L_2-Convergence of Bootstrap Particle Filter

Ömer Deniz Akyildiz

Imperial College London.

May 6, 2024

Abstract

These notes are prepared for Summer School on Bayesian Filtering (SSBF) 2024, complementary to the slides, and contain the proofs.

As explained in the lecture, we will prove L_2 convergence bounds for perfect Monte Carlo, self-normalised importance sampling, and bootstrap particle filters. The proofs are known in the literature, but we extracted them from Akyildiz (2019).

1 Perfect Monte Carlo

Theorem 1 (Perfect Monte Carlo). *Let φ be a bounded function, i.e.*

$$\|\varphi\|_\infty = \sup_{x \in \mathsf{X}} |\varphi(x)| < \infty.$$

Then, for any $N \geq 1$,

$$\|(\varphi, \pi) - (\varphi, \pi^N)\|_2 \leq \frac{2\|\varphi\|_\infty}{\sqrt{N}}.$$

Proof. We rewrite the L_2 norm using its definition as,

$$\begin{aligned}\left\|(\varphi, \pi) - (\varphi, \pi^N)\right\|_2 &= \left\|(\varphi, \pi) - \frac{1}{N}\sum_{k=1}^N \varphi\left(x^{(k)}\right)\right\|_2 \\ &= \mathbb{E}\left[\left|(\varphi, \pi) - \frac{1}{N}\sum_{k=1}^N \varphi\left(x^{(k)}\right)\right|^2\right]^{1/2}.\end{aligned}$$

Writing explicitly, we have,

$$\mathbb{E}\left[\left|(\varphi, \pi) - \frac{1}{N}\sum_{k=1}^N \varphi\left(x^{(k)}\right)\right|^2\right] = \frac{1}{N^2}\mathbb{E}\left[\left|\sum_{i=1}^N \left(\varphi(x^{(i)}) - (\varphi, \pi)\right)\right|^2\right].$$

We define $S^{(i)} = \varphi(x^{(i)}) - (\varphi, \pi)$ and note that $\mathbb{E}[S^{(i)}] = 0$ and $S^{(i)}$ are independent random variables. We therefore have,

$$\mathbb{E}\left[\left|(\varphi, \pi) - \frac{1}{N}\sum_{k=1}^{N} \varphi\left(x^{(k)}\right)\right|^2\right] = \frac{1}{N^2}\mathbb{E}\left[\left|\sum_{i=1}^{N} S^{(i)}\right|^2\right],$$
$$= \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{E}\left[\left|S^{(i)}\right|^2\right] \le \frac{N4\|\varphi\|_\infty^2}{N^2},$$

since $\left|S^{(i)}\right| = \left|\varphi(x^{(i)}) - (\varphi, \pi)\right| \le 2\|\varphi\|_\infty$. Therefore, we have,

$$\left\|(\varphi, \pi) - (\varphi, \pi^N)\right\|_2 \le \frac{2\|\varphi\|_\infty}{\sqrt{N}},$$

□

2 Self-Normalised Importance Sampling

Theorem 2 (Self-Normalised Importance Sampling). *Assume $\|W\|_\infty < \infty$ and $\|\varphi\|_\infty < \infty$. Then, the L_2 error (i.e., set $p = 2$) is bounded by*

$$\|(\varphi, \pi) - (\varphi, \tilde{\pi}^N)\|_2 \le \frac{c_2\|\varphi\|_\infty}{\sqrt{N}}$$

where

$$c_2 = \frac{2\|W\|_\infty}{(W, q)}.$$

Proof. First note that

$$(\varphi, \pi) = \frac{(\varphi, \gamma)}{\int \gamma(x)\mathrm{d}x} = \frac{(\varphi W, q)}{(W, q)}.$$

Then note the following inequalities,

$$|(\varphi, \pi) - (\varphi, \tilde{\pi}^N)| = \left|\frac{(\varphi W, q)}{(W, q)} - \frac{(\varphi W, q^N)}{(W, q^N)}\right|$$
$$\le \frac{\left|(\varphi W, q) - (\varphi W, q^N)\right|}{|(W, q)|} + |(\varphi W, q^N)|\left|\frac{1}{(W, q)} - \frac{1}{(W, q^N)}\right|$$
$$= \frac{\left|(\varphi W, q) - (\varphi W, q^N)\right|}{|(W, q)|} + \|\varphi\|_\infty|(W, q^N)|\left|\frac{(W, q^N) - (W, q)}{(W, q)(W, q^N)}\right|$$
$$= \frac{\left|(\varphi W, q) - (\varphi W, q^N)\right|}{(W, q)} + \frac{\|\varphi\|_\infty|(W, q^N) - (W, q)|}{(W, q)}.$$

We take squares of both sides and apply the inequality $(a + b)^2 \le 2(a^2 + b^2)$ to further bound the rhs,

$$|(\varphi, \pi) - (\varphi, \tilde{\pi}^N)|^2 \le 2\frac{\left|(\varphi W, q) - (\varphi W, q^N)\right|^2}{(W, q)^2} + 2\frac{\|\varphi\|_\infty^2|(W, q^N) - (W, q)|^2}{(W, q)^2}$$

We can now take the expectation of both sides,

$$\begin{aligned}\mathbb{E}\left[\left((\varphi,\pi)-(\varphi,\tilde{\pi}^N)\right)^2\right] \leq & \frac{2\mathbb{E}\left[\left((\varphi W,q)-(\varphi W,q^N)\right)^2\right]}{(W,q)^2} + \\ & \frac{2\|\varphi\|_\infty^2\mathbb{E}\left[\left((W,q^N)-(W,q)\right)^2\right]}{(W,q)^2}.\end{aligned}$$

Note that, both terms in the right hand side are perfect Monte Carlo estimates of the integrals, therefore,

$$\begin{aligned}\mathbb{E}\left[\left((\varphi,\pi)-(\varphi,\tilde{\pi}^N)\right)^2\right] \leq & \frac{2\|\varphi W\|_\infty^2}{(W,q)^2 N} + \frac{2\|\varphi\|_\infty^2\|W\|_\infty}{(W,q)^2 N}, \\ \leq & \frac{4\|\varphi\|_\infty^2\|W\|_\infty^2}{(W,q)^2 N},\end{aligned}$$

which completes the proof. □

3 Bootstrap Particle Filter

Theorem 3. *Assume that the likelihood function is positive and bounded*

$$g_t(x_t) > 0 \qquad \textit{and} \qquad \|g_t\|_\infty = \sup_{x_t \in \mathsf{X}} g_t(x_t) < \infty,$$

for all $t \geq 1$. Let φ be a bounded function and π_t^N be particle filter approximations of π_t. Then, for any $N \geq 1$,

$$\|(\varphi,\pi_t)-(\varphi,\pi_t^N)\|_2 \leq \frac{c_t\|\varphi\|_\infty}{\sqrt{N}}.$$

where $c_t < \infty$ is a constant independent of N.

Proof. This is an induction based proof. At time $t=0$, particle filter just samples from the prior of the model π_0 and by perfect Monte Carlo result, we readily have

$$\|(\varphi,\pi_0)-(\varphi,\pi_0^N)\|_2 \leq \frac{c_0\|\varphi\|_\infty}{\sqrt{N}}.$$

where $c_0 = 2$. Therefore, as an induction hypothesis, we assume

$$\|(\varphi,\pi_{t-1})-(\varphi,\pi_{t-1}^N)\|_2 \leq \frac{c_{t-1}\|\varphi\|_\infty}{\sqrt{N}}.$$

Particle filter takes three steps. We need to bound them separately.

1) Prediction/sampling step: Recall the predictive measure

$$\xi_t(\mathrm{d}x_t) = \int \tau(\mathrm{d}x_t|x_{t-1})\pi(\mathrm{d}x_{t-1}).$$

We need to next prove that the predictive approximation

$$\xi_t^N(\mathrm{d}x_t) = \frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_t^{(i)}}(\mathrm{d}x_t),$$

where $\bar{x}_t^{(i)} \sim \tau(\mathrm{d}x_t | x_{t-1}^{(i)})$ satisfies the L_2 bound

$$\|(\varphi, \xi^N) - (\varphi, \xi)\|_2 \leq \frac{c_{1,t}\|\varphi\|_\infty}{\sqrt{N}}.$$

Let us denote $\xi_t = \tau_t\pi_{t-1} := \int \tau_t(\mathrm{d}x_t|x_{t-1})\pi_{t-1}(\mathrm{d}x_{t-1})$. We can write

$$\begin{aligned}
\|(\varphi, \xi_t^N) - (\varphi, \xi_t)\|_2 &= \left\|(\varphi, \xi_t^N) - (\varphi, \tau_t\pi_{t-1})\right\|_2 \\
&\leq \left\|(\varphi, \xi_t^N) - (\varphi, \tau_t\pi_{t-1}^N)\right\|_2 \\
&\quad + \left\|(\varphi, \tau_t\pi_{t-1}^N) - (\varphi, \tau_t\pi_{t-1})\right\|_2,
\end{aligned}$$

where

$$(\varphi, \tau_t\pi_{t-1}^N) = \int \varphi(x_t)\tau_t(x_t|x_{t-1})\mathrm{d}x_t\pi_{t-1}^N(\mathrm{d}x_{t-1}) = \frac{1}{N}\sum_{i=1}^N \int \varphi(x_t)\tau(x_t|x_{t-1}^{(i)})\mathrm{d}x_t = \frac{1}{N}\sum_{i=1}^N (\varphi, \tau_t^{x_{t-1}^{(i)}}).$$

We have to now separately bound two terms. For the first term, we introduce the σ-algebra generated by the random variables $x_{0:t}^{(i)}$ and $\bar{x}_{1:t}^{(i)}$, $i = 1, \ldots, N$, denoted $\mathcal{F}_t = \sigma(x_{0:t}^{(i)}, \bar{x}_{1:t}^{(i)}, i = 1, \ldots, N)$. Since π_{t-1}^N is measurable w.r.t. $\mathcal{F}_{t-1}$, we can write

$$\mathbb{E}[(\varphi, \xi_t^N)|\mathcal{F}_{t-1}] = \frac{1}{N}\sum_{i=1}^N (\varphi, \tau_t^{x_{t-1}^{(i)}}) = (\varphi, \tau_t\pi_{t-1}^N).$$

Next, we define the random variables $S_t^{(i)} = \varphi(\bar{x}_t^{(i)}) - (\varphi, \tau_t\pi_{t-1}^N)$ and note that, conditional on $\mathcal{F}_{t-1}$, $S_t^{(i)}$, $i = 1, \ldots, N$ are zero-mean and independent. Then, the approximation error of ξ_t^N can be written as,

$$\mathbb{E}[\left|(\varphi, \xi_t^N\ ) - (\varphi, \tau_t\pi_{t-1}^N)\right|^2 |\mathcal{F}_{t-1}] = \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N S_t^{(i)}\right|^2 \middle| \mathcal{F}_{t-1}\right].$$

Using the fact that $S_t^{(i)}$ are conditionally zero-mean and independent, we can write,

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N S_t^{(i)}\right|^2 \middle| \mathcal{F}_{t-1}\right] = \frac{1}{N^2}\mathbb{E}\left[\sum_{i=1}^N \left|S_t^{(i)}\right|^2 \middle| \mathcal{F}_{t-1}\right],$$

Moreover, since $\left|S_t^{(i)}\right| = \left|\varphi(\bar{x}_t^{(i)}) - (\varphi, \tau_t\pi_{t-1}^N)\right| \leq 2\|\varphi\|_\infty$, we have,

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N S_t^{(i)}\right|^2 \middle| \mathcal{F}_{t-1}\right] \leq \frac{1}{N^2}N4\|\varphi\|_\infty^2 = \frac{4\|\varphi\|_\infty^2}{N}.$$

If we take unconditional expectations on both sides of the equation above, then we arrive at

$$\|(\varphi, \xi_t^N) - (\varphi, \tau_t\pi_{t-1}^N)\|_2 \leq \frac{\tilde{c}_1\|\varphi\|_\infty}{\sqrt{N}}, \tag{1}$$

where $\tilde{c}_1 = 2$ is a constant independent of N.

To handle the second term, we define $(\bar{\varphi}, \pi_{t-1}) = (\varphi, \tau_t \pi_{t-1})$ where $\bar{\varphi} \in B(\mathsf{X})$ and given by,

$$\bar{\varphi}(x) = (\varphi, \tau_t^x) = \int \varphi(x_t)\tau_t(x_t|x)\mathrm{d}x_t.$$

We also write $(\bar{\varphi}, \pi_{t-1}^N) = (\varphi, \tau_t \pi_{t-1}^N)$. Since $\|\bar{\varphi}\|_\infty \leq \|\varphi\|_\infty$, the induction hypothesis leads,

$$\begin{aligned} \|(\varphi, \tau_t \pi_{t-1}^N) - (\varphi, \tau_t \pi_{t-1})\|_2 &= \|(\bar{\varphi}, \pi_{t-1}^N) - (\bar{\varphi}, \pi_{t-1})\|_2 \\ &\leq \frac{c_{t-1}\|\varphi\|_\infty}{\sqrt{N}}, \end{aligned} \tag{2}$$

where c_{t-1} is a constant independent of N. Combining (1) and (2) yields,

$$\left\|(\varphi, \xi_t^N) - (\varphi, \xi_t)\right\|_2 \leq \frac{c_{1,t}\|\varphi\|_\infty}{\sqrt{N}} \tag{3}$$

where $c_{1,t} = c_{t-1} + 2 < \infty$ is a constant independent of N.

2) Weighting step: Next, we aim at bounding $\|(\varphi, \pi_t) - (\varphi, \tilde{\pi}_t^N)\|_2$ using (3). We have the weighted random measure,

$$\tilde{\pi}_t^N = \sum_{i=1}^N w_t^{(i)} \delta_{\bar{x}_t^{(i)}} \quad \text{where} \quad w_t^{(i)} = \frac{g_t(\bar{x}_t^{(i)})}{\sum_{i=1}^N g_t(\bar{x}_t^{(i)})}.$$

The integrals computed with respect to the weighted measure $\tilde{\pi}_t^N$ takes the form,

$$(\varphi, \tilde{\pi}_t^N) = \frac{(\varphi g_t, \xi^N)}{(g_t, \xi_t^N)}. \tag{4}$$

On the other hand, using Bayes theorem, integrals with respect to the optimal filter can also be written in a similar form as,

$$(\varphi, \pi_t) = \frac{(\varphi g_t, \xi_t)}{(g_t, \xi_t)}. \tag{5}$$

Using a similar argument as in the proof of importance sampling

$$\begin{aligned} \left|(\varphi, \tilde{\pi}_t^N) - (\varphi, \pi_t)\right| \leq &\frac{1}{(g_t, \xi_t)} \left(\|\varphi\|_\infty \left|(g_t, \xi_t) - (g_t, \xi_t^N)\right| \right. \\ &\left. + \left|(\varphi g_t, \xi_t) - (\varphi g_t, \xi_t^N)\right| \right), \end{aligned} \tag{6}$$

where $(g_t, \xi_t) > 0$ by assumption. Using Minkowski's inequality, we can deduce from (6) that

$$\begin{aligned} \left\|(\varphi, \tilde{\pi}_t^N) - (\varphi, \pi_t)\right\|_2 \leq &\frac{1}{(g_t, \xi_t)} \left(\|\varphi\|_\infty \left\|(g_t, \xi_t) - (g_t, \xi_t^N)\right\|_2 \right. \\ &\left. + \left\|(\varphi g_t, \xi_t) - (\varphi g_t, \xi_t^N)\right\|_2 \right). \end{aligned} \tag{7}$$

Noting that we have $\|\varphi g_t\|_\infty \leq \|\varphi\|_\infty \|g_t\|_\infty$, (3) and (7) together yield,

$$\left\|(\varphi, \pi_t) - (\varphi, \tilde{\pi}_t^N)\right\|_2 \leq \frac{c_{2,t}\|\varphi\|_\infty}{\sqrt{N}}, \tag{8}$$

where

$$c_{2,t,p} = \frac{2\|g_t\|_\infty c_{1,t}}{(g_t, \xi_t)} < \infty$$

is a finite constant independent of N.

3) Resampling step: Finally, since the random variables which are used to construct π_t^N are sampled i.i.d from $\tilde{\pi}_t^N$, the argument for the base case can also be applied here to yield,

$$\left\|(\varphi, \tilde{\pi}_t^N) - (\varphi, \pi_t^N)\right\|_2 \leq \frac{c_{3,t}\|\varphi\|_\infty}{\sqrt{N}}, \tag{9}$$

where $c_{3,t} < \infty$ is a constant independent of N. Combining bounds (8) and (9) to obtain the final result, with $c_t = c_{2,t} + c_{3,t} < \infty$, concludes the proof. □

References

Ömer Deniz Akyildiz. *Sequential and adaptive Bayesian computation for inference and optimization*. PhD thesis, Universidad Carlos III de Madrid, March 2019. URL http://akyildiz.me/works/thesis.pdf.