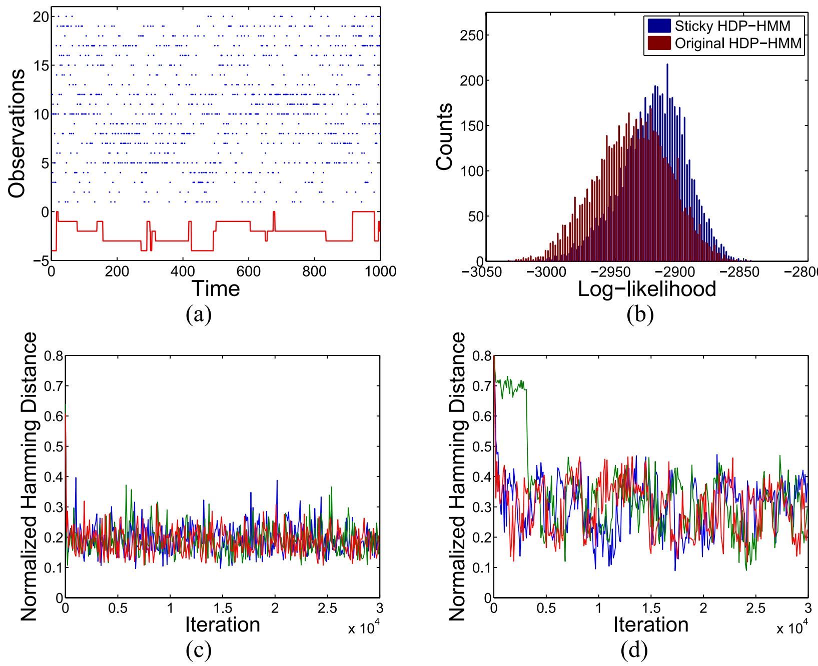

FIG. 8. (a) *Observation sequence* (*blue*) *and true state sequence* (*red*) *for a five-state HMM with multinomial observations*. (b) *Histogram of the predictive probability of test sequences using the inferred parameters sampled every* 100*th iteration from Gibbs iterations* 10,000–30,000 *for the sticky and original HDP-HMM*. *The Hamming distances over* 30,000 *Gibbs samples from three chains are shown for the* (c) *sticky HDP-HMM and* (d) *original HDP-HMM*.

In some applications, such as the speaker diarization problem that is explored in Section 8, one cares about the inferred segmentation of the data into a set of state labels. In this case, the advantage of incorporating the sticky parameter is clear. However, it is often the case that the metric of interest is the predictive power of the fitted model, not the accuracy of the inferred state sequence. To study performance under this metric, we simulated 10 test sequences using the same parameters that generated the training sequence. We then computed the likelihood of each of the test sequences under the set of parameters inferred at every 100th Gibbs iteration from iterations 10,000–30,000. This likelihood was computed by running the forward–backward algorithm of Rabiner (1989). We plot these results as a histogram in Figure 8(b). From this plot, we see that the fragmentation of data into redundant HMM states can also degrade the predictive performance of the inferred model. Thus, the sticky parameter plays an important role in the Bayesian nonparametric learning of HMMs even in terms of model averaging.

6.3. *Comparison to independent sparse Dirichlet prior.* We have alluded to the fact that the *shared* sparsity of the HDP-HMM induced by β is essential for