UPPSALA
UNIVERSITET

Sequential Monte Carlo methods

Lecture 10 – Properties of the likelihood estimator

Andreas Lindholm, Uppsala University

2019-08-27

Aim: Provide a better understanding for the properties of the particle filter likelihood estimator.

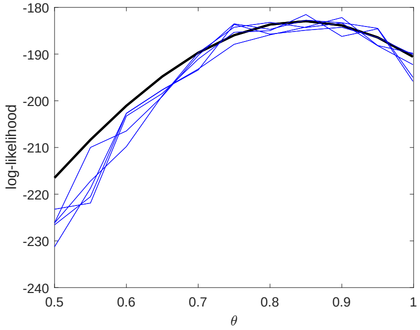
Outline:

1. The particle filter sampling distribution
2. Unbiasedness of the likelihood estimator
3. Central limit theorems

Simple LG-SSM,

$$X_t = \theta X_{t-1} + V_t, \qquad V_t \sim \mathcal{N}(0,1),$$
$$Y_t = X_t + E_t, \qquad E_t \sim \mathcal{N}(0,1).$$

Task: estimate $p(y_{1:100} \mid \theta)$ for a simulated data set. True $\theta^\star = 0.9$.



Black line – true likelihood computed using the Kalman filter.

Blue thin lines – 5 different likelihood estimates $\widehat{p}^N(y_{1:100} \mid \theta)$ computed using a bootstrap particle filter with $N = 100$ particles.

Bootstrap PF likelihood estimator

UPPSALA UNIVERSITET

The particle filter likelihood estimator,

$$\widehat{Z} = \prod_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \widetilde{W}_t^i \right\}$$

is a **random variable.**

Bootstrap PF likelihood estimator

UPPSALA UNIVERSITET

The particle filter likelihood estimator,

$$\widehat{Z} = \prod_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \widetilde{W}_t^i \right\}$$

is a **random variable.**

If we run the PF algorithm multiple times (with the same data $y_{1:T}$) we will get different realizations of this random variable, $\hat{z}[1], \hat{z}[2], \ldots,$ all of which estimate $p(y_{1:T} \mid \theta)$.

The particle filter likelihood estimator,

$$\widehat{Z} = \prod_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \widetilde{W}_t^i \right\}$$

is a **random variable.**

If we run the PF algorithm multiple times (with the same data $y_{1:T}$) we will get different realizations of this random variable, $\hat{z}[1], \hat{z}[2], \ldots,$ all of which estimate $p(y_{1:T} \mid \theta)$.

What can be said about the distribution and properties of the random variable $\widehat{Z}$?

Bootstrap PF likelihood estimator

The particle filter likelihood estimator,

$$\widehat{Z} = \prod_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \widetilde{W}_t^i \right\}$$

is a **random variable.**

If we run the PF algorithm multiple times (with the same data $y_{1:T}$) we will get different realizations of this random variable, $\hat{z}[1], \hat{z}[2], \dots,$ all of which estimate $p(y_{1:T} \mid \theta)$.

What can be said about the distribution and properties of the random variable $\widehat{Z}$?

N.B. From now on we consider the likelihood estimate for a fixed value of θ and thus drop θ from the notation $\Rightarrow$ task is to estimate $p(y_{1:T})$.

The particle filter sampling distribution

Use of random numbers in the particle filter

The particle filter uses random numbers to

1. initialize
2. resample and
3. propagate

the particles.

The weights, and **thus also the likelihood estimator**, are deterministic functions of these random numbers.

Particle filter sampling distribution

A particle filter that is run for time steps $t = 0, \ldots, T$ samples the random variables

$$\mathsf{X}_t = \{X_t^i\}_{i=1}^N, \qquad t = 0, \ldots, T,$$
$$\mathsf{A}_t = \{A_t^i\}_{i=1}^N, \qquad t = 1, \ldots, T,$$

with distributions (for the bootstrap PF):

$$\mathsf{X}_0 \sim \prod_{i=1}^N p(x_0^i) \qquad \text{(Initialization)}$$

$$\mathsf{A}_t \,|\, (\mathsf{X}_{t-1} = \mathsf{x}_{t-1}) \sim \prod_{i=1}^N w_{t-1}^{a_t^i} \qquad \text{(Resampling)}$$

$$\mathsf{X}_t \,|\, (\mathsf{X}_{t-1} = \mathsf{x}_{t-1}, \mathsf{A}_t = \mathsf{a}_t) \sim \prod_{i=1}^N p(x_t^i \,|\, x_{t-1}^{a_t^i}) \qquad \text{(Propagation)}$$

Particle filter sampling distribution

Let $\mathbf{X}_{0:T} = (\mathbf{X}_0, \dots \mathbf{X}_T)$ and $\mathbf{A}_{1:T} = (\mathbf{A}_1, \dots \mathbf{A}_T)$.

The distribution of **all the random variables** sampled by the bootstrap PF is thus,

$$\psi_{N,T}(\mathbf{x}_{0:T}, \mathbf{a}_{1:T}) = \left\{ \prod_{i=1}^{N} p(x_0^i) \right\} \prod_{t=1}^{T} \left\{ \prod_{i=1}^{N} w_{t-1}^{a_t^i} p(x_t^i \mid x_{t-1}^{a_t^i}) \right\},$$

with domain $\mathcal{X}^{N(T+1)} \times \{1, \dots, N\}^{NT}$.

Particle filter sampling distribution

Let $\mathbf{X}_{0:T} = (\mathbf{X}_0, \dots \mathbf{X}_T)$ and $\mathbf{A}_{1:T} = (\mathbf{A}_1, \dots \mathbf{A}_T)$.

The distribution of **all the random variables** sampled by the bootstrap PF is thus,

$$\psi_{N,T}(\mathbf{x}_{0:T}, \mathbf{a}_{1:T}) = \left\{ \prod_{i=1}^{N} p(x_0^i) \right\} \prod_{t=1}^{T} \left\{ \prod_{i=1}^{N} w_{t-1}^{a_t^i} p(x_t^i \mid x_{t-1}^{a_t^i}) \right\},$$

with domain $\mathcal{X}^{N(T+1)} \times \{1, \dots, N\}^{NT}$.

Executing the particle filter algorithm can be viewed as a way of generating **one sample** from this distribution!

Distribution of the likelihood estimator

The likelihood estimator $\widehat{Z}$ is a function of the random variables $\mathrm{X}_{0:T}$ and $\mathrm{A}_{1:T}$.

The distribution $\psi_{N,T}(\mathrm{x}_{0:T}, \mathrm{a}_{1:T})$ induces a distribution for $\widehat{Z}$ which we also denote by $\psi_{N,T}(z)$

$$\widehat{Z} \sim \psi_{N,T}(z), \qquad z \in \mathbb{R}_+.$$

The likelihood estimator $\widehat{Z}$ is a function of the random variables $\mathbf{X}_{0:T}$ and $\mathbf{A}_{1:T}$.

The distribution $\psi_{N,T}(\mathbf{x}_{0:T}, \mathbf{a}_{1:T})$ induces a distribution for $\widehat{Z}$ which we also denote by $\psi_{N,T}(z)$

$$\widehat{Z} \sim \psi_{N,T}(z), \qquad z \in \mathbb{R}_{+}.$$

Theorem: Unbiasedness of the likelihood estimator

The likelihood estimator $\widehat{Z}$ is unbiased, i.e.

$$\mathbb{E}_{\psi_{N,T}}\left[\widehat{Z}\right] = p(y_{1:T})$$

for **any number of particles** $N \geq 1$.

(Holds for the general auxiliary particle filter, though we have only discussed the bootstrap particle filter here.)

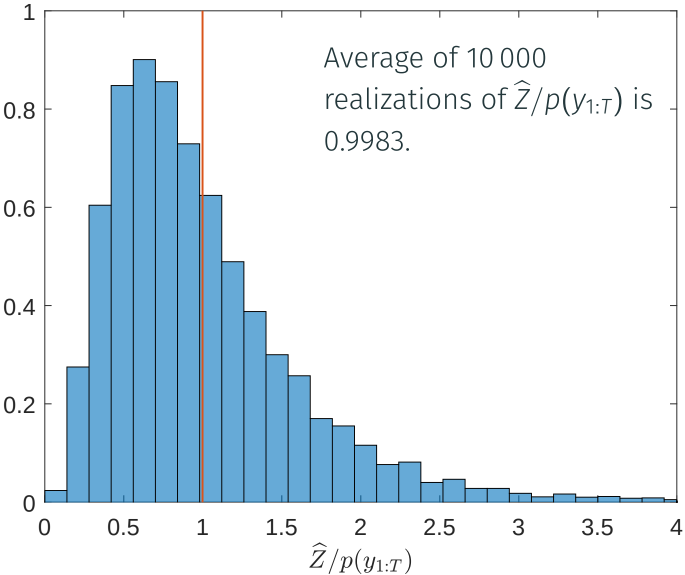
Simple LG-SSM,

$$
\begin{aligned}
X_t &= 0.9X_{t-1} + V_t, & V_t &\sim \mathcal{N}(0,1),\\
Y_t &= X_t + E_t, & E_t &\sim \mathcal{N}(0,1).
\end{aligned}
$$

Task: estimate $p(y_{1:T})$ for a **small** simulated data set consisting of $T = 20$ measurements.

Note that the “ground truth” can be computed using a Kalman filter.

Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 100$ particles.



Central limit theorems

Theorem: CLT for likelihood estimator

The likelihood estimator of the bootstrap particle filter satisfies a central limit theorem: With $\widehat{Z} \sim \psi_{N,T}(z)$,

$$\sqrt{N}\left(\frac{\widehat{Z}}{p(y_{1:T})} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sum_{t=0}^{T}\left\{\int \frac{p(x_t \mid y_{1:T})^2}{p(x_t \mid y_{1:t-1})} \mathrm{d}x_t - 1\right\}\right)$$

as $N \to \infty$.

Theorem: CLT for likelihood estimator

The likelihood estimator of the bootstrap particle filter satisfies a central limit theorem: With $\widehat{Z} \sim \psi_{N,T}(z)$,

$$\sqrt{N}\left(\frac{\widehat{Z}}{p(y_{1:T})} - 1\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \sum_{t=0}^{T}\left\{\int \frac{p(x_t \mid y_{1:T})^2}{p(x_t \mid y_{1:t-1})} \mathrm{d}x_t - 1\right\}\right)$$

as $N \to \infty$.

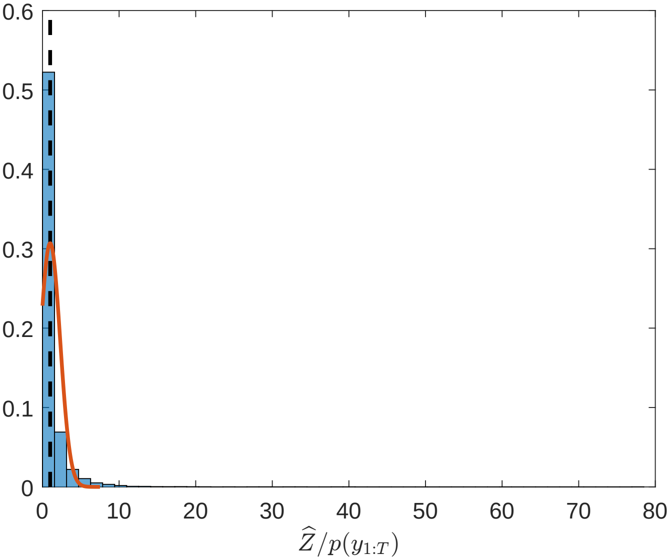
Under certain **exponential forgetting conditions** (recall lecture 5), one can show that the variance is

$$\mathrm{Var}_{\psi_{N,T}}\left[\frac{\widehat{Z}}{p(y_{1:T})}\right] \approx \frac{CT}{N}$$

for some constant $C < \infty$.

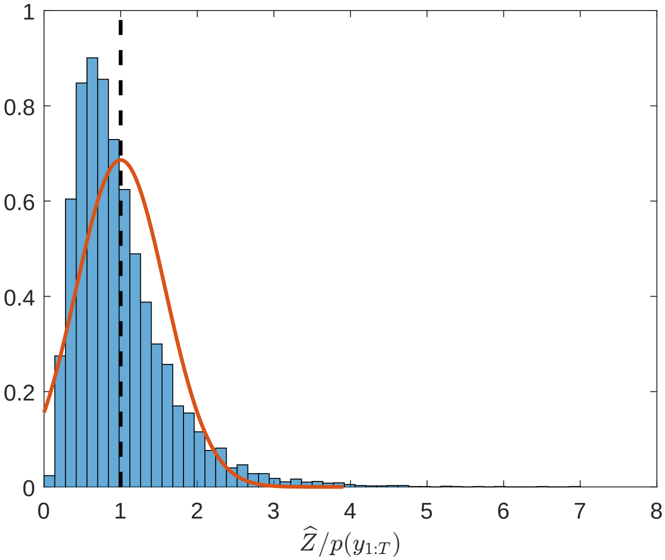
ex) Numerical illustration, cont'd

Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 20$ particles.

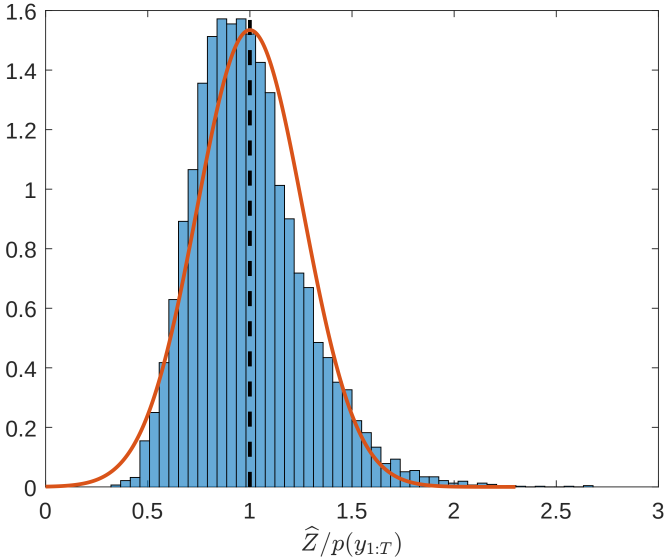


ex) Numerical illustration, cont'd

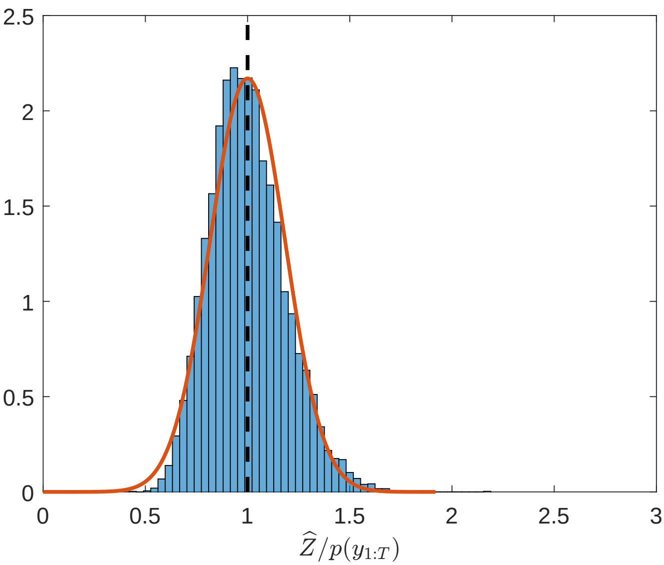
Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 100$ particles.



Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 500$ particles.



Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 1000$ particles.



Log-likelihood estimator

Alternatively, express the CLT in terms of $\log \widehat{Z}$.

Bias:

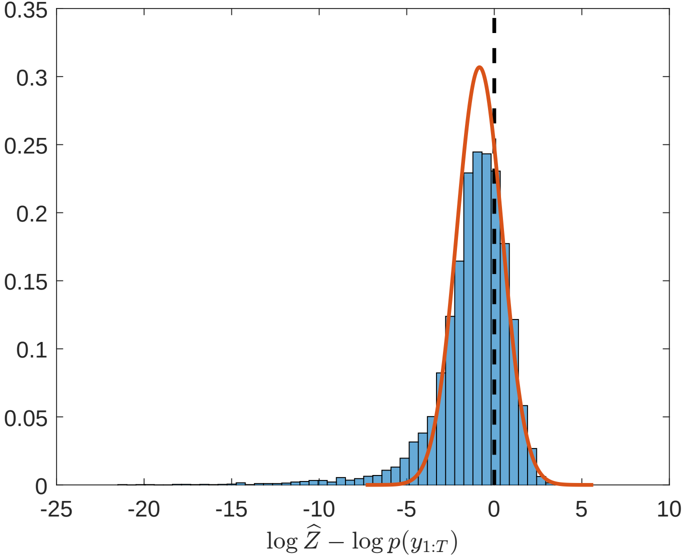
$$\mathbb{E}_{\psi_{N,T}}\left[\log \widehat{Z} - \log\{p(y_{1:T})\}\right] \approx -\frac{1}{2N}\sum_{t=0}^{T}\left\{\int \frac{p(x_t \mid y_{1:T})^2}{p(x_t \mid y_{1:t-1})}\mathrm{d}x_t - 1\right\}$$

Variance:

$$\mathrm{Var}_{\psi_{N,T}}\left[\log \widehat{Z}\right] \approx \frac{1}{N}\sum_{t=0}^{T}\left\{\int \frac{p(x_t \mid y_{1:T})^2}{p(x_t \mid y_{1:t-1})}\mathrm{d}x_t - 1\right\}$$

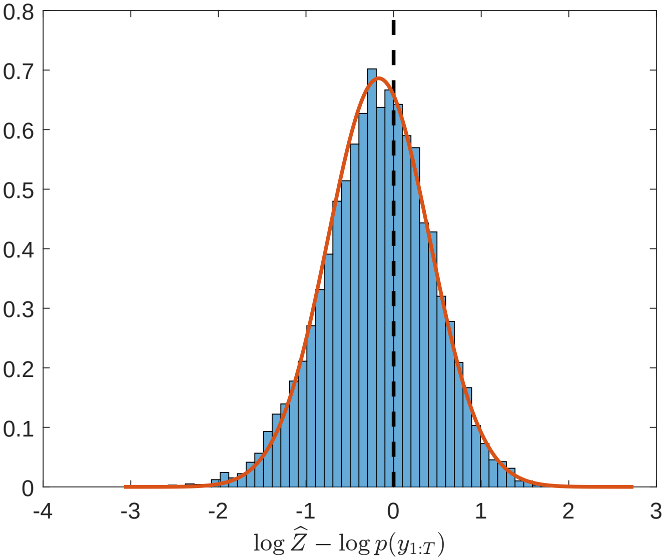
Note that the asymptotic variance is the same for $\widehat{Z}/p(y_{1:T})$ and $\log \widehat{Z}$.

Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 20$ particles.



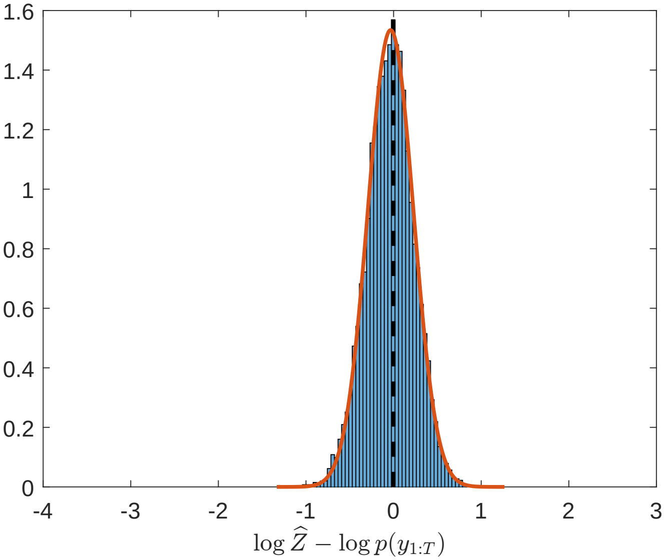
ex) Numerical illustration, cont'd

Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 100$ particles.



ex) Numerical illustration, cont'd

Histogram based on 10 000 independent realizations of $\widehat{Z} \sim \psi_{N,T}(z)$ using $N = 500$ particles.



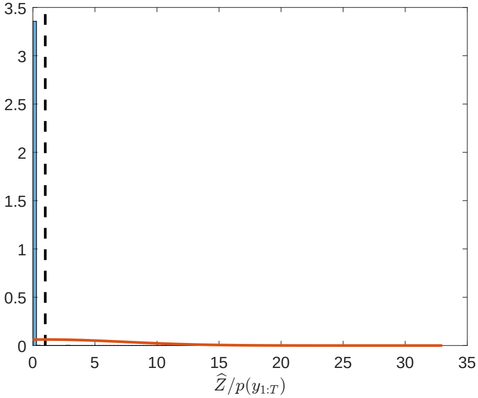
ex) Numerical illustration, cont'd

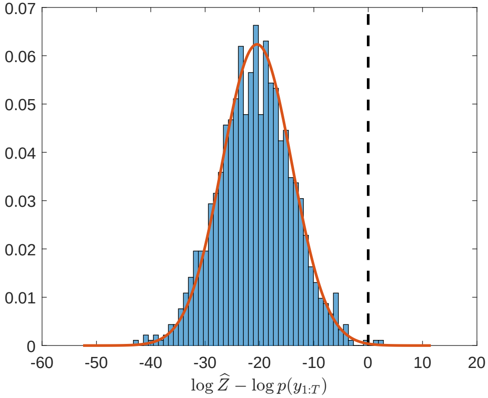
What happens if we increase T but keep N fixed?

UPPSALA UNIVERSITET

What happens if we increase T but keep N fixed?

Using $N = 100$ and $T = 1000$ (before: $T = 20$).





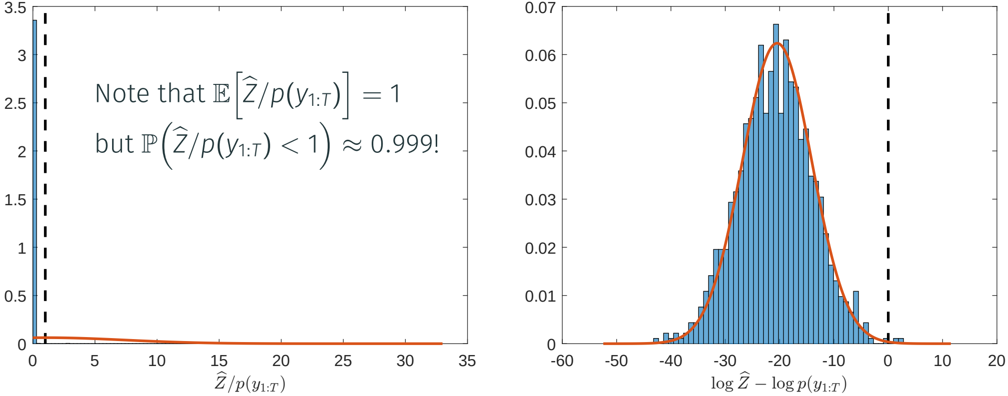
UPPSALA UNIVERSITET

What happens if we increase T but keep N fixed?

Using $N = 100$ and $T = 1000$ (before: $T = 20$).

Note that $\mathbb{E}\left[\widehat{Z}/p(y_{1:T})\right] = 1$

but $\mathbb{P}\left(\widehat{Z}/p(y_{1:T}) < 1\right) \approx 0.999!$



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

Particle filter sampling distribution: The joint distribution of all the random variables generated when running the particle filter.

Unbiasedness of the likelihood estimator: The expected value of the likelihood estimator, with respect to the randomness of the particle filter algorithm, is precisely the data likelihood. This property holds for any number of particles N.

Log-likelihood estimator: For numerical stability it is better to work with the logarithm of the likelihood estimator. The distribution of the log-likelihood estimator tends to converge more quickly to a Gaussian than that of the likelihood estimator.