encouraging a sparse, finite representation and flexible sharing among time series. The inherent conjugacy of the BP to the Bernoulli process allows for an analytic predictive distribution for a feature vector based on the feature vectors observed so far. As outlined in Section 6.1, this predictive distribution can be described via the Indian buffet process [Ghahramani, Griffiths and Sollich (2006)] under certain parameterizations of the BP. Computationally, this representation is key.

The beta process—Bernoulli process featural model. The BP is a special case of a general class of stochastic processes known as *completely random measures* [Kingman (1967)]. A completely random measure B is defined such that for any disjoint sets A_1 and A_2 (of some sigma algebra $\mathcal{A}$ on a measurable space Θ), the corresponding random variables $B(A_1)$ and $B(A_2)$ are independent. This idea generalizes the family of *independent increments processes* on the real line. All completely random measures can be constructed from realizations of a nonhomogenous Poisson process [up to a deterministic component; see Kingman (1967)]. Specifically, a Poisson rate measure ν is defined on a product space $\Theta \otimes \mathbb{R}$, and a draw from the specified Poisson process yields a collection of points $\{\theta_j, \omega_j\}$ that can be used to define a completely random measure:

$$B = \sum_{k=1}^{\infty} \omega_k \delta_{\theta_k}. \tag{5}$$

This construction assumes ν has infinite mass, yielding a countably infinite collection of points from the Poisson process. Equation (5) shows that completely random measures are discrete. Consider a rate measure defined as the product of an arbitrary sigma-finite *base measure* B_0, with total mass $B_0(\Theta) = \alpha$, and an improper beta distribution on the interval [0, 1]. That is, on the product space $\Theta \otimes [0, 1]$ we have the following rate measure:

$$\nu(d\omega, d\theta) = c\omega^{-1}(1-\omega)^{c-1}\, d\omega\, B_0(d\theta), \tag{6}$$

where $c > 0$ is referred to as a *concentration parameter*. The resulting completely random measure is known as the *beta process*, with draws denoted by $B \sim \mathrm{BP}(c, B_0)$. With this construction, the weights ω_k of the atoms in B lie in the interval (0, 1), thus defining our desired feature-inclusion probabilities.

The BP is conjugate to a class of *Bernoulli processes* [Thibaux and Jordan (2007)], denoted by $\mathrm{BeP}(B)$, which provide our desired feature representation. A realization

$$X_i \mid B \sim \mathrm{BeP}(B), \tag{7}$$

with B an atomic measure, is a collection of unit-mass atoms on Θ located at some subset of the atoms in B. In particular, $f_{ik} \sim \mathrm{Bernoulli}(\omega_k)$ is sampled independently for each atom θ_k in B, and then

$$X_i = \sum_k f_{ik} \delta_{\theta_k}. \tag{8}$$