the *hidden states* of the model, because we cannot *directly* observe which metastable state the system is in.

The hidden Markov model presumes the observed data $\mathbf{O} \equiv \{o_t^{(n)}\}$, where $n = 1, \ldots, N$ and $t = 0, \ldots, T_n$, was generated according to the following model dependent on parameters $\Theta \equiv \{\mathbf{T}, \mathbf{E}\}$ and prior information about the initial state distribution ~~$\rho(s)$~~:

ρ

$$s_0^{(n)} \sim \rho_{s_0^{(n)}}$$

$$s_t^{(n)} | s_{t-1}^{(n)}, \mathbf{T} \sim T_{s_{t-1}^{(n)}, s_t^{(n)}} \qquad t \geq 1 \,?$$

$$o_t^{(n)} | s_t^{(n)}, \mathbf{e}_{s_t^{(n)}} \sim \varphi(o_t^{(n)} | \mathbf{e}_{s_t^{(n)}}) \quad (n) \tag{1}$$

In diagrammatic form, the observed state data ~~o_t~~ $o^{(n)}$ and corresponding hidden state history ~~s_t~~ $S^{(n)}$ can be represented

$$\begin{array}{ccccccccc} \mathbf{s}^{(n)} & \equiv & s_0^{(n)} & \rightarrow & s_1^{(n)} & \rightarrow & s_2^{(n)} & \rightarrow \cdots & s_{T_n}^{(n)} \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \mathbf{o}^{(n)} & \equiv & o_0^{(n)} & & o_1^{(n)} & & o_2^{(n)} & & o_{T_n}^{(n)} \end{array} \tag{2}$$

depend on n? if so, should it be $s_0^{(n)} \sim \rho_{s_0^{(n)}}$?

Here, state transitions $(s_t^{(n)} \rightarrow s_{t+1}^{(n)})$ are governed by the discrete transition probability $T_{s_t^{(n)} s_{t+1}^{(n)}}$, while the "emission" of observables from each state $(s_t^{(n)} \rightarrow o_t^{(n)})$ is governed by the continuous ~~transition~~ emission? probability $\varphi(o_t^{(n)} | \mathbf{e}_{s_t^{(n)}})$. [JDC: We should cite Gerhard Hummer's transition matrix sampling, and mention this might be an alternative to using $\mathbf{T}$ here.]

The initial state distribution $\rho^{(n)}$ (which may itself be a function of the stationary distribution π of $\mathbf{T}$) simply reflects our knowledge of the initial conditions of the experiment that collected data $\mathbf{o}^{(n)}$. In the case that the experiment was prepared in equilibrium, ρ corresponds to the equilibrium distribution π of the model transition matrix $\mathbf{T}$. However, if the experiment was started out of equilibrium, perhaps restricted to some subset of states, then the prior might reflect simple ignorance as to which state the system initially started in by assigning each state in this subset an equal probability. Here, we presume that either ρ is known *a priori*, or that it is a function of the equilibrium distribution π determined by $\mathbf{T}$.

The *Markov* property of HMMs prescribes that the observation that the probability that a system in state i at time t is observed in state j at time $t+1$ is dependent only on knowledge of the state i, and given by the corresponding matrix element T_{ij} of the (row-stochastic) transition matrix $\mathbf{T}$.

wording is confusing

The probability that a particular value o of the observable is measured is dependent only on the current state s, and given by some model of the observable distribution for this state $\varphi(o|\mathbf{e}_s)$ parametrized by observable *emission* parameters $\mathbf{e}$.

For example, in the applications to force spectroscopy described in this paper, the observable denotes the measured force exerted on a bead in an optical trap, and the model is taken to be a simple Gaussian distribution parameterized by $\mathbf{e} \equiv \{\mu, \sigma^2\}$:

$$\varphi(o|\mathbf{e}) = \varphi(o|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(o-\mu)^2}{\sigma^2}\right]. \tag{3}$$

Given the HMM process specified in Eq. 1, the probability of observing the data $\mathbf{O}$ given the model parameters Θ is therefore

$$P(\mathbf{O}|\Theta) \tag{4}$$
$$= \sum_{\mathbf{S}} \prod_{n=1}^{N} \rho_{s_0^{(n)}} \varphi(o_0^{(n)} | \mathbf{e}_{s_0^{(n)}}) \prod_{t=1}^{T^{(n)}} T_{s_{(t-1)}^{(n)} s_t^{(n)}} \varphi(o_t^{(n)} | \mathbf{e}_{s_t^{(n)}})$$

have been using T_n but maybe they should all be $T^{(n)}$?

Still sort of conflicts w/ transition matrix $\mathbf{T}$. could notation use P for transition probability.

where the sum over hidden state histories $\mathbf{S}$ is shorthand for

$$\sum_{\mathbf{S}} \equiv \sum_{s_0^{(1)}=1}^{M} \cdots \sum_{s_{T_1}^{(1)}=1}^{M} \cdots \sum_{s_{T_N}^{(N)}=1}^{M}. \tag{5}$$

B. Maximum likelihood hidden Markov model (MLHMM)

The standard approach to construct an HMM from observed data is to compute the *maximum likelihood estimator* (MLE) for the model parameters Θ, which maximize the probability of the observed data $\mathbf{O}$ given the model,

$$\hat{\Theta} = \arg\max_{\Theta} P(\mathbf{O}|\Theta) \tag{6}$$

Once the MLE parameters $\hat{\Theta}$ are determined, the most likely hidden state history that produced the observations $\mathbf{O}$ can be determined using these parameters:

$$\hat{\mathbf{S}} = \arg\max_{\mathbf{S}} P(\mathbf{S}|\mathbf{O}, \hat{\Theta}) \quad \mathbf{S} \text{ or } \hat{\mathbf{S}}? \tag{7}$$

While the hidden state sequences $\hat{\mathbf{S}} \equiv \{\mathbf{s}^{(n)}\}_{n=1}^{N}$ corresponding to the observable traces $\mathbf{O}$ can provide the experimenter with insight into conformational dynamics, the transition matrix $\mathbf{T}$ captures the *statistical* aspects of the dynamics. [JDC: Talk about utility and drawbacks of MLHMM.]

I don't understand this distinction.

C. Bayesian hidden Markov model (BHMM)

Instead of simply determining the model that maximizes the likelihood of observing the data $\mathbf{O}$ given the model parameters Θ, we can make use of Bayes' theorem to compute the *posterior* distribution of model parameters given the observed data:

$$P(\Theta|\mathbf{O}) \propto P(\mathbf{O}|\Theta)P(\Theta). \tag{8}$$

Here, $P(\Theta)$ denotes a *prior* distribution, that encodes any *a priori* information we may have about the model parameters Θ. This might include both physical constraints (such as ensuring the transition matrix satisfy detailed balance) and prior rounds of inference from other independent experiments.

Making use of the likelihood (Eq. 4), the model posterior is then given by,

$$P(\Theta|\mathbf{O}) \propto P(\Theta) \tag{9}$$
$$\times \sum_{\mathbf{S}} \prod_{n=1}^{N} \rho_{s_0^{(n)}} \varphi(o_0^{(n)} | \mathbf{e}_{s_0^{(n)}}) \prod_{t=1}^{T^{(n)}} T_{s_{(t-1)}^{(n)} s_t^{(n)}} \varphi(o_t^{(n)} | \mathbf{e}_{s_t^{(n)}})$$

Drawing samples of Θ from this distribution will, in principle, allow the *confidence* with which individual parameters and combinations thereof are known, given the data (and subject to the validity of the model of Eq. 1 in correctly representing the process by which the observed data is generated). However, due to the sum over all hidden state histories $\mathbf{S}$ appearing in the posterior (Eq. 9), direct sampling of the model parameters Θ is difficult. Instead, we take the approach of introducing