

# Sequential Monte Carlo methods

## Lecture 10 – Properties of the likelihood estimator

---

Johan Alenlöv, Linköping University

2021-08-17

**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

## ex) Numerical illustration

Simple LG-SSM,

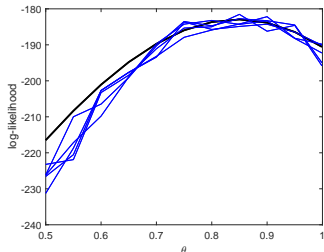
$$X_t = \theta X_{t-1} + V_t,$$

$$Y_t = X_t + E_t,$$

$$V_t \sim \mathcal{N}(0, 1),$$

$$E_t \sim \mathcal{N}(0, 1).$$

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



Black line – true likelihood computed using the Kalman filter.

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

## Bootstrap PF likelihood estimator

The particle filter likelihood estimator,

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

is a **random variable**.

## Bootstrap PF likelihood estimator

The particle filter likelihood estimator,

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

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  $p(y_{1:T} | \theta)$ .

## Bootstrap PF likelihood estimator

The particle filter likelihood estimator,

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

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  $p(y_{1:T} | \theta)$ .

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

## Bootstrap PF likelihood estimator

The particle filter likelihood estimator,

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

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  $p(y_{1:T} | \theta)$ .

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

**N.B.** From now on we consider the likelihood estimate for a fixed value of  $\theta$  and thus drop  $\theta$  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**
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  $t = 0, \dots, T$  samples the random variables

$$\begin{aligned} X_t &= \{X_t^i\}_{i=1}^N, & t &= 0, \dots, T, \\ A_t &= \{A_t^i\}_{i=1}^N, & t &= 1, \dots, T, \end{aligned}$$

with distributions (for the bootstrap PF):

$$X_0 \sim \prod_{i=1}^N p(x_0^i) \quad (\text{Initialization})$$

$$A_t | (X_{t-1} = x_{t-1}) \sim \prod_{i=1}^N w_{t-1}^{a_t^i} \quad (\text{Resampling})$$

$$X_t | (X_{t-1} = x_{t-1}, A_t = a_t) \sim \prod_{i=1}^N p(x_t^i | x_{t-1}^{a_t^i}) \quad (\text{Propagation})$$

# Particle filter sampling distribution

Let  $X_{0:T} = (X_0, \dots, X_T)$  and  $A_{1:T} = (A_1, \dots, A_T)$ .

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

$$\psi_{N,T}(x_{0:T}, a_{1:T}) = \left\{ \prod_{i=1}^N p(x_0^i) \right\} \prod_{t=1}^T \left\{ \prod_{i=1}^N w_{t-1}^{a_t^i} p(x_t^i | x_{t-1}^{a_t^i}) \right\},$$

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

## Particle filter sampling distribution

Let  $X_{0:T} = (X_0, \dots, X_T)$  and  $A_{1:T} = (A_1, \dots, A_T)$ .

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

$$\psi_{N,T}(x_{0:T}, a_{1:T}) = \left\{ \prod_{i=1}^N p(x_0^i) \right\} \prod_{t=1}^T \left\{ \prod_{i=1}^N w_{t-1}^{a_t^i} p(x_t^i | x_{t-1}^{a_t^i}) \right\},$$

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  $\hat{Z}$  is a function of the random variables  $X_{0:T}$  and  $A_{1:T}$ .

The distribution  $\psi_{N,T}(x_{0:T}, a_{1:T})$  induces a distribution for  $\hat{Z}$  which we also denote by  $\psi_{N,T}(z)$

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

## Distribution of the likelihood estimator

The likelihood estimator  $\hat{Z}$  is a function of the random variables  $X_{0:T}$  and  $A_{1:T}$ .

The distribution  $\psi_{N,T}(x_{0:T}, a_{1:T})$  induces a distribution for  $\hat{Z}$  which we also denote by  $\psi_{N,T}(z)$

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

### Theorem: Unbiasedness of the likelihood estimator

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

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

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.)



## ex) Numerical illustration

Simple LG-SSM,

$$X_t = 0.9X_{t-1} + V_t,$$

$$V_t \sim \mathcal{N}(0, 1),$$

$$Y_t = X_t + E_t,$$

$$E_t \sim \mathcal{N}(0, 1).$$

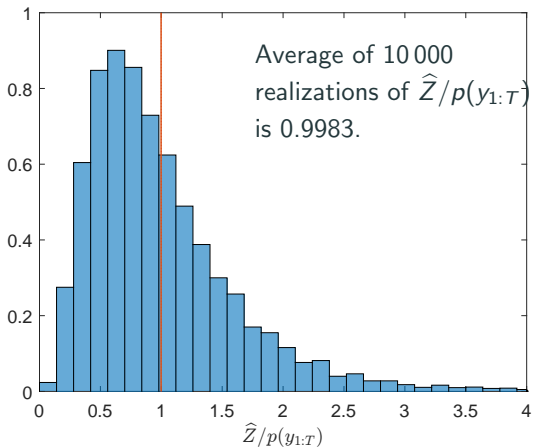
**Task:** estimate  $p(y_{1:T})$  for a **small** simulated data set consisting of  $T = 20$  measurements.

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



## ex) Numerical illustration

Histogram based on 10 000 independent realizations of  $\hat{Z} \sim \psi_{N,T}(z)$  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  $\hat{Z} \sim \psi_{N,T}(z)$ ,

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

as  $N \rightarrow \infty$ .

# Central limit theorem

## Theorem: CLT for likelihood estimator

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

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

as  $N \rightarrow \infty$ .

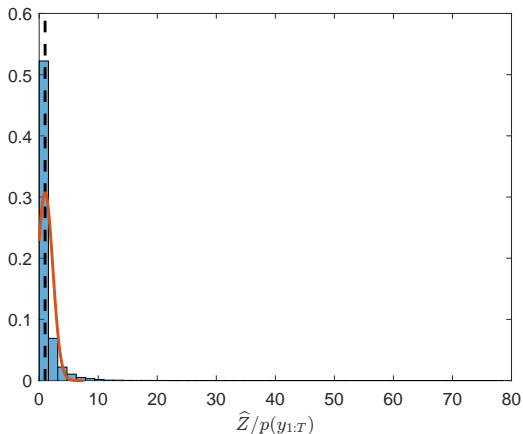
Under certain **exponential forgetting conditions** (recall lecture 5), one can show that the variance is

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

for some constant  $C < \infty$ .

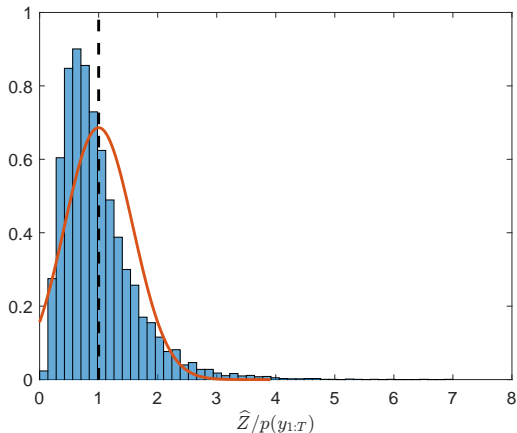
## ex) Numerical illustration, cont'd

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



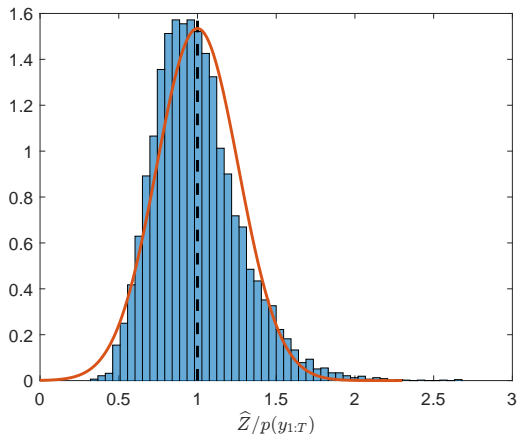
## ex) Numerical illustration, cont'd

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



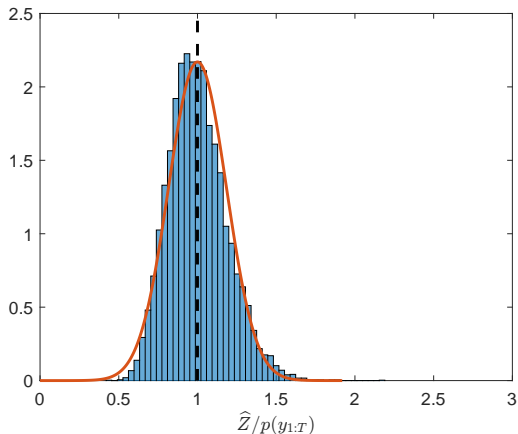
## ex) Numerical illustration, cont'd

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



## ex) Numerical illustration, cont'd

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





## Log-likelihood estimator

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

**Bias:**

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

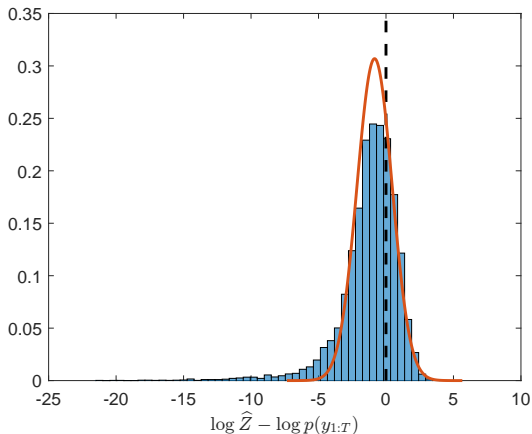
**Variance:**

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

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

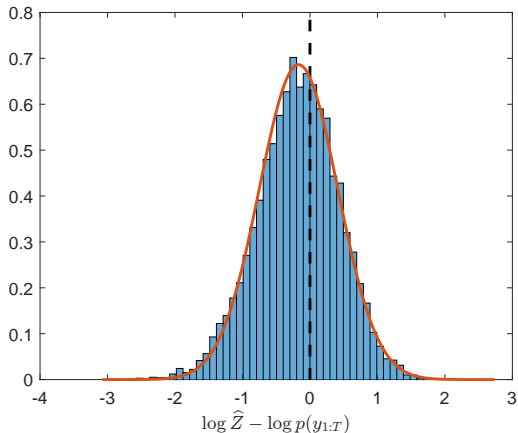
## ex) Numerical illustration, cont'd

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



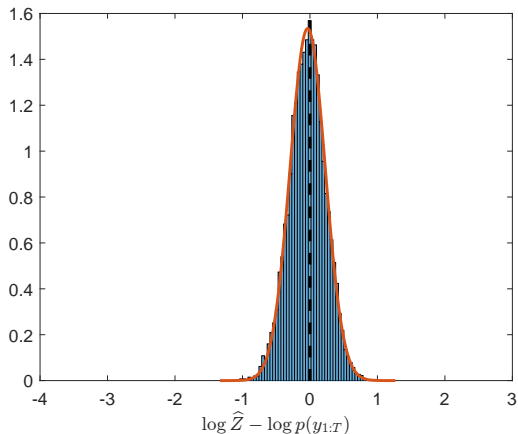
## ex) Numerical illustration, cont'd

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



## ex) Numerical illustration, cont'd

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



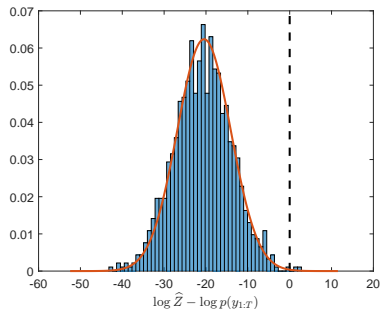
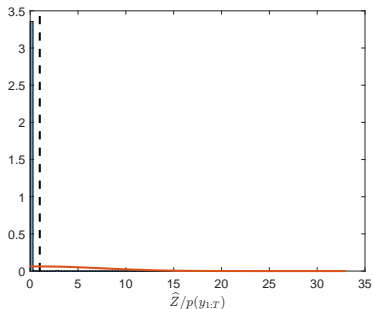
## ex) Numerical illustration, cont'd

What happens if we increase  $T$  but keep  $N$  fixed?

## ex) Numerical illustration, cont'd

What happens if we increase  $T$  but keep  $N$  fixed?

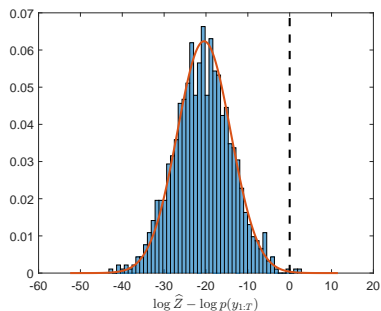
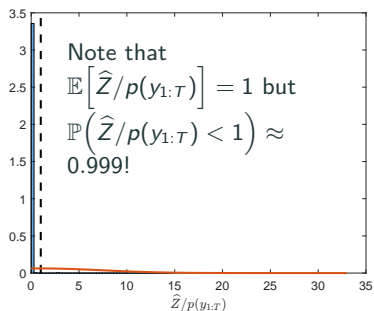
Using  $N = 100$  and  $T = 1000$  (before:  $T = 20$ ).



## ex) Numerical illustration, cont'd

What happens if we increase  $T$  but keep  $N$  fixed?

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



## 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



## A few concepts to summarize lecture 10

**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.