6.1. *Background*: *The Indian buffet process*. Sampling the features $\mathbf{F}$ requires some prerequisite knowledge. As shown by Thibaux and Jordan (2007), marginalizing over the latent beta process B in the beta process-Bernoulli process hierarchy and taking $c = 1$ induces a predictive distribution on feature indicators known as the Indian buffet process (IBP) [Ghahramani, Griffiths and Sollich (2006)].[4] The IBP is based on a culinary metaphor in which customers arrive at an infinitely long buffet line of dishes (*features*). The first arriving customer (*time series*) chooses Poisson(α) dishes. Each subsequent customer i selects a previously tasted dish k with probability m_k/i proportional to the number of previous customers m_k to sample it, and also samples Poisson(α/i) new dishes.

For a detailed derivation of the IBP from the beta process-Bernoulli process formulation of Section 4.1, see Supplement A of Fox et al. (2014).

6.2. *Sampling shared feature assignments*. We now consider sampling each sequence's binary feature assignment $\mathbf{f}_i$. Let $\mathbf{F}^{-ik}$ denote the set of all feature indicators excluding f_{ik}, and K_+^{-i} be the number of behaviors used by all other time series. Some of the K_+^{-i} features may also be shared by time series i, but those unique to this series are not included. For simplicity, we assume that these behaviors are indexed by $\{1, \dots, K_+^{-i}\}$. The IBP prior differentiates between this set of "shared" features that other time series have already selected and those "unique" to the current sequence and appearing nowhere else. We may safely alter sequence i's assignments to shared features $\{1, \dots, K_+^{-i}\}$ without changing the number of behaviors present in $\mathbf{F}$. We give a procedure for sampling these entries below. Sampling unique features requires adding or deleting features, which we cover in Section 6.6.

Given observed data $\mathbf{y}_{1:T_i}^{(i)}$, transition variables $\boldsymbol{\eta}^{(i)}$, and emission parameters $\boldsymbol{\theta}$, the feature indicators f_{ik} for the ith sequence's shared features $k \in \{1, \dots, K_+^{-i}\}$ have posterior distribution

$$p(f_{ik} | \mathbf{F}^{-ik}, \mathbf{y}_{1:T_i}^{(i)}, \boldsymbol{\eta}^{(i)}, \boldsymbol{\theta}) \propto p(f_{ik} | \mathbf{F}^{-ik}) p(\mathbf{y}_{1:T_i}^{(i)} | \mathbf{f}_i, \boldsymbol{\eta}^{(i)}, \boldsymbol{\theta}). \tag{14}$$

Here, the IBP prior implies that $p(f_{ik} = 1 | \mathbf{F}^{-ik}) = m_k^{-i}/N$, where m_k^{-i} denotes the number of sequences *other* than i possessing k. This exploits the exchangeability of the IBP [Ghahramani, Griffiths and Sollich (2006)], which follows from the BP construction [Thibaux and Jordan (2007)].

When sampling binary indicators like f_{ik}, Metropolis–Hastings proposals can mix faster [Frigessi et al. (1993)] and have greater efficiency [Liu (1996)] than standard Gibbs samplers. To update f_{ik} given $\mathbf{F}^{-ik}$, we thus use equation (14) to evaluate a Metropolis–Hastings proposal which flips f_{ik} to the binary complement

[4] Allowing any $c > 0$ induces a two-parameter IBP with a similar construction.