathcal{H}_\gamma} K_h!}\exp\left(-\alpha H_N\right)\prod_{k=1}^{K_\rho}\frac{(N-m_k)!(m_k-1)!}{N!}\frac{R}{\alpha}r_a \\
&\overset{r_\alpha=K_\alpha}{=} \frac{\alpha^{K_\rho}}{\Pi_{h=1}^{\mathcal{H}_\rho} K_h!}\exp\left(-\alpha H_N\right)\prod_{k=1}^{K_\rho}\frac{(N-m_k)!(m_k-1)!}{N!}\frac{R}{\alpha}r_a \\
&= g(\rho;\alpha)q_{\rho\gamma}
\end{aligned} \tag{6}$$

where $g(\gamma;\alpha)$ is the probability of a feature allocation γ under the IBP as defined in Equation (1), $q_{\gamma\rho}$ is the transition rate from state γ to state ρ, $\mathcal{H}_\gamma$, $\mathcal{H}_\rho$ are the number of distinct features in states γ and ρ respectively and r_a is the multiplicity (the times the feature is present at the current feature allocation) of feature a that dies. Detailed balance holds, and as such the process is reversible and the equilibrium distribution is $\mathrm{IBP}_{[N]}(\alpha)$. □

Assume that $(z(t))$ is a realization of the BDFP $(Z(t))$ over the finite interval $[0, T], T > 0$ and we write $(z(t))_{0\leq t\leq T}$. With probability one the sample path $(z(t))_{0\leq t\leq T}$ will only contain a finite number of jump events, each of which is either a birth or a death event. We write B and Q to denote the set of the features created or turned off by birth or death events respectively.

Proposition 1. *Writing $q(t) = q_{z(t)}$ to denote the total transition rate out of state $z(t)$, the probability of a realization $(z(t))$ under the law of the* BDFP *is:*

$$R^{|B|+|Q|}\frac{\alpha^{A-|B|-|Q|}}{\prod_{h=1}^{A^*-|B^*|}K_h!}\exp\left(-\alpha H_N\right)\exp\Big(-\int_0^T q(t)\mathrm{d}t\Big)\times\ldots$$
$$\prod_{b\in B\cup\{z(t=0)\}}\frac{(|b|-1)!(N-|b|)!}{N!}\prod_{d\in D}r_d \tag{7}$$

where $A = K_0 + |B| = K_T + |Q|$, $A^* = H_0 + |B^*| = H_T + |Q^*|$. B^*, Q^* are the sets of features with zero multiplicity at their creation time or with multiplicity of one at their death time respectively, and $\{z(t)\}$ denotes the set of features at time t.

4.1. Dependent Beta Process Construction

The BDFP process can be constructed using a nonhomogenous Poisson process $\mathbf{\Pi}$. Consider the Lévy measure $\nu(\mathrm{d}\omega\mathrm{d}x\mathrm{d}t_b\mathrm{d}t_\omega)$ on a product space $[0,1]\otimes\mathbb{X}\otimes\mathbb{R}\otimes[0,\infty)$. A sample corresponds to set of points $\mathbf{\Pi} = \{\omega_k, x_k, t_b^k, t_\omega^k\}_k$ where the range of k is countably infinite. Each atom corresponds to a feature and is associated with a weight $\omega_k \in [0,1]$, a location x_k, a birth time $t_b^k \in \mathbb{R}$ and a life-span $t_\omega^k \in [0,\infty)$ (Figure 1). The Lévy measure is of the form $\nu(\mathrm{d}\omega\mathrm{d}x\mathrm{d}t_b\mathrm{d}t_\omega) = \rho(\mathrm{d}\omega)\mu(\mathrm{d}x\mathrm{d}t_b\mathrm{d}t_\omega)$ and

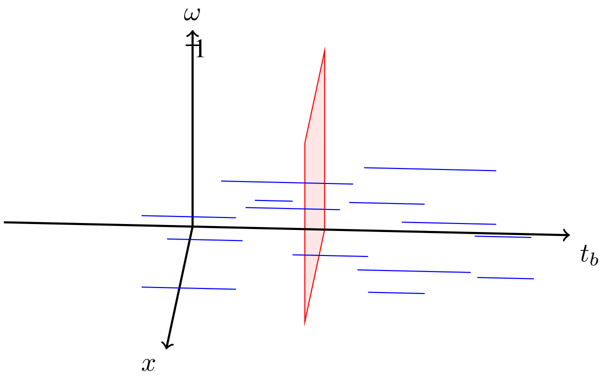


Figure 1. Cartoon for the dependent Beta process construction of the BDFP: a realisation of a Poisson process $\mathbf{\Pi}$ over the product space $[0,1]\otimes\mathbb{X}\otimes\mathbb{R}\otimes[0,\infty)$ is drawn. The t_ω dimension over the space $[0,\infty)$ is omitted in the axis representation. However, each t_ω corresponding to each point (feature) is drawn as a blue line of length t_ω starting at the associated birth time point.

corresponds to a Beta process on the combined space $\Theta = \mathbb{X}\otimes\mathbb{R}\otimes[0,\infty)$ with $\rho(\mathrm{d}\omega) = \alpha\omega^{-1}(1-\omega)^{\alpha-1}$ and *base measure* $\mu(\mathrm{d}\boldsymbol{\theta}) = \mu(\mathrm{d}x\mathrm{d}t_b\mathrm{d}t_w)$. Setting $g(\mathrm{d}t_b) = \mathrm{d}t_b$ and $\beta(\mathrm{d}t_\omega) = D\exp^{-Dt_\omega}\mathrm{d}t_\omega$, the base measure is $\mu(\mathrm{d}\boldsymbol{\theta}) = \mu_0(\mathrm{d}x)g(\mathrm{d}t_b)\beta(\mathrm{d}t_\omega) = \mu_0(\mathrm{d}x)\mathrm{d}t_b D\exp^{-Dt_\omega}\mathrm{d}t_\omega$, where D is the death rate. The constant measure $g(\mathrm{d}t_b)$ over the real line $\mathbb{R}$ is infinite but σ-finite, that is the total measure $g(\mathbb{R}) = \infty$, but there is a measurable partition (E_k) of $\mathbb{R}$ with each $g(E_k) < \infty$. Since $\nu(\mathrm{d}\omega\mathrm{d}\boldsymbol{\theta})$ integrates to infinity but satisfies $\int_{[0,1]}\int_\Theta(1\wedge|\omega|)\nu(\mathrm{d}\omega\mathrm{d}\boldsymbol{\theta}) < \infty$, a countably infinite number of i.i.d. random points $\{(\omega_k,\boldsymbol{\theta}_k)\}_{k=1}^\infty$ are obtained from the Poisson process and $\sum_{k=1}^\infty \omega_k$ is finite with probability one. A Beta process is a completely random measure (Kingman, 1967) and, as such, a sample can be expressed as $B = \sum_{k=1}^\infty \omega_k\delta_{\boldsymbol{\theta}_k}|\alpha,\mu \sim \mathrm{BP}(\alpha\mu)$, where the atoms $\boldsymbol{\theta}_k = \{x_k, t_b^k, t_\omega^k\} \in \Theta$ and weights $\omega_k \in [0,1]$.

Having drawn a sample B we can construct the feature allocations over an index space $\mathbb{R}$ as follows:

$$\begin{aligned} B &= \sum_{k=1}^\infty \omega_k\delta_{\boldsymbol{\theta}_k} \quad |\alpha,\mu \sim \mathrm{BP}(\alpha\mu) \\ S_{n:} &= \sum_{k=1}^\infty b_{nk}\delta_{\boldsymbol{\theta}} \quad |B \sim \mathrm{BeP}(\omega_k) \\ Z_{nk}(t) &= S_{nk}\mathbb{I}(t_b^k < t < t_b^k + t_\omega^k) \end{aligned} \tag{8}$$

with $b_{nk}|\omega_k \sim \mathrm{Bernoulli}(\omega_k)$ and $n = 1,\ldots,N$. The binary matrix $\mathbf{S}$ of dimension $N\times K$, is a *feature potential* matrix. Each binary element S_{nk} indicates whether object n possesses feature f_k. S is a global variable and doesn't depend on time t. At any time t, the feature allocation matrix $Z(t)$ is a deterministic function of the current features present at t, that is $\{f_k : t_b^k < t < t_b^k + t_w^k, k = 1,\ldots,\infty\}$ and the feature potential matrix S, i.e. $Z_{nk}(t) = 1$ iff $S_{nk} = 1$ and $t_b^k < t < t_b^k + t_\omega^k$.

The resulting feature allocation process $(z_n(t))_\mathbb{T}$ is equivalent to the following: every time a new feature f_k is created, each object n joins with probability ω_k, i.e. $z_{nk}(t_b^k)|\omega_k \sim \mathrm{Bernoulli}(\omega_k)$. If $z_{nk}(t_b^k) = 1$, object n will possess feature f_k until $t_b^k + t_\omega^k$. Repeat this process for all objects.

Proposition 2. *The* BDFP *is exchangeable and the Beta process* $\mathrm{BP}(\alpha\mu)$ *on* $\mathbb{X}\otimes\mathbb{R}\otimes[0,\infty)$ *describes its underlying mixing measure.*

Proof. *Consider a sequence of variables* $(z_n(t))_\mathbb{T}$ *with* $n = 1,2,\ldots,N$ *such that each* $(z_n(t))_\mathbb{T}$ *is the feature allocation evolution of object* n *over the index space* $\mathbb{T}$. *These variables are not independent since each* $(z_n(t))_\mathbb{T}$ *depends on the* $Z_{|[n-1]}(t) = (z_{1:(n-1)}(t))_\mathbb{T}$. *However, given a sample from the* $B \sim \mathrm{BP}(\alpha\mu)$ *described in Section 4.1, each variable* $(z_n(t))_\mathbb{T}$ *becomes conditionally independent and the following holds*

$$P((z_1(t))_\mathbb{T},(z_2(t))_\mathbb{T},\ldots,(z_N(t))_\mathbb{T}) = \int\prod_{n=1}^N P((z_n(t))_\mathbb{T}|B)\phi(\mathrm{d}B) \tag{9}$$

where $\phi = \mathrm{BP}(\alpha\mu)$.

Equation (9) is the de Finetti representation of the BDFP and as such the BDFP is exchangeable and the BP on $\Theta = \mathbb{X}\otimes\mathbb{R}\otimes[0,\infty)$ is its underlying mixing measure. Restricting our focus on each index t, the overall Beta process $\mathrm{BP}(\alpha\mu)$ on $\mathbb{X}\otimes\mathbb{R}\otimes[0,\infty)$ results in a set of **dependent** random measures over $\mathbb{X}$, one B_t for each $t \in \mathbb{T}$, such that each B_t is marginally a Beta process. Consider a fixed time point $t \in \mathbb{T}$ and the space $[0,1]\otimes\mathbb{X}$ (the red vertical plane in Figure 1). The point process on this plane (where blue lines intersect the plane) corresponds to features alive at time t, i.e. $t \in [t_b, t_b + t_\omega]$. The Lévy measure on this plane, is calculated by projecting the overall Lévy measure onto the plane,

$$\begin{aligned} \nu_t(\mathrm{d}\omega\mathrm{d}x) &= \int_0^\infty\int_{t-t_\omega}^t \nu(\mathrm{d}\omega\mathrm{d}x\mathrm{d}t_b\mathrm{d}t_\omega) \\ &= \alpha\omega^{-1}(1-\omega)^{\alpha-1}\frac{\mu_0(\mathrm{d}x)}{D} \end{aligned} \tag{10}$$

where ν_t is a measure over $[0,1]\otimes\mathbb{X}$ for a specific $t \in \mathbb{T}$. More specifically, it is the Lévy measure of a Beta process on $\mathbb{X}$ with $\rho(\mathrm{d}\omega) = \alpha\omega^{-1}(1-\omega)^{\alpha-1}$ and base measure $\mu_t(\mathrm{d}x) = \frac{\mu_0(\mathrm{d}x)}{D}$. Thus we have that marginally

$$B_t|\alpha,\mu_t \sim \mathrm{BP}(\alpha\mu_t), \forall t \in \mathbb{T}. \tag{11}$$

The restricted and projected measure at any index $t \in \mathbb{T}$ defines a Beta process. Two draws, B_t and B_s, with $t,s \in \mathbb{T}$, will be dependent with the amount of dependence decreasing as $|s-t|$ increases.

Proposition 3. *The dependent Beta process construction presented has* IBP *marginals at any* t.

Proof. *At any $t \in \mathbb{T}$, $B_t|\alpha, \mu_t \sim \mathrm{BP}(\alpha\mu_t)$. It is straightforward to see that, marginally, the feature allocation matrix Z_t obtained using the generative process in Equation (8) is equivalent to $Z_t|B_t \sim \mathrm{BeP}(B_t)$ and therefore $Z_t \sim \mathrm{IBP}(\alpha)$, $\forall t \in \mathbb{T}$.*

Corollary 1. *At any $t \in \mathbb{T}$, the feature allocation matrix Z_t can be generated by the following generative model as $K \to \infty$:*

$$\omega_k|\alpha \sim \mathrm{Beta}\left(\frac{R}{K}\right), \quad Z_{nk}|\omega_k \sim \mathrm{Bernoulli}(\omega_k) \qquad (12)$$

for $k = 1, \ldots, K$ and $n = 1, \ldots, N$

The proof of the corollary in included in the supplementary material. Note that the above is true only marginally, i.e. at time $t \in \mathbb{T}$ and it doesn't generste dependence structure between Z_t's.

We underline the dependence of Z_s and Z_t when $|s-t| \to 0$, $\forall s, t \in \mathbb{T}$. The closer s, t are, the more the atoms (features) B_s and B_t share. If we independently sampled $Z_s|B_s \sim \mathrm{BeP}(B_s)$ and $Z_t|B_t \sim \mathrm{BeP}(B_t)$ then Z_s, Z_t would be dependent, but not equal, even as $|s-t| \to 0$. However, in the BDFP the presence of the same features results in the *same* (not just similar) allocation as $|s-t| \to 0$. In both cases, the marginal distribution of the feature allocation matrix at any $t \in \mathbb{T}$ is $Z_t|B_t \sim \mathrm{BeP}(B_t)$ and $Z_t|\alpha \sim \mathrm{IBP}(\alpha)$. The BDFP results in a continuous evolution of the $Z(t)$ over $\mathbb{T}$: formally $Z_t \xrightarrow{d} Z_s$ as $t \to s$.

This construction of the BDFP resembles the spatial normalised Gamma process (SNΓP) by (Rao & Teh, 2009). The main difference lies in the marginal distribution; the SNΓP admits DP marginals as opposed to the Beta process marginals of the dependent Beta process as shown in Equation (11).

Proposition 4. *The feature allocation process described by Equation (8) with $B \sim \mathrm{BP}(\alpha\mu)$, has the same birth and death rates as the* BDF *process.*

5. Finite Model

For the BDFP, the inference simplifies considerably if we consider a finite approximation which gives the countably infinite model in the limit. Consider the space $\mathbb{S} = [0,1] \otimes \mathbb{X} \otimes [0,T] \otimes [0,\infty)$, where we restrict the space of t_b to be $[0,T]$ instead of the whole real line $\mathbb{R}$. This accounts for typical applications of the model where we observe data at distinct times over a finite time range. Consider the Lévy measure $\nu(\mathrm{d}\omega \mathrm{d}x \mathrm{d}t_b \mathrm{d}t_\omega)$ on the space $\mathbb{S}$. Then, under the dependent Beta process representation (see section 4.1), the expected number of atoms present in $\mathbb{S}$ is $\int_{\mathbb{S}} \nu(\mathrm{d}\omega \mathrm{d}x \mathrm{d}t_b \mathrm{d}t_\omega) = \int_0^1 \rho(\mathrm{d}\omega) \int_{\mathbb{X}} \mu_0(\mathrm{d}x) \int_0^T g(\mathrm{d}t_b) \int_0^\infty \beta(\mathrm{d}t_\omega) = KT$, where $K \to \infty$ since $\int_0^1 \rho(\mathrm{d}\omega) = \infty$. By considering finite K we allow inference on a finite model which approximates the infinite case with increasing fidelity as $K \to \infty$.

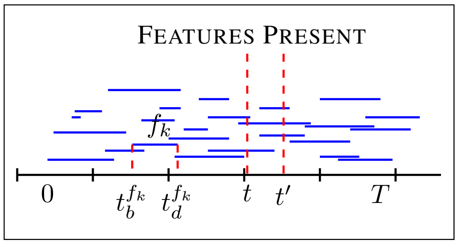


Figure 2. Cartoon for the Beta event construction of the BDFP: A $\mathrm{Poisson}(KT)$ number of features are uniformly distributed across the time range $[0,T]$ (blue lines). Each feature is assigned a weight sampled from $\mathrm{Beta}(\frac{R}{K}, 1)$. The leftmost point of each line corresponds to the time of birth of that feature, while the length of each line indicates the life span of each feature sampled from $\mathrm{Exponential}(D)$. To sample feature allocations from the process, we consider random time points across time, e.g. t, t' and draw imaginary red lines. The feature allocation matrix at t involves the features that are crossing the red line at t. The membership of the objects $n = 1, \ldots, N$ to those features is defined by the values of the corresponding elemnts in the potential matrix S.

The process is depicted in Figure 2 and the infinite case can be derived as the limit $K \to \infty$ of the following:

- Consider a time range $[0,T]$ and a set of features $\mathcal{F}$, such that $|\mathcal{F}| \sim \mathrm{Poisson}(KT)$. Assign to each feature $f_k \in \mathcal{F}$, $k = 1, \ldots |\mathcal{F}|$ a weight ω, such that $\omega_k \sim \mathrm{Beta}(\frac{R}{K}, 1)$ and $\mathbf{\Omega} = [\omega_1, \omega_2 \ldots \omega_{|\mathcal{F}|}]$.
- Associate each feature $f_k \in \mathcal{F}$, $k = 1, \ldots |\mathcal{F}|$ with a birth time t_b^k uniformly sampled in $[0,T]$; $t_b^k \sim \mathcal{U}(0,T)$ and $\mathbf{t}_b = [t_b^1 \ldots t_b^{|\mathcal{F}|}]$.
- For each $f_k \in \mathcal{F}$, sample its life span $t_w^k \sim \mathrm{Exponential}(D)$, where D is the death rate. Define the time of death t_d^k as $t_d^k = t_b^k + t_w^k$ and $\mathbf{t}_w = [t_w^1 \ldots t_w^{|\mathcal{F}|}]$.

We call the sequence of the above steps **Beta Event Process (BEP)**. Putting everything together, generate a sample $B = \{\mathcal{F}, \mathbf{\Omega}, \mathbf{t}_b, \mathbf{t}_w\} \sim \mathrm{BEP}(\alpha, R, K, T)$ as follows:

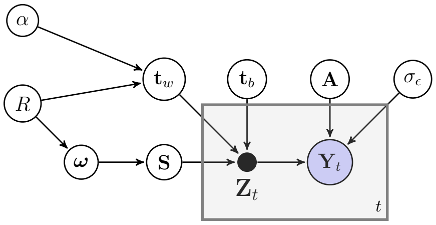
$$|\mathcal{F}| \sim \mathrm{Poisson}(KT)$$

$$\omega_k \sim \mathrm{Beta}\left(\frac{R}{K}, 1\right), t_b^k \sim \mathcal{U}(0,T), t_\omega^k \sim \mathrm{Exponential}(D) \qquad (13)$$

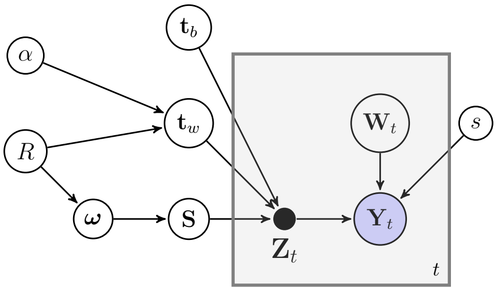
for $k = 1, \ldots, |\mathcal{F}|$. Having drawn a sample B from the BEP, we can construct the feature allocations over time as follows

$$\begin{aligned} S_{nk}|\omega_k &\sim \mathrm{Bernoulli}(\omega_k) \\ Z_{nk}(t) &= S_{nk}\mathbb{I}(t_b^k < t < t_b^k + t_\omega^k) \qquad (14) \end{aligned}$$

where $n = 1, \ldots, N$. The feature potential matrix (as defined in section 4.1) has now $N \times |\mathcal{F}|$ dimensions. Moreover, each $Z(t)$ for $t \in \mathbb{T}$ is a matrix of dimensions $N \times F^{(t)}$ and $F^{(t)} \leq |\mathcal{F}|$. Figure 3(a) show the graphical model for the BEP.



(a)



(b)

Figure 3. Graphical representation of the BEP for a time point t and for (a) a linear-Gaussian likelihood and (b) a sigmoid likelihood. The time series $\mathbf{Z}$ and $\mathbf{Y}$ are represented as single nodes indexed by the time location t. The birth and life span times of the total KT features are depicted using vector notation $\mathbf{t}_b$ and $\mathbf{t}_w$. The black ($\mathbf{Z}_t$) and grey ($\mathbf{Y}_t$) nodes indicate deterministic and observed parameters respectively.

Proposition 5. *In the finite model, the expected number of features present at any* $t \in \mathbb{T}$ *is* $\mathbb{E}[N_f] = \frac{K}{D}$ *and for* $D = \frac{R}{\alpha}$ *we have* $\mathbb{E}[N_f] = \frac{K\alpha}{R}$*.*

Hyperpriors. We put gamma priors on α and R.

Likelihood models. We consider two different likelihood models: *linear-Gaussian* for real data and *logistic* for binary network data.

For the linear-Gaussian likelihood model, consider a sequence of observations $\{Y_t \in \mathcal{Y} : t = 1, \ldots, L\}$ generated as

$$\mathbf{Y}_t = \mathbf{Z}_t\mathbf{A} + \epsilon_t \tag{15}$$

where $\mathbf{Y}_t$ is a $N \times M$ observation matrix at each time $t = 1, \ldots, L$, $\mathbf{A}$ is a factor loading matrix of dimension $|\mathcal{F}| \times M$ shared across time and $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is Gaussian white noise. We choose a Gaussian prior over $\mathbf{A}$, i.e $A_{fm} \sim \mathcal{N}(0, 1)$.

In the case of dynamic binary network data we extend the latent feature relational model (LFRM) proposed by (Miller et al., 2009). Let Y_t be the $N \times N$ binary matrix that contains links, i.e. $y_{tij} = \mathbf{Y}_t(i, j) = 1$ iff we observe a link from entity i to entity j at time t. We assume that the matrices $\mathbf{Y}_t$ are symmetric and ignore diagonal elements (self-links). The probability of a link from one entity to another is determined by the combined effect of all pairwise feature interactions. Let $\mathbf{W}_t$ be a $|\mathcal{F}| \times |\mathcal{F}|$ real-valued weight matrix where $W_t(k, k')$ is the weight that affects the probability of there being a link from entity i to entity j if entity i has feature k on, i.e. $Z_{tik} = Z_t(i, k) = 1$ and entity j has feature k' on, i.e. $Z_{tjk'} = Z_t(j, k') = 1$. The links are independent conditioned on $\mathbf{Z}_t$ and $\mathbf{W}_t$, and only the features that are on for the entities i and j at time t influence the probability of a link between those entities at that time (see Figure 3(b)). Formally,

$$P(y_{tij} = 1|\mathbf{Z}_t, \mathbf{W}_t) = \sigma\Big(\sum_{kl} Z_{tik}Z_{tjl}W_{tkl} + s\Big) \tag{16}$$

for $k, l = 1, \ldots, |\mathcal{F}|$, where s is a bias term and $\sigma(x) = (1 + e^{-x})^{-1}$ is the sigmoid function. For completeness, we assume the priors $w_t(k, l) \sim \mathcal{N}(\mu_w, \sigma_w^2)$ and $s \sim \mathcal{N}(\mu_s, \sigma_s^2)$.

5.1. Inference

As with many other Bayesian models, exact inference is intractable so we employ Markov Chain Monte Carlo (MCMC) for posterior inference over the latent variables of the finite model. A detailed description is provided in the supplementary material.

6. Experiments

We experimentally evaluate the BEP model on real-world genomics and social network data. To evaluate the model fit, we compared the BEP model to independent models at each time point.

6.1. Circadian Rhythm Dataset

Here we used a subset of the gene expression data from Piechota et al. (2010), including $N = 500$ genes in $D = 4$ different conditions (exposure to different drugs) over $L = 24$ time intervals. The measurements indicate how active a gene is at different times. We created 7 train-test splits holding out 20% of the data, and ran 700 MCMC iterations. We see that in terms of predictive performance the BEP outperforms independent IBP models (Table 1). The genes belonging to each factor show enrichment for different known biological pathways (Figure 4). Of particular note are the tryptophan metabolism genes enriched in factor 2, given tryptophan's suspected effects on drowsiness;

the vasopressin regulated water reabsorption, given this hormone's known circadian regulation (Earnest & Sladek, 1986; Yamaguchi et al., 2013); and the regulation of insulin producing beta cells, another hormone with circadian variation (Shi et al., 2013).

Table 1. Circadian dataset results using 20% held out data, a truncation level of $K = 10$, $|\mathcal{F}| = 24$, 700 iterations and a burnin of 500. Results are the average over 7 MCMC chains.

	BEP	INDEPENDENT IBP
TRAIN ERROR	$\mathbf{0.0917 \pm 0.0368}$	0.0983 ± 0.0012
TEST ERROR	$\mathbf{0.0948 \pm 0.0343}$	1.3380 ± 0.5155
TRAIN LOG LIKELIHOOD	6.508 ± 0.7715	$\mathbf{6.6871 \pm 0.0217}$
TEST LOG LIKELIHOOD	$\mathbf{1.5661 \pm 0.1583}$	-8.6861 ± 4.0670

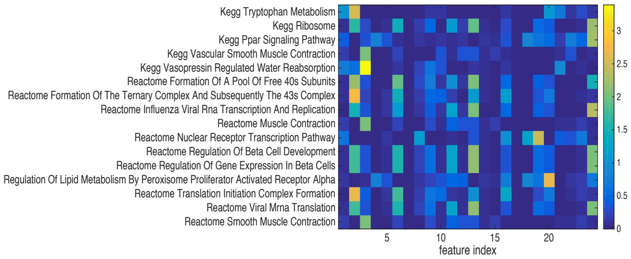


Figure 4. Circadian dataset: many of the features uncovered show enrichment in known biological pathways from Reactome and KEGG. Values here are $-\log_{10} p$ from a hypergeometric test for enrichment of the genes in each factor against the 500 background genes.

6.2. ChIP-seq Epigenetic Marks

For this experiment we used ChIP-seq (chromatin immunoprecipitation sequencing) data downloaded from the ENCODE project (Consortium, 2007), representing histone modifications and transcription factor binding in human neural crest cell lines (see (Park, 2009) for a nice review).

The observations involve counts associated with $N = 14$ (human) cell lines and $D = 10$ proteins. The counts indicate what proteins, with what chemical modifications, are bound to DNA along the genome. The measurements are stored in $N \times D$ matrix of counts Y_t: for each cell line, how many reads for each of the 10 proteins mapped to bin t (100 base pair (bp) region of the genome). $t = 1, \ldots, 500$ bins were considered at the start of chromosome 1 (50K bp in total). In Figure 5(a) each subfigure corresponds to one of the 10 proteins and in each subfigure the counts for the $N = 14$ cell lines are plotted over the genome section of length 50Kbp. Before inference, the raw counts were square-root transformed (a standard variance stabilizing transform for Poisson data) to make the Gaussian likelihood appropriate. We ran 7 different held-out tests, holding out a different 20% of the data each time. Results, using 700 MCMC iterations, are presented in Table 2. The BEP outperforms the independent IBP model in both test likelihood and error with a statistically significant difference. The independent IBP appears to have better results in train error and likelihood, again suggesting overfitting. Comparing the plots of the true measurements to the learnt ones by the BEP and independent IBP model in Figure 5 we see that both models successfully reproduce the data but the BEP reconstructions provide a cleaned up picture of the meaningful signal.

The features found by the model in the different genome locations correspond to different states associated with the specific genome location. Genes and regulatory DNA elements such as enhancers, silencers and insulators are embedded in genomes. These genomic elements on the DNA have footprints for the transacting proteins involved in transcription, either for the positioning or regulation of the transcriptional machinery. For instance, promoters are regions of DNA which recruit proteins required to initiate transcription of a particular gene and located near the transcription start sites. Enhancers are regions of DNA that can be bound by proteins which activate transcription of a distal gene. So a cell line, at specific genome location (recall that here each location corresponds to 100 base pairs), will have underlying feature membership (some promoters and some enhancer for example) that determines whether particular protein are found there using ChIP-seq.

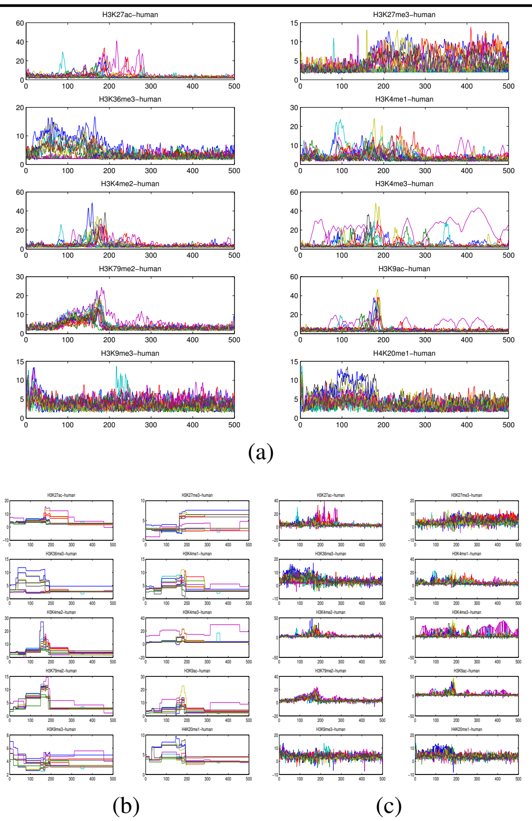
Genomic annotations, from ChromHMM (Ernst et al., 2011), are shown in Figure 8 in the supplementary document for the region we model. Different levels of the marks in these different regions are much easier to see in the reconstructed signal using BEP in Figure 5(b).

Table 2. Quantitative results for the ChIP-seq dataset . 20% held out data, a truncation level of $K = 3$, $|\mathcal{F}| = 21$, 700 iterations and a burnin of 500. Results are the average over 7 held out sets.

	BEP	INDEPENDENT IBP
TRAIN ERROR	0.4459 ± 0.0229	$\mathbf{0.032 \pm 0.0089}$
TEST ERROR	$\mathbf{0.4574 \pm 0.018}$	0.7746 ± 0.013
TRAIN LOG LIKELIHOOD	-12.4979 ± 0.1439	$\mathbf{-0.5916 \pm 0.0979}$
TEST LOG LIKELIHOOD	$\mathbf{-3.1666 \pm 0.0318}$	-175.7968 ± 4.49

7. van de Bunt's Dataset

In van de Bunt et al. (1999), 32 university freshman students in a given discipline at a Dutch university were surveyed at seven time points about who in their class they considered as friends. Initially, i.e. t_1, most of the students were unknown to each other. The first four time points are three weeks apart, whereas the last three time points are six weeks apart as showin in Figure 11 in the supplementary matrial. We symmetrise the matrix by assuming friendship if either individual reported it. We test the performance of BEP using the sigmoid likelihood model as in Equation



(a)

(b)

(c)

Figure 5. ChIP-seq data: The observed (a) and reconstructed observations (b), (c). The BEP reconstructions smooth out the noise making the meaning signal much easier to visualize. In both models, the noise signal was removed from the reconstructions.

(16) by holding out 10% of all links across all time points. We ran each model for 1000 MCMC iterations. The results are shown in Table 3. The independent network LFR models outperform BEP in the train setting and the test error while BEP outperforms in the test likelihood. However, here the results are comparable. Looking at Figure 6, both models provide the same picture of the allocation. It is possible the stationary assumption hurts the BEP: in the VDB dataset the number of links almost exclusively increases over time.

Table 3. van de Bunt's dataset results using 10% held out data, a truncation level of $K = 4$, $|\mathcal{F}| = 20$, 1000 iterations and a burnin of 200. Results are the average over 7 MCMC chains.

	BEP	INDEPENDENT LFRM
TRAIN ERROR	1.7009 ± 0.0850	$\mathbf{1.3413 \pm 0.1147}$
TEST ERROR	1.9107 ± 0.1321	$\mathbf{1.7891 \pm 0.1131}$
TRAIN LOG LIKELIHOOD	-1044.4943 ± 41.6363	$\mathbf{-839.4544 \pm 56.9877}$
TEST LOG LIKELIHOOD	$\mathbf{-345.7038 \pm 49.9882}$	-438.5848 ± 74.6396

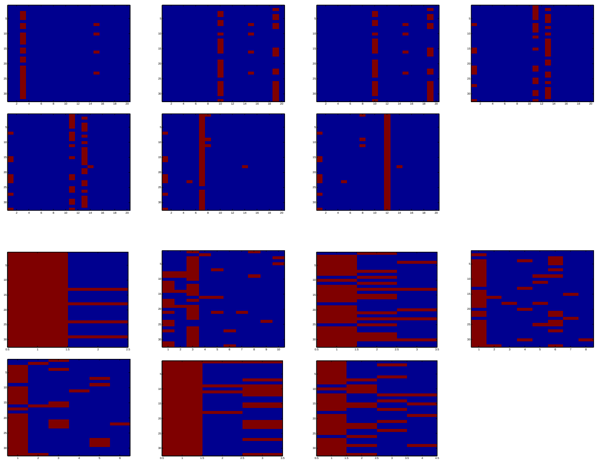


Figure 6. Inferred feature allocation matrices for the seven time points (from left to right) in the van De Bunt friendship dataset. **First two rows:** Feature allocation matrices inferred by BEP. **Last two rows:** Feature allocation matrices inferred by independent LFRM.

8. Discussion

Many modern machine learning and statistics tasks involve multidimensional data positioned along some linear covariate: we have shown functional genomics data where the covariate is position in the genome, and network data where links change over time. To model such data we need priors that utilize the dependencies through time, while handling high dimensionality. The BDFP is an expressive new Bayesian non-parametric prior that fulfills these criteria. It outputs time-evolving feature allocations, which can then be effectively used to model high-dimensional time-series data. Since the number of latent features is unbounded, like other Bayesian non-parametric methods, the model can adapt its complexity to the data. While the combinatorial BDFP may seem like a complex object to handle computationally, our theoretical results showing that the de Finetti measure underlying the BDFP is a specific beta process, which can be well approximated by a finite K model, the BEP. Our experimental results, compared to independent feature allocations, provides evidence that effectively modeling dependency in the feature allocation through the birth-death mechanism is appropriate for a wide range of statistical applications. Moreover, the BEP provides an interpretable structure using parameters not found, to the best of our knowledge, in existing models, i.e. birth and death rate of features. We are interested in scaling inference under the BEP to larger datasets, for example using (stochastic) variational inference methods that have been successful for the IBP (Doshi et al., 2009).

Acknowledgements Konstantina's research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617411.

References

Aldous, D J. Exchangeability and related topics. In *Ecole d'Ete de Probabilities de Saint-Flour*, volume XIII, pp. 1–198. Springer, 1983.

Consortium, The ENCODE Project. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816, 06 2007.

de Finetti, B. *Funzione Caratteristica Di un Fenomeno Aleatorio*, pp. 251–299. 6. Memorie. Academia Nazionale del Linceo, 1931.

Doshi, F., Miller, K. T., Van Gael, J., and Teh, Y. W. Variational inference for the Indian buffet process. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, volume 12, 2009.

Earnest, David J and Sladek, Celia D. Circadian rhythms of vasopressin release from individual rat suprachiasmatic explants in vitro. *Brain research*, 382(1):129–133, 1986.

Ernst, Jason, Kheradpour, Pouya, Mikkelsen, Tarjei S, Shoresh, Noam, Ward, Lucas D, Epstein, Charles B, Zhang, Xiaolan, Wang, Li, Issner, Robbyn, Coyne, Michael, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, 2011.

Ghahramani, Zoubin and Jordan, Michael. Factorial hidden markov models. *Machine Learning*, 29(2-3):245–273, 1997.

Griffiths, Thomas L. and Ghahramani, Zoubin. Infinite latent feature models and the indian buffet process. In *In NIPS*, pp. 475–482. MIT Press, 2005.

Griffiths, Thomas L. and Ghahramani, Zoubin. The indian buffet process: An introduction and review. *Journal of Machine Learning Research*, 12:1185–1224, July 2011.

Hjort, N. L. Nonparametric Bayes estimators based on Beta processes in models for life history data. *Annals of Statistics*, 18: 1259–1294, 1990.

Kingman, J.F.C. The coalescent. *Stochastic Processes and their Applications*, 13(3):235 – 248, 1982.

Kingman, John F. C. Completely Random Measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

Miller, Kurt, Griffiths, Thomas, and Jordan, Michael. Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems 22*, 2009.

Neal, Radford M. Density Modeling and Clustering Using Dirichlet Diffusion Trees. In *Bayesian Statistics 7*, pp. 619–629, 2003.

Palla, Konstantina, Knowles, David A., and Ghahramani, Zoubin. A dependent partition-valued process for multitask clustering and time evolving network modelling, 2013.

Park, Peter J. Chip–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*, 10(10):669–680, 2009.

Piechota, Marcin, Korostynski, Michal, Solecki, Wojciech, Gieryk, Agnieszka, Slezak, Michal, Bilecki, Wiktor, Ziolkowska, Barbara, Kostrzewa, Elzbieta, Cymerman, Iwona, Swiech, Lukasz, et al. The dissection of transcriptional modules regulated by various drugs of abuse in the mouse striatum. *Genome Biology*, 11(5):R48, 2010.

Rao, Vinayak and Teh, Yee Whye. Spatial normalized gamma processes. 2009.

Rasmussen, Carl Edward and Williams, Christopher K I. *Gaussian processes for machine learning*. MIT Press, 2006.

Shi, Shu-qun, Ansari, Tasneem S, McGuinness, Owen P, Wasserman, David H, and Johnson, Carl Hirschie. Circadian disruption leads to insulin resistance and obesity. *Current Biology*, 23(5):372–381, 2013.

Teh, Y. W., Elliott, L. T., and Blundell, C. Bayesian nonparametric modelling of genetic variations using fragmentation-coagulation processes. Submitted, 2013.

Thibaux, Romain and Jordan, Michael I. Hierarchical beta processes and the indian buffet process. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 2, pp. 564–571, 2007.

van de Bunt, Gerhard G, Van Duijn, Marijtje AJ, and Snijders, Tom AB. Friendship networks through time: An actor-oriented dynamic statistical network model. *Computational & Mathematical Organization Theory*, 5:167–192, 1999.

Van Gael, J., Teh, Y. W., and Ghahramani, Z. The infinite factorial hidden Markov model. In *Advances in Neural Information Processing Systems*, volume 21, 2009.

Williamson, Sinead, Orbanz, Peter, and Ghahramani, Zoubin. Dependent Indian buffet processes. In *Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics, AISTATS*, 2010.

Yamaguchi, Yoshiaki, Suzuki, Toru, Mizoro, Yasutaka, Kori, Hiroshi, Okada, Kazuki, Chen, Yulin, Fustin, Jean-Michel, Yamazaki, Fumiyoshi, Mizuguchi, Naoki, Zhang, Jing, et al. Mice genetically deficient in vasopressin v1a and v1b receptors are resistant to jet lag. *Science*, 342(6154):85–90, 2013.