UiO : Matematisk institutt

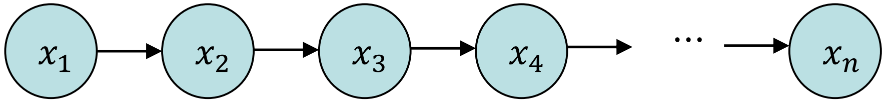
Det matematisk-naturvitenskapelige fakultet

STK-4051/9051 Computational Statistics Spring 2021
Sequential Monte Carlo (some retake)

Instructor: Odd Kolbjørnsen, oddkol@math.uio.no

Recap

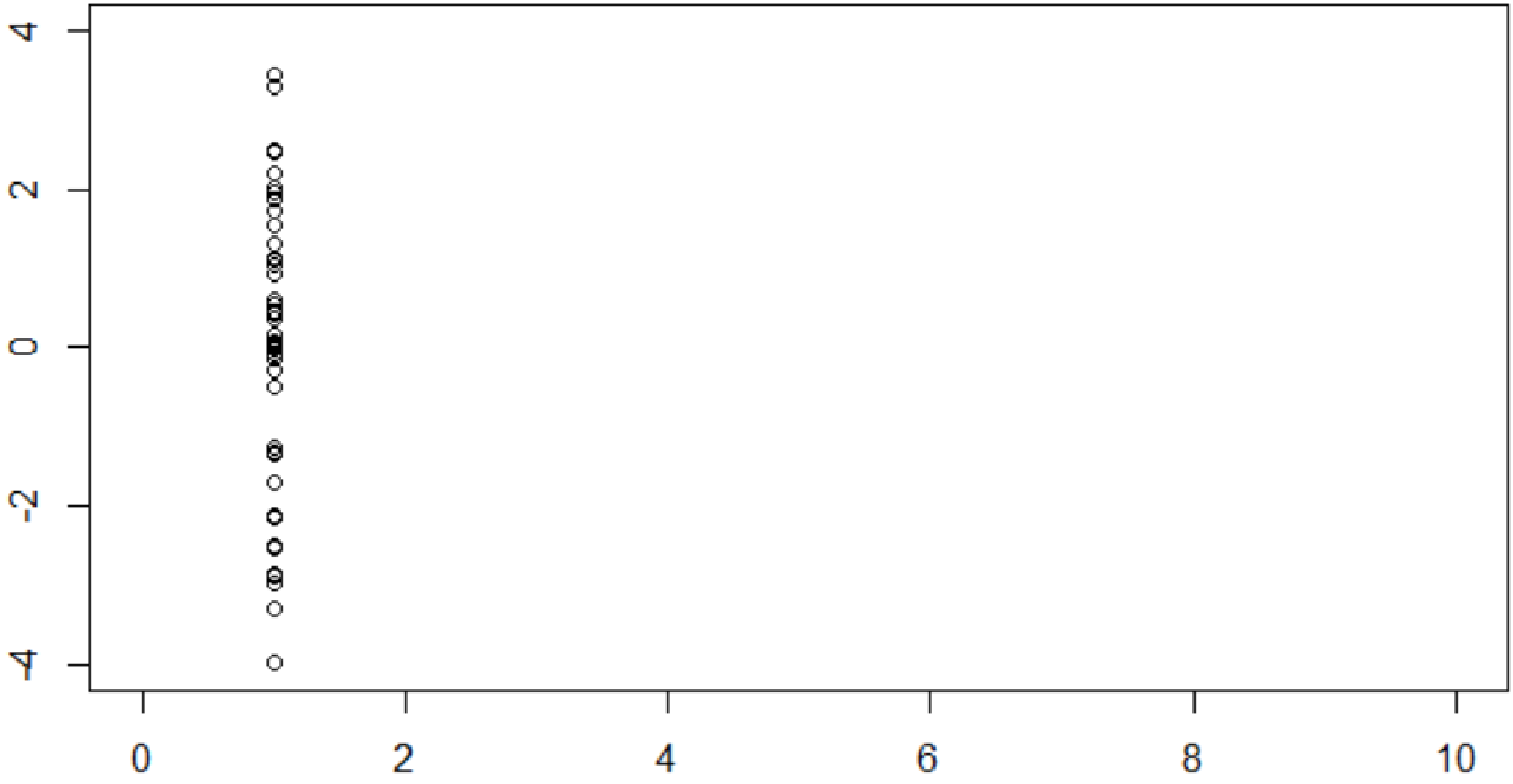
- Sequential Mote Carlo $x_t \sim f(x_t|x_{t-1})$,
  - $x_t^i \sim g(x_t^i|x_{t-1}^i)$, $w_t^i = w_{t-1}^i \frac{f(x_t^i|x_{t-1}^i)}{g(x_t^i|x_{t-1}^i)}$
  - Problem: weigth accumulation, $N_{\text{eff}} \rightarrow 1$
  - Particle filter - Resampling
  - New problem: Degeneracy



$$f(\boldsymbol{x}) = f(x_1)f(x_2|x_1)f(x_3|x_2)\cdots f(x_n|x_{n-1})$$

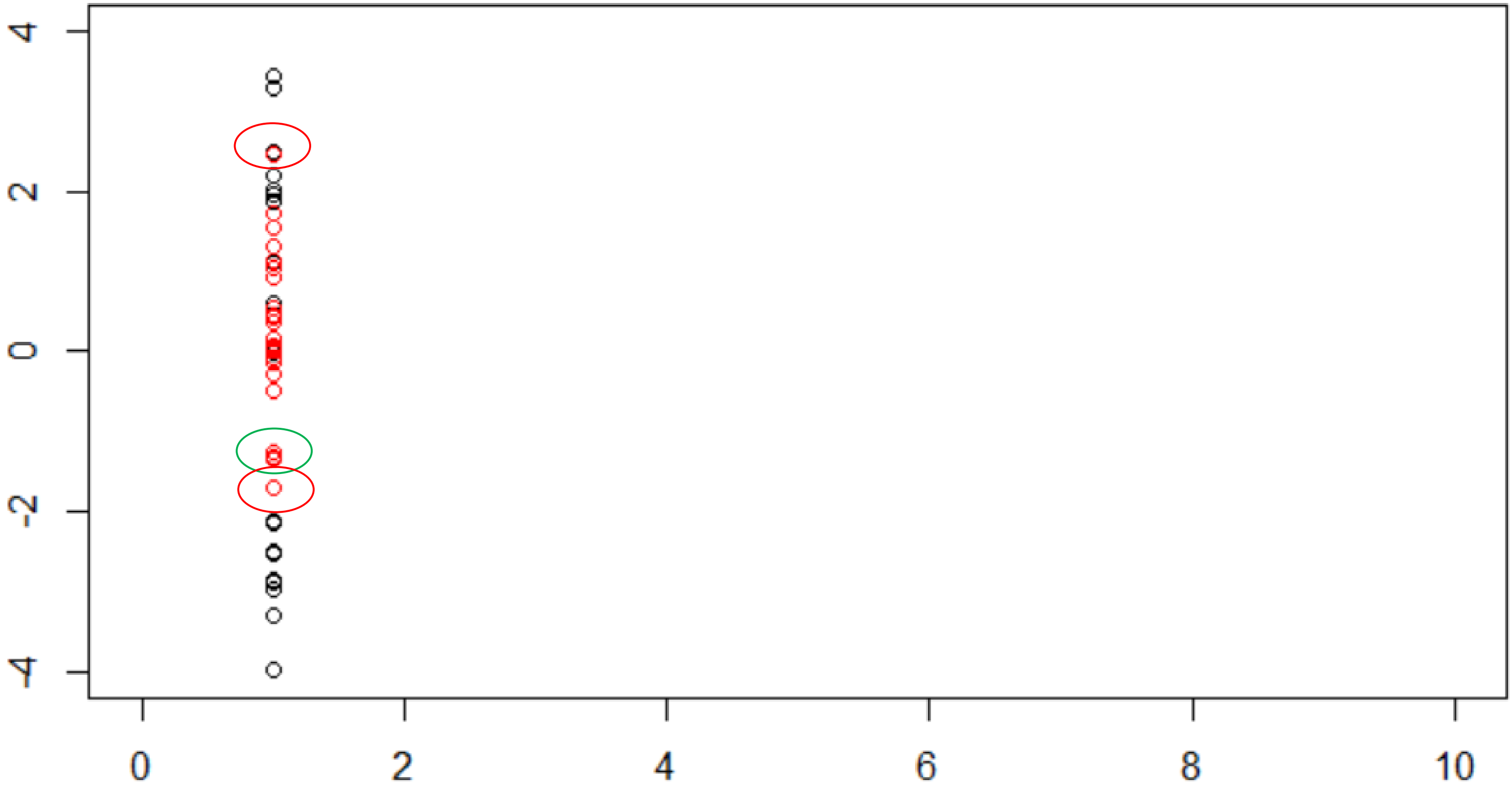
$$g(\boldsymbol{x}) = g(x_1)g(x_2|x_1)g(x_3|x_2)\cdots g(x_n|x_{n-1})$$

Proposal



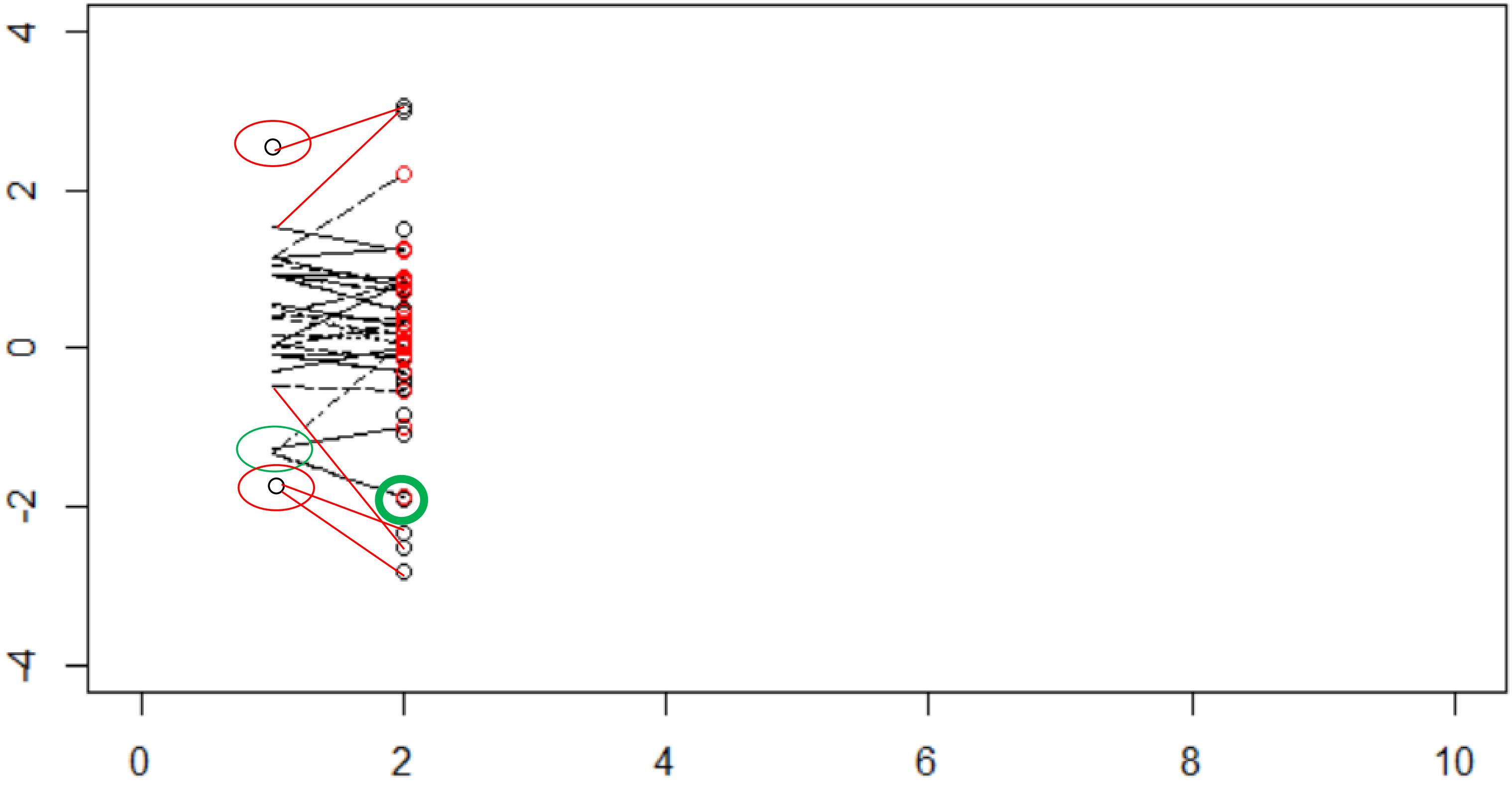
Resample =red

Red is a good approximation to $p(x_1)$



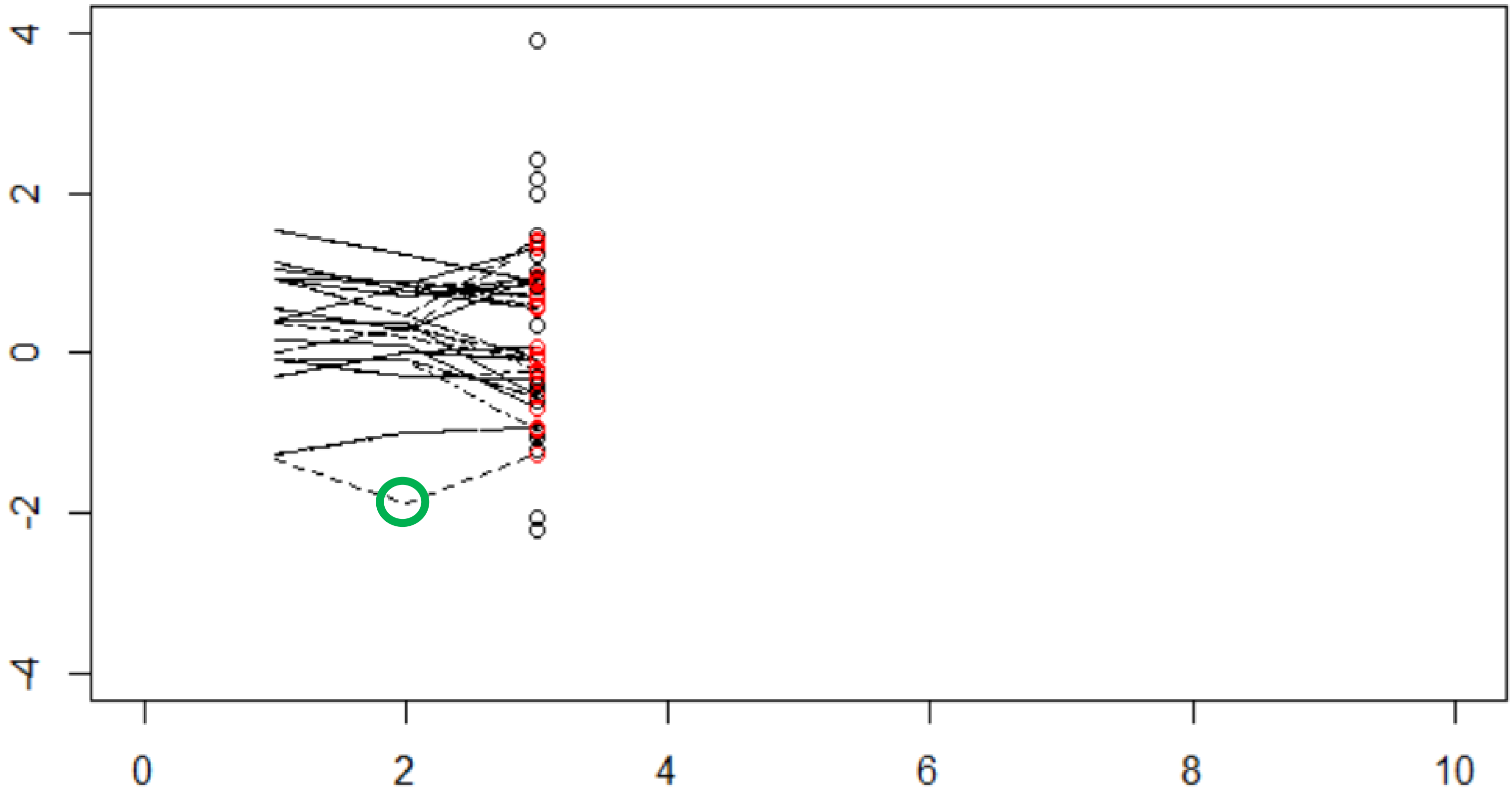
Step 2

Red is a good approximation to $p(x_2)$



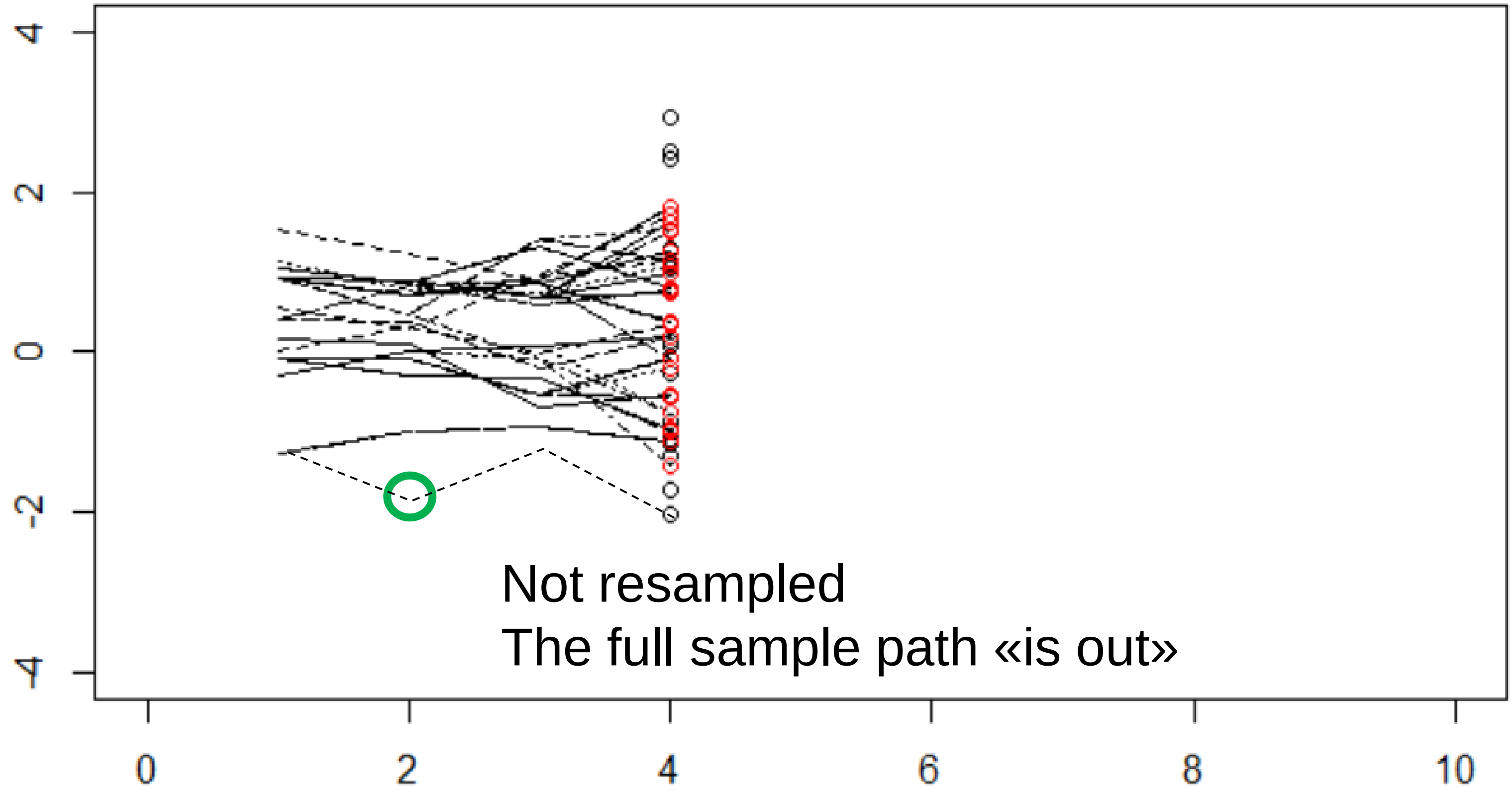
Step 3

Red is a good approximation to $p(x_3)$



Step 4

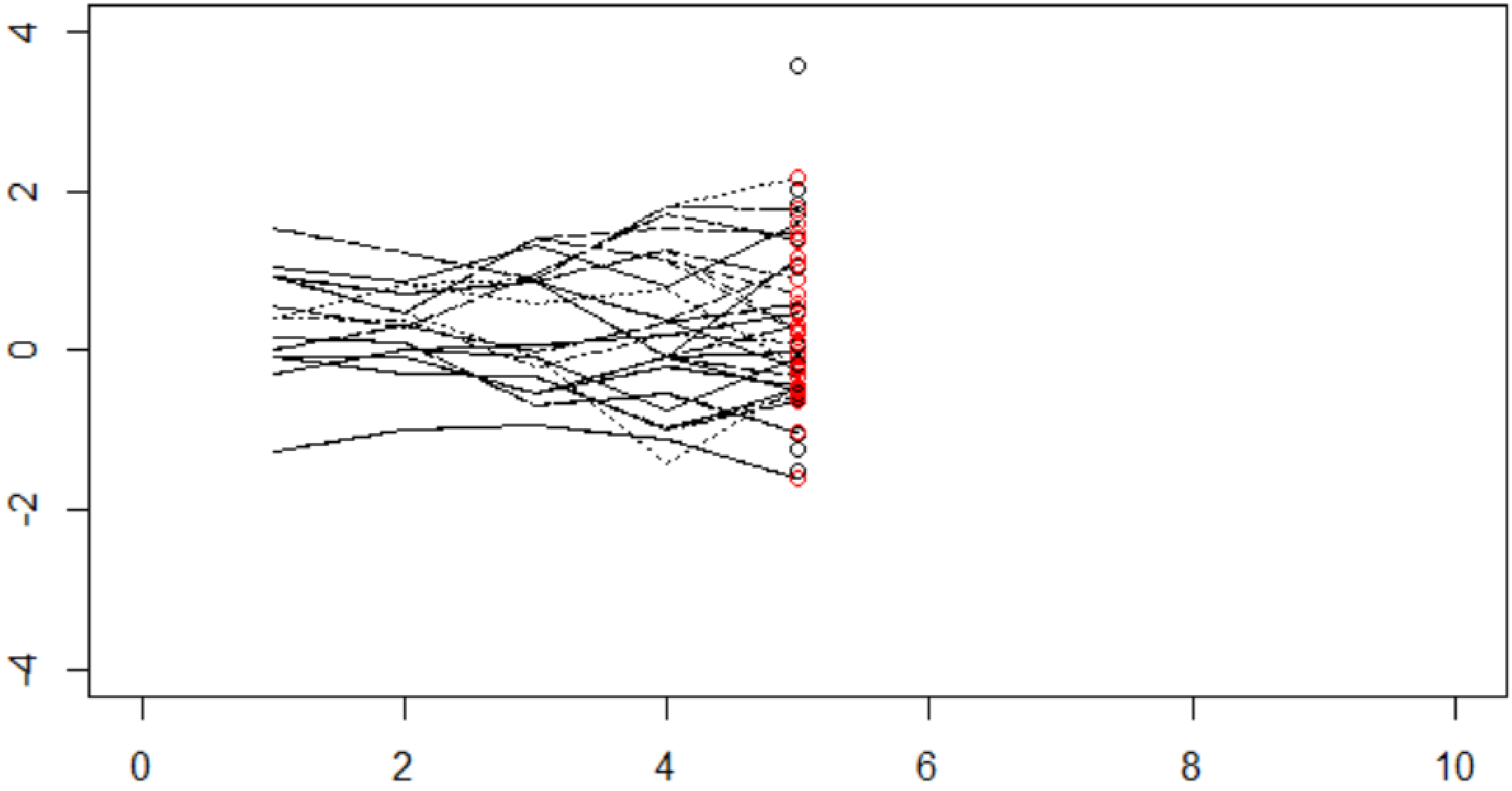
Red is a good approximation to $p(x_4)$



Not resampled
The full sample path «is out»

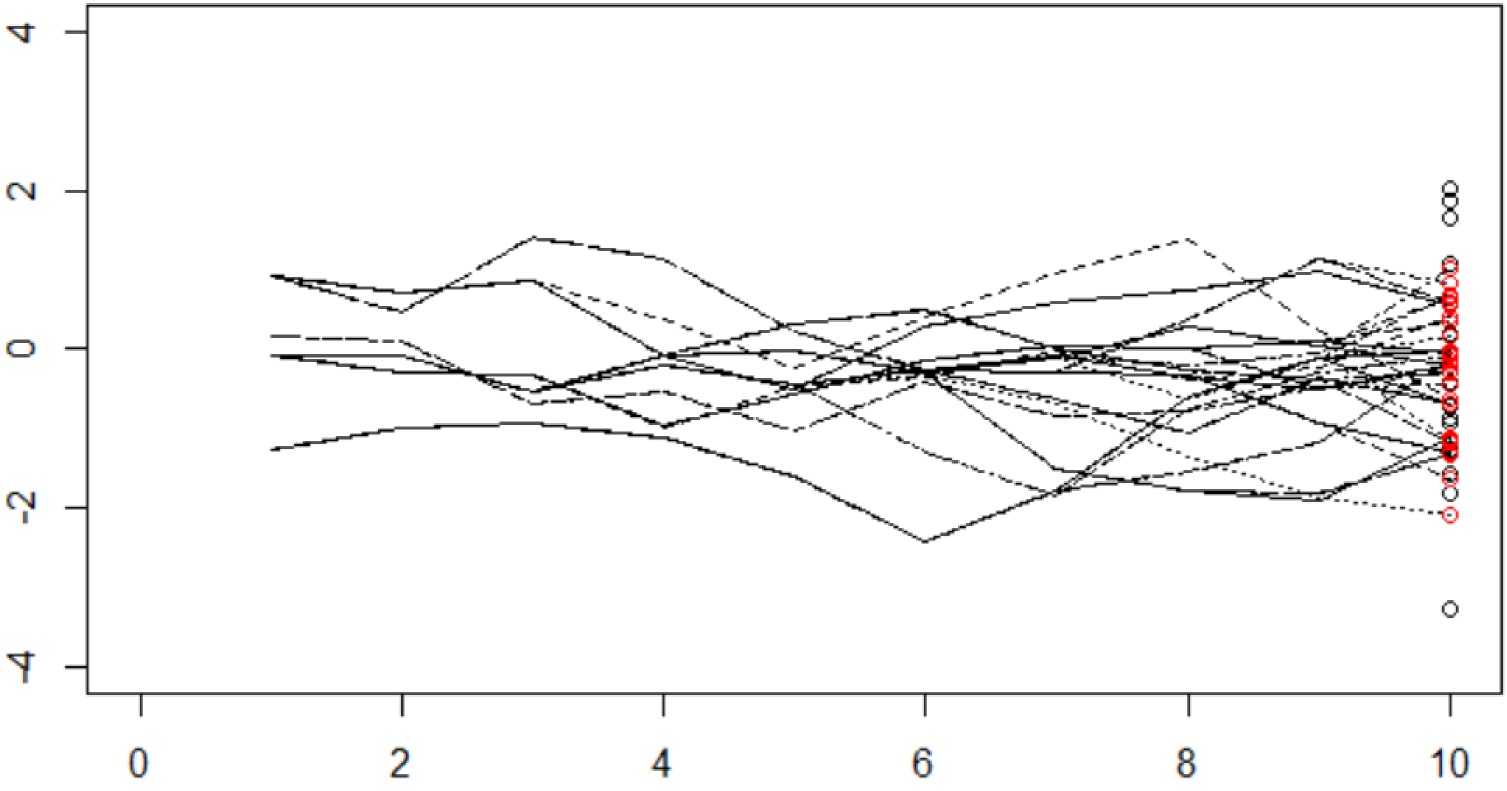
Step 5

Red is a good approximation to $p(x_5)$

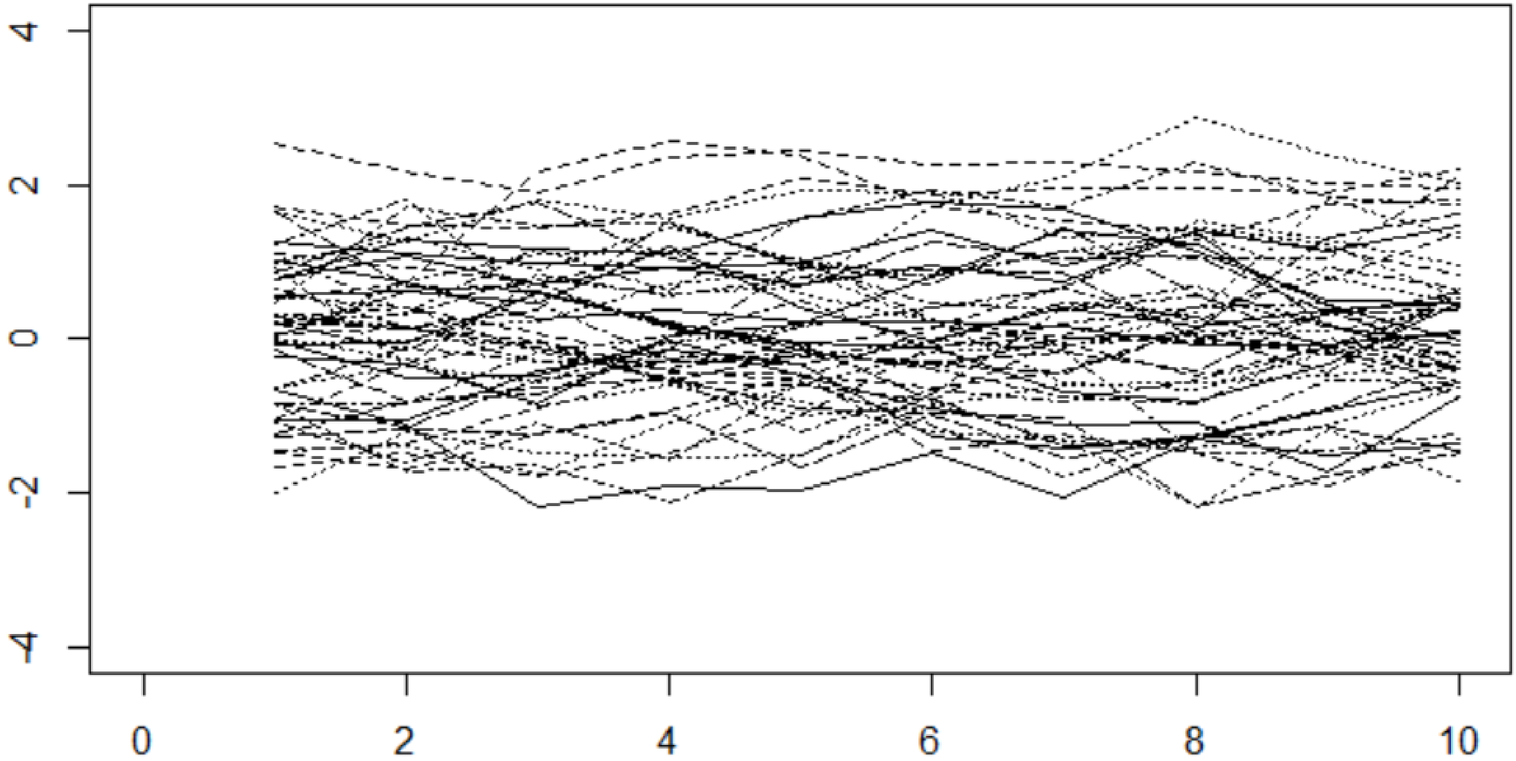


50 samples using resampling

Red is a good approximation to $p(x_{10})$



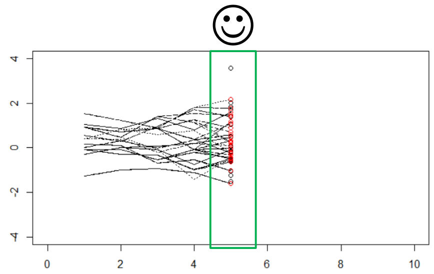
How 50 samples should look



When is what a good approximation?

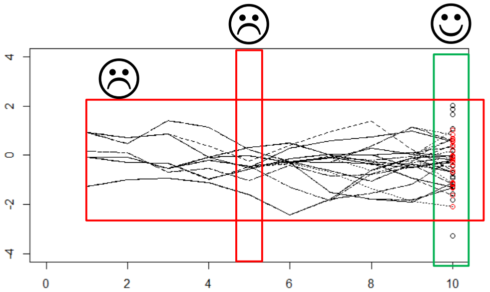
Sequential Monte Carlo particle filter:

- Sampling along the way is a good approximation for the marginal distribution at each step, $f(x_t)$



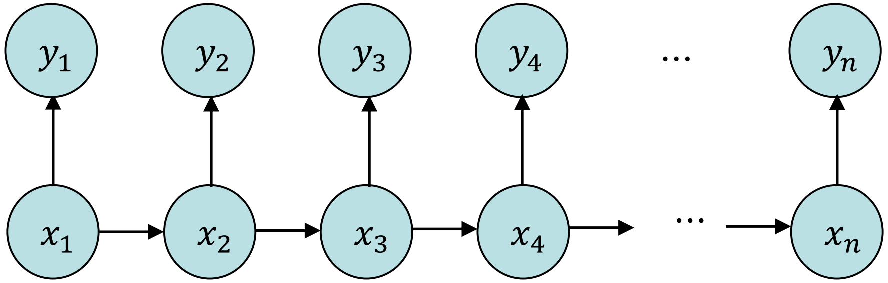
- At the end you can trace back the sample path for the final sample, this could have been an approximation for a sample from full distribution, but it is degenerated. Few samples in the beginning of the path. $f(x_1, \ldots, x_t, \ldots, x_n)$ is not good (as seen)

- If you try to use the final sample to get $f(x_t)$, this is not a good sample



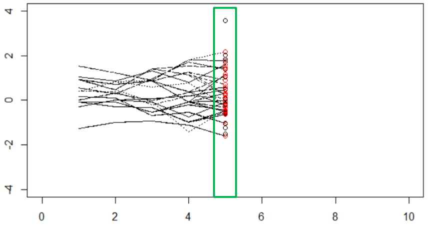
Last time cont…

- Sequential Monte Carlo for Hidden Markov Model
  - Want to sample from $f(x_t) \propto p(x_t^i | x_{t-1}^i)p(y_t | x_t^i)$
  - General $x_t^i \sim q\left(x_t | x_{t-1}^i, y_n\right), \ w_t^i = w_{t-1}^i \frac{p(x_t^i | x_{t-1}^i)p(y_t | x_t^i)}{q\left(x_t^i | x_{t-1}^i, y_n\right)}$
  - Bootstrap filter $x_t^i \sim p(x_t | x_{t-1}^i), \ \ w_t^i = w_{t-1}^i p(y_t | x_t^i)$
- Same problems as before:
  - Weigth accumulation, $N_{\text{eff}} \rightarrow 1$
  - Resampling → Degeneracy
  - Terrain navigation

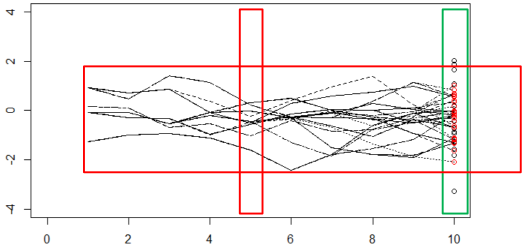


When is what a good approximation?

Sequential Monte Carlo bootstrap filter:



- Sampling along the way is a good approximation for the marginal distribution at each step, $f(x_t|y_{1:t})$
- At the end you can trace back the sample path for the final sample, this could have been an approximation for a sample from full distribution, but it is degenerated. Few samples in the beginning of the path. $f(x_{1:n}|y_{1:n})$ is not good (as seen)

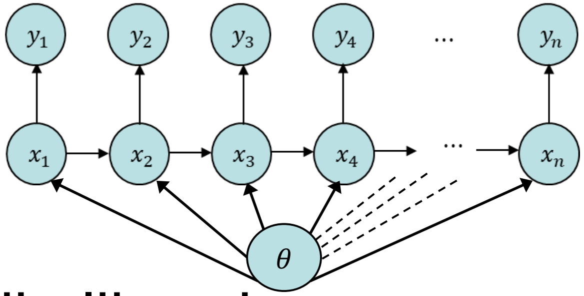


- If you try to use the final sample to get $f(x_t|y_{1:n})$, this is not a good sample
- So you solve the filter problem, $f(x_t|y_{1:t})$
not the smoothing problem, $f(x_t|y_{1:n})$
not the full conditioning problem $f(x_{1:n}|y_{1:n})$

The difference is which data you condition to

Last time cont…

- Inference in Sequential Markov models



- Maximum likelihood
  - Likelihood by sampling (bootstrap filter)

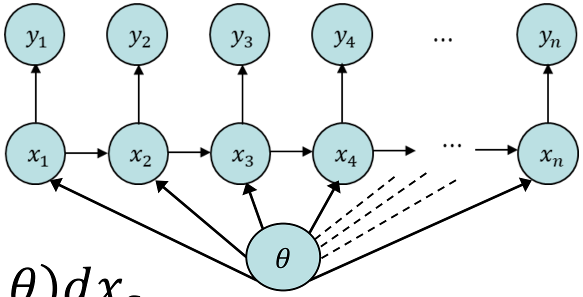
θ =?

$$p(\mathbf{y}_{1:t}|\theta) = p(y_1|\theta)\prod_{s=2}^{t}\underbrace{p(y_s|\mathbf{y}_{1:s-1};\theta)}\xrightarrow{s=2,\dots,t}\approx\sum_{i=1}^{N} w_{t-1}^{i}\,p(y_s|x_s^{i};\theta)$$

- Bayesian approach
  - Direct approach
  - Dynamic approach
  - Sufficient statistics ← Repeat

$p(\theta)$ prior
Want $p(\theta|y_{1:n})$
(posterior)

Maximum likelihood in SMC



$$p(\mathbf{y}_{1:t}|\theta) = p(y_1|\theta)\prod_{s=2}^{t} p(y_s|\mathbf{y}_{1:s-1};\theta)$$

$$p(y_s|y_{1:s-1},\theta) = \int_{x_s} p(x_s|y_{1:s-1},\theta)p(y_s|x_s,\theta)dx_s$$

$$= \int_{x_s}\int_{x_{s-1}} p(x_s,x_{s-1}|y_{1:s-1},\theta)p(y_s|x_s,\theta)dx_{s-1}dx_s$$

$$= \int_{x_s}\int_{x_{s-1}} p(x_s|x_{s-1},\theta)\underbrace{p(x_{s-1}|y_{1:s-1},\theta)}p(y_s|x_s,\theta)dx_{s-1}dx_s$$

We sample this
To get x_s^i

We have this for $s-1$ with:
$(x_{s-1}^i, w_{s-1}^i),\ i = 1,\dots,M$

We get the weight
update for x_s^i

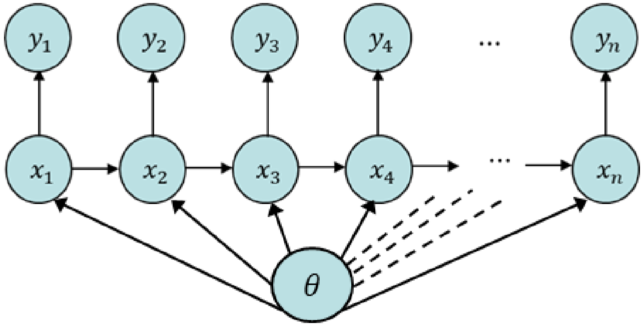
$$u_s^i = p\left(y_s \middle| x_s^i,\theta\right)$$

Sample for s: $\boxed{(x_s^i, u_s^i w_{s-1}^i),\ i = 1,\dots,M}$

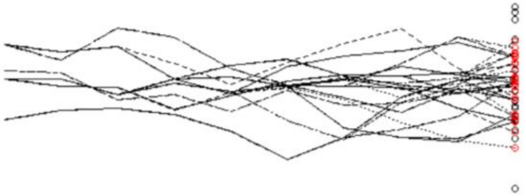
Issue: Direct approach

Direct approach samples theta initially and
is never able to resample it thus at each resampling
it is a thinning/depletion of unique elements

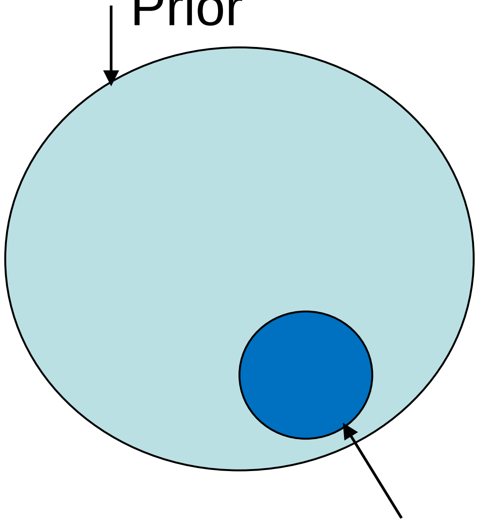
$$p(x_t, \theta|\mathbf{y}_{1:t}) \approx c \cdot \sum_{i=1}^{N} w_{t-1}^i p(x_t|x_{t-1}^i, \theta^i)\delta_\theta(\theta^i)p(y_t|x_t, \theta^i)$$



θ is sampled
here

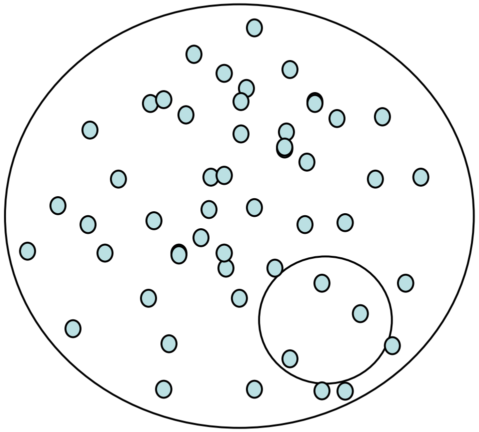


Prior



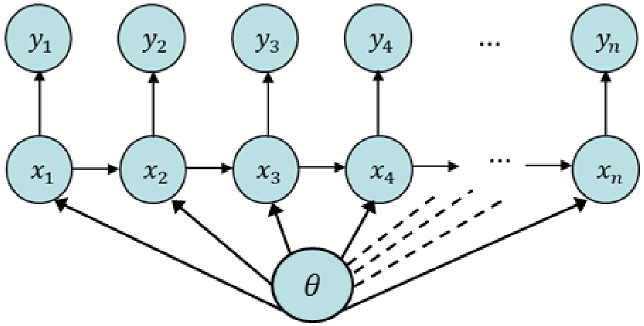
Posterior

Sample from prior

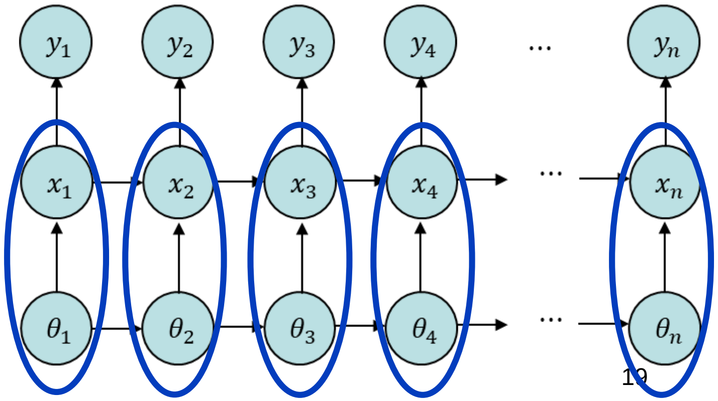


Issue: Dynamic approach

Dynamic approach. We can resample θ_t at each time step
This solves the problem of degeneration
(but we solve a different problem…)



Original problem

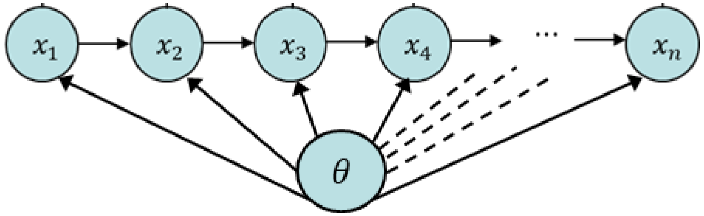


New problem

This problem is equivalet with
the «no θ problem»

$$\tilde{x}_t = (\underbrace{x_t, \theta_t}_{\boxed{\text{«Dynamic duo»}}})$$

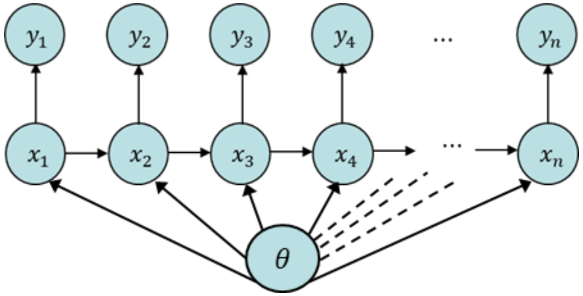
Approach using sufficient statistics



- We can update the statistics recursively
- We can compute sufficient statistics for θ

$p(\theta|\mathbf{x}_{1:t}) = p(\theta|S_t)$, S_t sufficient statistic.

- This opens for resampling of theta ☺
- We can expand this to the hidden Markov model as long as the hidden variable is the one influenced by theta
- This keeps the original conditioning structure (as x_t “protect” θ from y_t)



SMC and sufficient statistics general

- Assume $p(y_t|x_t)$ do not depend on θ.
- Assume $p(\theta|\mathbf{x}_{1:t}) = p(\theta|S_t)$, S_t sufficient statistic.
- Assume $S_t = h(S_{t-1}, x_{t-1}, x_t)$
- Fearnhead (2002) and Storvik (2002): Focus on $p(\overbrace{x_t, S_t}^{\text{«Dynamic duo»}}|\mathbf{y}_{1:t})$, not $p(x_t, \theta|\mathbf{y}_{1:t})$.
- Assume a properly weighted sample $\{(x_{t-1}^i, S_{t-1}^i, w_{t-1}^i), i = 1, \ldots, N\}$ with respect to $p(x_{t-1}, S_{t-1}|\mathbf{y}_{1:t-1})$
- Similar recursions as before:

$$
\begin{aligned}
p(x_t, S_t|\mathbf{y}_{1:t-1}) &= \int_{x_{t-1}} p(x_t, S_t|x_{t-1}, S_{t-1})p(x_{t-1}, S_{t-1}|\mathbf{y}_{1:t-1})dx_{t-1}dS_{t-1} \\
&\approx \sum_{i=1}^{N} w_{t-1}^i p(x_t, S_t|x_{t-1}^i, S_{t-1}^i) \\
p(x_t, S_t|\mathbf{y}_{1:t}) &\approx c \cdot \sum_{i=1}^{N} w_{t-1}^i p(x_t, S_t|x_{t-1}^i, S_{t-1}^i)p(y_t|x_t).
\end{aligned}
$$

- Simulation from $p(x_t, S_t|x_{t-1}^i, S_{t-1}^i)$ (possible proposal function)
  1. Simulate $\theta^i \sim p(\theta|x_{t-1}^i, S_{t-1}^i) = p(\theta|S_{t-1}^i)$.
  2. Simulate $x_t^i \sim p(x_t|x_{t-1}^i, \theta^i)$.
  3. Put $S_t^i = h(S_{t-1}^i, x_{t-1}^i, x_t^i)$.

Bootstrap filter
for the «dynamic duo»…

Algorithm Storvik filter

Algorithm 3 SMC with parameter updating

```
Simulate $\theta^i \sim p(\theta)$ for $i = 1, \dots, N$.                ▷ Initialization
Simulate $x_1^i \sim p(x_1|\theta^i)$ for $i = 1, \dots, N$.
Put weights $w_1^i = p(y_1|x_1^i)$.
Put $S_1^i = 0$ for $i = 1, \dots, N$.
for $t = 2, 3, \dots$ do                                   ▷ Sequential Monte Carlo
    Simulate $\theta^i \sim p(\theta|S_{t-1}^i)$ for $i = 1, \dots, N$.
    Simulate $x_t^i \sim p(x_t|x_{t-1}^i, \theta^i)$ for $i = 1, \dots, N$.
    Put weights $w_t^i = w_{t-1}^i p(y_t|x_t^i)$.
    Put $S_t^i = h(S_{t-1}^i, x_{t-1}^i, x_t^i)$.
    if $\hat{N}_{eff}$ is small then                                  ▷ Resampling
        Resample $(x_t^i, S_t^i)$ with probabilities proportional to $w_t^i$.
        Put $w_t^i = 1/N$.
    end if
end for
```

`SMC_lin_bin_parest_suff.R`

Example Lemmings evolving parameter: a