Sequential Monte Carlo without likelihoods

November 23, 2020

ABC setup

ABC setup

- Prior $\pi(\theta)$

ABC setup

- Prior $\pi(\theta)$
- Likelihood $\pi(y|\theta) = ?$

ABC setup

- Prior $\pi(\theta)$
- Likelihood $\pi(y|\theta) = ?$
- Observed data y_{obs}

ABC setup

- Prior $\pi(\theta)$
- Likelihood $\pi(y|\theta) = ?$
- Observed data y_{obs}
- We can't find the posterior $\pi(\theta|y_{\text{obs}}) \propto \pi(\theta)\pi(y_{\text{obs}}|\theta) = ?$

ABC setup

- Prior $\pi(\theta)$
- Likelihood $\pi(y|\theta) = ?$
- Observed data y_{obs}
- We can't find the posterior $\pi(\theta|y_{\text{obs}}) \propto \pi(\theta)\pi(y_{\text{obs}}|\theta) = ?$
- Instead find the approximate posterior

$$\pi_\epsilon(\theta|y_{\text{obs}}) \propto \pi(\theta) \int \pi(y|\theta) K_\epsilon(\rho(S(y), S(y_{\text{obs}})))\mathrm{d}y$$

ABC setup

- Prior $\pi(\theta)$
- Likelihood $\pi(y|\theta) = ?$
- Observed data y_{obs}
- We can't find the posterior $\pi(\theta|y_{\text{obs}}) \propto \pi(\theta)\pi(y_{\text{obs}}|\theta) = ?$
- Instead find the approximate posterior

$$\pi_\epsilon(\theta|y_{\text{obs}}) \propto \pi(\theta) \int \pi(y|\theta) K_\epsilon(\rho(S(y), S(y_{\text{obs}})))\mathrm{d}y$$

$$\pi_\epsilon(\theta, y|y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta) K_\epsilon(\rho(S(y), S(y_{\text{obs}})))$$

ABC Rejection sampling

- We want to sample from
  $$f(\theta, y) = \pi_\epsilon(\theta, y | y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta)K_\epsilon(\rho(S(y), S(y_{\text{obs}})))$$

ABC Rejection sampling

- We want to sample from
$$f(\theta, y) = \pi_\epsilon(\theta, y|y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta)K_\epsilon(\rho(S(y), S(y_{\text{obs}})))$$
- We can sample from
$$g(\theta, y) \propto g(\theta)\pi(y|\theta)$$

ABC Rejection sampling

- We want to sample from

$$f(\theta, y) = \pi_\epsilon(\theta, y|y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta)K_\epsilon(\rho(S(y), S(y_{\text{obs}})))$$

- We can sample from

$$g(\theta, y) \propto g(\theta)\pi(y|\theta)$$

Sampling:
For $i = 1, \ldots, N$:

1. Generate $\theta^{(i)} \sim g(\theta)$ from sampling density g.
2. Generate $y^{(i)} \sim p(y|\theta^{(i)})$ from the model.
3. Compute summary statistic $s^{(i)} = S(y^{(i)})$.
4. Accept $\theta^{(i)}$ with probability $\frac{K_h(\|s^{(i)} - s_{obs}\|)\pi(\theta^{(i)})}{Mg(\theta^{(i)})}$ where $M \geq K_h(0) \max_\theta \frac{\pi(\theta)}{g(\theta)}$.
   Else go to 1.

Output:
A set of parameter vectors $\theta^{(1)}, \ldots, \theta^{(N)} \sim \pi_{ABC}(\theta|s_{obs})$.

ABC MCMC

- We want to sample from
  $$f(\theta, y) = \pi_\epsilon(\theta, y|y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta)K_\epsilon(\rho(S(y), S(y_{\text{obs}})))$$

ABC MCMC

- We want to sample from
$$f(\theta, y) = \pi_\epsilon(\theta, y|y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta)K_\epsilon(\rho(S(y), S(y_{\text{obs}})))$$
- When the proposal distribution is
$$q((\theta', y')|(\theta, y)) \propto q(\theta'|\theta)\pi(y'|\theta')$$
the likelihoods cancel in the MH ratio.

Sampling:
For $i = 1, \dots, N$:

1. Generate candidate vector $\theta' \sim g(\theta^{(i-1)}, \theta)$ from the proposal density g.
2. Generate $y' \sim p(y|\theta')$ from the model and compute summary statistics $s' = S(y')$.
3. With probability:0
$$\min\left\{1, \frac{K_h(\|s' - s_{obs}\|)\pi(\theta')g(\theta', \theta^{(i-1)})}{K_h(\|s^{(i-1)} - s_{obs}\|)\pi(\theta^{(i-1)})g(\theta^{(i-1)}, \theta')}\right\}$$
set $(\theta^{(i)}, s^{(i)}) = (\theta', s')$. Otherwise set $(\theta^{(i)}, s^{(i)}) = (\theta^{(i-1)}, s^{(i-1)})$.

Output:
A set of correlated parameter vectors $\theta^{(1)}, \dots, \theta^{(N)}$ from a Markov chain with stationary distribution $\pi_{ABC}(\theta|s_{obs})$.

Toy example

Toy example

- Prior $\pi(\theta) \sim U[-10, 10]$

Toy example

- Prior $\pi(\theta) \sim U[-10, 10]$
- Likelihood $\pi(y|\theta) = \frac{1}{2}\mathcal{N}(y; \theta, 1) + \frac{1}{2}\mathcal{N}(y; \theta, \frac{1}{100})$

Toy example

- Prior $\pi(\theta) \sim U[-10, 10]$
- Likelihood $\pi(y|\theta) = \frac{1}{2}\mathcal{N}(y; \theta, 1) + \frac{1}{2}\mathcal{N}(y; \theta, \frac{1}{100})$
- Observed data $y_{\text{obs}} = 0$

Toy example

- Prior $\pi(\theta) \sim U[-10, 10]$
- Likelihood $\pi(y|\theta) = \frac{1}{2}\mathcal{N}(y; \theta, 1) + \frac{1}{2}\mathcal{N}(y; \theta, \frac{1}{100})$
- Observed data $y_{\text{obs}} = 0$
- Posterior

$$\pi(\theta|y_{\text{obs}}) \propto 1_{[-10,10]}(\theta) \times \left(\frac{1}{2}\mathcal{N}(\theta; 0, 1) + \frac{1}{2}\mathcal{N}(\theta; 0, \frac{1}{100})\right)$$

Toy example

- Prior $\pi(\theta) \sim U[-10, 10]$
- Likelihood $\pi(y|\theta) = \frac{1}{2}\mathcal{N}(y; \theta, 1) + \frac{1}{2}\mathcal{N}(y; \theta, \frac{1}{100})$
- Observed data $y_{\text{obs}} = 0$
- Posterior

$$\pi(\theta|y_{\text{obs}}) \propto 1_{[-10,10]}(\theta) \times \left(\frac{1}{2}\mathcal{N}(\theta; 0, 1) + \frac{1}{2}\mathcal{N}(\theta; 0, \frac{1}{100})\right)$$

- Instead find the approximate posterior

$$\pi_\epsilon(\theta|y_{\text{obs}}) \propto \pi(\theta) \int \pi(y|\theta) 1(|y - y_{\text{obs}}| \leq \epsilon)) \mathrm{d}y$$

Toy example

- Prior $\pi(\theta) \sim U[-10, 10]$
- Likelihood $\pi(y|\theta) = \frac{1}{2}\mathcal{N}(y; \theta, 1) + \frac{1}{2}\mathcal{N}(y; \theta, \frac{1}{100})$
- Observed data $y_{\text{obs}} = 0$
- Posterior

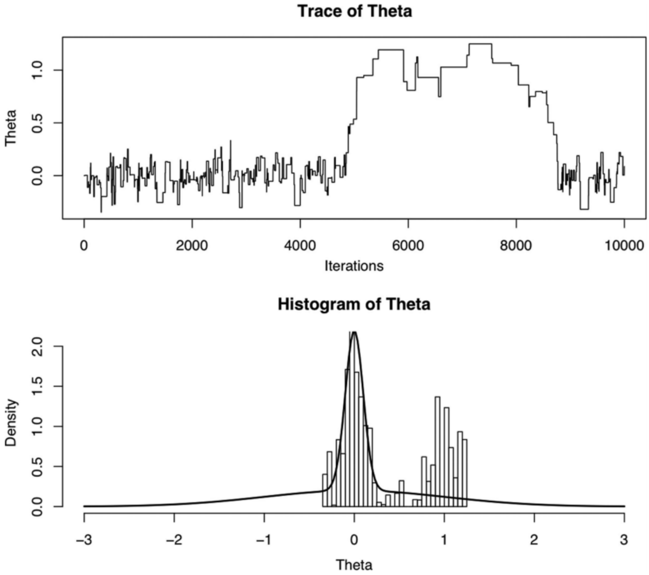
$$\pi(\theta|y_{\text{obs}}) \propto 1_{[-10,10]}(\theta) \times \left(\frac{1}{2}\mathcal{N}(\theta; 0, 1) + \frac{1}{2}\mathcal{N}(\theta; 0, \frac{1}{100})\right)$$

- Instead find the approximate posterior

$$\pi_\epsilon(\theta|y_{\text{obs}}) \propto \pi(\theta) \int \pi(y|\theta) 1(|y - y_{\text{obs}}| \leq \epsilon)) \mathrm{d}y$$

$$\pi_\epsilon(\theta, y|y_{\text{obs}}) \propto \pi(\theta)\pi(y|\theta) 1(|y - y_{\text{obs}}| \leq \epsilon)$$

Toy example



Trace of Theta
Theta
0.0
0.5
1.0
0
2000
4000
6000
8000
10000
Iterations
Histogram of Theta
Density
0.0
0.5
1.0
1.5
2.0
–3
–2
–1
0
1
2
3
Theta

Rejection Control

- Problem - regular importance sampling setup gives low weights

Rejection Control

- Problem - regular importance sampling setup gives low weights
- Solution - remove samples with small weights

Rejection Control

- Problem - regular importance sampling setup gives low weights
- Solution - remove samples with small weights
- Setup: target distribution is $f(\theta)$ and proposal distribution is $g(\theta)$

Rejection Control

- Problem - regular importance sampling setup gives low weights
- Solution - remove samples with small weights
- Setup: target distribution is $f(\theta)$ and proposal distribution is $g(\theta)$
- Sample $\theta^{(i)} \sim g(\theta)$ and set weights $w^{(i)} = \frac{f(\theta)}{g(\theta)}$

Rejection Control

- Problem - regular importance sampling setup gives low weights
- Solution - remove samples with small weights
- Setup: target distribution is $f(\theta)$ and proposal distribution is $g(\theta)$
- Sample $\theta^{(i)} \sim g(\theta)$ and set weights $w^{(i)} = \frac{f(\theta)}{g(\theta)}$
- Pick a threshold $c > 0$

Rejection Control

- Problem - regular importance sampling setup gives low weights
- Solution - remove samples with small weights
- Setup: target distribution is $f(\theta)$ and proposal distribution is $g(\theta)$
- Sample $\theta^{(i)} \sim g(\theta)$ and set weights $w^{(i)} = \frac{f(\theta)}{g(\theta)}$
- Pick a threshold $c > 0$
- Accept sample $\theta^{(i)}$ w.p. $\min\left\{1, \frac{w^{(i)}}{c}\right\}$ and recalculate weights

Rejection Control

- Problem - regular importance sampling setup gives low weights
- Solution - remove samples with small weights
- Setup: target distribution is $f(\theta)$ and proposal distribution is $g(\theta)$
- Sample $\theta^{(i)} \sim g(\theta)$ and set weights $w^{(i)} = \frac{f(\theta)}{g(\theta)}$
- Pick a threshold $c > 0$
- Accept sample $\theta^{(i)}$ w.p. $\min\left\{1, \frac{w^{(i)}}{c}\right\}$ and recalculate weights
- Resulting distribution $g^*(\theta)$ is closer to $f(\theta)$.

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \ldots, N$ through importance sampling

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \ldots, N$ through importance sampling
- Start with a good approximation ν_1 and sample population $\theta_1^{(i)} \sim \nu_1$

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \ldots, N$ through importance sampling
- Start with a good approximation ν_1 and sample population $\theta_1^{(i)} \sim \nu_1$
- Compute weights $w_1^{(i)} = \frac{\pi_1(\theta_1^{(i)})}{\nu_1(\theta_1^{(i)})}$

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \ldots, N$ through importance sampling
- Start with a good approximation ν_1 and sample population $\theta_1^{(i)} \sim \nu_1$
- Compute weights $w_1^{(i)} = \frac{\pi_1(\theta_1^{(i)})}{\nu_1(\theta_1^{(i)})}$
- For $n > 1$ sample from $\nu_n(\theta) = \int \nu_{n-1}(\theta') K_n(\theta', \theta) \mathrm{d}\theta'$ by

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \ldots, N$ through importance sampling
- Start with a good approximation ν_1 and sample population $\theta_1^{(i)} \sim \nu_1$
- Compute weights $w_1^{(i)} = \frac{\pi_1(\theta_1^{(i)})}{\nu_1(\theta_1^{(i)})}$
- For $n > 1$ sample from $\nu_n(\theta) = \int \nu_{n-1}(\theta')K_n(\theta', \theta)\mathrm{d}\theta'$ by

$$\theta_n^{(i)} \sim K_n(\theta_{n-1}^{(i)}, \cdot)$$

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \ldots, N$ through importance sampling
- Start with a good approximation ν_1 and sample population $\theta_1^{(i)} \sim \nu_1$
- Compute weights $w_1^{(i)} = \frac{\pi_1(\theta_1^{(i)})}{\nu_1(\theta_1^{(i)})}$
- For $n > 1$ sample from $\nu_n(\theta) = \int \nu_{n-1}(\theta')K_n(\theta', \theta)\mathrm{d}\theta'$ by
$$\theta_n^{(i)} \sim K_n(\theta_{n-1}^{(i)}, \cdot)$$
- Recompute the weights and continue

Sequential Importance Sampling Sequential Monte Carlo Sampler

- Our goal is to sample from π_n for $n = 1, \dots, N$ through importance sampling
- Start with a good approximation ν_1 and sample population $\theta_1^{(i)} \sim \nu_1$
- Compute weights $w_1^{(i)} = \frac{\pi_1(\theta_1^{(i)})}{\nu_1(\theta_1^{(i)})}$
- For $n > 1$ sample from $\nu_n(\theta) = \int \nu_{n-1}(\theta')K_n(\theta', \theta)\mathrm{d}\theta'$ by
$$\theta_n^{(i)} \sim K_n(\theta_{n-1}^{(i)}, \cdot)$$
- Recompute the weights and continue
- Problem - sometimes the weights are intractable
- SMC sampler solves it by introducing a backwards in time kernel L_n

ABC-PRC

- Combines partial rejection control with SMC

ABC-PRC

- Combines partial rejection control with SMC
- Sets $\pi_n(\theta) = \pi_{\epsilon_n}(\theta, y|y_{\text{obs}})$

ABC-PRC

- Combines partial rejection control with SMC
- Sets $\pi_n(\theta) = \pi_{\epsilon_n}(\theta, y|y_{\text{obs}})$

PRC1 Initialize $\epsilon_1, \ldots, \epsilon_T$, and specify initial sampling distribution μ_1.
Set population indicator $t = 1$.

PRC2 Set particle indicator $i = 1$.

PRC2.1 If $t = 1$ sample $\theta^{**} \sim \mu_1(\theta)$ independently from μ_1.
If $t > 1$ sample θ^* from the previous population $\{\theta_{t-1}^{(i)}\}$ with weights $\{W_{t-1}^{(i)}\}$, and perturb the particle to $\theta^{**} \sim K_t(\theta \mid \theta^*)$ according to a transition kernel K_t.

Generate a data set $x^{**} \sim f(x \mid \theta^{**})$.
If $\rho(S(x^{**}), S(x_0)) \geq \epsilon_t$ then go to PRC2.1.

PRC2.2 Set

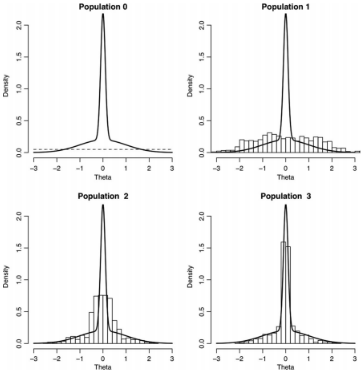
$$\theta_t^{(i)} = \theta^{**} \quad \text{and} \quad W_t^{(i)} = \begin{cases} \pi\left(\theta_t^{(i)}\right) / \mu_1\left(\theta_t^{(i)}\right) & \text{if} \quad t = 1 \\ \pi\left(\theta_t^{(i)}\right) / \sum_{j=1}^{N} W_{t-1}\left(\theta_{t-1}^{(j)}\right) K_t\left(\theta_t^{(i)} | \theta_{t-1}^{(j)}\right) & \text{if} \quad t > 1 \end{cases}$$

where $\pi(\theta)$ denotes the prior distribution for θ.
If $i < N$, increment $i = i + 1$ and go to PRC2.1.

PRC3 Normalize the weights so that $\sum_{i=1}^{N} W_t^{(i)} = 1$.
If $ESS = \left[\sum_{i=1}^{N}(W_t^{(i)})^2\right]^{-1} < E$ then resample with replacement, the particles $\{\theta_t^{(i)}\}$ with weights $\{W_t^{(i)}\}$ to obtain a new population $\{\theta_t^{(i)}\}$, and set weights $\{W_t^{(i)} = 1/N\}$.

PRC4 If $t < T$, increment $t = t + 1$ and go to PRC2.

Toy example results



Toy example results

			ABC-REJ	
t	ϵ_t	ABC-PRC	Prior	Posterior
1	2.000	4.907	–	–
2	0.500	4.899	–	–
3	0.025	66.089	400.806	21.338
	Total	75.895	400.806	21.338

Case study for Tuberculosis Transmission rates

t	ϵ	ABC-PRC	ABC-REJ
1	1.000	2.595	
2	0.5013	8.284	
3	0.2519	8.341	
4	0.1272	7.432	
5	0.0648	10.031	
6	0.0337	17.056	
7	0.0181	34.178	
8	0.0102	72.704	
9	0.0064	171.656	
10	0.0025	1,089.006	7,206.333
Total		1,421.283	7,206.333