The reversibility constraint can limit the effectiveness of our proposal framework. Even when a proposed configuration Ψ^* results in better joint probability, its Hastings factor can be small enough to cause rejection. For example, consider any merge proposal. Reversing this merge requires returning to the original configuration of the feature matrix $\mathbf{F}$ via a split proposal. Ignoring anchor sequence constraints for simplicity, split moves can produce roughly $3^{|\mathcal{S}|}$ possible feature matrices, since each sequence in the active set $\mathcal{S}$ could have its new features k_a, k_b set to [0 1], [1 0], or [1 1]. Returning to the exact original feature matrix out of the many possibilities can be very unlikely. Even though our proposals use data wisely, the vast space of possible split configurations means the Hastings factor will always be biased toward rejection of a merge move.

As a remedy, we recommend *annealing* the Hastings factor in the acceptance ratio of both split-merge and data-driven birth–death moves. That is, we use a modified acceptance ratio

$$\rho = \frac{p(\mathbf{y}, \Psi^*)}{p(\mathbf{y}, \Psi)} \left[\frac{q(\Psi | \Psi^*, \mathbf{y})}{q(\Psi^* | \Psi, \mathbf{y})} \right]^{1/T_s}, \tag{29}$$

where T_s indicates the "temperature" at iteration s. We start with a temperature that is very large, so that $\frac{1}{T_s} \approx 0$ and the Hastings factor is ignored. The resulting greedy stochastic search allows rapid improvement from the initial configuration. Over many iterations, we gradually decrease the temperature toward 1. After a specified number of iterations we fix $\frac{1}{T_s} = 1$, so that the Hastings factor is fully represented and the sampler is reversible.

In practice, we use an annealing schedule that linearly interpolates $\frac{1}{T_s}$ between 0 and 1 over the first several thousand iterations. Our experiments in Section 9 demonstrate improvement in mixing rates based on this annealing.

8. Related work. Defining the number of dynamic regimes presents a challenging problem in deploying Markov switching processes such as the AR-HMM. Previously, Bayesian nonparametric approaches building on the hierarchical Dirichlet process (HDP) [Teh et al. (2006)] have been proposed to allow uncertainty in the number of regimes by defining Markov switching processes on infinite state spaces [Beal, Ghahramani and Rasmussen (2001), Teh et al. (2006), Fox et al. (2011a, 2011b)]. See Fox et al. (2010) for a recent review. However, these formulations focus on a single time series, whereas in this paper our motivation is analyzing *multiple* time series. A naïve approach to this setting is to simply couple all time series under a shared HDP prior. However, this approach assumes that the state spaces of the multiple Markov switching processes are *exactly* shared, as are the transitions among these states. As demonstrated in Section 9 as well as our extensive toy data experiments in Supplement H of Fox et al. (2014), such strict sharing can limit the ability to discover unique dynamic behaviors and reduces predictive performance.