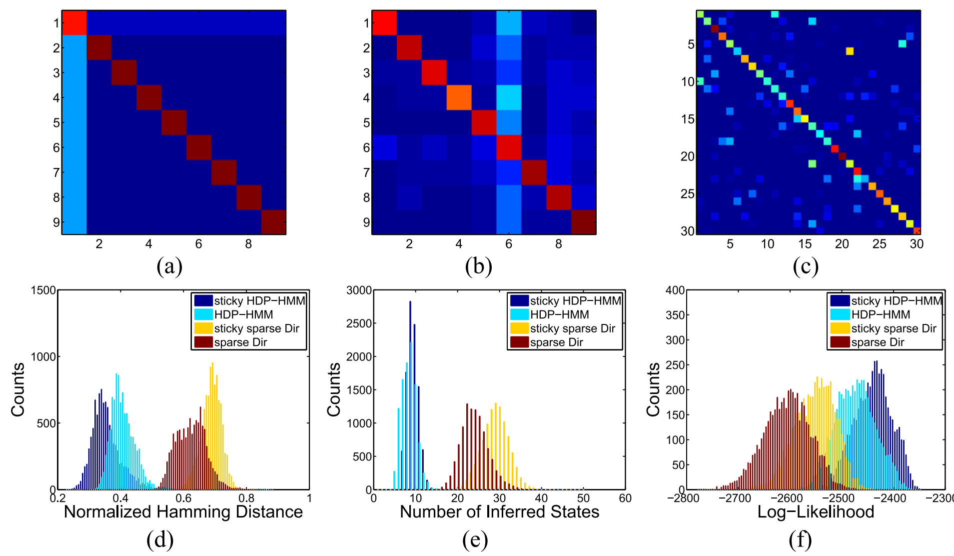

FIG. 10. (a) *The true transition probability matrix* (*TPM*) *associated with the state transition diagram of Figures* 9. (b) *and* (c) *The inferred TPM at the* 30,000*th Gibbs iteration for the sticky HDP-HMM and sticky sparse Dirichlet model*, *respectively*, *only examining those states with more than* 1% *of the assignments*. *For the HDP-HMM and sparse Dirichlet model*, *with and without the sticky parameter*, *we plot*: (d) *the Hamming distance error over* 10,000 *Gibbs iterations*, (e) *the inferred number of states with more than* 1% *of the assignments*, *and* (f) *the predictive probability of test sequences using the inferred parameters sampled every* 100*th iteration from Gibbs iterations* 5000–10,000.

Note that the results of Figure 10(f) also motivate the use of the sticky parameter in the more classical setting of a finite HMM with a standard Dirichlet sparsity prior. A motivating example of the use of sparse Dirichlet priors for finite HMMs is presented in Johnson (2007).

7. Multimodal emission densities. In many application domains, the data associated with each hidden state may have a complex, multimodal distribution. We propose to model such emission distributions nonparametrically, using a DP mixture of Gaussians. This formulation is related to the nested DP [Rodriguez, Dunson and Gelfand (2008)], which uses a Dirichlet process to partition data into groups, and then models each group via a Dirichlet process mixture. The bias toward self-transitions allows us to distinguish between the underlying HDP-HMM states. If the model were free to both rapidly switch between HDP-HMM states and associate multiple Gaussians per state, there would be considerable posterior uncertainty. Thus, it is only with the sticky HDP-HMM that we can effectively fit such models.

We augment the HDP-HMM state z_t with a term s_t indexing the mixture component of the z_tth emission density. For each HDP-HMM state, there is a unique