THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Adam M. Johansen, Sumeetpal S. Singh, Arnaud Doucet and Ba-Ngu Vo
Article Title: Convergence of the SMC Implementation of the PHD Filte
Year of publication: 2006
Link to published article:
http://dx.doi.org/10.1007/s11009-006-8552-y
Publisher statement: The original publication is available at www.springerlink.com

Methodology and Computing in Applied Probability manuscript No.
(will be inserted by the editor)

Convergence of the SMC Implementation of the PHD Filter

Adam M. Johansen[1], Sumeetpal S. Singh[1], Arnaud Doucet[2] and Ba-Ngu Vo[3]

[1] Department of Engineering, Trumpington Street, Cambridge, CB2 1PZ, UK
e-mail: amj26@cam.ac.uk and e-mail: sss40@cam.ac.uk

[2] Department of Computer Science, 201-2366 Main Mall and Department of Statistics, 333-6356 Agricultural Road, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada e-mail: arnaud@cs.ubc.ca

[3] Department of Electrical and Electronic Engineering, University of Melbourne, VIC 3010, Australia e-mail: bv@ee.unimelb.edu.au

The date of receipt and acceptance will be inserted by the editor

Abstract The probability hypothesis density (PHD) filter is a first moment approximation to the evolution of a dynamic point process which can be used to approximate the optimal filtering equations of the multiple-object tracking problem. We show that, under reasonable assumptions, a sequential Monte Carlo (SMC) approximation of the PHD filter converges in mean of order $p \geq 1$, and hence almost surely, to the true PHD filter. We also present a central limit theorem for the SMC approximation, show that the variance is finite under similar assumptions and establish a recursion for the asymptotic variance. This provides a theoretical justification for this implementation of a tractable multiple-object filtering methodology and generalises some results from sequential Monte Carlo theory.

AMS 2000 Subject Classification Numbers
Primary: 60F05 Secondary: 60F25; 62P30; 93E11

Key words central limit theorem–filtering–sequential Monte Carlo–finite random sets

1 Introduction

In a standard Hidden Markov Model (HMM), the state and measurement at time k are two vectors of possibly different dimensions, belonging to E and F respectively. These vectors evolve randomly over time but their dimensions are fixed. The aim is to compute recursively in time the distribution of the hidden state given all the observations that have been received so far. In *multi-object filtering*, recently introduced and studied by the data-fusion and tracking community [9, 11], the aim is to perform filtering when the state and observation variables are the finite subsets of E and F. Conceptually, this problem can be thought of as that of performing filtering when the state and observation spaces are the disjoint unions, $\uplus_{i=0}^{\infty} E^i$ and $\uplus_{i=0}^{\infty} F^i$, respectively. We remark that developing efficient computational tools to propagate the posterior density is extremely difficult in this setting [8].

An alternative which is easier to approximate computationally, the Probability Hypothesis Density (PHD) filter, has recently been proposed [11]. The PHD filter is a recursive algorithm that propagates the first moment, also referred to as the *intensity* [4], of the multi-object posterior. The first moment is an appropriately defined measure on E (although we also use the term to refer to the Radon-Nikodým derivative of this measure with respect to some appropriately defined dominating measure on the same space). While the first moment is now a function on E, i.e. the dimension of the "state space" is now fixed, the PHD filter recursion still involves multiple integrals that have no closed form expressions in general. An SMC implementation of the PHD filter was proposed in [14].

The aim of this paper is to analyse the convergence of the sequential Monte Carlo (SMC) implementation of the PHD filter proposed in [14]. SMC is a class of computational methods for the sequential approximation of integrals via a sequential importance sampling and resampling strategy [5, 7, 12]. Although numerous convergence results and central limit theorems have been obtained for particle systems which approximate Feynman-Kac flows [5] (including the optimal filtering equations), the PHD filter, being a first moment of the multi-object posterior, is an unnormalised density that does not obey the standard Bayes recursion. Thus, convergence results and central limit theorems which have been derived for Feynman-Kac flows do not apply to the SMC approximation of the PHD filter. Our contribution is to extend existing results to this system which has a number of added difficulties, particularly that the total mass of the filter is a time-varying quantity and the recursions are non-standard.

2 Background and Problem Formulation

2.1 Notation and Conventions

It is convenient, at this stage, to summarise the notation used throughout the remainder of this report, and the conventions which have been adopted. It is assumed throughout that the particle system first introduced in section 2.3.1 is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. All expectations and probabilities which are not explicitly associated with some other measure are taken with respect to $\mathbb{P}$.

For some measurable space $(E, \mathcal{E})$, let the set of measurable functions on E be denoted by $\mathcal{B}(E)$, the space of bounded measurable functions (endowed with the supremum norm, $||\xi||_\infty = \sup_{u \in E} |\xi(u)|$ for any $\xi : E \to \mathbb{R}$) by $\mathcal{B}_b(E)$, and the set of finite measures by $\mathcal{M}(E)$. The symbol $\mathbf{1}$ is used to denote the unit function on any space. We have assumed throughout that all measures admit a density with respect to some dominating measure, $\lambda(dx)$, and used the same symbol to represent a density and its associated measure, i.e. for some measure $\mu \in \mathcal{M}(E)$,

$$\mu(dx) = \mu(x)\lambda(dx)$$

Given a measure μ the integral of a function, f, with respect to μ is denoted $\mu(f)$.

A Markov Kernel K from E to E induces two operators. One acts upon functions in $\mathcal{B}_b(E)$ and takes values in $\mathcal{B}_b(E)$ and may be defined as:

$$\forall u \in E \quad \forall f \in \mathcal{B}_b(E) \quad K(f)(u) = \int_E K(u, dv) f(v)$$

and the other acts upon finite measures on E and takes values in $\mathcal{M}(E)$ and is defined by

$$\forall \mu \in \mathcal{M}(E) \quad \mu K(\cdot) = \int_E \mu(du) K(u, \cdot)$$

Given two transition kernels K and L which admit a density with respect to a suitable dominating measure, where L is absolutely continuous with respect to K, $K >> L$, we define $\frac{L}{K}(u,v) = \frac{dL(u,\cdot)}{dK(u,\cdot)}(v)$ (i.e. the Radon-Nikodým derivative). Given a transition kernel K and a non-negative function $g : E \times E \to \mathbb{R}^+$ we define the new kernel $K \times g$ by $K \times g(u, dv) = K(u, dv) g(u, v)$. Similarly, for two measures μ and ν on E, we define $\frac{\mu}{\nu}(u)$ to be $\frac{d\mu}{d\nu}(u)$. If μ and ν both admit densities with respect to the same dominating measure λ then $\frac{d\mu}{d\nu}(u)$ is simply the ratio of those densities evaluated at the point u. For any two functions $f, g : E \to \mathbb{R}$ we write fg for the standard multiplication of these functions.

Where it is necessary to describe matrices in terms of their components, we write $A = [a_{ij}]$ where a_{ij} is the expression for component i, j of matrix A.

When dealing with random finite sets, the convention in the literature is to use capital Greek letters to refer to a random set, a capital Roman letter to refer to a realisation of such a set and a lower case Roman letter to refer to an element of a realisation. We have followed this convention wherever possible.

Finally, we have considered the evolution of the PHD as an unnormalised density on a general space E. It is usual, but not entirely necessary, to assume that $E = \mathbb{R}^d$ and that the dominating measure $\lambda(dx)$ is Lebesgue measure. For the target tracking application described in section 2.4, this is, indeed, the case.

2.2 Multiple Object Filtering

We remark that although the description below is removed from any particular application, the model is popular with the data fusion and target tracking community [9,11,8]. Our intention in giving this abstract presentation is to emphasise the generality of the model with the intention of arousing the interest of other scientific communities.

The multi-object state evolves over time in a Markovian fashion and at each time k, a multi-object measurement is generated based upon the state at time k only. The multi-object state and multi-object measurement at time k are naturally represented as finite subsets $X_k \subset E$ and $Z_k \subset F$ respectively. For example, at time k, let X_k have $M(k)$ elements, i.e.,

$$X_k = \{x_{k,1}, \ldots, x_{k,M(k)}\} \in \mathcal{T}(E)$$

where $\mathcal{T}(E)$ denotes the collection of all finite subsets of the space E. Similarly, if $N(k)$ observations $z_{k,1}, \ldots, z_{k,N(k)}$ from F are received at time k, then

$$Z_k = \{z_{k,1}, \ldots, z_{k,N(k)}\} \in \mathcal{T}(F)$$

is the multi-object measurement. Analogous to the standard HMM case, in which uncertainty is characterised by modelling the states and measurements by random vectors, uncertainty in a multi-object system is characterised by modelling multi-object states and multi-object measurements as random finite sets (RFS) Ξ_k and Σ_k in E and F respectively. We denote particular realisations of Ξ_k and Σ_k by X_k and Z_k respectively. Conditioned upon a realisation X_{k-1} of the state at time $k-1$, Ξ_{k-1} the state evolution satisfies

$$\Xi_k = \Xi_k^S(X_{k-1}) \cup \Xi_k^B(X_{k-1}) \cup \Gamma, \tag{1}$$

where $\Xi_k^S(X_{k-1})$ denotes the RFS of elements that have 'survived' to time k and the other terms are RFSs of new elements, which are decomposed as $\Xi_k^B(X_{k-1})$ of elements spawned (spawning is a term used in the tracking literature for the process by which a large target, such as an aircraft carrier, emits a number of smaller targets, such as aircraft) from X_{k-1} and the RFS Γ_k of elements that appear spontaneously at time k. Note that the state evolution model incorporates individual element motion, element birth, death and spawning, as well as interactions between the elements. Similarly, given a realisation X_k of Ξ_k at time k, the observation Σ_k is modelled by

$$\Sigma_k = \Theta_k(X_k) \cup \Lambda_k, \tag{2}$$

where $\Theta_k(X_k)$ denotes the RFS of measurements generated by X_k, and Λ_k denotes the RFS of measurements that do not originate from any element in X_k, such as false measurements due to sensor noise or objects other than the class of objects of interest. The observation process so defined can capture element measurement noise, element-dependent probability of occlusions and false measurements.

The multi-object filtering problem concerns the estimation of the multi-object state X_k at time step k given the collection $Z_{1:k} \equiv (Z_1, ..., Z_k)$ of all observations up to time k. The object of interest is the posterior probability density of Ξ_k.

The above description of the dynamics of $\{\Xi_k\}$ and $\{\Sigma_k\}$ was a constructive one, while in filtering one needs to specify the state transition and observation density, that is, the densities of the following measures,

$$P(\Xi_k \in A | \Xi_{k-1} = X_{k-1}),$$
$$P(\Sigma_k \in B | \Xi_k = X_k),$$

where $A \subset \mathcal{T}(E)$ and $B \subset \mathcal{T}(F)$ are the measurable sets of their respective spaces. As this paper is concerned with the propagation of the first moment of the filtering density, we refer the reader to [14, 11] for details on the state transition and observation densities. We have also omitted details on how the RFSs of survived elements $\Xi_k^S(X_{k-1})$, spawned elements $\Xi_k^B(X_{k-1})$ and spontaneously spawned elements Γ_k are constructed. Similarly, details on the RFSs of true (or element generated) observations $\Theta_k(X_k)$ and false measurements Λ_k were omitted. Naturally, the construction of these sets are application specific and a simple numerical example provided in Section 2.4 below aims to clarify the ideas. We refer the reader to [9,11] for the constructions for applications in target tracking.

As was observed previously, it is extremely difficult to perform the computations involved in the filtering problem for this type of model. SMC methods cannot operate efficiently when direct importance sampling on a very high dimensional space is involved. Thus it is important to consider computationally tractable principled approximations. This leads us to the PHD filter, one such approximation which has become popular among the tracking community [11,14].

2.3 The PHD Filter

The PHD filter is a method of updating a measure, $\tilde{\alpha}_{k-1}$ given a random set of observations, Z_k, which can be interpreted as a first moment approximation of the usual Bayesian filtering equation. Within this framework, the quantity of interest is the intensity measure of a point process. Whilst it can be described by a measure, it is not in general a probability measure and it is necessary to maintain an estimate of both the total mass and the distribution of that mass. Details now follow.

Before summarising the mathematical formulation of the PHD filtering recursion, we briefly explain what is meant by the *first moment* of a random finite set. A finite subset $X \in \mathcal{T}(E)$ can also be equivalently represented by the counting measure N_X (on the measurable subsets of E) defined, for all measurable sets, A, by $N_X(A) = \sum_{x \in X} \mathbf{1}_A(x) = |A \cap X|$. Consequently, the random finite set Ξ can also be represented by a random counting measure N_Ξ defined by $N_\Xi(A) = |\Xi \cap A|$. This representation is commonly used in the point process literature [4].

The first moment of a random vector is simply the expectation of that random vector under a suitable probability measure. As there is no concept of set addition, an exact analogue of this form of moment is not possible in the RFS case. However, using the random counting measure representation, the 1st moment or *intensity measure* of a RFS Ξ is the first moment of its associated counting measure, i.e.,

$$\tilde{\alpha}(A) = \mathbb{E}\left[N_\Xi(A)\right].$$

The intensity measure of a set A gives the expected number of elements of Ξ that are in A. Although the intensity measure $\tilde{\alpha}$ is an integral of the counting measures, it is not itself a counting measure and hence does not necessarily have a finite set representation.

The density of the intensity measure with respect to a suitable dominating measure λ, when it exists, is also denoted $\tilde{\alpha}$ and is termed the *intensity function*[1]. In the tracking literature, $\tilde{\alpha}$ is also known as the Probability Hypothesis Density (PHD).

The PHD is the first moment of a RFS and hence tells us, for any region, the expected number of elements within that region. In the context of multi-object filtering, the PHD recursion described below propagates the density of the intensity measure $\tilde{\alpha}_k(A) := \mathbb{E}\left[N_{\Xi_k}(A) \middle| Z_1, \ldots, Z_k\right]$ for $k \geq 0$. This is clearly a useful

[1] As a reminder, we use the same notation for a measure and its density throughout.

representation for multi-object filtering and other applications, as it provides a simultaneous description of the number of elements of Ξ_k within the space, and their locations.

The PHD recursion can be described in terms of *prediction* and *update* steps, just as the optimal filtering recursion can. The derivation of the update step cannot be reproduced here due to space constraints, but the most elegant approach involves considering the evolution of the probability generating functional associated with a Poisson process under the action of the update step. All of this is presented in detail by [11]:

$$\alpha_k(dx) = (\Phi_k \tilde{\alpha}_{k-1})(dx) = (\tilde{\alpha}_{k-1}\phi_k)(dx) + \gamma_k(dx) \quad (3)$$

$$\tilde{\alpha}_k(dx) = (\Psi_k \alpha_k)(dx) = \left(\nu_k(x) + \sum_{z \in Z_k} \frac{\psi_{k,z}(x)}{\kappa_k(z) + \alpha_k(\psi_{k,z})}\right) \alpha_k(dx) \quad (4)$$

The prediction operator Φ_k is described in terms of a kernel, ϕ_k, which does not in general integrate to 1, and an additive measure, γ_x. The prediction kernel, ϕ_k describes the dynamics of existing elements and can be decomposed as: $\phi_k(x, dy) = e_k(x) f_k(x, dy) + b_k(x, dy)$ where $e_k(x)$ is the probability of an element at x at time $k-1$ surviving to time k, $f_k(x, dy)$ is a Markov kernel which describes the dynamics of the surviving elements and $b_k(x, dy)$ is a "spawning" kernel which describes the probability of an element at x at time $k-1$ giving rise to a new element in a neighbourhood dy at time k.

The update operator Ψ_k is a nonlinear operator which resembles a linear combination of Boltzmann-Gibbs operators (one of which describes the update equation of Bayesian filtering) with different associated potentials. However, there are some subtle differences which prove to be significant. The term Z_k denotes the random set of observations at time k and $\psi_{k,z}$ is the "likelihood" function associated with an observation at z at time k. $\kappa_k(z)$ is the intensity of the false measurement process at z. Finally, $\nu_k(x)$ is the probability of failing to observe an element at x at time k.

Note the correspondence between the terms in the PHD recursion and the sets in the constructive description of the multi-object filtering problem in section 2.2. The pairing of the terms are as follows: (Ξ_k^B, b_k) describe object birth including spawning, (Γ_k, γ_k) describe spontaneous births, and $(\Xi_k^S, e_k f_k)$ describe the dynamics of surviving objects. The measurement model has a more subtle relationship, Θ_k incorporates all of the information of $\psi_{k,z}$ and ν_k while the effect of Λ_k on the first moment is described by κ_k.

2.3.1 An SMC Implementation of the PHD Filter We consider essentially the algorithm proposed in [14] which describes a sequential Monte Carlo method for approximating the evolution of the PHD filter. It is assumed that the filter is initialised at time zero by sampling a set of L_0 particles from the true PHD (or, rather, the probability measure obtained by appropriately normalising it) and weighting them according to the total mass at time zero such that each particle has weight $w_0^{(i)} = \tilde{\alpha}_0(\mathbf{1})/L_0$. The following recursion is then used to predict the particles description at the next time step and then to update the estimate based upon the next observation set, just as in the standard filtering case. It is understood that the importance densities used may be conditioned upon the current observation set in addition to the previous particle position. We omit the dependency on the observation set in our notation.

Assume that a particle approximation consisting of L_{k-1} weighted particles is available at time $k-1$, with associated empirical measure $\tilde{\alpha}_{k-1}^{L_{k-1}}$.

1. Prediction:
   – Propagate forward the particles which survived the previous iteration to account for the dynamics of existing objects. For $i = 1, \ldots, L_{k-1}$, sample $Y_k^{(i)}$ from some importance distribution $q_k(X_{k-1}^{(i)}, \cdot)$ and calculate the importance weights

$$\tilde{w}_k^{(i)} = \frac{\phi_k\left(X_{k-1}^{(i)}, Y_k^{(i)}\right)}{q_k\left(X_{k-1}^{(i)}, Y_k^{(i)}\right)} w_{k-1}^{(i)} \tag{5}$$

   – Generate some new particles to account for spontaneous births. For $i = L_{k-1}+1, \ldots, L_{k-1}+J_k$, sample $Y_k^{(i)}$ from some importance distribution $p_k(\cdot)$ and calculate the importance weights

$$\tilde{w}_k^{(i)} = \frac{1}{J_k} \frac{\gamma_k(Y_k^{(i)})}{p_k(Y_k^{(i)})} \tag{6}$$

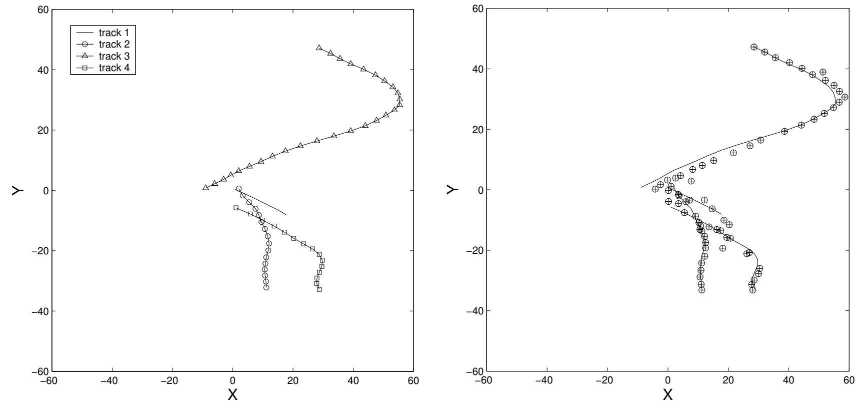
   – Let $M_k = J_k + L_{k-1}$ and let $\alpha_k^{M_k}$ denote the particle approximation to the predicted PHD filter at time k comprising these two weighted particle sets
2. Update:
   – Compute the empirical estimate of the normalising constant associated with each observation,

$$C_k(z) = \kappa_k(z) + \sum_{i=1}^{L_{k-1}+J_k} \tilde{w}_k^{(i)} \psi_{k,z}(Y_k^{(i)})$$

   – Adjust the particle weights to reflect the most recent observations. Update all the particle weights with:

$$\hat{w}_k^{(i)} = \left[\nu\left(Y_k^{(i)}\right) + \sum_{z \in Z_k} \frac{\psi_{k,z}\left(Y_k^{(i)}\right)}{C_k(z)}\right] \tilde{w}_k^{(i)}$$

3. Resampling:
   – Estimate the total mass: $\hat{N}_k = \sum_{j=1}^{L_{k-1}+J_k} \hat{w}_k^{(j)}$
   – Resample to reduce sample impoverishment (that is, the presence of a large (and increasing in time) number of particles with very small weights) and to prevent exponential growth of the size of the particle ensemble. Starting from the particle/weight pairs $\left\{\frac{\hat{w}_k^{(i)}}{\hat{N}_k}, Y_k^{(i)}\right\}_{i=1}^{L_{k-1}+J_k}$ sample L_k particles



(a) Ground truth (b) PHD filter output estimates ($\oplus$)

Fig. 1 Plots of 4 superimposed tracks over 40 time steps.

from the empirical probability distribution obtained by suitably normalising it, to obtain a set of L_k particles of equal weight $\left\{w_k^{(i)}/\hat{N}_k, X_k^{(i)}\right\}_{i=1}^{L_k}$

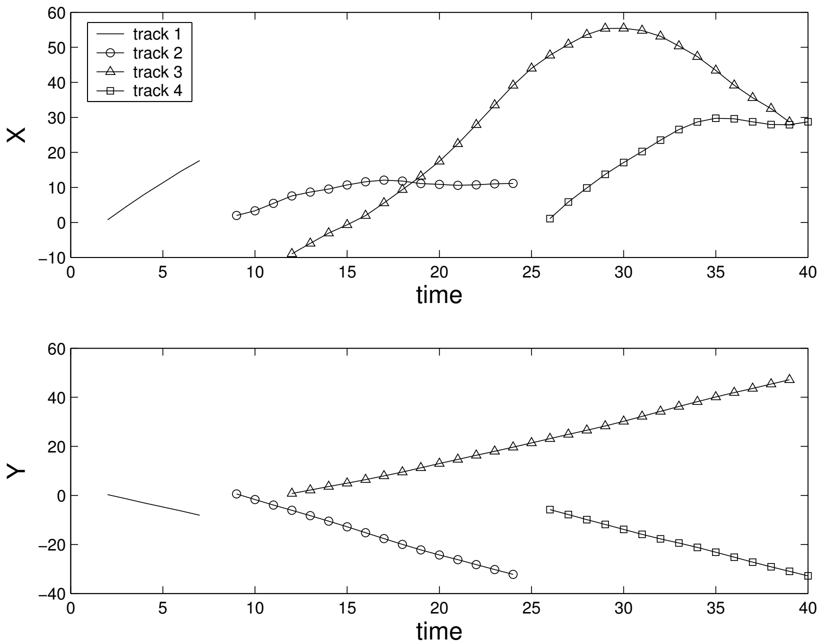
– Rescale the weights to reflect the total mass of the system (i.e. multiply the particle weights by a factor of $\hat{N}_k$) giving the particle/weight ensemble $\left\{w_k^{(i)}, X_k^{(i)}\right\}_{i=1}^{L_k}$ which defines $\tilde{\alpha}_k^{L_k}$.

2.4 A Motivating Example

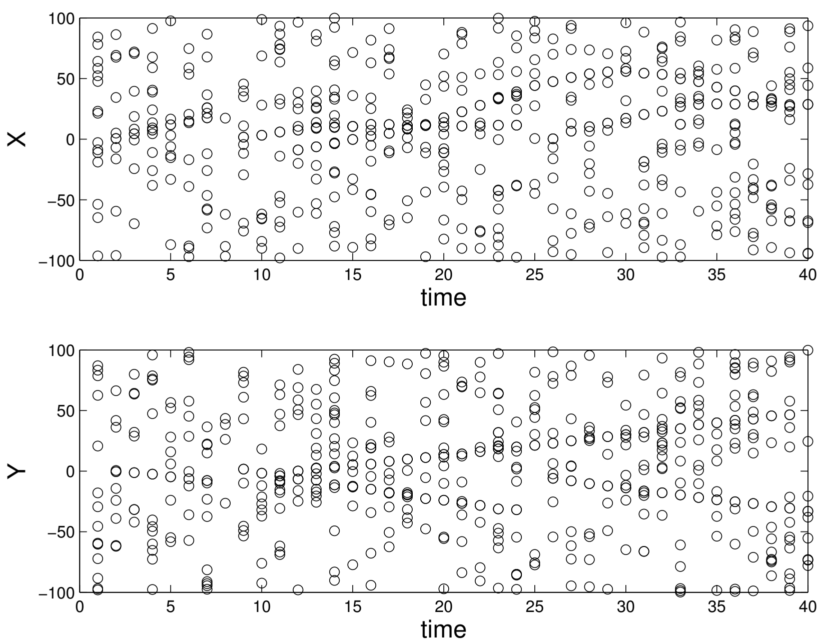
We present a brief example (which is taken from [14]) to illustrate the utility of the multi-object filtering framework and the SMC implementation of the PHD filter. Consider the problem of tracking an unknown number of targets that evolve in $\mathbb{R}^4$. For instance, in a two dimensional tracking example, each target could be described by its x and y coordinates as well as its velocity in these directions. Existing targets can leave the surveillance area and new targets can enter the scene. At time k, a realisation of the state is $X_k = \{x_{k,1}, \ldots, x_{k,M(k)}\} \subset \mathbb{R}^4$. As for the observations, each target generates one observation with a certain probability (i.e. each target generates at most one observation) and, the sensors can measure false observations that are not associated with any target, i.e., clutter. Assume that sensors measure a noisy value of the x and y coordinate of a target. A realisation of the observation would be $Z_k = \{z_{k,1}, \ldots, z_{k,N(k)}\} \subset \mathbb{R}^2$ where measurement $z_{k,i}$ could either correspond to an element in X_k or be a false measurement. Note that the number of observations need not coincide with the number of targets.

We now demonstrate the results of tracking the targets using the SMC implementation of the PHD filter. In our example each target moves according to a standard linear Gaussian model. Each existing target has a probability of survival that is independent of its position and velocity, i.e., a target at time $k-1$ survives to time k with probability 0.95. For simplicity no spawning is considered. At each time k, new targets can appear spontaneously according to a Poisson point process with an intensity function γ_k set to $0.2\mathcal{N}(\cdot;\bar{x},Q)$, where $\mathcal{N}(\cdot;\bar{x},Q)$ denotes a normal density with mean $\bar{x}$ and uncertainty corresponding to the covariance, Q. This corresponds to one new target being created every five time steps around a location $\bar{x}$ with covariance Q. As for the observations, each target generates a noisy observation of its position with certain probability. Additionally, false measurements are generated according to a Poisson point process with a uniform intensity function.

The peaks of $\tilde{\alpha}_k$ are points in E with the highest local concentration of the expected number of targets, and hence may be used to generate estimates for the location of the elements of Ξ. Since the total mass of the intensity measure gives the expected number of targets, the simplest approach is to round the particle estimate of this quantity to the closest integer, $\hat{N}_k$ and then to select the $\hat{N}_k$ largest peaks as target locations. This was the approach adopted in this numerical example, for which the positions of 4 targets over 40 time steps are displayed in Figure 1(b). These 4 targets start in the vicinity of the origin and move radially outwards. The start and finish times of each target can be seen from Figure 2(a), which plots the individual x and y components of each track against time. The x and y coordinates of the observations Z_k for all 40 time steps are shown in Figure 2(b). Figure 1(b) shows the position estimates superimposed on the true tracks over the 40 time steps. Observe the close proximity of the estimated positions to the true tracks even though the tracks of the targets were not strictly generated according to the assumed model.



(a) Ground truth: plots of x and y components of the 4 true tracks against time, showing the different start and finish times of the tracks.



(b) x and y components of position observations immersed in clutter of rate $r = 10$.

Fig. 2 True target positions and generated observations as a function of time.

3 Convergence Study

It is shown that the integral of any bounded test function under the SMC approximation of the PHD filter converges to the integral of that function under the true PHD filter in mean of order p (for all integer p) and hence almost surely. As observed by one referee, the restriction that test functions must be bounded seems more reasonable in the context of the PHD filter than the standard optimal filter as one is typically interested in the integrals of indicator functions. The result is shown to hold recursively by decomposing the evolution of the filter into a number of steps at each time. A number of additional points need to be considered in the present case. We assume throughout that the observation record $\{Z_k\}_{k\geq 0}$ is fixed and generates the PHD recursion.

Remark 1 As a preliminary, we need to show that both the true and approximate filters have finite mass at all times. In the case of the true filter this follows by assuming that the mass is bounded at time zero and that $||\phi_k||_\infty$ is finite. Proceeding by induction we have:

$$\begin{aligned}
\tilde{\alpha}_k(\mathbf{1}) &= \Psi_k \Phi_k \tilde{\alpha}_{k-1}(\mathbf{1}) \\
\Phi_k \tilde{\alpha}_{k-1}(\mathbf{1}) &\leq \gamma_k(\mathbf{1}) + ||\phi_k||_\infty \tilde{\alpha}_{k-1}(\mathbf{1}) \\
\tilde{\alpha}_k(\mathbf{1}) &\leq |Z_k| + \gamma_k(\mathbf{1}) + ||\phi_k||_\infty \tilde{\alpha}_{k-1}(\mathbf{1})
\end{aligned} \tag{7}$$

whilst, in the case of the particle approximation, it can always be shown to hold from the convergence towards the true filter at the previous time. Note that, whenever we have a result of the form (10) or (11) together with (7) the total mass of the approximate filter must be finite with probability one and a finite upper bound upon the mass can be obtained immediately (consider the $\mathbb{L}_1$ convergence result obtained by setting $p = 1$ in (10) or (11)).

We make extensive use of [5, Lemma 7.3.3], the relevant portion of which is reproduced here.

Lemma 1 (Del Moral, 2004) *Given a sequence of probability measures $(\mu_i)_{i\geq 1}$ on a given measurable space $(E, \mathcal{E})$ and a collection of independent random variables, one distributed according to each of those measures, $(X_i)_{i\geq 1}$, where $\forall i, X_i \sim \mu_i$, together with any sequence of measurable functions $(h_i)_{i\geq 1}$ such that $\mu_i(h_i) = 0$ for all $i \geq 1$, we define for any $N \in \mathbb{N}$,*

$$m_N(X)(h) = \frac{1}{N}\sum_{i=1}^{N} h_i(X_i) \;\; \textit{and} \;\; \sigma_N^2(h) = \frac{1}{N}\sum_{i=1}^{N} (\sup(h_i) - \inf(h_i))^2$$

If the h_i have finite oscillations (i.e., $\sup(h_i) - \inf(h_i) < \infty \; \forall i \geq 1$) then we have:

$$\sqrt{N}\mathbb{E}\left[|m_N(X)(h)|^p\right]^{1/p} \leq d(p)^{1/p} \sigma_N(h)$$

with, for any pair of integers n, p such that $n \geq p \geq 1$, denoting $(n)_p = n!/(n-p)!$:

$$d(2n) = (2n)_n 2^{-n} \;\; \textit{and} \;\; d(2n-1) = \frac{(2n-1)_n}{\sqrt{n - \frac{1}{2}}} 2^{-(n-\frac{1}{2})}$$

We begin by showing that as the number of particles used to approximate the PHD filter tends towards infinity, the estimate of the integral of any bounded measurable function under the empirical measure associated with the particle approximation converges towards the integral under the true PHD filter in terms of $\mathbb{L}_p$ norm and that the two integrals are $\mathbb{P} -$ a.s. equal in the limit of infinitely many particles. The principal result of this section is theorem 1 which establishes the first result and leads directly to the second.

Throughout this section we assume that a particle approximation consisting of L_{k-1} weighted particles is available at time $k-1$, with associated empirical measure $\tilde{\alpha}_{k-1}^{L_{k-1}}$. These particles are propagated forwards according to the algorithm described previously, and an additional J_k particles are introduced to account for the possibility of new objects appearing at time k. This gives us an $M_k = J_k + L_{k-1}$ particle approximation, denoted $\alpha_k^{M_k}$ to the PHD filter at time k, which is subsequently re-weighted (corresponding to the update step of the exact algorithm) and resampled to provide a sample of L_k particles at this time, $\tilde{\alpha}_k^{L_k}$. This leads to a recursive algorithm and provides a convenient decomposition of the error introduced at each time-step into quantities which can be straightforwardly bounded. We assume that J_k and M_k are chosen in a manner independent of the evolution of the particle system, but which may be influenced by such factors as the number of observations.

3.1 Conditions

As a final precursor to the convergence study, we present a number of weak conditions which are sufficient for the convergence results below to hold. The following conditions are assumed to hold throughout:

- The particle filter is initialised with some finite mass by *iid* sampling from a tractable distribution $\tilde{\alpha}_0$.
- The observation set is finite, $|Z_k| < \infty \forall k$.
- All of the importance ratios are bounded above:
$$\sup_{(x,y)\in E\times E} \left| \frac{\phi_k(x,y)}{q_k(x,y)} \right| < R_1 < \infty \quad \sup_{x\in E} \left| \frac{\gamma_k(x)}{p_k(x)} \right| < R_2 < \infty \tag{8}$$
and that at least one of these ratios is also strictly positive.
- The individual object likelihood function is bounded above and strictly positive:
$$0 < \psi_{k,z}(x) < R_3 < \infty \tag{9}$$
- The number of particles used at each time step are not dependent upon the particle approximation at that time step. In the case of the convergence results we allow for fairly general behaviour, requiring only that the number of particles at each stage is proportional to the number used at the previous step in the algorithm, $L_k \propto M_k = L_{k-1} + J_k$ and $J_k \propto L_{k-1}$; in the central limit theorem we assume that N particles are propagated forward at each time step and some additional fraction η_k are introduced at each time k to describe the spontaneous birth density (this is done for convenience rather than through necessity).

– Resampling is done according to a multinomial scheme, that is the number of representatives of each particle which survives is sampled from a multinomial distribution with parameters proportional to the particle weights.

The first of these conditions simply constrain the initialisation of the particle approximation, the next is a weak finiteness requirement placed upon the true system, the next two are implementation issues and are required to ensure that the importance weights and that the filter density remains finite. The penultimate condition prevents unstable interactions between the filter mass and the particle approximation.

3.2 $\mathbb{L}_p$ Convergence and Almost Sure Convergence

The following theorem is the main result of this section and is proved by induction. It is shown that each step of the algorithm introduces an error (in the $\mathbb{L}_p$ sense) whose upper bound converges to zero as the number of particles tends to infinity and that the errors accumulated by the evolution of the algorithm have the same property.

Theorem 1 ($\mathbb{L}_p$ Convergence)

Under the conditions specified in section 3.1, there exist finite constants such that for any $\xi \in \mathcal{B}_b(E)$, $\xi : E \to \mathbb{R}$ the following holds for all times k:

$$\mathbb{E}\left[\left|\alpha_k^{M_k}(\xi) - \alpha_k(\xi)\right|^p\right]^{1/p} \leq \bar{c}_{k,p} \frac{||\xi||_\infty}{\sqrt{M_k}} \tag{10}$$

$$\mathbb{E}\left[\left|\tilde{\alpha}_k^{L_k}(\xi) - \tilde{\alpha}_k(\xi)\right|^p\right]^{1/p} \leq c_{k,p} \frac{||\xi||_\infty}{\sqrt{L_k}} \tag{11}$$

Convergence in an $\mathbb{L}_p$ sense directly implies convergence in probability, so we also have:

$$\alpha_k^{M_k}(\xi) \xrightarrow{p} \alpha_k(\xi)$$
$$\tilde{\alpha}_k^{L_k}(\xi) \xrightarrow{p} \tilde{\alpha}_k(\xi)$$

Furthermore, by a Borel-Cantelli argument, the particle approximation of the integral of any function with finite fourth moment converges almost surely to the integral under the true PHD filter as the number of particles tends towards infinity.

Proof.

Equation (11) holds at time 0 by lemma 2.

Now, if equation (11) holds at time $k-1$ then, by lemmas 3 and 4, equation (10) holds at time k.

Similarly, if equation (10) holds at time k then by lemmas 5 and 6, equation (11) also holds at time k.

The theorem follows by induction. □

Lemma 2 (Initialisation) *If, at time zero, the particle approximation, $\tilde{\alpha}_0^{L_0}$, is obtained by taking L_0 iid samples from $\tilde{\alpha}_0/\tilde{\alpha}_0(\mathbf{1})$ and weighting each by $\tilde{\alpha}_0(\mathbf{1})/L_0$, then there exists a finite constant $c_{0,p}$ such that, for all $p \geq 1$ and for all test functions ξ in $\mathcal{B}_b(E)$:*

$$\mathbb{E}\left[\left|\tilde{\alpha}_0^{L_0}(\xi) - \tilde{\alpha}_0(\xi)\right|^p\right]^{1/p} \leq c_{0,p}\frac{||\xi||_\infty}{\sqrt{L_0}}$$

Proof. This can be seen to be true directly by applying lemma 1. □

Lemma 3 (Prediction) *If, for some finite constant $c_{k-1,p}$, and all test functions ξ in $\mathcal{B}_b(E)$:*

$$\mathbb{E}\left[\left|\tilde{\alpha}_{k-1}^{L_{k-1}}(\xi) - \tilde{\alpha}_{k-1}(\xi)\right|^p\right]^{1/p} \leq c_{k-1,p}\frac{||\xi||_\infty}{\sqrt{L_{k-1}}}$$

Then there exists some finite constant $\hat{c}_{k,p}$ such that, for all test functions ξ in $\mathbb{L}_p(E)$:

$$\mathbb{E}\left[\left|\Phi_k\tilde{\alpha}_{k-1}^{L_{k-1}}(\xi) - \Phi_k\tilde{\alpha}_{k-1}(\xi)\right|^p\right]^{1/p} \leq \hat{c}_{k,p}\frac{||\xi||_\infty}{\sqrt{L_{k-1}}}$$

Proof. From the definition of the prediction operator:

$$\begin{aligned}
&\mathbb{E}\left[\left|\Phi_k\tilde{\alpha}_{k-1}^{L_{k-1}}(\xi) - \Phi_k\tilde{\alpha}_{k-1}(\xi)\right|^p\right]^{1/p} \\
&= \mathbb{E}\left[\left|\tilde{\alpha}_{k-1}^{L_{k-1}}\phi_k(\xi) - \tilde{\alpha}_{k-1}\phi_k(\xi)\right|^p\right]^{1/p} \\
&= \mathbb{E}\left[\left|\left(\tilde{\alpha}_{k-1}^{L_{k-1}} - \tilde{\alpha}_{k-1}\right)\phi_k(\xi)\right|^p\right]^{1/p}
\end{aligned}$$

Hence, by the assumption of the lemma:

$$\begin{aligned}
\mathbb{E}\left[\left|\Phi_k\tilde{\alpha}_{k-1}^{L_{k-1}}(\xi) - \Phi_k\tilde{\alpha}_{k-1}(\xi)\right|^p\right]^{1/p} &\leq c_{k-1,p}\frac{\sup_\zeta|\phi_k(\zeta,\xi)|}{\sqrt{L_{k-1}}} \\
&\leq c_{k-1,p}\frac{\sup_{\zeta,x}\phi_k(\zeta,x)\,||\xi||_\infty}{\sqrt{L_{k-1}}}
\end{aligned}$$

Which gives us the claim of the lemma with: $\hat{c}_{k,p} = c_{k-1,p}\sup_{\zeta,x}\phi_k(x,\zeta)$ □

Lemma 4 (Sampling) *If, for some finite constant, $\hat{c}_{k,p}$:*

$$\mathbb{E}\left[\left|\Phi_k\tilde{\alpha}_{k-1}^{L_{k-1}}(\xi) - \Phi_k\tilde{\alpha}_{k-1}(\xi)\right|^p\right]^{1/p} \leq \hat{c}_{k,p}\frac{||\xi||_\infty}{\sqrt{L_{k-1}}}$$

Then, there exists a finite constant $\tilde{c}_{k,p}$ such that:

$$\mathbb{E}\left[\left|\alpha_k^{M_k}(\xi) - \alpha_k(\xi)\right|^p\right]^{1/p} \leq \tilde{c}_{k,p}\frac{||\xi||_\infty}{\sqrt{M_k}}$$

Proof. Let $\mathcal{G}_{k-1}$ be the $\sigma-$field generated by the set of all particles until time $k-1$. By a conditioning argument on $\mathcal{G}_{k-1}$, we may view $\left(Y_k^{(i)}\right)_{i\geq 1}$ as independent samples with respective distributions $\left(q_k\left(Y_k^{(i)},\cdot\right)\right)_{i\geq 1}$. Let $\ddot{\alpha}_k^{L_{k-1}}$ be the empirical measure associated with the particles $\left(Y_k^{(i)}\right)_{i\geq 1}$ after the re-weighting step in equation (5), i.e.,

$$\ddot{\alpha}_k^{L_{k-1}} = \sum_{i=1}^{L_{k-1}} \tilde{w}_k^{(i)} \delta_{Y_k^{(i)}}$$

and define the sequence of functions $h_i(\cdot) = \frac{\phi_k(X_{k-1}^{(i)},\cdot)\xi(\cdot)}{q_k(X_{k-1}^{(i)},\cdot)} - \phi_k(\xi)\left(X_{k-1}^{(i)}\right)$ and associated measures $\mu_i(\cdot) = q_k(X_k^{(i)},\cdot)$ such that $\mu_i(h_i) = 0$. It is clear that:

$$\frac{\ddot{\alpha}_k^{L_{k-1}}(\xi) - \tilde{\alpha}_k^{L_{k-1}}\phi_k(\xi)}{\tilde{\alpha}_{k-1}^{L_{k-1}}(\mathbf{1})} = \sum_{i=1}^{L_{k-1}} \frac{w_{k-1}^{(i)} h_i(Y_k^{(i)})}{\tilde{\alpha}_{k-1}^{L_{k-1}}(\mathbf{1})}$$

Which allows us to write:

$$\begin{aligned}
&\mathbb{E}\left[\left|\ddot{\alpha}_k^{L_{k-1}}(\xi) - \tilde{\alpha}_k^{L_{k-1}}\phi_k(\xi)\right|^p\right] \\
&= \mathbb{E}\left[\left|\tilde{\alpha}_{k-1}^{L_{k-1}}(\mathbf{1})\right|^p \mathbb{E}\left[\left|\sum_{i=1}^{L_{k-1}} \frac{w_{k-1}^{(i)} h_i(Y_k^{(i)})}{\tilde{\alpha}_{k-1}^{L_{k-1}}(\mathbf{1})}\right|^p \middle| \mathcal{G}_{k-1}\right]\right] \\
&\leq \mathbb{E}\left[\left|\tilde{\alpha}_{k-1}^{L_{k-1}}(\mathbf{1})\right|^p\right] \frac{2^p d(p)\left(\left|\left|\frac{\phi_k}{q_k}\right|\right|_\infty ||\xi||_\infty\right)^p}{\left(\sqrt{L_{k-1}}\right)^p}
\end{aligned}$$

where the final inequality follows from an application of lemma 1. This gives us the bound:

$$\mathbb{E}\left[\left|\ddot{\alpha}_k^{L_{k-1}}(\xi) - \tilde{\alpha}_k^{L_{k-1}}\phi_k(\xi)\right|^p\right]^{1/p} \leq \frac{2d(p)^{1/p} C_{k,p}^{\tilde{\alpha}} R_1 ||\xi||_\infty}{\sqrt{L_{k-1}}}$$

Where $C_{k,p}^{\alpha}$ is the finite constant which bounds $\mathbb{E}\left[\left|\tilde{\alpha}_{k-1}^{L_{k-1}}(\mathbf{1})\right|^p\right]^{1/p}$ (see remark 1).

If we allow $\mathring{\alpha}_k^{J_k}$ be the particle approximation to γ_k obtained by importance sampling from p_k then it is straightforward to verify that, for some finite constant $\hat{B}_k^p$ obtained by using lemma 1 once again:

$$\mathbb{E}\left[\left|\mathring{\alpha}_k^{J_k}(\xi) - \gamma_k(\xi)\right|^p\right]^{1/p} \leq 2d(p)^{1/p}\left|\left|\frac{\gamma_k}{p_k}\right|\right|_\infty \frac{||\xi||_\infty}{\sqrt{J_k}}$$

And noting that $\alpha_k^{M_k} = \mathring{\alpha}_k^{J_k} + \ddot{\alpha}_k^{L_{k-1}}$ *we can apply Minkowski's inequality to obtain:*

$$\mathbb{E}\left[\left|\alpha_k^{M_k}(\xi) - \Phi_k \tilde{\alpha}_{k-1}^{L_k}\right|^p\right]^{1/p} \leq 2C_{k,P}^{\alpha} d(p)^{1/p} \frac{\left|\left|\frac{\phi_k}{q_k}\right|\right|_\infty ||\xi||_\infty}{\sqrt{L_{k-1}}} + 2d(p)^{1/p}\left|\left|\frac{\gamma_k}{p_k}\right|\right|_\infty \frac{||\xi||_\infty}{\sqrt{J_k}}$$

Defining $l_{k-1} = L_{k-1}/M_k$ *and* $j_k = J_k/M_k$ *for convenience, we arrive at the result of the lemma with (making use of (8)):*

$$\tilde{c}_{k,p} = 2d(p)^{1/p} C_{k,p}^{\alpha} \frac{R_1}{\sqrt{l_{k-1}}} + 2d(p)^{1/p} \frac{R_2}{\sqrt{j_k}}$$

□

Lemma 5 (Update) *If for some finite constant* $\tilde{c}_{k,p}$*:*

$$\mathbb{E}\left[\left|\alpha_k^{M_k}(\xi) - \alpha_k(\xi)\right|^p\right]^{1/p} \leq \tilde{c}_{k,p} \frac{||\xi||_\infty}{\sqrt{M_k}}$$

Then there exists a finite constant $\bar{c}_{k,p}$ *such that:*

$$\mathbb{E}\left[\left|\Psi_k \alpha_k^{M_k}(\xi) - \tilde{\alpha}_k(\xi)\right|^p\right]^{1/p} \leq \bar{c}_{k,p} \frac{||\xi||_\infty}{\sqrt{M_k}}$$

Proof. The proof follows by expanding the norm and using Minkowski's inequality to bound the overall norm. The individual constituents are bounded by the assumption of the lemma.

$$\begin{aligned}
&\mathbb{E}\left[\left|\Psi_k \alpha_k^{M_k}(\xi) - \tilde{\alpha}_k(\xi)\right|^p\right]^{1/p} \\
&= \mathbb{E}\left[\left|\alpha_k^{M_k}\left(\nu_k \xi + \sum_{z \in Z_k} \frac{\psi_{k,z}\xi}{\kappa_k(z) + \alpha_k^{M_k}(\psi_{k,z})}\right)\right.\right. \\
&\qquad \left.\left. - \alpha_k\left(\nu_k \xi + \sum_{z \in Z_k} \frac{\psi_{k,z}\xi}{\kappa_k(z) + \alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p} \\
&\leq \mathbb{E}\left[\left|\alpha_k^{M_k}(\nu_k \xi) - \alpha_k(\nu_k \xi)\right|^p\right]^{1/p} + \\
&\sum_{z \in Z_k} \mathbb{E}\left[\left|\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z) + \alpha_k^{M_k}(\psi_{k,z})}\right) - \alpha_k\left(\frac{\psi_{k,z}\xi}{\kappa_k(z) + \alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p}
\end{aligned}$$

Noting that ν_k *is a probability, the first term is trivially bounded by the assumption of the lemma:*

$$\mathbb{E}\left[\left|\alpha_k^{M_k}(\nu_k \xi) - \alpha_k(\nu_k \xi)\right|^p\right]^{1/p} \leq \tilde{c}_{k,p} \frac{||\nu_k||_\infty ||\xi||_\infty}{\sqrt{M_k}} \leq \tilde{c}_{k,p} \frac{||\xi||_\infty}{\sqrt{M_k}}$$

In order to bound the second term a little more effort is required, consider a single element of the summation:

$$\mathbb{E}\left[\left|\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k^{M_k}(\psi_{k,z})}\right)-\alpha_k\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p}$$

$$\leq \mathbb{E}\left[\left|\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k^{M_k}(\psi_{k,z})}\right)-\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p}+$$

$$\mathbb{E}\left[\left|\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)-\alpha_k\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p}$$

$$\leq \mathbb{E}\left[\left|\frac{\alpha_k^{M_k}\left(\psi_{k,z}\xi\right)\left[\left(\kappa_k(z)+\alpha_k^{M_k}(\psi_{k,z})\right)-\left(\kappa_k(z)+\alpha_k(\psi_{k,z})\right)\right]}{\left(\left(\kappa_k(z)+\alpha_k^{M_k}(\psi_{k,z})\right)\right)\left(\left(\kappa_k(z)+\alpha_k(\psi_{k,z})\right)\right)}\right|^p\right]^{1/p}+$$

$$\mathbb{E}\left[\left|\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)-\alpha_k\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p}$$

$$\leq 2\frac{\mathbb{E}\left[\left|\alpha_k^{M_k}(\psi_{k,z})-\alpha_k(\psi_{k,z})\right|^p\right]^{1/p}||\xi||_\infty}{\kappa_k(z)+\alpha_k(\psi_{k,z})}$$

Where the final line follows from the positivity assumptions placed upon one of the weight ratios and the likelihood function. This allows us to assert that:

$$\sum_{z\in Z_k}\mathbb{E}\left[\left|\alpha_k^{M_k}\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k^{M_k}(\psi_{k,z})}\right)-\alpha_k\left(\frac{\psi_{k,z}\xi}{\kappa_k(z)+\alpha_k(\psi_{k,z})}\right)\right|^p\right]^{1/p}$$

$$\leq 2Z_k\,||\xi||_\infty\sup_z\frac{\mathbb{E}\left[\left|\alpha_k^{M_k}(\psi_{k,z})-\alpha_k(\psi_{k,z})\right|^p\right]^{1/p}}{\kappa_k(z)+\alpha_k(\psi_{k,z})}$$

$$\leq \frac{2Z_k\tilde{c}_{k,p}\,||\xi||_\infty}{\sqrt{M_k}}\sup_z\frac{||\psi_{k,z}||_\infty}{\kappa_k(z)+\alpha_k(\psi_{k,z})}$$

Combining this with the previous result and assumption (9) gives the result of the lemma with:

$$\bar{c}_{k,p}=1+2Z_k\tilde{c}_{k,p}\sup_z\frac{R_3}{\kappa_k(z)+\alpha_k(\psi_{k,z})}$$

□

Lemma 6 (Resampling) *If, for some finite constant,* $\bar{c}_{k,p}$*:*

$$\mathbb{E}\left[\left|\Psi_k\alpha_k^{M_k}(\xi)-\tilde{\alpha}_k(\xi)\right|^p\right]^{1/p}\leq\bar{c}_{k,p}\frac{||\xi||_\infty}{\sqrt{M_k}}$$

and the resampling scheme is multinomial, then there exists a finite constant $c_{k,p}$ *such that:*

$$\mathbb{E}\left[\left|\tilde{\alpha}_k^{L_k}(\xi)-\tilde{\alpha}_k(\xi)\right|^p\right]^{1/p}\leq c_{k,p}\frac{||\xi||_\infty}{\sqrt{L_k}}$$

Proof.

By Minkowski's inequality,

$$\mathbb{E}\left[\left|\tilde{\alpha}_k^{L_k}(\xi) - \tilde{\alpha}_k(\xi)\right|^p\right]^{1/p}$$
$$\leq \mathbb{E}\left[\left|\tilde{\alpha}_k^{L_k}(\xi) - \Psi_k\tilde{\alpha}_k^{M_k}(\xi)\right|^p\right]^{1/p} + \mathbb{E}\left[\left|\Psi_k\tilde{\alpha}_k^{M_k}(\xi) - \tilde{\alpha}_k(\xi)\right|^p\right]^{1/p}$$

By the assumption of the lemma:

$$\mathbb{E}\left[\left|\Psi_k\tilde{\alpha}_k^{M_k}(\xi) - \tilde{\alpha}_k(\xi)\right|^p\right]^{1/p} \leq \bar{c}_{k,p}\frac{||\xi||_\infty}{\sqrt{M_k}}$$

We can bound the remaining term by taking the expectation conditioned upon the sigma algebra generated by the particle ensemble prior to resampling, noting that the resampled particle set is iid *according to the empirical distribution before resampling:*

$$\mathbb{E}\left[\left|\tilde{\alpha}_k^{L_k}(\xi) - \Psi_k\tilde{\alpha}_k^{M_k}(\xi)\right|^p\right]^{1/p} \leq C_{k,p}^D C_{k,p}^R \frac{||\xi||_\infty}{\sqrt{L_k}}$$

where $C_{k,p}^D$ is the upper bound on $\mathbb{E}\left[\left|\tilde{\alpha}_k^{M_k}\right|^p\right]^{1/p}$ approximation at the resampling stage (again, this must exist by remark 1) and $C_{k,p}^R$ is a constant given by Del Moral's $\mathbb{L}_p$-bound lemma, lemma 1.

Thus we have the result of the lemma with:

$$c_{k,p} = C_{k,p}^D C_{k,p}^R + \bar{c}_{k,p}\sqrt{\frac{M_k}{L_k}}$$

□

It would be convenient to establish time-uniform convergence results and the stability of the filter with respect to its initial conditions. However, the tools pioneered by [5] and subsequently [10,3] are not appropriate in the present case: the PHD filter is not a Feynman-Kac flow and decoupling the "prediction" and "update" steps of the filter is not straightforward due to the inherent nonlinearity and the absence of a linear unnormalised flow. It is not obvious how to obtain such results under realistic assumptions.

4 Central Limit Theorem

A number of people have published central limit theorems for SMC Methods [6, 5,3,10]. As the PHD filtering equations are somewhat different to the standard Bayes recursion, a number of significant differences need to be addressed in this case. Firstly, the total mass of the filter is variable and unknown rather than fixed at unity and secondly, two importance sampling steps are required at each time. The other main result of the paper is theorem 2 which shows that a central limit theorem holds for the SMC approximation of the PHD filter. We adopt an inductive approach to demonstrating that a central limit theorem applies to estimates of the integral of an arbitrary test function under the random measure associated with the particle approximation to the PHD filter.

4.1 Formulation

It is convenient to write the PHD slightly differently to equations (3) and (4) for the purposes of considering the central limit theorem. It is useful to describe the evolution of the PHD filter in terms of selection and mutation operations to allow the errors introduced at each time to be divided into the error propagated forward from earlier times and that introduced by sampling at the present time-step. The formulation used is similar to that employed in the analysis of Feynman-Kac flows [5] under an interacting-process interpretation.

We introduce a potential function, $G_{k,\alpha_k} : E \rightarrow \mathbb{R}$ and its associated selection operator $S_{k,\alpha_k} : E \times E \rightarrow \mathbb{R}$ and as the selection operator which we employ updates the measure based upon the full distribution at the previous time, we may define the measure $\hat{S}_{k,\alpha_k}(x) = \alpha_k(S_{k,\alpha_k}(\cdot, x))\alpha_k(G_{k,\alpha_k})/\alpha_k(\mathbf{1})$ which is obtained by applying the selection operator to the measure and renormalising to correctly reflect the evolution of the mass of the filter:

$$
\begin{aligned}
G_{k,\alpha_k}(\cdot) &= \nu_k(\cdot) + \sum_{z \in Z_k} \frac{\psi_{k,z}(\cdot)}{\kappa_k(z) + \alpha_k(\psi_{k,z})} \\
S_{k,\alpha_k}(x, y) &= \frac{\alpha_k(y) G_{k,\alpha_k}(y)}{\alpha_k(G_{k,\alpha_k})} \\
\hat{S}_{k,\alpha_k}(\cdot) &= \alpha_k(\cdot) G_{k,\alpha_k}(\cdot)
\end{aligned}
$$

For clarity of exposition, we have assumed in this section that N particles are propagated forward from each time step to the next and that $\eta_k N$ particles are introduced to account for spontaneous births at time k (i.e., in the notation of the previous section, $L_k = N$ and $J_k = \eta_k N$). The notation $N_k = (1 + \eta_k)N$ is also used for notational convenience.

The interpretation of this formulation is slightly different and perhaps more intuitive. Update and resampling occur simultaneously and comprise the selection step, while prediction follows a mutation operation. Here we use α_k to refer to the predicted filter as in (3), and it is not necessary to make any reference to the updated filter. We separate the spontaneous birth component of the measure from that which depends upon the past and write the PHD recursion as:

$$
\begin{aligned}
\alpha_k(\xi) &= \hat{\alpha}_k(\xi) + \mathring{\alpha}_k(\xi) \\
\hat{\alpha}_k(\xi) &= \hat{S}_{k-1,\alpha_{k-1}} \phi_k(\xi) \\
\mathring{\alpha}_k(\xi) &= \gamma_k(\xi)
\end{aligned}
$$

4.1.1 The Particle Approximation Within this section, the particle approximation described previously can be restated as the following iterative procedure. This provides an alternative view of the algorithm given in section 2.3.1, with the additional assumption that the number of particles propagated forward at each time step is constant, and no explicit reference to $\tilde{\alpha_k}^{L_k}$. As we are concerned with asymptotic results the increased clarity more than compensates for the slight reduction in generality.

1. Let the particle approximation prior to resampling at time $k-1$ be of the form

$$\alpha_{k-1}^{N_{k-1}} = \frac{1}{N} \sum_{i=1}^{N_{k-1}} \tilde{w}_{k-1}^{(i)} \delta_{X_{k-1}^{(i)}}$$

2. Sample N particles to propagate forward via the selection operator:

$$\left\{ Y_k^{(i)} \leftarrow S_{k-1,\alpha_{k-1}^{N_{k-1}}}(\cdot) \right\}_{i=1}^{N}$$

3. Mutate these N particles.

$$\left\{ X_k^{(i)} \leftarrow q_k(Y_k^{(i)}, \cdot) \right\}_{i=1}^{N}$$

4. Introduce $\eta_k N$ particles to account for the possibility of births.

$$\left\{ X_k^{(i)} \leftarrow p_k(\cdot) \right\}_{i=N+1}^{N_k}$$

5. Define the particle approximation at time k as $\alpha_k^{N_k} = \hat{\alpha}_k^N + \mathring{\alpha}_k^{\eta_k N}$ where:

$$\hat{\alpha}_k^N = \frac{1}{N} \sum_{i=1}^{N} \tilde{w}_k^{(i)} \delta_{X_k^{(i)}} \text{ and } \mathring{\alpha}_k^{\eta_k N} = \frac{1}{N} \sum_{i=N+1}^{N_k} \tilde{w}_k^{(i)} \delta_{X_k^{(i)}}$$

   and the weights are given by:

$$\tilde{w}_k^{(i)} = \begin{cases} \alpha_{k-1}^{N_{k-1}} \left(G_{k-1,\alpha_{k-1}^{N_{k-1}}} \right) \frac{\phi_k(Y_k^{(i)}, X_k^{(i)})}{q_k(Y_k^{(i)}, X_k^{(i)})} & i \in \{1, \ldots, N\} \\ \frac{1}{\eta_k} \frac{\gamma_k(X_k^{(i)})}{p_k(X_k^{(i)})} & i \in \{N+1, \ldots, N_k\} \end{cases}$$

4.2 Variance Recursion

Theorem 2 (Central Limit Theorem) *The particle approximation to the PHD filter follows a central limit theorem with some finite variance for all continuous bounded test functions* $\xi : E \to \mathbb{R}^d$*, at all times* $k \geq 0$*:*

$$\lim_{N\to\infty} \sqrt{N} \left[\alpha_k^{N_k}(\xi) - \alpha_k(\xi) \right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_k^2(\xi)\right)$$

provided that the result holds at time 0*, which it does, for example, if the filter is initialised by obtaining samples from a normalised version of the true filter by importance sampling and weighting them correctly.*

In all cases we prove the case for scalar-valued test functions and the generalisation to the vector-valued case follows directly via the Cramer-Wold device [2, p.397].

Proof. By assumption, the result of the theorem holds at time 0*. Using induction the result can be shown to hold for all times by the sequence of lemmas, lemma 7-10, that follow.*

The core of the proof is the following decomposition:

$$\begin{aligned}\alpha_k^{N_k}(\xi) - \alpha_k(\xi) = \hat{\alpha}_k^N(\xi) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi) + \\ \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi) - \hat{\alpha}_k(\xi) + \\ \mathring{\alpha}_k^{(\eta_k N)}(\xi) - \mathring{\alpha}_k(\xi)\end{aligned}$$

Consistent with the notation defined in section 2.1, $\hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)$ *is to be understood as*

$$\int \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}(du) \int q_k(u,dv)\frac{\phi_k(u,v)}{q_k(u,v)}\xi(v)$$

i.e., $q_k \times \frac{\phi_k}{q_k}$ *defines a new transition kernel from* E *to* E*.*

The first term in this decomposition accounts for errors introduced at time k *by using a particle approximation of the prediction step and this is shown to converge to a centred normal distribution of variance* $\hat{V}_k(\xi)$ *in lemma 7. The second term describes the errors propagated forward from previous times, and is shown to follow a central limit theorem with variance* $\ddot{V}_k(\xi)$ *in lemma 8. The final term corresponds to sampling errors in the spontaneous birth components of the filter and this is shown to follow a central limit theorem with variance* $\mathring{V}_k(\xi)$ *in lemma 9.*

Lemma 10 shows that the result of combining the three terms of the decomposition is a random variable which itself follows a central limit theorem with variance:

$$\sigma_k^2(\xi) = \hat{V}_k(\xi) + \ddot{V}_k(\xi) + \mathring{V}_k(\xi)$$

which is precisely the result of the theorem for scalar test functions.

In the case of vector test functions, the result follows by the Cramer-Wold device, applied to any linear combination of their components, and the covariance matrix is denoted $\Sigma_k(\xi) = [\Sigma_k(\xi_i, \xi_j)]$*.* $\square$

Lemma 7 (Selection-prediction Sampling Errors) *The selection-prediction sampling error (due to steps 2 and 3) at time* k *converges to a normally distributed random variable of finite variance as the size of the particle ensemble tends towards infinity:*

$$\lim_{N\to\infty} \sqrt{N}\left(\hat{\alpha}_k^N(\xi) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)\right) \xrightarrow{d} \mathcal{N}\left(0, \hat{V}_k(\xi)\right)$$

Proof. Consider the term under consideration:

$$\begin{aligned}&\hat{\alpha}_k^N(\xi) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)\\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{w}_k^{(i)}\xi(X_k^{(i)}) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)\right)\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_{k,i}^N\end{aligned}$$

x where

$$U_{k,i}^N = \frac{\tilde{w}_k^{(i)}\xi(X_k^{(i)}) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)}{\sqrt{N}}$$

Let $\mathcal{H}_k^N = \sigma\left(\left\{X_n^{(i)}, \tilde{w}_n^{(i)}\right\}_{i=1}^{N_n} : n = 0, \ldots, k\right)$ *be the sigma algebra generated by the particle ensembles occurring at or before time* k *and further let* $\mathcal{H}_{k,j}^N = \sigma\left(\mathcal{H}_{k-1}^N, \left\{X_k^{(i)}, \tilde{w}_k^{(i)}\right\}_{i=1}^{j}\right)$.

It is evident that conditioned upon $\mathcal{H}_{k-1}^N$, $\left\{Y_k^{(i)}, X_k^{(i)}\right\}_{i=1}^{N}$ *are* iid *samples from the product distribution* $S_{k-1,\alpha_{k-1}^{N_{k-1}}}(y)q_k(y,x)$ *and, therefore:*

$$\mathbb{E}\left[U_{k,i}^N \middle| \mathcal{H}_{k,i-1}^N\right] = \mathbb{E}\left[U_{k,i}^N \middle| \mathcal{H}_{k-1}^N\right] = 0$$

Furthermore, conditionally, $U_{k,i}^N$ *has finite variance, which follows from assumption (8) and the assumption that the observation set and the initial mass of the filter are finite:*

$$\begin{aligned}\mathbb{E}\left[\left(U_{k,i}^N\right)^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(U_{k,i}^N\right)^2 \middle| \mathcal{H}_{k,i-1}^N\right]\right]\\ &= \frac{1}{N}\mathbb{E}\left[\left(\tilde{w}_k^{(i)}\xi(X_k^{(i)})\right)^2 - \left(\hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)\right)^2\right]\\ &< \frac{2\mathbb{E}\left[\left(\alpha_{k-1}^{N_{k-1}}(\mathbf{1}) + |Z_{k-1}|\right)^2\right] R_1^2 \,||\xi||_\infty^2}{N} < \infty\end{aligned}$$

Noting that, by the $\mathbb{L}_p$ *convergence result of theorem 1 the expectation may be bounded above uniformly in N. We have that* $\forall t \in [0,1], \epsilon > 0$*:*

$$\lim_{N\to\infty}\sum_{i=1}^{\lfloor Nt\rfloor}\mathbb{E}\left[\left(U_{k,i}^N\right)^2 \mathbb{I}_{|U_{k,i}^N|>\epsilon} \middle| \mathcal{H}_{k,i-1}^N\right] \xrightarrow{p} 0 \tag{12}$$

By noting that the following convergence result holds (and this can be seen by expanding each term and using theorem 1, noting that if two sequences of bounded random variables converge to two finite limits, then the product of those sequences *converges to the product of their respective limits and that for nonzero random variables the same is true of the quotient of those sequences)*

$$\alpha_{k-1}^{N_{k-1}}\left(G_{k,\alpha_{k-1}^{N_{k-1}}}\right)\alpha_{k-1}^{N_{k-1}}\left(G_{k,\alpha_{k-1}^{N_{k-1}}}\left(\phi_k \times \frac{\phi_k}{q_k}\right)(\xi)\right) - \alpha_{k-1}^{N_{k-1}}\left(G_{k,\alpha_{k-1}^{N_{k-1}}}\phi_k(\xi)\right)^2 \tag{13}$$

$$\xrightarrow{p} \alpha_{k-1}\left(G_{k,\alpha_{k-1}}\right)\alpha_{k-1}\left(G_{k,\alpha_{k-1}}\left(\phi_k \times \frac{\phi_k}{q_k}\right)(\xi)\right) - \alpha_{k-1}\left(G_{k,\alpha_{k-1}}\phi_k(\xi)\right)^2 \tag{14}$$

it is apparent (as (13) is equal to $\frac{N}{\lfloor Nt \rfloor}$ times (15) and (14) to $\frac{1}{t}$ times (16)) that

$$\sum_{k=1}^{\lfloor Nt \rfloor}\mathbb{E}\left[\left(U_{k,i}^N\right)^2\middle|\mathcal{H}_{k,i-1}^N\right] = \frac{\lfloor Nt \rfloor}{N}\alpha_{k-1}^{N_{k-1}}\left(G_{k-1,\alpha_{k-1}^{N_{k-1}}}\right)^2 \times \left[S_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(\phi_k \times \frac{\phi_k}{q_k}\right)(\xi^2) - S_{k-1,\alpha_{k-1}^{N_{k-1}}}(\phi_k(\xi))^2\right] \tag{15}$$

$$\xrightarrow{p} t\alpha_{k-1}\left(G_{k-1,\alpha_{k-1}}\right)^2 \times \left[S_{k-1,\alpha_{k-1}}\left(\phi_k \times \frac{\phi_k}{q_k}\right)(\xi^2) - S_{k-1,\alpha_{k-1}}(\phi_k(\xi))^2\right] \tag{16}$$

From this, it can be seen that for each N, the sequence

$$\left(U_{k,i}^N, \mathcal{H}_{k,i}^N\right), \quad 1 \leq i \leq N$$

is a square-integrable martingale difference which satisfies the Lindeberg condition (12) and hence a martingale central limit theorem may be invoked (see, for example, [13, page 543])) to show that:

$$\lim_{N\to\infty}\sqrt{N}\left(\hat{\alpha}_k^N(\xi) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)\right) \xrightarrow{d} \mathcal{N}\left(0, \hat{V}_k(\xi)\right)$$

where,

$$\hat{V}_k(\xi) = \alpha_{k-1}\left(G_{k-1,\alpha_{k-1}}\right)^2\left[S_{k-1,\alpha_{k-1}}\left(\phi_k \times \frac{\phi_k}{q_k}\right)(\xi^2) - S_{k-1,\alpha_{k-1}}(\phi_k(\xi))^2\right]$$
$$= \alpha_{k-1}(G_{k-1,\alpha_{k-1}})\hat{S}_{k-1,\alpha_{k-1}}\left(\phi_k \times \frac{\phi_k}{q_k}\right)(\xi^2) - \hat{S}_{k-1,\alpha_{k-1}}(\phi_k(\xi))^2$$

□

Lemma 8 (Propagated Errors) *The error resulting from propagating the particle approximation forward rather than the true filter has an asymptotically normal distribution with finite variance.*

$$\hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi) - \hat{\alpha}_k(\xi) \xrightarrow{d} \mathcal{N}\left(0, \ddot{V}_k(\xi)\right)$$

Proof. Direct expansion of the potential allows us to express this difference as:

$$\begin{aligned}
&\hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi) - \hat{\alpha}_k(\xi) \\
&= \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\phi_k(\xi) - \hat{S}_{k-1,\alpha_{k-1}}\phi_k(\xi) \\
&= \alpha_{k-1}^{N_{k-1}}\left(\phi_k(\xi) G_{k-1,\alpha_{k-1}^{N_{k-1}}}\right) - \alpha_{k-1}\left(\phi_k(\xi) G_{k-1,\alpha_{k-1}}\right) \\
&= \alpha_{k-1}^{N_{k-1}}\left(\phi_k(\xi)\nu_{k-1}\right) - \alpha_{k-1}\left(\phi_k(\xi)\nu_{k-1}\right) + \\
&\quad \sum_{z \in Z_{k-1}} \frac{\alpha_{k-1}^{N_{k-1}}(\Delta_{k-1,z}) - \alpha_{k-1}(\Delta_{k-1,z})}{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})}
\end{aligned}$$

where $\Delta_{k-1,z} = \psi_{k-1,z}\phi_k(\xi) - \frac{\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))}{\kappa_{k-1}(z)+\alpha_{k-1}(\psi_{k-1,z})}\psi_{k-1,z}$ *and the final equality can be shown to hold by considering a single term in the summation thus:*

$$\begin{aligned}
&\frac{\alpha_{k-1}^{N_k}(\psi_{k-1,z}\phi_k(\xi))}{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})} - \frac{\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))}{\kappa_{k-1}(z) + \alpha_{k-1}(\psi_{k-1,z})} \\
&= \frac{1}{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})}\Big[\alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z}\phi_k(\xi)) - \alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi)) \\
&\quad + \alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi)) - \frac{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})}{\kappa_{k-1}(z) + \alpha_{k-1}(\psi_{k-1,z})}\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))\Big] \\
&= \frac{1}{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})}\Big[\alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z}\phi_k(\xi)) - \alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi)) \\
&\quad + \frac{\alpha_{k-1}(\psi_{k-1,z}) - \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})}{\kappa_{k-1}(z) + \alpha_{k-1}(\psi_{k-1,z})}\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))\Big] \\
&= \frac{\alpha_{k-1}^{N_{k-1}}\left(\psi_{k-1,z}\phi_k(\xi) - \frac{\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))\psi_{k-1,z}}{\kappa_{k-1}(z)+\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))\psi_{k-1,z}}\right)}{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})} \\
&\quad - \frac{\alpha_{k-1}\left(\psi_{k-1,z}\phi_k(\xi) - \frac{\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))\psi_{k-1,z}}{\kappa_{k-1}(z)+\alpha_{k-1}(\psi_{k-1,z}\phi_k(\xi))\psi_{k-1,z}}\right)}{\kappa_{k-1}(z) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,z})}
\end{aligned}$$

If we set

$$\Delta_{k-1} = \left[\nu_k\phi_k(\xi), \Delta_{k-1,Z_{k-1,1}}, \ldots, \Delta_{k-1,Z_{k-1,|Z_{k-1}|}}\right]$$

where Z_{k-1}^i *denotes the* i^{th} *element of the set* Z_{k-1}*, and,*

$$\begin{aligned}
\rho_{k-1}^{N_{k-1}} = \Bigg[1, &\frac{1}{\kappa_{k-1}(Z_{k-1,1}) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,Z_{k-1,1}})}, \ldots, \\
&\frac{1}{\kappa_{k-1}(Z_{k-1,|Z_{k-1}|}) + \alpha_{k-1}^{N_{k-1}}(\psi_{k-1,Z_{k-1,|Z_{k-1}|}})}\Bigg]^T
\end{aligned}$$

Then the quantity of interest may be written as an inner product:

$$\left\langle \rho_{k-1}^{N_{k-1}}, \alpha_{k-1}^{N_{k-1}}\left(\Delta_{k-1}\right) - \alpha_{k-1}\left(\Delta_{k-1}\right) \right\rangle$$

We know from theorem 1 that $\rho_{k-1}^{N_{k-1}} \xrightarrow{p} \rho_{k-1}$, *where*

$$\rho_{k-1} = \left[1, \frac{1}{\kappa_{k-1}(Z_{k-1,1}) + \alpha_{k-1}(\psi_{k-1,Z_{k-1,1}})}, \ldots, \frac{1}{\kappa_{k-1}(Z_{k-1,|Z_{k-1}|}) + \alpha_{k-1}(\psi_{k-1,Z_{k-1,|Z_{k-1}|}})}\right]^T$$

And furthermore, we know by the induction assumption that each $\alpha_{k-1}^{N_{k-1}}\left(\Delta_{k-1}\right) - \alpha_{k-1}\left(\Delta_{k-1}\right)$ *is asymptotically normal with zero mean and some known variance,* $\Sigma_{k-1}(\Delta_{k-1})$. *By Slutzky's theorem, therefore, the quantity of interest converges to a normal distribution of mean zero and variance* $\ddot{V}_k(\xi) = \rho_{k-1}^T \Sigma_{k-1}(\Delta_{k-1})\rho_{k-1}$.
□

Lemma 9 (Spontaneous Births) *The error in the particle approximation to the spontaneous birth element of the PHD converges to a normal distribution with finite variance:*

$$\lim_{N\to\infty} \sqrt{N}\left[\mathring{\alpha}_k^{\eta_k N}(\xi) - \mathring{\alpha}_k(\xi)\right] \xrightarrow{d} \mathcal{N}\left(0, \mathring{V}_k(\xi)\right)$$

Proof.

$$\mathring{\alpha}_k^{\eta_k N}(\xi) - \mathring{\alpha}_k(\xi) = \frac{\gamma_k(\mathbf{1})}{\eta_k N} \sum_{j=N+1}^{N_k} \left(\frac{\gamma_k(X_k^{(j)})}{\gamma_k(\mathbf{1})p_k(X_k^{(j)})}\xi(X_k^{(j)}) - \frac{\gamma_k(\xi)}{\gamma_k(\mathbf{1})} \right)$$

Of course, the particles appearing within this sum are iid *according to* p_k *and this corresponds to* $\gamma_k(\mathbf{1})$ *multiplied by the importance sampling estimate giving us the standard result:*

$$\sqrt{\eta_k N}\left[\frac{\mathring{\alpha}_k^{\eta_k N}(\xi) - \mathring{\alpha}_k(\xi)}{\gamma_k(\mathbf{1})}\right] \xrightarrow{d} \mathcal{N}\left(0, \mathit{Var}_{p_k}\left(\frac{\gamma_k}{\gamma_k(\mathbf{1})p_k}\xi\right)\right)$$

which is precisely the result of the lemma with:

$$\mathring{V}_k(\xi) = \frac{1}{\eta_k}\left[\gamma_k\left(\frac{\gamma_k}{p_k}\xi^2\right) - \gamma_k(\xi)^2\right]$$

□

Lemma 10 (Combining Terms) *Using the results of lemmas 7–9 it follows that* $\alpha_k^{N_k}(\xi) - \alpha_k(\xi)$ *satisfies the central limit theorem:*

$$\lim_{N\to\infty} \sqrt{N}\left(\alpha_k^{N_k}(\xi) - \alpha_k(\xi)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_k(\xi)\right)$$

where the asymptotic variance is given by $\sigma_k(\xi) = \hat{V}_k(\xi) + \ddot{V}_k(\xi) + \mathring{V}_k(\xi)$.

Proof. The proof follows the method of [10]. The characteristic function of the random variable of interest is

$$\Upsilon_k(t) = \mathbb{E}\left[\exp\left(it\sqrt{N}\left(\alpha_k^{N_k}(\xi) - \alpha_k(\xi)\right)\right)\right]$$

As the particles associated with the spontaneous birth term of the PHD are independent of those propagated forward from the previous time we can write:

$$\Upsilon_k(t) = \mathbb{E}\left[\exp\left(it\sqrt{N}\left(\mathring{\alpha}_k^{\eta_k N}(\xi) - \mathring{\alpha}_k(\xi)\right)\right)\right] \\ \times\mathbb{E}\left[\exp\left(it\sqrt{N}\left(\hat{\alpha}_k^{N}(\xi) - \hat{\alpha}_k(\xi)\right)\right)\right]$$

The first term of this expansion is the characteristic function of a normal random variable, so all that remains is to show that the same is true of the second term. Using the same decomposition as above, we may write:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(it\sqrt{N}\left(\hat{\alpha}_k^{N}(\xi) - \hat{\alpha}_k(\xi)\right)\right)\right] \\
&= \mathbb{E}\Bigg[\overbrace{\mathbb{E}\left[\exp\left(it\sqrt{N}\left\{\hat{\alpha}_k^N(\xi) - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi)\right\}\right)\middle|\mathcal{H}_{k-1}^N\right]}^{\mathsf{A}} \\
&\qquad \underbrace{\times\exp\left(it\sqrt{N}\left\{\hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}}\left(q_k \times \frac{\phi_k}{q_k}\right)(\xi) - \hat{\alpha}_k(\xi)\right\}\right)}_{\mathsf{B}}\Bigg] \\
&= \mathbb{E}\left[\left(\mathsf{A} - \exp\left(-\frac{t^2\hat{V}_k(\xi)}{2}\right)\right)\mathsf{B}\right] + \exp\left(-\frac{t^2\hat{V}_k(\xi)}{2}\right)\mathbb{E}[\mathsf{B}]
\end{aligned}$$

All that remains is to show that the first term in this expansion vanishes and we will have shown that the characteristic function of interest $\Upsilon_k(t)$ corresponds to a Gaussian distribution as it can be expressed as the product of three Gaussian characteristic functions. Furthermore, it must have variance equal to the sum of the variances of the three constituent Gaussians which is exactly the result which we wish to prove.

By the conditionally iid *nature of the particles, we can write:*

$$\mathsf{A} = \mathbb{E}\left[\exp\left(it\sum_{j=1}^{N} U_{k,j}^N\right)\middle|\mathcal{H}_{k-1}^N\right] = \mathbb{E}\left[\exp\left(itU_{k,1}^N\right)\middle|\mathcal{H}_{k-1}^N\right]^N$$

Hence:

$$\left|\mathsf{A} - \exp\left(-\frac{t^2\hat{V}_k(\xi)}{2}\right)\right| = \left|\mathbb{E}\left[\exp\left(itU_{k,1}^N\right)\middle|\mathcal{H}_{k-1}^N\right]^N - \exp\left(-\frac{t^2\hat{V}_k(\xi)/N}{2}\right)^N\right|$$

Using the same result as [10] (i.e. that $|u^N - v^N| \leq N|u-v| \forall |u| \leq 1, |v| \leq 1$) we obtain:

$$\left| \mathsf{A} - \exp\left(-\frac{t^2 \hat{V}_k(\xi)}{2}\right)\right| \leq N \left| \mathbb{E}\left[\exp\left(itU_{k,1}^N\right)\middle| \mathcal{H}_{k-1}^N\right] - \exp\left(-\frac{t^2 \hat{V}_k(\xi)/N}{2}\right)\right|$$

The following decomposition can be used to show that this difference converges to zero as $N \to \infty$:

$$\mathbb{E}\left[\exp\left(itU_{k,1}^N\right) - \exp\left(-\frac{t^2 \hat{V}_k(\xi)/N}{2}\right)\middle| \mathcal{H}_{k-1}^N\right]$$

$$= \mathbb{E}\left[\exp\left(itU_{k,1}^N\right)\middle| \mathcal{H}_{k-1}^N\right] - \left(1 - \frac{t^2 (U_{k,1}^N)^2}{2}\right) + \tag{17}$$

$$\left(1 - \frac{t^2 (U_{k,1}^N)^2}{2}\right) - \exp\left(-\frac{t^2}{2}\mathbb{E}\left[\left(U_{k,1}^N\right)^2\middle| \mathcal{H}_{k-1}^N\right]\right) + \tag{18}$$

$$\exp\left(-\frac{t^2}{2}\mathbb{E}\left[\left(U_{k,1}^N\right)^2\middle| \mathcal{H}_{k-1}^N\right]\right) - \exp\left(-\frac{t^2 \hat{V}_k(\xi)/N}{2}\right) \tag{19}$$

We now show that the product of N and the expectation of each of these terms converges to zero. First, consider (17). We can represent e^{iy} as $1 + iy - \frac{y^2}{2} + \frac{|y|^3}{3!}\theta(y)$ for some suitable function $\theta(y), |\theta| < 1$. Thus, as $U_{k,1}^N$ is a martingale increment:

$$\left|\mathbb{E}\left[\exp\left(itU_{k,1}^N\right) - \left(1 - \frac{t^2 (U_{k,1}^N)^2}{2}\right)\middle| \mathcal{H}_{k-1}^N\right]\right|$$

$$\leq \frac{t^3}{6}\mathbb{E}\left[\left|U_{k,1}^N\right|^3\middle| \mathcal{H}_{k-1}^N\right]$$

$$\leq \frac{t^3}{6N^{3/2}} 2^3 \left|\left|\xi\right|\right|_\infty^3 \left(\alpha_{k-1}^{N_{k-1}}(\mathbf{1})R_1 + |Z_{n-1}|R_1\right)^3$$

And N times the expectation of this quantity converges to zero as $N \to \infty$.

To deal with (18) note that $1 - u \leq exp(-u) \leq 1 - u + u^2 \quad \forall u \geq 0$. Setting

$$u = \frac{t^2}{2}\mathbb{E}\left[\left(U_{k,1}^N\right)^2\middle| \mathcal{H}_{k-1}^N\right]$$

one obtains:

$$\mathbb{E}\left[\left|\left(1 - \frac{t^2 (U_{k,1}^N)^2}{2}\right) - \exp\left(-\frac{t^2}{2}\mathbb{E}\left[\left(U_{k,1}^N\right)^2\middle| \mathcal{H}_{k-1}^N\right]\right)\right|\middle| \mathcal{H}_{k-1}^N\right]$$

$$\leq \frac{t^4}{4}\mathbb{E}\left[\left(U_{k,1}^N\right)^2\middle| \mathcal{H}_{k-1}^N\right]^2$$

$$\leq \frac{t^4}{4}\frac{1}{N^2} 4 \left|\left|\xi\right|\right|_\infty^4 \left(\alpha_{k-1}^{N_{k-1}}(\mathbf{1})R_1 + |Z_{k-1}|R_1\right)^4$$

and once again, the expectation of N times the quantity of interest converges to zero.

Finally, (19) can be shown to vanish by considering the following exponential bound. For $v \geq u \geq 0$, we can write $|e^{-u} - e^{-v}| \leq |1 - e^{u-v}| \leq |u - v|$ where the final inequality corresponds to [1, 4.2.32] and this gives us:

$$\left| \exp\left(-\frac{t^2}{2} \mathbb{E}\left[\left(U_{k,1}^N\right)^2 \middle| \mathcal{H}_{k-1}^N \right] \right) - \exp\left(-\frac{t^2 \hat{V}_k(\xi)/N}{2} \right) \right|$$
$$\leq \frac{t^2}{2N} \left| N\mathbb{E}\left[\left(U_{k,1}^N\right)^2 \middle| \mathcal{H}_{k-1}^N \right]^2 - \hat{V}_k(\xi) \right|$$

which can be exploited by noting that:

$$N\mathbb{E}\left[\left(U_{k,i}^N\right)^2 \middle| \mathcal{H}_{k,i-1}^N \right] = \left[\alpha_{k-1}^{N_{k-1}} (G_{k-1,\alpha_{k-1}^{N_{k-1}}}) \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}} \left(\phi_k \times \frac{\phi_k}{q_k} \right) (\xi^2) \right.$$
$$\left. - \hat{S}_{k-1,\alpha_{k-1}^{N_{k-1}}} (\phi_k(\xi))^2 \right] \quad (20)$$
$$\hat{V}_k(\xi) = \left[\alpha_{k-1}(G_{k-1,\alpha_{k-1}}) \hat{S}_{k-1,\alpha_{k-1}} \left(\phi_k \times \frac{\phi_k}{q_k} \right) (\xi^2) \right.$$
$$\left. - \hat{S}_{k-1,\alpha_{k-1}} (\phi_k(\xi))^2 \right] \quad (21)$$

As (20) converges to (21) in probability (cf lemma 8) and (20) is bounded above, (20) converges to (21) in $\mathbb{L}_1$ and the result we seek follows. Consequently, (19) vanishes and we have the result of the lemma. □

References

1. M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover, ninth printing edition, 1970.
2. P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, second edition, 1986.
3. N. Chopin. Central limit theorem for sequential Monte Carlo methods and its applications to Bayesian inference. *Annals of Statistics*, 32(6):2385–2411, December 2004.
4. D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*, volume I: Elementary Theory and Methods of *Probability and Its Applications*. Springer, New York, second edition, 2003.
5. P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Probability and Its Applications. Springer Verlag, New York, 2004.
6. P. Del Moral and A. Guionnet. Central limit theorem for non linear filtering and interacting particle systems. *Annals of Applied Probability*, 9(2):275–297, 1999.
7. A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer Verlag, New York, 2001.
8. I. Gentil, B. Rémillard, and P. Del Moral. Filtering of images for detecting multiple targets trajectories. In *Statistical Modeling and Analysis for Complex Data Problems*. Springer, April 2005. To Appear.

9. I. R. Goodman, R. P. S. Mahler, and H. T. Nguyen. *Mathematics of Data Fusion*. Kluwer Academic Publishers, 1997.
10. H. R. Künsch. Recursive Monte Carlo filters: Algorithms and theoretical analysis. Research Report 112, ETH: Seminar Für Statistik, 2003. To appear in *Annals of Statistics*.
11. R. P. S. Mahler. Multitarget Bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1152, October 2003.
12. C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Verlag, New York, second edition, 2004.
13. A. N. Shiryaev. *Probability*. Number 95 in Graduate Texts in Mathematics. Springer Verlag, New York, second edition, 1995.
14. B. Vo, S.S. Singh, and A. Doucet. Sequential Monte Carlo methods for multi-target filtering with random finite sets. *IEEE Transactions on Aerospace and Electronic Systems*, June 2005. To Appear. http://www2.ee.unimelb.edu.au/staff/bv/vo/phdfilterAESrevfinal2cola.pdf.