The expected transition counts $\mathbf{C} \equiv (c_{ij})$ are then computed from the $\Xi^{(n)}$ as

$$c_{ij} = \sum_{n=1}^{N} \sum_{t=0}^{T^{(n)}-1} \xi_{tij}^{(n)} \tag{30}$$

This count matrix is used to update the maximum-likelihood transition matrix $\mathbf{T}$ using the method of Prinz et al. [23] described in the previous section.

The state observable distribution parameters $\mathbf{E}$ are then updated from the γ_{ti} as

$$\mu_i' = \frac{\sum_{n=1}^{N} \sum_{t=0}^{T_n} o_t^{(n)} \gamma_{ti}^{(n)}}{\sum_{n=1}^{N} \sum_{t=0}^{T^{(n)}} \gamma_{ti}^{(n)}}$$
$$\sigma_i'^2 = \frac{\sum_{n=1}^{N} \sum_{t=0}^{T^{(n)}} (o_t^{(n)} - \mu_i')^2 \gamma_{ti}^{(n)}}{\sum_{n=1}^{N} \sum_{t=0}^{T^{(n)}} \gamma_{ti}^{(n)}} \tag{31}$$

very slightly strange that we're now using i to index the state instead of n. Can't think of an easy fix though

Once the model parameters have been fitted by iteration of the above update procedure to convergence (which may only converge to a local maximum of the likelihood), the most likely hidden state sequence can be determined given the observations $\mathbf{O}$ and the MLE model $\hat{\Theta}$ using the Viterbi algorithm [13]. Like the forward-backward algorithm employed in the Baum-Welch procedure, the Viterbi algorithm also has a forward recursion component that is applied independently to each trace n,

$$\delta_{jt} = \begin{cases} \rho_j \varphi(o_t|\mathbf{e}_j) & t = 0 \\ \varphi(o_t|\mathbf{e}_j) \max_i \delta_{i(t-1)} T_{ij} & t = 1, \ldots, T^{(n)} \end{cases} \tag{32}$$

$$\Phi_{jt} = \begin{cases} 1 & t = 0 \\ \arg\max_i \delta_{i(t-1)} T_{ij} & t = 1, \ldots, T^{(n)} \end{cases} \tag{33}$$

as well as a reverse reconstruction component to compute the most likely state sequence $\hat{\mathbf{S}}$,

$$\hat{s}_t = \begin{cases} \arg\max_i \delta_{iT} & t = T^{(n)} \\ \Phi_{\hat{s}_{t+1}(t+1)} & t = T - 1, \ldots, 0 \end{cases} \tag{34}$$

[JDC: Again, more confusion with the indices and what to do with n sub- or superscipt.]

I think they're looking much better, just need to decide between T_n and $T^{(n)}$

C. Sampling from the posterior of the BHMM

Sampling from the posterior of the BHMM (Eq. 9) proceeds by rounds of Gibbs sampling, where each round consists of an update of the augmented model parameters $\{\mathbf{T}, \mathbf{E}, \mathbf{S}\}$ by sampling

$$\begin{aligned} \mathbf{S}'|\mathbf{T}, \mathbf{E}, \mathbf{O} &\sim P(\mathbf{S}'|\mathbf{T}, \mathbf{E}, \mathbf{O}) \\ \mathbf{T}'|\mathbf{S}' &\sim P(\mathbf{T}'|\mathbf{S}') \\ \mathbf{E}'|\mathbf{S}', \mathbf{O} &\sim P(\mathbf{E}'|\mathbf{S}', \mathbf{O}) \end{aligned}$$

where the conditional probabilities are given by Eq. 12.

1. Updating the hidden state sequences

In the first part of each sampling round, we use a modified form of the Viterbi process [JDC: Something to cite here?] to generate an independent sample of the hidden state history $\mathbf{S}$ given the transition probabilities $\mathbf{T}$, state observable distribution parameters $\mathbf{E}$, and observed data $\mathbf{O}$. ~~Instead of computing the most *likely* state sequence, Like~~ the Viterbi scheme, a forward recursion algorithm (Eq. ??) is applied to each trace $\mathbf{o}^{(n)}$ separately, but instead of computing the most *likely* state history on the reverse pass, a new state history is drawn from the distribution $P(\mathbf{s}|\mathbf{o}, \mathbf{T}, \mathbf{E})$. The forward recursion is identical to the Viterbi case:

$$\delta_{jt} = \begin{cases} \rho_j \varphi(o_t|\mathbf{e}_j) & t = 0 \\ \varphi(o_t|\mathbf{e}_j) \max_i \delta_{i(t-1)} T_{ij} & t = 1, \ldots, T^{(n)} \text{ ?} \end{cases} \tag{35}$$

The hidden state sequence $\mathbf{s}_t$ corresponding to observation trace $\mathbf{o}^{(n)}$ is then sampled according to $P(s_t|s_{t+1}, \ldots, s_T)$ in order from $t = T$ down to $t = 0$:

$$P(s_t|s_{t+1}, \ldots, s_T) \propto \begin{cases} \delta_{iT} & t = T^{(n)} \\ \delta_{it} T_{i s_{t+1}} & t = T - 1, \ldots, 0 \end{cases} \tag{36}$$

[JDC: Perhaps we should avoid the notation δ_{it}, since this could be confused with the Kronecker delta.] AGREE.

need to link $i = s_t$, right?

$$P(s_t|s_{t+1}, \ldots, s_T) \propto \begin{cases} \delta_{s_t T} & t = T \\ \delta_{s_t t} T_{s_t s_{t+1}} & t = \end{cases} \;??$$

2. Updating the transition probabilities

If no detailed balance constraint is used and the prior $P(\mathbf{T})$ is Dirichlet in each row of the transition matrix $\mathbf{T}$, it is possible to generate an independent sample of the transition matrix from the conditional distribution $P(\mathbf{T}'|\mathbf{S}')$ by sampling each row of the transition matrix from the conjugate Dirichlet posterior using the transition counts from the sampled state sequence $\mathbf{S}'$ [21]. However, because physical systems in the absence of energy input through an external driving force should satisfy detailed balance, we make use of this constraint in updating our transition probabilities, since this has been demonstrated to substantially reduce parameter uncertainty in the low-data regime [21].

The transition matrix is updated using the reversible transition matrix sampling scheme of Noé [21, 25]. Here, an adjusted count matrix $\mathbf{C} \equiv (c_{ij})$ is computed using the updated hidden state sequence $\mathbf{S}'$:

$$c_{ij} = b_{ij} + \sum_{n=1}^{N} \sum_{t=1}^{T_n} \delta_{i, s_{t-1}^{(n)}} \delta_{j, s_t^{(n)}} \tag{37}$$

where the Kronecker $\delta_{i,j} = 1$ if $i = j$ and zero otherwise, and $\mathbf{B} \equiv (b_{ij})$ is a matrix of prior pseudocounts, which we take to be zero following the work of Noé et al. [10]. Using the adjusted count matrix $\mathbf{C}$, a Metropolis-Hastings Monte Carlo procedure [26] is used to update the matrix and produce a new sample from $P(\mathbf{T}|\mathbf{S})$. Two move types are attempted, selected with equal probability, and 1000 moves are attempted to generate a new sample $\mathbf{T}'$ that is approximately uncorrelated from the previous $\mathbf{T}$. Prior to starting the Monte Carlo procedure, the vector of equilibrium probabilities for all states π is computed according to

$$\mathbf{T}\pi = \pi \tag{38}$$

$P(T'|S')$?