One can visualize this process as walking along the atoms of a discrete measure B and, at each atom θ_k, flipping a coin with probability of heads given by ω_k. Since the rate measure ν is σ-finite, Campbell's theorem [Kingman (1993)] guarantees that for α finite, B has finite expected measure resulting in a finite set of "heads" (active features) in each X_i.

Computationally, Bernoulli process realizations X_i are often summarized by an infinite vector of binary indicator variables $\mathbf{f}_i = [f_{i1}, f_{i2}, \ldots]$. Using the BP measure B to tie together the feature vectors encourages the X_i to share similar features while still allowing significant variability.

4.2. *Feature-constrained transition distributions.* We seek a prior for transition distributions $\boldsymbol{\pi}^{(i)} = \{\pi_k^{(i)}\}$ defined on an infinite-dimensional state space, but with positive support restricted to a finite subset specified by $\mathbf{f}_i$. Motivated by the fact that Dirichlet-distributed probability mass functions can be generated via normalized gamma random variables, for each time series i we define a doubly-infinite collection of random variables:

$$\eta_{jk}^{(i)} | \gamma, \kappa \sim \operatorname{Gamma}(\gamma + \kappa\delta(j,k), 1). \tag{9}$$

Here, the Kronecker delta function is defined by $\delta(j,k) = 0$ when $j \neq k$ and $\delta(k,k) = 1$. The hyperparameters γ, κ govern Markovian state switching probabilities. Using this collection of *transition weight* variables, denoted by $\boldsymbol{\eta}^{(i)}$, we define time-series-specific, feature-constrained transition distributions:

$$\pi_j^{(i)} = \frac{[\eta_{j1}^{(i)} \quad \eta_{j2}^{(i)} \quad \cdots] \odot \mathbf{f}_i}{\sum_{k|f_{ik}=1} \eta_{jk}^{(i)}}, \tag{10}$$

where $\odot$ denotes the element-wise, or Hadamard, vector product. This construction defines $\pi_j^{(i)}$ over the full set of positive integers, but assigns positive mass only at indices k where $f_{ik} = 1$, constraining time series i to only transition among behaviors indicated by its feature vector $\mathbf{f}_i$. See Figure 2.

The preceding generative process can be equivalently represented via a sample $\tilde{\pi}_j^{(i)}$ from a finite Dirichlet distribution of dimension $K_i = \sum_k f_{ik}$, containing the nonzero entries of $\pi_j^{(i)}$:

$$\tilde{\pi}_j^{(i)} | \mathbf{f}_i, \gamma, \kappa \sim \operatorname{Dir}([\gamma, \ldots, \gamma, \gamma + \kappa, \gamma, \ldots, \gamma]). \tag{11}$$

This construction reveals that κ places extra expected mass on the self-transition probability of each state, analogously to the sticky HDP-HMM [Fox et al. (2011b)]. We also use the representation

$$\pi_j^{(i)} | \mathbf{f}_i, \gamma, \kappa \sim \operatorname{Dir}([\gamma, \ldots, \gamma, \gamma + \kappa, \gamma, \ldots] \odot \mathbf{f}_i), \tag{12}$$

implying $\pi_j^{(i)} = [\pi_{j1}^{(i)} \quad \pi_{j2}^{(i)} \quad \cdots]$ has only a finite number of nonzero entries $\pi_{jk}^{(i)}$. This representation is an abuse of notation since the Dirichlet distribution is not defined for infinitely many parameters. However, the notation of