2

Particle Filters - A Theoretical Perspective

Dan Crisan

2.1 Introduction

The purpose of this chapter is to present a rigorous mathematical treatment of the convergence of particle filters. In general, we follow the notation and settings suggested by the editors, any extra notation being defined in the next section. Section 2.3.1 contains the main results of the paper: Theorems 2.3.1 and 2.3.2 provide *necessary and sufficient* conditions for the convergence of the particle filter to the posterior distribution of the signal. As an application of these results, we prove the convergence of a certain class of particle filters. This class includes several known filters (such as those presented in (Carpenter, Clifford and Fearnhead 1999b, Crisan, Del Moral and Lyons 1999, Gordon et al. 1993), but is by no means the most general one. Finally, we discuss some of the issues that are relevant in applications and which arise from the theoretical analysis of these methods.

We have tried to make the material as self-contained as possible, and therefore we have included in an appendix some useful definitions and results regarding conditional expectations/probabilities. We also include an elementary proof of the recurrence formula for the conditional distribution of the signal.

2.2 Notation and terminology

Let $\mathbb{R}^d$ be the d-dimensional Euclidean space and $\mathcal{B}\left(\mathbb{R}^d\right)$ be the σ-algebra of Borel subsets of $\mathbb{R}^d$. Hereafter we will use the following sets of functions:

$B\left(\mathbb{R}^d\right)$ - the set of bounded $\mathcal{B}\left(\mathbb{R}^d\right)$-measurable functions defined on $\mathbb{R}^d$.
$\mathcal{C}_b(\mathbb{R}^d)$ - the set of bounded continuous functions defined on $\mathbb{R}^d$.
$\mathcal{C}_k(\mathbb{R}^d)$ - the set of compactly supported continuous functions defined on

A. Doucet et al. (eds.), *Sequential Monte Carlo Methods in Practice*

$\mathbb{R}^d$.
If $f \in \mathcal{C}_b(\mathbb{R}^d)$ we denote by $||f||$ the supremum norm of f

$$||f|| \triangleq \sup_{x \in \mathbb{R}^d} |f(x)|.$$

We also consider the following sets of measures:

$\mathcal{M}_F(\mathbb{R}^d)$ - the set of finite measures over $\mathcal{B}(\mathbb{R}^d)$.
$\mathcal{P}(\mathbb{R}^d)$- the set of probability measures over $\mathcal{B}(\mathbb{R}^d)$.

If $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ (or $\mu \in \mathcal{P}(\mathbb{R}^d)$) and $f \in B(\mathbb{R}^{n_x})$, the integral of f with respect to μ is denoted by μf,

$$\mu f \triangleq \int_{\mathbb{R}^d} f(x)\, \mu(dx).$$

We endow $\mathcal{M}_F(\mathbb{R}^d)$, respectively, $\mathcal{P}(\mathbb{R}^d)$, with the weak topology. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of finite measures, then we say that μ_n converges (weakly) or in the weak topology to $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ and write $\lim_{n\to\infty} \mu_n = \mu$ if

$$\lim_{n\to\infty} \mu_n f = \mu f, \text{ for all } f \in \mathcal{C}_b(\mathbb{R}^d),$$

and we have the same definition if $(\mu_n)_{n=1}^{\infty}$ is a sequence of probability measures. Also, we denote the Dirac measure concentrated at $a \in \mathbb{R}^d$ by δ_a and the constant function 1 by $\bar{\mathbf{1}}$.

2.2.1 Markov chains and transition kernels

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X = \{X_t, t \in \mathbb{N}\}$ be a stochastic process defined on $(\Omega, \mathcal{F}, P)$ with values in $\mathbb{R}^{n_x}$. Let $\mathcal{F}_t^X$ be the σ-algebra generated by the process, i.e., $\mathcal{F}_t^X \triangleq \sigma(X_s, s \in [0,t])$. Then X is a *Markov chain* if, for all $t \in \mathbb{N}$ and $A \in \mathcal{B}(\mathbb{R}^{n_x})$,

$$P\left(X_{t+1} \in A \,\middle|\, \mathcal{F}_t^X\right) = P\left(X_{t+1} \in A \,|\, X_t\right). \tag{2.1}$$

The transition kernel of the Markov chain X is the function $K_t(\cdot,\cdot)$ defined on $\mathbb{R}^{n_x} \times \mathcal{B}(\mathbb{R}^{n_x})$ such that, for all $t \in \mathbb{N}$ and $x \in \mathbb{R}^{n_x}$,

$$K_t(x, A) = P\left(X_{t+1} \in A \,|\, X_t = x\right). \tag{2.2}$$

The transition kernel K_t has the following properties:

- $K_t(x,\cdot)$ is a probability measure on $\mathbb{R}^{n_x}$, for all $t \in \mathbb{N}$ and $x \in \mathbb{R}^{n_x}$.
- $K_t(\cdot, A) \in B(\mathbb{R}^{n_x})$, for all $t \in \mathbb{N}$ and $A \in \mathcal{B}(\mathbb{R}^{n_x})$.

The distribution of X is uniquely determined by its initial distribution and its transition kernel. Let us denote by q_t the distribution of the random variable X_t,

$$q_t(A) \triangleq P(X_t \in A).$$

Then, from (2.2), we deduce that q_t satisfies the recurrence formula $q_{t+1} = q_t K_t$, where $q_t K_t$ is the measure defined as

$$(q_t K_t)(A) \triangleq \int_{\mathbb{R}^{n_x}} K_t(x, A)\, q_t(dx).$$

Hence, $q_t = q_0 K_0 K_1 ... K_{t-1}$. We say that the transition kernel K_t satisfies the Feller property if, for all $t > 0$, the function $K_t f : \mathbb{R}^{n_x} \to \mathbb{R}$ defined as

$$K_t f(x) \triangleq \int_{\mathbb{R}^{n_x}} f(y) K_t(x, dy)$$

is continuous for every $f \in \mathcal{C}_b(\mathbb{R}^d)$. If K_t has the Feller property, then $K_t f \in \mathcal{C}_b(\mathbb{R}^d)$ for all $f \in \mathcal{C}_b(\mathbb{R}^d)$.

2.2.2 The filtering problem

Let $X = \{X_t, t \in \mathbb{N}\}$ be an $\mathbb{R}^{n_x}$-valued Markov process (called the signal process) with a Feller transition kernel $K_t(x, dy)$. Let also $Y = \{Y_t, t \in \mathbb{N}\}$ be an $\mathbb{R}^{n_y}$-valued stochastic process (called the observation process) defined as

$$Y_t \triangleq h(t, X_t) + W_t, \quad t > 0, \tag{2.3}$$

and $Y_0 = 0$. In (2.3), $h : \mathbb{N} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ is a Borel measurable function with the property that $h(t, \cdot)$ is continuous on $\mathbb{R}^{n_x}$ for all $t \in \mathbb{N}$ and, for all $t > 0$, $W_t : \Omega \to \mathbb{R}^{n_y}$ are independent random vectors and their laws are absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^{n_y}$. We denote by $g(t, \cdot)$ the density of W_t with respect to λ, and we further assume that $g(t, \cdot)$ is bounded and continuous. The filtering problem consists of computing the conditional distribution of the signal given the $\sigma-$algebra generated by the observation process from time 0 to the current time; in other words, one is interested in computing the (random) probability measure π_t, where

$$\pi_t(A) \triangleq P(X_t \in A \,|\sigma(Y_{0:t})), \quad \pi_t f = \mathbb{E}[f(X_t)\,|\sigma(Y_{0:t})] \tag{2.4}$$

for all $f \in B(\mathbb{R}^{n_x})$ and $A \in \mathcal{B}(\mathbb{R}^{n_x})$, where $Y_{0:t} \triangleq (Y_0, Y_1, \ldots, Y_t)$. Then $\pi_t = \pi_t^{Y_{0\ t}}$, where

$$\pi_t^{\mathbf{y}_{0\ t}}(A) \triangleq P(X_t \in A \,|Y_{0:t} = \mathbf{y}_{0:t}), \quad \pi_t^{\mathbf{y}_{0\ t}} f = \mathbb{E}[f(X_t)\,|Y_{0:t} = \mathbf{y}_{0:t}] \tag{2.5}$$

and $\mathbf{y}_{0:t} \triangleq (\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_t) \in (\mathbb{R}^{n_y})^{t+1}$. Notice that, while π_t is a random probability measure, $\pi_t^{\mathbf{y}_{0.t}}$ is a deterministic probability measure. We also introduce p_t and $p_t^{\mathbf{y}_{0\ t-1}}$, the predicted conditional probability measures for $t > 0$, where $p_t = p_t^{Y_{0\ t-1}}$ and

$$p_t^{\mathbf{y}_{0\ t-1}}(A) = P(X_t \in A\,|Y_{0:t-1} = \mathbf{y}_{0:t-1}), \ p_t^{\mathbf{y}_{0\ t-1}} f = \mathbb{E}[f(X_t)\,|Y_{0:t-1} = \mathbf{y}_{0:t-1}].$$

In the appendix we include a short introduction to conditional probabilities/expectations.

We have the following recurrence relations for π_t, respectively, $\pi_t^{\mathbf{y}_0\,t}$:

$$\begin{cases} \frac{d\pi_t}{dp_t} &= \frac{g_t^{Y_t}}{\int g_t^{Y_t}(x)p_t(d\mathbf{x})} \\ p_{t+1} &= \pi_t K_t \end{cases}, \qquad \begin{cases} \frac{d\pi_t^{\mathbf{y}_0\,t}}{dp_t^{\mathbf{y}_0\,t-1}} &= \frac{g_t^{\mathbf{y}_t}}{\int g_t^{\mathbf{y}_t}(x)p_t^{\mathbf{y}_0\,t-1}(d\mathbf{x})} \\ p_{t+1}^{\mathbf{y}_0\cdot t} &= \pi_t^{\mathbf{y}_0\,t} K_t \end{cases}, \quad (2.6)$$

where $g_t^{\mathbf{y}_t} \in \mathcal{C}_b(\mathbb{R}^{n_x})$ is defined by $g_t^{\mathbf{y}_t} := g(\mathbf{y}_t - h(t,\cdot))$ and, since $Y_0 = 0$, π_0 is the law of X. An elementary proof of (2.6) is included in the appendix. In the general case there is no closed solution for the system (2.6). In section 2.4.1 we present a generic class of particle filters that can be used to solve (2.6) numerically.

2.2.3 Convergence of measure-valued random variables

Essentially, the result of any algorithm to solve the filtering problem based on a sequential Monte Carlo method is a random measure which approximates π_t. In order to establish whether the algorithm is good or bad, we would need to define in what way a random measure or, more precisely, a sequence of random measures, can approximate another measure.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mu^n)_{n=1}^\infty$ be a sequence of random measures, $\mu^n : \Omega \to \mathcal{M}_F(\mathbb{R}^d)$ and $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ be a deterministic finite measure. As we shall see, in the case of approximations obtained using particle filters, n represents the number of particles used in the approximating particle system. Below we study two types of convergence:

1. $\lim_{n\to\infty} \mathbb{E}[|\mu^n f - \mu f|] = 0$ for all $f \in \mathcal{C}_b(\mathbb{R}^d)$;
2. $\lim_{n\to\infty} \mu^n = \mu$, P–a.s..

We will denote the first type of convergence by $\mathrm{e}\lim_{n\to\infty} \mu^n = \mu$. If there exists an integrable random variable $w : \Omega \to \mathbb{R}$ such that $\mu^n \bar{\mathbf{1}} \leq w$ for all n, then $\lim_{n\to\infty} \mu^n = \mu$, P–a.s., implies $\mathrm{e}\lim_{n\to\infty} \mu^n = \mu$ by then Dominated Convergence Theorem. The extra condition is satisfied if $(\mu^n)_{n=1}^\infty$ is a sequence of random *probability* measures because, in this case, $\mu^n \bar{\mathbf{1}} = 1$ for all n.

Remark 1. *If* $\mathrm{e}\lim_{n\to\infty} \mu^n = \mu$, *then sequences* $n(m)$ *exist such that* $\lim_{m\to\infty} \mu^{n(m)} = \mu$, P*–a.s..*

Proof. Since $\mathbb{R}^d$ is a locally compact separable metric space, there exists a countable set $\mathcal{M}$ which is dense in $\mathcal{C}_k(\mathbb{R}^d)$. Then $\mathcal{M} \cup \{\bar{\mathbf{1}}\}$ is a convergence determining set, i.e., if ν_n, $n = 1, 2...$, and ν are finite measures and $\lim_{n\to\infty} \nu_n f = \nu f$ for all $f \in \mathcal{M} \cup \{\bar{\mathbf{1}}\}$, then $\lim_{n\to\infty} \nu_n = \nu$. Since $\lim_{n\to\infty} \mathbb{E}[|\mu^n f - \mu f|] = 0$ for all $f \in \mathcal{M} \cup \{\bar{\mathbf{1}}\}$ and $\mathcal{M}$ is countable, one can find a subsequence $n(m)$ such that, with probability 1, $\lim_{m\to\infty} \mu^{n(m)} f = \mu f$ for all $f \in \mathcal{M} \cup \{\bar{\mathbf{1}}\}$. □

Moreover, if the rate of convergence for $\mathbb{E}\left[|\mu^n f - \mu f|\right]$ is known, then these sequences can be explicitly specified. For instance, if for all $f \in \mathcal{M} \cup \{\bar{\mathbf{1}}\}$, $\mathbb{E}\left[|\mu^n f - \mu f|\right] \leq c_f n^{-\frac{1}{2}}$, for all $n > 0$, then using a Borel-Cantelli argument one can prove that $\lim_{m\to\infty} \mu^{m^4} = \mu$, P–a.s..

If $\mathcal{M} \in \mathcal{C}_k(\mathbb{R}^d)$ is the set defined above, then the following

$$d_{\mathcal{M}}(\mu,\nu) \triangleq |\mu\bar{\mathbf{1}} - \nu\bar{\mathbf{1}}| + \sum_{f_k \in \mathcal{M}} \frac{|\mu f_k - \nu f_k|}{2^k \|f_k\|}, \tag{2.7}$$

is a distance on $\mathcal{M}_F\left(\mathbb{R}^d\right)$ (or $\mathcal{P}\left(\mathbb{R}^d\right)$), which generates the weak topology

$$\lim_{n\to\infty} \nu_n = \nu \Longleftrightarrow \lim_{n\to\infty} d_{\mathcal{M}}(\nu_n,\nu) = 0.$$

Using $d_{\mathcal{M}}$, the almost sure convergence **2.** is equivalent to

2'. $\lim_{n\to\infty} d_{\mathcal{M}}(\mu^n,\mu) = 0$, P–a.s..

Also, if there exists an integrable random variable $w : \Omega \to \mathbb{R}$ such that $\mu^n\bar{\mathbf{1}} \leq w$ for all n, then, similarly, **1.** implies

1'. $\lim_{n\to\infty} \mathbb{E}\left[d_{\mathcal{M}}(\mu^n,\mu)\right] = 0$.

However, a stronger condition (such as tightness) would be needed in order to ensure that **1.** is equivalent to **1'.**

The same definitions are valid for the case when the limiting measure μ is not deterministic, but random, $\mu : \Omega \to \mathcal{M}_F\left(\mathbb{R}^d\right)$. If μ is random we have, just as before, that **2.** is equivalent to **2'.**. Also, **2.** implies **1.** and **1.** implies **1'.**, under the condition that there exists an integrable random variable $w : \Omega \to \mathbb{R}$ such that $\mu^n\bar{\mathbf{1}} \leq w$ and $\mu\bar{\mathbf{1}} \leq w$ for all n.

The limiting measures that we are interested in are $\pi_t^{\mathbf{y}_{0\ t}}$ and π_t, hence deterministic and, random *probability* measures.

2.3 Convergence theorems

2.3.1 *The fixed observation case*

We shall look first at the case in which the observation process has an arbitrary but fixed value $\mathbf{y}_{0:T}$, where T is a finite but large time horizon. We assume that the recurrence formula (2.6) for $\pi_t^{\mathbf{y}_{0\ t}}$ – the conditional distribution of the signal given the event $\{Y_{0:t} = \mathbf{y}_{0:t}\}$ – holds true for this particular value for all $0 \leq t \leq T$ (remember that (2.6) is valid $P_{Y_{0:t}}$-almost surely). Then (2.6) requires the use of an intermediate step, the predicted conditional probability measure $p_t^{\mathbf{y}_{0\ t-1}}$:

$$\pi_{t-1}^{\mathbf{y}_{0:t-1}} \longrightarrow p_t^{\mathbf{y}_{0\ t-1}} \longrightarrow \pi_t^{\mathbf{y}_{0\ t}}.$$

Therefore it is natural to study algorithms that provide recursive approximations for $\pi_t^{\mathbf{y}_0\,t}$ by using intermediate approximations for $p_t^{\mathbf{y}_0\,t-1}$. We shall denote by $(\pi_t^n)_{n=1}^{\infty}$ and $(p_t^n)_{n=1}^{\infty}$ the approximating sequence for $\pi_t^{\mathbf{y}_0\,t}$ and $p_t^{\mathbf{y}_0\,t-1}$ and assume that π_t^n and p_t^n are random measures, not necessarily probabilities, such that $p_t^n \neq 0$, $\pi_t^n \neq 0$ (none of them is trivial) and $p_t^n g_t^{\mathbf{y}_t} > 0$ for all $n > 0$, $0 \leq t \leq T$. Let also $\bar{\pi}_t^n$ be defined as a (random) probability measure absolutely continuous with respect to p_t^n for $t \in \mathbb{N}$ and $n \geq 1$ such that

$$\frac{d\bar{\pi}_t^n}{dp_t^n} = \frac{g_t^{\mathbf{y}_t}}{p_t^n g_t^{\mathbf{y}_t}}. \tag{2.8}$$

The following theorems give us *necessary and sufficient* conditions for the convergence of p_t^n and π_t^n to $p_t^{\mathbf{y}_0\,t-1}$ and, respectively, $\pi_t^{\mathbf{y}_0\,t}$. In order to simplify notation, for the remainder of this subsection we suppress the dependence on $\mathbf{y}_{0:t}$ and we denote $\pi_t^{\mathbf{y}_0\,t}$ by π_t, $p_t^{\mathbf{y}_0\,t-1}$ by p_t, and $g_t^{\mathbf{y}_t}$ by g_t, but always keep in mind that the observation process is a given fixed path $\mathbf{y}_{0:T}$

Theorem 2.3.1. *The sequences p_t^n, π_t^n converge to p_t, respectively, to π_t with convergence taken to be of type **1.** if and only if the following three conditions are satisfied:*
a1. *For all $f \in C_b(\mathbb{R}^{n_x})$, $\lim_{n\to\infty} \mathbb{E}\left[|\pi_0^n f - \pi_0 f|\right] = 0$*
b1. *For all $f \in C_b(\mathbb{R}^{n_x})$, $\lim_{n\to\infty} \mathbb{E}\left[\left|p_t^n f - \pi_{t-1}^n K_{t-1} f\right|\right] = 0$*
c1. *For all $f \in C_b(\mathbb{R}^{n_x})$, $\lim_{n\to\infty} \mathbb{E}\left[|\pi_t^n f - \bar{\pi}_t^n f|\right] = 0$.*

Proof. The sufficiency part is proved by mathematical induction. The theorem is true for $n = 0$ from the condition **a1.** We need to show that, if p_{t-1}^n, π_{t-1}^n converge to p_{t-1}, respectively, to π_{t-1} then p_t^n, π_t^n converge to p_t, respectively, to π_t. Since $p_t = \pi_{t-1}K_{t-1}$, we have, for all $f \in C_b(\mathbb{R}^{n_x})$,

$$|p_t^n f - p_t f| \leq \left|p_t^n f - \pi_{t-1}^n K_{t-1} f\right| + \left|\pi_{t-1}^n K_{t-1} f - \pi_{t-1} K_{t-1} f\right|. \tag{2.9}$$

By taking the expectation of both sides of equation (2.9) we obtain the following $\lim_{n\to\infty} \mathbb{E}\left[|p_t^n f - p_t f|\right] = 0$, since the expected value of the first term on the right hand side of (2.9) converges to 0 from **b1.** and the expected value of the second one converges to 0 from the induction hypothesis as $K_{t-1}\ f \in C_b(\mathbb{R}^{n_x})$ for all $f \in C_b(\mathbb{R}^{n_x})$ – the Feller property of the transition kernel.
We then use

$$\begin{aligned}
|\bar{\pi}_t^n f - \pi_t f| &= \left|\frac{p_t^n f g_t}{p_t^n g_t} - \frac{p_t f g_t}{p_t g_t}\right| \\
&\leq \left|\frac{p_t^n f g_t}{p_t^n g_t} - \frac{p_t^n f g_t}{p_t g_t}\right| + \left|\frac{p_t^n f g_t}{p_t g_t} - \frac{p_t f g_t}{p_t g_t}\right| \\
&\leq \frac{||f||}{p_t g_t} |p_t^n g_t - p_t g_t| + \frac{1}{p_t g_t} |p_t^n f g_t - p_t f g_t|, \qquad (2.10)
\end{aligned}$$

therefore

$$\mathbb{E}\left[|\bar{\pi}_t^n f - \pi_t f|\right] \le \frac{||f||}{p_t g_t}\mathbb{E}\left[|p_t^n g_t - p_t g_t|\right] + \frac{1}{p_t g_t}\mathbb{E}\left[|p_t^n f g_t - p_t f g_t|\right], \quad (2.11)$$

and both terms on the right hand side of (2.11) converge to 0. Finally,

$$|\pi_t^n f - \pi_t f| \le |\pi_t^n f - \bar{\pi}_t^n f| + |\bar{\pi}_t^n f - \pi_t f|. \quad (2.12)$$

As the expected value of the first term on the right hand side of (2.12) converges to 0 using **c1.** , and the expected value of the second term converges to 0 using (2.11), we find that $\lim_{n\to\infty} \mathbb{E}\left[|\pi_t^n f - \pi_t f|\right] = 0$.

We now prove the necessity part. Assume that for all $t \ge 0$ and for all $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, $\lim_{n\to\infty} \mathbb{E}\left[|p_t^n f - p_t f|\right] = 0$ and $\lim_{n\to\infty} \mathbb{E}\left[|\pi_t^n f - \pi_t f|\right] = 0$ (this, in particular implies **a1.**). From (2.11), we see that $\lim_{n\to\infty} \mathbb{E}\left[|\pi_t f - \bar{\pi}_t^n f|\right] = 0$, and since

$$\mathbb{E}\left[|\pi_t^n f - \bar{\pi}_t^n f|\right] \le \mathbb{E}\left[|\pi_t^n f - \pi_t f|\right] + \mathbb{E}\left[|\pi_t f - \bar{\pi}_t^n f|\right], \quad (2.13)$$

we obtain **c1.**. Finally, since $p_t = \pi_{t-1} K_{t-1}$, we have, for all $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$,

$$\mathbb{E}\left[\left|p_t^n f - \pi_{t-1}^n K_{t-1} f\right|\right] \le \mathbb{E}\left[|p_t^n f - p_t f|\right] + \mathbb{E}\left[\left|\pi_{t-1} K_{t-1} f - \pi_{t-1}^n K_{t-1} f\right|\right], \quad (2.14)$$

which implies **b1.**. □

The corresponding theorem for the almost sure convergence of p_t^n, π_t^n to p_t, respectively, to π_t holds true.

Theorem 2.3.2. *The sequences p_t^n, π_t^n converge almost surely to p_t and π_t, i.e. with convergence of type 2., if and only if the following three conditions are satisfied;*
a2. $\lim_{n\to\infty} \pi_0^n = \pi_0$, *$P-$a.s.*
b2. $\lim_{n\to\infty} d_{\mathcal{M}}\left(p_t^n, \pi_{t-1}^n K_{t-1}\right) = 0$, *$P-$a.s.*
c2. $\lim_{n\to\infty} d_{\mathcal{M}}\left(\pi_t^n, \bar{\pi}_t^n\right) = 0$, *$P-$a.s.*.

Proof. The sufficiency part of the theorem is proved as above by mathematical induction using inequalities (2.9), (2.10) and (2.12). The necessity part of the theorem results from the following argument. Assume that for all $t \ge 0$ p_t^n and π_t^n converge almost surely to p_t and, respectively, to π_t. This implies that $\pi_{t-1}^n K_{t-1}$ converges almost surely to p_t $(= \pi_{t-1} K_{t-1})$ and, using (2.10), that $\bar{\pi}_t^n$ converges almost surely to π_t. Hence, almost surely $\lim_{n\to\infty} d_{\mathcal{M}}\left(p_t^n, p_t\right) = 0$, $\lim_{n\to\infty} d_{\mathcal{M}}\left(\pi_t^n, \pi_t\right) = 0$, $\lim_{n\to\infty} d_{\mathcal{M}}\left(\pi_{t-1}^n K_{t-1}, p_t\right) = 0$ and $\lim_{n\to\infty} d_{\mathcal{M}}\left(\bar{\pi}_t^n, \pi_t\right) = 0$, where $d_{\mathcal{M}}$ is the distance defined in (2.7). Finally, using the triangle inequalities

$$\begin{aligned} d_{\mathcal{M}}\left(p_t^n, \pi_{t-1}^n K_{t-1}\right) &\le d_{\mathcal{M}}\left(p_t^n, p_t\right) + d_{\mathcal{M}}\left(p_t, \pi_{t-1}^n K_{t-1}\right) \\ d_{\mathcal{M}}\left(\pi_t^n, \bar{\pi}_t^n\right) &\le d_{\mathcal{M}}\left(\pi_t^n, \pi_t\right) + d_{\mathcal{M}}\left(\pi_t, \bar{\pi}_t^n\right), \end{aligned}$$

we get **b2.** and **c2.**. □

Remark 2. *Theorems 2.3.1 and 2.3.2 are very natural. They say that, we obtain approximations of p_t and π_t for all $t \geq 0$. if and only if we have to start from an approximation of π_0, and then 'follow closely' the recurrence formula (2.6) for p_t and π_t.*

Now, the question arises whether we can lift the results to the case when the observation process is random and not just fixed to a given observation path.

2.3.2 The random observation case

In the previous section both the converging sequences and the limiting measures depend on the fixed value of the observation. Let us make this explicit by writing that

$$\lim_{n\to\infty} p_t^{n,\mathbf{y}_{0:t-1}} = p_t^{\mathbf{y}_{0:t-1}}, \qquad \lim_{n\to\infty} \pi_t^{n,\mathbf{y}_{0:t}} = \pi_t^{\mathbf{y}_{0:t}}, \tag{2.15}$$

where the limits in (2.15) are either of type **1.** or of type **2.**. Since $p_t = p_t^{Y_{0:t-1}}$ and $\pi_t = \pi_t^{Y_{0:t}}$, we expect that

$$\lim_{n\to\infty} p_t^{n,Y_{0:t-1}} = p_t, \qquad \lim_{n\to\infty} \pi_t^{n,Y_{0:t}} = \pi_t.$$

Let us look first at the convergence of type **1.**. We have

$$\begin{aligned}
\mathbb{E}\left[\left|p_t^{n,Y_{0:t-1}} f - p_t f\right|\right] &= \int_{(\mathbb{R}^{n_y})^t} \mathbb{E}\left[\left|p_t^{n,\mathbf{y}_{0:t-1}} f - p_t^{\mathbf{y}_{0:t}} f\right|\right] P_{Y_{0:t-1}}\left(d\mathbf{y}_{0:t-1}\right) \\
\mathbb{E}\left[\left|\pi_t^{n,Y_{0:t}} f - p_t f\right|\right] &= \int_{(\mathbb{R}^{n_y})^{t+1}} \mathbb{E}\left[\left|\pi_t^{n,\mathbf{y}_{0:t}} f - \pi_t^{\mathbf{y}_{0:t}} f\right|\right] P_{Y_{0:t}}\left(d\mathbf{y}_{0:t}\right).
\end{aligned}$$

Therefore, if $\mathrm{e}\lim_{n\to\infty} p_t^{n,\mathbf{y}_{0:t-1}} = p_t^{\mathbf{y}_{0:t-1}}$ for $P_{Y_{0:t-1}}$-almost all values $\mathbf{y}_{0:t-1}$ and $\mathrm{e}\lim_{n\to\infty} \pi_t^{n,\mathbf{y}_{0:t}} = \pi_t^{\mathbf{y}_{0:t}}$ for $P_{Y_{0:t}}$-almost all values $\mathbf{y}_{0:t}$ and we have two functions $v\left(\mathbf{y}_{0:t-1}\right)$ and $w\left(\mathbf{y}_{0:t}\right)$ such that, for all $n \geq 0$,

$$\mathbb{E}\left[\left|p_t^{n,\mathbf{y}_{0:t-1}} f - p_t^{\mathbf{y}_{0:t-1}} f\right|\right] \leq v_f\left(\mathbf{y}_{0:t-1}\right), \quad P_{Y_{0:t-1}}\text{-a.s.} \tag{2.16}$$

$$\mathbb{E}\left[\left|\pi_t^{n,\mathbf{y}_{0:t}} f - \pi_t^{\mathbf{y}_{0:t}} f\right|\right] \leq w_f\left(\mathbf{y}_{0:t}\right), \quad P_{Y_{0:t}}\text{-a.s.}, \tag{2.17}$$

then, by the Dominated Convergence Theorem, we have the following limits $\mathrm{e}\lim_{n\to\infty} p_t^{n,Y_{0:t-1}} = p_t$ and $\mathrm{e}\lim_{n\to\infty} \pi_t^{n,Y_{0:t}} = \pi_t$. Conditions (2.16) and (2.17) are trivially satisfied for probability approximations; in this case, $v_f = w_f = 2\|f\|$. However, we cannot obtain a necessary and sufficient condition in this way, since $\mathrm{e}\lim_{n\to\infty} p_t^{n,Y_{0:t}} = p_t$ and $\mathrm{e}\lim_{n\to\infty} \pi_t^{n,Y_{0:t}} = \pi_t$ doesn't imply $\mathrm{e}\lim_{n\to\infty} p_t^{n,\mathbf{y}_{0:t}} = p_t^{\mathbf{y}_{0:t}}$ $P_{Y_{0:t-1}}$-a.s. and, respectively, $\mathrm{e}\lim_{n\to\infty} \pi_t^{n,\mathbf{y}_{0:t}} = \pi_t^{\mathbf{y}_{0:t}}$ $P_{Y_{0:t}}$-a.s.. We have, though, the following proposition:

Proposition 2.3.3. *Provided for all $t \geq 0$, there exists a constant $c_t > 0$ such that $p_t g_t \geq c_t$, the sequences $p_t^{n,Y_{0:t-1}}, \pi_t^{n,Y_{0:t}}$ converge to p_t and to π_t*

with convergence taken to be of type 1. if and only, if for all $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$,

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}\left[\left|\pi_0^n f - \pi_0 f\right|\right] &= 0 \\ \lim_{n\to\infty} \mathbb{E}\left[\left|p_t^{n,Y_{0\,t-1}} f - \pi_{t-1}^{n,Y_{0\,t}} K_{t-1} f\right|\right] &= 0 \\ \lim_{n\to\infty} \mathbb{E}\left[\left|\pi_t^{n,Y_{0\,t}} f - \bar{\pi}_t^{n,Y_{0\,t}} f\right|\right] &= 0. \end{aligned}$$

Proof. The proof is identical to that of Theorem 2.3.1, so we omit it. □

We turn now to the almost sure convergence of $p_t^{n,Y_{0\,t-1}}$ and $\pi_t^{n,Y_{0\cdot t}}$ to p_t and, respectively, to π_t. In this case everything carries through smoothly.

Proposition 2.3.4. *The sequences* $p_t^{n,Y_{0\,t-1}}, \pi_t^{n,Y_{0.t}}$ *converge almost surely to* p_t*, respectively, to* π_t*, for all* $t \geq 0$ *if and only if, for all* $t \geq 0$*,*

$$\begin{aligned} \lim_{n\to\infty} \pi_0^n &= \pi_0, P\text{-a.s.}. \\ \lim_{n\to\infty} d_{\mathcal{M}}\left(p_t^{n,Y_{0\,t-1}}, \pi_{t-1}^{n,Y_{0\,t-1}} K_{t-1}\right) &= 0, \quad P\text{-a.s.}. \\ \lim_{n\to\infty} d_{\mathcal{M}}\left(\pi_t^{n,Y_{0\,t}}, \bar{\pi}_t^{n,Y_{0.t}}\right) &= 0, \quad P\text{-a.s.}. \end{aligned}$$

Proof. Again, the proof for this theorem is similar to that of Theorem 2.3.2, the only difference being the proof that $\lim_{n\to\infty} p_t^{n,Y_{0:t-1}} = p_t$, $P-$a.s. implies $\lim_{n\to\infty} \bar{\pi}_t^{n,Y_{0\,t}} = \pi_t$, $P-a.s.$ which is as follows. Let $\mathcal{M}$ be a convergence determining set of functions in $\mathcal{C}_b(\mathbb{R}^{n_x})$, for instance the set we used to construct the distance $d_{\mathcal{M}}$. Then one can find a subset $\Omega' \subset \Omega$ such that $P(\Omega') = 1$ and for all $\omega \in \Omega'$ $\lim_{n\to\infty} p_t^{n,Y_{0\cdot t-1}(\omega)} g_t = p_t g_t(\omega)$ and $\lim_{n\to\infty} p_t^{n,Y_{0\,t-1}(\omega)}(g_t f) = p_t(g_t f)(\omega)$ for all $f \in \mathcal{M}$. Hence for all $\omega \in \Omega'$

$$\lim_{n\to\infty} \bar{\pi}_t^{n,Y_{0:t}} f(\omega) = \lim_{n\to\infty} \frac{p_t^{n,Y_{0.t}(\omega)}(g_t f)}{p_t^{n,Y_{0:t}(\omega)} g_t} = \frac{p_t(g_t f)}{p_t g_t}(\omega) = \pi_t f(\omega), \quad \forall f \in \mathcal{M},$$

which implies $\lim_{n\to\infty} \bar{\pi}_t^{n,Y_{0\cdot t}} = \pi_t$, P-a.s. □

In the next section we shall present a class of particle filters which satisfies the conditions for these results. This is not the most general class of particle filter and the reader will find other particle filters in the remaining chapters of the book. Our intention is not to exhaust all classes of particle filters, but simply to exemplify the use of the results.

2.4 Examples of particle filters

2.4.1 Description of the particle filters

The algorithms presented below involve the use of a system of n particles which evolve (mutate) according to the law of a given Markov process and, at fixed times, give birth to a number of offspring. Several possible branching mechanisms are described and, after imposing some weak restrictions on those branching mechanisms, the empirical measure associated to the particle systems is proved to converge (as n tends to ∞) to the conditional distribution of the signal, given the observation.

Just as we did above, we shall denote by π_t^n the approximation to π_t and by p_t^n the approximation to p_t. The particle filter will have the following description.

Initialisation. The filter starts with π_0^n – the empirical measure associated to a sample of size n from π_0 – in other words, with n random particles of mass $\frac{1}{n}$, with positions $\mathbf{x}_0^{(i)}$, $i = 1, 2, ..., n$ such that for all i, $\mathbf{x}_0^{(i)}$ has distribution π_0:

$$\pi_0^n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\left\{\mathbf{x}_0^{(i)}\right\}}.$$

Obviously, $\mathrm{e}\lim_{n\to\infty} \pi_0^n = \pi_0$ and also $\lim_{n\to\infty} \pi_0^n = \pi_0$, P-a.s..

Iteration. We describe how to obtain π_t^n from π_{t-1}^n (hence this is a recursive algorithm). The approximation π_{t-1}^n will have the form $\frac{1}{n} \sum_{i=1}^{n} \delta_{\left\{\mathbf{x}_{t-1}^{(i)}\right\}}$, i.e., it will be the empirical measure associated to a system of n random particles. The first step is to move each particle using the transition kernel of the signal. If $\mathbf{x}_{t-1}^{(i)}$ is the position of the i^{th} particle, then $\bar{\mathbf{x}}_t^{(i)}$ – the new position – has the distribution $K_{t-1}\left(\mathbf{x}_{t-1}^{(i)}, \cdot\right)$ and the particles move independent of each other, p_t^n will be the empirical distribution associated with the new cloud of particles

$$p_t^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\left\{\bar{\mathbf{x}}_t^{(i)}\right\}}.$$

For each particle, we compute the weight $w_t^{(i)} = \frac{n g_t(\bar{\mathbf{x}}_0^{(i)})}{\sum_{j=1}^{n} g_t(\bar{\mathbf{x}}_0^{(j)})}$ and obviously $\frac{1}{n} \sum_{i=1}^{n} w_t^{(i)} \delta_{\left\{\bar{\mathbf{x}}_t^{(i)}\right\}}$ is exactly the measure $\bar{\pi}_t^n$ as defined in (2.8),

$$\bar{\pi}_t^n = \frac{1}{n} \sum_{i=1}^{n} w_t^{(i)} \delta_{\left\{\bar{\mathbf{x}}_t^{(i)}\right\}}.$$

Each particle is then replaced by a number of offspring – $\xi_t^{(i)}$ – with the mean number of offspring being $w_t^{(i)}$ and of finite variance. We impose the condition that the total number of particles does not change, i.e., $\sum_{i=1}^{n} \xi_t^{(i)} = n$. We denote the positions of the newly obtained particles by $\mathbf{x}_t^{(i)}$, $i = 1, ..., n$ and define π_t^n as

$$\pi_t^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\left\{\mathbf{x}_t^{(i)}\right\}}.$$

Let A_t^n be the covariance matrix of the random vector $\xi_t := (\xi_t^{(i)})_{i=1}^n$, $A_t^n := \mathbb{E}[(\xi_t - w_t)^T(\xi_t - w_t)]$, where T denotes the transpose and let $w_t := (w_t^{(i)})_{i=1}^n$ be the vector of weights or means. Then we assume that there exists a constant c_t, such that

$$q^T A_t^n q \leq n c_t, \text{ for all } q \in \mathbb{R}^{n_x},\ q = \left(q^{(i)}\right)_{i=1}^n,\ \left|q^{(i)}\right| \leq 1\ i = 1, ..., n. \tag{2.18}$$

The condition that the total number of particles remains constant is not essential. Indeed in (Crisan et al. 1999) we present a similar algorithm with variable number of particles. Theorems 2.3.1 and 2.3.2 and their corollaries can be used in order to prove the convergence of the algorithm in (Crisan et al. 1999).

Remark 3. *An alternative way to obtain p_t^n from π_{t-1}^n is to sample n times from the measure $\pi_{t-1}^n K_{t-1}$ and define p_t^n to be the empirical measure associated with this sample.*

Remark 4. *Condition (2.18) is equivalent to*

$$q^T A_t^n q \leq n \bar{c}_t, \text{ for all } q \in \mathbb{R}^{n_x},\ q = \left(q^{(i)}\right)_{i=1}^n,\ 0 \leq q^{(i)} \leq 1\ i = 1, ..., n \tag{2.19}$$

for a fixed constant $\bar{c}_t$.

Proof. Obviously (2.18) implies (2.19), so we need only show the inverse implication. Let $q \in \mathbb{R}^{n_x}$ be an arbitrary vector such that $q = \left(q^{(i)}\right)_{i=1}^n$, $\left|q^{(i)}\right| \leq 1,\ i = 1, ..., n$. Let also $q_+ = \left(q_+^{(i)}\right)_{i=1}^n$, $q_- = \left(q_-^{(i)}\right)_{i=1}^n$, $0 \leq q_+^{(i)}, q_-^{(i)} \leq 1\ i = 1, ..., n$, $q_+^{(i)} \triangleq \max\left(q^{(i)}, 0\right)$, $q_-^{(i)} \triangleq \max\left(-q^{(i)}, 0\right)$. We define $||\cdot||_A$ to be the semi-norm associated to the matrix A (if all its eigenvalues are positive, then $||\cdot||_A$ is a genuine norm), $||q||_A \triangleq \sqrt{q^T A q}$. Then $q = q_+ - q_-$ and, using the triangle inequality and (2.19),

$$||q||_{A_t^n} \leq ||q_+||_{A_t^n} + ||q_-||_{A_t^n} \leq 2\sqrt{n\bar{c}_t},$$

which implies that (2.18) holds with $c_t = 4\bar{c}_t$. □

The only part not yet specified is the branching mechanism ξ_t. Below we introduce two branching mechanisms that leave the total number of particles constant and satisfy (2.18).

2.4.2 Branching mechanisms

We introduce two possible branching mechanisms, the multinomial branching mechanism and the tree-based branching mechanism. Neither of the mechanisms is new; both can be found in the literature under various names (see, for instance, (Crisan et al. 1999) and the references therein).

The multinomial branching mechanism.

In this case, we choose $\xi_t = \text{Multinomial}\left(n, \frac{w_t^{(1)}}{n}, \dots, \frac{w_t^{(n)}}{n}\right)$. We have $\mathbb{E}\left[\xi_t^{(i)}\right] = w_t^{(i)}$, $\mathbb{E}\left[\left(\xi_t^{(i)} - w_t^{(i)}\right)^2\right] = w_t^{(i)}\left(1 - \frac{w_t^{(i)}}{n}\right)$ and, in addition, we have $\mathbb{E}\left[\left(\xi_t^{(i)} - w_t^{(i)}\right)\left(\xi_t^{(j)} - w_t^{(j)}\right)\right] = -\frac{w_t^{(i)}w_t^{(j)}}{n}$. Then for all $q \in \mathbb{R}^{n_x}$, $q = \left(q^{(i)}\right)_{i=1}^n$, $\left|q^{(i)}\right| \leq 1$, $i = 1, ..., n$,

$$\begin{aligned}
q^T A_t^n q &= \sum_{i=1}^n w_t^{(i)}\left(1 - \frac{w_t^{(i)}}{n}\right)\left(q^{(i)}\right)^2 - 2\sum_{\substack{i,j=1\\ i\neq j}}^n \frac{w_t^{(i)}w_t^{(j)}}{n} q^{(i)}q^{(j)} \\
&= \sum_{i=1}^n w_t^{(i)}\left(q^{(i)}\right)^2 - \frac{1}{n}\left(\sum_{i=1}^n w_t^{(i)}q^{(i)}\right)^2 \\
&\leq \sum_{i=1}^n w_t^{(i)}
\end{aligned}$$

and, since $\sum_{i=1}^n w_t^{(i)} = n$, (2.18) holds with $c_t = 1$. The properties of the multinomial branching mechanism have been extensively studied in the literature (again, see, for instance, (Crisan et al. 1999) and the references therein).

The tree-based branching mechanism.

We consider now a binary tree with n leaves and denote by r the root of the tree (see Figure 2.1). To the i^{th} particle we associate the i^{th} leaf of the tree and attach to it the corresponding weight w_i. Let m be an arbitrary node of the tree which is not a leaf, m_1 and m_2 be its direct offspring and I_m (respectively, $I_{m_1}$and I_{m_2}) be the set of leaves/particles which are the descendants of m (respectively, m_1 and m_2). Let w_m be the weight attached to the node m defined as the sum of the weights of all the leaves which are its descendants

$$w_m \triangleq \sum_{i\in I_m} w_i.$$

Obviously, $w_r = n$ because I_r comprises the whole set of leaves/particles and $w_m = w_{m_1} + w_{m_2}$, since I_{m_1} and I_{m_2} represent a partition of I_m.

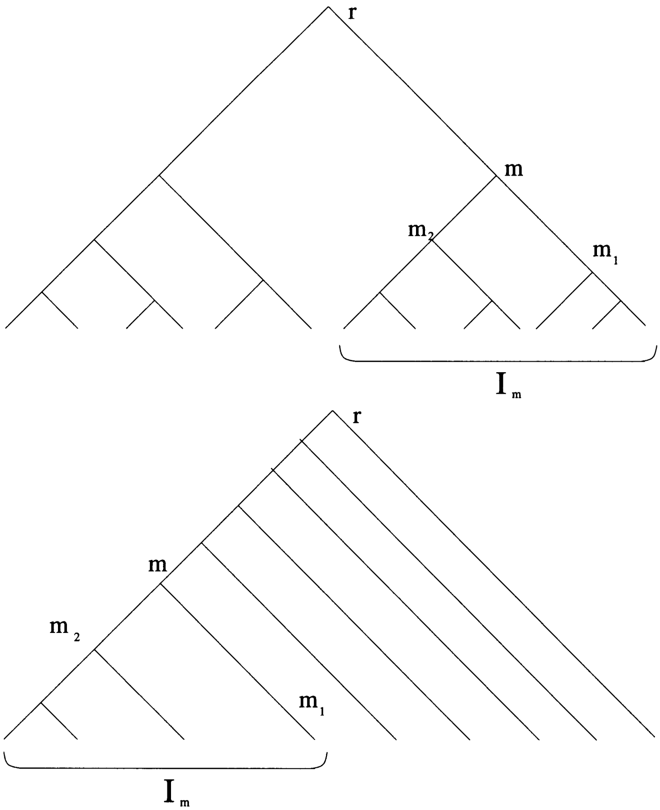

Figure 2.1. Branching binary trees.

We construct next a set of random variables ξ_m for all the nodes of the tree including the leaves such that

$$\xi_m = \begin{cases} [w_m] & \text{with probability} \quad 1 - \{w_m\} \\ [w_m] + 1 & \text{with probability} \quad \{w_m\} \end{cases}, \tag{2.20}$$

where $[x]$ is the integer of part $x \in \mathbb{R}$ and $\{x\} \triangleq x - [x]$. The random variables associated to the leaves will constitute the branching mechanisms of the corresponding particles. The construction is done recursively starting from the root and finishing with the leaves such that *the sum of the random variables associated to the children nodes is equal to the random variable associated to the parent node.*

We define ξ_r to be identically equal to n. We show next the rule of recursion.

Suppose that we have constructed ξ_m with the distribution given by (2.20). First we assign $[w_{m_1}]$ to m_1 and $[w_{m_2}]$ to m_2 and observe that $\xi_m - [w_{m_1}] - [w_{m_2}]$ is equal to $0, 1$ or 2. These are the only possibilities since ξ_m can have only two values, $[w_m]$ or $[w_m]+1$and $[w_m]$ $-[w_{m_1}] - [w_{m_2}]$ can be either 0 or 1. If $\xi_m - [w_{m_1}] - [w_{m_2}] = 2$ we assign one unit to each of the nodes, if $\xi_m - [w_{m_1}] - [w_{m_2}] = 0$ we do nothing, and if $\xi_m - [w_{m_1}] - [w_{m_2}] = 1$ we assign the unit randomly to one of the nodes such that the resulting random variables will have the right distribution. More precisely, if $[w_m]$ $= [w_{m_1}] + [w_{m_2}]$, we define

$$\begin{aligned}
\xi_{m_1} &\triangleq [w_{m_1}] + (\xi_m - [w_m])\eta_m \\
\xi_{m_2} &\triangleq [w_{m_2}] + (\xi_m - [w_m])(1 - \eta_m),
\end{aligned}$$

where

$$\eta_m = \begin{cases} 0 & \text{with probability } \frac{\{w_{m_2}\}}{\{w_m\}} \\ 1 & \text{with probability } \frac{\{w_{m_1}\}}{\{w_m\}} \end{cases}. \tag{2.21}$$

The distribution of the auxiliary random variable is well defined because

$$\begin{array}{l} w_m = w_{m_1} + w_{m_2} \\ [w_m] = [w_{m_1}] + [w_{m_2}] \end{array} \Longrightarrow \frac{\{w_{m_1}\}}{\{w_m\}} + \frac{\{w_{m_2}\}}{\{w_m\}} = 1.$$

If $[w_m] = [w_{m_1}] + [w_{m_2}] + 1$, we define

$$\begin{aligned}
\xi_{m_1} &\triangleq [w_{m_1}] + 1 + (\xi_m - [w_m] - 1)\eta_m \\
\xi_{m_2} &\triangleq [w_{m_2}] + 1 + (\xi_m - [w_m] - 1)(1 - \eta_m),
\end{aligned}$$

where

$$\eta_m = \begin{cases} 0 & \text{with probability } \frac{1-\{w_{m_2}\}}{1-\{w_m\}} \\ 1 & \text{with probability } \frac{1-\{w_{m_1}\}}{1-\{w_m\}} \end{cases}. \tag{2.22}$$

The distribution of the auxiliary random variable is well defined because

$$\begin{array}{l} w_m = w_{m_1} + w_{m_2} \\ [w_m] = [w_{m_1}] + [w_{m_2}] + 1 \end{array} \Longrightarrow \frac{1-\{w_{m_1}\}}{1-\{w_m\}} + \frac{1-\{w_{m_2}\}}{1-\{w_m\}} = 1.$$

We take all the random variables η_m independent of each other. In either case

$$\xi_{m_1} \triangleq [w_{m_1}] + a_m + (\xi_m - b_m)\eta_m \tag{2.23}$$

$$\xi_{m_2} \triangleq [w_{m_2}] + a_m + (\xi_m - b_m)(1 - \eta_m), \tag{2.24}$$

where $a_m = [w_m] - [w_{m_1}] + [w_{m_2}] = \{w_{m_1}\} + \{w_{m_2}\} - \{w_m\}$ and $b_m =$ $[w_m] - a_m$.

Let $\xi_\imath$ $i = 1, ..., n$ be the resulting random variables corresponding to the leaves of the tree, and therefore to the particles of the system. From (2.20) we see that

1. $E[\xi_i] = w_i$
2. $\mathrm{Var}(\xi_i) = E[\xi_i^2] - E[\xi_i]^2 = \{w_i\}(1 - \{w_i\}) \leq \frac{1}{4}$.

We also have

$$\begin{aligned} E[(\xi_{m_1} - E[\xi_{m_1}]) |\xi_m] &= E[(a_m - \{w_{m_1}\} + (\xi_m - [w_m] - a_m)\eta_m) |\xi_m] \\ &= \frac{a_m - \{w_{m_1}\}}{a_m - \{w_m\}}(\xi_m - w_m), \end{aligned}$$

and

$$\begin{aligned} & E[(\xi_{m_1} - E[\xi_{m_1}])(\xi_{m_2} - E[\xi_{m_2}]) |\xi_m] \\ &= E\left[\prod_{i=1}^{2} (a_m - \{w_{m_i}\} + (\xi_m - [w_m] - a_m)\eta_m) |\xi_m\right] \\ &= (a_m - \{w_{m_1}\})(a_m - \{w_{m_2}\}) \left(\frac{2(\xi_m - w_m)}{a_m - \{w_m\}} - 1\right). \end{aligned}$$

Hence, if i and j are two different particles or leaves, then

$$\begin{aligned} & E\left[(\xi_i - E[\xi_{m_i}])(\xi_j - E[\xi_{m_j}])\right] \\ &= -\prod_{s=1}^{h_i-1} \frac{a_{m_s} - \{w_{m_{s-1}}\}}{a_{m_s} - \{w_{m_s}\}} \prod_{t=1}^{h_j-1} \frac{a_{n_t} - \{w_{n_{t-1}}\}}{a_{n_t} - \{w_{n_t}\}} \left(a_{m_{h_i}} - \left\{w_{m_{h_i-1}}\right\}\right)\left(a_{n_{h_j}} - \left\{w_{n_{h_j-1}}\right\}\right), \end{aligned}$$

where $m_{h_i} = m_{h_j}$ is the most recent common ancestor of i and j, h_i and h_j are the number of generations from the most common ancestor to i, respectively, j, $m_1, m_2, ..., m_{h_i}$ is the genealogical line of i, and $m_1, m_2, ..., m_{h_i}$ is the genealogical line of j. Now, since the following apply $\mathrm{Var}(\xi_i) \leq \frac{1}{4}$ and $E\left[(\xi_i - E[\xi_{m_i}])(\xi_j - E[\xi_{m_j}])\right] \leq 0$, we have that $q^T A_t^n q \leq \frac{1}{4}n$, for all $q \in \mathbb{R}^{n_x}$, $q = \left(q^{(i)}\right)_{i=1}^n$, $0 \leq q^{(i)} \leq 1$, $i = 1, .., n$. Hence by Remark 4, (2.18) holds with $c_t = \frac{1}{2}$.

2.4.3 Convergence of the algorithm

We fix first the observation process to an arbitrary value $\mathbf{y}_{0:T}$, where T is a finite time horizon. We prove first that the random measures resulting from the class of algorithms described above converge to $\pi_t^{\mathbf{y}_0\,t}$ (respectively $p_t^{\mathbf{y}_0\,t-1}$) for all $0 \leq t \leq T$. For this we introduce the following σ-algebras

$$\begin{aligned} \mathcal{F}_t &= \sigma\left(\bar{\mathbf{x}}_s^{(i)}, \mathbf{x}_s^{(i)}, s \leq t, \quad i = 1, .., n\right) \\ \bar{\mathcal{F}}_t &= \sigma\left(\bar{\mathbf{x}}_s^{(i)}, \mathbf{x}_s^{(i)}, s < t, \quad \bar{\mathbf{x}}_t^{(i)}, \quad i = 1, .., n\right). \end{aligned}$$

Theorem 2.4.1. *Let $(p_t^n)_{n=1}^\infty$ and $(\pi_t^n)_{n=1}^\infty$ be the measure valued sequences produced by the class of algorithms described above. Then, for all $0 \leq t \leq T$, we have*

$$\mathrm{e}\lim_{n\to\infty} p_t^n = p_t^{\mathbf{y}_0\,t-1} \qquad \mathrm{e}\lim_{n\to\infty} \pi_t^n = \pi_t^{\mathbf{y}_0\,t}.$$

Proof. We apply Theorem 2.3.1. Since **a1.** is clearly satisfied, we only need to show that **b1.** and **c1.** hold true. If $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, then

$$\mathbb{E}\left[f\left(\bar{\mathbf{x}}_t^{(i)}\right)|\mathcal{F}_{t-1}\right] = K_{t-1}f\left(\mathbf{x}_{t-1}^{(i)}\right) \quad \text{for all } i = 1, ..., n.$$

Hence, $\mathbb{E}\left[p_t^n f\,|\mathcal{F}_t\right] = \pi_{t-1}^n K_{t-1}f$ and, using the independence of the motion of the particles, we have

$$\mathbb{E}\left[\left(p_t^n f - \pi_{t-1}^n K_{t-1}f\right)^2|\mathcal{F}_{t-1}\right] = \frac{1}{n}\pi_{t-1}^n\left(K_{t-1}f^2 - \left(K_{t-1}f\right)^2\right).$$

Therefore, $\mathbb{E}\left[\left(p_t^n f - \pi_{t-1}^n K_{t-1}f\right)^2\right] \leq \frac{||f||^2}{n}$ and **b1.** is satisfied. Now, since $\pi_t^n = \frac{1}{n}\sum_{i=1}^n \xi_t^{(i)}\delta_{\left\{\bar{\mathbf{x}}_t^{(i)}\right\}}$, we have

$$\begin{aligned} \mathbb{E}\left[\pi_t^n f\,|\bar{\mathcal{F}}_t\right] &= \bar{\pi}_t^n f \\ \mathbb{E}\left[\left(\pi_t^n f - \bar{\pi}_t^n f\right)^2|\bar{\mathcal{F}}_t\right] &= \frac{1}{n^2}\left(q_t^n\right)^T A_t^n q_t^n, \end{aligned} \tag{2.25}$$

where q_t^n is the vector with entries $\left(q_t^n\right)^{(i)} = f\left(\bar{\mathbf{x}}_t^{(i)}\right)$. From (2.18) and (2.25) we see that

$$\mathbb{E}\left[\left(\pi_t^n f - \bar{\pi}_t^n f\right)^2|\bar{\mathcal{F}}_t\right] \leq \frac{c_t\,||f||^2}{n}.$$

Therefore, $\mathbb{E}\left[\left(\pi_t^n f - \bar{\pi}_t^n f\right)^2\right] \leq c_t\frac{||f||^2}{n}$ and **c1.** is satisfied. □

Theorem 2.4.2. *Let $(p_t^n)_{n=1}^\infty$ and $(\pi_t^n)_{n=1}^\infty$ be the measure-valued sequences produced by the algorithm with multinomial branching mechanism described above. Then, for all $0 \leq t \leq T$, we have*

$$\lim_{n\to\infty} p_t^n = p_t^{\mathbf{y}_{0\cdot t-1}} \qquad \lim_{n\to\infty} \pi_t^n = \pi_t^{\mathbf{y}_{0\cdot t}} \quad P\text{-a.s.}$$

Proof. We apply Theorem 2.3.2. Let $\mathcal{M} \in \mathcal{C}_b(\mathbb{R}^d)$ be the countable, convergence determining set of functions defined in the previous section. Since $\mathbb{E}\left[f\left(\bar{\mathbf{x}}_t^{(i)}\right)|\mathcal{F}_{t-1}\right] = K_{t-1}f\left(\mathbf{x}_{t-1}^{(i)}\right)$ and using the independence of $\bar{\mathbf{x}}_t^{(1)}$, $\bar{\mathbf{x}}_t^{(2)}$, ..., $\bar{\mathbf{x}}_t^{(n)}$ given $\mathcal{F}_t$, we have

$$\begin{aligned} &\mathbb{E}\left[\left(p_t^n f - \pi_{t-1}^n K_{t-1}f\right)^4|\mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n\left(f\left(\bar{\mathbf{x}}_t^{(i)}\right) - K_{t-1}f\left(\mathbf{x}_{t-1}^{(i)}\right)\right)\right)^4|\mathcal{F}_{t-1}\right] \\ &= \frac{1}{n^4}\sum_{i=1}^n\mathbb{E}\left[\left(f\left(\bar{\mathbf{x}}_t^{(i)}\right) - K_{t-1}f\left(\mathbf{x}_{t-1}^{(i)}\right)\right)^4|\mathcal{F}_{t-1}\right] \\ &+ \frac{2}{n^4}\sum_{1\leq i<j\leq n}\mathbb{E}\left[\left(f\left(\bar{\mathbf{x}}_t^{(i)}\right) - K_{t-1}f\left(\mathbf{x}_{t-1}^{(i)}\right)\right)^2\left(f\left(\bar{\mathbf{x}}_t^{(j)}\right) - K_{t-1}f\left(\mathbf{x}_{t-1}^{(j)}\right)\right)^2|\mathcal{F}_{t-1}\right]. \end{aligned} \tag{2.26}$$

Hence, by taking expectation in both terms of equation (2.26), we have $\mathbb{E}\left[\left(p_t^n f - \pi_{t-1}^n K_{t-1} f\right)^4\right] \le \frac{16\|f\|^4}{n^2}$ and using a Borel Cantelli argument we get $\lim_{n\to\infty}\left|p_t^n f - \pi_{t-1}^n K_{t-1} f\right| = 0$ for all $f \in \mathcal{M}$, P-a.s.. Therefore $\lim_{n\to\infty} d_{\mathcal{M}}\left(p_t^n, \pi_{t-1}^n K_{t-1}\right) = 0$, and hence we have **b2.**. Similarly, one proves that, for all $f \in \mathcal{M}$,

$$\mathbb{E}\left[\left(\pi_t^n f - \bar{\pi}_t^n f\right)^4 |\mathcal{F}_t\right] \le \frac{16\,\|f\|^4}{n^2}, \tag{2.27}$$

which implies, as above, that $\lim_{n\to\infty} d_{\mathcal{M}}\left(\pi_t^n, \bar{\pi}_t^n\right) = 0$, i.e., **c2.**. □

The tree-based branching mechanism has more complex correlations between higher moments of the individual branching mechanisms, and we have not been able so far to prove that (2.27) holds true. However, we conjecture that this is the case and therefore we have almost sure convergence of for this algorithm as well.

We turn now to the case in which the observation process is random. With similar arguments one uses Propositions 2.3.3 and 2.3.4 to prove the following corollaries

Corollary 2.4.3. *Provided for all* $t \ge 0$, *there exists a constant* $c_t > 0$ *such that* $p_t g_t \ge c_t$, *we have*

$$\mathrm{e}\lim_{n\to\infty} p_t^{n,Y_{0\ t-1}} = p_t \qquad \mathrm{e}\lim_{n\to\infty} \pi_t^{n,Y_{0\cdot t}} = \pi_t.$$

Corollary 2.4.4. *Let* $(p_t^n)_{n=1}^\infty$ *and* $(\pi_t^n)_{n=1}^\infty$ *be the measure-valued sequences produced by the algorithm with multinomial branching mechanism described above. Then, for all* $0 \le t \le T$, *we have*

$$\lim_{n\to\infty} p_t^{n,Y_{0\cdot t-1}} = p_t \qquad \lim_{n\to\infty} \pi_t^{n,Y_{0\ t}} = \pi_t. \ \ P--a.s..$$

2.5 Discussion

The remarkable feature of Theorems 2.3.1 and 2.3.2 is their provision of efficient techniques for proving convergence of particle algorithms. The *necessary and sufficient* conditions **a1(2).**, **b1(2).**, **c1(2).** are natural and easy to verify, as can be seen from the proofs of Theorems 2.4.1 and 2.4.2. Of course, these theorems can be applied when the algorithms studied provide not only π_t^n – the approximation to π_t – but also p_t^n – the intermediate approximation to p_t. However, this may not always be the case, and one can have algorithms where π_t^n is obtained from π_{t-1}^n without using intrinsically the intermediate step p_t^n. In this case one *defines* p_t^n to be $\pi_{t-1}^n K_{t-1}$, or indeed any measure which is 'close' to $\pi_{t-1}^n K_{t-1}$ (so that **b1(2).** is satisfied) and then tries to apply the theorems. This implies that one need only check **c1(2).**, i.e., whether π_t^n is closed to $\bar{\pi}_t^n$, where $\bar{\pi}_t^n$ is the probability measure defined in (2.8).

The results presented in this chapter are directed towards proving convergence of the particle filters, their rates of convergence are not discussed. One can easily prove that, for the class of algorithms presented here, we have

$$\mathbb{E}\left[|\pi_t^n f - \pi_t f|\right] \leq \frac{m_t}{\sqrt{n}} \,||f||, \qquad \text{for all } f \in C_b(\mathbb{R}^d),$$

where m_t are constants independent of n, for all $t \geq 0$ (see (Crisan et al. 1999) for a proof of this result).

The two branching algorithms presented are both unbiased, in the sense that the mean of the offspring distribution of each particle is its corresponding weight. Also, the variance of the offspring distribution has the same order. However, of the two, the tree-based branching algorithm has minimal variance for individual particles. If ξ_i is the number of offspring of the particle i, then ξ_i has the prescribed mean w_i (the corresponding weight) and is an integer-valued random variable with the minimal variance among all the integer-valued random variables with the given mean w_i. So if, say, $w_i = 3.4$, then ξ_i takes only the values 3 (with probability 0.6) or 4 (with probability 0.4); hence, it stays as close as possible to the required w_i. By contrast, when one uses the multinomial branching mechanism, ξ_i can have any value between 0 and n with positive probability, meaning extra randomness is introduced in the system.

The minimal variance property for the tree-based branching algorithms holds true not only for individual particles, but also for groups of particles. If m is an arbitrary node of the tree and I_m is the set of particles corresponding to the offspring of m, then $\xi_m = \sum_{i \in I_m} \xi_i$ has the required mean $\sum_{i \in I_m} w_i$ and has minimal variance. Moreover, we prove in a forthcoming paper that in a discrete time/space set-up the tree-based branching algorithm (or, rather, a variant of it) gives the best approximation to π_t, when the measure of comparison is the relative entropy with respect to π_t.

Although we think that the tree-based branching mechanism is optimal for any correction/resampling procedure that we might undertake, we also believe that the updating procedure has to be greatly improved. The raw Monte Carlo method used in the class of algorithms presented is clearly not optimal because it introduces too much randomness. The answer may lie in applying a PDE approach – discretising the state-space and moving the particles using a quasi-deterministic law which approximates K_t.

Of course, there are other routes by which one may introduce vast improvements to these methods, such as resampling out of the current population using the new observation and then updating the position of the particles. Undoubtedly, there is virgin territory to be explored in this area and exciting developments await both practitioners and theoreticians.

2.6 Appendix

2.6.1 Conditional probabilities and conditional expectations

In this section, we state briefly some definitions and results that we need in the article. One can find details and proofs of these results in any monograph on probability theory, for example in (Shiryaev 1996) or (Williams 1991). We include this section for reasons of completeness, and also because we want to rigorously justify why π_t and $\pi_t^{\mathbf{y}_{0:t}}$ as defined in (2.4) and (2.5) are (random, respectively, deterministic) probability measures, and why we have $\pi_t f = \mathbb{E}\left[f(X_t)\left|\sigma(Y_{0:t})\right.\right]$ and $\pi_t^{\mathbf{y}_{0:t}} f = \mathbb{E}\left[f(X_t)\left|Y_{0:t} = \mathbf{y}_{0:t}\right.\right]$, respectively. We also state the results on conditional probabilities and expectations needed to prove the recurrence formula (2.6) in the following subsection.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \in \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of an integrable $\mathcal{F}$-measurable random variable ξ given $\mathcal{G}$ is defined as the integrable $\mathcal{G}$-measurable random variable, denoted by $\mathbb{E}\left[\xi\left|\mathcal{G}\right.\right]$, with the property

$$\int_A \xi dP = \int_A \mathbb{E}\left[\xi\left|\mathcal{G}\right.\right] dP, \tag{2.28}$$

for all $A \in \mathcal{G}$. Then $\mathbb{E}\left[\xi\left|\mathcal{G}\right.\right]$ exists and is *almost surely* unique. By this we mean that if $\bar{\xi}$ is another $\mathcal{G}$-measurable integrable random variable such that $\int_A \xi dP = \int_A \bar{\xi} dP$ for all $A \in \mathcal{G}$, then $\mathbb{E}\left[\xi\left|\mathcal{G}\right.\right] = \bar{\xi}$, $P-$a.s..

Here are some of the properties of the conditional expectation.

a. If $\alpha_1, \alpha_2 \in \mathbb{R}$ and ξ_1, ξ_2 are $\mathcal{F}$-measurable, then

$$\mathbb{E}\left[\alpha_1\xi_1 + \alpha_2\xi_2\left|\mathcal{G}\right.\right] = \alpha_1\mathbb{E}\left[\xi_1\left|\mathcal{G}\right.\right] + \alpha_2\mathbb{E}\left[\xi_2\left|\mathcal{G}\right.\right], \quad P\text{-a.s..}$$

b. If $\xi \geq 0$, then $\mathbb{E}\left[\xi\left|\mathcal{G}\right.\right] \geq 0$, P-a.s..
c. If $0 \leq \xi_n \nearrow \xi$, then $\mathbb{E}\left[\xi_n\left|\mathcal{G}\right.\right] \nearrow \mathbb{E}\left[\xi\left|\mathcal{G}\right.\right]$, P-a.s..
d. If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbb{E}\left[\mathbb{E}\left[\xi\left|\mathcal{G}\right.\right]\left|\mathcal{H}\right.\right] = \mathbb{E}\left[\xi\left|\mathcal{H}\right.\right]$, P-a.s..
e. If ξ is $\mathcal{G}$-measurable, then $\mathbb{E}\left[\xi\eta\left|\mathcal{G}\right.\right] = \xi\mathbb{E}\left[\eta\left|\mathcal{G}\right.\right]$, P-a.s..
f. If $\mathcal{H}$ is independent of $\sigma\left(\sigma(\xi), \mathcal{G}\right)$, then $\mathbb{E}\left[\xi\left|\sigma(\mathcal{G}, \mathcal{H})\right.\right] = \mathbb{E}\left[\xi\left|\mathcal{G}\right.\right]$, P-a.s..

The conditional probability of a set $A \in \mathcal{F}$ with respect to the σ-algebra $\mathcal{G}$ is the random variable denoted by $P(A|\mathcal{G})$defined as $P(A|\mathcal{G}) \triangleq \mathbb{E}\left[I_A\left|\mathcal{G}\right.\right]$, where I_A is the indicator function of the set A. From (2.28) we deduce that $P(A \cap B) = \int_B P(A|\mathcal{G})\, dP$ for all $B \in \mathcal{G}$. Let $\eta_1, ..., \eta_k$ be $\mathcal{F}$-measurable random variables. Then $\mathbb{E}\left[\xi\left|\eta_1, ..., \eta_k\right.\right]$ – the conditional expectation of ξ with respect to $\eta_1, ..., \eta_k$ – is the conditional expectation of ξ with respect to the σ-algebra generated by $\eta_1, ..., \eta_k$, i.e., $\mathbb{E}\left[\xi\left|\eta_1, ..., \eta_k\right.\right] = \mathbb{E}\left[\xi\left|\sigma(\eta_1, ..., \eta_k)\right.\right]$ and we have the analogue definition of $P(A|\eta_1, ..., \eta_k)$ – the conditional probability of A with respect to $\eta_1, ..., \eta_k$.

The fact that $P(A|\mathcal{G})$ is not pointwise uniquely defined may be troublesome. It implies that, for all $A \in \mathcal{B}(\mathbb{R}^{n_x})$, $\pi_t(A)$ is not pointwise uniquely defined but only almost sure. If $A_1, A_2, ... \in \mathcal{B}(\mathbb{R}^{n_x})$ is a sequence of pairwise disjoint sets, then, by properties a. and c.,

$$\pi_t\left(\bigcup_n A_n|\mathcal{G}\right) = \sum_n \pi_t(A_n|\mathcal{G}), \quad P- \text{a.s..}$$

However, for a given ω, the conditional probability $\pi_t(\cdot)(\omega)$ cannot be considered a-priori a measure on $\mathbb{R}^{n_x}$. One might suppose that $\pi_t(\cdot)(\omega)$ is a measure except for a set $\mathcal{N} \in \Omega$ of measure 0. This may not be the case for the following reason. Let $\mathcal{N}(A_1, A_2, ...)$ be the set of points ω such that the countable additivity property fail for these $A_1, A_2, ...$ Then the excluded set is $\mathcal{O}$, where $\mathcal{O} = \bigcup \mathcal{N}(A_1, A_2, ...)$, and the union is taken over all sequences $(A_i)_{i=1}^{\infty}$, such that for all $i > 0$, $A_i \in \mathcal{B}(\mathbb{R}^{n_x})$ and although the P-measure of $\mathcal{N}(A_1, A_2, ...)$ is zero, the P-measure of $\mathcal{N}$ may not be zero because of the uncountable union. However, it is natural to request that the conditional probability of the signal $\pi_t(\cdot)(\omega)$ be a (probability) measure.

Definition 2.6.1. *Let (Ω, F, P) be a probability, $(E, \mathcal{E})$ be a measurable space, $X : \Omega \to E$ be an $\mathcal{E}/\mathcal{F}$-measurable random element, and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. A function $Q(\omega, B)$ defined for all $\omega \in \Omega$ and $B \in \mathcal{E}$ is a regular conditional distribution/probability of X with respect to $\mathcal{G}$ if*
(a) for each $\omega \in \Omega$, $Q(\omega, \cdot)$ is a probability measure on $(E, \mathcal{E})$.
(b) for each $B \in \mathcal{F}$, $Q(\cdot, B)$ is $\mathcal{G}$-measurable and $Q(\cdot, B) = P(X \in B|\mathcal{G})$, $P-$a.s..

Definition 2.6.2. *A measurable space $(E, \mathcal{E})$ is a Borel space if there exists a one-to-one mapping $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $f(E) \in \mathcal{B}(\mathbb{R})$, f is $\mathcal{E}$-measurable and f^{-1} is $\mathcal{B}(\mathbb{R})/\mathcal{E}$- measurable.*

We state the following theorem without proof. The reader can find a proof in (Parthasarathy 1967) pp. 146-150.

Theorem 2.6.3. *Let $X = X(\omega)$ be a random element with values in a Borel space $(E, \mathcal{E})$. There then exists a regular conditional distribution of X with respect to $\mathcal{G}$.*

Since $(\mathbb{R}^{n_x}, \mathcal{B}(\mathbb{R}^{n_x}))$ is a Borel space, there exists a regular conditional distribution of X_t with respect to $\sigma(Y_{0:t})$. Therefore, if for all $A \in \mathcal{B}(\mathbb{R}^{n_x})$, we assign to $\pi_t(A)$ the value $Q(\cdot, A)$ (we can do this because $\pi_t(A)$ is defined only almost surely), where Q is the regular conditional distribution of X_t with respect to $\sigma(Y_{0:t})$, then π_t is a probability measure.

Remark 5. *If π_t is defined as above, then the identity in (2.4), $\pi_t f = \mathbb{E}[f(X_t)|\sigma(Y_{0:t})]$ holds true, $P-$a.s., for all $\mathcal{B}(\mathbb{R}^{n_x})$-measurable functions f.*

Proof. If $f = I_B$, where I_B is the characteristic function of an arbitrary set $B \in \mathcal{B}(\mathbb{R}^{n_x})$ the required formula holds by Definition 2.6.1(b). Consequently, it holds for simple functions. Let $\xi \geq 0$ be an arbitrary non-negative function and $0 \leq f_n \nearrow f$, where f_n are simple functions. Using property c. of the conditional expectation we have $\mathbb{E}[f(X_t)|\sigma(Y_{0:t})] = \lim_{n\to\infty}\mathbb{E}[f_n(X_t)|\sigma(Y_{0:t})]$, P-a.s.. But, since π_t is a probability measure for each $\omega \in \Omega$, we have, by Monotone Convergence Theorem, $\lim_{n\to\infty}\pi_t f_n = \pi_t f$. Hence, the identity holds for all non-negative measurable functions. The general case reduces to this one by using the representation $\xi = \xi_+ - \xi_-$. □

Let ξ, η be $\mathcal{F}$-measurable functions. Since $\mathbb{E}[\xi|\eta]$ is a $\sigma(\eta)$-measurable random variable, there exists a function $m = m(y) : \mathbb{R} \to \mathbb{R}$ such that $m(\eta) = \mathbb{E}[\xi|\eta]$. We denote $m(y)$ by $\mathbb{E}[\xi|\eta = y]$ and call it the conditional expectation of ξ with respect to the event $\{\eta = y\}$. Therefore, via the change of variable formula, we have, for all $A \in \mathcal{B}(\mathbb{R})$,

$$\int_{\{\omega;\eta\in A\}} \xi(\omega) P(d\omega) = \int_{\{\omega;\eta\in A\}} m(\eta(\omega))\, dP(d\omega) = \int_A m(y) P_\eta(dy), \tag{2.29}$$

where P_η is the probability distribution of η. We can use (2.29) as the defining formula for conditional expectation of ξ with respect to the event $\{\eta = y\}$. Indeed, the conditional expectation of ξ with respect to the event $\{\eta = y\}$, denoted by $\mathbb{E}[\xi|\eta = y]$ is the $\mathcal{B}(\mathbb{R})$-measurable random variable such that

$$\int_{\{\omega;\eta\in A\}} \xi dP = \int_A \mathbb{E}[\xi|\eta = y] P_\eta(dy) \tag{2.30}$$

holds true for all $A \in \mathcal{B}(\mathbb{R})$ (the definition has the obvious extension to ξ, η random vectors). Again, the function $y \to \mathbb{E}[\xi|\eta = y]$ is P_η-*almost surely* unique.

If we know $\mathbb{E}[\xi|\eta = y]$, *then we can deduce* $\mathbb{E}[\xi|\eta]$, *and conversely from* $\mathbb{E}[\xi|\eta]$ *we can reconstruct the conditional expectation* $\mathbb{E}[\xi|\eta = y]$. The expectation $\mathbb{E}[\xi|\eta = y]$ satisfies the following identity P_η–a.s.

$$\mathbb{E}[\xi f(\eta)|\eta = y] = f(y)\mathbb{E}[\xi|\eta = y] \tag{2.31}$$

for all $f \in \mathcal{B}(\mathbb{R})$. Moreover, if ξ and η are independent and $g \in \mathcal{B}(\mathbb{R}^2)$, then P_η–a.s.,

$$\mathbb{E}[\xi|\eta = y] = \mathbb{E}[\xi] \tag{2.32}$$

$$\mathbb{E}[g(\xi,\eta)|\eta = y] = \mathbb{E}[g(\xi, y)]. \tag{2.33}$$

The conditional probability of the event given by $A \in \mathcal{F}$ with respect to the event/under the condition that $\{\eta = y\}$ (notation $P(A|\eta = y)$) is defined as $\mathbb{E}[I_A|\eta = y]$. $P(A|\eta = y)$ is the $\mathcal{B}(\mathbb{R})$-measurable random variable such that

$$P(A \cap \{\eta \in B\}) = \int_B P(A \,|\eta = y) P_\eta(dy) \tag{2.34}$$

for all $B \in \mathcal{B}(\mathbb{R})$.

Now, if π_t is the regular conditional distribution of X_t with respect to $Y_{0:t}$ then, for all $A \in \mathcal{B}(\mathbb{R}^{n_x})$, $\pi_t(A)$ is $Y_{0:t}$-measurable. Hence, there exists a function $m = m(A, \mathbf{y}_{0,t}) : \mathcal{B}(\mathbb{R}^{n_x}) \times \mathrm{Im}(Y_{0:t}) \to \mathbb{R}$, such that, *pointwise* (and not just almost surely),

$$\pi_t(A)(\omega) = m(A, Y_{0:t}(\omega)).$$

Since, for all $\omega \in \Omega$, $\pi_t(\cdot)(\omega)$ is a probability measure, it follows that for all $\mathbf{y}_{0,t} \in \mathrm{Im}(Y_{0:t})$, $m(\cdot, \mathbf{y}_{0,t})$ is a probability measure on $\mathcal{B}(\mathbb{R}^{n_x})$, Then, just as above, we assign to $\pi_t^{\mathbf{y}_{0.t}}(A)$ the value $m(A, \mathbf{y}_{0,t})$ and we have that $\pi_t^{\mathbf{y}_{0\ t}}$ is a probability measure and $\pi_t^{\mathbf{y}_{0\ t}} f = \mathbb{E}[f(X_t)\,|Y_{0:t} = \mathbf{y}_{0:t}]$ for all $f \in B(\mathbb{R}^{n_x})$.

2.6.2 The recurrence formula for the conditional distribution of the signal

We first need to prove the following lemma

Lemma 2.6.4. *Let $P_{Y_{s\ t}} \in \mathcal{P}\left((\mathbb{R}^{n_y})^{t-s+1}\right)$ be the probability distribution of $Y_{s:t}$ and λ be the Lebesgue measure λ on $\left(((\mathbb{R}^{n_y})^{t-s+1}, \mathcal{B}\left(((\mathbb{R}^{n_y})^{t-s+1}\right)\right)$. Then, for all $0 < s \le t < \infty$, $P_{Y_{s\cdot t}}$ is absolutely continuous with respect to λ and its Radon-Nikodym derivative is*

$$\frac{dP_{Y_{s\ t}}}{d\lambda}(\mathbf{y}_{s:t}) = \Upsilon(\mathbf{y}_{s:t}) \triangleq \int_{(\mathbf{R}^{n_x})^{t-s+1}} \prod_{i=s}^{t} g_i(\mathbf{y}_i - h(i, \mathbf{x}_i)) P_{X_{s,t}}(d\mathbf{x}_{s:t}).$$

Proof. Let $C_{s:t} = C_s \times ... \times C_t$, where C_r are arbitrary Borel sets, $C_r \in \mathcal{B}(\mathbb{R}^{n_y})$ for all $s \le r \le t$. We need to prove that

$$P_{Y_{s\ t}}(C_{s:t}) = P(\{Y_{s:t} \in C_{s:t}\}) = \int_{C_{s\ t}} \Upsilon(\mathbf{y}_{s:t})\, d\mathbf{y}_s ... d\mathbf{y}_t. \tag{2.35}$$

By (2.34), or rather by its vector-valued analogue.

$$P(\{Y_{s:t} \in C_{s:t}\}) = \int_{(\mathbf{R}^{n_x})^{t-s+1}} P(Y_{s:t} \in C_{s:t}\,|X_{s:t} = \mathbf{x}_{s:t}) P_{X_{s\ t}}(d\mathbf{x}_{s:t}), \tag{2.36}$$

and using the fact that, for all i, X_i and W_i are independent and also $W_s, ..., W_t$ are independent, we have from (2.33)

$$\begin{aligned} P\left(Y_{s:t} \in C_{s:t} \left| X_{s:t} = \mathbf{x}_{s:t}\right.\right) &= \mathbb{E}\left[\prod_{i=s}^{t} I_{\{C_i\}}\left(h(i, X_i) + W_i\right) \left| X_{0,t} = \mathbf{x}_{0:t}\right.\right] \\ &= \mathbb{E}\left[\prod_{i=s}^{t} I_{\{\mathbf{C}_i\}}\left(h(i, \mathbf{x}_i) + W_i\right)\right] \\ &= \prod_{i=s}^{t} \mathbb{E}\left[I_{\{\mathbf{C}_i\}}\left(h(i, \mathbf{x}_i) + W_i\right)\right] \\ &= \prod_{i=s}^{t} \int_{C_i} g_i(\mathbf{y}_i - h\left(i, \mathbf{x}_i\right))d\mathbf{y}_i. \end{aligned} \tag{2.37}$$

By combining (2.36) and (2.37) and applying Fubini, we obtain (2.35). □

Proposition 2.6.5. *The conditional distribution of the signal satisfies the following recurrence relations, for $t \geq 0$:*

$$\begin{cases} \frac{d\pi_t^{\mathbf{y}_{0\ t}}}{dp_t^{\mathbf{y}_{0\ t-1}}} &= \frac{g_t^{\mathbf{y}_t}}{\int_{\mathbb{R}^{n_x}} g_t^{\mathbf{y}_t}(\mathbf{x}_t) p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t)} \\ p_{t+1}^{\mathbf{y}_{0\ t}} &= \pi_t^{\mathbf{y}_{0\ t}} K_t \end{cases} .$$

where $g_t^{\mathbf{y}_t} \triangleq g(\mathbf{y}_t - h(t, \cdot))$ and the recurrence is satisfied $P_{Y_{0\ t}}$-almost surely or, equivalently, λ-almost surely.

Proof. We prove first the second identity since it is the simpler of the two. For all $f \in B(\mathbb{R}^{n_x})$, we have, using the Markov property of X, $\mathbb{E}\left[f\left(X_{t+1}\right) \left|\mathcal{F}_t^X\right.\right] = \mathbb{E}\left[f\left(X_{t+1}\right)\left|X_t\right.\right] = K_t f\left(X_t\right)$. Now, since $W_{0:t}$ is independent of $X_{0:t+1}$,we have, using property f. of conditional expectations,

$$\mathbb{E}\left[f\left(X_{t+1}\right)\left|\sigma\left(\mathcal{F}_t^X, \sigma\left(W_{0:t}\right)\right)\right.\right] = \mathbb{E}\left[f\left(X_{t+1}\right)\left|\mathcal{F}_t^X\right.\right].$$

Hence, using property d. of conditional expectations,

$$\begin{aligned} p_{t+1} f &= \mathbb{E}\left[f\left(X_{t+1}\right)\left|Y_{0:t}\right.\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[f\left(X_{t+1}\right)\left|\sigma\left(\mathcal{F}_t^X, \sigma\left(W_{0:t}\right)\right)\right.\right]\left|\sigma\left(Y_{0:t}\right)\right.\right] \\ &= \mathbb{E}\left[K_t f\left(X_t\right)\left|\sigma\left(Y_{0:t}\right)\right.\right] \\ &= \pi_t K_t f \end{aligned}$$

which implies $p_{t+1}^{\mathbf{y}_{0\ t}} = \pi_t^{\mathbf{y}_{0\ t}} K_t$. We prove now the first identity. Let $C_{0:t} = C_0 \times ... \times C_t$ where C_r are arbitrary Borel sets, $C_r \in \mathcal{B}\left(\mathbb{R}^{n_y}\right)$, for all $0 \leq r \leq t$. We need to prove that

$$\int_{C_{0\ t}} \pi_t^{\mathbf{y}_{0\ t}}\left(A\right) P_{Y_{0.t}}\left(d\mathbf{y}_{0:t}\right) = \int_{C_{0\cdot t}} \frac{\int_A g_t^{\mathbf{y}_t}(\mathbf{x}_t) p_t^{\mathbf{y}_{0\ t-1}}\left(d\mathbf{x}_t\right)}{\int_{\mathbb{R}^{n_x}} g_t^{\mathbf{y}_t}(\mathbf{x}_t) p_t^{\mathbf{y}_{0\ t-1}}\left(d\mathbf{x}_t\right)} P_{Y_{0:t}}\left(d\mathbf{y}_{0:t}\right). \tag{2.38}$$

By (2.34) the first term in (2.38) is equal to $P(\{X_t \in A\} \cap \{Y_{0:t} \in C_{0:t}\})$. So, we need to prove the same thing for the second term. Since we have $\sigma(X_{0:t}, W_{0:t-1}) \supset \sigma(X_t, Y_{0:t-1})$, one obtains, using property f. of conditional expectations,

$$P(Y_t \in A_t | X_t, Y_{0:t-1}) = P(P(Y_t \in A_t | X_{0:t}, W_{0:t-1}) | X_t, Y_{0:t-1}), \quad (2.39)$$

and using property d. of conditional expectations and (2.37),

$$\begin{aligned} P(Y_t \in A_t | X_{0:t}, W_{0:t-1}) &= P(Y_t \in A_t | X_{0:t}) \\ &= P\left(Y_{0:t} \in (\mathbb{R}^{n_y})^t \times A_t | X_{0:t}\right) \\ &= \int_{A_t} g_t(\mathbf{y}_t - h(t, X_t)) d\mathbf{y}_t. \end{aligned} \quad (2.40)$$

From (2.39) and (2.40), we derive $P(Y_t \in A_t | X_t, Y_{0:t-1}) = \int_{A_t} g_t(\mathbf{y}_t - h(t, X_t)\, d\mathbf{y}_t$, which gives us

$$P(Y_t \in A_t | X_t = \mathbf{x}_t, Y_{0:t-1} = \mathbf{y}_{0:t-1}) = \int_{A_t} g_t^{\mathbf{y}_t}(\mathbf{x}_t) d\mathbf{y}_t. \quad (2.41)$$

Hence,

$$\begin{aligned} P_{Y_{0\ t}}(A_{0:t}) &= P(\{Y_t \in A_t\} \cap \{X_t \in \mathbf{R}^{n_x}\} \cap \{Y_{0:t-1} \in A_{0:t-1}\}) \\ &= \int_{\mathbf{R}^{n_x} \times A_{0\ t-1}} P(Y_t \in A_t | X_t = \mathbf{x}_t, Y_{0:t-1} = \mathbf{y}_{0:t-1}) P_{X_t, Y_{0\ t-1}}(d\mathbf{x}_t d\mathbf{y}_{0:t-1}) \\ &= \int_{\mathbf{R}^{n_x} \times A_{0\ t-1}} \int_{A_t} g_t^{\mathbf{y}_t}(\mathbf{x}_t) d\mathbf{y}_t p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t) P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}) \\ &= \int_{A_{0\ t}} \int_{\mathbf{R}^{n_x}} g_t^{\mathbf{y}_t}(\mathbf{x}_t) p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t) d\mathbf{y}_t P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}). \end{aligned} \quad (2.42)$$

In (2.42), we used the identity

$$P_{X_t, Y_{0\cdot t-1}}(d\mathbf{x}_t d\mathbf{y}_{0:t-1}) = p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t) P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}), \quad (2.43)$$

which is, again, a consequence of the vector-valued equivalent of (2.34), since for all $A \in \mathcal{B}(\mathbb{R}^{n_x})$ we have

$$\begin{aligned} P((X_t, Y_{0:t-1}) \in A \times C_{0:t-1}) &= \int_{\mathbf{C}_{0\ t-1}} P(X_t \in A | Y_{0:t-1} \in \mathbf{y}_{0:t-1}) P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}) \\ &= \int_{A \times \mathbf{C}_{0\ t-1}} p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t) P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}). \end{aligned}$$

From (2.42) we see that

$$P_{Y_{0\ t}}(d\mathbf{y}_{0:t}) = \int_{\mathbf{R}^{n_x}} g_t^{\mathbf{y}_t}(\mathbf{x}_t) p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t) d\mathbf{y}_t P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}).$$

Hence, the second term in (2.38) is equal to

$$\Xi \triangleq \int_{C_{0\ t}} \int_A g_t^{\mathbf{y}_t}(\mathbf{x}_t) p_t^{\mathbf{y}_{0\ t-1}}(d\mathbf{x}_t) d\mathbf{y}_t P_{Y_{0\ t-1}}(d\mathbf{y}_{0:t-1}),$$

which, in turn, using again (2.41) and (2.43), is equal to

$$
\begin{aligned}
\Xi &= \int_{A \times C_{0\ t-1}} \left(\int_{C_t} g_t^{\mathbf{y}_t}(\mathbf{x}_t) d\mathbf{y}_t \right) p_t^{\mathbf{y}_{0\ t-1}} (d\mathbf{x}_t)\, P_{Y_{0\ t-1}} (d\mathbf{y}_{0:t-1}) \\
&= \int_{A \times C_{0\ t-1}} P\left(Y_t \in C_t | X_t = \mathbf{x}_t, Y_{0:t-1} = \mathbf{y}_{0:t-1}\right) P_{X_t, Y_{0\ t-1}} (d\mathbf{x}_t d\mathbf{y}_{0:t-1}) \\
&= P\left(\{X_t \in A\} \cap \{Y_{0:t} \in C_{0:t}\}\right).
\end{aligned}
$$

□

Acknowledgment. The author would like to thank Jessica Gaines for her careful reading of the manuscript and many useful suggestions.