Sequential Monte Carlo

Rich Turner
Why Sequential Monte Carlo is important: flexible approximate inference

Goal: Approximate inference in non-linear sequence models

\[
p(y_{1:T}, x_{1:T} | \theta) = \prod_{t=1}^{T} p(x_t | x_{1:t-1}, \theta)p(y_t | x_t, \theta)
\]

\[
p(x_{1:T} | y_{1:T}, \theta) = \frac{1}{Z} p(x_{1:T}, y_{1:T} | \theta)
\]

e.g. 1. non-linear dynamical systems

\[
p(x_t | x_{1:t-1}, \theta) = p(x_t | x_{t-1}, \theta) = \mathcal{N}(x_t; f_\theta(x_{t-1}), \Sigma_\theta(x_{t-1}))
\]

e.g. 2. non-parametric methods defined through stochastic processes

Incorporating constraints into inferences

Leveraging a physical simulator for inference

Fundamental Monte Carlo method

- rejection sampling
- importance sampling
- sequential Monte Carlo
- MCMC
- particle-MCMC
- pseudo-marginal
Importance sampling review: normalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)  

Goal is not: \( x \sim p(x) \)

nasty: cannot integrate  

cannot sample  

can point-wise evaluate
Importance sampling review: normalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)  
Goal is not: \( x \sim p(x) \)

\( F = \int f(x)p(x)\frac{q(x)}{q(x)}dx \)
Importance sampling review: normalised target density

Goal: \( \hat{F} \approx \int f(x)p(x) \, dx = F \)

\[
F = \int f(x)p(x) \frac{q(x)}{q(x)} \, dx
\]

Goal is not: \( x \sim p(x) \)

\( x^{(n)} \sim q(x^{(n)}) \)

nasty: cannot integrate
cannot sample
can point-wise evaluate
Importance sampling review: normalised target density

**Goal:** \( \hat{F} \approx \int f(x)p(x)dx = F \)

\[
F = \int f(x)p(x)\frac{q(x)}{q(x)}dx
\]

\[
\approx \frac{1}{N} \sum_{n=1}^{N} f(x^{(n)})\frac{p(x^{(n)})}{q(x^{(n)})} = \hat{F}
\]

\( x^{(n)} \sim q(x^{(n)}) \)

**Goal is not:** \( x \sim p(x) \)

- nasty: cannot integrate
- cannot sample
- can point-wise evaluate
- unbiased
Importance sampling review: normalised target density

Goal: \[ \hat{F} \approx \int f(x)p(x) \, dx = F \]

Goal is not: \( x \sim p(x) \)

\[ F = \int f(x)p(x) \frac{q(x)}{q(x)} \, dx \]

\[ \approx \frac{1}{N} \sum_{n=1}^{N} f(x^{(n)}) \frac{p(x^{(n)})}{q(x^{(n)})} = \hat{F} \]

Choose \( q(x) \) to min variance in \( \hat{F} \) (often intractable) or close to \( p(x) \)

nasty: cannot integrate
cannot sample
can point-wise evaluate
unbiased
Importance sampling review: normalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)

\[ F = \int f(x)p(x)\frac{q(x)}{q(x)}dx \]

\[ \approx \frac{1}{N} \sum_{n=1}^{N} f(x^{(n)})\frac{p(x^{(n)})}{q(x^{(n)})} = \hat{F} \]

\( x^{(n)} \sim q(x^{(n)}) \)

Goal is not: \( x \sim p(x) \)

Choose \( q(x) \) to min variance in \( \hat{F} \) (often intractable) or close to \( p(x) \)

\[ \hat{F} = \sum_{n=1}^{N} w^{(n)} f(x^{(n)}) \]

\[ w^{(n)} = \frac{1}{N} \frac{p(x^{(n)})}{q(x^{(n)})} \]
Importance sampling review: normalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)\,dx = F \)

Goal is not: \( x \sim p(x) \)

\[
F = \int f(x)p(x)\frac{q(x)}{q(x)}\,dx \\
\approx \frac{1}{N} \sum_{n=1}^{N} f(x^{(n)})\frac{p(x^{(n)})}{q(x^{(n)})} = \hat{F}
\]

\( x^{(n)} \sim q(x^{(n)}) \)

Choose \( q(x) \) to min variance in \( \hat{F} \) (often intractable) or close to \( p(x) \)

\[
\hat{F} = \sum_{n=1}^{N} w^{(n)} f(x^{(n)}) \quad w^{(n)} = \frac{1}{N} \frac{p(x^{(n)})}{q(x^{(n)})} \\
p(x) \approx \sum_{n=1}^{N} \delta(x - x^{(n)})w^{(n)}
\]
Importance sampling review: unnormalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)

\[ p(x) = \frac{1}{Z} p^*(x) \]

Goal is not: \( x \sim p(x) \)

can evaluate

nasty: cannot integrate
cannot sample
cannot evaluate
Importance sampling review: unnormalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)

Goal is not: \( x \sim p(x) \)

\[ p(x) = \frac{1}{Z} p^*(x) \]

idea: apply IS approximation twice

\[ Z = \int p^*(x)dx \]

\[ F = \frac{1}{Z} \int f(x)p^*(x)dx \]
Importance sampling review: unnormalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)  

Goal is not: \( x \sim p(x) \)

\[ p(x) = \frac{1}{Z}p^*(x) \]

idea: apply IS approximation twice

can evaluate

cannot integrate

cannot sample

cannot evaluate

\[ Z = \int p^*(x)dx \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p^*(x^{(n)})}{q(x^{(n)})} = \sum_{n=1}^{N} w^{(n)} = \hat{Z} \]

\[ x^{(n)} \sim q(x^{(n)}) \]

\[ F = \frac{1}{Z} \int f(x)p^*(x)dx \]
Importance sampling review: unnormalised target density

Goal: \[ \hat{F} \approx \int f(x)p(x)dx = F \]

Goal is not: \[ x \sim p(x) \]

\[ p(x) = \frac{1}{Z}p^*(x) \]

can evaluate

cannot integrate
cannot sample
cannot evaluate

idea: apply IS approximation twice

\[ Z = \int p^*(x)dx \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p^*(x^{(n)})}{q(x^{(n)})} = \sum_{n=1}^{N} w^{(n)} = \hat{Z} \]

\[ x^{(n)} \sim q(x^{(n)}) \]

\[ F = \frac{1}{\hat{Z}} \int f(x)p^*(x)dx \approx \frac{1}{\hat{Z}} \sum_{n=1}^{N} f(x^{(n)})w^{(n)} = \hat{F} \]

use same samples for both approximations
Importance sampling review: unnormalised target density

Goal: \( \hat{F} \approx \int f(x)p(x)dx = F \)  
Goal is not: \( x \sim p(x) \)

\[
p(x) = \frac{1}{Z} p^*(x)
\]

idea: apply IS approximation twice

\[
Z = \int p^*(x)dx \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p^*(x^{(n)})}{q(x^{(n)})} = \sum_{n=1}^{N} w^{(n)} = \hat{Z}  
\]

\[
F = \frac{1}{Z} \int f(x)p^*(x)dx \approx \frac{1}{\hat{Z}} \sum_{n=1}^{N} f(x^{(n)})w^{(n)} = \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x^{(n)})
\]

use same samples for both approximations

normalised weights: \( \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_{n} w^{(n)}} \)

nasty: cannot integrate  
cannot sample  
cannot evaluate  
can evaluate
Importance sampling review: unnormalised target density

Goal: \( \hat{F} \approx \int f(x) p(x) dx = F \)  
Goal is not: \( x \sim p(x) \)

\[ p(x) = \frac{1}{Z} p^*(x) \]

idea: apply IS approximation twice

can evaluate

cannot integrate

cannot sample

cannot evaluate

Ratio of two estimates: biased, asymptotically unbiased

\[
\hat{F} = \frac{1}{\hat{Z}} \int f(x) p^*(x) dx \approx \frac{1}{\hat{Z}} \sum_{n=1}^{N} f(x^{(n)}) \tilde{w}^{(n)} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x^{(n)}) = \hat{F} = \sum_{n=1}^{N} \frac{w^{(n)}}{\sum_n w^{(n)}} f(x^{(n)})
\]

use same samples for both approximations

normalised weights: \( \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}} \)
Importance sampling review: time-series latent variable models

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z}p(x_{1:T}, y_{1:T})dx_{1:T} = F \)
Importance sampling review: time-series latent variable models

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T})dx_{1:T} = F \)

idea: apply IS approximation twice \( x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \)

\[
Z = \int p(x_{1:T}, y_{1:T})dx_{1:T}
\]

\[
F = \frac{1}{Z} \int f(x_{1:T})p(x_{1:T}, y_{1:T})dx_{1:T}
\]
Importance sampling review: time-series latent variable models

Goal: \( \hat{F} \approx \int f(x_{1:T}) p(x_{1:T} | y_{1:T}) dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T}) dx_{1:T} = F \)

idea: apply IS approximation twice \( x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \)

\[
Z = \int p(x_{1:T}, y_{1:T}) dx_{1:T} \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p(x_{1:T}^{(n)}, y_{1:T})}{q(x_{1:T}^{(n)})} = \sum_{n=1}^{N} w^{(n)} = \hat{Z}
\]

\[
F = \frac{1}{Z} \int f(x_{1:T}) p(x_{1:T}, y_{1:T}) dx_{1:T}
\]
Importance sampling review: time-series latent variable models

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T})dx_{1:T} = F \)

can evaluate \( p^*(x_{1:T}) \)

idea: apply IS approximation twice \( x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \)

\( Z = \int p(x_{1:T}, y_{1:T})dx_{1:T} \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p(x_{1:T}^{(n)}, y_{1:T})}{q(x_{1:T}^{(n)})} = \sum_{n=1}^{N} w^{(n)} = \hat{Z} \)

use same samples for both approximations

\( F = \frac{1}{Z} \int f(x_{1:T})p(x_{1:T}, y_{1:T})dx_{1:T} \approx \frac{1}{\hat{Z}} \sum_{n=1}^{N} f(x_{1:T}^{(n)})w^{(n)} = \hat{F} \)
Importance sampling review: time-series latent variable models

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T})dx_{1:T} = F \)

idea: apply IS approximation twice

\( x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \)

\[ Z = \int p(x_{1:T}, y_{1:T})dx_{1:T} \approx \frac{1}{N} \sum_{n=1}^{N} \frac{p(x_{1:T}^{(n)}, y_{1:T})}{q(x_{1:T}^{(n)})} = \sum_{n=1}^{N} w^{(n)} = \hat{Z} \]

use same samples for both approximations

\[ F = \frac{1}{Z} \int f(x_{1:T})p(x_{1:T}, y_{1:T})dx_{1:T} \approx \frac{1}{\hat{Z}} \sum_{n=1}^{N} f(x_{1:T}^{(n)})w^{(n)} = \hat{F} \]

\[ x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x_{1:T}^{(n)}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}} \]
Sequential Importance Sampling

Goal: $\hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T})\frac{1}{Z}p(x_{1:T}, y_{1:T})dx_{1:T} = F$

$x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x_{1:T}^{(n)}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}}$

$w_T = \frac{p(x_{1:T}^{(n)}, y_{1:T})}{q(x_{1:T}^{(n)})}$
Sequential Importance Sampling

Goal: \( \hat{F} \approx \int f(x_{1:T}) p(x_{1:T} | y_{1:T}) \, dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T}) \, dx_{1:T} = F \)

\[
p(x_{1:T}, y_{1:T}) = \prod_{t=1}^{T} p(x_t | x_{1:t-1}) p(y_t | x_t) \quad q(x_{1:T}) = \prod_{t=1}^{T} q(x_t | x_{1:t-1})
\]

\( x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x_{1:T}^{(n)}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}} \)

\( \omega_T = \frac{p(x_{1:T}^{(n)}, y_{1:T})}{q(x_{1:T}^{(n)})} \)
Sequential Importance Sampling

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T})dx_{1:T} = F \)

\[
p(x_{1:T}, y_{1:T}) = \prod_{t=1}^{T} p(x_t|x_{1:t-1})p(y_t|x_t) \quad q(x_{1:T}) = \prod_{t=1}^{T} q(x_t|x_{1:t-1})
\]

compute weights via a recursion:

\[
w_T = \frac{p(x_{1:T}, y_{1:T})}{q(x_{1:T})}
\]

\[
x^{(n)}_{1:T} \sim q(x^{(n)}_{1:T}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x^{(n)}_{1:T}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}}
\]

\[
w_T = \frac{p(x^{(n)}_{1:T}, y_{1:T})}{q(x^{(n)}_{1:T})}
\]
Sequential Importance Sampling

Goal: \( \hat{F} \approx \int f(x_{1:T}) p(x_{1:T} | y_{1:T}) \, dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z} p(x_{1:T}, y_{1:T}) \, dx_{1:T} = F \)

\[
p(x_{1:T}, y_{1:T}) = \prod_{t=1}^{T} p(x_t | x_{1:t-1}) p(y_t | x_t) \quad q(x_{1:T}) = \prod_{t=1}^{T} q(x_t | x_{1:t-1})
\]

compute weights via a recursion:

\[
w_T = \frac{p(x_{1:T}, y_{1:T})}{q(x_{1:T})} = \frac{p(x_T | x_{1:T-1}) p(y_T | x_T)}{q(x_T | x_{1:T-1})} \frac{p(x_{1:T-1}, y_{1:T-1})}{q(x_{1:T-1})}
\]

\[
x_{1:T}^{(n)} \sim q(x_{1:T}^{(n)}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x_{1:T}^{(n)}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}}
\]
Sequential Importance Sampling

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})\,dx_{1:T} = \int f(x_{1:T})\frac{1}{Z}p(x_{1:T},y_{1:T})\,dx_{1:T} = F \)

\[
p(x_{1:T},y_{1:T}) = \prod_{t=1}^{T} p(x_t|x_{1:t-1})p(y_t|x_t) \quad q(x_{1:T}) = \prod_{t=1}^{T} q(x_t|x_{1:t-1})
\]

compute weights via a recursion:

\[
w_T = \frac{p(x_{1:T},y_{1:T})}{q(x_{1:T})} = \frac{p(x_T|x_{1:T-1})p(y_T|x_T)}{q(x_T|x_{1:T-1})} \frac{p(x_{1:T-1},y_{1:T-1})}{q(x_{1:T-1})}
\]

\[
= \frac{p(x_T|x_{1:T-1})p(y_T|x_T)}{q(x_T|x_{1:T-1})} w_{T-1}
\]

simple recursion

\[
w_T = \frac{p(x_{1:T},y_{1:T})}{q(x_{1:T})}
\]

\[
x^{(n)}_{1:T} \sim q(x^{(n)}_{1:T}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x^{(n)}_{1:T}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}}
\]
Sequential Importance Sampling

Goal: \( \hat{F} \approx \int f(x_{1:T})p(x_{1:T}|y_{1:T})dx_{1:T} = \int f(x_{1:T}) \frac{1}{Z}p(x_{1:T}, y_{1:T})dx_{1:T} = F \)

\[
p(x_{1:T}, y_{1:T}) = \prod_{t=1}^{T} p(x_t|x_{1:t-1})p(y_t|x_t) \quad q(x_{1:T}) = \prod_{t=1}^{T} q(x_t|x_{1:t-1})
\]

compute weights via a recursion:

\[
w_T = \frac{p(x_{1:T}, y_{1:T})}{q(x_{1:T})} = \frac{p(x_T|x_{1:T-1})p(y_T|x_T)}{q(x_T|x_{1:T-1})} \frac{p(x_{1:T-1}, y_{1:T-1})}{q(x_{1:T-1})} = \frac{p(x_T|x_{1:T-1})p(y_T|x_T)}{q(x_T|x_{1:T-1})} w_{T-1}
\]

simple case:

\[
q(x_T|x_{1:T-1}) = p(x_T|x_{1:T-1}) \Rightarrow w_t = p(y_t|x_t)w_{t-1} \quad w_T = \frac{p(x^{(n)}_{1:T}, y_{1:T})}{q(x^{(n)}_{1:T})}
\]

\[
x^{(n)}_{1:T} \sim q(x^{(n)}_{1:T}) \quad \hat{F} = \sum_{n=1}^{N} \tilde{w}^{(n)} f(x^{(n)}_{1:T}) \quad \tilde{w}^{(n)} = \frac{w^{(n)}}{\sum_n w^{(n)}}
\]
Sequential Importance Sampling

approximation at t-1

\[ p(x_{1:t-1}|y_{1:t-1}) \approx \sum_{n=1}^{N} \tilde{w}_{t-1}^{(n)} \delta(x_{1:t-1} - x_{1:t-1}^{(n)}) \]
Sequential Importance Sampling

approximation at t-1

\[
p(x_{1:t-1}|y_{1:t-1}) \approx \sum_{n=1}^{N} \tilde{w}_{t-1}^{(n)} \delta(x_{1:t-1} - x_{1:t-1}^{(n)})
\]

sample new locations

\[
q(x_t|x_{1:t-1}) = p(x_t|x_{1:t-1})
\]
Sequential Importance Sampling

approximation at $t-1$

$$p(x_{1:t-1}|y_{1:t-1}) \approx \sum_{n=1}^{N} \tilde{w}_{t-1}^{(n)} \delta(x_{1:t-1} - x_{1:t-1}^{(n)})$$

sample new locations

$$q(x_{t}|x_{1:t-1}) = p(x_{t}|x_{1:t-1})$$

reweight using likelihood and renormalise

$$w_{t}^{(n)} = \tilde{w}_{t-1}^{(n)} p(y_{t}|x_{t}^{(n)})$$

$$\tilde{w}_{t}^{(n)} = \frac{w_{t}^{(n)}}{\sum_{n} w_{t}^{(n)}}$$
Sequential Importance Sampling

approximation at $t-1$

$$p(x_{1:t-1}|y_{1:t-1}) \approx \sum_{n=1}^{N} \tilde{w}_{t-1}^{(n)} \delta(x_{1:t-1} - x_{1:t-1}^{(n)})$$

sample new locations

$$q(x_t|x_{1:t-1}) = p(x_t|x_{1:t-1})$$

reweight using likelihood and renormalise

$$w_t^{(n)} = \tilde{w}_{t-1}^{(n)} p(y_t|x_t^{(n)})$$

$$\tilde{w}_t^{(n)} = \frac{w_t^{(n)}}{\sum_n w_t^{(n)}}$$

approximation at $t$

$$p(x_{1:t}|y_{1:t}) \approx \sum_{n=1}^{N} \tilde{w}_t^{(n)} \delta(x_{1:t} - x_{1:t}^{(n)})$$
Sequential Importance Sampling

approximation at \( t-1 \)

\[
p(x_{1:t-1}|y_{1:t-1}) \approx \sum_{n=1}^{N} \tilde{w}^{(n)}_{t-1} \delta(x_{1:t-1} - x^{(n)}_{1:t-1})
\]

sample new locations

\[
q(x_t|x_{1:t-1}) = p(x_t|x_{1:t-1})
\]

reweight using likelihood and renormalise

\[
w_t^{(n)} = \tilde{w}_{t-1}^{(n)} p(y_t|x_t^{(n)}) \quad \tilde{w}_t^{(n)} = \frac{w_t^{(n)}}{\sum_n w_t^{(n)}}
\]

approximation at \( t \)

\[
p(x_{1:t}|y_{1:t}) \approx \sum_{n=1}^{N} \tilde{w}_t^{(n)} \delta(x_{1:t} - x^{(n)}_{1:t})
\]

repeat
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$.
E.g. \( N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, y_t = x_t^2 + 0.1x_t + \eta_t, \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. \( N = 100, \; x_t = 0.9 x_{t-1} + 0.4 \epsilon_t, \; y_t = x_t^2 + 0.1 x_t + \eta_t, \; \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. \( N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, \ y_t = x_t^2 + 0.1x_t + \eta_t, \ \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0,1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. \( N = 100, \ x_t = 0.9x_{t-1} + 0.4\epsilon_t, \ y_t = x_t^2 + 0.1x_t + \eta_t/10, \ \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. \( N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, y_t = x_t^2 + 0.1x_t + \eta_t/10, \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
Sequential Importance Resampling

\[
p(x_{1:t-1}|y_{1:t-1}) \approx \sum_{n=1}^{N} \tilde{w}_{t-1}^{(n)} \delta(x_{1:t-1} - x_{1:t-1}^{(n)})
\]

if \( \text{Neff} < \text{thresh} \), resample \( N \) times from approx. and reset weights

\[
\text{Neff} = \frac{1}{\sum_{n=1}^{N} \left( \tilde{w}_{t-1}^{(n)} \right)^2}
\]

sample new locations

\[
q(x_t|x_{1:t-1}) = p(x_t|x_{1:t-1})
\]

reweight using likelihood and renormalise

\[
w_t^{(n)} = \tilde{w}_{t-1}^{(n)} p(y_t|x_t^{(n)}) \quad \tilde{w}_t^{(n)} = \frac{w_t^{(n)}}{\sum_n w_t^{(n)}}
\]

\[
p(x_{1:t}|y_{1:t}) \approx \sum_{n=1}^{N} \tilde{w}_t^{(n)} \delta(x_{1:t} - x_{1:t}^{(n)})
\]
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0,1)$
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$.
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. \( N = 100, \ x_t = 0.9x_{t-1} + 0.4\epsilon_t, \ y_t = x_t^2 + 0.1x_t + \eta_t/10, \ \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. \( N = 100, \ x_t = 0.9x_{t-1} + 0.4\epsilon_t, \ y_t = x_t^2 + 0.1x_t + \eta_t/10, \ \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. $N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, y_t = x_t^2 + 0.1x_t + \eta_t/10, \epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. \( N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, \ y_t = x_t^2 + 0.1x_t + \eta_t/10, \ \epsilon_t, \eta_t \sim \mathcal{N}(0,1) \)
E.g. $N = 100$, $x_t = 0.9x_{t-1} + 0.4\epsilon_t$, $y_t = x_t^2 + 0.1x_t + \eta_t/10$, $\epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
E.g. \( N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, y_t = x_t^2 + 0.1x_t + \eta_t/10, \epsilon_t, \eta_t \sim \mathcal{N}(0, 1) \)
E.g. $N = 100, x_t = 0.9x_{t-1} + 0.4\epsilon_t, y_t = x_t^2 + 0.1x_t + \eta_t/10, \epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$
better ways to resample (e.g. residual resampling and stratified sampling, Doucet 2001), adaptive proposals (e.g. neural adaptive sequential SMC)

lots of convergence results and analysis Doucet and Johansen 2011 (asymptotically unbiased, converges at rate $1/\sqrt{N}$)

path degeneracy and particle smoothing: smoothing estimates have very high-variance (new ways to smooth)

combining SMC and MCMC: particle MCMC methods (application to e.g. non-parametric models)

other ways of combining importance sampling and MCMC (e.g. annealed importance sampling)

combination with variational inference methods (3 simultaneous papers: Variational Sequential Monte Carlo, Filtering Variational Objectives, Auto-encoding sequential Monte Carlo)