Supplement: Memory (and Time) Efficient Sequential Monte Carlo

1 Notation used in the supplement and more background on standard SMC

Let $\mathcal{X}_r = E_1 \times \dots E_r$ be the target space at each generation (SMC iteration) $r \in \{1, \dots, R\}$ with corresponding product σ-algebra $\mathcal{F}_r$. We denote the target measures for each generation by $\pi_r : \mathcal{F}_r \to \mathbb{R}^+$. Given a test function ϕ and a measure π, we denote by $\pi\phi$ the integral of ϕ under π, $\pi\phi = \int \phi \, \mathrm{d}\pi$.

The two main inference questions we are interested in are to compute the data likelihood (normalization), and to compute expectations under the normalized measure. To succinctly study the approximation of both of these quantities, we introduce the following notations: given any positive measure λ on $\mathcal{X}$, we write $\|\lambda\| = \lambda(\mathcal{X})$ and $\bar{\lambda}(A) = \lambda(A)/\|\lambda\|$. Note that with this notation, posterior expectations take the form $\bar{\pi}_R\phi$, and the data likelihood is $\pi_R(\mathcal{X}_R)$ or, equivalently, $\pi_R\phi_1$, where $\phi_1 \equiv 1$. Note that our notation in the supplement is slightly different than in the main paper, where the notation π was already assumed to be normalized.

We assume that the target distribution is known only up to an unnormalized density, γ_r (the pointwise product of a prior times a likelihood in Bayesian applications) and we assume that a regular proposal distribution $\nu_{r,x}(A)$ with density $q_x(y)$ is provided (and we drop the dependency on r to simplify the notation). In order to concisely describe our method, we define the following random operators on measures:

Definition 1 *Given the arbitrary positive measure λ, the output of the random operator $\mathrm{res}_K\,\lambda$ (respectively $\mathrm{prop}_K\,\lambda$) is a new measure defined as follows: for any test function ϕ,*

$$
\begin{aligned}
(\mathrm{res}_K\,\lambda)(\phi) &= \|\lambda\| \frac{1}{K} \sum_{k=1}^{K} \phi(S_k), \\
(\mathrm{prop}_K\,\lambda)(\phi) &= \|\lambda\| \frac{1}{K} \sum_{k=1}^{K} w(S_k, S'_k)\phi(S'_k),
\end{aligned}
$$

where $S_i \sim \bar{\lambda}$, and $S'_i|S_i \sim \nu_{S_i}$ independently, and w is the standard SMC weight formula for state space models:

$$w_r(x, y) = \frac{\gamma_r(y)}{\gamma_{r-1}(x)} \frac{1}{q_x(y)}. \tag{1}$$

With this notation, the standard particle filter described in the Background section can be describe succinctly. We call the output of each generation of the filter a *particle population*, formally a discrete measure, $\pi_{r,K}$ defined on the product space $\mathcal{X}_r$. This measure can be written as:

$$\pi_{r,K} = \mathrm{prop}_K \, \pi_{r-1,K},$$

since prop_K incorporates both the resampling step (via the sampling of the S_i from $\bar{\pi}_{r-1,K}$), and the proposal step (via the sampling of S'_i from ν_{S_i}). Here K is constant and is a tuning parameter, the number of particle, which controls the accuracy of the method.

2 Measure-theoretic formulation of IPSMC

Note that with the notation of the previous section, our algorithm can be succinctly described as follows:

$$\pi^{\mathrm{IP}}_{r,K} = \mathrm{res}_K \left(\mathrm{prop}_{N(K,M)} \, \pi^{\mathrm{IP}}_{r-1,K} \right),$$

where M and $N(K, M)$ are described in the paper.

There are two resampling stages in the above expression, one in the form of the outer res operator, the other one included inside prop. Note that these two resampling stages are needed: since at iteration r we do not know in advance how many times the output discrete measure will be sampled from at iteration $r + 1$, we use the res_K operator to ensure that the memory bound K will be respected no matter what is the behavior at iteration $r + 1$. While this would be detrimental in a standard particle filter, this is not a problem in our scheme since the number of proposal calls at each iteration is suitably expanded via the stopping time N. This comes at a cost of a theoretically higher running time, but as our experiments demonstrate, in practice the running time of our method is competitive with standard SMC methods due to the lower frequency of memory writes.

3 Stream-based resampling

In the second task group, it is useful to identify which of the N implicit particles are selected by the contraction operator in a streaming fashion.

To do this, first recall that sampling K times from a multinomial is equivalent to arranging the normalized weights into a partition of a stick of unit length, to sample K independent uniform random variables and to look how many

variables fall in each stick segment. Note that equivalently, we can generate the spacings $T_1, T_2, \ldots, T_K$ between consecutive uniform variables.

In other words, if $V_1, V_2, \ldots, V_K$ are independent uniform random variables, and $V_{(1)}, V_{(2)}, \ldots, V_{(K)}$ are the sorted uniforms, then $T_1 = V_{(1)}$, and $T_k = V_{(k)} - V_{(k-1)}$ for $k \geq 2$. Note that $T_k | T_1, \ldots, T_{k-1}$ can be easily simulated as $(1 - \sum_{j=1}^{k-1} T_j) B_k$, where $B_k \sim \text{Beta}(1, K - k + 1)$.

See Algorithm 3 to see how this method is integrated in Task Group 2.

4 Algorithms

Algorithm 1 : **IP-SMC**

$\pi^{\text{IP}}_{0,K} \leftarrow$ **init**()
$z \leftarrow 1$
for $r = 1, 2, \ldots, R$ **do**
  $U_\cdot, U'_\cdot \leftarrow$ **seeds**(r)
  $\xi \leftarrow$ **task-group-1**$(\pi^{\text{IP}}_{r-1,K}, K, U_\cdot)$
  $(\pi^{\text{IP}}_{r,K}, I_r, z_r) \leftarrow$ **task-group-2**$(\pi^{\text{IP}}_{r-1,K}, K, U_\cdot, \xi, \phi_r)$
  $z \leftarrow z \times z_r$
end for
return $(z, I_1, \ldots, I_R)$

Algorithm 2 : **task-group-1**$(\pi^{\text{IP}}_{r-1,K}, K, U_\cdot)$

$\xi \leftarrow (0, 0, 0)$
for $n = 1, 2, \ldots, N^*$ **do**
  $x_{r-1} \sim \pi^{\text{IP}}_{r-1,K}[U_{2n}]$ {The bracket notation means that the seed of the sampling is fixed by U_{2k}}
  $x_r | x_{r-1} \sim \nu_{x_{r-1}}[U_{2n-1}]$
  $w \leftarrow w(x_{r-1}, x_r)$
  $\xi \leftarrow \xi + (w^0, w^1, w^2)$ {Note that higher, but still finite dimensional terms might be needed in the context of a large number of particles (see Section 3.4 of the main text)}
  if bin-approx$(m) > \alpha K$ **then**
    return ξ
  end if
end for

Algorithm 3 : **task-group-2**$(\pi^{\mathrm{IP}}_{r-1,K}, K, U_\cdot, \xi, \phi_r)$

$j \leftarrow 1$
$t \leftarrow 0$ {sum of uniform spacings visited so far}
$s \leftarrow 0$ {sum of weights visited so far}
$I \leftarrow 0$ {Monte Carlo partial sum}
$z \leftarrow 0$ {Estimate of the normalization}
for $n = 1, 2, \ldots, \xi_0$ **do**
  $x_{r-1} \sim \pi^{\mathrm{IP}}_{r-1,K}[U_{2n}]$ {The bracket notation means that the seed of the sampling is fixed by U_{2k}}
  $x_r | x_{r-1} \sim \nu_{x_{r-1}}[U_{2n-1}]$
  $w \leftarrow w(x_{r-1}, x_r)/\xi_1$
  $I \leftarrow I + w \times \pi(x_r)$
  $z \leftarrow z + w \times \xi_1/\xi_0$
  $r \leftarrow s + w$
  for $k \leq K$ **do**
    $t' \leftarrow t +$ **sample-interval**(k, K, t) {See Section 3.}
    if $t' < r$ **then**
      Add $(x_r, 1/K)$ to $\pi^{\mathrm{IP}}_{r,K}$
      $k \leftarrow k + 1$
      $t \leftarrow t'$
    else
      Break inner loop
    end if
  end for
  $s \leftarrow r$
end for
return $(\pi^{\mathrm{IP}}_{r,K}, I, z)$

Algorithm 4 : **bin-approx**(ξ)

return $\xi_0 - \binom{\xi_0}{2} \frac{\xi_2}{(\xi_1)^2}$ {See Section 3.4 of the main text for higher order approximations}

5 Proofs

5.1 Expectation

We now prove Proposition 1.

Proof: Let T_n be the number of times particle n is sampled. Then, we have that the number of distinct samples is:

$$\begin{aligned} D &:= |\{S_1, \ldots, S_K\}| \\ &= \sum_{n=1}^{N} \mathbf{1}(T_n > 0). \end{aligned}$$

Using $\mathbb{P}(T_n > 0) = 1 - (1 - \bar{w}_n)^K$, we can write the expected value of D as:

$$\begin{aligned}
\mathbb{E}D &= \mathbb{E}\sum_{n=1}^{N} T_n = \sum_{n=1}^{N} \mathbb{E}T_n \\
&= \sum_{n=1}^{N} \mathbb{P}(T_n = 1) \\
&= \sum_{n=1}^{N} 1 - (1 - \bar{w}_n)^K \\
&= N - \sum_{n=1}^{N} (1 - \bar{w}_n)^K.
\end{aligned}$$

■

5.2 Concentration Inequality

We also provide here a basic concentration inequality showing that the above expectation is a reasonable approximation of the realized number of unique particles after resampling.

Proposition 2 *Let Y_k denote the indicator that sample $S_k \in (1, \ldots, N)$ picks a particle not yet picked by any samples from $1, 2, \ldots, k-1$. Let $S = \sum_{k=1}^{K} Y_k$, the number of distinct particles sampled from a multinomial distribution. Then $\mathbf{V}ar(S) \leq 3K$ whenever $\max_i \bar{w}_i < 1/2$. And therefore using Chebyshev's inequality, for $\epsilon > 0$,*

$$P(|S - \mathbb{E}S| \geq \epsilon \mathbb{E}S) \leq \frac{3K}{\epsilon^2 (\mathbb{E}S)^2}$$

Proof:
We can express the variance as:

$$\mathbf{V}ar(S) = \mathbf{V}ar\left(\sum_{k=2}^{K} Y_k\right) = \sum_{k=2}^{K} \mathbf{V}ar(Y_k) + 2\sum_{k=2}^{K}\sum_{\Delta=1}^{K-k} Cov(Y_k, Y_{k+\Delta})$$

Using $\mathbb{P}(Y_k = 1) = \sum_{i=1}^{n} \bar{w}_i (1 - \bar{w}_i)^{k-1}$, we can bound the covariance:

$$
\begin{aligned}
Cov(Y_k, Y_{k+\Delta}) &= \mathbb{E}Y_k Y_{k+\Delta} - \mathbb{E}Y_k \mathbb{E}Y_{k+\Delta} \\
&= \sum_{i_1 \neq i_2} (1 - \bar{w}_{i_1} - \bar{w}_{i_2})^{k-1} \bar{w}_{i_1} (1 - \bar{w}_{i_2})^{\Delta - 1} \bar{w}_{i_2} \\
&- \sum_{i_1, i_2} (1 - \bar{w}_{i_1})^{k-1} \bar{w}_{i_1} (1 - \bar{w}_{i_2})^{k+\Delta-1} \bar{w}_{i_2} \\
&= \sum_{i_1, i_2} \bar{w}_{i_1}^{k+1} \bar{w}_{i_2}^{k+1} (1 - \bar{w}_{i_2})^{\Delta - 1} - \sum_{i=1}^{n} (1 - 2\bar{w}_i)^{k-1} (1 - \bar{w}_i)^{\Delta - 1} \bar{w}_i^2 \\
&\leq \sum_{i_1, i_2} \bar{w}_{i_1} \bar{w}_{i_2} (\bar{w}^*)^k = (\bar{w}^*)^k
\end{aligned}
$$

the inequality from the second last line to the last line is true because the second term in the second last line is greater than equal to 0. Therefore,

$$
\begin{aligned}
2 \sum_{k=2}^{K} \sum_{\Delta=1}^{K-k} Cov(Y_k, Y_{k+\Delta}) &\leq \sum_{2 \leq k \neq k' \neq K} (\bar{w}^*)^{\min\{k, k'\}} \\
&\leq K \sum_{k=2}^{K} (\bar{w}^*)^k \\
&\leq KC'
\end{aligned}
$$

Here, we can use $C' = 2$ using the assumption that $\max_i \bar{w}_i < 1/2$.

Therefore, we have $\mathbf{V}ar(S) \leq K + KC'$ and setting $C = 1 + C'$, we have the desired result. ■

Next, we show that if $N^* > N(K)$ and $K > 10$, the condition $\max_i \bar{w}_i < 1/2$ is automatically satisfied, by construction of the stopping time:

Proof: Suppose the contrary. Then, since the large weight is sampled at least half of the time in expectation, we have:

$$
\psi((w_1, \ldots, w_N), K) \leq \text{ceiling} \left(\frac{K}{2} \right) + 1.
$$

But on the other hand, the stopping criterion implies that

$$
\psi((w_1, \ldots, w_{N+1}), K) > \alpha K.
$$

This is a contradiction for $\alpha = 1 - (1 - 1/K)^K \approx 1 - \exp(-1)$ and $K > 10$, since:

$$
|\psi((w_1, \ldots, w_N), K) - \psi((w_1, \ldots, w_{N+1}), K| \leq 1.
$$

■

5.3 Consistency

Lemma 3 *For any positive measure λ with $\|\lambda\| < \infty$, we have*

$$\mathbb{E}[(\mathrm{prop}_K\,\lambda)\phi] = (\mathrm{prop}\,\lambda)\phi, \tag{2}$$

$$\mathbb{E}[(\mathrm{res}_K\,\lambda)\phi] = \lambda\phi, \tag{3}$$

where:

$$(\mathrm{prop}\,\lambda)\phi = \int \lambda(\,\mathrm{d}x) \int \nu_x^+(\,\mathrm{d}y) w(x,y)\phi(y).$$

Proof: We obtain Equation (2) by linearity of expectation:

$$\begin{aligned}\mathbb{E}[(\mathrm{prop}_K\,\lambda)\phi] &= \|\lambda\|\frac{K}{K}\mathbb{E}[w(S_1,S_1')\phi(S_1')] \\ &= \|\lambda\| \int \bar{\lambda}(\,\mathrm{d}x) \int \nu_x(\,\mathrm{d}y) w(x,y)\phi(y) = (\mathrm{prop}\,\lambda)\phi.\end{aligned}$$

Equation (3) follows by the same argument. ■

Lemma 4 *For any positive measure λ with $\|\lambda\| < \infty$, we have:*

$$\mathbb{E}\left[(\mathrm{prop}_K\,\lambda)\phi - (\mathrm{prop}\,\lambda)\phi\right]^2 \leq \frac{(C_1C_2)^2\|\lambda\|^2}{K} \tag{4}$$

$$\mathbb{E}\left[(\mathrm{res}_K\,\lambda)\phi - \lambda\phi\right]^2 \leq \frac{C_1^2\|\lambda\|^2}{K} \tag{5}$$

Proof: From Lemma 3, we can rewrite the left-hand sides as variances of sums. Next, using independence of (S_k, S_k') and $(S_{k'}, S_{k'}')$, $k \neq k'$ in the definition of prop_K, we have:

$$\mathbb{E}\left[(\mathrm{prop}_K\,\lambda)\phi - (\mathrm{prop}\,\lambda)\phi\right]^2 = \frac{\|\lambda\|^2}{K}\mathbf{V}ar[w(S_k,S_k')\phi(S_k')] \leq \frac{(C_1C_2)^2\|\lambda\|^2}{K}.$$

Equation (5) follows by the same argument. ■

Corollary 5 *For any stopping time $N(K)$ with $N(K) \geq K$, we have:*

$$\mathbb{E}\left[(\mathrm{prop}_{N(K)}\,\lambda)\phi - (\mathrm{prop}\,\lambda)\phi\right]^2 \leq \frac{(C_1C_2)^2\|\lambda\|^2}{K} \tag{6}$$

Note that the condition $N(K) \geq K$ is satisfied when $N(K) = N(K, \alpha K)$ with $\alpha = 1 - (1 - 1/K)^K \approx 1 - \exp(-1)$: in this case, we will have $N(K)$ minimized when the weights are uniform, in which case Proposition 2 implies that $N(K) = K$.

Proof: Using the assumption $N(K) \geq K$ and Lemma 4, we get:

$$\begin{aligned}
\mathbb{E}\left[(\mathrm{prop}_{N(K)}\,\lambda)\phi - (\mathrm{prop}\,\lambda)\phi\right]^2 &= \sum_{n=K}^{\infty} \mathbb{P}(N(K)=n)\ \mathbb{E}\left[(\mathrm{prop}_n\,\lambda)\phi - (\mathrm{prop}\,\lambda)\phi\right]^2 \\
&\leq \sum_{n=K}^{\infty} \mathbb{P}(N(K)=n)\ \frac{(C_1C_2)^2\|\lambda\|^2}{n} \\
&\leq \sum_{n=K}^{\infty} \mathbb{P}(N(K)=n)\ \frac{(C_1C_2)^2\|\lambda\|^2}{K} \\
&= \frac{(C_1C_2)^2\|\lambda\|^2}{K}
\end{aligned}$$

■

Next, the following lemma follows from Lemma 9 in [3]:

Lemma 6 *For all r, $\mathrm{prop}\,\pi_r = \pi_{r+1}$.*

Lemma 7 *If for all bounded measurable ϕ,*

$$\pi_{r,K}\phi \xrightarrow{\mathbf{L}^2} \pi_r\phi, \tag{7}$$

then we also have:

$$(\mathrm{prop}\,\pi_{r,K})\phi \xrightarrow{\mathbf{L}^2} (\mathrm{prop}\,\pi_r)\phi. \tag{8}$$

Moreover, by Lemma 6 the right-hand side of Equation (8) is equal to $\pi_{r+1}\phi$.

Proof: Let $\tilde{\phi}(x) = \int_A \nu_x^+(\,\mathrm{d}y) w(x,y)\phi(y)$. As $w < C_2$, $|\phi| < C$ implies $|\tilde{\phi}'| < CC_2$, we can use the test function $\tilde{\phi}$ in Equation (7) to obtain Equation (8). ■

We can now prove the main proposition:

Proof: We proceed by induction, showing for $r \geq 0$, and for all bounded ϕ, we have $\pi^{\mathrm{IP}}_{r,K}\phi \xrightarrow{\mathbf{L}^2} \pi_r\phi$. The base case is trivial, since $\pi^{\mathrm{IP}}_{0,K}$ and π_0 are equal to a Dirac delta on the same atom. To prove the induction hypothesis, we first decompose the L^2 norm using Minkowski inequality, and control each term separately:

$$\begin{aligned}
\mathbb{E}^{1/2}\left[\pi^{\mathrm{IP}}_{r+1,K}\phi - \pi_{r+1}\phi\right]^2 \leq\ & \mathbb{E}^{1/2}\left[\pi^{\mathrm{IP}}_{r+1,K}\phi - \left(\mathrm{prop}_{N(K)}\,\pi^{\mathrm{IP}}_{r,K}\right)\phi\right]^2 && (9)\\
&+ \mathbb{E}^{1/2}\left[\left(\mathrm{prop}_{N(K)}\,\pi^{\mathrm{IP}}_{r,K}\right)\phi - \left(\mathrm{prop}\,\pi^{\mathrm{IP}}_{r,K}\right)\phi\right]^2 && (10)\\
&+ \mathbb{E}^{1/2}\left[\left(\mathrm{prop}\,\pi^{\mathrm{IP}}_{r,K}\right)\phi - \pi_{r+1}\phi\right]^2 && (11)
\end{aligned}$$

For the first term, we have,

$$\begin{aligned}
\text{Equation (9)} &= \mathbb{E}^{1/2}\left[\left(\text{res}_K\left(\text{prop}_{N(K)}\,\pi^{\text{IP}}_{r,K}\right)\right)\phi - \left(\text{prop}_{N(K)}\,\pi^{\text{IP}}_{r,K}\right)\phi\right]^2 \\
&= \mathbb{E}^{1/2}\left[\text{res}_K\,\lambda_K\phi - \lambda_K\phi\right]^2 \\
&= \left(\mathbb{E}\left[\mathbb{E}\left[\left(\left(\text{res}_K\,\lambda_K\right)\phi - \lambda_K\phi\right)^2 |\lambda_K\right]\right]\right)^{1/2} \\
&\leq \left(\mathbb{E}\left[\frac{C_1^2\|\lambda_K\|^2}{K}\right]\right)^{1/2} \\
&\leq \frac{C_1 C_2^r}{\sqrt{K}},
\end{aligned}$$

where $\lambda_K = \text{prop}_{N(K)}\,\pi^{\text{IP}}_{r,K}$, and we use Lemma 4 to obtain the bound in the penultimate line.

For the second term:

$$\begin{aligned}
\text{Equation (10)} &= \left(\mathbb{E}\left[\mathbb{E}\left[\left(\left(\text{prop}_{N(K)}\,\pi^{\text{IP}}_{r,K}\right)\phi - \left(\text{prop}\,\pi^{\text{IP}}_{r,K}\right)\phi\right)^2 |\pi^{\text{IP}}_{r,K}\right]\right]\right)^{1/2} \\
&\leq \frac{C_1 C_2^r}{\sqrt{K}},
\end{aligned}$$

where we have used Corollary 5.

Finally, by Lemma 7 and the induction hypothesis, Equation (11) also goes to zero as $K \to \infty$.

■

6 Details on Experiments

6.1 Ising Example Posterior Estimates

In this section, we show the posterior estimates of $P(X_1 = +1)$ for the Ising model experiment described in Section 4.1 of the main paper. Shown in Figure 1 is an average over three different runs; it can be seen that IPSMC (red) and the standard SMC (blue) estimates approach to the true value of 0.5.

6.2 Phylogenetic Data Simulation

In this section, we explain the data simulation process for the phylogenetic experiments in Section 4.2 of the main paper.

The data generation process requires as an input the number of taxa N and the number of sites S. It also requires sampling of the phylogenetic tree that defines the ancestral relationship between the N taxa as well as the rate matrix Q, which describes the evolutionary process that takes place along the branches of the tree.

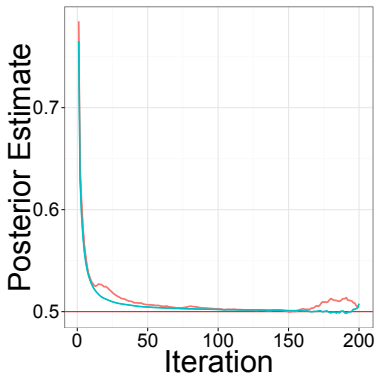

Figure 1: Posterior estimates of $P(X_1 = +1)$ obtained from three different runs (averaged) of the Ising model experiment for IPSMC (red) and the standard SMC (blue). It can be seen that both methods approach to the true value of 0.5.

First, we sample a binary tree t along with the branch lengths, that has N leaves from the coalescent [2]. We refer to t as a *phylogenetic tree*. Then, we fix a rate matrix Q using HKY model [1] and compute the stationary distribution by solving for π:

$$\pi Q = 0$$

To generate data in the phylogenetic experiments in Section 4.2 of the main paper, we sample a rate matrix according to the HKY model and fix it for each of the 5 runs.

Once we are equipped with the phylogenetic tree t and a CTMC defined by Q and π, we generate the sequences for each node in t in the pre-order traversal. One way to view the pre-order traversal of the tree is as a node labelling process by the integers $i \in \{1, \ldots, 2N-1\}$ (where $2N-1$ is the total number of nodes in the tree). For example, the root node would be labelled 1 whereas the right most leaf node would be labelled $2N-1$.

For each node i, let us denote its parent node by pa_i. Furthermore, let $X_i(s)$ denote the sequence at site s of node i and let $X_{\text{pa}_i}(s)$ denote the sequence at site s of node pa_i.

When $i = 1$, we generate the sequence $X_i(s) \sim \text{Multinomial}(\pi)$ for $s \in \{1, \ldots, S\}$. For $i \in \{2, \ldots, 2N-1\}$, we compute the transition probability matrix by exponentiating the rate matrix,

$$P(b_i) = e^{Qb_i}$$

where b_i denotes the branch length from node i to its parent pa_i. Then for each $s \in \{1, \ldots, S\}$,

$$X_i(s)|X_{\mathrm{pa}_i}(s) \sim \mathrm{Multinomial}(P_{X_{\mathrm{pa}_i(s)},\cdot}(b_i))$$

where $P_{X_{\mathrm{pa}_i(s)},\cdot}(b_i)$ denotes the row of the transition rate matrix corresponding to state $X_{\mathrm{pa}_i}(s)$.

References

[1] J. Felsenstein. *Inferring phylogenies*. Sinauer Associates, 2003.

[2] Y. W. Teh, H. Daumé III, and D. M. Roy. Bayesian agglomerative clustering with coalescents. In *Advances in Neural Information Processing Systems (NIPS)*, 2008.

[3] Liangliang Wang. *Bayesian phylogenetic inference via Monte Carlo methods*. PhD thesis, University of British Columbia, 2012.