Aim: Show how SMC can be used for a much wider class of problems than inference in state-space models.

Outline:

1. Examples of probabilistic models
2. General SMC formulation
3. Locally optimal proposals
Examples of probabilistic models
A phylogenetic (evolutionary) tree shows the inferred evolutionary relationships among various species based upon similarities and differences in their physical or genetic characteristics.

\[
p(T_r) = \begin{cases} 
2 \prod_{c \in \text{Ch}(r)} p(T_c) & \text{if } r \text{ is the root node}, \\
2\lambda e^{-(\lambda+\mu)\Delta r} \prod_{c \in \text{Ch}(r)} p(T_c) & \text{if } r \text{ is a speciation}, \\
\mu e^{-(\lambda+\mu)\Delta r} & \text{if } r \text{ is an extinction}, \\
e^{-(\lambda+\mu)\Delta r} & \text{if } r \text{ is an extant species},
\end{cases}
\]

(More on Lectures 17-18.)
A **probabilistic graphical model** (PGM) is a probabilistic model where a graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \) represents the conditional independency structure between random variables,

1. a set of **vertices** \( \mathcal{V} \) (nodes) represents the random variables
2. a set of **edges** \( \mathcal{E} \) containing elements \( (i, j) \in \mathcal{E} \) connecting a pair of nodes \( (i, j) \in \mathcal{V} \times \mathcal{V} \)
The Gaussian process (GP) is a non-parametric and probabilistic model for nonlinear functions.

**Non-parametric** means that it does not rely on any particular parametric functional form to be postulated.

\[
X_t = f(X_{t-1}) + V_t, \quad \text{s.t.} \quad f(X) \sim \mathcal{GP}(0, \kappa, \eta, f(x, x')),
\]

\[
Y_t = g(X_t) + E_t, \quad \text{s.t.} \quad g(X) \sim \mathcal{GP}(0, \kappa, \eta, g(x, x')).
\]

The model functions \( f \) and \( g \) are assumed to be realizations from Gaussian process priors and \( V_t \sim \mathcal{N}(0, Q) \), \( E_t \sim \mathcal{N}(0, R) \).

**Task:** Compute the posterior \( p(f, g, Q, R, \eta, x_{0:T} \mid y_{1:T}) \).
The **Dirichlet process** is a Bayesian nonparametric model, which can be used to construct mixture models with an **unknown** and possibly **infinite** number of components.

**Generative process:**

\[
p(x_{k+1} = j \mid x_{1:k}) = \begin{cases} \frac{n_{k,j}}{k+\alpha} & \text{for } j = 1, \ldots, J_k, \\ \frac{\alpha}{k+\alpha} & \text{for } j = J_k + 1, \end{cases},
\]

\[
\theta_k \sim F(\theta), \quad k = 1, 2, \ldots,
\]

\[
p(y_k \mid x_k, \{\theta_k\}_{k \geq 1}) = G(y_k \mid \theta_{x_k}),
\]
General SMC formulation
SMC can be used to approximate a sequence of probability distributions on a sequence of probability spaces of increasing dimension.

Let \( \{\pi_k(x_{1:k})\}_{k \geq 1} \) be an arbitrary sequence of target distributions

\[
\pi_k(x_{1:k}) = \frac{\tilde{\pi}_k(x_{1:k})}{Z_k}
\]

- The domain of \( x_k \) is \( \mathcal{X}_k \), and \( \mathcal{X}_{1:k} = \mathcal{X}_k \times \mathcal{X}_{1:k-1} \) for all \( k \).
- \( \tilde{\pi}_k(x_{1:k}) \) can be evaluated pointwise.
- The normalizing constant \( Z_k \) may be unknown.

Common tasks:
1. Approximate the normalization constant \( Z_k \).
2. Approximate \( \pi_k(x_k) \) and compute \( \int \phi(x_k) \pi_k(x_k) dx_k \).
The sequence of target distributions \( \{ \pi_k(x_{0:k}) \}_{k=1}^n \) can be constructed in many different ways.

The most basic construction arises from chain-structured graphs, such as the state space model (SSM).

\[
\begin{align*}
\pi_t(x_{0:t}) &= \frac{\tilde{\pi}_t(x_{0:t})}{p(x_{0:t}, y_{1:t})} \\
p(x_{0:t} | y_{1:t}) &= \frac{\tilde{p}(x_{0:t})}{p(y_{1:t})} \\
Z_t &= \int \tilde{\pi}(x_{0:t}) dx_{0:t}
\end{align*}
\]
Sequential Monte Carlo approximates

\[ \pi_k(x_{0:k}) \approx \sum_{i=1}^{N} w_k^i \delta_{x_{0:k}}(x_{0:k}). \]

The weighted particle populations \( \{x_{0:k}^i, w_k^i\}_{i=1}^{N} \) are generated sequentially for \( k = 1, 2, \ldots \).
Assume that we have obtained \( \{x^{i}_{0:k-1}, w^{i}_{k-1}\}^{N}_{i=1} \).

**Resampling:** Sample \( a^{i}_{k} \) with \( \mathbb{P}(a^{i}_{k} = j) = \nu^{j}_{k-1}, j = 1, \ldots, N. \)

**Propagation:** \( x^{i}_{k} \sim q_{k}(x_{k} | x^{a^{i}_{k}}_{1:k-1}) \) and \( x^{i}_{0:k} = (x^{a^{i}_{k}}_{0:k-1}, x^{i}_{k}). \)

**Weighting:** \( w^{i}_{k} \propto \frac{w^{a^{i}_{k}}_{k-1}}{\nu^{a^{i}_{k}}_{k-1}} \frac{\tilde{\pi}_{k}(x^{i}_{0:k})}{\tilde{\pi}_{k-1}(x^{a^{i}_{k}}_{0:k-1})q_{k}(x^{i}_{k} | x^{a^{i}_{k}}_{0:k-1})} \).

The result is a new weighted set of particles \( \{x^{i}_{0:k}, w^{i}_{k}\}^{N}_{i=1} \).
SMC for probabilistic graphical models
A **probabilistic graphical model** (PGM) is a probabilistic model where a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the conditional independency structure between random variables,

1. a set of **vertices** $\mathcal{V}$ (nodes) represents the random variables
2. a set of **edges** $\mathcal{E}$ containing elements $(i, j) \in \mathcal{E}$ connecting a pair of nodes $(i, j) \in \mathcal{V} \times \mathcal{V}$
SMC methods are used to approximate a sequence of probability distributions on a sequence of spaces of increasing dimension.

**Key idea:**
1. Introduce a **sequential decomposition** of the PGM.
2. Each **subgraph** induces an intermediate target dist.
3. Apply SMC to the sequence of intermediate target dist.

Using an artificial sequence of intermediate target distributions for an SMC method is a powerful (quite possibly underutilized) idea.

**Key question:** Exactly how do we define $\{\tilde{\pi}_k(x_{1:k})\}_{k \geq 1}$?
Using a 2D lattice model from statistical physics, $x \in (-\pi, \pi]$.

\[
p(x_{\mathcal{V}}) \propto e^{-\beta H(x_{\mathcal{V}})}, \quad H(x_{\mathcal{V}}) = - \sum_{(i,j) \in \mathcal{E}} J_{ij} \cos (x_i - x_j),
\]

The **intermediate sequence** of target distributions can be chosen

\[
\tilde{\pi}_k(x_{\mathcal{L}_k}) \propto \tilde{\pi}_{k-1}(x_{\mathcal{L}_{k-1}}) e^{\kappa(x_{\mathcal{L}_{k-1}}) \cos (x_k - \mu(x_{\mathcal{L}_{k-1}}))}.
\]

$\mathcal{L}_k$ – index to the nodes in the $k^{th}$ intermediate target $\tilde{\pi}_k(x_{\mathcal{L}_k})$. 
Algorithm SMC for graphical models

1. Initialize \((k = 1)\):
   (a) Draw \(x_{\mathcal{L}_1}^i \sim q_1(\cdot)\).
   (b) Set \(w_1^i = W_1(x_{\mathcal{L}_1}^i)\).

2. For \(k = 2 \text{ to } K\) do:
   (a) Resampling: Draw \(a_k^i, \mathbb{P}(a_k^i = j) = \tilde{w}_k^{i-1}/\sum_l \tilde{w}_k^{l-1}\).
   (b) Propagation: Draw \(\xi_k^i \sim q_k(\cdot|x_{\mathcal{L}_{k-1}}^{a_k^i})\), set \(x_{\mathcal{L}_k}^i = x_{\mathcal{L}_{k-1}}^{a_k^i} \cup \xi_k^i\).
   (c) Weighting: Set \(\tilde{w}_k^i = W_k(x_{\mathcal{L}_k}^i)\).

3. End

\(\mathcal{L}_k\) – index to the nodes in the \(k^{th}\) intermediate target \(\tilde{\pi}_k(x_{\mathcal{L}_k})\).
\(\xi_k^i\) – nodes added at step \(k\).

Also provides an unbiased estimate of the normalizing constant!
This model is borrowed from

How to select the targets?

We want to sample from $p(x_V) \Rightarrow \textbf{Require: } \pi_K(x_{\mathcal{L}_K}) = p(x_V)$, but...

...all \textbf{intermediate targets} $\{\pi_k(x_{\mathcal{L}_k})\}_{k=1}^{K-1}$ are design choices!

\textbf{Idea:}

$$\tilde{\pi}_k^{\psi}(x_{\mathcal{L}_k}) = \tilde{\pi}_k(x_{\mathcal{L}_k}) \psi_k(x_{\mathcal{L}_k})$$

From graph decomp.

- In theory, possible to select $\psi_k$ to get exact samples from $p(x_V)$ at iteration $K$.
- Use \textbf{deterministic inference method} to approximate optimal $\psi_k$

\textbf{SMC works as a post-correction of the biases} associated with the deterministic inference method of choice.
Model:

- Precision: $Q_{tt} = 0.1 \times \{|Ne(t)| + 1\}$ and $Q_{tt'} = -0.1$ if $t \sim t'$
- $y_t \sim \text{Binomial}(10, \text{logit}^{-1}(x_t))$
- Spatial structure $\sim$ regions in Germany, $T = 544$
What about stability?

**Recall:** for a state-space model we need exponential forgetting for the particle filter to be stable.

**The same is true in the general case!**
If there are **strong** and **long-ranging** dependencies among the variables $X_{1:k}$ under the distribution $\pi_k$, then the asymptotic variance of SMC may be exponential in $k$.

However,

- In many applications we *do* have fast enough forgetting (though, it can be difficult to verify theoretically)
- Even if this is not the case, SMC can give good results for moderate values of $k$
Short history of SMC

- Bootstrap particle filter invented around 1992–1993
- Auxiliary particle filter, 1999
- Convergence theory: many results in the early 2000 but still an active research area
- SMC Samplers, 2006 (similar ideas going back to at least 2002)
- Particle Markov chain Monte Carlo, around 2010
- SMC for PPL, graphical models, etc. 2010–present