or eliminate one of the existing features in $\boldsymbol{f}_{+i}$. Some previous work has applied RJMCMC to finite binary feature models [3, 27], but not to the IBP. Our proposal distribution factors as follows:

$$q(\boldsymbol{f}'_{+i}, \boldsymbol{\theta}'_+, \boldsymbol{\eta}'_+ \mid \boldsymbol{f}_{+i}, \boldsymbol{\theta}_+, \boldsymbol{\eta}_+) = q_f(\boldsymbol{f}'_{+i} \mid \boldsymbol{f}_{+i}) q_\theta(\boldsymbol{\theta}'_+ \mid \boldsymbol{f}'_{+i}, \boldsymbol{f}_{+i}, \boldsymbol{\theta}_+) q_\eta(\boldsymbol{\eta}'_+ \mid \boldsymbol{f}'_{+i}, \boldsymbol{f}_{+i}, \boldsymbol{\eta}_+). \quad (13)$$

Let $n_i = \sum_k f_{+ik}$. The feature proposal $q_f(\cdot \mid \cdot)$ encodes the probabilities of birth and death moves: a new feature is created with probability 0.5, and each of the n_i existing features is deleted with probability $0.5/n_i$. For parameters, we define our proposal using the generative model:

$$q_\theta(\boldsymbol{\theta}'_+ \mid \boldsymbol{f}'_{+i}, \boldsymbol{f}_{+i}, \boldsymbol{\theta}_+) = \begin{cases} b_0(\theta'_{+,n_i+1}) \prod_{k=1}^{n_i} \delta_{\theta_{+k}}(\theta'_{+k}), & \text{birth of feature } n_i+1; \\ \prod_{k \neq \ell} \delta_{\theta_{+k}}(\theta'_{+k}), & \text{death of feature } \ell, \end{cases} \quad (14)$$

where b_0 is the density associated with $\alpha^{-1}B_0$. The distribution $q_\eta(\cdot \mid \cdot)$ is defined similarly, but using the gamma prior on transition variables of Eq. (7). The MH acceptance probability is then

$$\rho(\boldsymbol{f}'_{+i}, \boldsymbol{\theta}'_+, \boldsymbol{\eta}'_+ \mid \boldsymbol{f}_{+i}, \boldsymbol{\theta}_+, \boldsymbol{\eta}_+) = \min\{r(\boldsymbol{f}'_{+i}, \boldsymbol{\theta}'_+, \boldsymbol{\eta}'_+ \mid \boldsymbol{f}_{+i}, \boldsymbol{\theta}_+, \boldsymbol{\eta}_+), 1\}. \quad (15)$$

Canceling parameter proposals with corresponding prior terms, the acceptance ratio $r(\cdot \mid \cdot)$ equals

$$\frac{p(\mathbf{y}^{(i)}_{1:T_i} \mid [\boldsymbol{f}_{-i}\, \boldsymbol{f}'_{+i}], \theta_{1:K_+^{-i}}, \boldsymbol{\theta}'_+, \boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}'_+) \text{ Poisson}(n'_i \mid \alpha/N)\; q_f(\boldsymbol{f}_{+i} \mid \boldsymbol{f}'_{+i})}{p(\mathbf{y}^{(i)}_{1:T_i} \mid [\boldsymbol{f}_{-i}\, \boldsymbol{f}_{+i}], \theta_{1:K_+^{-i}}, \boldsymbol{\theta}_+, \boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}_+) \text{ Poisson}(n_i \mid \alpha/N)\; q_f(\boldsymbol{f}'_{+i} \mid \boldsymbol{f}_{+i})}, \quad (16)$$

with $n'_i = \sum_k f'_{+ik}$. Because our birth and death proposals do not modify the values of existing parameters, the Jacobian term normally arising in RJMCMC algorithms simply equals one.

4.2 Sampling dynamic parameters and transition variables

Posterior updates to transition variables $\boldsymbol{\eta}^{(i)}$ and shared dynamic parameters θ_k are greatly simplified if we instantiate the mode sequences $z^{(i)}_{1:T_i}$ for each object i. We treat these mode sequences as *auxiliary variables*: they are sampled given the current MCMC state, conditioned on when resampling model parameters, and then discarded for subsequent updates of feature assignments $\boldsymbol{f}_i$.

Given feature-constrained transition distributions $\boldsymbol{\pi}^{(i)}$ and dynamic parameters $\{\theta_k\}$, along with the observation sequence $\mathbf{y}^{(i)}_{1:T_i}$, we *jointly* sample the mode sequence $z^{(i)}_{1:T_i}$ by computing backward messages $m_{t+1,t}(z^{(i)}_t) \propto p(\mathbf{y}^{(i)}_{t+1:T_i} \mid z^{(i)}_t, \tilde{\mathbf{y}}^{(i)}_t, \boldsymbol{\pi}^{(i)}, \{\theta_k\})$, and then recursively sampling each $z^{(i)}_t$:

$$z^{(i)}_t \mid z^{(i)}_{t-1}, \mathbf{y}^{(i)}_{1:T_i}, \boldsymbol{\pi}^{(i)}, \{\theta_k\} \sim \pi^{(i)}_{z^{(i)}_{t-1}}(z^{(i)}_t) \mathcal{N}\big(\mathbf{y}^{(i)}_t; \boldsymbol{A}_{z^{(i)}_t} \tilde{\mathbf{y}}^{(i)}_t, \Sigma_{z^{(i)}_t}\big) m_{t+1,t}(z^{(i)}_t). \quad (17)$$

Because Dirichlet priors are conjugate to multinomial observations $z^{(i)}_{1:T}$, the posterior of $\pi^{(i)}_j$ is

$$\pi^{(i)}_j \mid \boldsymbol{f}_i, z^{(i)}_{1:T}, \gamma, \kappa \sim \text{Dir}([\gamma + n^{(i)}_{j1}, \dots, \gamma + n^{(i)}_{jj-1}, \gamma + \kappa + n^{(i)}_{jj}, \gamma + n^{(i)}_{jj+1}, \dots] \otimes \boldsymbol{f}_i). \quad (18)$$

Here, $n^{(i)}_{jk}$ are the number of transitions from mode j to k in $z^{(i)}_{1:T}$. Since the mode sequence $z^{(i)}_{1:T}$ is generated from feature-constrained transition distributions, $n^{(i)}_{jk}$ is zero for any k such that $f_{ik} = 0$. Thus, to arrive at the posterior of Eq. (18), we only update $\eta^{(i)}_{jk}$ for instantiated features:

$$\eta^{(i)}_{jk} \mid z^{(i)}_{1:T}, \gamma, \kappa \sim \text{Gamma}(\gamma + \kappa\delta(j,k) + n^{(i)}_{jk}, 1), \quad k \in \{\ell \mid f_{i\ell} = 1\}. \quad (19)$$

We now turn to posterior updates for dynamic parameters. We place a conjugate matrix-normal inverse-Wishart (MNIW) prior [26] on $\{\boldsymbol{A}_k, \Sigma_k\}$, comprised of an inverse-Wishart prior $\text{IW}(S_0, n_0)$ on Σ_k and a matrix-normal prior $\mathcal{MN}(\boldsymbol{A}_k; M, \Sigma_k, K)$ on $\boldsymbol{A}_k$ given Σ_k. We consider the following sufficient statistics based on the sets $\boldsymbol{Y}_k = \{\mathbf{y}^{(i)}_t \mid z^{(i)}_t = k\}$ and $\tilde{\boldsymbol{Y}}_k = \{\tilde{\mathbf{y}}^{(i)}_t \mid z^{(i)}_t = k\}$ of observations and lagged observations, respectively, associated with behavior k:

$$S^{(k)}_{\tilde{y}\tilde{y}} = \sum_{(t,i) \mid z^{(i)}_t = k} \tilde{\mathbf{y}}^{(i)}_t \tilde{\mathbf{y}}^{(i)^T}_t + \boldsymbol{K} \qquad S^{(k)}_{y\tilde{y}} = \sum_{(t,i) \mid z^{(i)}_t = k} \mathbf{y}^{(i)}_t \tilde{\mathbf{y}}^{(i)^T}_t + \boldsymbol{M}\boldsymbol{K}$$

$$S^{(k)}_{yy} = \sum_{(t,i) \mid z^{(i)}_t = k} \mathbf{y}^{(i)}_t \mathbf{y}^{(i)^T}_t + \boldsymbol{M}\boldsymbol{K}\boldsymbol{M}^T \qquad S^{(k)}_{y|\tilde{y}} = S^{(k)}_{yy} - S^{(k)}_{y\tilde{y}} S^{-(k)}_{\tilde{y}\tilde{y}} S^{(k)^T}_{\tilde{y}\tilde{y}}.$$

Following Fox et al. [5], the posterior can then be shown to equal

$$\boldsymbol{A}_k \mid \Sigma_k, \boldsymbol{Y}_k \sim \mathcal{MN}\left(\boldsymbol{A}_k; S^{(k)}_{y\tilde{y}} S^{-(k)}_{\tilde{y}\tilde{y}}, \Sigma_k, S^{(k)}_{\tilde{y}\tilde{y}}\right), \quad \Sigma_k \mid \boldsymbol{Y}_k \sim \text{IW}\left(S^{(k)}_{y|\tilde{y}} + S_0, |\boldsymbol{Y}_k| + n_0\right).$$