Comparison of Resampling Schemes for Particle Filtering

Randal Douc
Ecole Polytechnique
91128 Palaiseau, France
douc at cmapx.polytechnique.fr

Olivier Cappé
Centre National de la Recherche Scientifique
46 rue Barrault, 75634 Paris, France
cappe at tsi.enst.fr

Eric Moulines
GET Télécom Paris
46 rue Barrault, 75634 Paris, France
moulines at tsi.enst.fr

Abstract

This contribution is devoted to the comparison of various resampling approaches that have been proposed in the literature on particle filtering. It is first shown using simple arguments that the so-called residual and stratified methods do yield an improvement over the basic multinomial resampling approach. A simple counter-example showing that this property does not hold true for systematic resampling is given. Finally, some results on the large-sample behavior of the simple bootstrap filter algorithm are given. In particular, a central limit theorem is established for the case where resampling is performed using the residual approach.

1 Introduction

The terms *particle filtering* or *Sequential Monte Carlo* (henceforth abbreviated to SMC), refer to a class of techniques which have demonstrated a strong potential for signal and image processing applications [7], [17]. Schematically, the principle behind sequential Monte Carlo may be viewed as the combination of two main elements: *sequential importance sampling*, which dates back to [16, 12], and *resampling*, whose importance in the context of SMC was first demonstrated by [11], based on ideas of [18]. In this contribution, we focus on the second aspect and consider the comparison of several techniques that have been proposed to implement the resampling step.

To fix the notations, we briefly describe the basic SMC approach known as *sequential importance sampling with resampling* (or SISR). The algorithm proceeds as follows:

- At time 0, draw m particles $\{\xi_0^i\}_{1\leq i\leq m}$ from a common probability density r_0 and compute the associated importance weights $\omega_0^i = \nu_0(\xi_0^i)g_0(\xi_0^i)/r_0(\xi_0^i)$.
- For successive time indices and for $i = 1, \ldots, m$, simulate ξ_{k+1}^i independently from the past according to a transition density function[1] $r(\xi_k^i, \cdot)$ and update the weights as

$$\omega_{k+1}^i = \omega_k^i q(\xi_k^i, \xi_{k+1}^i) g_{k+1}(\xi_{k+1}^i)/r(\xi_k^i, \xi_{k+1}^i).$$

In the context of filtering, ν_0 us the initial distribution of the state variable, q is the transition density function corresponding to the, possibly non-linear, state equation (supposed here to be time-homogeneous), and g_k is the conditional likelihood of the observation at index k given the corresponding state, viewed as a function of the state variable. Then, the self-normalized importance sampling estimator $\sum_{i=1}^m \omega_k^i f(\xi_k^i) / \sum_{j=1}^m \omega_k^j$ is an estimator of the filtered state moment, that is the expectation of f applied to the non-observable state variable at time k given all observations up to time k. Not that the choice $r = q$ is particular in that the weight update formula then reduces to $\omega_{k+1}^i = \omega_k^i \, g_{k+1}(\xi_{k+1}^i)$ and thus depends only on the previous weight and new particle position; when used in conjunction with resampling ideas to be discussed below this choice ($r = q$) is known as the *bootstrap filter* [11].

The method sketched so far corresponds to the sequential importance sampling algorithm, whose drawback is that it becomes unstable as k increase due to the discrepancy between the weights – a phenomenon sometimes referred to as *weight degeneracy* [1, Chapter 7]. To stabilize the algorithm it is necessary to perform resampling sufficiently often. In the following, we denote by $\{\xi^i, \omega^i\}_{1\leq i\leq m}$ the set of particle positions and associated weights at some generic time index k (which is omitted from our notations) and by $\mathcal{G}^n$ the σ-field generated by the generations of particles and weights up to time k, included. We also assume that the

[1] In this contribution it is assumed that all transition kernels $K(x, dy)$ may be written as $k(x, y)\lambda(dy)$, where λ is a fixed reference measure (which we usually do not specify); k is referred to as a *transition density function*. When ν is a probability density function and f a function, we will use the usual notations $\nu(f) = \int \nu(x)f(x)\lambda(dx)$, $kf(x) = \int k(x, x')f(x')\lambda(dx')$, $\nu k(x) = \int \nu(x')\lambda(dx')k(x', x)$, and,
$\nu kf = \int \nu(x)kf(x)\lambda(dx) = \int \nu k(x)f(x)\lambda(dx)$
$= \iint \nu(x)k(x, x')f(x')\lambda(dx)\lambda(dx')$.

weights have already been normalized, i.e., that $\sum_{i=1}^m \omega^i = 1$. Resampling consists in selecting new particle positions and weights $\{\tilde{\xi}^i, \tilde{\omega}^i\}_{i=1,\ldots,\tilde{M}}$ such that the discrepancy between the resampled weights $\{\tilde{\omega}^i\}_{i=1,\ldots,\tilde{M}}$ is reduced. Of course, it is also necessary that the resampled particle system be as good an approximation to $\{\xi^i, \omega^i\}_{1\leq i\leq m}$ as possible, in some suitable sense. There are a number of options for performing resampling and we focus here on the most widely used class of resampling techniques in which the resampling is random and subject to the constraints

$$\tilde{M} = n, \tag{1}$$

$$\tilde{\omega}_k^i = 1/n, \tag{2}$$

$$\mathrm{E}\left[N^i \middle| \mathcal{G}^n\right] = n\omega_k^i, \quad \text{for } i = 1, \ldots, m, \tag{3}$$

where n is a non-random integer and $N^i \stackrel{\text{def}}{=} \#\{j, 1 \leq j \leq n : \tilde{\xi}^j = \xi^i\}$ are the particle duplication counts. The third constraint is sometimes known as the "unbiasedness" or "proper weighting" condition [15]. Of course, it is in general most natural to keep the population size fixed and n is often taken to be equal to m. In some situations however it does make sense to consider resampling scenarios in which n and m are different, at least for some time indices, and we thus keep separate notations for these two quantities.

Note that we do not consider here some important resampling algorithms that are either such that the population size varies (randomly) after resampling [4] or such that the weights are not constrained to be equal after resampling [10]. Our aim with the present contribution is to complement the results previously published on resampling in [15, 9, 14, 3] as well as to discuss some conjectures.

The rest of the paper is organized as follows: Section 2 briefly describes the four main resampling methods that have been proposed in the literature which satisfy the constraints mentioned above. Section 3 shows that residual and stratified resampling, as well as the combination of both, improve over multinomial resampling in the sense of having lower conditional variance. We also provide a counterexample which shows that the same property does not hold for systematic resampling, although its empirical performance is generally found to be close to that of residual and stratified resampling. Finally, we consider in Section 4 the large sample (i.e., when n increases) behavior of particle filtering methods which use these various forms of resampling. We are currently able to show that, in general, central limit theorems hold with the residual resampling approach, although the target and proposal distributions must satisfy a non trivial condition.

2 Description of Resampling Algorithms

2.1 Multinomial Resampling

The simplest approach to resampling is based on an idea at the core of the bootstrap method [8] that consists in drawing, conditionally upon $\mathcal{G}^n$, the new positions $\{\tilde{\xi}^i\}_{1\leq i\leq n}$ independently from the common point mass distribution $\sum_{j=1}^m \omega_j \delta_{\xi_j}$. In practice, this is achieved by repeated uses of the inversion method:

1. Draw n independent uniforms $\{U^i\}_{1\leq i\leq n}$ on the interval $(0, 1]$;
2. Set $I^i = D_\omega^{\mathrm{inv}}(U^i)$ and $\tilde{\xi}^i = \xi^{I^i}$, for $i = 1, \ldots, n$, where D_ω^{inv} is the inverse of the cumulative distribution function associated with the (normalized) weights $\{\omega^i\}_{1\leq i\leq m}$, that is, $D_\omega^{\mathrm{inv}}(u) = i$ for $u \in (\sum_{j=1}^{i-1} \omega^j, \sum_{j=1}^{i} \omega^j]$. When needed, we will denote by $\xi : \{1, \ldots, m\} \to \mathsf{X}$ the function such that $\xi(i) = \xi^i$, so that $\tilde{\xi}^i$ may also be written as $\xi \circ D_\omega^{\mathrm{inv}}(U^i)$.

This form of resampling is generally known as *multinomial resampling* since the duplication counts $N^1, \ldots, N^m$ are by definition distributed according to the multinomial distribution $\mathrm{Mult}(n; \omega^1, \ldots, \omega^m)$.

2.2 Residual Resampling

Residual resampling, or *remainder resampling*, is mentioned by [19], [15] as an efficient means to decrease the variance due to resampling. In this approach, for $i = 1, \ldots, m$, we have

$$N^i = \lfloor n\omega^i \rfloor + \bar{N}^i, \tag{4}$$

where $\lfloor\ \rfloor$ denotes the integer part and $\bar{N}^1, \ldots, \bar{N}^n$ are distributed according to the multinomial distribution $\mathrm{Mult}(n - R; \bar{\omega}^1, \ldots, \bar{\omega}^n)$ with $R = \sum_{i=1}^m \lfloor n\omega^i \rfloor$ and

$$\bar{\omega}^i = \frac{n\omega^i - \lfloor n\omega^i \rfloor}{n - R}, \quad i = 1, \ldots, m. \tag{5}$$

This scheme obviously satisfy (3). In practice, the multinomial counts $\bar{N}^1, \ldots, \bar{N}^n$ from the residual multinomial distribution are generated as in the multinomial resampling approach described above.

2.3 Stratified Resampling

Stratified resampling – see [13] and [9, Section 5.3] – is based on ideas used in survey sampling and consists in pre-partitioning the $(0, 1]$ interval into n disjoint sets, $(0, 1] = (0, 1/n] \cup \cdots \cup (\{n-1\}/n, 1]$. The U^is are then drawn independently in each of these sub-intervals: $U^i \sim \mathrm{U}\left(\left(\{i-1\}/n, i/n\right]\right)$, where $\mathrm{U}([a, b])$ denotes the uniform distribution on the interval $[a, b]$. Then the inversion method is used as in multinomial resampling. It is easily checked that, as was the case for residual sampling, the difference between the duplication count N^i and its target value $n\omega^i$ is less than one in absolute value (for all is). In addition,

$$\mathrm{E}\left[\sum_{i=1}^n f(\tilde{\xi}^i) \middle| \mathcal{G}^n\right] = \mathrm{E}\left[\sum_{i=1}^n f \circ \xi \circ D_\omega^{\mathrm{inv}}(U^i) \middle| \mathcal{G}^n\right]$$
$$= n \sum_{i=1}^n \int_{(i-1)/n}^{i/n} f \circ \xi \circ D_\omega^{\mathrm{inv}}(u)\, du = n \sum_{i=1}^m \omega^i f(\xi^i),$$

for all integrable functions f, showing that this algorithm also satisfies (3).

2.4 Systematic Resampling

Systematic resampling takes the previous method one step further by deterministically linking all the variables drawn in the sub-intervals. This is achieved by setting

$$U^i = (i-1)/n + U,$$

where U is a *single* random draw from the $\mathrm{U}((0,1/n])$ distribution. Since the U^is generated this way obviously have the same marginal distribution as those used in the stratified resampling approach, the method still satisfies (3). It was introduced in the particle filter literature by [2] as "stratified" sampling but it is also mentioned by [19] under the name of *universal* sampling. It is often preferred due to its computational simplicity and good empirical performance. As pointed out by [14] however, it is the only resampling method for which the resulting particle positions $\tilde{\xi}^i$ are no more independent given $\mathcal{G}^n$. Thus, studying its performance is much harder than for other methods.

A final remark of some importance is that both stratified and systematic resampling are sensitive to the order in which the particles are ordered: a simple permutation of the indices of the particles before resampling changes the distribution of the new resampled set of particles. In contrast, residual resampling behaves more like the basic multinomial resampling approach in that it disregards the order in which the particles are numbered.

3 Basic Properties of Sampling Schemes

3.1 Multinomial Resampling

For multinomial resampling, the selection indices $I^1, \ldots, I^n$ are conditionally i.i.d. given $\mathcal{G}^n$ and thus the conditional variance is given by

$$\begin{aligned}\mathrm{Var}&\left[\frac{1}{n}\sum_{i=1}^n f(\tilde{\xi}^i)\middle|\mathcal{G}^n\right]\\ &= \frac{1}{n}\left\{\sum_{i=1}^m \omega^i f^2(\xi^i) - \left[\sum_{i=1}^m \omega^i f(\xi^i)\right]^2\right\}. \quad (6)\end{aligned}$$

3.2 Residual Resampling

The residual sampling estimator may be decomposed into

$$\frac{1}{n}\sum_{i=1}^n f(\tilde{\xi}^i) = \sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n} f(\xi^i) + \frac{1}{n}\sum_{i=1}^{n-R} f(\xi^{\bar{I}^i}), \quad (7)$$

where $\bar{I}^1, \ldots, \bar{I}^{n-R}$ are conditionally independent given $\mathcal{G}^n$ with distribution $\mathrm{P}(\bar{I}^i = j \,|\, \mathcal{G}^n) = \bar{\omega}^j$ for $i = 1, \ldots, n-R$ and $j = 1, \ldots, m$. Because the residual resampling estimator is the sum of one term that, given $\mathcal{G}^n$, is deterministic and one term that involves conditionally i.i.d. draws, the conditional variance of residual resampling is given by

$$\begin{aligned}\frac{1}{n^2}\mathrm{Var}&\left[\sum_{i=1}^{n-R} f(\xi^{\bar{I}^i})\middle|\mathcal{G}^n\right] = \frac{n-R}{n^2}\mathrm{Var}\left[f(\xi^{\bar{I}^1})\middle|\mathcal{G}^n\right] \quad (8)\\ &= \frac{1}{n}\sum_{i=1}^m \omega^i f^2(\xi^i)\\ &\quad - \sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n^2} f^2(\xi^i) - \frac{n-R}{n^2}\left\{\sum_{i=1}^m \bar{\omega}^i f(\xi^i)\right\}^2.\end{aligned}$$

To compare (8) with (6), first write

$$\sum_{i=1}^m \omega^i f(\xi^i) = \sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n} f(\xi^i) + \frac{n-R}{n}\sum_{i=1}^m \bar{\omega}^i f(\xi^i).$$

Then note that the sum of the m numbers $\lfloor n\omega^i\rfloor/n$ plus $(n-R)/n$ equals one, whence this sequence of $m+1$ numbers can be viewed as a probability distribution. Thus Jensen's inequality applied to the square of the right-hand side of the previous display yields

$$\begin{aligned}&\left\{\sum_{i=1}^m \omega^i f(\xi^i)\right\}^2\\ &\quad\le \sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n} f^2(\xi^i) + \frac{n-R}{n}\left\{\sum_{i=1}^m \bar{\omega}^i f(\xi^i)\right\}^2.\end{aligned}$$

Combining with (8), this shows that the conditional variance of residual sampling is always smaller than that of multinomial sampling given by (6).

3.3 Stratified Resampling

Because $U^1, \ldots, U^n$ are still conditionally independent given $\mathcal{G}^n$ for this method,

$$\begin{aligned}&\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n f(\xi^{I^i})\middle|\mathcal{G}^n\right] =\\ &\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left[f\circ\xi\circ D_\omega^{\mathrm{inv}}(U^i)\middle|\mathcal{G}^n\right] =\\ &\frac{1}{n}\sum_{i=1}^m \omega^i f^2(\xi^i) - \frac{1}{n}\sum_{i=1}^n\left[n\int_{(i-1)/n}^{i/n} f\circ\xi\circ D_\omega^{\mathrm{inv}}(u)du\right]^2.\end{aligned}$$

By Jensen's inequality,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\left[n\int_{(i-1)/n}^{i/n} f\circ\xi\circ D_\omega^{\mathrm{inv}}(u)du\right]^2 \ge\\ &\quad\left[\sum_{i=1}^n\int_{(i-1)/n}^{i/n} f\circ\xi\circ D_\omega^{\mathrm{inv}}(u)du\right]^2 = \left[\sum_{i=1}^m \omega^i f(\xi^i)\right]^2,\end{aligned}$$

showing that the conditional variance of stratified sampling is always smaller than that of multinomial sampling. Note that stratified sampling may be coupled with the residual sampling method discussed previously: the proof above shows that using stratified sampling on the R residual indices that are indeed drawn randomly can then only decrease the conditional variance. It is also clear that the fact that the conditional variance is reduced does not depend on the particular choice of the sub-intervals (as being the intervals $(\{i-1\}/n, i/n]$), more general partitions could be considered as well.

3.4 Systematic Resampling

For this last sampling scheme, it is much more complicated to provide a usable expression of the conditional variance due to all the resampled particles being (conditionally) dependent [14]. We can however provide a simple counter-example to the frequently encountered conjecture that systematic resampling dominates multinomial resampling in terms of conditional variance.

Consider the case where the initial population of particles $\{\xi^i\}_{1\leq i\leq n}$ is composed of the interleaved repetition of only two distinct values x_0 and x_1, with identical multiplicities (assuming n to be even). In other words,

$$\{\xi^i\}_{1\leq i\leq n} = \{x_0, x_1, x_0, x_1, \ldots, x_0, x_1\}.$$

We denote by $2\omega/n$ the common value of the normalized weight ω^i associated to the $n/2$ particles ξ^i that satisfy $\xi^i = x_1$, so that the remaining ones (which are such that $\xi^i = x_0$) share a common weight of $2(1-\omega)/n$. Without loss of generality, we assume that $1/2 \leq \omega < 1$ and denote by $|f| = |f(x_1) - f(x_0)|$.

Under multinomial resampling, (6) shows that the conditional variance of the estimate $n^{-1}\sum_{i=1}^n f(\xi^i)$ is given by

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n f(\tilde{\xi}^i_{\mathrm{mult}})\middle|\mathcal{G}^n\right] = \frac{1}{n}(1-\omega)\omega|f|^2. \quad (9)$$

In this particular example, it is straightforward to verify that residual and stratified resampling are equivalent – which is not the case in general – and amount to deterministically setting $n/2$ particles to the value x_1 (because the value $2\omega/n$ is assumed to be larger than $1/n$), whereas the $n/2$ remaining ones are drawn by $n/2$ *conditionally independent* Bernoulli trials with probability of picking x_1 equal to $2\omega-1$. Hence the conditional variance, for both the residual and stratified schemes, is equal to $n^{-1}(2\omega-1)(1-\omega)|f|^2$. It is hence always smaller than (9), as expected from the general study of these two methods. Note that for specific configurations of the weights, such as when ω gets close to 0.5, the resampling becomes quasi-deterministic when using residual or stratified resampling and the improvement over the basic multinomial scheme becomes all the more significant.

In contrast, systematic resampling also deterministically sets $n/2$ of the $\tilde{\xi}^i$ to be equal to x_1 but depending on the draw of the initial shift, *all* the $n/2$ remaining particles are either set to x_1, with probability $2\omega-1$, or to x_0, with probability $2(1-\omega)$. Hence the variance is that of a *single* Bernoulli draw scaled by $n/2$, that is,

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n f(\tilde{\xi}^i_{\mathrm{syst}})\middle|\mathcal{G}^n\right] = (\omega-1/2)(1-\omega)|f|^2.$$

note that in this case, the conditional variance of systematic resampling is not only larger than (9) for most values of ω (except when ω is very close to $1/2$), but it does not even decrease to zero as n grows! Clearly, this observation is dependent on the order in which the initial population of particles is presented. It is easy to verify (using simulations) that, in this example, systematic resampling becomes very similar to residual/stratified resampling if the particles are randomly permuted before resampling. Hence, the above counter-example probably correspond to a "rare" situation. It does however show that systematic resampling is a variance reduction method which is not as robust as systematic and residual resampling and also suggest that theoretical study of the behavior of systematic resampling probably is a very hard task.

4 Large-Sample Behavior of Resampling

We now come to the question of assessing the large sample behavior of particle filtering methods based on various forms of resampling. The behavior of basic particle filtering methods when using the multinomial resampling has been extensively studied in [5]. For reasons of space and simplicity we only consider here the case of the bootstrap filter (i.e., when the transition kernel q of the hidden chain is used as proposal) where resampling is performed at each time index. In this basic case, each iteration of the particle filtering algorithm may be decomposed into two successive steps:

Prediction Given the population of unweighted particles at time index k, $\{\tilde{\xi}^i_k\}_{1\leq i\leq m}$, extend each trajectory conditionally independently according to $\xi^i_{k+1} \sim q(\tilde{\omega}^i_k, .)$;

Filtering After computing the weights as

$$\omega^i_{k+1} = g_{k+1}(\xi^i_{k+1}) / \sum_{j=1}^m g_{k+1}(\xi^j_{k+1}),$$

perform resampling to obtain the new unweighted population of particles $\{\tilde{\xi}^i_{k+1}\}_{1\leq i\leq n}$.

The choice of a particular resampling approach does obviously impact only on the second of these two steps.

To establish central limit theorems for the algorithm above, one can use repeatedly the two theorems below which are adapted from [1, Chapter 9] where the corresponding results are stated under slightly more general assumptions. The current population of particle is assumed to satisfy the following assumptions.

Assumption 1.

(i) $\{\xi^i\}_{1\le i\le m}$ *are consistent (in probability) and satisfy a central limit theorem (as* $m\to\infty$*) for a density* ν *and all bounded functions* f*, where* $\sigma^2(f)$ *denotes the asymptotic variance, that is,*

$$\frac{1}{m}\sum_{i=1}^m f(\xi^i) \xrightarrow{\mathrm{P}} \nu(f)$$

and

$$\sqrt{m}\left[\frac{1}{m}\sum_{i=1}^m f(\xi^i) - \nu(f)\right] \xrightarrow{\mathcal{D}} \mathrm{N}(0,\sigma^2(f))$$

for all bounded functions f*.*

(ii) The weights are given by $\omega^i = g(\xi^i)/\sum_{j=1}^m g(\xi^j)$*, where* $g(x) = \mu(x)/\nu(x)$ *for a probability density function* μ*;* g *is bounded from above and may be known up to a constant only.*

Theorem 2. *Under Assumption 1–(i), new particles* $\{\xi_+^i\}_{1\le i\le m}$ *distributed conditionally independently under* $\xi_+^i \sim q(\xi^i,\cdot)$ *are consistent for* νq *and all bounded functions* f *with asymptotic variance*

$$\sigma_+^2(f) = \nu\left[qf^2 - (qf)^2\right] + \sigma^2(qf) \tag{10}$$

Theorem 3. *Under Assumption 1, if* (a) *the resampled particles are conditionally independent given* $\mathcal{G}^n$*,* (b) $n\to\infty$ *with* $n/m\to\alpha$*, and,* (c)

$$n\,\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n f(\tilde\xi^i)\,\middle|\,\mathcal{G}^n\right] \xrightarrow{\mathrm{P}} \kappa(f) \tag{11}$$

that is deterministic, then $\{\tilde\xi^i\}_{1\le i\le n}$ *are consistent and satisfy a central limit theorem for* μ *and all bounded functions* f *with asymptotic variance*

$$\tilde\sigma^2(f) = \kappa(f) + \alpha\,\sigma^2\left(\frac{\mu}{\nu}[f-\mu(f)]\right) \tag{12}$$

Following the argument of [14, 3], by repeatedly applying Theorems 2 and 3 one may prove that the particle filter, when considered at any finite time index k, does satisfy a central limit theorem. The variance formula in (10) is a simple instance of the Rao-Blackwell theorem whereas (12) shows that the limit of the conditional variance of resampling gets added to the variance of (self-normalized or Bayesian) importance sampling scaled by the factor α. This latter factor is interesting as it shows that using $n \ll m$ may render the variance of the particle estimator almost independent of what happened in previous steps. This phenomenon should not be over-interpreted however as it only occurs because the sum is normalized by n, and not m (or $m+n$) which is more connected with the actual number of operations required to implement the method. Note that the requirement that g be bounded, which is not very restrictive in the filtering context, may be relaxed – see [1, Chapter 9] for details.

With multinomial resampling, (6) and the consistency directly implies that $\kappa(f) = \mu(f^2) - [\mu(f)]^2$ that is the variance under the target density μ. For other resampling schemes however, showing that (11) holds is all but trivial. We consider in the sequel the case of residual resampling. By (8),

$$\begin{aligned} &n\,\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n f(\tilde\xi^i)\,\middle|\,\mathcal{G}^n\right] \qquad (13)\\ &= \sum_{i=1}^m\left(\omega^i - \frac{\lfloor n\omega^i\rfloor}{n}\right) f^2(\xi^i) - \frac{n-R}{n}\left\{\sum_{i=1}^m \bar\omega^i f(\xi^i)\right\}^2\\ &= \sum_{i=1}^m\left(\omega^i - \frac{\lfloor n\omega^i\rfloor}{n}\right) f^2(\xi^i)\\ &\quad - \left\{\sum_{i=1}^m\left(\omega^i - \frac{\lfloor n\omega^i\rfloor}{n}\right) f(\xi^i)\right\}^2 \Bigg/ \left(1 - \sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n}\right). \end{aligned}$$

Under Assumption 1, for all bounded function f,

$$\sum_{i=1}^m \omega^i f^2(\xi^i) = \frac{m^{-1}\sum_{i=1}^m \frac{\mu}{\nu}(\xi^i) f^2(\xi^i)}{m^{-1}\sum_{i=1}^m \frac{\mu}{\nu}(\xi^i)} \xrightarrow{\mathrm{P}} \mu(f^2)$$

and $\sum_{i=1}^m \omega^i f(\xi^i) \xrightarrow{\mathrm{P}} \mu(f)$. However the case of sums that involve integer parts cannot be handled similarly and require the following technical lemma.

Lemma 4. *Under Assumption 1, if* $n\to\infty$ *with* $n/m\to\alpha$ *and* $\mu\left(\mathbb{1}_{\left\{x:\,\alpha\frac{\mu}{\nu}(x)\in\mathbb{N}\right\}}\right) = 0$*, then for all bounded function* f*,*

$$\sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n} f(\xi^i) \xrightarrow{\mathrm{P}} \nu\left\{\frac{1}{\alpha}\left\lfloor\frac{\alpha\mu}{\nu}\right\rfloor f\right\}.$$

Proof. Recall that $\omega^i = g(\xi^i)/\sum_{j=1}^m g(\xi^j)$ with $g(x) = \mu(x)/\nu(x)$. For any $K\ge 1$, define the set $\mathcal{B}_K = \bigcup_{j=0}^\infty [j - 1/K, j+1/K]$.

$$\begin{aligned} &\sum_{i=1}^m \frac{\lfloor n\omega^i\rfloor}{n} f(\xi^i)\mathbb{1}_{\{\alpha g(\xi^i)\in(K,\infty)\cup((0,K)\cap\mathcal{B}_K)\}}\\ &\le \sum_{i=1}^m \omega^i f(\xi^i)\mathbb{1}_{\{\alpha g(\xi^i)\in(K,\infty)\cup((0,K)\cap\mathcal{B}_K)\}}\\ &\xrightarrow{\mathrm{P}} \int f(x)\mathbb{1}_{\{\alpha g(x)\in(K,\infty)\cup((0,K)\cap\mathcal{B}_K)\}}\mu(x)\lambda(dx), \end{aligned}$$

where the notation $\mathbb{1}$ stands for the indicator function. The limit on the right-hand side of the last display can be made arbitrarily small by taking K sufficiently large because $\int f(x)\mathbb{1}_{\{\alpha g(x)\in\mathbb{N}\}}\mu(dx)\lambda(dx) = 0$ and g is bounded by Assumption 1. For any $K\ge 1$, there exists $\eta > 0$ such that

$$\begin{aligned} &\mathbb{1}_{\left\{\left|\frac{n}{\sum_{j=1}^m g(\xi^i)} - \alpha\right|\le\eta\right\}}\\ &\quad\times \mathbb{1}_{\{\alpha g(\xi^i)\in(0,K)\setminus\mathcal{B}_K\}}\left(\lfloor n\omega^i\rfloor - \lfloor\alpha g(\xi^i)\rfloor\right) = 0. \end{aligned}$$

Combining the above with $n/\sum_{j=1}^m g(\xi^i) \xrightarrow{\mathrm{P}} \alpha$ and

$$\sum_{i=1}^m \frac{\lfloor \alpha g(\xi^i) \rfloor}{n} f(\xi^i) \mathbb{1}_{\{\alpha g(\xi^i) \in (0,K) \setminus \mathcal{B}_K\}} \xrightarrow{\mathrm{P}} \int \frac{\lfloor \alpha g(x) \rfloor}{\alpha} f(x) \mathbb{1}_{\{\alpha g(x) \in (0,K) \setminus \mathcal{B}_K\}} \nu(x) \lambda(dx),$$

yields

$$\sum_{i=1}^m \frac{\lfloor n\omega^i \rfloor}{n} f(\xi^i) \mathbb{1}_{\{\alpha g(\xi^i) \in (0,K) \setminus \mathcal{B}_K\}} \xrightarrow{\mathrm{P}} \int \frac{\lfloor \alpha g(x) \rfloor}{\alpha} f(x) \mathbb{1}_{\{\alpha g(x) \in (0,K) \setminus \mathcal{B}_K\}} \nu(x) \lambda(dx).$$

The proof follows by letting $K \to \infty$. □

Corollary 5. *Under Assumption 1 and assuming that* $\mu\left(\mathbb{1}_{\{x:\, \alpha \frac{\mu}{\nu}(x) \in \mathbb{N}\}}\right) = 0$,

$$n \operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^n f(\tilde{\xi}^i) \middle| \mathcal{G}^n\right] \xrightarrow{\mathrm{P}} \kappa(f) = \nu\left\{\left(\frac{\mu}{\nu} - \frac{1}{\alpha}\left\lfloor \frac{\alpha\mu}{\nu} \right\rfloor\right) f^2\right\} - \left[\nu\left\{\left(\frac{\mu}{\nu} - \frac{1}{\alpha}\left\lfloor \frac{\alpha\mu}{\nu} \right\rfloor\right) f\right\}\right]^2 \Bigg/ \left(1 - \nu\left\{\frac{1}{\alpha}\left\lfloor \frac{\alpha\mu}{\nu} \right\rfloor\right\}\right)$$

for the residual sampling method. Hence, the resampled particles satisfy a central limit theorem with limiting variance given by (12).

The variance formula given in Corollary 5 was first derived in [3] which however lacked a rigorous proof of Lemma 4 and the necessity of the support condition – see [6] for a counter-example showing that this condition is indeed necessary and non-trivially satisfied. Note also that the asymptotic variance found in Corollary 5 is obtained as the (rescaled) limit of the conditional variance and is thus smaller than in the case where multinomial resampling is used (see Section 3.2).

5 Conclusions

In practical applications of sequential Monte Carlo methods, residual, stratified, and systematic resampling are generally found to provide comparable results. Despite the lack of complete theoretical analysis of its behavior, systematic resampling is often preferred because it is the simplest method to implement. From a theoretical point of view however only the residual and stratified resampling methods (as well as the combination of both) may be shown to dominate the basic multinomial resampling approach, in the sense of having lower conditional variance for all configurations of the weights. A central limit theorem as been established for the residual sampling approach. It is likely that a similar result can be obtained for stratified sampling, based on Theorem 3. The situation is however somewhat more involved in this latter case due to the fact that the new resampled particles, although still conditionally independent, have a distribution which depend on the order in which the particles are initially labelled.

References

[1] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

[2] J. Carpenter, P. Clifford, and P. Fearnhead. An improved particle filter for non-linear problems. *IEE Proc., Radar Sonar Navigation*, 146:2–7, 1999.

[3] N. Chopin. Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004.

[4] D. Crisan, P. Del Moral, and T. Lyons. Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields*, 5(3):293–318, 1999.

[5] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[6] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo. Preprint, May 2005.

[7] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

[8] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.

[9] P. Fearnhead. *Sequential Monte Carlo methods in filtter theory*. PhD thesis, University of Oxford, 1998.

[10] P. Fearnhead and P. Clifford. On-line inference for hidden Markov models via particle filters. *J. Roy. Statist. Soc. Ser. B*, 65:887–899, 2003.

[11] N. Gordon, D. Salmond, and A. F. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.*, 140:107–113, 1993.

[12] J. Handschin and D. Mayne. Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering. In *Int. J. Control*, volume 9, pages 547–559, 1969.

[13] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1:1–25, 1996.

[14] H. R. Künsch. Recursive Monte-Carlo filters: algorithms and theoretical analysis, 2003. Preprint ETHZ, seminar für statistics.

[15] J. Liu and R. Chen. Sequential Monte-Carlo methods for dynamic systems. 93:1032–1044, 1998.

[16] D. Q. Mayne. A solution of the smoothing problem for linear dynamic systems. *Automatica*, 4:73–92, 1966.

[17] B. Ristic, M. Arulampalam, and A. Gordon. *Beyond Kalman Filters: Particle Filters for Target Tracking*. Artech House, 2004.

[18] D. B. Rubin. 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). 82:543–546, 1987.

[19] D. Whitley. A genetic algorithm tutorial. *Stat. Comput.*, 4:65–85, 1994.