UPPSALA UNIVERSITET

Sequential Monte Carlo methods

Lecture 10 – Properties of the likelihood estimator

Fredrik Lindsten, Uppsala University

2017-08-25

Outline – Lecture 10

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

The particle filter sampling distribution

Bootstrap PF likelihood estimator

The likelihood estimator of the **bootstrap particle filter**,

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

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, $z[1], z[2], \ldots,$ all of which estimate $p(y_{1:T} \mid \theta)$.

What can be said about the distribution and properties of the random variable Z?

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

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

with distributions (for the bootstrap PF):

$$\mathbf{X}_0 \sim \prod_{i=1}^N p(x_0^i) \qquad \text{(Initialization)}$$
$$\mathbf{A}_t \,|\, (\mathbf{X}_{t-1} = \mathbf{x}_{t-1}) \sim \prod_{i=1}^N w_{t-1}^{a_t^i} \qquad \text{(Resampling)}$$
$$\mathbf{X}_t \,|\, (\mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \mathbf{A}_t = \mathbf{a}_t) \sim \prod_{i=1}^N p(x_t^i \,|\, x_{t-1}^{a_t^i}) \qquad \text{(Propagation)}$$

Particle filter sampling distribution

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

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

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

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

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 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 Z which we also denote by $\psi_{N,T}(z)$ (by abuse of notation),

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

Theorem: Unbiasedness of the likelihood estimator

The likelihood estimator Z is unbiased, i.e.

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

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,

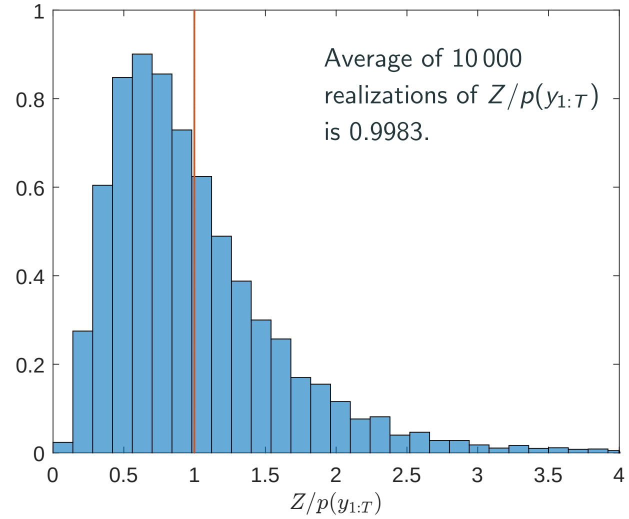
$$X_t = 0.9X_{t-1} + V_t, \qquad V_t \sim \mathcal{N}(0,1),$$
$$Y_t = X_t + E_t, \qquad E_t \sim \mathcal{N}(0,1).$$

Task: estimate $p(y_{1:T})$ 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 $Z \sim \psi_{N,T}(z)$ using $N = 100$ particles.



Central limit theorems

Central limit theorem

Theorem: CLT for likelihood estimator

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

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

as $N \to \infty$.

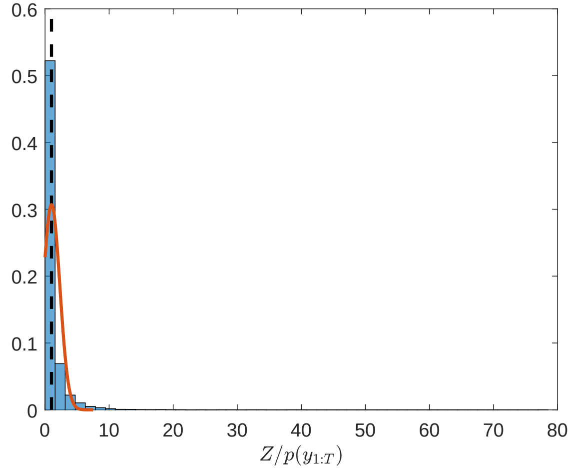
Under certain **exponential forgetting conditions** (recall lecture 5), one can show that the variance is

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

for some constant $C < \infty$.

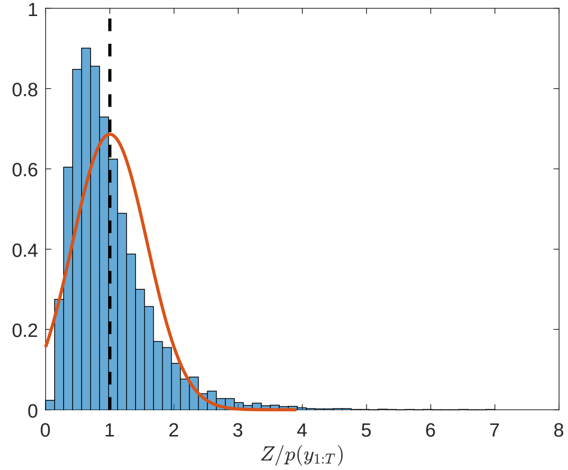
ex) Numerical illustration, cont'd

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



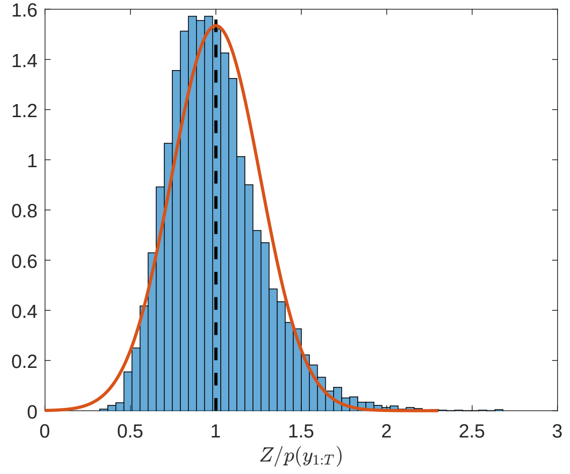
ex) Numerical illustration, cont'd

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



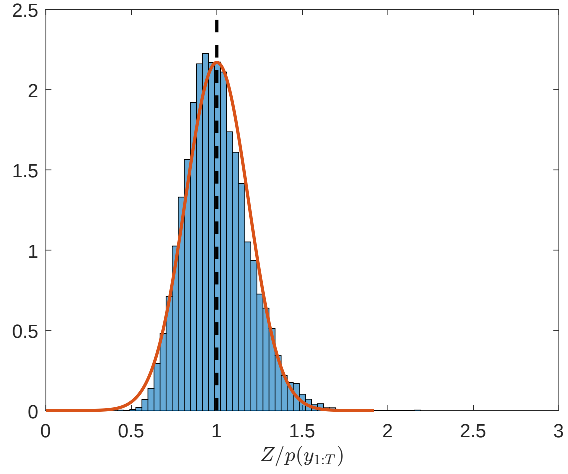
ex) Numerical illustration, cont'd

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



ex) Numerical illustration, cont'd

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



Log-likelihood estimator

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

Bias:

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

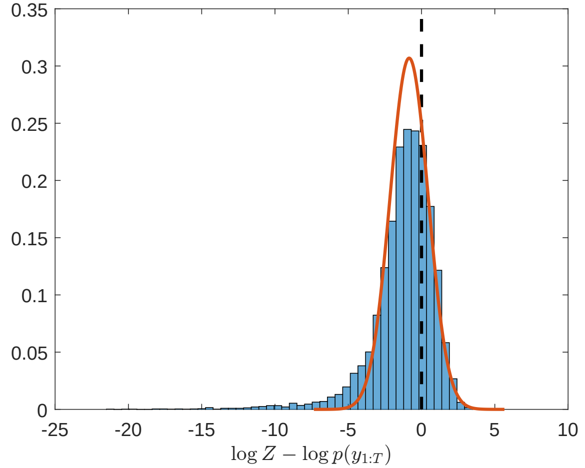
Variance:

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

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

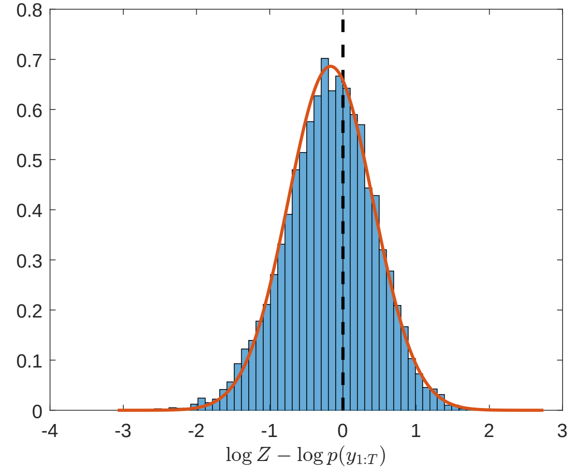
ex) Numerical illustration, cont'd

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



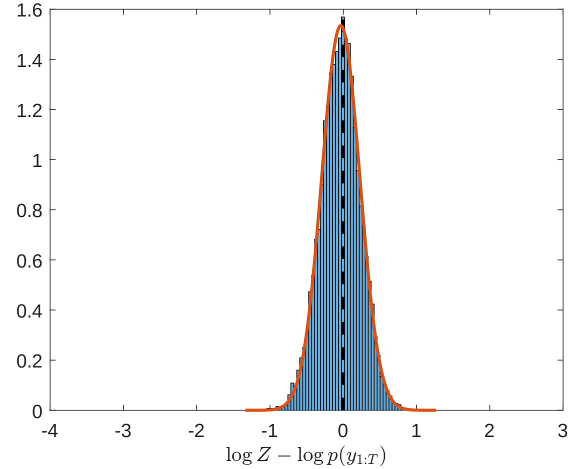
ex) Numerical illustration, cont'd

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



ex) Numerical illustration, cont'd

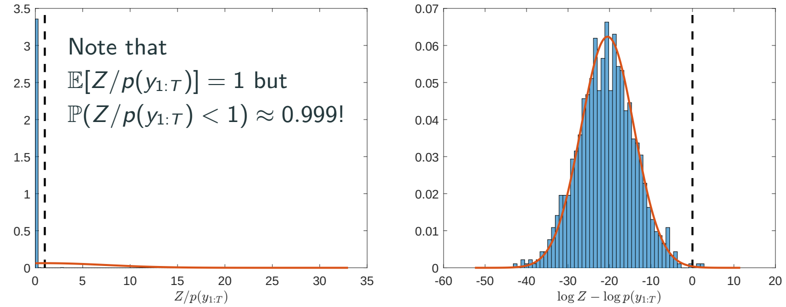
Histogram based on 10 000 independent realizations of $Z \sim \psi_{N,T}(z)$ using $N = 500$ particles.



ex) Numerical illustration, cont'd

What happens if we increase T but keep N fixed?

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



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.