ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES
UMR CNRS 7641

Limit theorems for weighted samples with applications to Sequential Monte Carlo Methods

R. Douc, E. Moulines

R.I. 580 *july 2005*

Limit theorems for weighted samples with applications to Sequential Monte Carlo Methods

By R. Douc[1] and E. Moulines[2].

Abstract

In the last decade, sequential Monte-Carlo methods (SMC) emerged as a key tool in computational statistics (see for instance Doucet et al. (2001), Liu (2001), Künsch (2001)). These algorithms approximate a sequence of distributions by a sequence of weighted empirical measures associated to a weighted population of particles. These particles and weights are generated recursively according to elementary transformations: mutation and selection. Examples of applications include the sequential Monte-Carlo techniques to solve optimal non-linear filtering problems in state-space models, molecular simulation, genetic optimization, etc.

Despite many theoretical advances (see for instance Gilks and Berzuini (2001), Künsch (2003), Del Moral (2004), Chopin (2004)), the asymptotic property of these approximations remains of course a question of central interest. In this paper, we analyze sequential Monte Carlo methods from an asymptotic perspective, that is, we establish law of large numbers and invariance principle as the number of particles gets large. We introduce the concepts of *weighted sample* consistency and asymptotic normality, and derive conditions under which the mutation and the selection procedure used in the sequential Monte-Carlo build-up preserve these properties. To illustrate our findings, we analyze SMC algorithms to approximate the filtering distribution in state-space models. We show how our techniques allow to relax restrictive technical conditions used in previously reported works and provide grounds to analyze more sophisticated sequential sampling strategies.

Short title: Limit theorems for SMC.

1 Introduction

Sequential Monte Carlo (SMC) refer to a class of methods designed to approximate a *sequence of probability distributions* over a *sequence of probability space* by a set of points, termed *particles* that each have an assigned non-negative weight and are updated recursively in time. SMC methods can be seen as a combination of the sequential importance sampling introduced method in Handschin and Mayne (1969) and the sampling importance resampling algorithm proposed in Rubin (1987); it uses a combination of *mutation* and *selection* steps. In the *mutation* step, the particles are propagated forward in time using proposal kernels

and their importance weights are updated taking into account the targeted distribution. In the *selection* (or *resampling*) step, particles multiply or die depending on their *fitness* measured by their importance weights. Many algorithms have been proposed since, which differ in the way the particles and the importance weights evolve and adapt.

SMC methods have a long history in molecular simulations, where they have been found to be a one of the most powerful means for the simulation and optimization of chain polymers (see for instance Landau and Binder (2000)). SMC methods have more recently emerged as a key tool to solve *on-line* prediction / filtering / smoothing problems in a dynamic system. Simple yet flexible SMC methods have been shown to overcome the numerical difficulties and pitfalls typically encountered with traditional methods based on approximate non-linear filtering (such as the extended Kalman filter or gaussian-sum filters); see for instance Liu and Chen (1998), Liu (2001), Doucet et al. (2001) and Ristic et al. (2004) and the references therein. More recently, SMC methods have been shown to be a promising alternative to Markov Chain Monte Carlo techniques for sampling complex distributions over large dimensional spaces; see for instance Gilks and Berzuini (2001) and Cappé et al. (2005).

In this paper, we focus on the asymptotic behavior of the *weighted particle* approximation as the number of particles tend to infinity. Because the particles interact during the selection steps, they are not independent which make the analysis of particle approximation a challenging area of research. This topic has attracted in recent years a great deal of efforts in recent years making it a daunting task to give credit every contribution. The first rigorous convergence result was obtained in Del Moral (1996), who established the almost-sure convergence of an elementary SMC algorithm (the so-called *bootstrap filter*). A central limit theorem for this algorithm was derived in Del Moral and Guionnet (1999) and refined in Del Moral and Miclo (2000). The proof of the CLT was later simplified and extended to more general SMC algorithms by Künsch (2003) and Chopin (2004). Bounds on the fluctuations of the particle approximations for different norms were reported in Crisan and Lyons (1997), Del Moral and Miclo (2000) and Crisan and Doucet (2002). Del Moral (2004) provides an up-to-date and thorough coverage of recent theoretical developments in this area.

With few exceptions (see Chopin (2004) and to a lesser extent, Crisan and Doucet (2002) and Künsch (2003)), these results apply under simplifying assumptions on the way particles are mutated and selected which restrict the scope of applicability of the results only to the most elementary SMC implementations. In particular, all these results assume that selection is performed at each iteration which implies that the weights are not propagated. This is clearly an annoying limitations since it has been noticed in practice that resampling the particle system at each time step is most often not a clever choice. As discussed for example in (Liu and Chen, 1998, section 2), when the weights are nearly constants, resampling only reduces the number of distinct particles and introduces extra Monte Carlo variations. Resampling should only be applied when the weights are very skewed: carrying many particles with very small importance weights is indeed a waste. Resampling provide chances to *good* particles to replicate and

hence rejuvenate the sampler to produce better *future* particles.

The main purpose of this paper is to derive an asymptotic theory of weighted system of particles. To our best knowledge, limit theorems for such weighted approximations were only considered in Liu and Chen (1998), who mostly sketched consistency proofs. In this paper, we establish both law of large numbers and central limit theorems, under assumptions that are presumably closed from being minimal. These results apply not only to the many different implementations of the SMC algorithms, including rather sophisticated schemes such as the resample and move algorithm Berzuini and Gilks (2001) or the auxiliary particle filter by Pitt and Shephard (1999). They cover resampling schedules (when to resample) that can be either deterministic or dynamic, i.e. based on the distribution of the importance weights at the current iteration. They also cover sampling scheme that can be either simple random sampling (with weights) but also residual sampling (Liu and Chen, 1998) or auxiliary sampling (Pitt and Shephard, 1999). We do not impose a specific structure on the sequence of the target probability measure; therefore our results apply not only to sequential filtering or smoothing of state-space contexts, but also to recent algorithms developed for population Monte Carlo or for molecular simulation.

The paper is organized as follows. In section 2, we introduce the definitions of weighted sample consistency and asymptotic normality; we then discuss the meaning of these definitions in a simple situation. In section 3, we present and discuss the conditions upon which consistency or / and asymptotic normality of a weighted sample is preserved during the mutation and selection steps. In section 4, we apply the result to the estimation of the joint smoothing distribution for a state-space model. In particular, we establish a central limit theorem for a SMC method involving a dynamic resampling scheme. These results are based on general results on triangular array of martingale increments (in the sense of (Shiryaev, 1996, Section II.7)) which are established in section A.

2   Notations and Definitions

All the random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$. A state space X is said to be *general* if it is equipped with a countably generated σ-field $\mathcal{B}(\mathsf{X})$. For a general state space X, we denote by $\mathcal{P}(\mathsf{X})$ the set of probability measures on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and $\mathbb{B}(\mathsf{X})$ (resp. $\mathbb{B}^+(\mathsf{X})$) the set of all $\mathcal{B}(\mathsf{X})/\mathcal{B}(\mathbb{R})$-measurable (resp. non-negative) functions from X to $\mathbb{R}$ equipped with the Borel σ-field $\mathcal{B}(\mathbb{R})$. For any $\mu \in \mathcal{P}(\mathsf{X})$ and $f \in \mathbb{B}(\mathsf{X})$ satisfying $\int_{\mathsf{X}} \mu(dx)|f(x)| < \infty$, $\mu(f)$ denotes $\int_{\mathsf{X}} f(x)\mu(dx)$. Let X and Y be two general state spaces. A kernel V from $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ to $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$ is a map from $\mathsf{X} \times \mathcal{B}(\mathsf{Y})$ into $[0, 1]$ such that for each $A \in \mathcal{B}(\mathsf{Y})$, $x \mapsto V(x, A)$ is a nonnegative bounded measurable function on X and for each $x \in \mathsf{X}$, $A \mapsto V(x, A)$ is a measure on $\mathcal{B}(\mathsf{Y})$. We say that V is finite if $V(x, \mathsf{Y}) < \infty$ for any $x \in \mathsf{X}$; it is Markovian if $V(x, \mathsf{X}) \equiv 1$ for any $x \in \mathsf{X}$. For any function $f \in \mathbb{B}(\mathsf{X} \times \mathsf{Y})$ such that $\int_{\mathsf{Y}} V(x, dy)|f(x, y)| < \infty$ we denote by $V(\cdot, f)$ the function

$$V(\cdot, f) : x \mapsto V(x, f) \stackrel{\text{def}}{=} \int_{\mathsf{Y}} V(x, dy)f(x, y) \ . \tag{1}$$

3

The function $V(\cdot, f)$ belongs to $\mathbb{B}(\mathsf{X})$. We sometimes use the abridged notation Vf instead of $V(\cdot, f)$. For ν a measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, we denote by νV the measure on $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$ defined for any $A \in \mathcal{B}(\mathsf{Y})$ by $\nu V(A) = \int_{\mathsf{X}} \nu(dx) V(x, A)$.

Let Ξ be a general state space, μ be a probability measure on $(\Xi, \mathcal{B}(\Xi))$, $\{M_N\}_{N \geq 0}$ be a sequence of integers, and C be a subset of $\mathsf{L}^1(\Xi, \mu)$. We approximate the probability measure μ by points $\xi_{N,i} \in \Xi$, $i = 1, \ldots, M_N$ associated to non-negative weights $\omega_{N,i} \geq 0$.

Definition 1. *A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ on Ξ is said to be consistent for the probability measure μ and the set $\mathsf{C} \subseteq \mathsf{L}^1(\Xi, \mu)$ if for any $f \in \mathsf{C}$, as $N \to \infty$,*

$$\Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \xrightarrow{\mathrm{P}} \mu(f) , \tag{2}$$

$$\Omega_N^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} \xrightarrow{\mathrm{P}} 0 \quad where \quad \Omega_N = \sum_{i=1}^{M_N} \omega_{N,i} . \tag{3}$$

This definition of weighted sample consistency is similar to the notion of *properly weighted sample* introduced in Liu and Chen (1998). The difference stems from the condition (3) which implies that the contribution of each individual term in the sum vanishes in the limit as $N \to \infty$, a condition often referred as *smallness* in the literature.

Example 1 (Importance Sampling). To illustrate the meaning of these conditions, consider the importance sampling estimator. Let μ (the *target distribution*) and ν (the *proposal distribution*) be a *known* (perhaps up to a normalizing constant) probability distribution on $(\Xi, \mathcal{B}(\Xi))$. Suppose that μ is absolutely continuous with respect to ν and denote by W the *importance function*, $W = \beta \frac{d\mu}{d\nu}$ for some $\beta > 0$. Then, the weighted sample $\{(\xi_{N,i}, W(\xi_{N,i}))\}_{1 \leq i \leq N}$, where $\{\xi_{N,i}\}_{1 \leq i \leq N}$ are i.i.d. ν-distributed is consistent for $\{\mu, \mathsf{L}^1(\Xi, \mu)\}$. Eq. (2) follows from the law of large numbers. For any $\epsilon, C > 0$,

$$N^{-1} \mathrm{E}\left[\max_{1 \leq i \leq N} \omega_{N,i} \right] \leq \epsilon + N^{-1} \sum_{i=1}^{N} \mathrm{E}\left[\omega_{N,i} \mathbf{1}_{\{\omega_{N,i} \geq \epsilon N\}} \right] \leq \epsilon + \nu\left(W \mathbf{1}\{W \geq C\}\right)$$

for sufficiently large N. Because $\nu(W) = \beta$, $\nu(W\mathbf{1}\{W \geq C\})$ converges to zero as $C \to \infty$, and (3) follows because $\Omega_N/N \xrightarrow{\mathrm{P}} \beta$.

Of course, importance sampling is only one technique of the many techniques that can be used to obtain a consistent weighted samples; other approaches will be considered below. For more complex sampling schemes C can be a proper subset of $\mathsf{L}^1(\Xi, \mu)$. In order to obtain sensible results we restrict our attention to classes of sets which are sufficiently rich.

Definition 2 (Proper Set). *A set C of real-valued measurable functions on Ξ is said to be* proper *if the following conditions are satisfied.*

(i) C is a linear space: for any f and g in C and reals α and β, $\alpha f + \beta g \in \mathsf{C}$,

4

(ii) If $g \in \mathsf{C}$ and f is measurable with $|f| \leq |g|$, then $|f| \in \mathsf{C}$,

(iii) For all c, the constant function $f \equiv c$ belongs to C.

For any function f, define the positive and negative parts of it by

$$f^+ \stackrel{\text{def}}{=} f \vee 0 \quad \text{and} \quad f^- \stackrel{\text{def}}{=} (-f) \vee 0 \; ,$$

and note that f^+ and f^- are both dominated by $|f|$. Thus, if $|f| \in \mathsf{C}$ then f^+ and f^- both belong to C and so does $f = f^+ - f^-$. It is easily seen that for any $p \geq 0$ and any measure μ on $(\Xi, \mathcal{B}(\Xi))$, the set $\mathsf{L}^p(\Xi, \mu)$ is proper.

A classical way to strengthen consistency it is to consider distributional convergence of the normalized difference. Recall that (see Aldous and Eagleson (1978), (Hall and Heyde, 1981, chapter 3)) if a sequence of rv's converges in distribution to X, (written $X_N \stackrel{\mathcal{D}}{\longrightarrow} X$), the convergence is said to be *stable* (written $X_N \stackrel{\mathcal{D}}{\longrightarrow} X$ (stably)), if for any set $B \in \mathcal{F}$ and for a countable dense set of points $x \in \mathbb{R}$,

$$\mathrm{P}(X_N \leq x, B) \quad \text{exists}.$$

In other word, for all events $B \in \mathcal{F}$ such that $\mathrm{P}(B) > 0$, the distribution of X_N conditional on B converges in distribution to some (proper) distribution which may depend on B. The convergence is *mixing* (written $X_N \stackrel{\mathcal{D}}{\longrightarrow} X$ (mixing)) if in a stable limit result $X_N \stackrel{\mathcal{D}}{\longrightarrow} X^*$ (stably) the limit random variable X^* can be chosen to be independent of $\mathcal{F}$. Mixing is in particular useful when studying random sum limit theorems.

Let μ be a probability measure on $(\Xi, \mathcal{B}(\Xi))$, γ be a finite measure on $(\Xi, \mathcal{B}(\Xi))$, $\mathsf{A} \subseteq \mathsf{L}^1(\Xi, \mu)$ and $\mathsf{W} \subseteq \mathsf{L}^1(\Xi, \gamma)$ be sets of real-valued measurable functions on Ξ, σ a real non negative function on A, and $\{a_N\}$ be a non-decreasing real sequence diverging to infinity.

Definition 3. *A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ on Ξ is said to be asymptotically normal for $(\mu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\})$ if*

$$a_N \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} \{f(\xi_{N,i}) - \mu(f)\} \stackrel{\mathcal{D}}{\longrightarrow} \mathrm{N}\{0, \sigma^2(f)\} \quad \textit{(mixing)} \qquad \textit{for any } f \in \mathsf{A} \; , \tag{4}$$

$$a_N^2 \Omega_N^{-2} \sum_{i=1}^{M_N} \omega_{N,i}^2 f(\xi_{N,i}) \stackrel{\mathrm{P}}{\longrightarrow} \gamma(f) \qquad\qquad\qquad \textit{for any } f \in \mathsf{W} \tag{5}$$

$$a_N \Omega_N^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} \stackrel{\mathrm{P}}{\longrightarrow} 0 \tag{6}$$

We stress that the rate $\{a_N\}$ can be different from $\sqrt{M_N}$ because of the dependence among the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ introduced by the different transformations undergone by the weighted sample.

Example 2 (Importance Sampling (continued)). The rationale for the conditions (5) and (6) is best understood by considering again the importance sampling example. Assume that $\nu(W^2) < \infty$, where as above $W = \beta \frac{d\mu}{d\nu}$. Define the subsets $\mathsf{A} \subset \mathbb{B}(\Xi)$ and $\mathsf{W} \subset \mathbb{B}(\Xi)$,

$$\mathsf{A} \overset{\text{def}}{=} \left\{ f \in \mathbb{B}(\Xi) : \nu(W^2 f^2) < \infty \right\} \quad \text{and} \quad \mathsf{W} \overset{\text{def}}{=} \left\{ f \in \mathbb{B}(\Xi) : \nu(W^2|f|) < \infty \right\} .$$

For $f \in \mathsf{A}$, define $S_N(f) \overset{\text{def}}{=} \Omega_N^{-1} \sum_{i=1}^{N} \omega_{N,i}[f(\xi_{N,i}) - \mu(f)]$. Using $\Omega_N/N \overset{\text{P}}{\longrightarrow} \beta$ and (Aldous and Eagleson, 1978, Theorem 2), $N^{1/2}S_N(f) \overset{\mathcal{D}}{\longrightarrow} S$ (mixing), where S is Gaussian random variable with zero-mean and variance

$$\sigma^2(f) \overset{\text{def}}{=} \nu \left\{ \left(\frac{d\mu}{d\nu} \right)^2 [f - \mu(f)]^2 \right\} . \tag{7}$$

The law of large numbers for i.i.d. sequence implies that for any $f \in \mathsf{W}$,

$$N\Omega_N^{-2} \sum_{i=1}^{N} \omega_{N,i}^2 f(\xi_{N,i}) = N\Omega_N^{-2} \sum_{i=1}^{N} W^2(\xi_{N,i}) f(\xi_{N,i}) \overset{\text{P}}{\longrightarrow} \gamma(f) = \nu \left[\left(\frac{d\mu}{d\nu} \right)^2 f \right] ,$$

showing (5). For any $\epsilon, C > 0$,

$$N^{-1} \mathrm{E} \left[\max_{1 \leq i \leq N} \omega_{N,i}^2 \right] \leq \epsilon^2 + N^{-1} \sum_{i=1}^{N} \mathrm{E} \left[\omega_{N,i}^2 \mathbf{1}_{\{\omega_{N,i}^2 \geq \epsilon^2 N\}} \right] \leq \epsilon^2 + \nu \left(W^2 \mathbf{1}\{W^2 \geq C\} \right)$$

for sufficiently large N. Because A is a proper set, $\nu \left(W^2 \mathbf{1}\{W \geq C\} \right)$ go to zero as $C \to \infty$. Using again $\Omega_N/N \overset{\text{P}}{\longrightarrow} \beta$, the previous display implies that

$$N^{1/2} \Omega_N^{-1} \max_{1 \leq i \leq N} \omega_{N,i} \overset{\text{P}}{\longrightarrow} 0 , \tag{8}$$

showing (6): any individual term in the sum is small compared to the square root of the variance, a condition which is known to be necessary for a central limit theorem to hold for triangular arrays of independent random variables, see Petrov (1995). Stability in the limit can be used to deal with situations where the number of terms in the sum are random (see Csörgő and Fischler (1971)). This is considered in a companion paper.

3 Main results

To analyze the sequential Monte Carlo methods discussed in the introduction, we now need to study how the mutation step and the resampling step affects the consistent or / and asymptotically normal weighted sample.

3.1 Mutation

To study SISR algorithms, we need first to show that when moving the particles using a Markov kernel and then assigning them appropriately defined importance

weights, we transform a weighted sample consistent (or asymptotically normal) for a distribution ν on a general state space $(\Xi, \mathcal{B}(\Xi))$ into a weighted sample consistent (or asymptotically normal) for a distribution μ on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Let L be a kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ such that $\nu L(\tilde{\Xi}) > 0$ and for any $f \in \mathbb{B}(\tilde{\Xi})$,

$$\mu(f) = \frac{\nu L(f)}{\nu L(\tilde{\Xi})} = \frac{\iint \nu(d\xi) L(\xi, d\tilde{\xi}) f(\tilde{\xi})}{\iint \nu(d\xi) L(\xi, \tilde{\Xi})} \ . \tag{9}$$

There exist of course many such kernels: one may set for example $L(\xi, \tilde{A}) = \mu(\tilde{A})$ for any $\xi \in \Xi$, but, as we will see below, this is not usually the most appropriate choice.

We wish to transform a weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ targeting the distribution ν on $(\Xi, \mathcal{B}(\Xi))$ into a weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ targeting μ on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. The number of particles $\tilde{M}_N$ is set to be $\tilde{M}_N = \alpha_N M_N$ where α_N is the number of offsprings of each particle. The use of multiple offspring for a particle has been suggested in the context of SIR by Rubin (1987): when the mutation step is followed by a resampling step, an increase in the number of distinct particles in the mutation step will typically increase the number of distinct particles **after** the resampling step. In the sequential context, this operation is regarded as a practical mean for contending particle impoverishment. These offsprings are proposed using Markov kernels R, $k = 1, \ldots, \alpha_N$, from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Implicitly, we suppose that sampling from the proposal kernels R is doable.

Most importantly, we assume that for any $\xi \in \Xi$, the probability measure $L(\xi, \cdot)$ on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ is absolutely continuous with respect to R, which we denote $L(\xi, \cdot) \ll R(\xi, \cdot)$ and define

$$W(\xi, \tilde{\xi}) \stackrel{\text{def}}{=} \frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\tilde{\xi}) \tag{10}$$

The new weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ is constructed as follows. We draw new particle positions $\{\tilde{\xi}_{N,j}\}_{1 \leq j \leq \tilde{M}_N}$ conditionally independently given $\mathcal{F}_{N,0} = \sigma(\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N})$ with distribution given for $i = 1, \ldots, M_N$, $k = 1, \ldots, \alpha_N$ and $A \in \mathcal{B}(\tilde{\Xi})$ by

$$\mathrm{P}\left(\tilde{\xi}_{N,\alpha_N(i-1)+k} \in A \,\middle|\, \mathcal{F}_{N,0}\right) = R(\xi_{N,i}, A) \tag{11}$$

and associate to each new particle positions the importance weight:

$$\tilde{\omega}_{N,\alpha_N(i-1)+k} = \omega_{N,i} W(\xi_{N,i}, \tilde{\xi}_{N,\alpha_N(i-1)+k}) \ , \ \text{where} \ i = 1, \ldots, M_N \ , \ k = 1, \ldots, \alpha_N \ . \tag{12}$$

The mutation step is *unbiased* in the sense that, for any $f \in \mathcal{B}(\tilde{\Xi})$ and $i = 1, \ldots, M_N$,

$$\sum_{j=\alpha_N(i-1)+1}^{\alpha_N i} \mathrm{E}\left[\tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \,\middle|\, \mathcal{F}_{N,j-1}\right] = \alpha_N \omega_{N,i} L(\xi_{N,i}, f) \ , \tag{13}$$

7

where for $j = 1, \ldots, \tilde{M}_N$, $\mathcal{F}_{N,j} \stackrel{\text{def}}{=} \mathcal{F}_{N,0} \vee \sigma(\{\tilde{\xi}_{N,l}\}_{1 \leq l \leq j})$. The following theorems state conditions under which the mutation step described above preserves the weighted sample consistency.

Theorem 1. *Let ν be a probability measure on $(\Xi, \mathcal{B}(\Xi))$, μ be a probability measure on $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Let L be a finite kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ satisfying $\nu L(\tilde{\Xi}) > 0$ any $\tilde{A} \in \mathcal{B}(\tilde{\Xi})$, $\mu(\tilde{A}) = \nu L(\tilde{A})/\nu L(\tilde{\Xi})$. Assume that*

   (i) the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, C), where $\mathsf{C} \subset \mathsf{L}^1(\Xi, \nu)$ is a proper set ,

   (ii) the function $L(\cdot, \tilde{\Xi})$ belongs to C ,

   (iii) for any $\xi \in \Xi$, $L(\xi, \cdot) \ll R(\xi, \cdot)$.

Then, $\tilde{\mathsf{C}}$ is a proper set and the weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ defined by (11) *and* (12) *is consistent for $(\mu, \tilde{\mathsf{C}})$, where*

$$\tilde{\mathsf{C}} \stackrel{\text{def}}{=} \{f \in \mathsf{L}^1(\tilde{\Xi}, \mu), \, L(\cdot, |f|) \in \mathsf{C}\} , \tag{14}$$

We now turn to prove the asymptotic normality.

Theorem 2. *Suppose that the assumptions of Theorem 1 hold. Let γ be a finite measure on $(\Xi, \mathcal{B}(\Xi))$, $\mathsf{A} \subseteq \mathsf{L}^1(\Xi, \nu)$, and $\mathsf{W} \subseteq \mathsf{L}^1(\Xi, \gamma)$ be proper sets, σ be a non negative function on A. Assume in addition that*

   (i) the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is asymptotically normal for $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{M_N^{1/2}\})$,

   (ii) the function $R(\cdot, W^2)$ belongs to W .

Then, if $\alpha_N \to \alpha$, the weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{1 \leq i \leq \tilde{M}_N}$ is asymptotically normal for $(\mu, \tilde{\mathsf{A}}, \tilde{\mathsf{W}}, \tilde{\sigma}, \tilde{\gamma}, \{M_N^{1/2}\})$ with

$$\tilde{\mathsf{A}} \stackrel{\text{def}}{=} \left\{ f \in \mathsf{L}^1(\tilde{\Xi}, \mu) : L(\cdot, |f|) \in \mathsf{A} \quad and \quad R(\cdot, W^2 f^2) \in \mathsf{W} \right\} ,$$

$$\tilde{\mathsf{W}} \stackrel{\text{def}}{=} \left\{ f \in \mathsf{L}^1(\tilde{\Xi}, \mu) : R(\cdot, W^2 |f|) \in \mathsf{W} \right\} ,$$

$$\tilde{\sigma}^2(f) \stackrel{\text{def}}{=} \sigma^2\{\tilde{L}[f - \mu(f)]\} + \alpha^{-1} \gamma R\{[\tilde{W}f - R(\cdot, \tilde{W}f)]^2\} , \qquad for \ all \ f \in \tilde{\mathsf{A}} ,$$

$$\tilde{\gamma}(f) \stackrel{\text{def}}{=} \alpha^{-1} \gamma R(\tilde{W}^2 f) , \qquad\qquad\qquad for \ all \ f \in \tilde{\mathsf{W}} ,$$

where $\tilde{L} \stackrel{\text{def}}{=} [\nu L(\tilde{\Xi})]^{-1} L$ and $\tilde{W} = [\nu L(\tilde{\Xi})]^{-1} W$. Moreover, $\tilde{\mathsf{A}}$ and $\tilde{\mathsf{W}}$ are proper sets.

Remark 1. The use of different proposal kernels R_k is sometimes advisable, to add variety in the type of moves that are applied to the particle system. The idea of using multiple type of moves is exploited in particular in adaptive importance sampling (see Douc et al. (2005) for an illustration). The results above can be adapted directly when the number of offsprings $\alpha_N = \alpha$ does not depend

on N. To handle this case, define the average kernel $R = \alpha^{-1} \sum_{k=1}^{\alpha} R_k$ and suppose that for any $\xi \in \Xi$, $L(\xi, \cdot) \ll R(\xi, \cdot)$ (note that the individual proposal kernels need not be absolutely continuous with respect to L). For $i = 1, \ldots, M_N$, $k = 1, \ldots, \alpha_N$ and $A \in \mathcal{B}(\tilde{\Xi})$, we draw the new particle positions according to

$$\mathrm{P}\left(\tilde{\xi}_{N, \alpha_N(i-1)+k} \in A \,\Big|\, \mathcal{F}_{N,0}\right) = R_k(\xi_{N,i}, A) \ . \tag{15}$$

We associate to each particle the importance weight,

$$\tilde{\omega}_{N, \alpha_N(i-1)+k} = \omega_{N,i} W(\xi_{N,i}, \tilde{\xi}_{N, \alpha_N i + k}) \ , \quad \text{where } i = 1, \ldots, M_N \ , \ k = 1, \ldots, \alpha_N \ . \tag{16}$$

where W is defined as in (10). With these definitions and notations, the results of Theorems 1 and 2 continue to hold.

To illustrate the meaning of these statements, consider again the elementary example of importance sampling. It turns out that reweighting the particles without moving them is a particular case of mutation. Thus, Theorem 1 and 2 may apply in this context.

Example 3. Let ν and μ be two probability measures on $(\Xi, \mathcal{B}(\Xi))$. Assume that $\mu \ll \nu$. Set $L(\xi, \cdot) = \beta \frac{d\mu}{d\nu}(\xi)\delta_\xi$, for some $\beta > 0$, where δ_ξ is the Dirac measure. With this definition of L, we obviously have $\mu(\cdot) = \nu L(\cdot)/\nu L(\Xi)$. If $\{\xi_{N,i}\}_{1 \leq i \leq M_N}$ are independent ν-distributed random variables, then the weighted sample $\{(\xi_{N,i}, 1)\}_{1 \leq i \leq M_N}$ is consistent for $(\nu, \mathsf{L}^1(\Xi, \nu))$. Put for $\xi \in \Xi$ and $A \in \mathcal{B}(\Xi)$, $R(\xi, A) = \delta_\xi(A)$. Then, Theorem 1 shows that $\{(\tilde{\xi}_{N,i}, W(\xi_{N,i}, \tilde{\xi}_{N,i}))\}_{1 \leq i \leq M_N}$ where $W(\xi, \tilde{\xi}) = \beta \frac{d\mu}{d\nu}(\tilde{\xi})$ is consistent for $\{\mu, \mathsf{L}^1(\Xi, \mu)\}$. We now turn to check the asymptotic normality. From the CLT for i.i.d. sequence, the weighted sample is $\{(\xi_{N,i}, 1)\}_{1 \leq i \leq M_N}$ is asymptotically normal for $(\nu, \mathsf{L}^2(\mathsf{X}, \nu), \mathsf{L}^1(\mathsf{X}, \nu), \sigma^2, \nu, M_N^{1/2}\}$, where for $f \in \mathsf{A} \stackrel{\text{def}}{=} \mathsf{L}^2(\mathsf{X}, \nu)$, $\nu(f) = \nu\{[f - \nu(f)]^2\}$. Suppose that $\nu(W^2) < \infty$. Because $R(\cdot, W^2) = W^2 \in \mathsf{W} = \mathsf{L}^1(\mathsf{X}, \nu)$, then Theorem 2 shows that $\{(\tilde{\xi}_{N,i}, W(\xi_{N,i}, \tilde{\xi}_{N,i}))\}_{1 \leq i \leq M_N}$ is asymptotically normal for $(\mu, \tilde{\mathsf{A}}, \tilde{\mathsf{C}}, \tilde{\sigma}, \tilde{\gamma}, \sqrt{M_N}\}$, where $\tilde{\mathsf{A}} = \{f \in \mathsf{L}^1(\Xi, \mu), \nu(W^2 f^2) < \infty\}$, $\tilde{\mathsf{W}} = \{f \in \mathsf{L}^1(\Xi, \mu), \nu(W^2|f|) < \infty\}$, $\tilde{\gamma}(f) = \nu[W^2 f]$ for $f \in \tilde{\mathsf{W}}$, and $\tilde{\sigma}^2(f) = \mathrm{Var}_\nu[Wf]$ for $f \in \tilde{\mathsf{A}}$. These expressions are particularly simple because, in this case, $Wf = R(\cdot, Wf)$, which is of course false in general.

3.2 Resampling

Resampling is the second basic type of transformation used in sequential Monte-Carlo methods; resampling converts a weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ targeting a distribution ν into an equally weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ targeting the same distribution ν on $(\Xi, \mathcal{B}(\Xi))$. Note that the number of resampled particles, denoted $\tilde{M}_N$, might well differ from the initial number of particles M_N. In this section, we will focus on importance sampling estimators that satisfy the following *unbiasedness condition*:

$$\text{for any } f \in \mathbb{B}(\Xi), \quad \mathrm{E}\left[\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} f(\tilde{\xi}_{N,i}) \,\Bigg|\, \mathcal{F}_N\right] = \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \ , \tag{17}$$

9

where $\mathcal{F}_N = \sigma(\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$. The unbiasedness condition (17) implies that resampling can only increase the variability of the importance sampling estimator and has thus, in one step, an adversary effect. In the sequential context however, resampling is nevertheless essential by removing particles with small importance weights and producing multiple copies of particles with large importance weights, that help in generating better future samples. There are many different unbiased resampling procedures described in the literature. The most elementary procedure is the so-called *multinomial* sampling in which we draw conditionally independently given $\mathcal{F}_N$ integer-valued random variables $\{I_{N,k}\}_{1 \leq k \leq \tilde{M}_N}$ with distribution

$$\mathrm{P}\left(I_{N,k} = i \mid \mathcal{F}_N\right) = \Omega_N^{-1} \omega_{N,i}, \quad i = 1, \ldots, M_N, \tag{18}$$

and set

$$\tilde{\xi}_{N,i} = \xi_{N,I_{N,i}}, \quad \text{for } i = 1, \ldots, \tilde{M}_N. \tag{19}$$

To bring down the extra-variability incurred by multinomial sampling, other resampling strategies have been considered in the literature. A possible solution, investigated in Liu and Chen (1995), consists in using a combination of deterministic plus residual sampling. This scheme consists in retaining at least $\lfloor \Omega_N^{-1} \tilde{M}_N \omega_{N,i} \rfloor$, $i = 1, \ldots, M_N$ copies of the particles and then reallocating the remaining particles by applying the multinomial sampling procedure but on the residual importance weights, $\tilde{M}_N \Omega_N^{-1} \omega_{N,i} - \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor$. Define $J_{N,0} = 0$ and for $k \geq 1$, $J_{N,k} = \sum_{i=1}^{k} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor$. The residual resampling algorithm is defined as follows:

(i) For $i = 1, \ldots, M_N$, assign

$$\tilde{\xi}_{N,j} = \xi_{N,i}, \quad \text{for all } j = J_{N,i-1} + 1, \ldots, J_{N,i}. \tag{20}$$

(ii) Draw, conditionally independently given $\mathcal{F}_N = \sigma(\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ random variables $\{I_{N,k}\}_{1 \leq k \leq \tilde{M}_N - \bar{M}_N}$ where

$$\bar{M}_N \overset{\text{def}}{=} J_{N,M_N} = \sum_{i=1}^{M_N} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor,$$

$$\mathrm{P}\left(I_{N,k} = i \mid \mathcal{F}_N\right) = \frac{\tilde{M}_N \Omega_N^{-1} \omega_{N,i} - \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor}{\tilde{M}_N - \bar{M}_N} \overset{\text{def}}{=} \tilde{\omega}_{N,i}, \quad i = 1, \ldots, M_N. \tag{21}$$

and set $\tilde{\xi}_{N, \bar{M}_N + k} = \xi_{N, I_{N,k}}$ for all $1 \leq k \leq \tilde{M}_N - \bar{M}_N$.

If the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, C), where C is a proper subset of $\mathbb{B}(\mathsf{X})$, it is a natural question to ask whether $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ is consistent for ν and, if so, what an appropriately defined class of functions on Ξ might be. It happens that a fairly general result can be obtained in this case.

Theorem 3. *Let ν a probability measure on $(\Xi, \mathcal{B}(\Xi))$ and $\mathsf{C} \subseteq \mathsf{L}^1(\Xi, \nu)$ be a proper set of functions. Assume that the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$*

is consistent for (ν, C), *where* $\mathsf{C} \subseteq \mathbb{B}(\Xi)$ *is a proper set. Then, the uniformly weighted sample* $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ *obtained using **any** unbiased resampling scheme (i.e. satisfying (17)) is consistent for* (ν, C).

It is also sensible in this discussion to strengthen the requirement of consistency into asymptotic normality, and again prove that the sampling operation transform an asymptotically normal weighted sample for ν into an asymptotically normal sample for ν (for appropriately defined class of functions, normalizing factors, etc.). We consider first the multinomial sampling algorithm.

Theorem 4. *Suppose that the assumptions of Theorem 3 hold. Let* γ *be a finite measure on* $(\Xi, \mathcal{B}(\Xi))$, $\mathsf{A} \subseteq \mathsf{L}^1(\Xi, \nu)$ *and* $\mathsf{W} \subseteq \mathsf{L}^1(\Xi, \gamma)$ *be proper sets,* σ *be a non negative function on* A *and* $\{a_N\}$ *be a non-negative sequence. Define*

$$\tilde{\mathsf{A}} \overset{\text{def}}{=} \left\{ f \in \mathsf{A}, \, f^2 \in \mathsf{C} \right\} , \tag{22}$$

Assume in addition that

(i) *the weighted sample* $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ *is asymptotically normal for* $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\})$

(ii) $\lim_{N \to \infty} a_N = \infty$ *and* $\lim_{N \to \infty} a_N^2 / \tilde{M}_N = \beta$, *where* $\beta \in [0, \infty]$.

Then $\tilde{\mathsf{A}}$ *is a proper set and the following holds true for the resampled system* $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ *defined as in (18) and (19).*

(i) *If* $\beta < 1$, *then* $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ *is asymptotically normal for* $(\nu, \tilde{\mathsf{A}}, \mathsf{C}, \tilde{\sigma}, \tilde{\gamma}, \{a_N\})$ *with*

$$\tilde{\sigma}^2(f) = \beta \operatorname{Var}_\nu(f) + \sigma^2(f) \quad \text{for any } f \in \tilde{\mathsf{A}} ,$$
$$\tilde{\gamma} = \beta \nu .$$

(ii) *If* $\beta \geq 1$, *then* $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ *is asymptotically normal for* $(\nu, \tilde{\mathsf{A}}, \mathsf{C}, \tilde{\sigma}, \tilde{\gamma}, \{\tilde{M}_N^{1/2}\})$ *with*

$$\tilde{\sigma}^2(f) = \operatorname{Var}_\nu(f) + \beta^{-1} \sigma^2(f) \quad \text{for any } f \in \tilde{\mathsf{A}} ,$$
$$\tilde{\gamma} = \nu .$$

Example 4. Let $\nu, \mu \in \mathcal{P}(\Xi)$ and suppose that $\mu \ll \nu$. As shown in the preceding example, if $\{\xi_{N,i}\}_{1 \leq i \leq M_N}$ is an i.i.d. sample from ν, then the weighted sample $\{(\xi_{N,i}, W(\xi_{N,i}))\}_{1 \leq i \leq M_N}$, where $W \propto d\mu/d\nu$ is consistent for $\{\mu, \mathsf{L}^1(\Xi, \mu)\}$ and asymptotically normal for $(\mu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \sqrt{M_N}\}$, where $\mathsf{A} = \{f \in \mathsf{L}^1(\Xi, \mu), \nu [W^2 f^2] < \infty\}$, $\mathsf{W} = \{f \in \mathbb{B}(\Xi), \nu [W^2 |f|] < \infty\}$, $\gamma(f) = \nu [W^2 f]$ for $f \in \mathsf{W}$, and $\sigma^2(f) = \operatorname{Var}_\nu [Wf]$ for $f \in \mathsf{A}$. The sampling importance resampling procedure outlined in Rubin (1987) consists in resampling the weighted sample $\{(\xi_{N,i}, W(\xi_{N,i}))\}_{1 \leq i \leq M_N}$. It is recommended that the number $\tilde{M}_N$ of resampled particles be much less than M_N the number of elements in the importance weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ (in (Gelman and Rubin, 1992, pp.

459) it is suggested to sample $\tilde{M}_N = 10$ out of $M_N = 1000$ elements). This suggestion is supported by the results above. Theorem 3 shows that the uniformly weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \le i \le \tilde{M}_N}$ is consistent for $\{\mu, \mathsf{L}^1(\Xi, \mu)\}$. Assume that $\beta \ge 1$, Theorem 4 shows that $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \le i \le \tilde{M}_N}$ is asymptotically normal for $(\mu, \tilde{\mathsf{A}}, \mathsf{L}^1(\Xi, \mu), \tilde{\sigma}, \mu, \sqrt{\tilde{M}_N})$ where $\tilde{\mathsf{A}} \stackrel{\text{def}}{=} \{f \in \mathsf{L}^1(\Xi, \mu), \nu((W + W^2)f^2) < \infty\}$ and $\tilde{\sigma}$ is defined by $\tilde{\sigma}^2(f) = \mathrm{Var}_\mu(f) + \beta^{-1} \mathrm{Var}_\nu[Wf]$. Suppose now that $\lim_{N \to \infty} M_N / \tilde{M}_N = \infty$ and thus $\beta = \infty$. Then, the limiting variance is the basic Monte Carlo variance $\mathrm{Var}_\mu(f)$: the resampled particles can be thought of as an i.i.d. sample from the target distribution μ, providing thus a simple theoretical justification for the SIR procedure.

The analysis of the deterministic plus residual sampling is more involved. To carry out the analysis, it is required to specify more precisely the importance weight.

Theorem 5. *Let k be an integer, ν be probability measure on $(\Xi, \mathcal{B}(\Xi))$, γ be a finite measure on $(\Xi, \mathcal{B}(\Xi))$, $\mathsf{A} \subseteq \mathsf{L}^1(\Xi, \nu)$, $\mathsf{C} \subseteq \mathsf{L}^1(\Xi, \nu)$ and $\mathsf{W} \subseteq \mathsf{L}^1(\mathsf{X}, \gamma)$ be proper sets, σ be a non negative function on A and Φ be a non negative function on Ξ. Assume that*

(i) $\{(\xi_{N,i}, \Phi(\xi_{N,i}))\}_{1 \le i \le M_N}$ is consistent for (ν, C) and asymptotically normal for $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\})$,

(ii) $\lim_{N \to \infty} \tilde{M}_N / M_N = \ell$, $\lim_{N \to \infty} a_N = \infty$ and $\lim_{N \to \infty} a_N^2 / \tilde{M}_N = \beta$, where $\ell \in [0, \infty]$ and $\beta \in [0, \infty]$,

(iii) $1/\Phi \in \mathsf{C}$, and $\nu(\ell\nu(1/\Phi)\Phi \in \mathbb{N} \cup \{\infty\}) = 0$.

Define

$$W[\ell, \Phi] \stackrel{\text{def}}{=} \begin{cases} 1 - \frac{\ell\nu(1/\Phi)\Phi}{[\ell\nu(1/\Phi)\Phi]} & \ell < \infty \\ 0 & \ell = \infty . \end{cases} \tag{23}$$

Then, the following holds true for the uniformly weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \le i \le \tilde{M}_N}$ defined by the deterministic-plus-residual sampling ((20) and (21))

(i) If $\beta < 1$, then $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \le i \le \tilde{M}_N}$ is asymptotically normal for $(\nu, \tilde{\mathsf{A}}, \mathsf{C}, \tilde{\sigma}, \tilde{\gamma}, \{a_N\})$ where $\tilde{\mathsf{A}}$ is given by (22) and

$$\tilde{\sigma}^2(f) \stackrel{\text{def}}{=} \beta\nu \left\{ W[\ell, \Phi] \left(f - \frac{\nu\{W[\ell, \Phi]f\}}{\nu\{W[\ell, \Phi]\}} \right)^2 \right\} + \sigma^2(f) , \quad f \in \tilde{\mathsf{A}} ,$$

$$\tilde{\gamma} \stackrel{\text{def}}{=} \beta\nu .$$

(ii) If $\beta \ge 1$, then $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \le i \le \tilde{M}_N}$ is asymptotically normal for $(\nu, \tilde{\mathsf{A}}, \mathsf{C}, \tilde{\sigma}, \tilde{\gamma}, \{\tilde{M}_N^{1/2}\})$ where $\tilde{\mathsf{A}}$ is given by (22) and

$$\tilde{\sigma}^2(f) \stackrel{\text{def}}{=} \nu \left\{ W[\ell, \Phi] \left(f - \frac{\nu\{W[\ell, \Phi]f\}}{\nu\{W[\ell, \Phi]\}} \right)^2 \right\} + \beta^{-1}\sigma^2(f) , \quad f \in \tilde{\mathsf{A}} ,$$

$$\tilde{\gamma} \stackrel{\text{def}}{=} \nu ,$$

12

Because $W[\ell, \Phi] \leq 1$, for any $f \in \tilde{\mathsf{A}}$,

$$\nu\left\{W[\ell, \Phi]\left(f - \frac{\nu\{W[\ell, \Phi]f\}}{\nu\{W[\ell, \Phi]\}}\right)^2\right\} = \inf_{c \in \mathbb{R}} \nu\left\{W[\ell, \Phi](f - c)^2\right\}$$

$$\leq \inf_{c \in \mathbb{R}} \nu\left\{(f - c)^2\right\} = \mathrm{Var}_\nu(f) ,$$

showing that the variance of the residual plus deterministic sampling is always lower than that of the multinomial sampling. These results extend (Chopin, 2004, Theorem 2) that derive an expression of the variance of the residual sampling in a specific case. Note however the assumption Theorem 5-(iii) is missing in the statement of (Chopin, 2004, Theorem 2). This assumption cannot be relaxed, as shown in Section D.

4 An application to state-space models

In this section, we apply the results developed above to state-space models. State-space model has become a powerful tool for dynamic systems. In this model, an underlying state of interest changes over time and measurements are taken to enable inferences to be made about the state. The state process $\{X_k\}_{k \geq 1}$ is a Markov chain on a general state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ with initial distribution χ and kernel Q. The observations $\{Y_k\}_{k \geq 1}$ are random variables taking value in a general state space $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$ that are independent conditionally on the state sequence $\{X_k\}_{k \geq 1}$; in addition, there exists a measure λ on $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$, and a transition density function $x \mapsto g(x, y)$, referred to as the likelihood, such that $\mathrm{P}\left(Y_k \in A \mid X_k = x\right) = \int_A g(x, y)\lambda(dy)$, for all $A \in \mathsf{Y}$. The kernel Q and the likelihood functions $x \mapsto g(x, y)$ are assumed to be known. These quantities could be time-dependent. Such models have been used in a wide range of applications, including quantitative finance, engineering and natural sciences and consequently have become of increasing interest to statisticians (see for instance (Liu, 2001, chapter 3,4) and Künsch (2001) for an introduction to that field).

In this paper, we are primarily concerned with the recursive estimation of the (joint) smoothing distribution, i.e. the conditional distribution of the state sequence $X_{1:k} \overset{\mathrm{def}}{=} (X_1, \ldots, X_k)$ given the σ-algebra generated by the observed process from time 1 to k, i.e. one is interested in estimating the conditional expectation

$$\phi_{\chi,k}(y_{1:k}, f) \overset{\mathrm{def}}{=} \mathrm{E}\left[f(X_{1:k}) \mid Y_{1:k} = y_{1:k}\right] , \quad \text{where } f \in \mathbb{B}(\mathsf{X}^k) . \qquad (24)$$

We shall consider the case in which the observations have an arbitrary but fixed value $y_{1:k}$, and we drop them from the notations. We denote $g_k(x) = g(x, y_k)$. A straightforward application of the Bayes formula, shows that, for any $f \in \mathcal{B}^+(\mathsf{X}^k)$,

$$\phi_{\chi,k}(f) = \frac{\int \cdots \int \phi_{\chi,k-1}(d\tilde{x}_{1:k-1}) Q(\tilde{x}_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(\tilde{x}_{1:k})}{\int \cdots \int \phi_{\chi,k-1}(d\tilde{x}_{1:k-1}) Q(\tilde{x}_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k)} . \qquad (25)$$

In practice, these computations can only be performed in closed-form for linear Gaussian models and for finite state-space models. Many approximations schemes have been proposed in the literature to tackle this problem, but most of the known solutions either suffer from poor performance and / or instability (Extended Kalman Filter, Gaussian sum filter, etc.), or are prohibitively expensive to implement (grid-based solutions). In the Monte-Carlo framework, we approximate the posterior distribution $\phi_{\chi,k}$ at each iteration k by means of a weighted sample $\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq k \leq M_N}$, where the superscript k indicates the iteration index. Note that $\phi_{\chi,k}$ is defined on $(\mathsf{X}^{k+1}, \mathcal{B}(\mathsf{X}^{k+1}))$ and thus that the points $\xi_{N,i}^{(k)}$ belong to X^{k+1} (these are often referred to as *path particle* in the literature; see Del Moral (2004)).

To apply the results presented in section 3, it is first required to define a transition kernel L_{k-1} satisfying (9) with $\nu = \phi_{\chi,k-1}$, $(\Xi, \mathcal{B}(\Xi)) = (\mathsf{X}^k, \mathcal{B}(\mathsf{X}^k))$, $\mu = \phi_{\chi,k}$ and $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi})) = (\mathsf{X}^{k+1}, \mathcal{B}(\mathsf{X}^{k+1}))$, i.e. for any $f \in \mathbb{B}^+(\mathsf{X}^{k+1})$,

$$\phi_{\chi,k}(f) = \frac{\int \cdots \int \phi_{\chi,k-1}(dx_{1:k-1}) L_{k-1}(x_{1:k-1}, d\tilde{x}_{1:k}) f(\tilde{x}_{1:k})}{\int \cdots \int \phi_{\chi,k-1}(dx_{1:k-1}) L_{k-1}(dx_{1:k-1}, \mathsf{X}^{k+1})} \qquad (26)$$

In the second step, we must choose a proposal kernel R_{k-1} satisfying

$$L_{k-1}(x_{1:k-1}, \cdot) \ll R_{k-1}(x_{1:k-1}, \cdot), \quad \text{for any } x_{1:k-1} \in \mathsf{X}^k. \qquad (27)$$

There are many possible choices, which will be associated with different algorithms proposed in the literature. The first obvious choice consists in setting, for any $f \in \mathcal{B}(\mathsf{X}^k)$,

$$L_{k-1}(x_{1:k-1}, f) = T_{k-1}(x_{1:k-1}, f) = \int Q(x_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(x_{1:k-1}, \tilde{x}_k) . \qquad (28)$$

Note that, by construction, the kernel $T_{k-1}(x_{1:k-1}, \cdot)$ leaves the coordinates $x_{1:k-1}$. This corresponds to the so-called sequential importance sampling algorithm. The first obvious choice is that of setting $R_{k-1} = Q_{k-1}$, where the kernel Q_{k-1} is defined, for any $f \in \mathbb{B}^+(\mathsf{X}^k)$,

$$Q_{k-1}(x_{1:k-1}, f) = \int Q(x_{k-1}, d\tilde{x}_k) f(x_{1:k-1}, \tilde{x}_k) . \qquad (29)$$

With this particular choice, for any $x_{1:k-1} \in \mathsf{X}^k$ and $\tilde{x}_{1:k} \in \mathsf{X}^k$,

$$\frac{dT_{k-1}(x_{1:k-1}, \cdot)}{dQ_{k-1}(x_{1:k-1}, \cdot)}(\tilde{x}_{1:k}) \propto g_k(\tilde{x}_k) . \qquad (30)$$

Note that the incremental weight $g_k(\tilde{x}_k)$ *does not depend on* $x_{1:k-1}$, that is, on the past path particle. The use of the prior kernel $R_{k-1} = Q_{k-1}$ is popular because sampling from the prior kernel Q_{k-1} is often straightforward, and computing the incremental weight simply amounts to evaluating the conditional likelihood of the new observation given the current particle position. Often, significant gain can be expected by taking into account the new observation in the mutation kernel R_{k-1} (see for instance Liu and Chen (1998), Doucet et al. (2000) and Doucet

et al. (2001)). A possible choice is to set the instrumental kernel $R_{k-1}(x_{1:k-1}, \cdot)$ as the conditional distribution of the state X_k given the previous state and the current observation, i.e. to set $R_{k-1} = T_{k-1}^\star$, where for any $f \in \mathbb{B}^+(\mathsf{X}^k)$ and $x_{1:k-1} \in \mathsf{X}^{k-1}$,

$$T_{k-1}^\star(x_{1:k-1}, f) = \frac{\int Q(x_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(x_{1:k-1}, \tilde{x}_k)}{\int Q(x_{k-1}, dx_k) g_k(x_k)} \ ,$$

which is often termed the *optimal kernel*. The incremental importance weight is thus given by

$$\frac{dT_{k-1}(x_{1:k-1}, \cdot)}{dT_{k-1}^\star(x_{1:k-1}, \cdot)}(\tilde{x}_{1:k}) = \int Q(x_{k-1}, dx_k) g_k(x_k) \ , \tag{31}$$

which depends on the last coordinate of the past value of the path particle $x_{1:k-1}$ but not on the current value of the particle offspring $\tilde{x}_{1:k}$. Most often, it is not easy to sample from $T_{k-1}^\star$ or to compute the importance weights $T_{k-1}(\xi_{N,j}^{(k-1)}, \mathsf{X})$. A possible solution consists in sampling from $T_{k-1}^\star$ by accept-reject (see Tanizaki (1999) and Künsch (2003)); another solution is to approximate the optimal kernel by using more or less sophisticated tricks (see for instance Shephard and Pitt (1997), Doucet et al. (2000), Tanizaki (2001)).

Other choices for L_{k-1} and R_{k-1} are possible. For example, Gilks and Berzuini (2001) have introduced a variant of this algorithm, the *resample-move* algorithm, in which the whole path of the particles are mutated. This technique allows to combat the progressive impoverishment of the system of particles as the dynamic process evolves. The construction goes as follows. Let P_{k-1} be a kernel on $(\mathsf{X}^k, \mathcal{B}(\mathsf{X}^k))$ such that $\phi_{\chi,k-1}$ is an invariant distribution for P_{k-1}, i.e. $\phi_{\chi,k-1} P_{k-1} = \phi_{\chi,k-1}$. Such kernel can be constructed easily by using for example the Metropolis-Hastings construction (which automatically guarantees the detailed balance condition). Then, define the kernel L_{k-1}, for any $f \in \mathcal{B}(\mathsf{X}^k)$ and $x_{1:k-1} \in \mathsf{X}^{k-1}$, by

$$L_{k-1}(x_{1:k-1}, f) = \int \cdots \int P_{k-1}(x_{1:k-1}, d\tilde{x}_{1:k-1}) Q(\tilde{x}_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(\tilde{x}_{1:k}) \ . \tag{32}$$

Because P_{k-1} is invariant for $\phi_{\chi,k-1}$, for any $f \in \mathcal{B}(\mathsf{X}^k)$,

$$\int \cdots \int \phi_{\chi,k-1}(dx_{1:k-1}) P_{k-1}(x_{1:k-1}, d\tilde{x}_{1:k-1}) Q(\tilde{x}_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(\tilde{x}_{1:k}) =$$

$$\int \cdots \int \phi_{\chi,k-1}(d\tilde{x}_{1:k-1}) Q(\tilde{x}_{k-1}, d\tilde{x}_k) g_k(\tilde{x}_k) f(\tilde{x}_{1:k}) \ ,$$

showing that (26). With this definition of L_{k-1}, the condition (27) is satisfied for example by the kernel R_{k-1} given, for any $f \in \mathcal{B}(\mathsf{X}^k)$,

$$R_{k-1}(x_{1:k-1}, f) = \int \cdots \int P_{k-1}(x_{1:k-1}, d\tilde{x}_{1:k-1}) Q(\tilde{x}_{k-1}, d\tilde{x}_k) f(\tilde{x}_{1:k}) \ .$$

Other possible choices and implementations issues are given in Gilks and Berzuini (2001).

15

We proceed from the weighted sample $\{(\xi_{N,i}^{(k-1)}, \omega_{N,i}^{(k-1)})\}_{1 \leq i \leq M_N}$ targeting $\phi_{\chi,k-1}$ to $\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$ targeting $\phi_{\chi,k}$ as follows. To keep the discussion simple, it is assumed that each particle gives birth to a single offspring. In the mutation step, we draw $\{\tilde{\xi}_{N,i}^{(k)}\}_{1 \leq i \leq M_N}$ conditionally independently given $\mathcal{F}_N^{(k-1)}$ with distribution given, for any $f \in \mathbb{B}^+(\mathsf{X}^k)$ by

$$\mathrm{E}\left[f(\tilde{\xi}_{N,i}^{(k)}) \,\Big|\, \mathcal{F}_N^{(k-1)} \right] = R_{k-1}(\xi_{N,i}^{(k-1)}, f) = \int \cdots \int R_{k-1}(\xi_{N,i}^{(k-1)}, dx_{1:k}) f(x_{1:k}) , \tag{33}$$

where $i = 1, \ldots, M_N$. Next we assign to the particle $\tilde{\xi}_{N,i}^{(k)}$, $i = 1, \ldots, M_N$, the importance weight

$$\tilde{\omega}_{N,i}^{(k)} = \omega_{N,i}^{(k-1)} W_{k-1}(\xi_{N,i}^{(k-1)}, \tilde{\xi}_{N,i}^{(k)}) \ \text{with} \ W_{k-1}(x_{1:k-1}, \tilde{x}_{1:k}) = \frac{dL_{k-1}(x_{1:k-1}, \cdot)}{dR_{k-1}(x_{1:k-1}, \cdot)}(\tilde{x}_{1:k}) . \tag{34}$$

Instead of resampling at each iteration (which is the assumption upon which most of the asymptotic analysis have been carried out so far), we rejuvenate the particle system only when the importance weights are too skewed. As discussed in (Kong et al., 1994, section 4), a sensible approach is to try to monitor the coefficient of variations of weight, defined by

$$\left[\mathrm{CV}_N^{(k)} \right]^2 \overset{\text{def}}{=} \frac{1}{M_N} \sum_{i=1}^{M_N} \left(\frac{M_N \tilde{\omega}_{N,i}^{(k)}}{\widetilde{\Omega}_N^{(k)}} - 1 \right)^2 .$$

The coefficient of variation is minimal when the normalized importance weights $\tilde{\omega}_{N,i}^{(k)}/\widetilde{\Omega}_N^{(k)}$, $i = 1, \ldots, M_N$, are all equal to $1/M_N$, in which case $\mathrm{CV}_N^{(k)} = 0$. The maximal value of $\mathrm{CV}_N^{(k)}$ is $\sqrt{M_N - 1}$, which corresponds to one of the normalized weights being one and all others being null. Therefore, the coefficient of variation is often interpreted as a measure of the number of ineffective particles (those that do not significantly contribute to the estimate). A related criterion with a simpler interpretation is the so-called *effective sample size* $\mathrm{ESS}_N^{(k)}$ (Liu, 1996), defined as

$$\mathrm{ESS}_N^{(k)} = \left[\sum_{i=1}^{M_N} \left(\frac{\tilde{\omega}_{N,i}^{(k)}}{\widetilde{\Omega}_N^{(k)}} \right)^2 \right]^{-1} , \tag{35}$$

which varies between 1 (all weights null but one) and M_N (equal weights). It is straightforward to check the relation

$$\mathrm{ESS}_N^{(k)} = \frac{M_N}{1 + \left[\mathrm{CV}_N^{(k)} \right]^2} .$$

The effective sample size may be understood as a proxy for the equivalent number of i.i.d. samples at time k. Some additional insights and heuristics about the coefficient of variation are given by Liu and Chen (1995). If the coefficient of variation of the importance weights (or equivalently, if the ratio of the effective sample size

to the total number of particles, $\mathrm{ESS}_N^{(k)}/M_N$) crosses a threshold we rejuvenate the particle system. More precisely if at time k, $\mathrm{CV}_N^{(k)} \geq \kappa$ (such time index are called *dynamic check points* in Liu et al. (2001)), we draw $I_k^{N,1}, \ldots, I_k^{N,M_N}$ conditionally independently given $\widetilde{\mathcal{F}}_N^{(k)} = \mathcal{F}_N^{(k-1)} \vee \sigma\left(\{\widetilde{\xi}_{N,i}^{(k)}, \widetilde{\omega}_{N,i}^{(k)}\}_{1 \leq i \leq M_N}\right)$, with distribution

$$\mathrm{P}\left(I_k^{N,i} = j \,\middle|\, \widetilde{\mathcal{F}}_N^{(k)}\right) = \widetilde{\omega}_{N,j}^{(k)}/\widetilde{\Omega}_N^{(k)}, \quad i = 1, \ldots, M_N, j = 1, \ldots, M_N \qquad (36)$$

and we set

$$\xi_{N,i}^{(k)} = \widetilde{\xi}_{N,I_k^{N,i}}^{(k)} \quad \text{and} \quad \omega_{N,i}^{(k)} = 1, i = 1, \ldots, M_N. \qquad (37)$$

If $\mathrm{CV}_N^{(k)} < \kappa$, we simply copy the mutated path particles: $(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)}) = (\widetilde{\xi}_{N,i}^{(k)}, \widetilde{\omega}_{N,i}^{(k)})$, $i = 1, \ldots, M_N$. In both cases, we set $\mathcal{F}_N^{(k)} = \widetilde{\mathcal{F}}_N^{(k)} \vee \sigma(\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$. We consider here only multinomial resampling, but the deterministic plus residual sampling can be applied as well.

Theorem 6. *For any $k > 0$, let L_k and R_k be transition kernels from $(\mathsf{X}^k, \mathcal{B}(\mathsf{X}^k))$ to $(\mathsf{X}^{k+1}, \mathcal{B}(\mathsf{X}^{k+1}))$ satisfying (26) and (27), respectively. Assume that the equally weighted sample $\{(\xi_{N,i}^{(1)}, 1)\}_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,1}, \mathsf{L}^1(\mathsf{X}, \phi_{\chi,1})\}$ and asymptotically normal for $(\phi_{\chi,1}, \mathsf{A}_1, \mathsf{W}_1, \sigma_1, \phi_{\chi,1}, \{M_N^{1/2}\})$ where A_1 and W_1 are proper sets and define recursively (A_k) and (W_k) by*

$$\mathsf{A}_k \overset{\mathrm{def}}{=} \left\{ f \in \mathsf{L}^2\left(\mathsf{X}^k, \phi_{\chi,k}\right), \quad L_{k-1}(\cdot, f) \in \mathsf{A}_{k-1}, R_{k-1}(\cdot, W_{k-1}^2 f^2) \in \mathsf{W}_{k-1} \right\},$$

$$\mathsf{W}_k \overset{\mathrm{def}}{=} \left\{ f \in \mathsf{L}^1\left(\mathsf{X}^k, \phi_{\chi,k}\right), \quad R_{k-1}(\cdot, W_{k-1}^2 |f|) \in \mathsf{W}_{k-1} \right\}.$$

Assume in addition that for any $k \geq 1$, $R_k(\cdot, W_k^2) \in \mathsf{W}_k$. Then for any $k \geq 1$, (A_k) and (W_k) are proper sets and $\{(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,k}, \mathsf{L}^1(\mathsf{X}, \phi_{\chi,k})\}$ and asymptotically normal for $(\phi_{\chi,k}, \mathsf{A}_k, \mathsf{W}_k, \sigma_k, \gamma_k, \{M_N^{1/2}\})$, where the functions σ_k and the measure γ_k are given by

$$\sigma_k^2(f) = \varepsilon_k \,\mathrm{Var}_{\phi_{\chi,k}}(f)$$

$$+ \frac{\sigma_{k-1}^2(L_{k-1}\{f - \phi_{\chi,k}(f)\}) + \gamma_{k-1}R_{k-1}\left[\{W_{k-1}f - R_{k-1}(\cdot, W_{k-1}f)\}^2\right]}{\{\phi_{\chi,k-1}L_{k-1}(\mathsf{X}^k)\}^2}$$

$$\gamma_k(f) = \varepsilon_k \phi_{\chi,k} + (1 - \varepsilon_k)\frac{\gamma_{k-1}R_{k-1}(W_{k-1}^2 f)}{[\phi_{\chi,k-1}L_{k-1}(\mathsf{X}^k)]^2}$$

where W_k is defined in (34) and

$$\varepsilon_k \overset{\mathrm{def}}{=} \mathbf{1}\left\{[\phi_{\chi,k-1}L_{k-1}(\mathsf{X}^k)]^{-2}\gamma_{k-1}R_{k-1}(W_{k-1}^2) \geq 1 + \kappa^2\right\}$$

Proof. Recall that the algorithm proceeds as follows. The weighted sample $\{(\xi_{N,i}^{(k-1)}, \omega_{N,i}^{(k-1)})\}_{1 \leq i \leq M_N}$ are mutated into $\{(\widetilde{\xi}_{N,i}^{(k)}, \widetilde{\omega}_{N,i}^{(k)})\}_{1 \leq i \leq M_N}$ as described by (33) and (34) (see also section 3.1). The resulting particles are then resampled according to the multinomial resampling algorithm (as described in Section

17

3.2) so as to obtain a family of equally weighted particles that we denote by $\{(\bar{\xi}_{N,i}^{(k)}, 1)\}_{1 \leq i \leq M_N}$.

$$(\xi_{N,i}^{(k-1)}, \omega_{N,i}^{(k-1)}) \overset{\text{mutation}}{\longrightarrow} (\tilde{\xi}_{N,i}^{(k)}, \tilde{\omega}_{N,i}^{(k)}) \overset{\text{resampling}}{\longrightarrow} (\bar{\xi}_{N,i}^{(k)}, 1)$$

We then assign

$$(\xi_{N,i}^{(k)}, \omega_{N,i}^{(k)}) = \begin{cases} (\tilde{\xi}_{N,i}^{(k)}, \tilde{\omega}_{N,i}^{(k)}) & \text{if } \mathrm{CV}_N^{(k)} < \kappa \\ (\bar{\xi}_{N,i}^{(k)}, 1) & \text{otherwise} \end{cases}$$

The proof now follows by induction. Assume that for some $k > 1$, the weighted sample $\{(\xi_{N,i}^{(k-1)}, \omega_{N,i}^{(k-1)})\}_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,k-1}, \mathsf{L}^1\,(\mathsf{X}, \phi_{\chi,k-1})\}$ and asymptotically normal for $(\phi_{\chi,k-1},\, \mathsf{A}_{k-1},\, \mathsf{W}_{k-1},\, \sigma_{k-1},\, \gamma_{k-1},\, \{M_N^{1/2}\})$. By Theorem 1 and 2, $(\tilde{\xi}_{N,i}^{(k)}, \tilde{\omega}_{N,i}^{(k)})_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,k}, \mathsf{L}^1\,(\mathsf{X}, \phi_{\chi,k})\}$ and asymptotically normal for $(\phi_{\chi,k},\, \tilde{\mathsf{A}}_k,\, \tilde{\mathsf{W}}_k,\, \tilde{\sigma}_k,\, \tilde{\gamma}_k,\, \{M_N^{1/2}\})$ where $\tilde{\mathsf{A}}_k,\, \tilde{\mathsf{W}}_k,\, \tilde{\sigma}_k,\, \tilde{\gamma}_k$, are defined from $\mathsf{A}_k,\, \mathsf{W}_k,\, \sigma_k,\, \gamma_k$, using Theorem 2. And by Theorem 3 and 4, $(\bar{\xi}_{N,i}^{(k)}, 1)_{1 \leq i \leq M_N}$ is consistent for $\{\phi_{\chi,k}, \mathsf{L}^1\,(\mathsf{X}, \phi_{\chi,k})\}$ and asymptotically normal for $(\phi_{\chi,k},\, \bar{\mathsf{A}}_k,\, \bar{\mathsf{W}}_k,\, \bar{\sigma}_k,\, \bar{\gamma}_k,\, \{M_N^{1/2}\})$ where $\bar{\mathsf{A}}_k,\, \bar{\mathsf{W}}_k,\, \bar{\sigma}_k,\, \bar{\gamma}_k$, are defined from $\tilde{\mathsf{A}}_k$, $\tilde{\mathsf{W}}_k,\, \tilde{\sigma}_k,\, \tilde{\gamma}_k$, using Theorem 4. The asymptotic normality of $(\tilde{\xi}_{N,i}^{(k)}, \tilde{\omega}_{N,i}^{(k)})_{1 \leq i \leq M_N}$ and $(\bar{\xi}_{N,i}^{(k)}, 1)_{1 \leq i \leq M_N}$, combined with:

$$\left[\mathrm{CV}_N^{(k)}\right]^2 = M_N \sum_{i=1}^{M_N} \left(\frac{\tilde{\omega}_{N,i}^{(k)}}{\widetilde{\Omega}_N^{(k)}}\right)^2 - 1 \overset{\mathrm{P}}{\longrightarrow} \tilde{\gamma}_k(1) - 1 \quad \text{and} \quad \epsilon_k = \mathbf{1}\left\{\tilde{\gamma}_k(1) - 1 > \kappa^2\right\},$$

complete the proof.

$\square$

A   Weak Limits Theorems for Triangular Array

This section summarizes various limit theorems for triangular arrays of dependent random variables To keep the technical assumptions in our main theorems minimal, we derive these limit theorems under assumptions that are weaker than the one typically used in the literature (see Hall and Heyde (1980)). Through our general results can be obtained by weakening the assumptions and adapting the proofs for triangular array of dependent random variables given in Dvoretzky (1972) and McLeish (1974), we prefer to develop them here independently. We hope that the greater accessibility of the proofs will compensate for this sacrifice of brevity. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, let X be a random variable and let $\mathcal{G}$ be a a sub-σ field of $\mathcal{F}$. Define $X^+ \overset{\mathrm{def}}{=} \max(X, 0)$ and $X^- \overset{\mathrm{def}}{=} -\min(X, 0)$. Following (Shiryaev, 1996, Section II.7), if

$$\min(\mathrm{E}\left[X^+ \,\middle|\, \mathcal{G}\right], \mathrm{E}\left[X^- \,\middle|\, \mathcal{G}\right]) < \infty \ , \ \text{P-a.s.}$$

(a version of) the conditional expectation of X given $\mathcal{G}$ is defined by

$$\mathrm{E}\left[X \,\middle|\, \mathcal{G}\right] = \mathrm{E}\left[X^+ \,\middle|\, \mathcal{G}\right] - \mathrm{E}\left[X^- \,\middle|\, \mathcal{G}\right] \ ,$$

where, on the P-null-set of sample points for which $\mathrm{E}\left[X^+ \mid \mathcal{G}\right] = \mathrm{E}\left[X^- \mid \mathcal{G}\right] = \infty$, the difference $\mathrm{E}\left[X^+ \mid \mathcal{G}\right] - \mathrm{E}\left[X^- \mid \mathcal{G}\right]$ is given an arbitrary value, for instance, zero. In particular, if $\mathrm{E}\left[\|X\| \mid \mathcal{G}\right] < \infty$ P-a.s. then $\mathrm{E}\left[X^+ \mid \mathcal{G}\right] < \infty$ and $\mathrm{E}\left[X^- \mid \mathcal{G}\right] < \infty$ P-a.s. and we may always define the conditional expectation in this context.

Let $\{M_N\}_{N \geq 0}$ be a sequence of positive integers satisfying $\lim_{N \to \infty} M_N = \infty$. Without loss of generality, we will assume that $\{M_N\}_{N \geq 0}$ is non decreasing. Let $\{U_{N,i}\}_{1 \leq i \leq M_N}$ be a triangular array of random variables on $(\Omega, \mathcal{F}, \mathrm{P})$. Let $\{\mathcal{F}_{N,i}\}_{0 \leq i \leq M_N}$ be a triangular array of of sub-sigma-fields of $\mathcal{F}$ such that for each N and each $i = 1, \ldots, M_N$, $U_{N,i}$ is $\mathcal{F}_{N,i}$-measurable and $\mathcal{F}_{N,i-1} \subseteq \mathcal{F}_{N,i}$.

Proposition 7. *Assume that* $\mathrm{E}\left[\|U_{N,j}\| \mid \mathcal{F}_{N,j-1}\right] < \infty$ *P-a.s. for any N and any $j = 1, \ldots, M_N$, and*

$$
\sup_N \mathrm{P}\left(\sum_{j=1}^{M_N} \mathrm{E}\left[\|U_{N,j}\| \mid \mathcal{F}_{N,j-1}\right] \geq \lambda\right) \to 0 \qquad \text{as } \lambda \to \infty \tag{38}
$$

$$
\sum_{j=1}^{M_N} \mathrm{E}\left[\|U_{N,j}\| \mathbf{1}\{\|U_{N,j}\| \geq \epsilon\} \mid \mathcal{F}_{N,j-1}\right] \xrightarrow{\mathrm{P}} 0 \qquad \text{for any positive } \epsilon. \tag{39}
$$

Then, $\max_{1 \leq i \leq M_N} \left|\sum_{j=1}^{i} U_{N,j} - \sum_{j=1}^{i} \mathrm{E}\left[U_{N,j} \mid \mathcal{F}_{N,j-1}\right]\right| \xrightarrow{\mathrm{P}} 0$.

Proof. Assume first that for each N and each $i = 1, \ldots, M_N$, $U_{N,i} \geq 0$, P-a.s.. By (Dvoretzky, 1972, Lemma 3.5), we have that for any constants ϵ and $\eta > 0$,

$$
\mathrm{P}\left[\max_{1 \leq i \leq M_N} U_{N,i} \geq \epsilon\right] \leq \eta + \mathrm{P}\left[\sum_{i=1}^{M_N} \mathrm{P}\left(U_{N,i} \geq \epsilon \mid \mathcal{F}_{N,i-1}\right) \geq \eta\right].
$$

From the conditional version of the Chebyshev identity,

$$
\mathrm{P}\left[\max_{1 \leq i \leq M_N} U_{N,i} \geq \epsilon\right] \leq \eta + \mathrm{P}\left[\sum_{i=1}^{M_N} \mathrm{E}\left[U_{N,i} \mathbf{1}\{U_{N,i} \geq \epsilon\} \mid \mathcal{F}_{N,i-1}\right] \geq \eta\epsilon\right]. \tag{40}
$$

Let ϵ and $\lambda > 0$ and define $\bar{U}_{N,i}$ by

$$
\bar{U}_{N,i} \stackrel{\text{def}}{=} U_{N,i} \mathbf{1}\{U_{N,i} < \epsilon\} \mathbf{1}\left\{\sum_{j=1}^{i} \mathrm{E}\left[U_{N,j} \mid \mathcal{F}_{N,j-1}\right] < \lambda\right\}.
$$

For any $\delta > 0$,

$$
\mathrm{P}\left(\max_{1 \leq i \leq M_N} \left|\sum_{j=1}^{i} U_{N,j} - \sum_{j=1}^{i} \mathrm{E}\left[U_{N,j} \mid \mathcal{F}_{N,j-1}\right]\right| \geq 2\delta\right)
$$

$$
\leq \mathrm{P}\left(\max_{1 \leq i \leq M_N} \left|\sum_{j=1}^{i} \bar{U}_{N,j} - \sum_{j=1}^{i} \mathrm{E}\left[\bar{U}_{N,j} \mid \mathcal{F}_{N,j-1}\right]\right| \geq \delta\right)
$$

$$
+ \mathrm{P}\left(\max_{1 \leq i \leq M_N} \left|\sum_{j=1}^{i} U_{N,j} - \bar{U}_{N,j} - \sum_{j=1}^{i} \mathrm{E}\left[U_{N,j} - \bar{U}_{N,j} \mid \mathcal{F}_{N,j-1}\right]\right| \geq \delta\right).
$$

The second term in the RHS is bounded by

$$
P\left(\max_{1\leq i\leq M_N} U_{N,i}\geq\epsilon\right) + P\left(\sum_{j=1}^{M_N} \mathrm{E}\left[U_{N,j}\,|\,\mathcal{F}_{N,j-1}\right]\geq\lambda\right) +
$$

$$
P\left(\sum_{j=1}^{M_N} \mathrm{E}\left[U_{N,j}\mathbf{1}\{U_{N,j}\geq\epsilon\}\,|\,\mathcal{F}_{N,j-1}\right]\geq\delta\right)
$$

Eqs. (39) and (40) imply that the first and last terms in the last expression converge to zero for any $\epsilon>0$ and (38) implies that the second term may be arbitrarily small by choosing for λ sufficiently large. Now, by the Doob maximal inequality,

$$
P\left(\max_{1\leq i\leq M_N}\left|\sum_{j=1}^{i}\bar{U}_{N,j}-\mathrm{E}\left[\bar{U}_{N,j}\,|\,\mathcal{F}_{N,j-1}\right]\right|\geq\delta\right)
$$

$$
\leq\delta^{-2}\sum_{j=1}^{M_N}\mathrm{E}\left(\bar{U}_{N,j}-\mathrm{E}\left[\bar{U}_{N,j}\,|\,\mathcal{F}_{N,j-1}\right]\right)^2
$$

This last term does not exceed

$$
\delta^{-2}\sum_{i=1}^{M_N}\mathrm{E}(\bar{U}_{N,j}^2)\leq\delta^{-2}\epsilon\sum_{j=1}^{M_N}\mathrm{E}[\bar{U}_{N,j}]\leq\delta^{-2}\epsilon\,\mathrm{E}\left[\sum_{j=1}^{M_N}\mathrm{E}\left[\bar{U}_{N,j}\,|\,\mathcal{F}_{N,j-1}\right]\right]\leq\delta^{-2}\epsilon\lambda\,.
$$

Since ϵ is arbitrary, the proof follows for $U_{N,j}\geq 0$, P-a.s., for each N and $j=1,\ldots,M_N$. The proof extends to an arbitrary triangular array $\{U_{N,j}\}_{1\leq i\leq M_N}$ by applying the preceding result to $\{U_{N,j}^+\}_{1\leq i\leq M_N}$ and $\{U_{N,j}^-\}_{1\leq j\leq M_N}$. $\quad\square$

Lemma 8. *Assume that for all N and $i=1,\ldots,M_N$, $\sigma_{N,i}^2\stackrel{\mathrm{def}}{=}\mathrm{E}\left[U_{N,i}^2\,\middle|\,\mathcal{F}_{N,i-1}\right]<\infty$ P-a.s. and $\mathrm{E}\left[U_{N,i}\,|\,\mathcal{F}_{N,i-1}\right]=0$, and for all $\epsilon>0$,*

$$
\sum_{i=1}^{M_N}\mathrm{E}\left[U_{N,i}^2\mathbf{1}\{|U_{N,i}|\geq\epsilon\}\,\middle|\,\mathcal{F}_{N,0}\right]\stackrel{\mathrm{P}}{\longrightarrow}0\,. \tag{41}
$$

Then, if any of the two following conditions holds

(i) $\sum_{i=1}^{M_N}\sigma_{N,i}^2=1$,

(ii) $\sigma_{N,i}$ is $\mathcal{F}_{N,0}$-measurable and $\{\sum_{j=1}^{M_N}\sigma_{N,j}^2\}_{N\geq0}$ is tight,

then for any real u,

$$
\mathrm{E}\left[\exp\left(iu\sum_{j=1}^{M_N}U_{N,j}\right)\,\middle|\,\mathcal{F}_{N,0}\right]-\exp\left(-(u^2/2)\sum_{j=1}^{M_N}\sigma_{N,j}^2\right)\stackrel{\mathrm{P}}{\longrightarrow}0\,. \tag{42}
$$

Proof. Write the following decomposition (with the convention $\sum_{j=a}^{b} = 0$ if $a > b$):

$$\mathrm{e}^{\mathrm{i}u \sum_{j=1}^{M_N} U_{N,j}} - \mathrm{e}^{-\frac{u^2}{2} \sum_{j=1}^{M_N} \sigma_{N,j}^2} = \sum_{l=1}^{M_N} \mathrm{e}^{\mathrm{i}u \sum_{j=1}^{l-1} U_{N,j}} \left(\mathrm{e}^{\mathrm{i}u U_{N,l}} - \mathrm{e}^{-\frac{u^2}{2}\sigma_{N,l}^2} \right) \mathrm{e}^{-\frac{u^2}{2} \sum_{j=l+1}^{M_N} \sigma_{N,j}^2} .$$

Moreover, if (i) or (ii) holds, $\sum_{j=1}^{l-1} U_{N,j}$ and $\sum_{j=l+1}^{M_N} \sigma_{N,j}^2$ are $\mathcal{F}_{N,l-1}$-measurable random variables. To see this, write if (i) holds, $\sum_{j=l+1}^{M_N} \sigma_{N,j}^2 = 1 - \sum_{j=1}^{l} \sigma_{N,j}^2$ and if (ii) holds, note that $\sum_{j=l+1}^{M_N} \sigma_{N,j}^2$ is a $\mathcal{F}_{N,0}$-measurable random variable and $\mathcal{F}_{N,0} \subset \mathcal{F}_{N,l-1}$. This implies that

$$\left| \mathrm{E} \left[\exp \left(\mathrm{i}u \sum_{j=1}^{M_N} U_{N,j} \right) - \exp \left(-(u^2/2) \sum_{j=1}^{M_N} \sigma_{N,j}^2 \right) \middle| \mathcal{F}_{N,0} \right] \right|$$
$$\leq \sum_{l=1}^{M_N} \mathrm{E} \left[\left| \mathrm{E} \left[\exp(\mathrm{i}u U_{N,l}) \middle| \mathcal{F}_{N,l-1} \right] - \exp(-u^2 \sigma_{N,l}^2/2) \right| \middle| \mathcal{F}_{N,0} \right] . \quad (43)$$

For any $\epsilon > 0$,

$$\left| \mathrm{E} \left[\exp \left(\mathrm{i}u U_{N,l} \right) \middle| \mathcal{F}_{N,l-1} \right] - 1 - \frac{1}{2} u^2 \sigma_{N,l}^2 \right|$$
$$\leq \frac{1}{6} |u|^3 \, \mathrm{E} \left[|U_{N,l}|^3 \mathbf{1}\{|U_{N,l}| \leq \epsilon\} \middle| \mathcal{F}_{N,l-1} \right] + u^2 \, \mathrm{E} \left[U_{N,l}^2 \mathbf{1}\{|U_{N,l}| > \epsilon\} \middle| \mathcal{F}_{N,l-1} \right]$$
$$\leq \frac{1}{6} \epsilon |u|^3 \sigma_{N,l}^2 + u^2 \, \mathrm{E} \left[U_{N,l}^2 \mathbf{1}\{|U_{N,l}| > \epsilon\} \middle| \mathcal{F}_{N,l-1} \right] ,$$

and thus,

$$\mathrm{E} \left[\sum_{l=1}^{M_N} \left| \mathrm{E} \left[\exp \left(\mathrm{i}u U_{N,l} \right) - 1 - \frac{1}{2} u^2 \sigma_{N,l}^2 \middle| \mathcal{F}_{N,l-1} \right] \right| \middle| \mathcal{F}_{N,0} \right]$$
$$\leq \frac{1}{6} \epsilon |u|^3 \sum_{l=1}^{M_N} \sigma_{N,l}^2 + u^2 \sum_{l=1}^{M_N} \mathrm{E} \left[U_{N,l}^2 \mathbf{1}\{|U_{N,l}| \geq \epsilon\} \middle| \mathcal{F}_{N,0} \right] . \quad (44)$$

Using either (i) or (ii) and since $\epsilon > 0$ is arbitrary, it follows from (41) that the RHS tends in probability to 0 as $N \to \infty$. Using again a Taylor inequality,

$$\mathrm{E} \left[\sum_{l=1}^{M_N} \left| \mathrm{E} \left[\exp \left(-u^2 \sigma_{N,l}^2/2 \right) - 1 - \frac{1}{2} u^2 \sigma_{N,l}^2 \middle| \mathcal{F}_{N,l-1} \right] \right| \middle| \mathcal{F}_{N,0} \right]$$
$$\leq \frac{u^4}{8} \sum_{l=1}^{M_N} \mathrm{E} \left[\sigma_{N,l}^4 \middle| \mathcal{F}_{N,0} \right] \leq \frac{u^4}{8} \mathrm{E} \left[\max_{1 \leq l \leq M_N} \sigma_{N,l}^2 \middle| \mathcal{F}_{N,0} \right] \left(\sum_{l=1}^{M_N} \sigma_{N,l}^2 \right) , \quad (45)$$

since under (i) or (ii), $\sum_{l=1}^{M_N} \sigma_{N,l}^2$ is $\mathcal{F}_{N,0}$-measurable. Because, for any $\epsilon > 0$,

$$\mathrm{E} \left[\max_{1 \leq j \leq M_N} \sigma_{N,j}^2 \middle| \mathcal{F}_{N,0} \right] \leq \epsilon^2 + \sum_{j=1}^{M_N} \mathrm{E} \left[U_{N,j}^2 \mathbf{1}\{|U_{N,j}| \geq \epsilon\} \middle| \mathcal{F}_{N,0} \right] ,$$

21

it follows from (41) and assumptions (i) or (ii) that the RHS of (45) tends in probability to 0 as $N \to \infty$. Therefore the LHS of (43) tends in probability to 0 because the sum on the RHS of this display is bounded by (44) and (45). $\quad\square$

Proposition 9. *Assume that* $\sigma_{N,i}^2 \overset{\text{def}}{=} \mathrm{E}\left[U_{N,i}^2 \,\middle|\, \mathcal{F}_{N,i-1}\right] < \infty$ *P-a.s.* , $\mathrm{E}\left[U_{N,i} \,\middle|\, \mathcal{F}_{N,i-1}\right] = 0$ *for all* N *and* $i = 1, \ldots, M_N$ *and*

$$\sum_{i=1}^{M_N} \sigma_{N,i}^2 \overset{\mathrm{P}}{\longrightarrow} \sigma^2 \qquad\qquad\qquad \text{for some } \sigma^2 > 0 , \qquad (46)$$

$$\sum_{i=1}^{M_N} \mathrm{E}\left[U_{N,i}^2 \mathbf{1}\{|U_{N,i}| \geq \epsilon\} \,\middle|\, \mathcal{F}_{N,i-1}\right] \overset{\mathrm{P}}{\longrightarrow} 0 \qquad \text{for any } \epsilon > 0 . \qquad (47)$$

Then for any real u,

$$\mathrm{E}\left[\exp\left(\mathrm{i}u \sum_{i=1}^{M_N} U_{N,i}\right) \,\middle|\, \mathcal{F}_{N,0}\right] \overset{\mathrm{P}}{\longrightarrow} \exp\left(-\sigma^2 u^2 / 2\right) . \qquad (48)$$

Proof. Without loss of generality, we assume that $\sigma^2 = 1$. Define the stopping time τ_N by

$$\tau_N \overset{\text{def}}{=} \max\left\{1 \leq k \leq M_N : \sum_{j=1}^{k} \sigma_{N,j}^2 \leq 1\right\} ,$$

with the convention $\max \emptyset = 0$. Put $\bar{U}_{N,k} = U_{N,k}$ for $k \leq \tau_N$, $\bar{U}_{N,k} = 0$ for $\tau_N < k \leq M_N$ and $\bar{U}_{N,M_N+1} = \left(1 - \sum_{j=1}^{\tau_N} \sigma_{N,j}^2\right)^{1/2} Y_N$, where $\{Y_N\}$ are $\mathcal{N}(0,1)$ independent and independent of $\mathcal{F}_{N,M_N}$. By construction,

$$\mathrm{E}\left[\bar{U}_{N,M_N+1}^2 \,\middle|\, \mathcal{F}_{N,M_N}\right] = 1 - \sum_{j=1}^{\tau_N} \sigma_{N,j}^2 < \infty , \quad \mathrm{P}-a.s.$$

and $\mathrm{E}\left[\bar{U}_{N,M_N+1} \,\middle|\, \mathcal{F}_{N,M_N}\right] = 0$. The triangular array $\{\bar{U}_{N,k}\}_{1 \leq k \leq M_N+1}$ obviously satisfies the conditions of Lemma 8 for conditional means and variances. By construction,

$$\sum_{j=1}^{M_N} \mathrm{E}\left[\bar{U}_{N,j}^2 \mathbf{1}\{|\bar{U}_{N,j}| \geq \epsilon\} \,\middle|\, \mathcal{F}_{N,0}\right] = \mathrm{E}\left[\sum_{j=1}^{\tau_N} \mathrm{E}\left[U_{N,j}^2 \mathbf{1}\{|U_{N,j}| \geq \epsilon\} \,\middle|\, \mathcal{F}_{N,j-1}\right] \,\middle|\, \mathcal{F}_{N,0}\right] . \qquad (49)$$

Since $\sum_{j=1}^{\tau_N} \mathrm{E}\left[U_{N,j}^2 \mathbf{1}\{|U_{N,j}| \geq \epsilon\} \,\middle|\, \mathcal{F}_{N,j-1}\right] \leq 1$, then (47) shows that the RHS of (49) converges in probability to 0 as $N \to \infty$. On the other hand, if $\tau_N < M_N$,

$$0 \leq 1 - \sum_{j=1}^{\tau_N} \sigma_{N,j}^2 \leq \sigma_{N,\tau_N+1}^2 \leq \max_{1 \leq j \leq M_N} \sigma_{N,j}^2 .$$

For any $\epsilon > 0$,

$$\max_{1 \leq j \leq M_N} \sigma_{N,j}^2 \leq \epsilon^2 + \sum_{j=1}^{M_N} \mathrm{E}\left[U_{N,j}^2 \mathbf{1}\{|U_{N,j}| \geq \epsilon\} \,\middle|\, \mathcal{F}_{N,j-1}\right]$$

Since $\epsilon > 0$ is arbitrary, it follows from (47) that

$$1 - \sum_{j=1}^{\tau_N} \sigma_{N,j}^2 \xrightarrow{\mathrm{P}} 0 \ , \tag{50}$$

and since $1 - \sum_{j=1}^{\tau_N} \sigma_{N,j}^2 \leq 1$, this implies that $\mathrm{E}\left[\bar{U}_{N,M_N+1}^2 \,\middle|\, \mathcal{F}_{N,0}\right] \xrightarrow{\mathrm{P}} 0$. Therefore, $\{\bar{U}_{n,k}\}_{1 \leq k \leq M_N+1}$ satisfies (41). Put

$$\sum_{j=1}^{M_N} U_{N,j} = \sum_{j=1}^{M_N+1} \bar{U}_{N,j} - \bar{U}_{N,M_N+1} + \sum_{j=\tau_N+1}^{M_N} U_{N,j} \ . \tag{51}$$

The first term on the RHS is asymptotically $\mathcal{N}(0,1)$ and $\bar{U}_{N,M_N+1} \xrightarrow{\mathrm{P}} 0$. It remains to prove that $\sum_{j=\tau_N+1}^{M_N} U_{N,j} \xrightarrow{\mathrm{P}} 0$. First note that

$$\sum_{j=\tau_N+1}^{M_N} \sigma_{N,j}^2 = \sum_{j=1}^{M_N} \sigma_{N,j}^2 - 1 + \left(1 - \sum_{j=1}^{\tau_N} \sigma_{N,j}^2\right) \xrightarrow{\mathrm{P}} 0 \tag{52}$$

For any $\lambda > 0$,

$$\mathrm{E}\left(\sum_{j=\tau_N+1}^{M_N} U_{N,j}\mathbf{1}\left\{\sum_{i=\tau_N+1}^{j} \sigma_{N,i}^2 \leq \lambda\right\}\right)^2 = \mathrm{E}\left(\sum_{j=\tau_N+1}^{M_N} \sigma_{N,j}^2\mathbf{1}\left\{\sum_{i=\tau_N+1}^{j} \sigma_{N,i}^2 \leq \lambda\right\}\right) \ .$$

The term between brackets converges to 0 in probability by (52) and its absolute value is bounded by λ. Thus, by dominated convergence, this expectation converges to 0. Thus,

$$\sum_{j=\tau_N+1}^{M_N} U_{N,j}\mathbf{1}\left\{\sum_{i=\tau_N+1}^{j} \sigma_{N,i}^2 \leq \lambda\right\} \xrightarrow{\mathrm{P}} 0$$

Moreover,

$$\mathrm{P}\left(\sum_{j=\tau_N+1}^{M_N} U_{N,j}\mathbf{1}\left\{\sum_{i=\tau_N+1}^{j} \sigma_{N,i}^2 > \lambda\right\} \neq 0\right)$$

$$\leq \mathrm{P}\left(\exists j \in \{1, \ldots, M_N\}, \quad \sum_{i=\tau_N+1}^{j} \sigma_{N,i}^2 > \lambda\right) = \mathrm{P}\left(\sum_{i=\tau_N+1}^{M_N} \sigma_{N,i}^2 > \lambda\right) \ ,$$

which converges to 0 by (52). The proof is completed. $\qquad\square$

We will use the following technical Lemma which is a conditional version of (Dvoretzky, 1972, Lemma 3.3).

Lemma 10. *Let $\mathcal{G}$ a σ-field and X a random variable such that $\mathrm{E}\left[X^2 \,\middle|\, \mathcal{G}\right] < \infty$. Then, for any $\epsilon > 0$,*

$$4\,\mathrm{E}\left[\,|X|^2\mathbf{1}\{|X| \geq \epsilon\} \,\middle|\, \mathcal{G}\right] \geq \mathrm{E}\left[\,|X - \mathrm{E}\left[X \,\middle|\, \mathcal{G}\right]|^2\mathbf{1}\{|X - \mathrm{E}\left[X \,\middle|\, \mathcal{G}\right]| \geq 2\epsilon\} \,\middle|\, \mathcal{G}\right] \ ,$$

Proof. Let $Y = X - \mathrm{E}\,[X \,|\, \mathcal{G}]$. We have $\mathrm{E}\,[Y \,|\, \mathcal{G}] = 0$. It is equivalent to show that for any $\mathcal{G}$-measurable random variable c,

$$\mathrm{E}\,\big[Y^2 \mathbf{1}\{|Y| \geq 2\epsilon\} \,\big|\, \mathcal{G}\big] \leq 4\,\mathrm{E}\,\big[\,|Y + c|^2 \mathbf{1}\{|Y + c| \geq \epsilon\} \,\big|\, \mathcal{G}\big]$$

On the set $\{|c| < \epsilon\}$,

$$
\begin{aligned}
\mathrm{E}\,\big[Y^2 \mathbf{1}\{|Y| \geq 2\epsilon\} \,\big|\, \mathcal{G}\big] &\leq 2\,\mathrm{E}\,\big[((Y + c)^2 + c^2)\mathbf{1}\{|Y + c| \geq \epsilon\} \,\big|\, \mathcal{G}\big] \\
&\leq 2(1 + c^2/\epsilon^2)\,\mathrm{E}\,\big[(Y + c)^2 \mathbf{1}\{|Y + c| \geq \epsilon\} \,\big|\, \mathcal{G}\big] \\
&\leq 4\,\mathrm{E}\,\big[(Y + c)^2 \mathbf{1}\{|Y + c| \geq \epsilon\} \,\big|\, \mathcal{G}\big] \ .
\end{aligned}
$$

Moreover, on the set $\{|c| \geq \epsilon\}$, using that $\mathrm{E}\,[cY \,|\, \mathcal{G}] = c\,\mathrm{E}\,[Y \,|\, \mathcal{G}] = 0$.

$$
\begin{aligned}
\mathrm{E}\,\big[Y^2 \mathbf{1}\{|Y| \geq 2\epsilon\} \,\big|\, \mathcal{G}\big] &\leq \mathrm{E}\,\big[Y^2 + c^2 - \epsilon^2 \,\big|\, \mathcal{G}\big] \\
&\leq \mathrm{E}\,\big[(Y + c)^2 - \epsilon^2 \,\big|\, \mathcal{G}\big] \\
&\leq \mathrm{E}\,\big[(Y + c)^2 \mathbf{1}\{|Y + c| \geq \epsilon\} \,\big|\, \mathcal{G}\big] \ .
\end{aligned}
$$

The proof is completed. $\qquad\square$

Corollary 11. *Assume that for each N and $i = 1, \ldots, M_N$, $\mathrm{E}\,\Big[U_{N,i}^2 \,\Big|\, \mathcal{F}_{N,i-1}\Big] < \infty$ and*

$$\sum_{i=1}^{M_N} \{\mathrm{E}\,\big[U_{N,i}^2 \,\big|\, \mathcal{F}_{N,i-1}\big] - (\mathrm{E}\,[U_{N,i} \,|\, \mathcal{F}_{N,i-1}])^2\} \overset{\mathrm{P}}{\longrightarrow} \sigma^2 \quad \text{for some } \sigma^2 > 0 \ , \quad (53)$$

$$\sum_{i=1}^{M_N} \mathrm{E}\,\Big[U_{N,i}^2 \mathbf{1}_{\{|U_{N,i}| \geq \epsilon\}} \,\Big|\, \mathcal{F}_{N,i-1}\Big] \overset{\mathrm{P}}{\longrightarrow} 0 \qquad\qquad \text{for any } \epsilon > 0 \ . \qquad (54)$$

Then, for any real u,

$$\mathrm{E}\,\left[\exp\left(\mathrm{i}u \sum_{i=1}^{M_N} \{U_{N,i} - \mathrm{E}\,[U_{N,i} \,|\, \mathcal{F}_{N,i-1}]\}\right) \,\middle|\, \mathcal{F}_{N,0}\right] \overset{\mathrm{P}}{\longrightarrow} \exp(-(u^2/2)\sigma^2) \ . \quad (55)$$

If in addition $\mathcal{F}_{N,i} \subseteq \mathcal{F}_{N',i}$ for all $N' \geq N$ and $i \leq M_N$, then

$$\sum_{i=1}^{M_N} \{U_{N,i} - \mathrm{E}\,[U_{N,i} \,|\, \mathcal{F}_{N,i-1}]\} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, \sigma^2) \quad (\text{mixing}) \ . \qquad (56)$$

Proof. Set $\bar{U}_{N,i} = U_{N,i} - \mathrm{E}\,[U_{N,i} \,|\, \mathcal{F}_{N,i-1}]$. By construction $\mathrm{E}\,\big[\bar{U}_{N,i} \,\big|\, \mathcal{F}_{N,i-1}\big] = 0$ and because $\mathrm{E}\,\Big[\bar{U}_{N,i}^2 \,\Big|\, \mathcal{F}_{N,i-1}\Big] = \mathrm{E}\,\Big[U_{N,i}^2 \,\Big|\, \mathcal{F}_{N,i-1}\Big] - (\mathrm{E}\,[U_{N,i} \,|\, \mathcal{F}_{N,i-1}])^2$, (46) is fulfilled. The proof of (47) follows from Lemma 10. It remains to show (56). Let $\mathcal{F}_\infty = \vee_1^\infty \mathcal{F}_{N,M_N}$ the σ-field generated by $\cup_1^\infty \mathcal{F}_{N,M_N}$. For any $E' \in \mathcal{F}_\infty$ and any $\epsilon > 0$, there exists an m such that $E \in \mathcal{F}_{m,M_m}$ and $\mathrm{P}(E\Delta E') < \epsilon$ (where Δ denotes the symmetric difference). Denotes $V_{N,i} = U_{N,i} - \mathrm{E}\,[U_{N,i} \,|\, \mathcal{F}_{N,i-1}]$.

$$\left|\mathrm{E}\,\left[\exp\left(\mathrm{i}u \sum_{i=1}^{M_N} V_{N,i}\right)\mathbf{1}(E)\right] - \mathrm{E}\,\left[\exp\left(\mathrm{i}u \sum_{i=1}^{M_N} V_{N,i}\right)\mathbf{1}(E')\right]\right| \leq \mathrm{P}(E\Delta E') < \epsilon$$

Moreover, applying (55) to the triangular array $\{U_{N,M_m+i}\}_{1\le i\le M_N-M_m}$ (with N sufficiently large) associated to the σ-field $\{\mathcal{F}_{N,M_m+i}\}_{0\le i\le M_N-M_m}$,

$$\mathrm{E}\left[\exp\left(iu\sum_{i=m+1}^{M_N} V_{N,i}\right)\mathbf{1}(E)\right] \xrightarrow{\mathrm{P}} \exp(-u^2\sigma^2/2)\,\mathrm{P}(E)$$

and $\max_{1\le i\le M_N}|V_{N,i}| \xrightarrow{\mathrm{P}} 0$. Thus,

$$\mathrm{E}\left[\exp\left(iu\sum_{i=1}^{M_N} V_{N,i}\right)\mathbf{1}(E)\right] \xrightarrow{\mathrm{P}} \exp(-u^2\sigma^2/2)\,\mathrm{P}(E)$$

This implies that for any $E'\in\mathcal{F}_\infty$,

$$\mathrm{E}\left[\exp\left(iu\sum_{i=1}^{M_N} V_{N,i}\right)\mathbf{1}(E')\right] \xrightarrow{\mathrm{P}} \exp(-u^2\sigma^2/2)\,\mathrm{P}(E')$$

Now, let $E\in\mathcal{F}$,

$$E\left[\exp\left(iu\sum_{i=1}^{M_N} V_{N,i}\right)\mathbf{1}(E)\right] = \mathrm{E}\left[\exp\left(iu\sum_{i=1}^{M_N} V_{N,i}\right)\mathrm{E}\left[\mathbf{1}(E)\,|\,\mathcal{F}_\infty\right]\right]$$
$$\xrightarrow{\mathrm{P}} \exp(-u^2\sigma^2/2)\,\mathrm{E}(\mathrm{E}\left[\mathbf{1}(E)\,|\,\mathcal{F}_\infty\right]) = \exp(-u^2\sigma^2/2)\,\mathrm{P}(E).$$

The proof is completed. $\qquad\square$

B   Proof of Theorems 1 and 2

Proof of the Theorem 1. We set $\mathcal{F}_{N,0} = \sigma\left(\{(\xi_{N,i},\omega_{N,i})\}_{1\le i\le M_N}\right)$ and for $j = 1,\ldots,\tilde{M}_N$, $\mathcal{F}_{N,j} = \mathcal{F}_{N,0}\vee\sigma\left(\{\tilde{\xi}_{N,k}\}_{1\le k\le j}\right)$. Checking that $\tilde{\mathsf{C}}$ is proper is straightforward, so we turn to the consistency. We show first that for any $f\in\tilde{\mathsf{C}}$,

$$\frac{1}{\alpha_N\Omega_N}\sum_{j=1}^{\tilde{M}_N}\tilde{\omega}_{N,j}f(\tilde{\xi}_{N,j}) \xrightarrow{\mathrm{P}} \nu L(f)\,, \tag{57}$$

where $\tilde{\xi}_{N,j}$ and $\tilde{\omega}_{N,j}$ are defined in (11) and (12), respectively. Because the mutation step is unbiased (see (13)),

$$(\alpha_N\Omega_N)^{-1}\sum_{j=1}^{\tilde{M}_N}\mathrm{E}\left[\tilde{\omega}_{N,j}f(\tilde{\xi}_{N,j})\,\Big|\,\mathcal{F}_{N,j-1}\right] = \Omega_N^{-1}\sum_{i=1}^{M_N}\omega_{N,i}L(\xi_{N,i},f)\,.$$

Because the weighted sample $\{(\xi_{N,i},\omega_{N,i})\}_{1\le i\le M_N}$ is consistent for (ν,C) and for $f\in\tilde{\mathsf{C}}$, the function $L(\cdot,f)\in\mathsf{C}$,

$$\Omega_N^{-1}\sum_{i=1}^{M_N}\omega_{N,i}L(\xi_{N,i},f) \xrightarrow{\mathrm{P}} \nu L(f)\,,$$

25

it suffices to show that

$$(\alpha_N \Omega_N)^{-1} \sum_{j=1}^{\tilde{M}_N} \left\{ \tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) - \mathrm{E}\left[\tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \,\middle|\, \mathcal{F}_{N,j-1} \right] \right\} \xrightarrow{\mathrm{P}} 0 \,. \qquad (58)$$

Put $U_{N,j} = (\alpha_N \Omega_N)^{-1} \tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j})$ for $j = 1, \dots, \tilde{M}_N$ and appeal to Proposition 7. Just as above,

$$\sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,j}| \,\middle|\, \mathcal{F}_{N,j-1} \right] = \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} L(\xi_{N,i}, |f|) \xrightarrow{\mathrm{P}} \nu L(|f|) \,,$$

showing that the sequence $\left\{ \sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,j}| \,\middle|\, \mathcal{F}_{N,j-1} \right] \right\}_{N \geq 0}$ is tight (Proposition 7-Eq.(38)). For any $\epsilon > 0$, put $A_N \stackrel{\text{def}}{=} \sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,j}| \mathbf{1}\{|U_{N,j}| \geq \epsilon\} \,\middle|\, \mathcal{F}_{N,j-1} \right]$. We need to show that $A_N \xrightarrow{\mathrm{P}} 0$ (Proposition 7-Eq.(39)). For any positive C, $\xi \in \Xi$, $R(\xi, W|f|\mathbf{1}\{W|f| \geq C\}) \leq R(\xi, W|f|) = L(\xi, |f|)$. Because the function $L(\cdot, |f|)$ belongs to the proper set C, the function $R(\cdot, W|f|\mathbf{1}\{W|f| \geq C\})$ belongs to C. Hence for all C, $\epsilon > 0$,

$$A_N \mathbf{1} \left\{ (\alpha_N \Omega_N)^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} \leq \epsilon/C \right\}$$
$$\leq \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} \left[\alpha_N^{-1} \sum_{k=1}^{\alpha_N} R(\xi_{N,i}, W|f|\mathbf{1}\{W|f| \geq C\}) \right]$$
$$\xrightarrow{\mathrm{P}} \nu R(W|f|\mathbf{1}\{W|f| \geq C\}) \,.$$

By dominated convergence, the RHS can be made arbitrarily small by letting $C \to \infty$. Combining with $\Omega_N^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} \xrightarrow{\mathrm{P}} 0$, this shows that A_N tends to zero in probability, showing (39). Thus Proposition 7 applies and (57) holds. Under the stated assumptions, the function $L(\cdot, \tilde{\Xi})$ belongs to C, implying that the constant function $g \equiv 1$ satisfies (57); therefore, $(\alpha_N \Omega_N)^{-1} \sum_{j=1}^{\tilde{M}_N} \tilde{\omega}_{N,j} \xrightarrow{\mathrm{P}} \nu L(\Xi)$. Combined with (57) this shows that for any $f \in \tilde{\mathsf{C}}$,

$$\tilde{\Omega}_N^{-1} \sum_{j=1}^{\tilde{M}_N} \tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \xrightarrow{\mathrm{P}} \mu(f) \,,$$

To complete the proof of the consistency, it remains to prove that

$$\tilde{\Omega}_N^{-1} \max_{1 \leq j \leq \tilde{M}_N} \tilde{\omega}_{N,j} \xrightarrow{\mathrm{P}} 0 \,.$$

Because $(\alpha_N \Omega_N)^{-1} \tilde{\Omega}_N \xrightarrow{\mathrm{P}} \nu L(\Xi)$, it is actually sufficient to show that

$$(\alpha_N \Omega_N)^{-1} \max_j \tilde{\omega}_{N,j} \xrightarrow{\mathrm{P}} 0 \,.$$

For any $C > 0$,

$$(\alpha_N \Omega_N)^{-1} \max_{1 \le j \le \tilde{M}_N} \tilde{\omega}_{N,j} \mathbf{1}_{\{W(\tilde{\xi}_{N,j}) \le C\}} \le C (\alpha_N \Omega_N)^{-1} \max_{1 \le i \le M_N} \omega_{N,i} \xrightarrow{\mathrm{P}} 0$$

$$(\alpha_N \Omega_N)^{-1} \max_{1 \le j \le \tilde{M}_N} \tilde{\omega}_{N,j} \mathbf{1}_{\{W(\tilde{\xi}_{N,j}) > C\}} \le (\alpha_N \Omega_N)^{-1} \sum_{j=1}^{\tilde{M}_N} \tilde{\omega}_{N,j} \mathbf{1}_{\{W(\tilde{\xi}_{N,j}) > C\}}$$

$$\xrightarrow{\mathrm{P}} \nu L (\{W > C\})$$

The term in the RHS of the last equation goes to zero as $C \to \infty$ which concludes the proof. $\square$

Proof of the Theorem 2. First we note that by definition α is necessarily at least 1. Checking that $\tilde{\mathsf{A}}$ and $\tilde{\mathsf{W}}$ is proper is straightforward. Pick $f \in \tilde{\mathsf{A}}$ and assume, without loss of generality, that $\mu(f) = 0$. Write $\widetilde{\Omega}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i} f(\tilde{\xi}_{N,i}) = (\alpha_N \Omega_N / \widetilde{\Omega}_N) (A_N + B_N)$, with

$$A_N = (\alpha_N \Omega_N)^{-1} \sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[\tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \,\Big|\, \mathcal{F}_{N,j-1} \right] = \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} L(\xi_{N,i}, f) \ ,$$

$$B_N = (\alpha_N \Omega_N)^{-1} \sum_{j=1}^{\tilde{M}_N} \left\{ \tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) - \mathrm{E}\left[\tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \,\Big|\, \mathcal{F}_{N,j-1} \right] \right\} \ .$$

Because $\alpha_N \Omega_N / \widetilde{\Omega}_N$ converges to $1/\nu L(\tilde{\Xi})$ in probability (see in the proof of Theorem 1), the conclusion of the theorem follows from Slutsky's theorem if we prove that $M_N^{1/2}(A_N + B_N)$ converges weakly to $\mathrm{N}(0, \sigma^2(Lf) + \alpha^{-1}\eta^2(f))$ where

$$\eta^2(f) \stackrel{\mathrm{def}}{=} \gamma R\{[Wf - R(\cdot, Wf)]^2\} \ , \tag{59}$$

with W given in (10). The function $L(\cdot, f)$ belongs to A and $\nu L(f) = \mu(f) \nu L(\tilde{\Xi}) = 0$. Because $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \le i \le M_N}$ is asymptotically normal for $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{M_N^{1/2}\})$, $M_N^{1/2} A_N \xrightarrow{\mathcal{D}} \mathrm{N}(0, \sigma^2(Lf))$. Next we prove that for any real u,

$$\mathrm{E}\left[\exp(\mathrm{i}u \tilde{M}_N^{1/2} B_N) \,\Big|\, \mathcal{F}_{N,0} \right] \xrightarrow{\mathrm{P}} \exp\left(-(u^2/2)\eta^2(f) \right) \ .$$

where $\eta^2(f)$ is defined in (59). For that purpose we use corollary 11, and we thus need to check (53)-(54) with

$$U_{N,j} \stackrel{\mathrm{def}}{=} (\alpha_N \Omega_N)^{-1} \tilde{M}_N^{1/2} \tilde{\omega}_{N,j} f(\tilde{\xi}_{N,j}) \ , \quad j = 1, \ldots, \tilde{M}_N \ .$$

Under the stated assumptions, for $f \in \mathsf{A}$, the function $R(\cdot, W^2 f^2)$ belongs to W. By the Jensen inequality, for any $\xi \in \Xi$, $\{L(\xi, f)\}^2 = \{R(\xi, Wf)\}^2 \le R(\xi, W^2 f^2)$. Because the W is proper and the function $R(\cdot, W^2 f^2) \in \mathsf{W}$, the relation $\{L(\cdot, f)\}^2 \le R(\cdot, W^2 f^2)$ implies that the function $\{L(\cdot, f)\}^2$ also belongs

27

to W. Because $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \le i \le M_N}$ is asymptotically normal for $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{M_N^{1/2}\})$, (5) implies

$$\sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[U_{N,j}^2 \,\middle|\, \mathcal{F}_{N,j-1}\right] = \frac{M_N}{\Omega_N^2} \sum_{i=1}^{M_N} \omega_{N,i}^2 R(\xi_{N,i}, W^2 f^2) \xrightarrow{\mathrm{P}} \gamma \, R(W^2 f^2) \;,$$

$$\sum_{j=1}^{\tilde{M}_N} (\mathrm{E}\left[U_{N,j} \,\middle|\, \mathcal{F}_{N,j-1}\right])^2 = \frac{M_N}{\Omega_N^2} \sum_{i=1}^{M_N} \omega_{N,i}^2 \left\{L(\xi_{N,i}, f)\right\}^2 \xrightarrow{\mathrm{P}} \int \gamma(d\xi) \left\{L(\xi, f)\right\}^2 \;.$$

These displays imply that (53) holds. It remains to check (54). We have, for all C, $\epsilon > 0$,

$$\left(\sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[U_{N,j}^2 \mathbf{1}_{\{|U_{N,j}| \ge \epsilon\}} \,\middle|\, \mathcal{F}_{N,j-1}\right]\right) \mathbf{1}\left\{\frac{M_N \max_{1 \le i \le M_N} \omega_{N,i}^2}{\alpha_N \Omega_N^2} \le \left(\frac{\epsilon}{C}\right)^2\right\} \le$$

$$\frac{M_N}{\Omega_N^2} \sum_{i=1}^{M_N} \omega_{N,i}^2 R\left(\xi_{N,i}, W^2 f^2 \mathbf{1}\{|Wf| \ge C\}\right) \xrightarrow{\mathrm{P}} \gamma R\left(W^2 f^2 \mathbf{1}\{|Wf| \ge C\}\right)$$

which converges to 0 as C goes to infinity. Combining with

$$(M_N)^{1/2} \Omega_N^{-1} \max_{1 \le i \le M_N} \omega_{N,i} \xrightarrow{\mathrm{P}} 0$$

yields

$$\sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[U_{N,j}^2 \mathbf{1}\{|U_{N,j}| \ge \epsilon\} \,\middle|\, \mathcal{F}_{N,j-1}\right] \xrightarrow{\mathrm{P}} 0 \;,$$

and this is condition (54). Thus corollary 11 applies. It follows that

$$\begin{pmatrix} M_N^{1/2} A_N \\ \tilde{M}_N^{1/2} B_N \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}\left(0, \mathrm{diag}\{\sigma^2(Lf), \eta^2(f)\}\right) \;,$$

where $\eta^2(f)$ is given in (59). The proof of condition (4) in the definition of an asymptotically normal sample is now concluded upon writing $M_N^{1/2}(A_N + B_N) = M_N^{1/2} A_N + \alpha_N^{-1/2} \tilde{M}_N^{1/2} B_N$.

Consider now (5). Recalling that $\tilde{\Omega}_N / (\alpha_N \Omega_N) \xrightarrow{\mathrm{P}} \nu L(\Xi)$, it is sufficient to show that for $f \in \tilde{\mathsf{W}}$,

$$\frac{M_N}{(\alpha_N \Omega_N)^2} \sum_{j=1}^{\tilde{M}_N} \tilde{\omega}_{N,j}^2 f(\tilde{\xi}_{N,j}) \xrightarrow{\mathrm{P}} \alpha^{-1} \gamma R(h) \;, \tag{60}$$

where $h \stackrel{\mathrm{def}}{=} W^2 f$. Define $U_{N,j} = (\alpha_N \Omega_N)^{-2} M_N \tilde{\omega}_{N,j}^2 f(\tilde{\xi}_{N,j})$. Under the stated assumptions, for any $f \in \tilde{\mathsf{W}}$, the functions $R(\cdot, |h|)$ and $R(\cdot, h)$ belong to W. Be-

cause $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is asymptotically normal for $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{M_N^{1/2}\})$,

$$\sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,j}| \,|\, \mathcal{F}_{N,j-1}\right] = \frac{M_N}{\alpha_N \Omega_N^2} \sum_{i=1}^{M_N} \omega_{N,i}^2 R(\xi_{N,i}, |h|) \xrightarrow{\mathrm{P}} \alpha^{-1} \gamma R(|h|) , \quad (61)$$

$$\sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[U_{N,j} \,|\, \mathcal{F}_{N,j-1}\right] = \frac{M_N}{\alpha_N \Omega_N^2} \sum_{i=1}^{M_N} \omega_{N,i}^2 R(\xi_{N,i}, h) \xrightarrow{\mathrm{P}} \alpha^{-1} \gamma R(h) . \quad (62)$$

We appeal to Proposition 7. Eq. (61) shows that the tightness condition (38). For $\epsilon > 0$, set $A_N = \sum_{j=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,j}|\mathbf{1}_{\{|U_{N,j}| \geq \epsilon\}} \,\Big|\, \mathcal{F}_{N,j-1}\right]$. We need to show that $A_N \xrightarrow{\mathrm{P}} 0$. For any positive C and $\xi \in \Xi$, $R(\xi, |h|\mathbf{1}\{|h| \geq C\}) \leq R(\xi, |h|)$. Because the function $R(\cdot, |h|)$ belongs to W and the set W is proper, the function $R(\cdot, |h|\mathbf{1}\{|h| \geq C\})$ belongs to C. Hence for all $C, \epsilon > 0$,

$$A_N \mathbf{1}\left\{\max_i \frac{M_N \omega_{N,i}^2}{(\alpha_N \Omega_N)^2} \leq \frac{\epsilon}{C}\right\}$$

$$\leq \frac{M_N}{\Omega_N^2} \sum_{i=1}^{M_N} \omega_{N,i}^2 R\left(\xi_{N,i}, |h|\mathbf{1}\{|h| \geq C\}\right) \xrightarrow{\mathrm{P}} \gamma R\left(|h|\mathbf{1}\{|h| \geq C\}\right) .$$

By the dominated convergence theorem, the RHS can be made arbitrarily small by letting $C \to \infty$. Combining with $M_N(\alpha_N \Omega_N^2)^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i}^2 \xrightarrow{\mathrm{P}} 0$, this shows that A_N tends to zero in probability, showing (39). Thus Proposition 7 applies and condition (5) is proved. To complete the proof, it remains to prove (6). Combining with $\widetilde{\Omega}_N/(\alpha_N \Omega_N) \xrightarrow{\mathrm{P}} \nu L(\Xi)$ (see proof of Theorem 1) and $\tilde{M}_N = \alpha_N M_N$, it is actually sufficient to show that $\left(\alpha_N \Omega_N^2\right)^{-1} M_N \max_{1 \leq j \leq \tilde{M}_N} \tilde{\omega}_{N,j}^2 \xrightarrow{\mathrm{P}} 0$. For any $C > 0$,

$$M_N \frac{\max_{1 \leq j \leq \tilde{M}_N} \tilde{\omega}_{N,j}^2 \mathbf{1}_{\{W(\tilde{\xi}_{N,j}) \leq C\}}}{\alpha_N (\Omega_N)^2} \leq C^2 M_N \frac{\max_{1 \leq i \leq M_N} \omega_{N,i}^2}{\alpha_N (\Omega_N)^2} \xrightarrow{\mathrm{P}} 0 .$$

Applying (60) with $f \equiv 1$, it holds that

$$M_N \frac{\max_{1 \leq j \leq \tilde{M}_N} \tilde{\omega}_{N,j}^2 \mathbf{1}_{\{W(\tilde{\xi}_{N,j}) > C\}}}{\alpha_N (\Omega_N)^2} \leq \frac{M_N}{\alpha_N (\Omega_N)^2} \sum_{j=1}^{\tilde{M}_N} \tilde{\omega}_{N,j}^2 \mathbf{1}_{\{W(\tilde{\xi}_{N,j}) \geq C\}}$$

$$\xrightarrow{\mathrm{P}} \gamma R\left(W^2 \mathbf{1}\{W \geq C\}\right) .$$

The term on the RHS goes to zero as $C \to \infty$. This proves condition (6) and concludes the proof. $\qquad\square$

C   Proof of the Theorems 3 and 4

Proof of the Theorem 3. As above, we set $\mathcal{F}_{N,0} = \sigma\left(\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}\right)$ and for $j = 1, \ldots, \tilde{M}_N$, $\mathcal{F}_{N,j} = \mathcal{F}_{N,0} \vee \sigma\left(\{\tilde{\xi}_{N,k}\}_{1 \leq k \leq j}\right)$. Pick f in C. Since C is

proper, $|f|\mathbf{1}\{|f| \geq C\} \in \mathsf{C}$ for any $C \geq 0$. Because $\{(\xi_{N,i},\ \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, C), and C is a proper set of functions,

$$\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|f(\tilde{\xi}_{N,i})|\mathbf{1}_{\{|f(\tilde{\xi}_{N,i})| \geq C\}} \,\Big|\, \mathcal{F}_{N,i-1}\right]$$

$$= \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i}|f(\xi_{N,i})|\mathbf{1}_{\{|f(\xi_{N,i})| \geq C\}} \xrightarrow{\mathrm{P}} \nu(|f|\mathbf{1}_{\{|f| \geq C\}})\,, \quad (63)$$

We now check (38)–(39) of Proposition 7. For any $i = 1, \ldots, M_N$, $U_{N,i} \overset{\mathrm{def}}{=} \tilde{M}_N^{-1} f(\tilde{\xi}_{N,i})$. Taking $C = 0$ in (63):

$$\sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,i}|\,|\,\mathcal{F}_{N,i-1}\right] = \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|f(\tilde{\xi}_{N,i})|\,\Big|\,\mathcal{F}_{N,i-1}\right] \xrightarrow{\mathrm{P}} \nu(|f|) < \infty\,,$$

whence the sequence $\{\sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,i}|\,|\,\mathcal{F}_{N,i-1}\right]\}_{N \geq 0}$ is tight. Next, for any positive ϵ and C we have for sufficiently large N,

$$\sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|U_{N,i}|\mathbf{1}_{\{|U_{N,i}| \geq \epsilon\}}\,\Big|\,\mathcal{F}_{N,i-1}\right] = \frac{1}{\tilde{M}_N} \sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|f(\tilde{\xi}_{N,i})|\mathbf{1}_{\{|f|(\tilde{\xi}_{N,i}) \geq \epsilon \tilde{M}_N\}}\,\Big|\,\mathcal{F}_{N,i-1}\right]$$

$$\leq \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathrm{E}\left[|f(\tilde{\xi}_{N,i})|\mathbf{1}_{\{|f|(\tilde{\xi}_{N,i}) \geq C\}}\,\Big|\,\mathcal{F}_{N,i-1}\right] \xrightarrow{\mathrm{P}} \mu\left(|f|\mathbf{1}_{\{|f| \geq C\}}\right)\,.$$

By dominated convergence the RHS of this display tends to zero as $C \to \infty$. Thus the LHS of the display converges to zero in probability, showing (39). $\quad\square$

Proof of the Theorem 4. Pick $f \in \tilde{\mathsf{A}}$ and write $\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} f(\tilde{\xi}_{N,i}) - \nu(f) = A_N + B_N$ with

$$A_N = \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i}\{f(\xi_{N,i}) - \nu(f)\}\,,$$

$$B_N = \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \left\{f(\tilde{\xi}_{N,i}) - \mathrm{E}\left[f(\tilde{\xi}_{N,i})\,\Big|\,\mathcal{F}_{N,i-1}\right]\right\}\,,$$

We first prove that

$$\mathrm{E}\left[\exp\left(\mathrm{i}u\,\tilde{M}_N^{1/2}\,B_N\right)\,\Big|\,\mathcal{F}_{N,0}\right] \xrightarrow{\mathrm{P}} \exp\left(-(u^2/2)\,\mathrm{Var}_\nu(f)\right)\,. \quad (64)$$

We will appeal to corollary 11 and hence need to check (53)-(54) with $U_{N,i} \overset{\mathrm{def}}{=} \tilde{M}_N^{-1/2} f(\tilde{\xi}_{N,i})$. First, because $\{(\xi_{N,i}, \omega_{N,i})\}_{1 \leq i \leq M_N}$ is consistent for (ν, C) and

for $f \in \tilde{\mathsf{A}}$, $f^2 \in \mathsf{C}$,

$$
\sum_{j=1}^{\tilde{M}_N} \{ \mathrm{E} \left[U_{N,j}^2 \,\middle|\, \mathcal{F}_{N,j-1} \right] - (\mathrm{E} \left[U_{N,j} \,\middle|\, \mathcal{F}_{N,j-1} \right])^2 \}
$$

$$
= \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f^2(\xi_{N,i}) - \left\{ \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \right\}^2 \xrightarrow{\mathrm{P}} \nu(f^2) - \{\nu(f)\}^2 = \mathrm{Var}_\nu(f) ,
$$

showing (53). Pick $\epsilon > 0$. For any positive constant C,

$$
\sum_{j=1}^{\tilde{M}_N} \mathrm{E} \left[U_{N,j}^2 \mathbf{1}_{\{|U_{N,j}| \geq \epsilon\}} \,\middle|\, \mathcal{F}_N \right] \leq \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathrm{E} \left[f^2(\tilde{\xi}_{N,j}) \mathbf{1}\{|f(\tilde{\xi}_{N,j})| \geq C\} \,\middle|\, \mathcal{F}_N \right]
$$

$$
= \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f^2(\xi_{N,i}) \mathbf{1}\{|f|(\xi_{N,i}) \geq C\} ,
$$

where the inequality holds for sufficiently large N. Since f^2 belongs to the proper set $\mathsf{C} \subseteq \mathsf{L}^1(\Xi, \nu)$, we have $f^2 \mathbf{1}\{|f| \geq C\} \in \mathsf{C}$. This implies that the RHS of the above display converges in probability to $\nu(f^2 \mathbf{1}\{|f| \geq C\})$. Because $f^2 \in \mathsf{C} \subseteq \mathsf{L}^1(\Xi, \nu)$, $\nu(f^2 \mathbf{1}\{|f| \geq C\})$ tends to zero as $C \to \infty$, so that (54) is satisfied.

Combining (64) with $a_N A_N \xrightarrow{\mathcal{D}} \mathrm{N}\left(0, \sigma^2(f)\right)$, we find that for any real numbers u and v,

$$
\mathrm{E} \left[\exp(\mathrm{i}(u\tilde{M}_N^{1/2} B_N + v a_N A_N)) \right] = \mathrm{E} \left[\mathrm{E} \left[\exp(\mathrm{i}u\tilde{M}_N^{1/2} B_N) \,\middle|\, \mathcal{F}_{N,0} \right] \exp(\mathrm{i}v a_N A_N) \right]
$$

$$
\to \exp\left\{ -(u^2/2) \mathrm{Var}_\nu(f) \right\} \exp\left\{ -(v^2/2)\sigma^2(f) \right\} .
$$

Thus the bivariate characteristic function converges to the characteristic function of a bivariate normal, implying that

$$
\begin{pmatrix} a_N A_N \\ \tilde{M}_N^{1/2} B_N \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}\left(0, \mathrm{diag}\left[\sigma^2(f), \mathrm{Var}_\mu(f)\right]\right) .
$$

Put $b_N = a_N$ if $\alpha < 1$ and $b_N = \tilde{M}_N^{1/2}$ if $\alpha \geq 1$. The proof follows from

$$
b_N(A_N + B_N) = (b_N a_N^{-1}) a_N A_N + (b_N \tilde{M}_N^{-1/2}) \tilde{M}_N^{1/2} B_N.
$$

The condition (5) and (6) are obviously fulfilled using that the weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}_{1 \leq i \leq \tilde{M}_N}$ is consistent for (ν, C). $\qquad \square$

D   Proof of Theorem 5

Proof of Theorem 5. To apply Corollary 11, we just have to check (53) and (54) where $U_{N,i} = a_N \tilde{M}_N^{-1} f(\tilde{\xi}_{N,i})$ and $\{\mathcal{F}_{N,i}\}$ defined by $\mathcal{F}_{N,0} = \sigma\{(\xi_{N,i})_{1 \leq i \leq M_N}\}$

and for all $1 \le k \le \tilde{M}_N$, $\mathcal{F}_{N,k} = \mathcal{F}_{N,0} \vee \sigma\{(\tilde{\xi}_{N,i})_{1 \le i \le k}\}$. Noting that $\tilde{\xi}_{N,j}$ is $\mathcal{F}_{N,0}$ measurable for $j = 1, \ldots, \bar{M}_N$, we have

$$
\begin{aligned}
A_N &= \sum_{i=1}^{\tilde{M}_N} \{ \mathrm{E}\left[U_{N,i}^2 \,|\, \mathcal{F}_{N,i-1} \right] - (\mathrm{E}\left[U_{N,i} \,|\, \mathcal{F}_{N,i-1} \right])^2 \} \\
&= \sum_{i=\bar{M}_N+1}^{\tilde{M}_N} \{ \mathrm{E}\left[U_{N,i}^2 \,|\, \mathcal{F}_{N,i-1} \right] - (\mathrm{E}\left[U_{N,i} \,|\, \mathcal{F}_{N,i-1} \right])^2 \} \\
&= \frac{a_N^2}{\tilde{M}_N} \frac{\tilde{M}_N - \bar{M}_N}{\tilde{M}_N} \left\{ \sum_{i=1}^{M_N} \tilde{\omega}_{N,i} f^2(\xi_{N,i}) - \left(\sum_{i=1}^{M_N} \tilde{\omega}_{N,i} f(\xi_{N,i}) \right)^2 \right\} \quad (65)
\end{aligned}
$$

where the weights $\tilde{\omega}_{N,i}$ are given in (21). Note that

$$
\tilde{\omega}_{N,i} = \frac{\Omega_N^{-1} \omega_{N,i} - \tilde{M}_N^{-1} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor}{1 - \tilde{M}_N^{-1} \sum_{i=1}^{M_N} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor} \quad \text{and} \quad \frac{\tilde{M}_N - \bar{M}_N}{\tilde{M}_N} = 1 - \frac{\sum_{i=1}^{M_N} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor}{\tilde{M}_N} \; .
$$

By applying Lemma 12,

$$
A_N = \frac{a_N^2}{\tilde{M}_N} \sum_{i=1}^{M_N} \left(\Omega_N^{-1} \omega_{N,i} - \tilde{M}_N^{-1} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor \right) f^2(\xi_{N,i}) -
$$

$$
\frac{a_N^2}{\tilde{M}_N} \frac{\left(\sum_{i=1}^{M_N} \left(\Omega_N^{-1} \omega_{N,i} - \tilde{M}_N^{-1} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor \right) f(\xi_{N,i}) \right)^2}{1 - \tilde{M}_N^{-1} \sum_{i=1}^{M_N} \lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \rfloor} \xrightarrow{\mathrm{P}} \tilde{\sigma}^2(f),
$$

It remains to check (5) and (6). By Theorem 3, the weighted sample $\{(\tilde{\xi}_{N,i}, 1)\}$ is consistent for (ν, C), which implies,

$$
\frac{a_N^2}{\tilde{M}_N} \frac{1}{\tilde{M}_N} \sum_{i=1}^{\tilde{M}_N} f(\tilde{\xi}_{N,i}) \xrightarrow{\mathrm{P}} \beta \nu(f)
$$

and thus (5) is satisfied. (6) is trivially satisfied. $\qquad \square$

Lemma 12. *Under the assumptions of Proposition 5, for any $f \in \mathsf{C}$,*

$$
\frac{1}{\tilde{M}_N} \sum_{i=1}^{M_N} \left\lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \right\rfloor f(\xi_{N,i}) \xrightarrow{\mathrm{P}} \nu \left(f \frac{\lfloor \ell \nu(1/\Phi)\Phi \rfloor}{\ell \, \nu(1/\Phi)\Phi} \right) \; .
$$

Proof. For any $K \ge 1$, denotes $\mathcal{B}_K = \bigcup_{j=0}^{\infty} [j - 1/K, j + 1/K]$. Because the weighted sample $\{(\xi_{N,i}, \Phi(\xi_{N,i})\}_{1 \le i \le M_N}$ is consistent for (ν, C) and $\left\lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \right\rfloor \le \tilde{M}_N \Omega_N^{-1} \omega_{N,i}$, we have for any $f \in \mathsf{C}$

$$
\tilde{M}_N^{-1} \sum_{i=1}^{M_N} \left\lfloor \tilde{M}_N \Omega_N^{-1} \omega_{N,i} \right\rfloor f(\xi_{N,i}) \mathbf{1} \left\{ \ell \nu(1/\Phi) \omega_{N,i} \in (K, \infty) \cup ([0, K] \cap \mathcal{B}_K) \right\}
$$

$$
\le \frac{1}{\Omega_N} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \mathbf{1} \left\{ \ell \nu(1/\Phi) \Phi(\xi_{N,i}) \in (K, \infty) \cup ([0, K] \cap \mathcal{B}_K) \right\}
$$

$$
\xrightarrow{\mathrm{P}} \int f(\xi) \mathbf{1} \left\{ \ell \nu(1/\Phi) \Phi(\xi) \in (K, \infty) \cup ([0, K] \cap \mathcal{B}_K) \right\} \nu(d\xi) \; .
$$

The RHS of the previous display can be made arbitrarily small by taking K sufficiently because $\int f(\xi)\mathbf{1}\left\{\ell\nu(1/\Phi)\Phi(\xi) \in \{\infty\} \cup \mathbb{N}\right\}\nu(d\xi) = 0$. For any $K \geq 1$, there exists $\eta > 0$ such that for any $a, b \in \mathbb{R}$,

$$\mathbf{1}\{|a - \ell\nu(1/\Phi)| \leq \eta\}\mathbf{1}\left\{b \in [0, K] \setminus \mathcal{B}_K\right\}\left(\lfloor ab \rfloor - \lfloor \ell\nu(1/\Phi)b \rfloor\right) = 0$$

Combining the previous equality with

$$\tilde{M}_N \Omega_N^{-1} = \frac{\tilde{M}_N}{M_N}\left(\Omega_N^{-1}\sum_{i=1}^{M_N}\omega_{N,i}\frac{1}{\Phi(\xi_{N,i})}\right) \xrightarrow{\mathrm{P}} \ell\nu(1/\Phi)$$

and

$$\tilde{M}_N^{-1}\sum_{i=1}^{M_N}\left\lfloor \ell\nu(1/\Phi)\Phi(\xi_{N,i}) \right\rfloor f(\xi_{N,i})\mathbf{1}\left\{\ell\nu(1/\Phi)\Phi(\xi) \in [0, K] \setminus \mathcal{B}_K\right\}$$

$$\xrightarrow{\mathrm{P}} \int \frac{\lfloor \ell\nu(1/\Phi)\Phi(\xi) \rfloor}{\ell\,\nu(1/\Phi)\Phi(\xi)}f(\xi)\mathbf{1}\left\{\ell\nu(1/\Phi)\Phi(\xi) \in [0, K] \setminus \mathcal{B}_K\right\}\nu(d\xi)$$

yields

$$\tilde{M}_N^{-1}\sum_{i=1}^{M_N}\left\lfloor \tilde{M}_N\Omega_N^{-1}\omega_{N,i} \right\rfloor f(\xi_{N,i})\mathbf{1}\left\{\ell\nu(1/\Phi)\Phi(\xi_{N,i}) \in [0, K] \setminus \mathcal{B}_K\right\}$$

$$\xrightarrow{\mathrm{P}} \int \frac{\lfloor \ell\nu(1/\Phi)\Phi(\xi) \rfloor}{\ell\,\nu(1/\Phi)\Phi(\xi)}f(\xi)\mathbf{1}\left\{\ell\nu(1/\Phi)\Phi(\xi) \in [0, K] \setminus \mathcal{B}_K\right\}\nu(d\xi)\,.$$

The proof follows by letting $K \to \infty$. $\qquad\qquad\qquad\qquad\qquad\square$

The condition $\nu\left\{\ell\nu(1/\Phi)\Phi \in \mathbb{N} \cup \{\infty\}\right\} = 0$ in Proposition 5 and Lemma 12 is crucial. Assume that $\{\xi_{N,i}\}_{1 \leq i \leq N}$ is an i.i.d. μ-distributed sample where μ is the distribution on the set $\{1/2, 2\}$ given by: $\mu(\{1/2\}) = 2/3$ and $\mu(\{2\}) = 1/3$. Let ν be the distribution on $\{1/2, 2\}$ given by: $\nu(\{1/2\}) = 1/3$ and $\nu(\{2\}) = 2/3$. The weighted sample $\{(\xi_{N,i}, \xi_{N,i})\}_{1 \leq i \leq N}$ (i.e. where we have set $\Phi(\xi) = \xi$) is a consistent sample for ν: for any function $f \in \mathbb{B}(\{1/2, 2\}) \stackrel{\text{def}}{=} \{f : \{1/2, 2\} \to \mathbb{R}, |f(1/2)| < \infty \quad \text{and} \quad |f(2)| < \infty\}$,

$$\frac{\sum_{i=1}^{M_N}\xi_{N,i}f(\xi_{N,i})}{\sum_{i=1}^{M_N}\xi_{N,i}}$$
$$\xrightarrow{\mathrm{P}} \frac{(1/2)f(1/2)\mu(1/2) + 2f(2)\mu(2)}{(1/2)\mu(1/2) + 2\mu(2)} = (1/2)f(1/2) + 1/3f(2) = \nu(f)\,.$$

In this example, $\ell = 1$ and obviously $\nu(1/\Phi) = 1$. Moreover,

$$\nu\left\{\Phi \in \{\infty\} \cup \mathbb{N}\right\} = \nu\left\{\{1/2, 2\} \cap \mathbb{N}\right\} = \nu(\{2\}) = 2/3 \neq 0\,.$$

We will show that the convergence in Lemma 12 fails. More precisely, setting $f(\xi) = \xi$, we will show that

$$\frac{1}{M_N}\sum_{i=1}^{M_N}\left\lfloor M_N\frac{\omega_{N,i}}{\Omega_N} \right\rfloor f(\xi_{N,i}) = \frac{1}{M_N}\sum_{i=1}^{M_N}\left\lfloor M_N\frac{\xi_{N,i}}{\sum_{j=1}^{M_N}\xi_{N,j}} \right\rfloor f(\xi_{N,i}) \xrightarrow{\mathcal{D}} 4/3 - (2Z)/3\,,$$

$$(66)$$

33

where Z is a Bernoulli variable with parameter $1/2$. This would imply that $M_N^{-1} \sum_{i=1}^{M_N} \left\lfloor M_N \frac{\omega_{N,i}}{\Omega_N} \right\rfloor f(\xi_{N,i})$ does not converge in probability to a constant. The LLN and CLT for i.i.d. random variables imply that

$$M_N^{-1} \Omega_N = M_N^{-1} \sum_{i=1}^{M_N} \xi_{N,i} \xrightarrow{\mathrm{P}} 1$$

$$\begin{bmatrix} \mathbf{1}\{M_N \Omega_N^{-1} < 1\} \\ \mathbf{1}\{M_N \Omega_N^{-1} \geq 1\} \end{bmatrix} = \begin{bmatrix} \mathbf{1}\{(M_N)^{1/2}(M_N^{-1}\Omega_N - 1) > 0\} \\ \mathbf{1}\{(M_N)^{1/2}(M_N^{-1}\Omega_N - 1) \leq 0\} \end{bmatrix} \xrightarrow{\mathcal{D}} \begin{bmatrix} Z \\ 1 - Z \end{bmatrix}$$

where Z is a Bernoulli random variable with parameter $1/2$. Since $\omega_{N,i} = \Phi(\xi_{N,i}) = \xi_{N,i} \in \{1/2, 2\}$ and $f(\xi) = \xi$,

$$\mathbf{1}\left\{ \frac{1}{2} < \frac{M_N}{\Omega_N} < \frac{3}{2} \right\} \frac{1}{M_N} \sum_{i=1}^{M_N} \left\lfloor M_N \frac{\omega_{N,i}}{\Omega_N} \right\rfloor f(\xi_{N,i})$$

$$= \mathbf{1}\left\{ \frac{1}{2} < \frac{M_N}{\Omega_N} < \frac{3}{2} \right\} \frac{2}{M_N} \sum_{i=1}^{M_N} \left\lfloor \frac{2M_N}{\Omega_N} \right\rfloor \mathbf{1}\{\xi_{N,i} = 2\}$$

$$= \mathbf{1}\left\{ \frac{1}{2} < \frac{M_N}{\Omega_N} < 1 \right\} \frac{2}{M_N} \sum_{i=1}^{M_N} \mathbf{1}\{\xi_{N,i} = 2\}$$

$$+ \mathbf{1}\left\{ 1 \leq \frac{M_N}{\Omega_N} < \frac{3}{2} \right\} \frac{4}{M_N} \sum_{i=1}^{M_N} \mathbf{1}\{\xi_{N,i} = 2\}$$

$$\xrightarrow{\mathcal{D}} (2Z)/3 + 4(1 - Z)/3 = 4/3 - (2Z)/3 .$$

The proof of (66) is concluded by noting that $\frac{M_N}{\Omega_N} \xrightarrow{\mathrm{P}} 1$.

References

ALDOUS, D. J. and EAGLESON, G. K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6** 325–331.

BERZUINI, C. and GILKS, W. R. (2001). Resample-move filtering with cross-model jumps. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. De Freitas and N. Gordon, eds.). Springer.

CAPPÉ, O., GUILLIN, A., MARIN, J. and ROBERT, C. (2005). Population Monte Carlo. *J. Comput. Graph. Statist.* **13** 907–929.

CHOPIN, N. (2004). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Ann. Statist.* **32** 2385–2411.

CRISAN, D. and DOUCET, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **50** 736–746.

CRISAN, D. and LYONS, T. (1997). Nonlinear filtering and measure-valued processes. *Probab. Theory Related Fields* **109** 217–244.

Csörgő, M. and Fischler, R. (1971). On mixing and the central limit theorem. *Tôhoku Math. J. (2)* **23** 139–145.

Del Moral, P. (1996). Nonlinear filtering: interacting particle solution. *Markov Process. Related Fields* **2** 555–579.

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

Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9** 275–297.

Del Moral, P. and Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, vol. 1729 of *Lecture Notes in Math.* Springer, Berlin, 1–145.

Douc, R., Guillin, A., Marin, J.-M. and Robert, C. (2005). Convergence of adaptive sampling schemes. Tech. Rep. 2005-6, CEREMADE.

Doucet, A., De Freitas, N. and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice.* Springer, New York.

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10** 197–208.

Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory.* Univ. California Press, Berkeley, Calif.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 473–483.

Gilks, W. R. and Berzuini, C. (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. *J. Roy. Statist. Soc. Ser. B* **63** 127–146.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York, London.

Hall, P. and Heyde, C. C. (1981). Rates of convergence in the martingale central limit theorem. *Ann. Probab.* **9** 395–404.

Handschin, J. and Mayne, D. (1969). Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering. In *Int. J. Control*, vol. 9.

Kong, A., Liu, J. S. and Wong, W. (1994). Sequential imputation and Bayesian missing data problems. *J. Am. Statist. Assoc.* **89** 590–599.

KÜNSCH, H. R. (2001). State space and hidden markov models. In *Complex Stochastic Systems* (O. E. Barndorff-Nielsen, D. R. Cox and C. Klueppelberg, eds.). CRC Publisher, Boca raton, 109–173.

KÜNSCH, H. R. (2003). Recursive Monte-Carlo filters: algorithms and theoretical analysis. Preprint ETHZ, seminar für statistics.

LANDAU, D. P. and BINDER, K. (2000). *A guide to Monte Carlo simulations in statistical physics.* Cambridge University Press, Cambridge.

LIU, J. (2001). *Monte Carlo Strategies in Scientific Computing.* Springer, New York.

LIU, J. and CHEN, R. (1995). Blind deconvolution via sequential imputations. *J. Roy. Statist. Soc. Ser. B* **430** 567–576.

LIU, J. and CHEN, R. (1998). Sequential Monte-Carlo methods for dynamic systems. *J. Roy. Statist. Soc. Ser. B* **93** 1032–1044.

LIU, J., CHEN, R. and LOGVINENKO, T. (2001). A theoretical framework for sequential importance sampling and resampling. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. De Freitas and N. Gordon, eds.). Springer.

LIU, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Stat. Comput.* **6** 113–119.

McLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628.

PETROV, V. V. (1995). *Limit Theorems of Probability Theory.* Oxford University Press.

PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.* **94** 590–599.

RISTIC, B., ARULAMPALAM, M. and GORDON, A. (2004). *Beyond Kalman Filters: Particle Filters for Target Tracking.* Artech House.

RUBIN, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong). *J. Am. Statist. Assoc.* **82** 543–546.

SHEPHARD, N. and PITT, M. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84** 653–667. Erratum in 91:249–250, 2004.

SHIRYAEV, A. N. (1996). *Probability.* 2nd ed. Springer.

TANIZAKI, H. (1999). On the nonlinear and nonnormal filter using rejection sampling. *IEEE Trans. Automat. Control* **44** 314–319.

TANIZAKI, H. (2001). Nonlinear and non-Gaussian state space modeling using sampling techniques. *Ann. Inst. Statist. Math.* **53** 63–81.