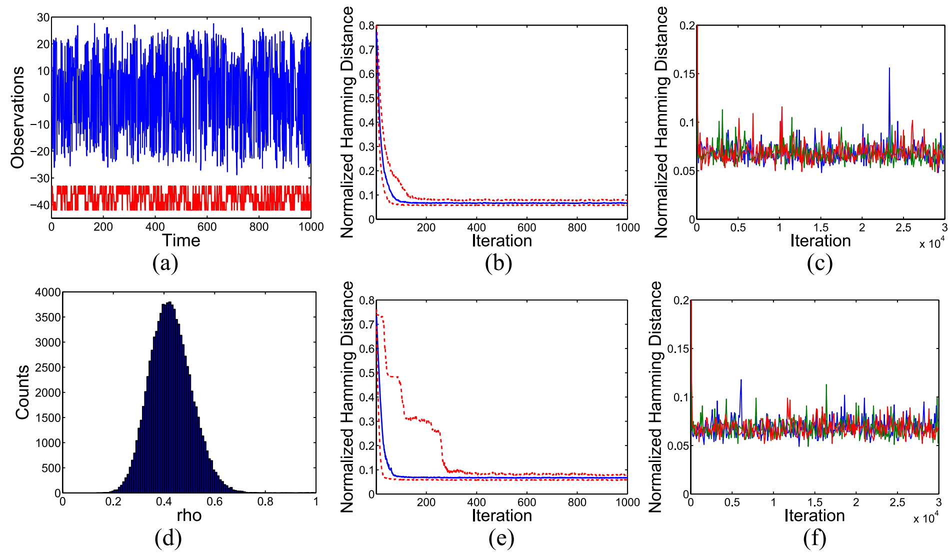

FIG. 7. (a) *Observation sequence* (*blue*) *and true state sequence* (*red*) *for a four-state HMM with fast state switching*. *For the original HDP-HMM using a blocked Gibbs sampler*: (b) *the median* (*solid blue*) *and* 10*th and* 90*th quantiles* (*dashed red*) *of Hamming distance between the true and estimated state sequences over the first* 1000 *Gibbs samples from* 200 *chains*, *and* (*c*) *Hamming distance over* 30,000 *Gibbs samples from three chains*. (d) *Histogram of the inferred self-transition parameter*, ρ, *for the sticky HDP-HMM blocked sampler*. (e) *and* (f) *Analogous plots to* (b) *and* (c) *for the sticky HDP-HMM*.

is no concept of similarity between nonidentical observation values. In contrast, Gaussian emissions have a continuous range of values in $\mathbb{R}^n$ with a clear notion of *closeness* between observations under the Lebesgue measure, aiding in grouping observations under a single HMM state's Gaussian emission distribution, even in the absence of a self-transition bias.

To demonstrate the increased posterior uncertainty with discrete observations, we generated data from a five-state HMM with multinomial emissions with a 0.98 probability of self-transition and equal probability of transitions to the other four states. The vocabulary, or range of possible observation values, was set to 20. The observation and true state sequences are shown in Figure 8(a). We placed a symmetric Dirichlet prior on the parameters of the multinomial distribution, with the Dirichlet hyperparameters equal to 2 [i.e., $\mathrm{Dir}(2, \dots, 2)$].

From Figure 8, we see that even after burn-in, many fast-switching state sequences have significant posterior probability under the nonsticky model, leading to sweeps through regions of larger Hamming distance error. A qualitative plot of one such inferred sequence after 30,000 Gibbs iterations is shown in Figure 1(c). Such sequences have negligible posterior probability under the sticky HDP-HMM formulation.