University of Alberta

A DISCRETE-TIME PARTICLE FILTER AND CENTRAL LIMIT THEOREM

by

Zi Ye

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Applied Mathematics

Department of Mathematical and Statistical Sciences

Spring 2014

Edmonton, Alberta

Abstract

The signal, we want to keep track of, is always modeled as a stochastic process. The filtering problem is that, due to some random noise, we sometimes can only access a distorted and corrupted partial observation of the signals. The objective of filtering is to find out the conditional distributions (expectations) of the signal process based on the history of the observation process, denoted as π_n. In the thesis, we always assume that the signal process X is a discrete-time stochastic process, and this type of problem is called the discrete-time filtering problem. Different from the previous solutions, we do not estimate π_n directly. Instead, we estimate the unnormalized filter σ_n, which is defined under a new fictitious probability measure Q. We will show the relation between π_n and σ_n as $\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)}$. If we can construct some particle systems, called particle filters, to estimate the unnormalized filter σ_n, it is enough to construct a particle system approximation to π_n.

In the weighted particle system, each particle is an independent copy from the signal process, and there is a weight associated to it. We prove the Strong Law of Large Numbers and the Central Limit Theorem for the weighted particle system. In addition, we calculate the random variance and the expected variance of the Central Limit Theorem.

The problem with the weighted particle system is that some particles do not behave like the signal process due to randomness. This problem manifests itself in a large random variance or expected variance of the Central Limit Theorem. To combat this problem, we will introduce another particle filter, that utilize resampling. Our key to analyzing this new particle filter mathematically is to simplify it to a fictitious particle system, which is mathematically simpler but can not be implemented on a computer. We prove the Strong Law of Large Numbers and the Central Limit Theorem for the fictitious particle system. In addition, we calculate the random variance and the expected variance of the Central Limit Theorem. An example is given where the expected variance of the weighted particle system is infinite while the fictitious system's is still finite, proving the need for resampling like that introduced within.

Acknowledgements

First of all, I would like to express my sincere gratitude to my master's supervisor, Professor Mike Kouritzin. He patiently provides me with his unreserved help and guidance, and leads me to finish the thesis step by step. He always inspires me to think in a different perspective, which is extremely helpful for my thesis. His encouragement and advice are of great significance for me to proceed through the master program as well as complete my thesis. He is a strong and supportive supervisor during my master study, but he always gives me great freedom to pursue independent study and research.

Moreover, I want to give gratitude to the rest of my committee members, Professor Yau Shu Wong, Professor Arno Berger and Professor Tahir Choulli, for their support, guidance and valuable suggestions.

Furthermore, I also would like to thank all of my friends. They give me continuous support and encouragement during my master study.

Last but not least, special thanks to my parents. They provide me with a carefree environment to grow up. In addition, their encouragement, support and love are always my driving forces and inspire me to pursue my goal.

Table of Contents

1 Introduction **1**

1.1 Nonlinear filtering 1

1.1.1 Background and motivation 1

1.1.2 Notations and definitions 2

1.1.3 Filtering problem 6

1.2 Particle filter method 10

1.2.1 Introduction 10

1.2.2 Resampled particle system 11

2 Unnormalized Filter **14**

2.1 Fictitious probability measure 14

2.2 Notations and unnormalized filter 18

3 Weighted Particle System **32**

3.1 Introduction of the particle system 32

3.2 Main results 36

3.2.1 Strong Law of Large Numbers 36

3.2.2 Central Limit Theorem 40

4 Fictitious Particle System **44**

4.1 Introduction of the particle system 44
4.2 Main results 51
4.2.1 Strong Law of Large Numbers 51
4.2.2 Central Limit Theorem 52

5 Summary and Future Work **63**

Bibliography **65**

Chapter 1

Introduction

1.1 Nonlinear filtering

1.1.1 Background and motivation

Filtering theory is an active research field with wide applications to real world problems in areas such as: signal processing, target detection and tracking, weather prediction and financial market.

To introduce it intuitively, we show an example to illustrate the objective of the filtering theory. This example comes from the wireless communication that was the main motivation for the filtering theory in the early stage. The *signal* (process) X_n has some randomness and is always a stochastic process. At time n, the signal is transmitted to a receiver and the receiver will receive the signal, defined as $h(X_{n-1})$, where h is a function. However, there are some random *noise* V_n during the transmission. Therefore, what we observe from the receiver is not only $h(X_{n-1})$ but also the noise term V_n. We define the observation model as

$$Y_n = h(X_{n-1}) + V_n, \tag{1.1}$$

where Y_n is called the *observation process*.

At time n, we will get the n^{th} observation Y_n from the receiver. The observation information $Y_1, Y_2, ..., Y_n$ at each observation time $1, 2, ..., n$ is known to us now. The objective of the filtering is to estimate the distribution of the signal process based on the previous observation information $Y_1, Y_2, ..., Y_n$.

In general, signals are always modeled as a stochastic processes and described by a stochastic dynamical system, which can not be solved directly and completely. Due to some random noise, we can only access a partial, distorted and corrupted observations of the signals. The goal of the filtering theory is to find the probabilistic distribution of the signals conditioning on the back observations.

1.1.2 Notations and definitions

In the thesis, we use $\mathbb{N}$ and $\mathbb{Z}$ to denote the set of natural numbers (including 0) and the set of integer numbers, respectively. Denote the product of d copies of the real numbers set $\mathbb{R}$ as $\mathbb{R}^d$, where $d \in \mathbb{N}$. The *Borel* $\sigma-$*algbra* defined on the set E is generated by all open sets in E, denoted by $\mathcal{B}(E)$. For example, $\mathcal{B}(\mathbb{R})$ is defined as the $\sigma-$algebra generated by all open sets in $\mathbb{R}$. The *Borel set* is any set in a topological space that can be formed from open sets. Let $E^P(\cdot)$ denote as the expectation with respect to the probability measure P.

A discrete-time *Markov process* $\{X_n, n \in \mathbb{N}\}$ with respect to the filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$ is defined as a sequence of random variables taking values in a measurable space $(E, \mathcal{E})$ satisfying that

1. X_n is $\mathcal{F}_n$-measurable for every $n \in \mathbb{N}$.

2. For all $B \in \mathcal{E}$, we have

$$P\left(X_{n+1} \in B \,|\mathcal{F}_n\right) = P\left(X_{n+1} \in B \,|X_n\right), \tag{1.2}$$

   for every $n \in \mathbb{N}$.

Sometimes, the discrete-time Markov process is called *Markov Chain*.

The *filtration* $\{\mathcal{F}_n, n \in \mathbb{N}\}$ is a sequence of σ-algebras satisfying that

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n \subset \cdots \tag{1.3}$$

As a Markov process, the conditional distribution of any future state X_{n+1} given the past states $X_0, X_1, \cdots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state.

A function $K : E \times \mathcal{E} \to \mathbb{R}$ is said to be a *transition kernel* if

1. For each $x \in E$, $A \to K(x, A)$ is a probability measure on the space $(E, \mathcal{E})$.

2. For each $A \in \mathcal{E}$, $x \to K(x, A)$ is a measurable function.

We say that $\{X_n, n \in \mathbb{N}\}$ is a Markov process with respect to the filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$ with the *Markov transition kernel* $\{K_n, n \in \mathbb{N}\}$ if

$$P\left(X_{n+1} \in A \,|\mathcal{F}_n\right) = K_{n+1}\left(X_n, A\right), \tag{1.4}$$

for a set $A \in \mathcal{E}$ and $n \in \mathbb{N}$.

A Markov Chain is said to be *time-homogeneous*, if the conditional probability $P(X_n = j\,|X_{n-1} = i) = P(X_1 = j\,|X_0 = i)$, for every $n > 1$ and any $i, j \in E$. That means, $P(X_n = j | X_{n-1} = i)$ is independent from the time index n and, the Markov transition kernel can be denoted as K at any time.

We denote $B(E)$ and $B(E)_+$ as the class of bounded measurable functions and non-negative bounded measurable functions defined on the space E respectively. Let $\overline{C}(E)$ and $\overline{C}(E)_+$ denote the class of continuous bounded functions and the class of non-negative continuous bounded functions respectively. Define the norm as $|f|_\infty = \sup_{x \in E} |f(x)|$. We also let $\mathcal{M}(E)$ and $\mathcal{P}(E)$ denote the space of finite measures and the space of probability measures defined on E, topologized by weak convergence. If $\eta \in \mathcal{M}(E)$ and f is an integrable function defined on E, $\eta(f)$ is defined as

$$\eta(f) = \int_E f(x)\, \eta(\mathrm{d}x).$$

Weak convergence means, for $\{\mu_n\}_{n=1}^{\infty}, \mu \in \mathcal{M}(E)$, $\mu_n \Rightarrow \mu$ if and only if $\mu_n(f) \rightarrow \mu(f)$ for all $f \in \overline{C}(E)$, when $n \rightarrow \infty$.

For a Markov transition kernel K and $\eta \in \mathcal{P}(E)$, $K\eta$ is defined as

$$K\eta(\mathrm{d}x) = \int_E K(z, \mathrm{d}x)\, \eta(\mathrm{d}z),$$

$K\eta$ is a probability measure defined on the space E, and

$$K^n \eta = K\left(K^{n-1}\eta\right).$$

Kf is defined as a function on the space E satisfying that

$$Kf(x) = \int_E f(z)\, K(x, \mathrm{d}z),$$

for any integrable function f and $x \in E$. It now follows that $Kf \in B(E)_+$ if $f \in B(E)_+$.

Then we have

$$(K\eta)(f) = \int_E f(x)(K\eta)(\mathrm{d}x) \quad (1.5)$$
$$= \int_E f(x) \int_E K(z, \mathrm{d}x)\,\eta(\mathrm{d}z)$$
$$= \int_E \int_E f(x) K(z, \mathrm{d}x)\,\eta(\mathrm{d}z),$$

and

$$\eta(K(f)) = \int_E \eta(\mathrm{d}z) K(f)(z) \quad (1.6)$$
$$= \int_E \eta(\mathrm{d}z) \int_E K(z, \mathrm{d}x) f(x)$$
$$= \int_E \int_E f(x) K(z, \mathrm{d}x)\,\eta(\mathrm{d}z).$$

We have shown that

$$(K\eta)(f) = \eta(K(f)), \quad (1.7)$$

for any integrable function f, Markov transition kernel K and $\eta \in \mathcal{P}(E)$.

A *Dirac measure* is a probability measure defined on some measurable space $(E, \mathcal{E})$ as

$$\delta_x(A) = 1_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases}, \quad (1.8)$$

for a given $x \in E$ and any measurable set $A \in \mathcal{E}$. 1_A is the indicator function of the set A.

A stochastic process $\{X_n, n \in \mathbb{N}\}$ is called a *martingale* with respect to the filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$ if for any $n \in \mathbb{N}$,

1. X_n is $\mathcal{F}_n$-measurable.
2. $E(|X_n|) < \infty$.
3. $E[X_n | \mathcal{F}_{n-1}] = X_{n-1}$.

1.1.3 Filtering problem

Filtering theory deals with estimating the current state of a non-observable signal X based upon the history of a distorted and corrupted partial observation process Y living on the same probability space $(\Omega, \mathcal{F}, P)$ as X. For many practical problems, the signal process is modeled as a time-homogeneous discrete-time Markov process $\{X_n, n \in \mathbb{N}\}$ with the initial distribution π_0 and the Markov transition kernel K. The signal process takes its values in some complete, separable metric space (E, ρ). By the definition of the Markov process, we can get

$$P(X_0 \in A) = \pi_0(A), \tag{1.9}$$

and

$$P\left(X_{n+1} \in A \middle| \mathcal{F}_n^X\right) = K(X_n, A), \tag{1.10}$$

for any set $A \in \mathcal{E}$ and $n \in \mathbb{N}$. $\mathcal{F}_n^X$ is the σ-algebra generated by $\{X_i, 0 \leq i \leq n\}$. We also have

$$\begin{aligned} E^P\left[f(X_n)\left|\mathcal{F}_{n-1}^X\right.\right] &= \int_E f(x)\, P\left(X_n \in \mathrm{d}x \left|\mathcal{F}_{n-1}^X\right.\right) \\ &= \int_E f(x)\, K(X_{n-1}, \mathrm{d}x) \\ &= Kf(X_{n-1}). \end{aligned} \tag{1.11}$$

Now, we give a concrete example for the signal process X. Suppose that the signal process $\{X_n, n \in \mathbb{N}\}$ is a simple random walk defined on the probability space $(\Omega, \mathcal{F}, P)$ and living on $\mathbb{Z}$. The initial state is $X_0 = 0$, which is the origin. At each time n, it moves either $+1$ or -1 with equal probability $\frac{1}{2}$. That means, X_{n+1} is either $X_n + 1$ or $X_n - 1$ both with probability $\frac{1}{2}$. We have

$$P(X_{n+1} = x + 1 \,|X_n = x) = \frac{1}{2} \text{ and } P(X_{n+1} = x - 1 \,|X_n = x) = \frac{1}{2}, \tag{1.12}$$

for every $x \in \mathbb{Z}$ and every $n \in \mathbb{N}$. Therefore, the random walk is a time-homogeneous Markov process. The initial distribution π_0 and the Markov transition kernel K are

$$\pi_0(\mathrm{d}x) = \delta_0(\mathrm{d}x), \tag{1.13}$$

and

$$\begin{aligned} K(X_{n-1}, \mathrm{d}x) &= P(X_n \in \mathrm{d}x \,|X_{n-1}) \\ &= \frac{1}{2}\delta_{(X_{n-1}+1)}(\mathrm{d}x) + \frac{1}{2}\delta_{(X_{n-1}-1)}(\mathrm{d}x), \end{aligned} \tag{1.14}$$

for any $n \geq 1$, respectively. $\mathrm{d}x$ is the infinitesimal neighborhood around the point x as

$$\mathrm{d}x = \left(\mathrm{d}x^1, \mathrm{d}x^2, \cdots, \mathrm{d}x^N\right). \tag{1.15}$$

The noise process $\{V_n, n \in \mathbb{N}\}$ is a sequence of independent and identically distributed random vectors with the common strictly positive probability density function g. V_n is defined on the same measurable space $(\Omega, \mathcal{F}, P)$ and takes its values in the set $\mathbb{R}^d$. The noise process $\{V_n, n \in \mathbb{N}\}$ is independent from the signal process $\{X_n, n \in \mathbb{N}\}$.

Notice that the probability density function g is always positive. For instance, the noise process $\{V_n, n \in \mathbb{N}\}$ can be a sequence of independent normally distributed random variables, where the probability density function g is defined as

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \sigma > 0, \tag{1.16}$$

for any $x \in (-\infty, \infty)$.

The noise process can be a sequence of independent random variables with the double exponential distribution. The probability density function g is defined as

$$g(x) = \begin{cases} \frac{1}{2}\exp(-x), & x \geq 0, \qquad (1.17) \\ \frac{1}{2}\exp(x), & x < 0. \qquad (1.17') \end{cases}$$

The noise process could also be a sequence of independent random variables with the standard Cauchy distribution. In this case, the probability density function g is defined as

$$g(x) = \frac{1}{\pi(1 + x^2)}, \tag{1.18}$$

for any $x \in (-\infty, \infty)$.

The sensor function h is a measurable mapping from E to $\mathbb{R}^d$. h could be a linear function, such as the polynomial function $h(x) = ax + b$, where a and b are constants. h could also be a nonlinear function, such as the exponential function $h(x) = e^x$. If h is a linear function, it is called the linear filtering problem. Otherwise, it is the nonlinear filtering problem.

Like the wireless communication example, the observation model for the filtering problem is defined as:

$$Y_n = h(X_{n-1}) + V_n, \tag{1.19}$$

for any $n \in \mathbb{N}$.

The observation process $\{Y_n, n \in \mathbb{N}\}$ is defined on the same measurable space $(\Omega, \mathcal{F}, P)$ and takes its values in the set $\mathbb{R}^d$.

The objective of the filtering problem is to compute the conditional probabilities

$$\pi_n(A) = P\left(X_n \in A \middle| \mathcal{F}_n^Y\right), \tag{1.20}$$

for all Borel sets A, or equivalently, the conditional expectations under the probability measure P

$$\pi_n(f) = E^P\left[f(X_n) \middle| \mathcal{F}_n^Y\right], \tag{1.21}$$

for any $f \in B(E)$ and $n \in \mathbb{N}$, where $\mathcal{F}_n^Y \doteq \sigma\{Y_k, k = 1, ..., n\}$ is the information obtained from the back observations.

While there are well-known mathematical formulae for π_n under many situations, these formulae are, with few exceptions, fundamentally infinitely dimensional and hence not implementable on a computer. Still, there are many ways to approximate these conditional distributions π_n in a computer workable manner.

1.2 Particle filter method

1.2.1 Introduction

To solve the filtering problem, we use the particle filter method invented in the 1960s. In the particle system, we create $N \in \mathbb{N}$ copies of the signal process $\{X_n, n \in \mathbb{N}\}$ and, each particle evolves independently of each other. Due to randomness, most particles may not behave like the signal. Therefore, historically, particle filters, also known as sequential Monte Carlo methods, were considered poor choices for most filtering problems until resampling techniques were invented that dramatically improved performance. Nowadays, resampled particle filters are relied upon in a wide variety of applications in such diverse areas as econometrics, target detecting and tracking.

For a resampled particle filter, at time 0, each particle is independently distributed as X_0. For any $n \geq 1$ and $n \in \mathbb{N}$, the particle filter is a two-step mechanism, resampling process and evolution process, from time $n-1$ to n. In the resampling process, each particle is relocated according to some mechanisms. Different particle filters have different resampling techniques. The resampling process will prevent some particles from deviating from the signal too much. In the evolution process, each particle evolves forward independently of each other according to the Markov transition kernel K of the signal process.

The original (resampled) interacting and branching particle filters have been intensely studied. However, Del Moral, Kouritzin and Miclo showed that the performance, and even the very success, of a particle filter depends heavily upon the type of resampling used and little theory is known about which resampling procedures should be used.

In the thesis, we will introduce and analyze the classical weighted particle filter without resampling and a new class of particle filter with resampling. In the new algorithm, the particles interact weakly through use of the total mass process in the resampling procedure as well as in the particle control step. The analysis of this algorithm is based upon a coupling to a fictitious particle system, corresponding to an idealized, unimplementable particle filter. The Strong Law of Large Numbers and the Central Limit Theorem are developed for this filter.

1.2.2 Resampled particle system

The new algorithm we will introduce and analyze in the thesis is explained in terms of a fixed number of particles $N \in \mathbb{N}$. We define the following branching Markov process $\left\{\mathbb{S}_n^N, n = 0, 1, ...\right\}$ as

Initialize: Particles $\left\{\mathbb{X}_0^i\right\}_{i=1}^N$ are independent and identically distributed random variables with the same distribution π_0. The weight of each particle is $\mathbb{L}_0 = 1$.

Repeat: For $n = 1, 2, ...$ do

1. Estimate: $\mathbb{S}_{n-1}^N = \frac{\mathbb{L}_{n-1}}{N} \sum_{i=1}^N \delta_{\mathbb{X}_{n-1}^i}$
2. Weight: $\mathbb{W}_n^i = \frac{g\left(Y_n - h\left(\mathbb{X}_{n-1}^i\right)\right)}{g(Y_n)}$ and $\hat{\mathbb{L}}_n^i = \mathbb{W}_n^i \mathbb{L}_{n-1}$ for $i = 1, ..., N$

3. Resampled Weight: $\mathbb{L}_n = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbb{L}}_n^i$

4. Offspring Numbers: $\left\{\mathbb{Z}_n^i\right\}_{i=1}^{N}$ are independent Bernoulli random variables with the probabilities $\left\{\frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} - \left\lfloor \frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} \right\rfloor\right\}_{i=1}^{N}$, respectively,

$$\mathbb{Z}_n^i = \begin{cases} 1, & \text{with probability } \frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} - \left\lfloor \frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} \right\rfloor, \\ 0, & \text{with probability } 1 - \frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} + \left\lfloor \frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} \right\rfloor. \end{cases} \tag{1.22}$$

   The offspring numbers is defined as $N_n^i = \left\lfloor \frac{\hat{\mathbb{L}}_n^i}{\mathbb{L}_n} \right\rfloor + \mathbb{Z}_n^i$ for $i = 1, ..., N$

5. Resample: $\hat{\mathbb{X}}_{n-1}^j = \mathbb{X}_{n-1}^i$ for $j \in \left\{N_n^1 + \cdots + N_n^{i-1} + 1, ..., N_n^1 + \cdots + N_n^i\right\}$ for $i = 1, ..., N$

6. Particle Control: Remove $\sum_{i=1}^{N} N_n^i - N$ randomly selected particles, or duplicate $N - \sum_{i=1}^{N} N_n^i$ randomly selected particles with replacement

7. Evolve Independently: $P\left(\mathbb{X}_n^i \in A_i \ \forall\, i \middle| \hat{\mathbb{X}}_{n-1}\right) = \prod_{i=1}^{N} K\left(\hat{\mathbb{X}}_{n-1}^i, A_i\right)$

The new resampled particle system is too complicated to analyze mathematically. To make it accessible, the analysis of the new algorithm is based upon a coupling to a fictitious particle system, that is mathematically simpler but unimplementable on a computer. In the thesis, we will show that the fictitious particle system satisfies the Strong Law of Large Numbers and the Central Limit Theorem. In addition, it can be a better filter approximation to σ_n, where σ_n is the unnormalized filter defined later, compared to the classical weighted particle system. In Chapter 2, we will state and prove the theoretical solution for the filtering problem. We will introduce the unnormalized filter σ_n and its relation with π_n. For comparison purpose, in Chapter 3, we will analyze the classical weighted particle system. The fictitious particle system is introduced and analyzed in Chapter 4. Conceptually, one can think of the above particle system as a weakly interacting one. In the fictitious particle system, the particle control is eliminated and $\mathbb{L}_n$ is replaced by $\sigma_n(1)$.

Chapter 2

Unnormalized Filter

2.1 Fictitious probability measure

The objective of the filtering is to find the conditional expectations with respect to the probability measure P

$$\pi_n(f) = E^P\left(f(X_n) \middle| \mathcal{F}_n^Y\right), \tag{2.1}$$

for any $f \in B(E)$. $\mathcal{F}_n^Y$ is the σ-algebra defined as

$$\mathcal{F}_n^Y \doteq \sigma\{Y_k, k = 1, 2, ..., n\}, \tag{2.2}$$

with the convention $\mathcal{F}_0^Y = \{\emptyset, \Omega\}$.

One of the best ways of constructing particle filters is to transfer all of the information obtained from the observations into a likelihood or weight function by the way of measure change.

In this reference probability method, a new fictitious probability measure Q is introduced under which the signal, observation process $\{(X_n, Y_{n+1}),\ n = 0, 1, 2, ...\}$ has the same distribution as the signal, noise process $\{(X_n, V_{n+1}),\ n = 0, 1, 2, ...\}$ does under P. In particular, this means that the observations become a sequence of independent and identically distributed random vectors with the common strictly positive density function g that are independent of the signal process X under the probability measure Q. In this case, all the observation information is absorbed into the weight or likelihood process $\{L_n,\ n = 1, 2, ...\}$ transforming Q back to P, which in our case has the form

$$\frac{\mathrm{d}P}{\mathrm{d}Q}\Big|_{\mathcal{F}_\infty^X \vee \mathcal{F}_n^Y} = L_n = \prod_{j=1}^{n} W_j, \tag{2.3}$$

where

$$\begin{aligned} W_j &= \alpha_j \left(X_{j-1}\right), \\ \alpha_j (x) &= \frac{g\left(Y_j - h(x)\right)}{g\left(Y_j\right)}, \end{aligned} \tag{2.4}$$

and therefore

$$L_n = \prod_{j=1}^{n} \frac{g\left(Y_j - h\left(X_{j-1}\right)\right)}{g\left(Y_j\right)}, \tag{2.5}$$

with the convention $L_0 = 1$. $\mathcal{F}_n^X$ is the σ-algebra defined as

$$\mathcal{F}_n^X \doteq \sigma\{X_k, k = 0, 1, ..., n\}, \tag{2.6}$$

with the convention $\mathcal{F}_{-1}^X = \{\Omega, \emptyset\}$. $\mathcal{F}_\infty^X$ is the σ-algebra defined as

$$\mathcal{F}_\infty^X \doteq \sigma\{X_k, k = 0, 1, 2, ...\}. \tag{2.7}$$

We can find that

$$L_n = \prod_{j=1}^{n} W_j = \prod_{j=1}^{n-1} W_j \cdot W_n = L_{n-1} W_n. \tag{2.8}$$

To prove the following theorem, we first introduce some new notations. The set E^∞ and $\left(\mathbb{R}^d\right)^\infty$ are defined as

$$\begin{aligned} E^\infty &= E \times E \times \cdots \times E \times \cdots, \\ \left(\mathbb{R}^d\right)^\infty &= \mathbb{R}^d \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d \times \cdots, \end{aligned} \tag{2.9}$$

which is the product of infinite copies of the set E and $\mathbb{R}^d$, respectively. As we have stated before, $\mathcal{B}(E^\infty)$ and $\mathcal{B}\left(\left(\mathbb{R}^d\right)^\infty\right)$ is the Borel $\sigma-$algebra defined on the set E and $\mathbb{R}^d$, respectively. Now, we introduce the Kolmogorov's consistency theorem.

Theorem 2.1: *For any set T and universally measurable space $(S_t, \mathcal{B}_t)_{t\in T}$, and any consistent family of laws $\{P_F, F \text{is finite}, F \subset T\}$, where P_F is defined on the product space $S_F = \prod_{t\in F} S_t$, there is a probability measure P_T on the product space $S_T = \prod_{t\in T} S_t$ with $P_T \circ f_{TF}^{-1} = P_F$ for all finite set $F \subset T$, where f_{TF}^{-1} is the natural projection from S_T onto S_F.*

This is Theorem 12.1.2 in [14].

The following standard result constructs the real probability measure P from the fictitious one.

Theorem 2.2: *Suppose that the signal $\{X_n, n = 0, 1, ...\}$ and the observation $\{Y_n, n = 1, 2, ...\}$ are independent stochastic processes defined on the canonical probability space $\left(\Omega = E^\infty \times \left(\mathbb{R}^d\right)^\infty, \mathcal{F} = \mathcal{B}(E^\infty) \otimes \mathcal{B}\left(\left(\mathbb{R}^d\right)^\infty\right), Q\right)$, the observations $\{Y_n, n = 1, 2, ...\}$ are a sequence of independent and identically distributed random vectors with* *strictly positive density function g defined on $\mathbb{R}^d$ and $V_n = Y_n - h(X_{n-1})$, for all $n = 1, 2, \ldots$. Then, there exists a probability measure P such that (2.3) holds, the noise $\{V_n,\ n \in \mathbb{N}\}$ are independent and identically distributed random vectors on $(\Omega, \mathcal{F}, P)$ with the common probability density function g and $\{X_n, n = 0, 1, \ldots\}$ is independent of $\{V_n, n = 1, 2, \ldots\}$ with the same law as on $(\Omega, \mathcal{F}, Q)$.*

Proof. Define P_n on $\left(\Omega = E^\infty \times \left(\mathbb{R}^d\right)^n, \mathcal{F} = \mathcal{B}(E^\infty) \otimes \mathcal{B}\left(\left(\mathbb{R}^d\right)^n\right)\right)$ by

$$\frac{\mathrm{d}P_n}{\mathrm{d}Q} = L_n, \tag{2.10}$$

and let $1 \le j_1 < j_2 < \cdots < j_k \le n$, $0 \le i_1 < i_2 < \cdots < i_l$. Then, by the independence of X and Y

$$\begin{aligned}
&E^{P_n}\left[\prod_{r=1}^{k} f_r\left(V_{j_r}\right)\prod_{p=1}^{l}\phi_p\left(X_{i_p}\right)\right] \qquad (2.11)\\
&= E^Q\left[\prod_{m=1}^{n}\frac{g(Y_m - h(X_{m-1}))}{g(Y_m)}\prod_{r=1}^{k} f_r\left(Y_{j_r} - h\left(X_{j_r-1}\right)\right)\prod_{p=1}^{l}\phi_p\left(X_{i_p}\right)\right]\\
&= E^Q\left[\prod_{p=1}^{l}\phi_p\left(X_{i_p}\right)\int_{\mathbb{R}^d} g_1(y_1 - h(X_0))\,\mathrm{d}y_1\cdots\int_{\mathbb{R}^d} g_n(y_n - h(X_{n-1}))\,\mathrm{d}y_n\right]\\
&= E^Q\left[\prod_{p=1}^{l}\phi_p\left(X_{i_p}\right)\right]\int_{\mathbb{R}^d} g_1(v_1)\,\mathrm{d}v_1\cdots\int_{\mathbb{R}^d} g_n(v_n)\,\mathrm{d}v_n\\
&= E^Q\left[\prod_{p=1}^{l}\phi_p\left(X_{i_p}\right)\right]\prod_{r=1}^{k}\int_{\mathbb{R}^d} f_r\left(v_{j_r}\right)g\left(v_{j_r}\right)\mathrm{d}v_{j_r},
\end{aligned}$$

where

$$g_i = \begin{cases} g f_r, & \text{if } i = j_r \\ g, & \text{if } i \notin \{j_1, \ldots, j_k\} \end{cases}. \tag{2.12}$$

By Theorem 2.1 and (2.11), $\{P_n\}$ are consistent. □

2.2 Notations and unnormalized filter

Under the new fictitious probability measure Q, we can define the unnormalized filter as

$$\sigma_n(f) = E^Q\left(L_n f(X_n)\middle|\mathcal{F}_n^Y\right), \tag{2.13}$$

for any $n \in \mathbb{N}$.

Lemma 2.1: *For the unnormalized filter $\sigma_n(f)$ and the conditional expectation $\pi_n(f)$, we have the following relation*

$$\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)}, \tag{2.14}$$

and

$$\sigma_0 = \pi_0, \tag{2.15}$$

for any $n \in \mathbb{N}$ and any $f \in B(E)$.

Proof. To prove that for any $f \in B(E)$,

$$\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)},$$

we have to show that

$$\pi_n(f) = E^P\left(f(X_n)\middle|\mathcal{F}_n^Y\right) = \frac{\sigma_n(f)}{\sigma_n(1)} = \frac{E^Q\left(L_n f(X_n)\middle|\mathcal{F}_n^Y\right)}{E^Q\left(L_n\middle|\mathcal{F}_n^Y\right)},$$

i.e., we need to prove that

$$E^P\left(f(X_n)\left|\mathcal{F}_n^Y\right.\right)E^Q\left(L_n\left|\mathcal{F}_n^Y\right.\right)=E^Q\left(L_nf(X_n)\left|\mathcal{F}_n^Y\right.\right),$$

for every $A\in\mathcal{F}_n^Y$, we have to show that

$$\begin{aligned}E^Q\left[E^P\left(f(X_n)\left|\mathcal{F}_n^Y\right.\right)E^Q\left(L_n\left|\mathcal{F}_n^Y\right.\right)1_A\right]=&E^Q\left[L_nf(X_n)1_A\right]\\=&E^P\left[f(X_n)1_A\right].\end{aligned}$$

However

$$\begin{aligned}LHS\ =&E^Q\left[E^P\left(f(X_n)\left|\mathcal{F}_n^Y\right.\right)E^Q\left(L_n\left|\mathcal{F}_n^Y\right.\right)1_A\right] &(2.16)\\=&E^Q\left[E^Q\left[L_n\cdot E^P\left(f(X_n)\left|\mathcal{F}_n^Y\right.\right)\cdot 1_A\left|\mathcal{F}_n^Y\right.\right]\right] &(2.17)\\=&E^Q\left[L_n\cdot E^P\left(f(X_n)\left|\mathcal{F}_n^Y\right.\right)\cdot 1_A\right]\\=&E^Q\left[L_n\cdot E^P\left(f(X_n)1_A\left|\mathcal{F}_n^Y\right.\right)\right]\\=&E^P\left[E^P\left[f(X_n)1_A\left|\mathcal{F}_n^Y\right.\right]\right]\\=&E^P\left[f(X_n)1_A\right]\\=&RHS.\end{aligned}$$

Since $E^P\left(f(X_n)\left|\mathcal{F}_n^Y\right.\right)$ and 1_A is $\mathcal{F}_n^Y$-measurable, we can get (2.17) from (2.16). Therefore, we have showed that

$$\pi_n(f)=\frac{\sigma_n(f)}{\sigma_n(1)},$$

for any $n\in\mathbb{N}$ and any $f\in B(E)$.

Since $L_0 = 1$, we have

$$\sigma_0(f) = E^Q\left(L_0 f(X_0) \middle| \mathcal{F}_0^Y\right) = E^Q\left(f(X_0) \middle| \mathcal{F}_0^Y\right),$$

and

$$\pi_0(f) = \frac{\sigma_0(f)}{\sigma_0(1)} = \frac{E^Q\left(L_0 f(X_0) \middle| \mathcal{F}_0^Y\right)}{E^Q\left(L_0 \middle| \mathcal{F}_0^Y\right)} = E^Q\left(f(X_0) \middle| \mathcal{F}_0^Y\right),$$

for any $f \in B(E)$.

Hence, we can get

$$\sigma_0(f) = \pi_0(f),$$

and since f is any function from $B(E)$

$$\sigma_0 = \pi_0.$$

□

By Lemma 2.1, it is enough to construct a particle filter approximation σ_n^N to the unnormalized filter σ_n, since we can then construct our filter approximation to π_n as $\pi_n^N(f) = \frac{\sigma_n^N(f)}{\sigma_n^N(1)}$.

Now, we introduce some lemmas about the unnormalized filter $\sigma_n(f)$.

As we have stated before, the signal process $\{X_n, n \in \mathbb{N}\}$ is a time-homogeneous discrete-time Markov process, where the initial distribution is π_0 and the Markov transition kernel is K for any time n. Since the signal process $\{X_n, n \in \mathbb{N}\}$ has the Markov transition kernel K under both the original probability measure P and the new fictitious probability measure Q, we have that

$$Q\left(X_{n+1} \in A \middle| \mathcal{F}_n^X\right) = P\left(X_{n+1} \in A \middle| \mathcal{F}_n^X\right) = K(X_n, A), \tag{2.18}$$

and by (1.11)

$$E^Q\left[f(X_n)\left|\mathcal{F}_{n-1}^X\right.\right] = E^P\left[f(X_n)\left|\mathcal{F}_{n-1}^X\right.\right] = Kf(X_{n-1}), \tag{2.19}$$

for any set $A \subset E$ and $n \in \mathbb{N}$.

Theorem 2.3: *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then*

$$E\left[E\left(X\left|\mathcal{G}\right.\right)\left|\mathcal{H}\right.\right] = E\left[X\left|\mathcal{H}\right.\right].$$

This is the *Tower Property* for the conditional expectation in [16].

By the pervious results, we can get the recursion formula for the unnormalized filter.

Lemma 2.2: *For the unnormalized filter, we have*

$$\sigma_n(f) = \sigma_{n-1}(A_n f), \tag{2.20}$$

where the operator A_n is defined as

$$A_n f(x) = \begin{cases} \frac{g(Y_n - h(x))}{g(Y_n)} Kf(x) = \alpha_n(x) Kf(x), & n = 1, 2, ... \\ f(x), & n = 0 \end{cases}, \tag{2.21}$$

for any $f \in B(E)$, $x \in E$ and $n \in \mathbb{N}$.

Proof. By the definition of the unnormalized filter, we can get

$$\begin{aligned}
\sigma_n(f) =& E^Q\left[L_n f(X_n)\left|\mathcal{F}_n^Y\right.\right] && (2.22)\\
=& E^Q\left[E^Q\left[L_n f(X_n)\left|\mathcal{F}_n^Y \vee \mathcal{F}_{n-1}^X\right.\right]\left|\mathcal{F}_n^Y\right.\right] && (2.23)\\
=& E^Q\left[E^Q\left[W_n L_{n-1} f(X_n)\left|\mathcal{F}_n^Y \vee \mathcal{F}_{n-1}^X\right.\right]\left|\mathcal{F}_n^Y\right.\right] && (2.24)\\
=& E^Q\left[W_n L_{n-1} E^Q\left[f(X_n)\left|\mathcal{F}_{n-1}^X\right.\right]\left|\mathcal{F}_n^Y\right.\right] && (2.25)\\
=& E^Q\left[W_n L_{n-1} K f(X_{n-1})\left|\mathcal{F}_n^Y\right.\right]\\
=& E^Q\left[L_{n-1}\frac{g(Y_n - h(X_{n-1}))}{g(Y_n)} K f(X_{n-1})\left|\mathcal{F}_n^Y\right.\right]\\
=& E^Q\left[L_{n-1} A_n f(X_{n-1})\left|\mathcal{F}_{n-1}^Y\right.\right]\\
=& \sigma_{n-1}(A_n f),
\end{aligned}$$

for any $f \in B(E)$ and $n \in \mathbb{N}$.

Due to the tower property of the conditional expectation, we can get (2.23) from (2.22). By the independence of X and Y under the new fictitious probability measure Q, we can have (2.25) from (2.24). □

If we define the composite operator $\{A_{i,n}, n \in \mathbb{N}, 1 \leq i \leq n+1\}$ as

$$A_{i,n} f(x) = \begin{cases} A_i(A_{i+1} \cdots (A_n f))(x) & \forall i \leq n \\ f(x) & i = n+1 \end{cases}, \tag{2.26}$$

by applying Lemma 2.2 repeatedly, we can get that

$$\sigma_n(f) = \sigma_0(A_{1,n} f) = \pi_0(A_{1,n} f). \tag{2.27}$$

This immediately implies that

$$\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)} = \frac{\sigma_{n-1}(A_n f)}{\sigma_{n-1}(A_n 1)} = \frac{\pi_0(A_{1,n} f)}{\pi_0(A_{1,n} 1)}. \tag{2.28}$$

Now, we discuss the moment condition about the unnormalized filter.

Lemma 2.3: *Suppose that* $f \in B(E)_+$, *then we have*

$$E^Q[\sigma_n(f)] = \pi_0(K^n f), \tag{2.29}$$

and

$$E^Q[\sigma_n(f)] < \infty,$$

for any $n \in \mathbb{N}$.

Proof. Notice that g is the probability density function. Taking expectations over Y, we have that

$$\begin{aligned} E^Q\left[\alpha_j(x)\right] =& E^Q\left[\frac{g\left(Y_j - h(x)\right)}{g\left(Y_j\right)}\right] \\ =& \int_{\mathbb{R}^d} \frac{g(y - h(x))}{g(y)} g(y)\, \mathrm{d}y \\ =& \int_{\mathbb{R}^d} g(y - h(x))\, \mathrm{d}y \\ =& \int_{\mathbb{R}^d} g(y')\, \mathrm{d}y' \quad \text{Let } y' = y - h(x) \\ =& 1, \end{aligned} \tag{2.30}$$

for any $j \in \mathbb{N}$.

Therefore, we can get

$$
\begin{aligned}
E^{Q}\left[\sigma_{n}(f)\right] =&E^{Q}\left[\pi_{0}\left(A_{1,n}f\right)\right] \qquad (2.31)\\
=&E^{Q}\left[\int_{E}A_{1,n}f(x)\,\pi_{0}(\mathrm{d}x)\right]\\
=&E^{Q}\left[\int_{E}A_{1}A_{2,n}f(x)\,\pi_{0}(\mathrm{d}x)\right]\\
=&\int_{E}E^{Q}\left[A_{1}A_{2,n}f(x)\right]\pi_{0}(\mathrm{d}x) \text{ By Fubini's Theroem}\\
=&\int_{E}E^{Q}\left[\frac{g\left(Y_{1}-h(x)\right)}{g\left(Y_{1}\right)}KA_{2,n}f(x)\right]\pi_{0}(\mathrm{d}x)\\
=&\int_{E}E^{Q}\left[\int_{E}A_{2,n}f(\mathrm{z})\,K(x,\mathrm{dz})\right]\pi_{0}(\mathrm{d}x)\\
=&\int_{E}\int_{E}E^{Q}[A_{2,n}f(\mathrm{z})]K(x,\mathrm{dz})\,\pi_{0}(\mathrm{d}x)\\
=&\int_{E}E^{Q}\left[A_{2,n}f(\mathrm{z})\right]K\pi_{0}(\mathrm{dz})\\
=&\int_{E}E^{Q}\left[A_{n+1,n}f(\mathrm{z})\right]K^{n}\pi_{0}(\mathrm{dz}) \text{ Apply the technique recursively}\\
=&\int_{E}f(\mathrm{z})\left(K^{n}\pi_{0}\right)(\mathrm{dz})\\
=&K^{n}\pi_{0}(f)\\
=&\pi_{0}\left(K^{n}f\right) \text{ By (1.7).}
\end{aligned}
$$

Since the function f is bounded, we have

$$E^{Q}\left[\sigma_{n}(f)\right] < \infty,$$

for any $n \in \mathbb{N}$. □

Now, it will be helpful to give a new definition for the next lemmas.

The *observation variability function* is defined as

$$\lambda(x,\xi) = \int \frac{g(y-h(x))\,g(y-h(\xi))}{g(y)}\mathrm{d}y.$$

It is also useful to define the single variable version

$$\bar{\lambda}(x) = \lambda(x,x) = \int \frac{[g(y-h(x))]^2}{g(y)}\mathrm{d}y.$$

We show the observation variability functions of two kinds of probability density functions.

1. Suppose that g is the probability density function of a normally distributed random variable X, *i.e.* $X \sim \mathcal{N}(m,\sigma)$, then

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-m)^2}{2\sigma^2}\right), \tag{2.32}$$

where $y \in (-\infty,\infty)$ and $\sigma > 0$.

Therefore we have

$$\begin{aligned}
\lambda(x,\xi) &= \int_{-\infty}^{\infty} \frac{g(y-h(x))\,g(y-h(\xi))}{g(y)}\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-h(x)-m)^2}{2\sigma^2} - \frac{(y-h(\xi)-m)^2}{2\sigma^2} + \frac{(y-m)^2}{2\sigma^2}\right)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[y-(h(x)+h(\xi)+m)]^2}{2\sigma^2} + \frac{2h(x)\,h(\xi)}{2\sigma^2}\right)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[y-(h(x)+h(\xi)+m)]^2}{2\sigma^2}\right)\exp\left(\frac{h(x)\,h(\xi)}{\sigma^2}\right)\mathrm{d}y \\
&= \exp\left(\frac{h(x)\,h(\xi)}{\sigma^2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[y-(h(x)+h(\xi)+m)]^2}{2\sigma^2}\right)\mathrm{d}y \\
&= \exp\left(\frac{h(x)\,h(\xi)}{\sigma^2}\right).
\end{aligned}$$

2. Suppose that g is the probability density function of a double exponential distributed random variable X, then

$$g(x) = \begin{cases} \frac{1}{2}e^{-x}, & x \geq 0, \quad (2.33) \\ \frac{1}{2}e^{x}, & x < 0, \quad (2.33') \end{cases}$$

and the sensor function h is non-negative.

Then, when $h(x) \leq h(\xi)$, we have that

$$\lambda(x,\xi) = \frac{1}{2}e^{-h(x)-h(\xi)}\left(\int_{-\infty}^{0} e^{y}\mathrm{d}y + \int_{0}^{h(x)} e^{3y}\mathrm{d}y\right) + \frac{1}{2}e^{h(x)-h(\xi)}\int_{h(x)}^{h(\xi)} e^{y}\mathrm{d}y + \frac{1}{2}e^{h(x)+h(\xi)}\int_{h(\xi)}^{\infty} e^{-y}\mathrm{d}y \quad (2.34)$$

$$= \frac{1}{2}e^{-h(x)-h(\xi)}\left[\frac{2}{3} + \frac{1}{3}e^{3h(x)}\right] + \frac{1}{2}e^{h(x)-h(\xi)}\left[e^{h(\xi)} - e^{h(x)}\right] + \frac{1}{2}e^{h(x)}$$

$$= \frac{1}{3}e^{-h(x)-h(\xi)} - \frac{1}{3}e^{2h(x)-h(\xi)} + e^{h(x)}.$$

Hence, by symmetry, we have that

$$\lambda(x,\xi) = \frac{1}{3}\left[e^{-h(x)-h(\xi)} - e^{2h(x)\wedge h(\xi)-h(\xi)\vee h(x)}\right] + e^{h(x)\wedge h(\xi)}, \quad (2.35)$$

where $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$, for any numbers a and b.

Many of our constraints and calculations are naturally placed upon the observation variability function λ and the Markov transition kernel K. Indeed, they will largely appear as the one and two variable combined kernels

$$\bar{K}_{\lambda}(x,\mathrm{d}z) = \bar{\lambda}(x)\,K(x,\mathrm{d}z), \quad (2.36)$$

and

$$K_\lambda(x,\xi,\mathrm{dz},\mathrm{d}\zeta) = \lambda(x,\xi)\,K(x,\mathrm{dz})\,K(\xi,\mathrm{d}\zeta)\,. \tag{2.37}$$

For any $f \in B(E)$, we have

$$\begin{aligned} K_\lambda(f \times f)(x,\xi) &= \int_E \int_E f(z)\, f(\zeta)\, K_\lambda(x,\xi,\mathrm{dz},\mathrm{d}\zeta) \\ &= \int_E \int_E f(z)\, f(\zeta)\, \lambda(x,\xi)\, K(x,\mathrm{dz})\, K(\xi,\mathrm{d}\zeta)\,. \end{aligned} \tag{2.38}$$

Lemma 2.4: *Suppose that $f \in B(E)_+$, then we have*

$$E^Q\left[L_n f(X_n)\right]^2 = \pi_0\left(\bar{K}_\lambda^n\left(f^2\right)\right), \tag{2.39}$$

and if $\pi_0\left(\bar{K}_\lambda^n\left(f^2\right)\right) < \infty$, then

$$E^Q\left[L_n f(X_n)\right]^2 < \infty,$$

for any $n \in \mathbb{N}$.

Proof. Taking expectation over Y, we have that

$$\begin{aligned} E^Q\left[\alpha_j(x)\cdot\alpha_j(\xi)\right] =& E^Q\left[\frac{\left(g\left(Y_j - h(x)\right)\right)\left(g\left(Y_j - h(\xi)\right)\right)}{g^2\left(Y_j\right)}\right] \\ =& \int \frac{\left(g(y-h(x))\right)\left(g(y-h(\xi))\right)}{g^2(y)} \cdot g(y)\,\mathrm{d}y \\ =& \int \frac{g(y-h(x))\,g(y-h(\xi))}{g(y)}\mathrm{d}y \\ =& \lambda(x,\xi)\,, \end{aligned} \tag{2.40}$$

for any $j \in \mathbb{N}$.

Then we can get

$$
\begin{aligned}
E^{Q}\left[L_n f(X_n)\right]^2 &= E^{Q}\left[L_n^2 f^2(X_n)\right] \\
&= E^{Q}\left[\prod_{l=1}^{n} \gamma_l^2 (X_{l-1})\, f^2(X_n)\right] \\
&= E^{Q}\left[\prod_{l=1}^{n} \bar{\lambda}(X_{l-1})\, f^2(X_n)\right] \\
&= E^{Q}\left[\prod_{l=1}^{n-1} \bar{\lambda}(X_{l-1})\, \bar{\lambda}(X_{n-1})\, K f^2(X_{n-1})\right] \\
&= E^{Q}\left[\prod_{l=1}^{n-1} \bar{\lambda}(X_{l-1})\, \bar{K}_{\lambda} f^2(X_{n-1})\right] \\
&= \pi_0\left(\bar{K}_{\lambda}^{n}\left(f^2\right)\right), \text{ Apply the technique recursively}
\end{aligned}
\tag{2.41}
$$

for any $n \in \mathbb{N}$.

If $\pi_0\left(\bar{K}_{\lambda}^{n}\left(f^2\right)\right) < \infty$, then

$$E^{Q}\left[L_n f(X_n)\right]^2 < \infty,$$

for any $n \in \mathbb{N}$. □

Remark: The condition

$$\pi_0\left(\bar{K}_{\lambda}^{n}\left(f^2\right)\right) < \infty$$

might seem hard to verify. However, if the observation variability function λ is bounded by B, then

$$\pi_0\left(\bar{K}_{\lambda}^{n}\left(f^2\right)\right) < |f|_{\infty}^{2} B^n.$$

It follows by the previous example 1, we have that λ is bounded if the observation noise is Gaussian distributed and the sensor function h is bounded. By the previous example 2, if the observation noise is double exponential distributed and the sensor function h is bounded, then the observation variability function is bounded.

Lemma 2.5: *Suppose that* $f \in B(E)_+$*, then we have*

$$E^Q\left[\sigma_n^2(f)\right] = \pi_0 \times \pi_0\left(K_\lambda^n(f \times f)\right), \tag{2.42}$$

and if $\pi_0 \times \pi_0\left(K_\lambda^n(f \times f)\right) < \infty$*, then*

$$E^Q\left[\sigma_n^2(f)\right] < \infty,$$

for any $n \in \mathbb{N}$.

Proof. By (2.27), we can get

$$\begin{aligned} E^Q\left[\sigma_n^2(f)\right] =& E^Q\left[\pi_0^2(A_{1,n}f)\right] \\ =& E^Q\left[\int_E A_{1,n}f(x)\,\pi_0(\mathrm{d}x)\cdot\int_E A_{1,n}f(\xi)\,\pi_0(\mathrm{d}\xi)\right] \\ =& \int_E\int_E E^Q\left[A_{1,n}f(x)\,A_{1,n}f(\xi)\right]\pi_0(\mathrm{d}x)\,\pi_0(\mathrm{d}\xi), \end{aligned} \tag{2.43}$$

and

$$\begin{aligned} E^Q\left[A_{i,n}f(x)\,A_{i,n}f(\xi)\right] =& E^Q\left[A_iA_{i+1,n}f(x)\,A_iA_{i+1,n}f(\xi)\right] \\ =& E^Q\left[\frac{g(Y_i - h(x))}{g(Y_i)}KA_{i+1,n}f(x)\,\frac{g(Y_i - h(\xi))}{g(Y_i)}KA_{i+1,n}f(\xi)\right] \\ =& E^Q\left[\gamma_i(x)\,\gamma_i(\xi)\int_E A_{i+1,n}f(\mathrm{z})\,K(x,\mathrm{dz})\int_E A_{i+1,n}f(\zeta)\,K(\xi,\mathrm{d}\zeta)\right] \\ =& \int_E\int_E \lambda(x,\xi)\,K(x,\mathrm{dz})\,K(\xi,\mathrm{d}\zeta)\,E^Q\left[A_{i+1,n}f(\mathrm{z})A_{i+1,n}f(\zeta)\right] \\ =& \int_E\int_E E^Q\left[A_{i+1,n}f(\mathrm{z})\,A_{i+1,n}f(\zeta)\right]K_\lambda(x,\xi,\mathrm{dz},\mathrm{d}\zeta), \end{aligned} \tag{2.44}$$

for any $i = 1, 2, ..., n$.

Therefore, we can get by substitution that

$$E^Q\left[\sigma_n^2(f)\right] = \pi_0 \times \pi_0\left(K_\lambda^n(f \times f)\right) < \infty,$$

for any $n \in \mathbb{N}$. $\square$

Remark: The condition

$$\pi_0 \times \pi_0\left(K_\lambda^n(f \times f)\right) < \infty,$$

might seem hard to verify. However, if the observation variability function λ is bounded by B, then

$$\pi_0 \times \pi_0\left(K_\lambda^n(f \times f)\right) < |f|_\infty^2 B^n.$$

It follows by the previous example 1, we have that λ is bounded if the observation noise is Gaussian distributed and the sensor function h is bounded. By the previous example 2, if the observation noise is double exponential distributed and the sensor function h is bounded, then the observation variability function is bounded.

Remark: By Lemma 2.3 and Lemma 2.5, the variance of the unnormalized filter $\sigma_n(f)$ under the new fictitious probability measure Q is

$$\begin{aligned} E^Q\left[\sigma_n(f) - E^Q(\sigma_n(f))\right]^2 =& E^Q\left[\sigma_n^2(f)\right] - \left[E^Q(\sigma_n(f))\right]^2 \\ =& \pi_0 \times \pi_0\left(K_\lambda^n(f \times f)\right) - \left(\pi_0\left(K^n f\right)\right)^2. \end{aligned} \tag{2.45}$$

By Lemma 2.3 and Lemma 2.5, we have showed that, under some conditions, the first moment and the second moment of the unnormalized filter $\sigma_n(f)$ with respect to the new fictitious probability measure Q are bounded. These lemmas establish sufficient regularity for our Strong Law of Large Numbers and Central Limit Theorem results

Define $\mathcal{F}_{\infty}^{Y} = \sigma\{Y_n,\ n = 1, 2, 3, ...\}$. In the sequel, we will fix an observation path, set

$$Q^{Y}(\cdot) = Q\left(\cdot \middle| \mathcal{F}_{\infty}^{Y}\right), \tag{2.46}$$

and let $E^{Y}[\cdot]$ denote the expectation with respect to Q^{Y}.

Chapter 3

Weighted Particle System

3.1 Introduction of the particle system

In the weighted particle filter, we create fixed number $N \in \mathbb{N}$ independent copies of the signal process, called the *particles* and simulate them simultaneously. The weighted particle system do not utilize resampling.
At time n, the particle system is defined as $\left\{X_n^i\right\}_{i=1}^N$. Each particle is a sample from the signal process. That means, each particle X_n^i is a time-homogeneous discrete-time Markov process defined on the measurable space $(\Omega, \mathcal{F}, Q)$, where the initial distribution is π_0 and the Markov transition kernel is K for any time. $\left\{X^i\right\}_{i=1}^N$ are independent from each other, and under the new fictitious probability measure Q, $\left\{X^i\right\}_{i=1}^N$ are also independent from the observation process Y. We note that $\left(X^i, Y\right)$ has the same distribution as (X, Y), for any $1 \leq i \leq N$.

The weight of each particle is defined as

$$L_n^i = \prod_{j=1}^{n} W_j^i, \tag{3.1}$$

$$W_j^i = \alpha_j\left(X_{j-1}^i\right) = \frac{g\left(Y_j - h\left(X_{j-1}^i\right)\right)}{g\left(Y_j\right)},$$

and

$$L_0^i = 1,$$

for any $n \in \mathbb{N}$, $1 \leq j \leq n$ and $1 \leq i \leq N$.

For each particle, we can define the single particle measures

$$\beta_n^k = L_n^k \delta_{X_n^k} \quad \text{and} \quad \beta_{-1}^k = \pi_0, \tag{3.2}$$

for any $n \in \mathbb{N}$ and $1 \leq k \leq N$.

Then we have the the following measure-valued evolution

$$\begin{aligned}
\beta_n^k(f) &= L_n^k f\left(X_n^k\right) \qquad (3.3)\\
&= L_{n-1}^k \frac{g\left(Y_n - h\left(X_{n-1}^k\right)\right)}{g(Y_n)} K f\left(X_{n-1}^k\right) + L_n^k\left[f\left(X_n^k\right) - E^Y\left(f\left(X_n^k\right)\middle|\mathcal{F}_{n-1}^X\right)\right]\\
&= \beta_{n-1}^k(A_n f) + L_n^k\left[f\left(X_n^k\right) - E^Y\left(f\left(X_n^k\right)\middle|\mathcal{F}_{n-1}^X\right)\right]\\
&= \beta_0^k(A_{1,n} f) + \sum_{l=1}^{n} L_l^k\left[A_{l+1,n} f\left(X_l^k\right) - E^Y\left(A_{l+1,n} f\left(X_l^k\right)\middle|\mathcal{F}_{l-1}^X\right)\right] \text{ By recursion}\\
&= \pi_0(A_{1,n} f) + M_n^{\beta^k}(f)\\
&= \sigma_n(f) + M_n^{\beta^k}(f),
\end{aligned}$$

where

$$
\begin{aligned}
M_n^{\beta^k}(f) &= \sum_{l=0}^{n} \left[L_l^k \left(A_{l+1,n} f\left(X_l^k\right) - E^Y\left[A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X \right] \right) \right] \\
&= \sum_{l=0}^{n} \left[L_l^k A_{l+1,n} f\left(X_l^k\right) - L_l^k E^Y\left(A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X \right) \right] \\
&= \sum_{l=0}^{n} \left[L_l^k A_{l+1,n} f\left(X_l^k\right) - L_l^k K A_{l+1,n} f\left(X_{l-1}^k\right) \right] \\
&= \sum_{l=0}^{n} \left[L_l^k A_{l+1,n} f\left(X_l^k\right) - L_{l-1}^k \frac{g\left(Y_l - h\left(X_{l-1}^k\right)\right)}{g\left(Y_l\right)} K A_{l+1,n} f\left(X_{l-1}^k\right) \right] \\
&= \sum_{l=0}^{n} \left[L_l^k A_{l+1,n} f\left(X_l^k\right) - L_{l-1}^k A_l A_{l+1,n} f\left(X_{l-1}^k\right) \right] \\
&= \sum_{l=0}^{n} \left[L_l^k A_{l+1,n} f\left(X_l^k\right) - L_{l-1}^k A_{l,n} f\left(X_{l-1}^k\right) \right] \\
&= \sum_{l=0}^{n} \left[\beta_l^k \left(A_{l+1,n} f\right) - \beta_{l-1}^k \left(A_{l,n} f\right) \right],
\end{aligned}
\tag{3.4}
$$

for any $n \in \mathbb{N}$ and $1 \leq k \leq N$.

If we average over the particles and define $\sigma_n^N(f)$ as

$$
\sigma_n^N(f) = \frac{1}{N} \sum_{i=1}^{N} L_n^i f\left(X_n^i\right), \tag{3.5}
$$

then

$$
\sigma_n^N(f) = \sigma_n(f) + M_n^N(f), \tag{3.6}
$$

where

$$
M_n^N(f) = \sum_{l=0}^{n} \left[\sigma_l^N \left(A_{l+1,n} f\right) - \sigma_{l-1}^N \left(A_{l,n} f\right) \right], \tag{3.7}
$$

for any $n \in \mathbb{N}$.

Lemma 3.1: $\left\{M_n^N(f), n \in \mathbb{N}\right\}$ *is a zero-mean martingale in n with respect to* Q^Y.

Proof. First, we calculate the expectation of $M_n^{\beta^k}(f)$.

Since L_l^k is $\mathcal{F}_l^Y \vee \mathcal{F}_{l-1}^X$-measurable for any $1 \le k \le N$, we have

$$\begin{aligned} E^Y\left[M_n^{\beta^k}(f)\right] =& E^Y\left[\sum_{l=0}^{n}\left[L_l^k\left(A_{l+1,n} f\left(X_l^k\right) - E^Y\left[A_{l+1,n} f\left(X_l^k\right)\middle|\mathcal{F}_{l-1}^X\right]\right)\right]\right] \qquad (3.8) \\ =& \sum_{l=0}^{n} E^Y\left[L_l^k\left(A_{l+1,n} f\left(X_l^k\right) - E^Y\left(A_{l+1,n} f\left(X_l^k\right)\middle|\mathcal{F}_{l-1}^X\right)\right)\right] \\ =& \sum_{l=0}^{n}\left[E^Y\left(L_l^k A_{l+1,n} f\left(X_l^k\right)\right) - E^Y\left[L_l^k E^Y\left(A_{l+1,n} f\left(X_l^k\right)\middle|\mathcal{F}_{l-1}^X\right)\right]\right] \\ =& \sum_{l=0}^{n}\left[E^Y\left(L_l^k A_{l+1,n} f\left(X_l^k\right)\right) - E^Y\left[E^Y\left(L_l^k A_{l+1,n} f\left(X_l^k\right)\middle|\mathcal{F}_{l-1}^X\right)\right]\right] \\ =& \sum_{l=0}^{n}\left[E^Y\left(L_l^k A_{l+1,n} f\left(X_l^k\right)\right) - E^Y\left(L_l^k A_{l+1,n} f\left(X_l^k\right)\right)\right] \\ =& 0. \end{aligned}$$

By the relation between $M_n^{\beta^k}(f)$ and $M_n^N(f)$

$$E^Y\left[M_n^N(f)\right] = \frac{1}{N}\sum_{k=1}^{N} E^Y\left[M_n^{\beta^k}(f)\right] = 0.$$

Therefore, the expectation of $\left\{M_n^N(f), n \in \mathbb{N}\right\}$ is zero with respect to Q^Y.

For any $l \in \mathbb{N}$ and $1 \leq k \leq N$, we have

$$
\begin{aligned}
& E^Y\left[M_l^{\beta^k}(f) - M_{l-1}^{\beta^k}(f) \middle| \mathcal{F}_{l-1}^X\right] \\
=& E^Y\left[L_l^k\left(A_{l+1,n} f\left(X_l^k\right) - E^Y\left(A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right)\right) \middle| \mathcal{F}_{l-1}^X\right] \\
=& E^Y\left[L_l^k A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right] - E^Y\left[L_l^k E^Y\left(A_{k+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right) \middle| \mathcal{F}_{l-1}^X\right] \\
=& E^Y\left[L_l^k A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right] - E^Y\left[E^Y\left(L_l^k A_{k+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right) \middle| \mathcal{F}_{l-1}^X\right] \\
=& E^Y\left[L_l^k A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right] - E^Y\left[L_l^k A_{l+1,n} f\left(X_l^k\right) \middle| \mathcal{F}_{l-1}^X\right] \\
=& 0.
\end{aligned}
\tag{3.9}
$$

Hence, we can get

$$
E^Y\left[M_l^{\beta^k}(f) \middle| \mathcal{F}_{l-1}^X\right] = M_{l-1}^{\beta^k}(f).
$$

By averaging over the particles, we have

$$
E^Y\left[M_l^N(f) \middle| \mathcal{F}_{l-1}^X\right] = M_{l-1}^N(f).
$$

That means, $\left\{M_n^N(f), n \in \mathbb{N}\right\}$ is a martingale in n with respect to Q^Y. □

By Lemma 3.1, we know that $\left\{M_n^N(f), n = 0, 1, ...\right\}$ is a zero-mean martingale in n as well as the average of N independent and identically distributed zero-mean random variables over i both with respect to Q^Y.

3.2 Main results

3.2.1 Strong Law of Large Numbers

First, we introduce a new definition and theorem stated in [1].

Definition: A class of functions $\mathcal{M}$, defined on the topological space E, *strongly separate points* $(s.s.p)$ if for every $x \in E$ and neighborhood O_x of x, there is a finite collection $\left\{g^1, ..., g^k\right\} \subset \mathcal{M}$ such that

$$\inf_{y \notin O_x} \max_{1 \leq l \leq k} \left| g^l(y) - g^l(x) \right| > 0. \tag{3.10}$$

Define a new class of functions as

$$A = \{f_i\}_{i=1}^{\infty} = \left\{ \prod_{j=1}^{l} \left(1 - d\left(x_j, \cdot\right)\right) \vee 0 : l \in \mathbb{N}, x_j \in \{y_k\}_{k=1}^{\infty} \right\}, \tag{3.11}$$

for some dense collection $\{y_k\}_{k=1}^{\infty} \subset E$. From [1], we know that the class of functions A $s.s.p$.

Theorem 3.1: *Suppose that* $(E, \mathcal{T})$ *is a topological space,* $\{P_n\} \cup \{P\} \subset \mathcal{P}(E)$ *and a class of functions* $\mathcal{M} \subset B(E)$ *that strongly separates points and is closed under multiplication, and*

$$\int_E g \mathrm{d}P_n \to \int_E g \mathrm{d}P, \quad \forall g \in \mathcal{M}. \tag{3.12}$$

If $\mathcal{M}$ *is countable, then*

$$P_n \Rightarrow P, \tag{3.13}$$

where $(\Rightarrow)$ *means weak convergence.*

Now, we can state and prove the Strong Law of Large Numbers for the weighted particle filter.

Theorem 3.2: *Suppose that the unnormalized filter is defined as*

$$\sigma_n(f) = E^Q\left(L_n f(X_n)\middle|\mathcal{F}_n^Y\right), \tag{3.14}$$

then for the weighted particle system,

$$\sigma_n^N \Rightarrow \sigma_n \quad a.s. \quad \left[Q^Y\right], \tag{3.15}$$

when $N \to \infty$.

Proof. By (3.7), $\left\{M_n^N(f), n \in \mathbb{N}\right\}$ is defined as

$$M_n^N(f) = \sum_{l=0}^{n}\left[\sigma_l^N(A_{l+1,n}f) - \sigma_{l-1}^N(A_{l,n}f)\right], \tag{3.16}$$

and it is the average of N independent and identically distributed random variables over i with respect to Q^Y.

The expectation of $\beta_n^k(f)$ is

$$\begin{aligned} E^Y\left[\beta_n^k(f)\right] &= E^Y\left[\sigma_n(f) + M_n^{\beta^k}(f)\right] \\ &= E^Y\left[\sigma_n(f)\right] + E^Y\left[M_n^{\beta^k}(f)\right] \\ &= E^Q\left[E^Q\left(L_n f(X_n)\middle|\mathcal{F}_n^Y\right)\middle|\mathcal{F}_\infty^Y\right] + 0 \\ &= E^Q\left[L_n f(X_n)\middle|\mathcal{F}_n^Y\right] \\ &= \sigma_n(f) \\ &= \pi_0(A_{1,n}f). \end{aligned} \tag{3.17}$$

Then the expectation of $\sigma_n^N(f)$ is

$$E^Y\left[\sigma_n^N(f)\right] = \frac{1}{N}\sum_{k=1}^{N} E^Y\left[\beta_n^k(f)\right] = \sigma_n(f) = \pi_0(A_{1,n}f). \tag{3.18}$$

Since by Lemma 2.3, we have

$$E^Y\left[\left|L_n^i f\left(X_n^i\right)\right|\right] = E^Y\left[\sigma_n(f)\right] = \pi_0(K^n f) < \infty, \tag{3.19}$$

for any $1 \le i \le N$ and any function $f \in B(E)_+$.

Then, the Strong Law of Large Numbers implies that

$$M_n^N(f) \to 0 \quad \text{a.s.} \quad \left[Q^Y\right], \tag{3.20}$$

so

$$\sigma_n^N(f) \to \sigma_n(f) \quad \text{a.s.} \quad \left[Q^Y\right], \tag{3.21}$$

for all $f \in B(E)_+$.

For the set A defined above, it satisfies that $A \subset \overline{C}(E)_+$ and A is countable. Notice that $A \subset B(E)_+$, then we have

$$\sigma_n^N(f) \to \sigma_n(f) \quad \text{a.s.} \quad \left[Q^Y\right], \tag{3.22}$$

for all $f \in A$. In addition, the set A strongly separate points and is closed under multiplication. Therefore, by Theorem 3.1, we can get

$$\sigma_n^N \Rightarrow \sigma_n \quad \text{a.s.} \quad \left[Q^Y\right]. \tag{3.23}$$

□

3.2.2 Central Limit Theorem

Define an operator $A_j^{(2)}$ as

$$A_j^{(2)} f(x) = \begin{cases} \alpha_j^2(x) K f(x), & j = 1, 2, ... \\ f(x), & j = 0 \end{cases}, \tag{3.24}$$

and also the composite operator

$$A_{i,n}^{(2)} f(x) = \begin{cases} A_i^{(2)} \left(A_{i+1}^{(2)} \cdots \left(A_n^{(2)} f \right) \right)(x), \ \forall i \le n \\ f(x), & i = n+1 \end{cases}. \tag{3.25}$$

Now, we state and prove the Central Limit Theorem for the weighted particle system.

Theorem 3.3: *Let* $f \in B(E)_+$ *satisfy*

$$\pi_0 \left(\bar{K}_\lambda^n \left(f^2 \right) \right) < \infty, \tag{3.26}$$

and

$$\pi_0 \times \pi_0 \left(K_\lambda^n (f \times f) \right) < \infty. \tag{3.27}$$

Then, the weighted particle system satisfies

$$\sqrt{N} \left(\sigma_n^N(f) - \sigma_n(f) \right) \Rightarrow \mathcal{N} \left(0, \gamma_n^W(f) \right) \quad a.s. \quad \left[Q^Y \right], \tag{3.28}$$

where

$$\gamma_n^W(f) = \sum_{l=0}^{n} \pi_0 A_{1,l-1}^{(2)} \left[A_l^{(2)} (A_{l+1,n} f)^2 - (A_{l,n} f)^2 \right], \tag{3.29}$$

and

$$E^Q \left[\gamma_n^W(f) \right] = \sum_{l=0}^{n} \pi_0 \bar{K}_\lambda^{l-1} \left[\bar{K}_\lambda - K_\lambda \right] K_\lambda^{n-l} (f \times f). \tag{3.30}$$

Proof. As we have stated before, $\left\{ M_n^N(f), n \in \mathbb{N} \right\}$ is the average of N independent and identically distributed random variables over i with respect to Q^Y. In addition, by Lemma 2.4 and Lemma 2.5, we can learn that

$$E^Y \left[L_n^i f\left(X_n^i\right) - \sigma_n(f) \right]^2 \leq 2 \left\{ E^Y \left[L_n^i f\left(X_n^i\right) \right]^2 + E^Y \left[\sigma_n(f) \right]^2 \right\} < \infty, \tag{3.31}$$

for any $n \in \mathbb{N}$.

By the classical Central Limit Theorem, we have

$$\sqrt{N} \left(\sigma_n^N(f) - \sigma_n(f) \right) \Rightarrow \mathcal{N}\left(0, \gamma_n^W(f)\right) \quad \text{a.s.} \quad \left[Q^Y \right]. \tag{3.32}$$

Since for the weighted particle system, $\left(X^k, Y\right)$ has the same distribution with (X, Y), for any $1 \leq k \leq N$, we can just work with (X, Y), and denote $\beta_l = L_l \delta_{X_l}$. By the martingale property, the $\mathcal{F}_{\infty}^{Y}$-measurable random variance can be written as

$$
\begin{aligned}
\gamma_n^W(f) =& E^Y\left[M_n^{\beta^k}(f)\right]^2 - \left[E^Y\left(M_n^{\beta^k}(f)\right)\right]^2 \\
=& E^Y\left[M_n^{\beta^k}(f)\right]^2 \\
=& E^Y\left[\sum_{l=0}^{n}\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)\right]^2 \\
=& \sum_{l=0}^{n} E^Y\left[\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)\right]^2 \\
&+2\sum_{1\le i<j\le n} E^Y\left[(\beta_i(A_{i+1,n}f) - \beta_{i-1}(A_{i,n}f))\left(\beta_j\left(A_{j+1,n}f\right) - \beta_{j-1}\left(A_{j,n}f\right)\right)\right] \\
=& \sum_{l=0}^{n} E^Y\left[\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)\right]^2 \\
&+2\sum_{1\le i<j\le n} E^Y\left[\left(M_i^{\beta^k}(f) - M_{i-1}^{\beta^k}(f)\right)\left(M_j^{\beta^k}(f) - M_{j-1}^{\beta^k}(f)\right)\right] \\
=& \sum_{l=0}^{n} E^Y\left[\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)\right]^2 \\
&+2\sum_{1\le i<j\le n} E^Y\left[E^Y\left[\left(M_i^{\beta^k}(f) - M_{i-1}^{\beta^k}(f)\right)\left(M_j^{\beta^k}(f) - M_{j-1}^{\beta^k}(f)\right)\middle|\mathcal{F}_{j-1}^X\right]\right] \\
=& \sum_{l=0}^{n} E^Y\left[\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)\right]^2 \\
&+2\sum_{1\le i<j\le n} E^Y\left[\left(M_i^{\beta^k}(f) - M_{i-1}^{\beta^k}(f)\right)E^Y\left[\left(M_j^{\beta^k}(f) - M_{j-1}^{\beta^k}(f)\right)\middle|\mathcal{F}_{j-1}^X\right]\right] \\
=& \sum_{l=0}^{n} E^Y\left[\beta_l(A_{l+1,n}f) - \beta_{l-1}(A_{l,n}f)\right]^2 \\
=& \sum_{l=0}^{n}\left\{E^Y\left[\beta_l(A_{l+1,n}f)\right]^2 - E^Y\left[\beta_{l-1}(A_{l,n}f)\right]^2\right\} \\
=& \sum_{l=0}^{n}\left\{E^Y\left[L_l^2(A_{l+1,n}f)^2(X_l)\right] - E^Y\left[L_{l-1}^2(A_{l,n}f)^2(X_{l-1})\right]\right\} \\
=& \sum_{l=0}^{n}\pi_0 A_{1,l-1}^{(2)}\left[A_l^{(2)}(A_{l+1,n}f)^2 - (A_{l,n}f)^2\right].
\end{aligned}
\tag{3.33}
$$

Now, we move to calculate the expected value of $\gamma_n^W(f)$. By the fact that $\{Y_l\}$ are independent and identically distributed random vectors, and independent of X, we can find that

$$
\begin{aligned}
E^Q\left[\gamma_n^W(f)\right] &= \sum_{l=0}^{n} E^Q\left[\prod_{k=1}^{l}\bar{\lambda}(X_{k-1})\left(A_{l+1,n}f(X_l) - KA_{l+1,n}f(X_{l-1})\right)^2\right] \qquad (3.34)\\
&= \sum_{l=0}^{n} E^Q\left[\prod_{k=1}^{l}\bar{\lambda}(X_{k-1})\left(K\left(A_{l+1,n}f\right)^2(X_{l-1}) - \left(KA_{l+1,n}f(X_{l-1})\right)^2\right)\right]\\
&= \sum_{l=0}^{n} E^Q\left[\prod_{k=1}^{l-1}\bar{\lambda}(X_{k-1})\left(\int_E A_{l+1,n}f(x)\,A_{l+1,n}f(x)\,\bar{\lambda}(X_{l-1})\,K(X_{l-1},\mathrm{d}x)\right.\right.\\
&\quad - \left.\left.\int_E A_{l+1,n}f(x)\,K(X_{l-1},\mathrm{d}x)\int_E A_{l+1,n}f(\xi)\,K(X_{l-1},\mathrm{d}\xi)\,\lambda(X_{l-1},X_{l-1})\right)\right].\\
&= \sum_{l=0}^{n} E^Q\left[\prod_{k=1}^{l-1}\bar{\lambda}(X_{k-1})\left(\int_E A_{l+1,n}f(x)\,A_{l+1,n}f(x)\,\bar{\lambda}(X_{l-1})\,K(X_{l-1},\mathrm{d}x)\right.\right.\\
&\quad - \left.\left.\int_E\int_E A_{l+1,n}f(x)\,A_{l+1,n}f(\xi)\,K_\lambda(X_{l-1},X_{l-1},\mathrm{d}x,\mathrm{d}\xi)\right)\right].
\end{aligned}
$$

Hence, it follows by (2.44) that

$$
\begin{aligned}
E^Q\left[\gamma_n^W(f)\right] &= \sum_{l=0}^{n} E^Q\left[\prod_{k=1}^{l-1}\bar{\lambda}(X_{k-1})\left(\int_E K_\lambda^{n-l}(f\times f)(x,x)\,\bar{\lambda}(X_{l-1})\,K(X_{l-1},\mathrm{d}x)\right.\right.\\
&\quad - \left.\left.\left(K_\lambda^{n+1-l}(f\times f)\right)(X_{l-1},X_{l-1})\right)\right] \qquad (3.35)\\
&= \sum_{l=0}^{n} \pi_0\bar{K}_\lambda^{l-1}\left[\bar{K}_\lambda - K_\lambda\right]K_\lambda^{n-l}(f\times f).
\end{aligned}
$$

□

Chapter 4

Fictitious Particle System

4.1 Introduction of the particle system

The problem with the weighted particle system is that, due to randomness, most particles do not behave like the signal $\{X_n, n = 0, 1, ...\}$ so their weights become relatively small compared to the weights of very few good particles. This results in a particle filter that effectively consists of an average over only a very small proportion of the particles. This problem manifests itself theoretically in the large expected variance of the central limit theorem in the previous chapter and practically in the need to use a huge number of particles in most applications. To combat this effect, we introduce the resampling process.

Initially, we pretend herein that we have access to the actual unnormalized filter total mass $\{\sigma_n(1), \ n = 0, 1, 2, ...\}$ and consider an unimplementable fictitious particle system. In particular, we use the resampled algorithm given in the Chapter 1 with $\sigma_n^N(1)$ replaced with (the computer unworkable) $\sigma_n(1)$ and the particle control step is eliminated.

Suppose that $N_0 \in \mathbb{N}$ and we have the following random variables:

1. $\left\{\xi_n^{k,i;x} \,:\, k,i,n \in \mathbb{N}, x \in E\right\}$ are independent random variables with the distribution $K(x,\cdot)$,

2. $\left\{\chi^k\right\}_{k=1}^{\infty}$ are independent samples from π_0,

3. $\left\{U_n^{k,i} \,:\, k,i,n \in \mathbb{N}\right\}$ are independent Uniform[0, 1] random variables,

which are mutually independent.

We want to keep track of the fictitious particle system in terms of the first ancestor of each particle. In other words, our fictitious particle filter will be the average of N independent and identically distributed branching Markov processes $\left\{\mathcal{B}_n^k,\ n = 0,1,...\right\}$, each starting from an independent sample δ_{χ^k}. They evolve independently of each other only interacting with $\sigma_n(1)$, which is deterministic with respect to Q^Y. At any time, many of the $\mathcal{B}^k$ may have died out while others may have branched into multiple particles. For clarity, the particles at time n, that are offspring from the original particle χ^k, will be denoted as $\left\{\mathcal{X}_n^{k,i}\right\}_{i=1}^{N_n^k}$ and the weight of each particle after the resampling will only depend upon n and be denoted as $\mathcal{L}_n$. Then, the branching Markov process corresponding to the initial particle and the complete filter estimate will be

$$\mathcal{B}_n^k = \mathcal{L}_n \sum_{i=1}^{N_n^k} \delta_{\mathcal{X}_n^{k,i}} \quad \text{and} \quad \mathcal{S}_n^N = \frac{1}{N}\sum_{k=1}^{N} \mathcal{B}_n^k, \tag{4.1}$$

for any $1 \le k \le N$ and $n \in \mathbb{N}$, respectively.

Now, we define the branching Markov processes $\left\{\mathcal{B}_n^k,\ n = 0,1,...\right\}$ as follows:

Initialize: $\mathcal{X}_0^{k,1} = \chi^k$, $N_0^k = 1$ and $\mathcal{L}_0 = 1$, for $k = 1,2,...,N$.

Repeat: For $n = 1,2,...$ do

1. Estimate: $\mathcal{B}_{n-1}^k = \sum_{i=1}^{N_{n-1}^k} \mathcal{L}_{n-1}\delta_{\mathcal{X}_{n-1}^{k,i}}$

2. Weight: $\mathcal{W}_n^{k,i} = \frac{g\left(Y_n - h\left(\mathcal{X}_{n-1}^{k,i}\right)\right)}{g(Y_n)}$ and $\hat{\mathcal{L}}_n^{k,i} = \mathcal{W}_n^{k,i}\mathcal{L}_{n-1}$, for $i = 1, ..., N_{n-1}^k$

3. Resampled Weight: $\mathcal{L}_n = \sigma_n(1)$

4. Offspring Numbers: $\mathcal{Z}_n^{k,i} = 1_{U_n^{k,i} + \left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor \leq \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n}}$, $N_n^{k,i} = \left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor + \mathcal{Z}_n^{k,i}$, for $i = 1, ..., N_{n-1}^k$ and $N_n^k = N_n^{k,1} + \cdots + N_n^{k,N_{n-1}^k}$

5. Resample: Let $\hat{\mathcal{X}}_{n-1}^{k,j} = \mathcal{X}_{n-1}^{k,i}$, for $j \in \left\{N_n^{k,1} + \cdots + N_n^{k,i-1} + 1, ..., N_n^{k,1} + \cdots + N_n^{k,i}\right\}$

6. Evolve Independently: $\mathcal{X}_n^{k,i} = \xi_n^{k,i,\hat{\mathcal{X}}_{n-1}^{k,i}}$

For notational convenience, we define $\mathcal{N}_l^{k,i} = \sum_{j=1}^{i-1} N_l^{k,j}$, $\mathcal{F}_n^U = \sigma\left\{U_l^{k,i}, k, i \in \mathbb{N}, l \leq n\right\}$ and $\mathcal{F}_n^X \doteq \sigma\left\{\mathcal{X}_l^{k,i}, k, i \in \mathbb{N}, l \leq n\right\}$ for any $n \geq 0$, with the convention $\mathcal{F}_{-1}^U = \mathcal{F}_{-1}^X \doteq \{\emptyset, \Omega\}$. Define the $\sigma-$algebra $\mathcal{F}_n^{UX} \doteq \mathcal{F}_n^U \vee \mathcal{F}_n^X$, for any $n \geq -1$. After the resampling, we have $N_n^{k,i}$ particles at location $\mathcal{X}_{n-1}^{k,i}$ each with weight $\mathcal{L}_n = \sigma_n(1)$.
Hence, the effective weight at location $\mathcal{X}_{n-1}^{k,i}$ after the resampling satisfies:

$$
\begin{aligned}
E^Y\left[\mathcal{L}_n N_n^{k,i} \middle| \mathcal{F}_{n-1}^U \vee \mathcal{F}_n^X\right] =& \mathcal{L}_n\left(\left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor + 1\right)\left(\frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} - \left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor\right) + \mathcal{L}_n\left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor\left(1 - \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} + \left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor\right) \\
=& \mathcal{L}_n \cdot \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \qquad (4.2) \\
=& \hat{\mathcal{L}}_n^{k,i},
\end{aligned}
$$

which is the weight prior to the resampling. Therefore, the fictitious particle system is *unbiased*.

However, we need to go further and establish a martingale property. Averaging over the $U_n^{k,i}$, we can get that

$$
\begin{aligned}
&E^Y\left[\mathcal{L}_n \sum_{j=\mathcal{N}_n^{k,i}+1}^{\mathcal{N}_n^{k,i}+N_n^{k,i}} f\left(\mathcal{X}_n^{k,j}\right) \middle| \mathcal{F}_{n-1}^{U^{k,i}} \vee \mathcal{F}_n^X\right] \qquad (4.3)\\
&= E^Y\left[\mathcal{L}_n \sum_{j=\mathcal{N}_n^{k,i}+1}^{\mathcal{N}_n^{k,i}+\left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor} f\left(\mathcal{X}_n^{k,j}\right) + \mathcal{L}_n\left(\frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} - \left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor\right) f\left(\mathcal{X}_n^{k,\mathcal{N}_n^{k,i}+\left\lfloor \frac{\hat{\mathcal{L}}_n^{k,i}}{\mathcal{L}_n} \right\rfloor+1}\right) \middle| \mathcal{F}_{n-1}^{U^{k,i}} \vee \mathcal{F}_n^X\right],
\end{aligned}
$$

where the $\sigma-$algebra $\mathcal{F}_{n-1}^{U^{k,i}} = \sigma\left\{U_m^{l,j} : m \leq n, (l, j, m) \neq (k, i, n)\right\}$. Then N_{n-1}^k is $\mathcal{F}_{n-1}^{UX}$-measurable, for any $1 \leq k \leq N$. Using (4.3) plus the fact $N_{n-1}^k \in \mathcal{F}_{n-1}^{UX}$, we can get that

$$E^{Y}\left[\mathcal{B}_{n}^{k}(f)\middle|\mathcal{F}_{n-1}^{UX}\right] = E^{Y}\left[\sum_{j=1}^{N_{n}^{k}}\mathcal{L}_{n}f\left(\mathcal{X}_{n}^{k,j}\right)\middle|\mathcal{F}_{n-1}^{UX}\right] \tag{4.4}$$

$$= E^{Y}\left[\sum_{i=1}^{N_{n-1}^{k}}\sum_{j=\mathcal{N}_{l}^{k,i}+1}^{\mathcal{N}_{l}^{k,i}+N_{n}^{k,i}}\mathcal{L}_{n}f\left(\mathcal{X}_{n}^{k,j}\right)\middle|\mathcal{F}_{n-1}^{UX}\right]$$

$$= \sum_{i=1}^{N_{n-1}^{k}}E^{Y}\left[\mathcal{L}_{n}\sum_{j=\mathcal{N}_{l}^{k,i}+1}^{\mathcal{N}_{l}^{k,i}+N_{n}^{k,i}}f\left(\mathcal{X}_{n}^{k,j}\right)\middle|\mathcal{F}_{n-1}^{UX}\right]$$

$$= \sum_{j=1}^{N_{n-1}^{k}}E^{Y}\left[E^{Y}\left[\mathcal{L}_{n}\sum_{j=\mathcal{N}_{l}^{k,i}+1}^{\mathcal{N}_{l}^{k,i}+N_{n}^{k,i}}f\left(\mathcal{X}_{n}^{k,j}\right)\middle|\mathcal{F}_{n-1}^{U^{k,i}}\vee\mathcal{F}_{n}^{X}\right]\middle|\mathcal{F}_{n-1}^{UX}\right]$$

$$= \sum_{i=1}^{N_{n-1}^{k}}\mathcal{L}_{n}\frac{\hat{\mathcal{L}}_{n}^{k,i}}{\mathcal{L}_{n}}Kf\left(\mathcal{X}_{n-1}^{k,i}\right)$$

$$= \sum_{i=1}^{N_{n-1}^{k}}\hat{\mathcal{L}}_{n}^{k,i}Kf\left(\mathcal{X}_{n-1}^{k,i}\right)$$

$$= \sum_{i=1}^{N_{n-1}^{k}}\mathcal{W}_{n}^{k,i}\mathcal{L}_{n-1}Kf\left(\mathcal{X}_{n-1}^{k,i}\right)$$

$$= \sum_{i=1}^{N_{n-1}^{k}}\mathcal{L}_{n-1}A_{n}f\left(\mathcal{X}_{n-1}^{k,i}\right)$$

$$= \mathcal{B}_{n-1}^{k}(A_{n}f),$$

subject to

$$\mathcal{B}_{0}^{k}(f) = f\left(\chi^{k}\right), \tag{4.5}$$

for any $1 \le k \le N$.

By using (4.4) recursively, we can find that

$$\begin{aligned} E^Y\left[\mathcal{B}_n^k(f)\right] =& E^Y\left[E^Y\left(\mathcal{B}_n^k(f)\middle|\mathcal{F}_{n-1}^{UX}\right)\right] \\ =& E^Y\left[\mathcal{B}_{n-1}^k(A_n f)\right] \\ =& E^Y\left[B_0^k(A_{1,n} f)\right] \\ =& E^Y\left[A_{1,n} f\left(\chi^k\right)\right] \\ =& \pi_0(A_{1,n} f) \\ =& \sigma_n(f) \\ =& E^Y\left[\sigma_n(f)\right]. \end{aligned} \tag{4.6}$$

Applying (4.4), we have the following measure-valued evolution

$$\begin{aligned} \mathcal{B}_n^k(f) =& \mathcal{B}_{n-1}^k(A_n f) + \mathcal{B}_n^k(f) - E^Y\left[\mathcal{B}_n^k(f)\middle|\mathcal{F}_{n-1}^{UX}\right] \\ =& \mathcal{B}_0^k(A_{1,n} f) + \sum_{l=1}^{n}\left[\mathcal{B}_l^k(A_{l+1,n} f) - E^Y\left[\mathcal{B}_l^k(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right]\right] \\ =& \pi_0(A_{1,n} f) + M_n^{\mathcal{B}^k}(f) \\ =& \sigma_n(f) + M_n^{\mathcal{B}^k}(f), \end{aligned} \tag{4.7}$$

where

$$M_n^{\mathcal{B}^k}(f) = \sum_{l=0}^{n}\left[\mathcal{B}_l^k(A_{l+1,n} f) - E^Y\left[\mathcal{B}_l^k(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right]\right], \tag{4.8}$$

for any $1 \leq k \leq N$.

Averaging over the ancestral branches, we can find that

$$E^Y\left[\mathcal{S}_n^N(f)\middle|\mathcal{F}_{n-1}^{UX}\right] = \mathcal{S}_{n-1}^N(A_n f) \quad \text{subject to } \mathcal{S}_0^N(f) = \frac{1}{N}\sum_{k=1}^{N} f\left(\chi^k\right), \tag{4.9}$$

$$E^Y\left[\mathcal{S}_n^N(f)\right] = \sigma_n(f) = E^Y\left[\sigma_n(f)\right], \tag{4.10}$$

$$\mathcal{S}_n^N(f) = \sigma_n(f) + M_n^N(f), \tag{4.11}$$

where

$$M_n^N(f) = \sum_{l=0}^{n}\left[\mathcal{S}_l^N(A_{l+1,n} f) - E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right]\right]. \tag{4.12}$$

Lemma 4.1: $\left\{M_n^N(f), n \in \mathbb{N}\right\}$ *is a zero-mean martingale in n with respect to* Q^Y.

Proof. By (4.10), we have

$$E^Y\left[M_n^N(f)\right] = E^Y\left[\mathcal{S}_n^N(f) - \sigma_n(f)\right] = E^Y\left[\mathcal{S}_n^N(f)\right] - E^Y\left[\sigma_n(f)\right] = 0, \tag{4.13}$$

and

$$\begin{aligned} E^Y\left[M_l^N(f) - M_{l-1}^N(f)\middle|\mathcal{F}_{l-1}^{UX}\right] =& E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f) - E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right]\middle|\mathcal{F}_{l-1}^{UX}\right] \\ =& E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right] - E^Y\left[E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right]\middle|\mathcal{F}_{l-1}^{UX}\right] \\ =& E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right] - E^Y\left[\mathcal{S}_l^N(A_{l+1,n} f)\middle|\mathcal{F}_{l-1}^{UX}\right] \\ =& 0, \end{aligned} \tag{4.14}$$

for any $l \in \mathbb{N}$.

Therefore, $\left\{M_n^N(f), n \in \mathbb{N}\right\}$ is a zero-mean martingale in n with respect to Q^Y. □

4.2 Main results

4.2.1 Strong Law of Large Numbers

Now, we state and prove the Strong Law of Large Numbers for the fictitious particle system.

Theorem 4.1: *Suppose that the unnormalized filter is defined as*

$$\sigma_n(f) = E^Q\left(L_n f(X_n)\middle|\mathcal{F}_n^Y\right), \tag{4.15}$$

then for the fictitious particle system,

$$\mathcal{S}_n^N \Rightarrow \sigma_n \quad a.s. \quad \left[Q^Y\right], \tag{4.16}$$

where ($\Rightarrow$) *means weak convergence.*

Proof. Since by Lemma 2.3 and (4.6), we have $E^Y\left[\left|\mathcal{B}_n^k(f)\right|\right] < \infty$, for any $1 \le k \le N$ and any function $f \in B(E)_+$.

Hence, it follows by the Strong Law of Large Numbers for independent and identically distributed random variables that

$$\mathcal{S}_n^N(f) = \frac{1}{N}\sum_{k=1}^{N}\mathcal{B}_n^k(f) \to \sigma_n(f) \quad \text{a.s.} \quad \left[Q^Y\right],$$

for any $f \in B(E)_+$.

Using the same set A in Theorem 3.2, we can get

$$\mathcal{S}_n^N \Rightarrow \sigma_n \quad \text{a.s.} \quad \left[Q^Y\right]. \tag{4.17}$$

□

4.2.2 Central Limit Theorem

To establish the variance in the Central Limit Theorem for the fictitious particle system, we need to define the remainder function $R_l(x)$,

$$R_l(x) = \frac{\sigma_l(1)}{\sigma_{l-1}(1)} \left\{ \frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} - \left\lfloor \frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} \right\rfloor - \left(\frac{\sigma_{l-1}(1)\,\gamma_l(x)}{\sigma_l(1)} - \left\lfloor \frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} \right\rfloor \right)^2 \right\},$$

which is an artifact of our resampling procedure.

Let $M = \frac{\sigma_{l-1}(1)\alpha_l(x)}{\sigma_l(1)} - \left\lfloor \frac{\sigma_{l-1}(1)\alpha_l(x)}{\sigma_l(1)} \right\rfloor$. We can get a bound for $R_l(x)$

$$\begin{aligned} R_l(x) =& \frac{\sigma_l(1)}{\sigma_{l-1}(1)} \left\{ \frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} - \left\lfloor \frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} \right\rfloor - \left(\frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} - \left\lfloor \frac{\sigma_{l-1}(1)\,\alpha_l(x)}{\sigma_l(1)} \right\rfloor \right)^2 \right\} \\ =& \frac{\sigma_l(1)}{\sigma_{l-1}(1)} \left(M - M^2 \right) \qquad (4.18) \\ \leq& \frac{\sigma_l(1)}{4\sigma_{l-1}(1)}. \end{aligned}$$

However, by the definition of the floor function, we have $0 \leq M < 1$. Therefore

$$M - M^2 \geq 0, \qquad (4.19)$$

and

$$R_l(x) \geq 0. \qquad (4.20)$$

Now, we state and prove the Central Limit Theorem for the fictitious particle system.

Theorem 4.2: *The fictitious particle system satisfies*

$$\sqrt{N}\left(\mathcal{S}_n^N(f)-\sigma_n(f)\right)\Rightarrow\mathcal{N}\left(0,\gamma_n^R(f)\right)\quad a.s.\quad\left[Q^Y\right], \tag{4.21}$$

where

$$\gamma_n^R(f)=\sum_{l=0}^{n}\sigma_l(1)\pi_0A_{1,l-1}\left[A_l\left(A_{l+1,n}f\right)^2-\alpha_l\left(KA_{l+1,n}f\right)^2+R_l\left(KA_{l+1,n}f\right)^2\right], \tag{4.22}$$

for any f (defined in Theorem 3.3).

Proof. For the fictitious particle system, we first establish the required second moment condition

$$\begin{aligned}E^Y\left[\mathcal{B}_n^k(f)-\sigma_n(f)\right]^2&\leq 2\left\{E^Y\left[\mathcal{B}_n^k(f)\right]^2+E^Y\left[\sigma_n(f)\right]^2\right\} \\ &\leq 2\left|f\right|_\infty^2E^Y\left[\left(N_n^k\mathcal{L}_n\right)^2\right]+2\cdot\pi_0\times\pi_0\left(K_\lambda^n(f\times f)\right).\end{aligned} \tag{4.23}$$

Moreover, by (4.2),

$$\begin{aligned}E^Y\left[\left(\mathcal{L}_nN_n^k\right)^2\left|\mathcal{F}_{n-1}^U\vee\mathcal{F}_n^X\right.\right]&=\mathcal{L}_n^2\sum_{i=1}^{N_{n-1}^k}\left(E^Y\left[\left(N_n^{k,i}\right)^2\left|\mathcal{F}_{n-1}^U\vee\mathcal{F}_n^X\right.\right]-\left|E^Y\left[N_n^{k,i}\left|\mathcal{F}_{n-1}^U\vee\mathcal{F}_n^X\right.\right]\right|^2\right) \\ &+\sum_{i,j=1}^{N_{n-1}^k}E^Y\left[\mathcal{L}_nN_n^{k,i}\left|\mathcal{F}_{n-1}^U\vee\mathcal{F}_n^X\right.\right]E^Y\left[\mathcal{L}_nN_n^{k,j}\left|\mathcal{F}_{n-1}^U\vee\mathcal{F}_n^X\right.\right] \\ &=\mathcal{L}_n^2\sum_{i=1}^{N_{n-1}^k}\left\{\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}-\left[\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}\right]-\left(\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}-\left[\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}\right]\right)^2\right\} \\ &+\sum_{i,j=1}^{N_{n-1}^k}\mathcal{L}_{n-1}^2W_n^{k,i}W_n^{k,j}.\end{aligned} \tag{4.24}$$

However, by the non-negativity and boundness of the function g, there is a $c_Y>0$

such that $\sup_{x\in E, j\le n} \frac{g\left(Y_j-h(x)\right)}{g\left(Y_j\right)} \le c_Y$. Therefore, by (4.24)

$$E^Y\left[\left(\mathcal{L}_n N_n^k\right)^2\right] \le \frac{E^Y\left[\mathcal{L}_n^2 N_{n-1}^k\right]}{4} + c_Y^2 E^Y\left[\left(\mathcal{L}_{n-1} N_{n-1}^k\right)^2\right], \tag{4.25}$$

and

$$E^Y\left[N_l^k\right] = E^Y\left[N_{l-1}^k \frac{W_l^{k,i} L_{l-1}}{L_l}\right] \le c_Y \frac{L_{l-1}}{L_l} E^Y\left[N_{l-1}^k\right] \ \forall l = 1, ..., n-1. \tag{4.26}$$

Using recursion on (4.25) and (4.26), one finds that

$$E^Y\left[\left(\mathcal{L}_n N_n^k\right)^2\right] < \infty. \tag{4.27}$$

Therefore

$$E^Y\left[\mathcal{B}_n^k(f) - \sigma_n(f)\right]^2 < \infty. \tag{4.28}$$

It follows by the Central Limit Theorem for independent and identically distributed random variables that

$$\sqrt{N}\left(\mathcal{S}_n^N(f) - \sigma_n(f)\right) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \left(\mathcal{B}_n^k(f) - \sigma_n(f)\right) \Rightarrow \mathcal{N}\left(0, \gamma_n^R(f)\right). \tag{4.29}$$

Now, we calculate the random variance $\gamma_n^R(f)$. To simplify our notations, we abbreviate $M^k = M_n^{\mathcal{B}^k}(f)$. Therefore, by (4.4) and (4.8), we have

$$\begin{aligned} M^k &= \sum_{l=0}^{n} \left[\mathcal{B}_l^k(A_{l+1,n} f) - E^Y\left[\mathcal{B}_l^k(A_{l+1,n} f) \middle| \mathcal{F}_{l-1}^{UX}\right]\right] \\ &= \sum_{l=0}^{n} \left[\mathcal{B}_l^k(A_{l+1,n} f) - \mathcal{B}_{l-1}^k(A_{l,n} f)\right], \end{aligned} \tag{4.30}$$

with the convention $\mathcal{B}_1^k = \pi_0$, for any $1 \leq k \leq N$, and the martingale differences are

$$\begin{aligned}
&\mathcal{B}_l^k\left(A_{l+1,n}f\right) - \mathcal{B}_{l-1}^k\left(A_{l,n}f\right) \qquad (4.31)\\
=&\sum_{i=1}^{N_l^k} \mathcal{L}_l A_{l+1,n} f\left(\mathcal{X}_l^{k,i}\right) - \sum_{i=1}^{N_{l-1}^k} \mathcal{L}_{l-1} A_{l,n} f\left(\mathcal{X}_{l-1}^{k,i}\right)\\
=&\sum_{i=1}^{N_{l-1}^k}\left\{\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} \mathcal{L}_l A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right) - \mathcal{L}_{l-1}\frac{g\left(Y_l - h\left(\mathcal{X}_{l-1}^{k,i}\right)\right)}{g\left(Y_l\right)} K A_{l+1,n} f\left(\mathcal{X}_{l-1}^{k,i}\right)\right\}\\
=&\sum_{i=1}^{N_{l-1}^k}\left\{\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} \mathcal{L}_l A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right) - \hat{\mathcal{L}}_l^{k,i} K A_{l+1,n} f\left(\mathcal{X}_{l-1}^{k,i}\right)\right\}\\
=&\sigma_l(1)\sum_{i=1}^{N_{l-1}^k}\left\{\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right) - E^Y\left[\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right)\middle|\mathcal{F}_{l-1}^{UX}\right]\right\}.
\end{aligned}$$

Therefore, by the independence of the $\{U\}$ as well as the independence of the $\{\xi\}$

$$\begin{aligned}
&E^Y\left[\left(\mathcal{B}_l^k\left(A_{l+1,n}f\right) - \mathcal{B}_{l-1}^k\left(A_{l,n}f\right)\right)^2\middle|\mathcal{F}_{l-1}^{UX}\right] \qquad (4.32)\\
=&\ \sigma_l^2(1)\sum_{i_1,i_2=1}^{N_{l-1}^k}\left\{E^Y\left[\sum_{j_1=\mathcal{N}_l^{k,i_1}+1}^{\mathcal{N}_l^{k,i_1}+N_l^{k,i_1}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j_1}\right)\sum_{j_2=\mathcal{N}_l^{k,i_2}+1}^{\mathcal{N}_l^{k,i_2}+N_l^{k,i_2}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j_2}\right)\middle|\mathcal{F}_{l-1}^{UX}\right]\right.\\
&-\left.E^Y\left[\sum_{j_1=\mathcal{N}_l^{k,i_1}+1}^{\mathcal{N}_l^{k,i_1}+N_l^{k,i_1}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j_1}\right)\middle|\mathcal{F}_{l-1}^{UX}\right]E^Y\left[\sum_{j_2=\mathcal{N}_l^{k,i_2}+1}^{\mathcal{N}_l^{k,i_2}+N_l^{k,i_2}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j_2}\right)\middle|\mathcal{F}_{l-1}^{UX}\right]\right\}\\
=&\ \sigma_l^2(1)\sum_{i=1}^{N_{l-1}^k}\left\{E^Y\left[\left(\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right)\right)^2\middle|\mathcal{F}_{l-1}^{UX}\right] - \left(E^Y\left[\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right)\middle|\mathcal{F}_{l-1}^{UX}\right]\right)^2\right\}.
\end{aligned}$$

However, by the independence of the $\{\xi\}$ again

$$E^{Y}\left[\left(\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right)\right)^2 \middle| \mathcal{F}_{l-1}^{UX}\right] \tag{4.33}$$
$$= E^{Y}\left[N_l^{k,i}\left\{K\left(A_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right) - \left(KA_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right)\right\} + \left(N_l^{k,i} KA_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right) \middle| \mathcal{F}_{l-1}^{UX}\right]$$
$$= \frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}\left\{K\left(A_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right) - \left(KA_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right)\right\}$$
$$+ \left\{\left(\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}\right)^2 + \left(\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l} - \left\lfloor \frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l} \right\rfloor\right) - \left(\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l} - \left\lfloor \frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l} \right\rfloor\right)^2\right\}\left(KA_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right),$$

and

$$\left(E^{Y}\left[\sum_{j=\mathcal{N}_l^{k,i}+1}^{\mathcal{N}_l^{k,i}+N_l^{k,i}} A_{l+1,n} f\left(\mathcal{X}_l^{k,j}\right) \middle| \mathcal{F}_{l-1}^{UX}\right]\right)^2 = \left(E^{Y}\left[N_l^{k,i} KA_{l+1,n} f\left(\mathcal{X}_{l-1}^{k,i}\right) \middle| \mathcal{F}_{l-1}^{UX}\right]\right)^2 \tag{4.34}$$
$$= \left(\frac{\hat{\mathcal{L}}_l^{k,i}}{\mathcal{L}_l}\right)^2\left(KA_{l+1,n} f\right)^2\left(\mathcal{X}_{l-1}^{k,i}\right).$$

Combining the last three equations, we can find that

$$E^{Y}\left[\left(\mathcal{B}_l^k\left(A_{l+1,n} f\right) - \mathcal{B}_{l-1}^k\left(A_{l,n} f\right)\right)^2 \middle| \mathcal{F}_{l-1}^{UX}\right] \tag{4.35}$$
$$= \sigma_l(1)\, \mathcal{B}_{l-1}^k\left(A_l\left(A_{l+1,n} f\right)^2 - \alpha_l\left(KA_{l+1,n} f\right)^2 + R_l\left(KA_{l+1,n} f\right)^2\right).$$

By (4.4), we have

$$\begin{aligned}
\gamma_n^R(f) &= \sum_{l=0}^{n} E^Y\left[\left(\mathcal{B}_l^k(A_{l+1,n}f) - \mathcal{B}_{l-1}^k(A_{l,n}f)\right)^2\right] \qquad (4.36)\\
&= \sum_{l=0}^{n} E^Y\left[E^Y\left[\left(\mathcal{B}_l^k(A_{l+1,n}f) - \mathcal{B}_{l-1}^k(A_{l,n}f)\right)^2 \middle| \mathcal{F}_{l-1}^{UX}\right]\right]\\
&= \sum_{l=0}^{n} \sigma_l(1) E^Y\left[\mathcal{B}_{l-1}^k\left(A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2 + R_l(KA_{l+1,n}f)^2\right)\right]\\
&= \sum_{l=0}^{n} \sigma_l(1)\pi_0 A_{1,l-1}\left[A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2 + R_l(KA_{l+1,n}f)^2\right].
\end{aligned}$$

□

Remark: To make the calculations simpler, we cancel the remainder term of the variance of the Central Limit Theorem. By (4.6) and (2.44), we can get the expectation of the random variance $\gamma_n^R(f)$

$$\begin{aligned}
E^Q\left[\gamma_n^R(f)\right] &= \sum_{l=0}^{n} E^Q\left[\sigma_l(1)\mathcal{B}_{l-1}^k\left(A_l(A_{l+1,n}f)^2 - (KA_{l+1,n}f)^2\gamma_l\right)\right] \qquad (4.37)\\
&= \sum_{l=0}^{n} E^Q\left[\sigma_{l-1}(\gamma_l)\sigma_{l-1}\left[\left(K(A_{l+1,n}f)^2 - (KA_{l+1,n}f)^2\right)\gamma_l\right]\right]\\
&= \sum_{l=0}^{n} E^Q\Bigg\{\sigma_{l-1}(\gamma_l)\sigma_{l-1}\Bigg[\gamma_l(y)\Bigg(\int\int K_\lambda^{n-l}(f\times f)(z,z)K(y,\mathrm{d}z)\\
&\quad - \int\int K_\lambda^{n-l}(f\times f)(z,\zeta)K(y,\mathrm{d}z)K(y,\mathrm{d}\zeta)\Bigg)\Bigg]\Bigg\}\\
&= \sum_{l=0}^{n} \int\int (\pi_0\times\pi_0)K_\lambda^{l-1}(\mathrm{d}x,\mathrm{d}y)\left[\bar{K}_1 - K_1\right]K_\lambda^{n-l}(f\times f)(y,y)\lambda(x,y).
\end{aligned}$$

where K_1 and $\bar{K}_1$ are the combined kernels when $\lambda \equiv 1$.

Now, we calculate some particular examples to compare the expected variances of the Central Limit Theorem for both particle systems.

Suppose that the sensor function $h(x) = x$ and the probability density function

$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, then the observation variability function is

$$\lambda(x, \xi) = \exp(h(x) h(\xi)) = \exp(x\xi). \tag{4.38}$$

Let the Markov transition kernel as $K(x, \mathrm{dz}) = \sqrt{\frac{2}{\pi}} e^{-2(x-z)^2} \mathrm{dz}$, then we have

$$\begin{aligned} \bar{K}_\lambda(x, \mathrm{dz}) =& K(x, \mathrm{dz}) \lambda(x) \\ =& \sqrt{\frac{2}{\pi}} e^{-2(x-z)^2} \mathrm{dz} \cdot e^{x^2} \\ =& \sqrt{\frac{2}{\pi}} e^{-x^2+4xz-2z^2} \mathrm{dz}, \end{aligned} \tag{4.39}$$

and

$$\begin{aligned} K_\lambda(x, \xi, \mathrm{dz}, \mathrm{d}\zeta) =& K(x, \mathrm{dz}) K(\xi, \mathrm{d}\zeta) \lambda(x, \xi) \\ =& \frac{2}{\pi} e^{-2(x-z)^2} \mathrm{dz} \cdot e^{-2(\xi-\zeta)^2} \mathrm{d}\zeta \cdot e^{x\xi} \\ =& \frac{2}{\pi} e^{x\xi-2x^2+4xz-2z^2-2\xi^2+4\xi\zeta-2\zeta^2} \mathrm{dz}\mathrm{d}\zeta. \end{aligned} \tag{4.40}$$

Using $\bar{K}_\lambda^{l+1}(x, \mathrm{dz}) = \int \bar{K}_\lambda^l(\zeta, \mathrm{dz}) \bar{K}_\lambda(x, \mathrm{d}\zeta)$, we have

$$\begin{aligned} \bar{K}_\lambda^2(x, \mathrm{dz}) =& \int \bar{K}_\lambda(\zeta, \mathrm{dz}) \bar{K}_\lambda(x, \mathrm{d}\zeta) \\ =& \frac{2}{\pi} \int_{-\infty}^{\infty} e^{-x^2+4x\zeta-2\zeta^2} \mathrm{d}\zeta \cdot e^{-\zeta^2+4z\zeta-2z^2} \mathrm{dz} \\ =& \frac{2}{\pi} e^{-x^2-2z^2} \mathrm{dz} \int_{-\infty}^{\infty} e^{-3\zeta^2+4(x+z)\zeta} \mathrm{d}\zeta \\ =& \frac{2}{\sqrt{3\pi}} e^{-\frac{2}{3}z^2+\frac{8}{3}xz+\frac{1}{3}x^2} \mathrm{dz}, \end{aligned} \tag{4.41}$$

and

$$\bar{K}_\lambda^3(x,\mathrm{d}z) = \int \bar{K}_\lambda^2(\zeta,\mathrm{d}z)\,\bar{K}_\lambda(x,\mathrm{d}\zeta) \tag{4.42}$$
$$=\frac{2\sqrt{2}}{\sqrt{3}\pi}\int_{-\infty}^{\infty} e^{-\frac{2}{3}z^2+\frac{8}{3}\zeta z+\frac{1}{3}\zeta^2}\mathrm{d}z\cdot e^{-x^2+4x\zeta-2\zeta^2}\mathrm{d}\zeta$$
$$=\frac{2\sqrt{2}}{\sqrt{3}\pi}e^{-\frac{2}{3}z^2-x^2}\mathrm{d}z\int_{-\infty}^{\infty} e^{-\frac{5}{3}\zeta^2+\left(\frac{8}{3}z+4x\right)\zeta}\mathrm{d}\zeta$$
$$=\frac{2\sqrt{2}}{\sqrt{3}\pi}e^{\frac{6}{15}z^2+\frac{16}{5}xz+\frac{7}{3}x^2}\mathrm{d}z,$$

and

$$\bar{K}_\lambda^4(x,\mathrm{d}z) = \int \bar{K}_\lambda^3(\zeta,\mathrm{d}z)\,\bar{K}_\lambda(x,\mathrm{d}\zeta) \tag{4.43}$$
$$=\frac{4}{\sqrt{5}\pi}\int_{-\infty}^{\infty} e^{\frac{5}{16}z^2+\frac{16}{5}\zeta z+\frac{7}{3}\zeta^2}\mathrm{d}z\cdot e^{-x^2+4x\zeta-2\zeta^2}\mathrm{d}\zeta$$
$$=\frac{4}{\sqrt{5}\pi}e^{\frac{6}{15}z^2-x^2}\mathrm{d}z\int_{-\infty}^{\infty} e^{\frac{1}{3}\zeta^2+\left(4x+\frac{16}{5}z\right)\zeta}\mathrm{d}\zeta,$$

But the integration $\int_{-\infty}^{\infty} e^{\frac{1}{3}\xi^2+\left(4x+\frac{16}{5}z\right)\zeta}\mathrm{d}\zeta$ is infinite. By (3.35), the expected variance of the weighted particle system is

$$E^Q\left[\gamma_n^W(f)\right] = \sum_{l=0}^{n} \pi_0 \bar{K}_\lambda^{l-1}\left[\bar{K}_\lambda - K_\lambda\right]K_\lambda^{n-l}(f\times f). \tag{4.44}$$

When $n \geq 5$, the 5^{th} term of the sum is $\pi_0\bar{K}_\lambda^4\left[\bar{K}_\lambda - K_\lambda\right]K_\lambda^{n-5}(f\times f)$ and will be infinite. Therefore, the expected variance of the weighted particle system will be infinite.

Now, we calculate the expected variance of the fictitious particle system. Using

$$K_\lambda^{l+1}(x,\xi,\mathrm{d}z,\mathrm{d}\zeta) = \int\int K_\lambda^l(y,\theta,\mathrm{d}z,\mathrm{d}\zeta)\,K_\lambda(x,\xi,\mathrm{d}y,\mathrm{d}\theta), \tag{4.45}$$

we have,

$$K_\lambda^2(x,\xi,\mathrm{d}z,\mathrm{d}\zeta) = \int\int K_\lambda(y,\theta,\mathrm{d}z,\mathrm{d}\zeta)\,K_\lambda(x,\xi,\mathrm{d}y,\mathrm{d}\theta) \tag{4.46}$$

$$=\frac{2}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(y\theta-2y^2+4yz-2z^2-2\theta^2+4\theta\zeta-2\zeta^2\right)\mathrm{d}z\mathrm{d}\zeta$$

$$\cdot\frac{2}{\pi}\exp\left(x\xi-2x^2+4yx-2y^2-2\xi^2+4\xi\theta-2\theta^2\right)\mathrm{d}y\mathrm{d}\theta$$

$$=\frac{4}{\pi^2}\exp\left(-2z^2-2\zeta^2+x\xi-2x^2-2\xi^2\right)\mathrm{d}z\mathrm{d}\zeta$$

$$\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[y\theta-4y^2+4(x+z)y-4\theta^2+4(\zeta+\xi)\theta\right]\mathrm{d}y\mathrm{d}\theta.$$

The integration is calculated as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[y\theta-4y^2+4(x+z)y-4\theta^2+4(\zeta+\xi)\theta\right]\mathrm{d}y\mathrm{d}\theta \tag{4.47}$$

$$=\frac{2\pi}{\sqrt{63}}\exp\left[\frac{63}{64}(x+z)^2+\frac{63}{64}(\zeta+\xi)^2+\frac{16}{63}(x+z)(\zeta+\xi)\right]$$

$$\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\sqrt{63}}{2\pi}\exp\left\{-4\left[\theta-\left(\frac{4}{63}(x+z)+\frac{32}{63}(\xi+\zeta)\right)\right]^2-4\left[y-\left(\frac{32}{63}(x+z)+\frac{4}{63}(\xi+\zeta)\right)\right]^2\right.$$

$$\left.+\left[\theta-\left(\frac{4}{63}(x+z)+\frac{32}{63}(\xi+\zeta)\right)\right]\left[y-\left(\frac{32}{63}(x+z)+\frac{4}{63}(\xi+\zeta)\right)\right]\right\}\mathrm{d}y\mathrm{d}\theta,$$

Notice that the last two lines of (4.47) is a two-dimensional Gaussian distribution. Therefore, we have

$$K_\lambda^2(x,\xi,\mathrm{d}z,\mathrm{d}\zeta) \tag{4.48}$$

$$=\frac{4}{\pi^2}\exp\left(-2z^2-2\zeta^2+x\xi-2x^2-2\xi^2\right)\mathrm{d}z\mathrm{d}\zeta$$

$$\cdot\frac{2\pi}{\sqrt{63}}\exp\left[\frac{63}{64}(x+z)^2+\frac{63}{64}(\zeta+\xi)^2+\frac{16}{63}(x+z)(\zeta+\xi)\right]$$

$$=\frac{8}{\sqrt{63}\pi}\exp\left[-\frac{62}{63}\left(x^2+\xi^2+\zeta^2+z^2\right)+\frac{79}{63}x\xi+\frac{128}{63}(xz+\zeta\xi)+\frac{16}{62}(x\zeta+z\zeta+z\xi)\right]\mathrm{d}z\mathrm{d}\zeta.$$

The idea of calculating K_λ^2 is that: we try to divide the integration into two parts. The first part is the integration of the probability density function of a Gaussian distributed random vector. The second part is the integration of some remainder terms. By the same technique, we can calculate K_λ^3 and K_λ^4, respectively,

$$
\begin{aligned}
& K_\lambda^3(x,\xi,\mathrm{dz},\mathrm{d}\zeta) \\
=& \frac{16}{\sqrt{63}\pi^2}\exp\left(-\frac{62}{63}z^2 - -\frac{62}{63}\zeta^2 + \frac{16}{63}z\zeta + x\xi - 2x^2 - 2\xi^2\right)\cdot \\
& \frac{2\sqrt{63}\pi}{\sqrt{2145}}\exp\left[\frac{188}{2145}\left(\frac{128}{63}z + \frac{16}{63}\zeta + 4x\right)^2 + \frac{188}{2145}\left(\frac{128}{63}\zeta + \frac{16}{63}z + 4\xi\right)^2\right. \\
& \left. + \frac{79}{2145}\left(\frac{128}{63}z + \frac{16}{63}\zeta + 4x\right)\left(\frac{128}{63}\zeta + \frac{16}{63}z + 4\xi\right)\right]\mathrm{dzd}\zeta \\
=& \frac{32}{\sqrt{2145}\pi}\exp\left[-\frac{1282}{2145}\left(x^2 + \xi^2 + \zeta^2 + z^2\right) + \frac{1264}{2145}z\zeta + \frac{3409}{2145}x\xi\right. \\
& \left. + \frac{3136}{2145}(xz + \zeta\xi) + \frac{1024}{2145}(x\zeta + z\xi)\right]\mathrm{dzd}\zeta.
\end{aligned}
$$

and

$$
\begin{aligned}
& K_\lambda^4(x,\xi,\mathrm{dz},\mathrm{d}\zeta) \\
=& \frac{64}{\sqrt{2145}\pi^2}\exp\left(-\frac{1282}{2145}z^2 - -\frac{1282}{2145}\zeta^2 + \frac{1264}{2145}z\zeta + x\xi - 2x^2 - 2\xi^2\right)\mathrm{dzd}\zeta\cdot \\
& \frac{4290\pi}{\sqrt{112567455}}\exp\left[\frac{796}{7497}\left(\frac{3136}{2145}z + \frac{1024}{2145}\zeta + 4x\right)^2 + \frac{796}{7497}\left(\frac{3136}{2145}\zeta + \frac{1024}{2145}z + 4\xi\right)^2\right. \\
& \left. + \frac{487}{7497}\left(\frac{3136}{2145}z + \frac{1024}{2145}\zeta + 4x\right)\left(\frac{3136}{2145}\zeta + \frac{1024}{2145}z + 4\xi\right)\right]\cdot \\
=& \frac{128}{\sqrt{52479}\pi}\exp\left[-\frac{2258}{7497}\left(x^2 + \xi^2 + \zeta^2 + z^2\right) + \frac{6125}{7497}z\zeta + \frac{15289}{7497}x\xi\right. \\
& \left. + \frac{10240}{7497}(xz + \zeta\xi) + \frac{5888}{7497}(x\zeta + z\xi)\right]\mathrm{dzd}\zeta.
\end{aligned}
$$

By (4.37), the expected variance of the fictitious particle system is

$$E^{Q}\left[\gamma_n^R(f)\right] = \sum_{l=0}^{n} \int \int (\pi_0 \times \pi_0) K_\lambda^{l-1}(\mathrm{d}x, \mathrm{d}y) [\bar{K}_1 - K_1] K_\lambda^{n-l}(f \times f)(y, y) \lambda(x, y). \tag{4.49}$$

If $n = 5$, the expected variance of the fictitious particle system is

$$\sum_{l=0}^{5} \int \int (\pi_0 \times \pi_0) K_\lambda^{l-1}(\mathrm{d}x, \mathrm{d}y) [\bar{K}_1 - K_1] K_\lambda^{5-l}(f \times f)(y, y) \lambda(x, y). \tag{4.50}$$

By the previous calculations, we can see that K_λ^i is finite when $i \leq 4$. Therefore, the expected variance of the fictitious particle system is still finite when $n = 5$. However, as we have showed before, the expected variance of the weighted particle system is infinite when $n = 5$. Hence, we have showed that, under these conditions, the expected variance of the fictitious particle system is less than that of the weighted particle system.

Chapter 5

Summary and Future Work

In the thesis, we introduce two particle systems: weighted particle system and fictitious particle system, to approximate the unnormalized filter in the filtering problem. We prove the Strong Law of Large Numbers and the Central Limit Theorem for both systems, and calculate the variances as well as the expected variances of the Central Limit Theorem. Under some particular examples, we have showed that the expected variance of the fictitious particle system is much less than that of the classical classical weighted particle system.

In the previous calculations, we always eliminate the remainder term of the variance of the fictitious particle system. But in reality, the remainder term exists since the resampling process will create a lot of noise. In the future, it will be of great interest to find some methods to eliminate the remainder term. In addition, we could investigate more examples proving that the expected variance of the fictitious particle system is less than that of the weighted particle system.

In this thesis, we only discuss and analyze the fictitious particle system, which is mathematically simpler but can not be implemented on a computer. We know that the fictitious particle system is a coupling to the new resampled particle system introduced in Chapter 1. In the future, we could analyze the computer-implementable resampled particle system with the results of the thesis.

Bibliography

[1] D. Blount, M .A . Kouritzin (2010): On Convergence Determining and Separating Classes of Functions, *Stochastic Process. Appl.* **120**, 1898-1907

[2] P. Del Moral, M. Ledoux (2000): Convergence of Empirical Processes for Interacting Particle Systems with Application to Nonlinear Filtering, *Journal of Theoretical Probability* **13**, 225-257

[3] D .J . Ballatyne, J. Hailes, M .A . Kouritzin, Hongwei Long, J .H . Wiersma (2003): A Hybrid Weighted Interacting Particle Filter for Multi-target Tracking, *Proceedings of SPIE* **5096**, 244-255

[4] P. Del Moral, J. Jacod (2001): Interacting Particle Filtering with Discrete Observations, *In Sequential Monte Carlo Methods in Practice*, 43-75

[5] P. Del Moral, L. Miclo (2001): Branching and Interacting Particle Systems: Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering

[6] D. Crisan, P. Del Moral, T. J. Lyons (1999): Non Linear Filtering using Branching and Interacting Particle System, *Markov Processes and Related Fields* **5(3)**, 293-319

[7] P. Del Moral, M. A. Kouritzin, L. Miclo (2001): On a Class of Discrete Generation Interacting Particle System, *Electronic Journal of Probability* **6**, 1-26

[8] P. Del Moral (1996): Non-linear Filtering: Interacting Particle Resolution, *Markov Processes and Related Fields* **2(4)**, 555-581

[9] P. Del Moral (1998): Measure Valued Processes and Interacting Particle Systems. Application to Non Linear Filtering Problems, *The Annals of Applied Probability* **8(2)**, 438-495

[10] P. Del Moral, A. Guionnet(1998): Large Deviations for Interacting Particle Systems. Application to Non Linear Filtering Problems, *Stochastic Process. Appl.* **78**, 69-95

[11] P. Del Moral, A. Guionnet(1999): Central Limit Theorem for Non Linear Filtering and Interacting Particle Systems, *The Annals of Applied Probability* **9(2)**, 275-297

[12] Aad W. van der Vaart, J. A. Wellner (1996): *Weak Convergence and Empirical Processes with Application to Statistics*, Springer Series in Statistics, ISBN: 0387946403

[13] R. L. Dobrushin (1956): Central Limit Theorem for Nonstationary Markov Chains *I*, *Theory of Probility and its Applications* **1**, 66-80

[14] R. M. Dudley (2004): *Real Analysis and Probability*, Cambridge Mathematical Textbook, ISBN: 978432503859

[15] A. N. Shiryaev (1996): *Probability*, Springer Series in Mathematics, ISBN: 0387945490

[16] D. Williams (2008): *Probability with Martingales*, Cambridge Mathematical Textbooks, ISBN: 9787506292511

[17] P. E. Protter (1990): *Stochastic Integration and Differential Equations*, Springer Berlin Heidelberg New York, ISBN: 9787506291972

[18] P. Billingsley (1995): *Probability and Measure*, Wiley Series in Probability and Matheamtical Statistics, ISBN: 9787506282765

[19] B. Oksendal (2005): *Stochastic Differential Equations*, Springer Berlin Heidelberg New York, ISBN: 9787506273084

[20] S. M. Ross (2007): *Introduction to Probability Models*, Elsevier, ISBN: 9787115160232