discrete or continuous valued, are independent. The HMM is the most basic example of a Markov switching process and forms the building block for more complicated processes examined later.

**THERE IS A NATURAL INTERPLAY BETWEEN COMBINATORIAL STOCHASTIC PROCESSES AND STATE-SPACE DESCRIPTIONS OF DYNAMICAL SYSTEMS.**

### *FINITE HMM*

Let $z_t$ denote the mode of the Markov chain at time $t$, and $\pi_j$ the mode-specific transition distribution for mode $j$. Given the mode $z_t$, the observation $y_t$ is conditionally independent of the observations and modes at other time steps. The generative process can be described as

$$\begin{aligned} z_t \mid z_{t-1} &\sim \pi_{z_{t-1}} \\ y_t \mid z_t &\sim F(\theta_{z_t}) \end{aligned} \tag{1}$$

for an indexed family of distributions $F(\cdot)$ (e.g., multinomial for discrete data or multivariate Gaussian for real, vector-valued data), where $\theta_i$ are the emission parameters for mode $i$. The notation $x \sim F$ indicates that the random variable $x$ is drawn from a distribution $F$. We use bar notation $x \mid F \sim F$ to specify conditioned-upon random elements, such as a random distribution. The directed graphical model associated with the HMM is shown in Figure 1(a).

One can equivalently represent the HMM via a set of transition probability measures $G_j = \sum_{k=1}^{K} \pi_{jk}\delta_{\theta_k}$, where $\delta_\theta$ is a unit mass concentrated at $\theta$. Instead of employing transition distributions on the set of integers (i.e., modes) that index into the collection of emission parameters, we operate directly in the parameter space $\Theta$, and transition between emission parameters with probabilities $\{G_j\}$. Specifically, let $j_{t-1}$ be the unique emission parameter index $j$ such that $\theta'_{t-1} = \theta_j$. Then,

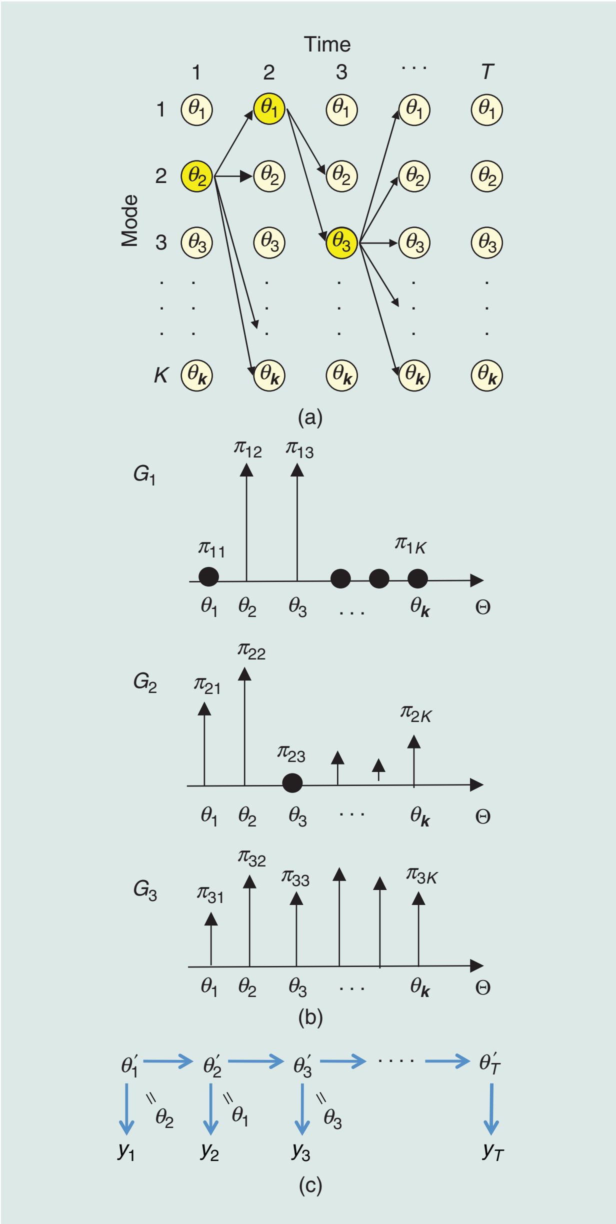
$$\begin{aligned} \theta'_t \mid \theta'_{t-1} &\sim G_{j_{t-1}} \\ y_t \mid \theta'_t &\sim F(\theta'_t). \end{aligned} \tag{2}$$

Here, $\theta'_t \in \{\theta_1, \ldots, \theta_K\}$ takes the place of $\theta_{z_t}$ in (1). A visualization of this process is shown by the trellis diagram of Figure 2.

One can consider a Bayesian HMM by treating the transition probability measures $G_j$ as random, and endowing them with a prior. Formally, a random measure on a measurable space $\Theta$, with sigma algebra $\mathcal{A}$, is defined as a stochastic process whose index set is $\mathcal{A}$. That is, $G(A)$ is a nonnegative random variable for each $A \in \mathcal{A}$. Since the probability measures are solely distinguished by their weights on the shared set of emission parameters $\{\theta_1, \ldots, \theta_K\}$, we consider a prior that independently handles these components. Specifically, take the weights $\pi_j = [\pi_{j1} \ldots \pi_{jK}]$ (i.e., transition distributions) to be independent draws from a $K$-dimensional Dirichlet distribution,

$$\pi_j \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_K) \quad j = 1, \ldots, K, \tag{3}$$

implying that $\sum_k \pi_{jk} = 1$, as desired. Then, assume that the atoms are drawn as $\theta_j \sim H$ for some base measure $H$ on the parameter space $\Theta$. Depending on the form of the emission distribution, various choices of $H$ lead to computational efficiencies via conjugate analysis.




**[FIG2] (a) Trellis representation of an HMM. Each circle represents one of the $K$ possible HMM emission parameters at various time steps. The highlighted circles indicate the selected emission parameter $\theta'_t$ at time $t$, and the arrows represent the set of possible transitions from that HMM mode to each of the $K$ possible next modes. The weights of these arrows indicate the relative probability of the transitions encoded by the mode-specific transition probability measures $G_j$. (b) Transition probability measures $G_1$, $G_2$, and $G_3$ corresponding to the example trellis diagram. (c) A representation of the HMM observations $y_t$, which are drawn from emission distributions parameterized by the highlighted nodes.**