Sequential Monte Carlo methods

Lecture 5 – Basic convergence theory

Johan Alenlöv, Uppsala University

2019-08-26

Aim: Provide some insight into the convergence and stability of the bootstrap particle filter.

Outline:

1. Central limit theorem for **importance sampling**
2. Central limit theorem for the **bootstrap particle filter**
3. **Stability** — key difference between the two

CLT for importance sampling

Importance sampling

Importance sampling,

Target: $\pi(x)$

Proposal: $q(x)$

Weight function: $\omega(x) = \frac{\pi(x)}{q(x)}$

Importance sampling

Importance sampling,

Target: $\pi(x)$

Proposal: $q(x)$

Weight function: $\omega(x) = \frac{\pi(x)}{q(x)}$

Procedure (for $i = 1, \ldots, N$)

1. Sample $x^i \sim q(x)$,
2. Compute $\widetilde{w}^i = \omega(x^i)$,
3. Normalize $w^i = \frac{\widetilde{w}^i}{\sum_{j=1}^{N} \widetilde{w}^i}$.

Importance sampling

Importance sampling,

Target: $\pi(x)$

Proposal: $q(x)$

Weight function: $\omega(x) = \frac{\pi(x)}{q(x)}$

Procedure (for $i = 1, \ldots, N$)

1. Sample $x^i \sim q(x)$,
2. Compute $\widetilde{w}^i = \omega(x^i)$,
3. Normalize $w^i = \frac{\widetilde{w}^i}{\sum_{j=1}^N \widetilde{w}^i}$.

N.B. Here, we define ω in terms of the normalized target – no difference algorithmically but simplifies analysis.

Importance sampling bias

From the black board we have,

$$\mathbb{E}\left[\widehat{I}_N^{\mathsf{IS}}(\varphi)\right] = I(\varphi) - \frac{\mathrm{Cov}_q[g(X), \omega(X)]}{N} + \frac{I(\varphi)\,\mathrm{Var}_q[\omega(X)]}{N} + O\left(\frac{1}{N^2}\right)$$

Importance sampling bias

From the black board we have,

$$\mathbb{E}\left[\widehat{I}_N^{\mathrm{IS}}(\varphi)\right] = I(\varphi) - \frac{\mathrm{Cov}_q[g(X), \omega(X)]}{N} + \frac{I(\varphi)\,\mathrm{Var}_q[\omega(X)]}{N} + O\left(\frac{1}{N^2}\right)$$

Thus, the **bias** in the importance sampling estimator, **for large N**, is

$$\begin{aligned} &\mathbb{E}\left[\widehat{I}_N^{\mathrm{IS}}(\varphi)\right] - I(\varphi) \\ &\qquad \approx -\frac{\mathrm{Cov}_q[g(X), \omega(X)]}{N} + \frac{I(\varphi)\,\mathrm{Var}_q[\omega(X)]}{N} \\ &\qquad\qquad = \cdots = -\frac{1}{N} \int \frac{\pi(x)^2}{q(x)} (\varphi(x) - I(\varphi)) \mathrm{d}x \end{aligned}$$

Importance sampling bias and variance

Importance sampling bias (large N):

$$\mathbb{E}\left[\widehat{I}_N^{\mathsf{IS}}(\varphi)\right] - I(\varphi) \approx -\frac{1}{N}\int \frac{\pi(x)^2}{q(x)}(\varphi(x) - I(\varphi))\mathrm{d}x$$

Importance sampling bias and variance

Importance sampling bias (large N):

$$\mathbb{E}\left[\widehat{I}_N^{\mathrm{IS}}(\varphi)\right] - I(\varphi) \approx -\frac{1}{N}\int \frac{\pi(x)^2}{q(x)}(\varphi(x) - I(\varphi))\mathrm{d}x$$

Importance sampling variance (large N):

$$\mathrm{Var}\left[\widehat{I}_N^{\mathrm{IS}}(\varphi)\right] \approx \frac{1}{N}\int \frac{\pi(x)^2}{q(x)}(\varphi(x) - I(\varphi))^2\mathrm{d}x$$

Importance sampling bias and variance

Importance sampling bias (large N):

$$\mathbb{E}\left[\widehat{I}_N^{\mathrm{IS}}(\varphi)\right] - I(\varphi) \approx -\frac{1}{N}\int \frac{\pi(x)^2}{q(x)}(\varphi(x) - I(\varphi))\mathrm{d}x$$

Importance sampling variance (large N):

$$\mathrm{Var}\left[\widehat{I}_N^{\mathrm{IS}}(\varphi)\right] \approx \frac{1}{N}\int \frac{\pi(x)^2}{q(x)}(\varphi(x) - I(\varphi))^2\mathrm{d}x$$

Mean-squared error $=$ bias^2 $+$ variance — Dominated by variance!

CLT for importance sampling

Asymptotically, as $N \to \infty$,

Central limit theorem (CLT) for importance sampler

$$\sqrt{N}\left(\sum_{i=1}^{N} W^i \varphi(X^i) - I(\varphi)\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \int \frac{\pi(x)^2}{q(x)} (\varphi(x) - I(\varphi))^2 \mathrm{d}x\right)$$

Importance sampling for filtering

Importance sampling for $\pi(x_{0:t}) = p(x_{0:t} \mid y_{1:t})$, where

$$\underbrace{p(x_{0:t} \mid y_{1:t})}_{\text{target}} = \frac{\overbrace{p(x_{0:t}, y_{1:t})}^{\text{unnormalized target}}}{\underbrace{p(y_{1:t})}_{\text{normalization}}} \propto p(y_{1:t} \mid x_{0:t}) p(x_{0:t})$$

Importance sampling for filtering

Importance sampling for $\pi(x_{0:t}) = p(x_{0:t} \mid y_{1:t})$, where

$$\underbrace{p(x_{0:t} \mid y_{1:t})}_{\text{target}} = \frac{\overbrace{p(x_{0:t}, y_{1:t})}^{\text{unnormalized target}}}{\underbrace{p(y_{1:t})}_{\text{normalization}}} \propto p(y_{1:t} \mid x_{0:t})p(x_{0:t})$$

Procedure: (for $i = 1, \ldots, N$)

1. Generate $x_{0:t}^i \sim p(x_{0:t})$ by simulating the system dynamics
2. Compute weights $\widetilde{w}_t^i = p(y_{1:t} \mid x_{0:t}^i)$ and normalize $\Rightarrow w_t^i$

ex) Importance sampling for filtering

ex) Very simple state space model where the states are independent over time (no dynamics),

$$X_t \sim \mathcal{N}(0, 1), \quad t = 0, 1, \ldots,$$
$$Y_t \,|\, (X_t = x_t) \sim \mathcal{N}(x_t, \sigma^2), \quad t = 1, 2, \ldots$$

ex) Importance sampling for filtering

ex) Very simple state space model where the states are independent over time (no dynamics),

$$X_t \sim \mathcal{N}(0,1), \quad t = 0,1,\dots,$$
$$Y_t \mid (X_t = x_t) \sim \mathcal{N}(x_t, \sigma^2), \quad t = 1,2,\dots$$

Asymptotic variance of importance sampler at time t is,

$$\left\{ \prod_{k=0}^{t-1} \int \frac{p(x_k \mid y_k)^2}{p(x_k)} \mathrm{d}x_k \right\} \int \frac{p(x_t \mid y_t)^2}{p(x_t)} (\varphi(x_t) - I_t(\varphi))^2 \mathrm{d}x_t$$

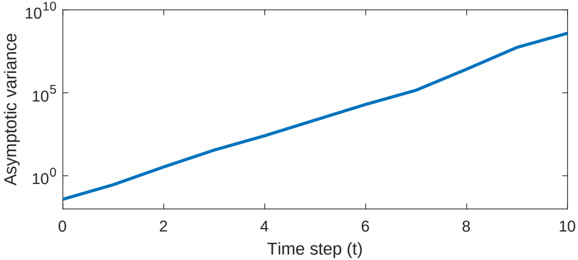
ex) Importance sampling for filtering

ex) Very simple state space model where the states are independent over time (no dynamics),

$$X_t \sim \mathcal{N}(0,1), \quad t = 0, 1, \ldots,$$
$$Y_t \mid (X_t = x_t) \sim \mathcal{N}(x_t, \sigma^2), \quad t = 1, 2, \ldots$$

Asymptotic variance of importance sampler at time t is,

$$\left\{ \prod_{k=0}^{t-1} \int \frac{p(x_k \mid y_k)^2}{p(x_k)} \mathrm{d}x_k \right\} \int \frac{p(x_t \mid y_t)^2}{p(x_t)} (\varphi(x_t) - I_t(\varphi))^2 \mathrm{d}x_t$$



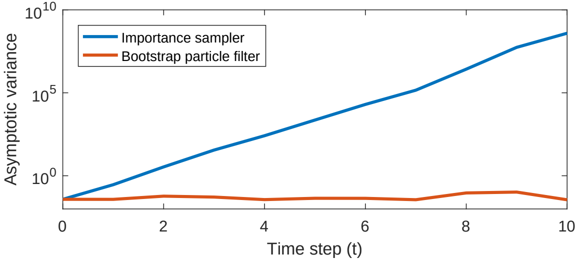
ex) Importance sampling for filtering

ex) Very simple state space model where the states are independent over time (no dynamics),

$$X_t \sim \mathcal{N}(0,1), \quad t = 0,1,\ldots,$$
$$Y_t \mid (X_t = x_t) \sim \mathcal{N}(x_t, \sigma^2), \quad t = 1,2,\ldots$$

Asymptotic variance of importance sampler at time t is,

$$\left\{ \prod_{k=0}^{t-1} \int \frac{p(x_k \mid y_k)^2}{p(x_k)} \mathrm{d}x_k \right\} \int \frac{p(x_t \mid y_t)^2}{p(x_t)} (\varphi(x_t) - I_t(\varphi))^2 \mathrm{d}x_t$$



CLT for bootstrap particle filter

CLT for bootstrap particle filter

Test function: $I_t(\varphi) = \mathbb{E}[\varphi(X_t) \mid y_{1:t}]$.

Theorem: CLT for bootstrap particle filter

$$\sqrt{N}\left(\sum_{i=1}^{N} W_t^i \varphi(X_t^i) - I_t(\varphi)\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_t(\varphi))$$

with

$$V_t(\varphi) = \sum_{k=0}^{t} \int \frac{p(x_k \mid y_{1:t})^2}{p(x_k \mid y_{1:k-1})} \left(I_{k,t}(\varphi \mid x_k) - I_t(\varphi)\right)^2 \mathrm{d}x_k$$

and

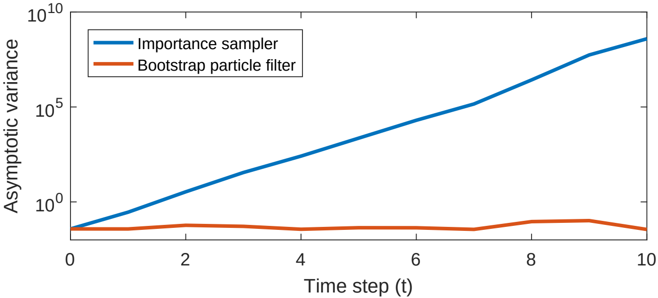
$$I_{k,t}(\varphi \mid x_k) = \mathbb{E}[\varphi(X_t) \mid y_{k+1:t}, x_k] \overset{k<t}{=} \int \varphi(x_t) p(x_t \mid x_k, y_{k+1:t}) \mathrm{d}x_t.$$

ex) Very simple model, cont'd

Simple model with $X_t \sim \mathcal{N}(0,1)$, independent over time.

$$I_{k,t}(\varphi \mid x_k) = \mathbb{E}[\varphi(X_t) \mid y_{k+1:t}, x_k] = \begin{cases} \mathbb{E}[\varphi(X_t) \mid y_t] & k < t, \\ \varphi(x_t) & k = t, \end{cases}$$

It follows that all terms $k < t$ in the definition of $V_t(\varphi)$ are zero!



Particle filter stability

Often the distant past has little effect on the future (and vice versa)
— referred to as **forgetting**

Particle filter stability

Often the distant past has little effect on the future (and vice versa) — referred to as **forgetting**

Exponential forgetting of **exact filter**:

$$\frac{1}{2}\int |p(x_t \mid x_k, y_{k+1:t}) - p(x_t \mid x'_k, y_{k+1:t})| \mathrm{d}x_t \leq \rho^{t-k}$$

Furthermore, it often holds that,

$$\frac{p(x_k \mid y_{1:t})^2}{p(x_k \mid y_{1:k-1})} \approx \frac{p(x_k \mid y_{1:k+\Delta})^2}{p(x_k \mid y_{1:k-1})}$$

Particle filter stability

Often the distant past has little effect on the future (and vice versa) — referred to as **forgetting**

Exponential forgetting of **exact filter**:

$$\frac{1}{2}\int |p(x_t \mid x_k, y_{k+1:t}) - p(x_t \mid x'_k, y_{k+1:t})| \mathrm{d}x_t \leq \rho^{t-k}$$

Furthermore, it often holds that,

$$\frac{p(x_k \mid y_{1:t})^2}{p(x_k \mid y_{1:k-1})} \leq A$$

Particle filter stability

Often the distant past has little effect on the future (and vice versa) — referred to as **forgetting**

Exponential forgetting of **exact filter**:

$$\frac{1}{2}\int |p(x_t \mid x_k, y_{k+1:t}) - p(x_t \mid x_k', y_{k+1:t})| \mathrm{d}x_t \leq \rho^{t-k}$$

Furthermore, it often holds that,

$$\frac{p(x_k \mid y_{1:t})^2}{p(x_k \mid y_{1:k-1})} \leq A$$

Thus, for bounded $|\varphi| < B$, it holds that $V_t(\varphi) \leq C$, independent of t!

Particle filter stability

Often the distant past has little effect on the future (and vice versa) — referred to as **forgetting**

Exponential forgetting of **exact filter**:

$$\frac{1}{2}\int |p(x_t \mid x_k, y_{k+1:t}) - p(x_t \mid x_k', y_{k+1:t})| \mathrm{d}x_t \leq \rho^{t-k}$$

Furthermore, it often holds that,

$$\frac{p(x_k \mid y_{1:t})^2}{p(x_k \mid y_{1:k-1})} \leq A$$

Thus, for bounded $|\varphi| < B$, it holds that $V_t(\varphi) \leq C$, independent of t!

The bootstrap particle filter is **stable**, in the sense that the estimator variance does not increase (unboundedly) with t.

Proof sketch

Three steps of the approximation

$\sum_{i=1}^{N} w_{t-1}^{i} \varphi(x_{t-1}^{i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Resampling: $a_t^i \sim \text{Discrete}(\{w_{t-1}^j\}_{j=1}^N)$

Propagation: $x_t^i \sim p(x_t \mid x_{t-1}^{a_t^i})$

Weighting: $\widetilde{w}_t^i = p(y_t \mid x_t^i)$ and normalize $\Rightarrow w_t^i$

Three steps of the approximation

$\sum_{i=1}^{N} w_{t-1}^{i} \varphi(x_{t-1}^{i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Resampling: $\frac{1}{N} \sum_{i=1}^{N} \varphi(x_{t-1}^{a_t^i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Propagation: $x_t^i \sim p(x_t \mid x_{t-1}^{a_t^i})$

Weighting: $\widetilde{w}_t^i = p(y_t \mid x_t^i)$ and normalize $\Rightarrow w_t^i$

Three steps of the approximation

$\sum_{i=1}^{N} w_{t-1}^{i} \varphi(x_{t-1}^{i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Resampling: $\frac{1}{N} \sum_{i=1}^{N} \varphi(x_{t-1}^{a_t^i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Propagation: $\frac{1}{N} \sum_{i=1}^{N} \varphi(x_{t}^{i})$ approximates $\mathbb{E}[\varphi(X_{t}) \mid y_{1:t-1}]$

Weighting: $\widetilde{w}_t^i = p(y_t \mid x_t^i)$ and normalize $\Rightarrow w_t^i$

Three steps of the approximation

$\sum_{i=1}^{N} w_{t-1}^{i} \varphi(x_{t-1}^{i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Resampling: $\frac{1}{N} \sum_{i=1}^{N} \varphi(x_{t-1}^{a_t^i})$ approximates $\mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]$

Propagation: $\frac{1}{N} \sum_{i=1}^{N} \varphi(x_{t}^{i})$ approximates $\mathbb{E}[\varphi(X_{t}) \mid y_{1:t-1}]$

Weighting: $\sum_{i=1}^{N} w_{t}^{i} \varphi(x_{t}^{i})$ approximates $\mathbb{E}[\varphi(X_{t}) \mid y_{1:t}]$

Inductive proof idea (I/II)

Inductive hypothesis:

$$\sqrt{N}\left(\sum_{i=1}^{N} W_{t-1}^{i}\varphi(X_{t-1}^{i}) - \mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_{t-1}(\varphi))$$

Resampling:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} \varphi(X_{t-1}^{A_t^i}) - \mathbb{E}[\varphi(X_{t-1}) \mid y_{1:t-1}]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \widetilde{V}_{t-1}(\varphi))$$

with $\widetilde{V}_{t-1}(\varphi) = V_{t-1}(\varphi) + \mathrm{Var}[\varphi(X_{t-1}) \mid y_{1:t-1}]$ follows from a conditional CLT.

Inductive proof idea (II/II)

Propagation:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^i) - \mathbb{E}[\varphi(X_t)\,|\,y_{1:t-1}]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \bar{V}_t(\varphi))$$

with $\bar{V}_t(\varphi) \;=\; \widetilde{V}_{t-1}(\mathbb{E}[\varphi(X_t)\,|\,x_{t-1}]) + \mathbb{E}[\mathrm{Var}[\varphi(X_t)\,|\,X_{t-1}]\,|\,y_{1:t-1}]$,
again, follows from a conditional CLT.

Inductive proof idea (II/II)

Propagation:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^i) - \mathbb{E}[\varphi(X_t) \mid y_{1:t-1}]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \bar{V}_t(\varphi))$$

with $\bar{V}_t(\varphi) = \widetilde{V}_{t-1}(\mathbb{E}[\varphi(X_t) \mid x_{t-1}]) + \mathbb{E}[\mathrm{Var}[\varphi(X_t) \mid X_{t-1}] \mid y_{1:t-1}]$, again, follows from a conditional CLT.

Weighting:

$$\sqrt{N}\left(\sum_{i=1}^{N} W_t^i\varphi(X_t^i) - \mathbb{E}[\varphi(X_t) \mid y_{1:t}]\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_t(\varphi))$$

with $V_t(\varphi) = \bar{V}_t\left(\frac{p(y_t \mid x_t)}{p(y_t \mid y_{1:t-1})} \cdot \{\varphi(x_t) - \mathbb{E}[\varphi(X_t) \mid y_{1:t}]\}\right)$ follows from the delta method.

References

A non-exhaustive list of references:

Arnaud Doucet and Adam M. Johansen. **A Tutorial on Particle Filtering and Smoothing: Fifteen years Later.** *The Oxford Handbook of Nonlinear Filtering, Oxford University Press*, 656–704, 2011.

Pierre Del Moral. **Feynman-Kac Formulae - Genealogical and Interacting Particle Systems with Applications.** *Springer*, 2004.

Nicolas Chopin. **Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference.** *The Annals of Statistics*, 32:2385–2411, 2004.

Nick Whiteley. **Stability properties of some particle filters.** *Annals of Applied Probability*, 23(6):2500–2537, 2013.

A few concepts to summarize lecture 5

Bias and variance: both of order $\frac{1}{N}$ — mean squared error dominated by variance! (Holds for both importance sampling and particle filter.)

Exponential forgetting: A property of the dynamical model — the influence of historical states on the future diminishes exponentially fast.

Particle filter stability: Under forgetting conditions, errors *do not* accumulate with time.

Next up

Practicals: From 15:15 — 17:00 in Room VIII and Room XI