The Probability Generating Functional for Finite Point Processes, and Its Application to the Comparison of PHD and Intensity Filters

ROY STREIT

Probability generating functionals (PGFLs) for finite point processes are used to derive the probability hypothesis density (PHD) filter and intensity filter (iFilter) for multitarget tracking. Presenting them in a common PGFL framework makes manifest their similarities and differences. A significant difference is their measurement model—the PHD filter uses an exogenous clutter model and the iFilter uses an endogenous scattering model.

Manuscript received January 1, 2012; revised July 1, 2012 and February 14, 2013; released for publication April 14, 2013.

Refereeing of this contribution was handled by Peter Willett.

Authors' address: Metron, Inc., 1818 Library St., Reston, VA 20190

1. INTRODUCTION

Many radar and sonar sensor systems generate several point measurements at every scan. Some measurements are due to targets and others are due to clutter, or scatterers, in the sensor field of view. The multi-target tracking problem is to estimate the number of targets and their states given the measurements. The multi-hypothesis tracking (MHT) method for solving this problem is based on two widely accepted assumptions: 1) targets are points; and 2) sensors generate at most one measurement per target per scan. The second is called the "at most one measurement per target" rule. It is the cause of the intrinsically high computational complexity of optimal MHT algorithms and, consequently, the reason so many diverse kinds of alternative suboptimal algorithms are widely studied.

This paper concerns the class of multitarget tracking filters based on finite point process models for multiple target states and sensor measurement sets. Two specific kinds of filters are discussed—the PHD (probability hypothesis density) filter and the iFilter (intensity filter). Many of the differences between these filters are due to the different models of the measurement set. Contrasting these two filters in this way has the added benefit of revealing the fundamental importance of the classical methods of finite point processes for tracking applications.

Section 2 provides background on finite point processes and reviews their application to multitarget tracking filters. The next two sections are largely didactic. Section 3 defines the probability generating functional (PGFL) of a single point process. Basic results related to the PGFL are derived there. PGFLs play a central role—they characterize the probability structures underpinning the filters. Section 4 defines the bivariate PGFL of two finite point processes. The general Bayes posterior point process is defined, and its PGFL is derived from the bivariate PGFL.

Section 5 derives the PHD filter and iFilter as examples of the general Bayes posterior point process. The PHD filter uses a traditional clutter model, while the iFilter uses a scattering model. These modeling differences manifest themselves in the PGFLs of the filters, thus exposing the similarities and differences between them. Conclusions and concluding remarks are given in Section 6.

2. BACKGROUND

PGFLs for finite point processes were introduced in 1962 by Moyal [11]. In this seminal paper, Moyal noted the connection between PGFLs and probability generating functions (PGFs) of discrete random variables. He defined functional derivatives of the PGFL and used them (see (15) below) to prove that the PGFL characterizes the point process. He defined the factorial moments using PGFLs. Moyal applied his functional calculus to stochastic population processes, establishing the connection to the classical theory of branching processes (see [1] for more background). Moyal investigated cluster processes and multiplicative processes, which are processes whose PGFL factors as in (42) below. He also studied time-dependent Markovian multiplicative population processes.

Branching processes and point process theory were studied extensively by Harris in 1963 [6]. According to the authoritative text by Daley and Vere-Jones [5, p. 1], point process theory "reached a definitive form in the now classic treatments by Moyal (1962) and Harris (1963)."

Mahler applied PGFLs to multitarget tracking problems in a series of papers; see [9] and [10] and the references therein. In this corpus he uses the FISST (*fi*nite set *st*atistics) calculus to derive the PHD filter. He introduced random finite set (RFS) models for multitarget state, as well as the idea of recursively approximating the Bayes posterior process by a Poisson point process (PPP). The term PHD was coined by Stein and Winter [14], who viewed the process of evidence accrual as additive, as opposed to multiplicative. The reformulation of the PHD using random finite sets is due to Mahler [8]. The PGFL of the Bayes posterior finite point process takes an attractive form (see (28) below). The same form was derived for the PHD tracking filter in [10, Sec. 14.8.2]; however, that result is specific to the tracking application.

An exact expression is given for the probability generating function (PGF) of the distribution of the number of points in the Bayes posterior process *before* the PPP approximation of the multitarget state. The result is a straightforward consequence of the connection between the PGF and the PGFL of the posterior process, but nonetheless it may be new. These discrete distributions provide insight into the nature of the exogenous and endogenous measurement models, as well as the PPP approximation to multitarget state.

The distinction drawn between exogenous and endogenous measurement models is perhaps new, but the use of the augmented state space, denoted below by S^+, in tracking applications dates to at least 1986 (see [7]). (More general augmented state spaces are used by Chen, et al. [3] for dynamic clutter modeling.) The iFilter was derived by Streit and Stone [15] using a direct enumeration of measurements to targets that avoids PGFLs. Their Bayesian method is based on well-known properties of PPPs [16]. The PGFL derivation of the iFilter presented in this paper is new. The iFilter was first referred to by that name in 2010.

The relationship between medical imaging algorithms and the PHD and iFilter was first discussed in [17]. The similarity between them and the famous Shepp-Vardi algorithm (1982) for positron emission tomography (PET) is remarkable. The relationship arises because PET uses PPP models for the image—the spatial distribution of a radioisotope, i.e., the intensity function of radioisotope decays. The connection to the classic Richardson-Lucy (1972/1974) algorithm for image restoration problems is also pointed out in [16].

3. PROBABILITY GENERATING FUNCTIONALS

The event space $\mathcal{E}(S)$ of the finite point process Ξ is the set of all ordered pairs of the form $\xi = (n, \{s_1, \ldots, s_n\})$, $s_i \in S$. For $n = 0$, the event is $(0, \emptyset)$. For $n \geq 1$ the event corresponds to $n!$ equally likely, ordered events of the form $(n, s_{\sigma(1)}, \ldots, s_{\sigma(n)})$, $\sigma \in Sym(n)$, where $Sym(n)$ denotes the set of all permutations of the first n positive integers. The space S can be very general, but is typically a specified subset of $\mathbb{R}^d$, $d \geq 1$. In physics, n is called the canonical number, the collection $\mathcal{E}_n(S)$ of all subsets of S with n points is the nth canonical ensemble, and the space $\mathcal{E}(S)$ is the grand canonical ensemble.

A functional is, in general, merely a name for an operator whose input is a function and output is a (real or complex) number. For example, definite integrals are functionals. PGFLs for general finite point processes were defined by Moyal [11, Sec. 4] as a generalization of PGFs for multivariate discrete random variables. He showed that PGFLs characterize the point process via its functional derivatives. The results presented in this section are due to Moyal. The presentation here is didactic in style and intended to be widely accessible.

3.1. Definition of the PGFL

Let Ξ be a random variable with outcomes $\xi \in \mathcal{E}(S)$. Define $\Xi = (N, X)$, where N is the canonical number and X is the set of points in the random canonical ensemble $\mathcal{E}_N(S)$. The PGFL of Ξ is defined for real-valued functions h on the state space S as

$$G^{\Xi}[h] = \sum_{n=0}^{\infty} p_N^{\Xi}(n) \int_{S^n} \left(\prod_{i=1}^{n} h(s_i) \right) p_{X|N}^{\Xi}(s_1, \ldots, s_n \mid n) ds_1 \cdots ds_n \tag{1}$$

where $p_N^{\Xi}(n)$ is the distribution (probability mass function or discrete pdf) of N, and $p_{X|N}^{\Xi}(s_1, \ldots, s_n \mid n)$ is the pdf of the points $(s_1, \ldots, s_n)$ conditioned on $N = n$. For $n = 0$, $p_{X|N}^{\Xi}(\cdot \mid n) = 1$ and $\prod_{i=1}^{n} h(s_i)$ is defined to be one. Simply put, the PGFL is the expectation of the random product $\prod_{i=1}^{N} h(s_i)$. The PGFL is evaluated only for functions h such that the integrals and the sum in are absolutely convergent. It is sufficient [11] to require that $|h(s)| \leq 1$ for $s \in S$. No physical units are associated with the values of $h(s)$, so the integrals in (1) are unitless and the sum is dimensionally consistent.

A finite point process Ξ is a PPP if the canonical number N is Poisson distributed with mean $\mu =$

$\int_S f^{\Xi}(s)ds < \infty$, where $f^{\Xi}(s) \geq 0$ is the intensity function, and points are independently and identically distributed in S with pdf $f^{\Xi}(s)/\mu$. Thus, $p_N^{\Xi}(n) = e^{-\mu}\mu^n/n!$ and $p_{X|N}^{\Xi}(s_1,\ldots,s_n \mid n) = \mu^{-n}\prod_{i=1}^n f^{\Xi}(s_i)$. Direct calculation shows that

$$G^{\Xi}[h] = \exp\left[-\int_S f^{\Xi}(s)ds + \int_S h(s)f^{\Xi}(s)ds\right]. \quad (2)$$

The PGFL (2) is log-linear, that is, $\log(G^{\Xi}[h]/G^{\Xi}[0])$ is linear in h. For further discussion of PPPs and their applications, see [16].

3.2. Functional Derivatives of the PGFL

The finite set statistics (FISST) calculus concerns functional differentiation of PGFLs, where functional differentiation has exactly the same meaning as in the Calculus of Variations. The functional derivative of $G^{\Xi}[h]$ with respect to the variation w is defined by

$$\frac{\partial G^{\Xi}}{\partial w}[h] = \lim_{\varepsilon\to 0^+}\frac{d}{d\varepsilon}G^{\Xi}[h+\varepsilon w] = \lim_{\varepsilon\to 0^+}\frac{G^{\Xi}[h+\varepsilon w]-G^{\Xi}[h]}{\varepsilon}. \quad (3)$$

Here, w is a bounded real-valued function on S. (It will be specified shortly.) From (1),

$$G^{\Xi}[h+\varepsilon w] = \sum_{n=0}^{\infty} p_N^{\Xi}(n)\int_{S^n}\prod_{i=1}^n [h(s_i)+\varepsilon w(s_i)] \times p_{X|N}^{\Xi}(s_1,\ldots,s_n \mid n)ds_1\cdots ds_n. \quad (4)$$

Moyal [11, Sec. 4] proves that (4) is an analytic function of ε in some open region of the complex plane containing the origin. Using (3) gives, since analyticity in ε justifies interchanging the sum and the derivative,

$$\frac{\partial G^{\Xi}}{\partial w}[h] = \sum_{n=1}^{\infty} p_N^{\Xi}(n)\sum_{k=1}^{n}\int_{S^n} w(s_k)\prod_{i=1,i\neq k}^{n} h(s_i) \times p_{X|N}^{\Xi}(s_1,\ldots,s_n \mid n)ds_1\cdots ds_n. \quad (5)$$

The outermost sum starts at $n = 1$ because the derivative with respect to ε of the $n = 0$ term is zero. The innermost sum over $i \neq k$ arises from the product rule for ordinary differentiation. The Dirac delta function $\delta_x(s) \equiv \delta(s-x)$ is called an "impulse (point mass) at $s = x \in S$." Specifying the variation to be $w(s) = \delta_x(s)$ gives the functional derivative

$$\frac{\partial G^{\Xi}}{\partial x}[h] \equiv \frac{\partial G^{\Xi}}{\partial \delta_x}[h] = \left.\frac{\partial G^{\Xi}}{\partial w}[h]\right|_{w(\cdot)=\delta_x(\cdot)} = \sum_{n=1}^{\infty} p_N^{\Xi}(n)\sum_{k=1}^{n}\int_{S^n}\delta_x(s_k)\prod_{i=1,i\neq k}^{n} h(s_i) \times p_{X|N}^{\Xi}(s_1,\ldots,s_n \mid n)ds_1\cdots ds_n. \quad (6)$$

(Alternatively, specifying the variation to be a function in a test sequence for the delta function and taking the limit gives the same result.) Using the sampling property of the Dirac delta function, the argument symmetries of $p_{X|N}^{\Xi}(\cdot)$, and relabeling arguments appropriately gives

$$\frac{\partial G^{\Xi}}{\partial x}[h] = \sum_{n=1}^{\infty} p_N^{\Xi}(n)n\int_{S^{n-1}}\prod_{i=2}^{n} h(s_i) \times p_{X|N}^{\Xi}(x,s_2,\ldots,s_n \mid n)ds_2\cdots ds_n \quad (7)$$

where the product in (7) is taken equal to one for $n = 1$. The integrals are over S^{n-1}, not S^n. Note that the derivative is a functional.

It is important to keep in mind that taking the variation w to be equal to the Dirac delta function δ_x makes the derivative $\partial G^{\Xi}[h]/\partial x$ depend on the point x even though $G_{\Xi}[h]$ itself does not. For this reason, the functional derivative (7) is referred to in this paper as the derivative with respect to an impulse at x, not simply as the derivative with respect to x.

Derivatives of the PGFL with respect to any finite number of distinct impulses extract, or decode, the pdf of Ξ from its PGFL. To find the functional derivative with respect to $x_2 \neq x_1$, start with (7) by replacing x with x_1 and h with $h + \varepsilon w$. This gives

$$\frac{\partial G^{\Xi}}{\partial x_1}[h+\varepsilon w] = \sum_{n=1}^{\infty} p_N^{\Xi}(n)n\int_{S^{n-1}}\prod_{i=2}^{n}[h(s_i)+\varepsilon w(s_i)] \times p_{X|N}^{\Xi}(x_1,s_2,\ldots,s_n \mid n)ds_2\cdots ds_n. \quad (8)$$

Differentiating with respect to ε and setting $\varepsilon = 0$ gives

$$\frac{\partial}{\partial w}\left(\frac{\partial G^{\Xi}}{\partial x_1}[h]\right) = \sum_{n=2}^{\infty} p_N^{\Xi}(n)n\sum_{k=2}^{n}\int_{S^{n-1}} w(s_k)\prod_{i=2,i\neq k}^{n} h(s_i) \times p_{X|N}^{\Xi}(x_1,s_2,\ldots,s_n \mid n)ds_2\cdots ds_n. \quad (9)$$

Substituting the variation $w(s) = \delta_{x_2}(s)$, where $x_2 \neq x_1$, and using symmetry properties of $p_{X|N}^{\Xi}(\cdot)$ gives the functional derivative,

$$\frac{\partial^2 G^{\Xi}}{\partial x_2 \partial x_1}[h] = \sum_{n=2}^{\infty} p_N^{\Xi}(n)n(n-1)\int_{S^{n-2}}\prod_{i=3}^{n} h(s_i) \times p_{X|N}^{\Xi}(x_1,x_2,s_3,\ldots,s_n \mid n)ds_3\cdots ds_n \quad (10)$$

where, for $n = 2$, the product is equal to one. The integrals are now over S^{n-2}.

Functional derivatives of the PGFL with respect to the variations $w_1,\ldots,w_n$ are defined recursively as above, or equivalently as

$$\frac{\partial^n G^\Xi}{\partial w_1 \cdots \partial w_n}[h] = \frac{\partial^n G^\Xi}{\partial \varepsilon_1 \cdots \partial \varepsilon_n}\left[h + \sum_{j=1}^n \varepsilon_j w_j\right]_{\varepsilon_1 = \cdots = \varepsilon_n = 0}. \tag{11}$$

The functional derivative with respect to impulses at distinct points $x_1,\ldots,x_n$ is

$$\begin{aligned}\frac{\partial^n G^\Xi}{\partial x_1 \cdots \partial x_n}[h] &\equiv \left.\frac{\partial^n G^\Xi}{\partial w_1 \cdots \partial w_n}[h]\right|_{w_1=\delta_{x_1},\ldots,w_n=\delta_{x_n}} \\ &= \sum_{k=n}^\infty p_N^\Xi(k)k(k-1)\cdots(k-n+1) \\ &\quad \times \int_{S^{k-n}}\left(\prod_{i=n+1}^k h(s_i)\right) \\ &\quad \times p_{X|N}^\Xi(x_1,\ldots,x_n,s_{n+1},\ldots,s_k \mid k)ds_{n+1}\cdots ds_k\end{aligned} \tag{12}$$

where for $k = n$ the product is equal to one. The order of differentiation is immaterial. For convenience, the derivative for $n = 0$ is defined to be $G^\Xi[h]$. The derivative (12) is first order with respect to the distinct points $x_1,\ldots,x_n$.

3.3. Event Likelihood

Evaluating (7) and (10) for $h(\cdot) \equiv 0$ gives, respectively,

$$\frac{\partial G^\Xi}{\partial x}[0] = p_N^\Xi(1)p_{X|N}^\Xi(x \mid N = 1) \tag{13}$$

and

$$\frac{\partial^2 G^\Xi}{\partial x_1 \partial x_2}[0] = \frac{\partial^2 G^\Xi}{\partial x_2 \partial x_1}[0] = 2!p_N^\Xi(2)p_{X|N}^\Xi(x_1,x_2 \mid N = 2). \tag{14}$$

In words, the derivative evaluated at $h \equiv 0$ is the pdf of the event $\xi = (1,\{x\})$, and the derivative with respect to impulses at x_1 and x_2 is the pdf of the event $\xi = (2,\{x_1,x_2\})$ or, equivalently, 2! times the pdf of the ordered event $(2,x_1,x_2)$. From (12), for $n \geq 1$ distinct impulses,

$$\begin{aligned}\frac{\partial^n G^\Xi}{\partial x_1 \cdots \partial x_n}[0] &= n!p_N^\Xi(n)p_{X|N}^\Xi(x_1,\ldots,x_n \mid n) \\ &= n!p^\Xi(n,x_1,\ldots,x_n) \\ &= p^\Xi(n,\{x_1,\ldots,x_n\})\end{aligned} \tag{15}$$

where $p^\Xi(n,\{x_1,\ldots,x_n\})$ is the pdf of Ξ for unordered events and $p^\Xi(n,x_1,\ldots,x_n)$ is pdf for the corresponding ordered event.

The derivatives (15) show that a finite point process is characterized by its PGFL. This fact is important because it means that a finite point process can be defined by deriving its PGFL.

3.4. Factorial Moments

The first moment of Ξ is the special case of (7) with $h(s) \equiv 1$:

$$\begin{aligned}m_{[1]}^\Xi(x) &= \frac{\partial G^\Xi}{\partial x}[1] \\ &= \sum_{n=1}^\infty n p_N^\Xi(n)\int_{S^{n-1}} p_{X|N}^\Xi(x,s_2,\ldots,s_n \mid n)ds_2\cdots ds_n.\end{aligned} \tag{16}$$

For PPPs it is straightforward to verify from the PGFL (2) that the intensity function $f^\Xi(x)$ is identical to the first moment, i.e., $f^\Xi(x) = m_{[1]}^\Xi(x)$. For this reason the first moment of a finite point process is often called the intensity function.

Substituting $h(s) \equiv 1$ into (12) gives the nth factorial moment,

$$\begin{aligned}m_{[n]}^\Xi&(x_1,\ldots,x_n) \\ &\equiv \frac{\partial^n G^\Xi}{\partial x_1 \cdots \partial x_n}[1] \\ &= \sum_{k=n}^\infty p_N^\Xi(k)k(k-1)\cdots(k-n+1) \\ &\quad \times \int_{S^{k-n}} p_{X|N}^\Xi(x_1,\ldots,x_n,s_{n+1},\ldots,s_k \mid k)ds_{n+1}\cdots ds_k\end{aligned} \tag{17}$$

where for $k = n$ the conditional pdf is $p_{X|N}^\Xi(x_1,\ldots,x_n \mid n)$. Factorial moments can be interpreted as multi-point intensity functions (when points are distinct with probability one). To see this, note that (17) can be written intuitively as [4, eq. (5.4.12)]

$$\begin{aligned}m_{[n]}^\Xi&(x_1,\ldots,x_n)dx_1\cdots dx_n \\ &= \Pr\left[\begin{array}{c}\text{exactly one point of the process is} \\ \text{located in each infinitesimal subset} \\ [x_i,x_i+dx_i),\ i=1,\ldots,n\end{array}\right].\end{aligned} \tag{18}$$

For $n = 1$ and $n = 2$, for distinct points $x,y \in S$,

$$\begin{aligned}m_{[1]}^\Xi&(x)dx \\ &= \Pr[\text{exactly one point in } [x,x+dx)] \\ m_{[2]}^\Xi&(x,y)dx\,dy \\ &= \Pr[\text{exactly one point in } [x,x+dx) \\ &\qquad \text{and one point in } [y,y+dy)].\end{aligned} \tag{19}$$

For PPPs the second probability is the product of $m_{[1]}^\Xi(x)dx$ and $m_{[1]}^\Xi(y)dy$, a result that follows from well-known independence properties of PPPs. In general, however, the second moment does not factor. The application of factorial moments in tracking applications is discussed in [2] but is outside the scope of the present paper.

3.5. Probability Generating Function of Canonical Number

The probability generating function (PGF) of N, denoted by $F^{\Xi}(x)$, is determined by evaluating the PGFL of Ξ for the constant function $h(s) \equiv x$. Substituting into (1) gives

$$F^{\Xi}(x) \equiv G^{\Xi}[h]|_{h(\cdot)\equiv x} \equiv G^{\Xi}[x] = \sum_{n=0}^{\infty} p_N^{\Xi}(n) x^n. \tag{20}$$

In the signal processing literature, $F^{\Xi}(z^{-1})$ is called the z-transform of the sequence of probabilities $(p_N^{\Xi}(n) : n = 0,1,\ldots)$. The probability $p_N^{\Xi}(n)$ is

$$p_N^{\Xi}(n) = \frac{1}{n!}\frac{d^n F^{\Xi}}{dx^n}(0) \tag{21}$$

where the nth derivative with respect to x is the ordinary derivative evaluated at $x = 0$. The probability $p_N^{\Xi}(n)$ is $n!$ times the integral of the ordered pdf $p^{\Xi}(n,x_1,x_2,\ldots,x_n)$ over all $x_1,x_2,\ldots,x_n$. The first derivative of the PGF evaluated at $x = 1$ is

$$\frac{dF^{\Xi}}{dx}(1) = \sum_{n=0}^{\infty} p_N^{\Xi}(n) n x^{n-1}\Bigg|_{x=1} \equiv E^{\Xi}[N] \tag{22}$$

where $E^{\Xi}[N]$ is the expected number of points in a realization of Ξ.

4. BAYES POSTERIOR POINT PROCESS

In this section the conditional, or posterior, point process $\Sigma \mid \Upsilon$ is defined using Bayes method in terms of the bivariate process (Υ,Ξ). The random variables are finite point processes, but this does not alter the Bayesian methodology. The PGFL of the Bayes posterior process $\Xi \mid \Upsilon$ and two summary statistics, namely, the intensity function and the distribution of the canonical number, are derived. Finally, Bayesian estimates are defined using the posterior point process and a specified loss function.

For tracking applications, Υ is the observation space and Ξ the multitarget state space. The points of a realization $\Upsilon = \upsilon$ are the measurements in a sensor scan. The joint pdf of the measurement and target processes is denoted by $p^{\Upsilon\Xi}(\upsilon,\xi)$, where $\Xi = \xi$ is a realization of the target process. The conditional pdf $p^{\Upsilon|\Xi}(\upsilon \mid \xi)$ is derived from physical models of the targets and the sensor likelihood function $p(y \mid s)$.

4.1. Bivariate PGFL

Let Υ be a finite point process with events $\upsilon = (m,\{y_1,\ldots,y_m\}) \in \mathcal{E}(Y)$, where the space Y is in general unrelated to the space S. Extending the definition of the PGFL for Ξ to the joint process (Υ,Ξ) with events in the Cartesian product space $\mathcal{E}(Y)\times\mathcal{E}(S)$ gives the bivariate PGFL as the expectation of the product of random products $\prod_{i=1}^{M} g(y_i)\prod_{j=1}^{N} h(s_j)$, that is,

$$\begin{aligned}G^{\Upsilon\Xi}[g,h] = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} p_{MN}^{\Upsilon\Xi}(m,n) &\times \int_{Y^m}\int_{S^n}\left(\prod_{i=1}^{m} g(y_i)\right)\left(\prod_{j=1}^{n} h(s_j)\right)\\ &\times p_{YX|MN}^{\Upsilon\Xi}(y_1,\ldots,y_m,s_1,\ldots,s_n \mid m,n)\\ &\times dy_1\cdots dy_m ds_1\cdots ds_n\end{aligned} \tag{23}$$

where $p_{MN}^{\Upsilon\Xi}(\cdot)$ and $p_{YX|MN}^{\Upsilon\Xi}(\cdot)$ are the discrete and continuous pdfs associated with the joint process (Υ,Ξ). If $m = 0$ or $n = 0$ in (23), the corresponding product is defined to be one. It is important to keep in mind that $g(\cdot)$ and $h(\cdot)$ are functions defined on Y and S, respectively.

Marginalizing the bivariate point process over one process yields the PGFL of other process. More formally,

$$G^{\Upsilon\Xi}[1,h] = G^{\Xi}[h] \qquad \text{and} \qquad G^{\Upsilon\Xi}[g,1] = G^{\Upsilon}[g]. \tag{24}$$

To obtain the first expression, substitute $g(\cdot) = 1$ in (23), integrate over $y_1,\ldots,y_m$, and sum over m. The other expression is obtained similarly.

4.2. PGFL of the Bayes Posterior Point Process

To write the PGFL of the Bayes posterior point process, note that the derivative of (23) with respect to impulses at the distinct points $\{y_1,\ldots,y_m\} \subset Y$ evaluated for $g(\cdot) = 0$ is

$$\begin{aligned}\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1\cdots\partial y_m}[0,h] &= m!\sum_{n=0}^{\infty} p_{MN}^{\Upsilon\Xi}(m,n)\int_{S^n}\left(\prod_{j=1}^{n} h(s_j)\right)\\ &\quad\times p_{YX|MN}^{\Upsilon\Xi}(y_1,\ldots,y_m,s_1,\ldots,s_n \mid m,n) ds_1\cdots ds_n.\end{aligned} \tag{25}$$

Evaluating the derivative of $G^{\Upsilon\Xi}[g,1] = G^{\Upsilon}[g]$ with respect to impulses at $y_1,\ldots,y_m$ for $g(\cdot) = 0$ gives

$$\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1\cdots\partial y_m}[0,1] = \frac{\partial^m G^{\Upsilon}}{\partial y_1\cdots\partial y_m}[0] = m! p_{MY}^{\Upsilon}(m,y_1,\ldots,y_m). \tag{26}$$

Dividing (25) by (26) yields

$$\frac{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,h]}{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,1]} = \sum_{n=0}^{\infty} \int_{S^n} \left(\prod_{j=1}^{n} h(s_j) \right) \times \frac{p_{MYNX}^{\Upsilon\Xi}(m,y_1,\ldots,y_m,n,s_1,\ldots,s_n)}{p_{MY}^{\Upsilon}(m,y_1,\ldots,y_m)} ds_1 \cdots ds_n$$
$$= \sum_{n=0}^{\infty} \int_{S^n} \left(\prod_{j=1}^{n} h(s_j) \right) \times p_{NX|MY}^{\Xi|\Upsilon}(n,s_1,\ldots,s_n \mid m,y_1,\ldots,y_m) ds_1 \cdots ds_n. \tag{27}$$

Substituting the Bayes factorization

$$p_{NX|MY}^{\Xi|\Upsilon}(n,s_1,\ldots,s_n \mid m,y_1,\ldots,y_m) = p_{N|MY}^{\Xi|\Upsilon}(n \mid m,y_1,\ldots,y_m) p_{X|NMY}^{\Xi|\Upsilon}(s_1,\ldots,s_n \mid m,y_1,\ldots,y_m)$$

into (27) and comparing the result with definition (1) shows that the ratio is the PGFL of the Bayes posterior process, that is,

$$G^{\Xi|\Upsilon}[h \mid y_1,\ldots,y_m] = \frac{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,h]}{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,1]}. \tag{28}$$

The PGFL (28) is valid for general finite point processes. The denominator of (28) is a constant that scales the PGFL of the numerator.

The probability structure of the Bayes posterior process is characterized by the functional derivatives of (28) with respect to impulses at the distinct points $\{x_1,\ldots,x_n\} \subset S$. A specialized version of (28) for multitarget tracking applications is derived in [10, p. 757], where it is described as the PGFL "form of the multitarget corrector."

4.3. Summary Statistics of the Bayes Posterior Process

Since the event space $\mathcal{E}(S)$ is very large, it is useful to provide summary statistics of the posterior process $\Xi \mid \Upsilon$. Two statistics are of interest here. The first is the intensity function $f^{\Xi|\Upsilon}(x)$ of $\Xi \mid \Upsilon$. It is found by the evaluating at $h(s) = 1$ the functional derivative of (28) with respect to an impulse at $x \in S$:

$$f^{\Xi|\Upsilon}(x) = \frac{\frac{\partial^{m+1} G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m \partial x}[0,1]}{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,1]}, \qquad x \in S. \tag{29}$$

The expression (29) holds for general finite point processes.

The other summary statistic is the distribution of $N^{\Xi|\Upsilon}$, the canonical number of points in the Bayes posterior process. The PGF of $N^{\Xi|\Upsilon}$ is the PGFL (28) evaluated for the constant function $h(s) = x$; that is,

$$F^{\Xi|\Upsilon}(x) = \frac{\left.\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,h]\right|_{h(\cdot)\equiv x}}{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,1]}. \tag{30}$$

The posterior pdf of the canonical number is, using (21),

$$p_N^{\Xi|\Upsilon}(n) = \frac{1}{n!}\frac{d^n}{dx^n} F^{\Xi|\Upsilon}(0) = \frac{\frac{1}{n!}\frac{d^n}{dx^n}\left(\left.\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,h]\right|_{h(\cdot)\equiv x} \right)_{x=0}}{\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,1]}. \tag{31}$$

From (22), the expected number of targets is $E[N^{\Xi|\Upsilon}] = (d/dx)F^{\Xi|\Upsilon}(1)$.

4.4. Bayesian and Pseudo-MAP Estimators

A Bayesian estimate of Ξ is determined using a specified loss function $L(\zeta \mid \xi)$. This function gives the loss associated with choosing the estimate $\zeta \in \mathcal{E}(S)$ for Ξ when the true realization is $\xi \in \mathcal{E}(S)$. The Bayes loss of ζ is the expected loss, $R(\zeta) \equiv E[L(\zeta \mid \xi)]$, where the expectation (see (1)) is with respect to the density $p^{\Xi|\Upsilon}(\xi \mid \upsilon)$ of the Bayes posterior process $\Xi \mid \Upsilon$. The Bayes estimate $\hat{\xi}_{\text{Bayes}} \in \mathcal{E}(S)$ for Ξ minimizes the Bayes loss:

$$\hat{\xi}_{\text{Bayes}} = \arg\min_{\zeta \in \mathcal{E}(S)} R(\zeta). \tag{32}$$

The Bayes estimate depends on the choice of the loss function $L(\zeta \mid \xi)$.

In many problems, $L(\zeta \mid \xi)$ can be specified so that the Bayes estimate reduces to the maximum a posteriori (MAP) estimate, $\arg\max_{\xi} p^{\Xi|\Upsilon}(\xi \mid \upsilon)$. However, the MAP estimate is undefined for the posterior pdf $p^{\Xi|\Upsilon}(\xi \mid \upsilon)$. To see this, it is only necessary to observe that $p^{\Xi|\Upsilon}(\xi_1 \mid \upsilon)$ and $p^{\Xi|\Upsilon}(\xi_2 \mid \upsilon)$ have different units when the realizations ξ_1 and ξ_2 have different numbers of points.

Pseudo-MAP estimates can be defined using the posterior distribution of the canonical number and intensity functions, or other summary statistics. The consistency of such estimates must be verified on a case by case basis. These topics are outside the scope of the paper.

4.5. Branching Process Form of the Bivariate PGFL

The fundamental Bayesian paradigm of traditional single target tracking is: The observation process is Y, the object/target process is X, the conditional process $Y \mid X$ is the connection between them, and Bayes Theorem is used to inference the Bayesian process $X \mid Y$. The same approach is adopted here, but with X replaced by Ξ and Y by Υ. This changes the mathematical details, but not the Bayesian paradigm.

Note that the Bayes posterior point process—its pdf, intensity, and canonical distribution—were obtained above without reference to the conditional measurement point process $\Upsilon \mid \Xi$. The conditional process is fundamental to the traditional understanding of how measurement processes are exploited to make inferences about target processes. This section follows the traditional Bayesian paradigm, but uses point process models. The approach to Bayesian tracking problems using point processes was first discussed in [9].

The bivariate PGFL is written in a branching process form, that is, as the composition of two functionals. This form lies at the root of many diverse applications, including the tracking applications discussed in the succeeding sections of the paper. Population and branching processes are an established part of the classic literature of probability [1] and [6].

From Bayes theorem,

$$p_{MYNX}^{\Upsilon\Xi}(\cdot) \equiv p_{NX}^{\Xi}(\cdot)p_{MY|NX}^{\Upsilon|\Xi}(\cdot) = p_{N}^{\Xi}(\cdot)p_{X|N}^{\Xi}(\cdot)p_{MY|NX}^{\Upsilon|\Xi}(\cdot). \tag{33}$$

Substituting this into (23) and interchanging the order of sums and integrals (justified by absolute convergence) gives

$$G^{\Upsilon\Xi}[g,h] = \sum_{n=0}^{\infty} p_N^{\Xi}(n) \int_{S^n} \left(\prod_{i=1}^{n} h(s_i)\right) G^{\Upsilon|\Xi}[g \mid s_1,\ldots,s_n] \times p_{X|N}^{\Xi}(s_1,\ldots,s_n \mid n) ds_1 \cdots ds_n \tag{34}$$

where

$$G^{\Upsilon|\Xi}[g \mid s_1,\ldots,s_n] = \sum_{m=0}^{\infty} \int_{Y^m} \left(\prod_{j=1}^{m} g(y_i)\right) \times p_{MY|NX}^{\Upsilon|\Xi}(m,y_1,\ldots,y_m \mid n,s_1,\ldots,s_n) dy_1 \cdots dy_m \tag{35}$$

is the PGFL of the conditional process $\Upsilon \mid \Xi$, as is seen by using the Bayes factorization $p_{MY|NX}^{\Upsilon|\Xi}(\cdot) = p_{M|NX}^{\Upsilon|\Xi}(\cdot)$ $p_{Y|MNX}^{\Upsilon|\Xi}(\cdot)$.

The pdf of the conditional process $\Upsilon \mid \Xi$ is found by functional differentiation (cf. (15)):

$$p^{\Upsilon|\Xi}(\upsilon \mid n,s_1,\ldots,s_n) = \frac{\partial^m}{\partial y_1 \cdots \partial y_m} G^{\Upsilon|\Xi}[0 \mid s_1,\ldots,s_n] \tag{36}$$

where $\upsilon = (m,\{y_1,\ldots,y_m\})$. Evaluating the functional derivatives (36) reveals the detailed structure of the likelihood function, including any enumerations that are inherent in the conditional process $\Upsilon \mid \Xi$. A target tracking example is given in the Appendix.

The expression (34) simplifies greatly if the conditional PGFL factors, that is, if it corresponds to the superposition of conditionally independent measurement processes. If it does, then

$$G^{\Upsilon|\Xi}[g \mid s_1,\ldots,s_n] = \prod_{i=1}^{n} T[g \mid s_i] \tag{37}$$

where $T[g \mid s]$, $s \in S$, is a specified functional. Substituting (37) into (34) gives the branching process form of the bivariate PGFL:

$$G^{\Upsilon\Xi}[g,h] = G^{\Xi}[hT[g \mid \cdot]] \tag{38}$$

where $G^{\Xi}[h]$ denotes the PGFL of Ξ. This is the branching process form of the PGFL. It is central to multitarget tracking applications.

5. SINGLE SENSOR MULTITARGET TRACKING

As noted above, the PGFL characterizes the finite point process. Therefore, in applications, the formulation of the PGFL is of first importance. The filters in this section and the next are derived directly from the relevant PGFL.

The PGFL approach is applied to two measurement models for multitarget tracking. Both models are consistent with the "at most one measurement per target" rule. The PHD filter uses an exogenous clutter model, that is, clutter arises spontaneously in the sensor measurement space and is superposed with the target measurement processes. This is a natural model if the outputs of the sensor signal processor are thresholded to produce point measurements that are fed to a post-processor. It is a standard model that is widely accepted in the tracking literature. The exogenous clutter process is assumed to be a PPP on S with intensity $\lambda(x)$, $x \in S$.

In contrast, the iFilter uses an endogenous model in which all measurements are attributed to scatterers in the augmented target state space $S^+ \equiv S \cup \phi$ defined below in Section 5.2. A target is a scatterer whose state is in S, and a clutter measurement corresponds to a scatterer whose state in S is unknown, i.e., it is φ. This is a natural model for sensor signal processors when distinctions between scatterers are not drawn. It is relatively unused in the tracking community. The different models lead to remarkably similar filters.

The endogenous and exogenous models are mathematically compatible; that is, they can be used together in the same problem. What is mathematically possible, however, must also be justified in the application. This possibility is not explored further in this paper.

The PGFL of the superposition of mutually independent finite point processes is the product of their PGFLs (see [5, Prop. 9.4.1.IX]). This property makes the PGFL useful in many problems involving enumeration, since crucial questions often revolve around learning which process gave rise to which point in a superposition of points—this is the measurement to target assignment problem.

The target process Ξ is interpreted throughout the paper as the point process that is predicted to the current time from the previous time step. Prediction involves independent thinning (see [16, Sec. 2.8]) and Markovian target motion, neither of which alters the character of the target process model—if the target process is a PPP at the previous time step, the predicted process is a PPP. The essential differences between the PHD filter and the iFilter are due the different measurement models, not the predicted target process.

For concreteness, denote the target process at the previous time step by Ξ^-. It is defined on S for exogenous models and on S^+ for endogenous models. For the filters considered here, Ξ^- is assumed to be a PPP to close the Bayesian recursion. Denote its intensity function by $f^{\Xi^-}(\cdot)$. Target motion from the previous time step to the current one is assumed to be Markovian. For exogenous models, denote the transition (motion) model by $F(x \mid x^-)$, x^-, $x \in S$. Thus, $\int_S F(x \mid x^-)dx = 1$ for all $x^- \in S$. Let $d(x)$ denote the probability that a target at x does not survive to the next time step, and let $B(x)$ denote the intensity of a new target PPP at x in the current time step. The predicted process Ξ is the process Ξ^- after it is thinned by $d(x)$ and transformed by $F(x \mid x^-)$, and with new targets superimposed. The process Ξ is a PPP on S, and its intensity is

$$f^{\Xi}(x) = B(x) + \int_S F(x \mid x^-)(1 - d(x^-)) f^{\Xi^-}(x^-)dx^-. \tag{39}$$

The result can be derived in several ways (see, e.g., [16]), but none are given here.

The motion model for endogenous models is denoted by $\Psi(x \mid x^-)$, $x^-, x \in S^+$. Integrals over the augmented space are defined as in (56) below. Thus,

$$1 = \int_{S^+} \Psi(x \mid x^-)dx \equiv \Psi(\phi \mid x^-) + \int_S \Psi(x \mid x^-)dx$$

for all $x^- \in S^+$. The survival probability $d(x)$ is defined for all $x \in S^+$. For convenience, $d(x)$ is assumed to be the same as for the exogenous model for $x \in S$, but the probability $d(\phi)$ is new and must be specified. After thinning and transformation by $\Psi(x \mid x^-)$, the predicted process Ξ is a PPP on S^+, and its intensity is

$$\begin{aligned} f^{\Xi}(x) &= \int_{S^+} \Psi(x \mid x^-)(1 - d(x^-)) f^{\Xi^-}(x^-)dx^- \\ &\equiv \Psi(x \mid \phi)(1 - d(\phi)) f^{\Xi^-}(\phi) \\ &\quad + \int_S \Psi(x \mid x^-)(1 - d(x^-)) f^{\Xi^-}(x^-)dx^-, \qquad x \in S^+. \end{aligned} \tag{40}$$

This result can be derived in the same manner as (39).

The transition model $\Psi(x \mid x^-)$ and intensity $f^{\Xi^-}(\phi)$ can be chosen to match (39) on the subspace S of S^+. To do this, let $d(\phi) = 0$ and set $f^{\Xi^-}(\phi) = \mu + \lambda$, where $\mu = \int_S B(x)dx$ and $\lambda = \int_Y \lambda(y)dy$, respectively, are the expected numbers of new targets and clutter (see Assumption 2 in Section 5.1 below) in the exogenous model. The transition function Ψ on S^+ is defined in terms of the parameters of the exogenous model via the partitioned matrix

$$\left[\begin{array}{c|c} \Psi(x \mid x^-) & \Psi(\phi \mid x^-) \\ \hline \Psi(x \mid \phi) & \Psi(\phi \mid \phi) \end{array}\right] = \left[\begin{array}{c|c} F(x \mid x^-) & 0 \\ \hline B(x)/(\mu + \lambda) & \lambda/(\mu + \lambda) \end{array}\right] \qquad \text{for} \quad x, x^- \in S. \tag{41}$$

For these choices, and complementary ones for the measurement likelihood function (see the paragraph following (63) below), (40) reduces to (39) on S.

5.1. Exogenous Clutter—the PHD Filter

The set of target states and the set of sensor measurements are modeled as finite point processes Ξ and Υ with outcomes $\xi \in \mathcal{E}(S)$ and $\upsilon \in \mathcal{E}(Y)$, respectively, where the target state space is S and sensor measurement space is Y. The PGFL of $\Upsilon \mid \Xi$ is obtained under three assumptions:

1. The target process Ξ is a PPP on S with intensity function $f^{\Xi}(s)$.
2. Conditioned on the event $\Xi = \xi = (n, \{s_1, \ldots, s_n\})$, the measurement process is the superposition of n mutually independent, identical, target-originated measurement processes and a PPP clutter process on Y with intensity function $\lambda(y)$ that is independent of targets.
3. No target generates more than one measurement in the space Y.

The exogenous clutter model is part of the second assumption. The target-originated measurement processes in the second assumption are finite point processes on Y. It leads to the factorization (42) below. The third assumption says that the target-originated processes have at most one point.

Bivariate PGFL for the PHD Filter

Assumptions 1–3 lead directly to the factored form (44) of the PGFL of the conditional process. The measurement process Υ is the superposition of a clutter process $\Upsilon_{\text{clutter}}$ and an unknown number of identical target-originated measurement processes Υ_{target}. Conditioned on $\Xi = \xi = (n, \{s_1, \ldots, s_n\})$, there are $N = n$ target-originated measurement processes. These n processes and the clutter process are conditionally independent, so the PGFL of the conditional process $\Upsilon \mid \Xi$ is the product of their individual PGFLs. Thus, for real-valued functions g defined on Y,

$$G^{\Upsilon \mid \Xi}[g \mid s_1, \ldots, s_n] = G^{\Upsilon_{\text{clutter}}}[g] \prod_{i=1}^{n} G^{\Upsilon_{\text{target}}}[g \mid s_i] \tag{42}$$

where the product is taken equal to one if $n = 0$.

The form of $G^{\Upsilon_{\text{target}}}[g \mid s]$ is derived from the assumption that a target at $s \in S$ generates at most one measurement. It is also assumed that a target at s is detected with probability $P^D(s)$ and not detected with probability $1 - P^D(s)$. Then, using (1) directly, the PGFL of target-originated measurement is a two term sum,

$$G^{\Upsilon_{\text{target}}}[g \mid s] = 1 - P^D(s) + P^D(s) \int_Y g(y) p(y \mid s) dy \tag{43}$$

where $p(y \mid s)$ is the (assumed known) sensor pdf of the point measurement $y \in Y$ given a target at $s \in S$. The clutter PGFL is a PPP on Y with intensity function $\lambda(y)$, so its PGFL is the same form as (2). The measurement set is the superposition of independent processes, by Assumption 2. Substituting (43) and the clutter PGFL into (42) gives

$$\begin{aligned} G^{\Upsilon|\Xi}&[g \mid s_1, \ldots, s_n] \\ &= \exp\left[-\int_Y \lambda(y) dy + \int_Y g(y)\lambda(y) dy\right] \\ &\quad \times \prod_{i=1}^{n} \left(1 - P^D(s_i) + P^D(s_i) \int_Y g(y) p(y \mid s_i) dy\right). \end{aligned} \tag{44}$$

In words, (44) says that the measurement set is the outcome of an insertion/deletion process, or "indel" process for short, because PPP clutter is randomly inserted and target-originated measurements are randomly deleted.

Explicit expressions for the pdf of the conditional process $\Upsilon \mid \Xi$ are not needed. These expressions are, however, very insightful because to write them down is to enumerate the measurement to target assignments. Examples are given in the Appendix.

Substituting (42) into (34) gives

$$\begin{aligned} G^{\Upsilon\Xi}[g,h] = G^{\Upsilon_{\text{clutter}}}[g] \sum_{n=0}^{\infty} p_N^{\Xi}(n) \int_{S^n} \prod_{i=1}^{n} \left(h(s_i) G^{\Upsilon_{\text{target}}}[g \mid s_i]\right) \\ \times p_{X|N}^{\Xi}(s_1, \ldots, s_n \mid n) ds_1 \cdots ds_n. \end{aligned} \tag{45}$$

The sum-integral in (45) is equal to the PGFL of the target process Ξ evaluated at the function $h(s) G^{\Upsilon_{\text{target}}}[g \mid s]$. By assumption, the PGFL of Ξ is a PPP with intensity f^{Ξ} and its PGFL is given by (2). Substituting $h(s) G^{\Upsilon_{\text{target}}}[g \mid s]$ into (2), and then substituting the PGFLs for target (43) and for PPP clutter, yields the PGFL of the joint process:

$$\begin{aligned} G^{\Upsilon\Xi}[g,h] = \exp\Bigg[&-\int_Y \lambda(y) dy + \int_Y g(y)\lambda(y) dy \\ &- \int_S f^{\Xi}(s) ds + \int_S h(s) f^{\Xi}(s) ds \\ &- \int_S h(s) P^D(s) f^{\Xi}(s) ds \\ &+ \int_S \int_Y g(y) h(s) p(y \mid s) P^D(s) f^{\Xi}(s) dy\, ds \Bigg]. \end{aligned} \tag{46}$$

Except for clutter, this PGFL is a special case of the general result (38).

First Summary Statistic—the Target Intensity Function

The intensity function of the Bayes posterior process $\Xi \mid \Upsilon = \upsilon = (m, y_1, \ldots, y_m)$ is obtained by substituting the functional derivatives of (46) with respect to impulses at the measurement points $y_1, \ldots, y_m$ into (29). The derivative of $G^{\Upsilon\Xi}[g,h]$ with respect to an impulse at $y_1 \in Y$ is

$$\begin{aligned} \frac{\partial G^{\Upsilon\Xi}}{\partial y_1}&[g,h] \\ &= G^{\Upsilon\Xi}[g,h] \left\{\lambda(y_1) + \int_S h(s) p(y_1 \mid s) P^D(s) f^{\Xi}(s) ds\right\}. \end{aligned} \tag{47}$$

The term in braces in (47) does not depend on $g(y)$, so its functional derivative with respect to an impulse at $y \neq y_1$ is zero. Thus, for distinct impulses at $y_1, \ldots, y_m$ in Y,

$$\begin{aligned} \frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}&[g,h] = \\ & G^{\Upsilon\Xi}[g,h] \prod_{i=1}^{m} \left(\lambda(y_i) + \int_S h(s) p(y_i \mid s) P^D(s) f^{\Xi}(s) ds\right). \end{aligned} \tag{48}$$

The functional derivative of (48) with respect to an impulse at $x \in S$ is

$$\begin{aligned} \frac{\partial^{m+1} G^{\Upsilon\Xi}}{\partial x \partial y_1 \cdots \partial y_m}[g,h] &= G^{\Upsilon\Xi}[g,h] f^{\Xi}(x) \left(1 - P^D(x) + P^D(x) \int_Y g(y) p(y \mid x) dy\right) \\ &\quad \times \prod_{i=1}^{m} \left(\lambda(y_i) + \int_S h(s) p(y_i \mid s) P^D(s) f^{\Xi}(s) ds\right) \\ &\quad + G^{\Upsilon\Xi}[g,h] P^D(x) f^{\Xi}(x) \sum_{i=1}^{m} p(y_i \mid x) \prod_{k=1, k\neq i}^{m} \left(\lambda(y_k) + \int_S h(s) p(y_k \mid s) P^D(s) f^{\Xi}(s) ds\right). \end{aligned} \tag{49}$$

Substituting $g(y) = 0$ and $h(x) = 1$ gives the unconditional pdf

$$p^{\Upsilon}(\upsilon) = \frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,1] = G^{\Upsilon\Xi}[0,1] \prod_{i=1}^{m} \left(\lambda(y_i) + \int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds \right) \tag{50}$$

and

$$\frac{\partial^{m+1} G^{\Upsilon\Xi}}{\partial x \partial y_1 \cdots \partial y_m}[0,1] = G^{\Upsilon\Xi}[0,1] f^{\Xi}(x) \left\{ (1 - P^D(x)) \prod_{i=1}^{m} \left(\lambda(y_i) + \int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds \right) + P^D(x) \sum_{i=1}^{m} p(y_i \mid x) \prod_{k=1, k\neq i}^{m} \left(\lambda(y_k) + \int_S p(y_k \mid s) P^D(s) f^{\Xi}(s) ds \right) \right\}. \tag{51}$$

Substituting (51) and (50) into (29) gives the intensity function, that is, the PHD:

$$f_{\mathrm{PHD}}^{\Xi|\Upsilon}(x) = f^{\Xi}(x) \left[1 - P^D(x) + P^D(x) \sum_{i=1}^{m} \frac{p(y_i \mid x)}{\lambda(y_i) + \int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds} \right]. \tag{52}$$

The expected number of targets in S is

$$\hat{N}_{\mathrm{PHD}}(S) = \int_S f_{\mathrm{PHD}}^{\Xi|\Upsilon}(x) dx. \tag{53}$$

The number $\hat{N}_{\mathrm{PHD}}(S)$ is an expectation over the grand canonical ensemble and is typically not an integer.

Second Summary Statistic—Distribution of Target Canonical Number

The Bayes posterior target process $\Xi \mid (\Upsilon = \upsilon)$ is a PPP only for $m = 0$. This follows from the form of the PGF of $N^{\Xi|\Upsilon}$,

$$F_{\mathrm{PHD}}^{\Xi|\Upsilon}(x) = \exp\left[(x-1) \int_S (1 - P^D(s)) f^{\Xi}(s) ds \right] \times \prod_{i=1}^{m} \frac{\lambda(y_i) + x \int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds}{\lambda(y_i) + \int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds} \tag{54}$$

where the product is equal to one for $m = 0$. This expression is obtained by substituting (48) into the general result (30). The PGF (54) is the product of two PGFs. One is the PGF of the number of clutter points, which is Poisson distributed with mean $\int_S (1 - P^D(s)) f^{\Xi}(s) ds$. The other is the PGF of the number of heads, i.e., targets detected, in a coin tossing experiment. The experiment uses m non-identical coins, each tossed only once, where the probability of a target detection for the ith coin is

$$\Pr[\text{target detection} \mid \text{coin } i] = \frac{\int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds}{\lambda(y_i) + \int_S p(y_i \mid s) P^D(s) f^{\Xi}(s) ds}. \tag{55}$$

The distribution $p_N^{\Xi|\Upsilon}(n)$ of canonical number is obtained by differentiating $F^{\Xi|\Upsilon}(x)$ with respect to x. The mean number of targets is the sum of the means of the factors in (54), and this is clearly identical to (53).

The PGF (54) of the Bayes canonical number distribution is an immediate consequence of the connection between PGFs and PGFLs. Nonetheless, it appears to be new to the PHD literature.

The PHD filter approximates the Bayes posterior process by a PPP with intensity (52). The PGF of the Bayes posterior distribution of canonical number is approximated by the Poisson distribution whose mean is (53). The mean canonical numbers of the Bayes posterior process and the PPP approximation are equal, but the distributions are mismatched.

Post-processing decision procedures estimate the actual number of targets, decide which measurements correspond to targets and which to clutter, and compute point estimates and areas of uncertainty (AOUs) for detected targets. The point estimates are interpreted as pseudo-MAP estimates of target states, as discussed in Section 3, and the AOUs are surrogates for error covariance matrices. These important topics are outside the scope of the present paper.

5.2. Endogenous Scattering—the iFilter

An endogenous measurement model is a model of the spatial distribution of scatterers. It makes no distinction between scatterers that are targets and those that are clutter; such distinctions are relegated to a post-processing classification decision procedure. The predicted target process Ξ is defined on S^+, and its intensity is given by (40).

To compare the endogenous measurement model to the standard exogenous model, interpret a scatterer with state $s \in S$ as a target in the same state. Scatterers whose state is ϕ are clutter-generators in the exogenous model. Because Ξ is a PPP on S^+, more than one point in the realization $\xi = (n, \{s_1, \ldots, s_n\})$ can be equal to ϕ. Thus, the model accommodates a variable number of clutter points by varying the number of scatterers with state ϕ. Assumptions 1–3 above are modified as follows:

1′. The scattering process Ξ is a PPP on $S^+ = S \cup \phi$ with intensity function $f^{\Xi}(s)$, $s \in S^+$, where the state ϕ is assigned to scatterers whose state $s \in S$ is unknown.

2′. Conditioned on the scattering event $\Xi = \xi = (n, \{s_1, \ldots, s_n\})$, the measurement process is the superposition of n mutually independent identical scatterer-originated measurement processes.

3′. No scatterer generates more than one measurement in the space Y.

The third assumption says that a scatterer-originated process generates at most one measurement regardless of the scatterer state. It leads to the two term expression (57) below. The scatterer-originated measurement processes in the second assumption are finite point processes on Y. It leads to the factorization (58) below.

Bivariate PGFL for the iFilter

The Markovian transition model that determines the predicted intensity on S^+ is an essential ingredient of the ability of the iFilter to distinguish between scatterers in state $s \in S$ and those in state ϕ. Transition models on S^+ incorporate within themselves models for initiation and termination of tracks; however, the details are not discussed in this paper to focus attention on the PGFL aspects of the point process theory. It is also necessary that the temporal stability of measurements from scatterers in state $s \in S$ (i.e., the targets) be greater than that of measurements from scatterers in state ϕ (i.e., the clutter). This depends on the character of the data. For further discussion of iFilters and an application to real data, see [12] and [13].

The sensor likelihood function $p(y \mid s)$, detection probability $P^D(s)$, and intensity function $f^{\Xi}(s)$ are extended from S to S^+. The density $p(y \mid \phi)$ is interpreted as the pdf of y given that it arises from a scatterer with state ϕ; the probability of detection $P^D(\phi)$ is the probability that a scatterer with state ϕ generates a measurement; and the intensity $f^{\Xi}(\phi)$ is the expected number of scatterers with state ϕ. The number $f^{\Xi}(\phi)$ is dimensionless. Integrals over S^+ are defined for real-valued functions $h(s)$, $s \in S^+$, by

$$\int_{S^+} h(s)ds \equiv h(\phi) + \int_S h(s)ds. \tag{56}$$

This definition is used in the PGFL. Functional derivatives extend to the space S^+ by defining the Dirac delta function so that $\delta_\phi(\phi) = 1$ and $\delta_x(\phi) = \delta_\phi(s) = 0$ for $s, x \in S$. Repeated functional derivatives with respect to impulses at ϕ are used to show that the PGFL characterizes the point process; however, as will be seen, the iFilter requires only one such derivative. Further details are straightforward and are omitted.

Detected scatterers contribute measurements to the measurement set, but undetected scatterers do not. The PGFL of a scatterer is

$$G^{\Upsilon_{\text{scatter}}}[g \mid s] = 1 - P^D(s) + P^D(s)\int_Y g(y)p(y \mid s)dy, \quad s \in S^+ \tag{57}$$

an expression identical to (43) except that it holds on S^+. The PGFL of the measurement set is, from the conditional independence assumptions,

$$G^{\Upsilon \mid \Xi}[g \mid s_1, \ldots, s_n] = \prod_{i=1}^{n}\left(1 - P^D(s_i) + P^D(s_i)\int_Y g(y)p(y \mid s_i)dy\right). \tag{58}$$

A separate clutter model is not used because such points are modeled as scatterers whose state is $\phi \in S^+$. The joint PGFL is obtained using the conditional process (58) in the same manner as before (see (45)). The result is

$$\begin{aligned} G^{\Upsilon\Xi}[g,h] = \exp\Bigg[&-\int_{S^+} f^{\Xi}(s)ds + \int_{S^+} h(s)f^{\Xi}(s)ds \\ &- \int_{S^+} h(s)P^D(s)f^{\Xi}(s)ds \\ &+ \int_{S^+}\int_Y g(y)h(s)p(y \mid s)P^D(s)f^{\Xi}(s)dy\,ds\Bigg]. \end{aligned} \tag{59}$$

The PGFLs (59) and (46) fully characterize the similarities and differences between the scattering and clutter models, respectively.

First Summary Statistic—the Scatterer Intensity Function

The derivatives are

$$\frac{\partial^m G^{\Upsilon\Xi}}{\partial y_1 \cdots \partial y_m}[0,h] = G^{\Upsilon\Xi}[0,h]\prod_{i=1}^{m}\int_{S^+} h(s)p(y_i \mid s)P^D(s)f^{\Xi}(s)ds \tag{60}$$

and

$$\begin{aligned} \frac{\partial^{m+1} G^{\Upsilon\Xi}}{\partial x \partial y_1 \cdots \partial y_m}[0,h] &= G^{\Upsilon\Xi}[0,h]f^{\Xi}(x)(1 - P^D(x)) \\ &\quad \times \prod_{i=1}^{m}\int_{S^+} h(s)p(y_i \mid s)P^D(s)f^{\Xi}(s)ds \\ &\quad + G^{\Upsilon\Xi}[0,h]P^D(x)f^{\Xi}(x)\sum_{i=1}^{m} p(y_i \mid x) \\ &\quad \times \prod_{k=1,k\neq i}^{m}\int_{S^+} h(s)p(y_k \mid s)P^D(s)f^{\Xi}(s)ds. \end{aligned} \tag{61}$$

The intensity function of the iFilter is the ratio of (61) to (60) evaluated at $h(s) = 1$:

$$
\begin{aligned}
f^{\Xi|\Upsilon}_{\mathrm{iFilter}}(x) &= f^{\Xi}(x)\left[1 - P^D(x) + P^D(x)\sum_{i=1}^{m}\frac{p(y_i \mid x)}{\int_{S^+} p(y_i \mid s)P^D(s)f^{\Xi}(s)ds}\right] \\
&= f^{\Xi}(x)\left[1 - P^D(x) + P^D(x)\sum_{i=1}^{m}\frac{p(y_i \mid x)}{\hat{\lambda}(y_i) + \int_{S} p(y_i \mid s)P^D(s)f^{\Xi}(s)ds}\right]
\end{aligned} \tag{62}
$$

where, for any $y \in Y$,

$$
\hat{\lambda}(y) = p(y \mid \phi)P^D(\phi)f^{\Xi}(\phi) \tag{63}
$$

is the estimated measurement intensity at $y \in Y$ due to scatterers with state ϕ. Since (62) holds for $x \in S^+$, the updated intensity $f^{\Xi|\Upsilon}_{\mathrm{iFilter}}(\phi)$ is (62) evaluated for $x = \phi$.

The likelihood function can be chosen so that the posterior intensity $\hat{\lambda}(y)$ matches the specified exogenous clutter intensity $\lambda(y)$. In addition to the parameter choices made in (41), let the detection probability be $P^D(\phi) = 1$ and define $p(y \mid \phi) = \lambda(y)/\lambda$. From (41), the predicted intensity is $f^{\Xi}(\phi) = \Psi(\phi \mid \phi)f^{\Xi^-}(\phi) = \lambda$, so that $\hat{\lambda}(y) = \lambda(y)$.

Second Summary Statistic—the Scatterer Canonical Distribution

The canonical number is the number of scatterers in all of S^+. Note that the count necessarily includes scatterers with state ϕ. The PGF of the canonical number is, using (60) and (30),

$$
\begin{aligned}
F^{\Xi|\Upsilon}_{\mathrm{iFilter}}(x) &= x^m \exp\left[(x-1)\int_{S^+}(1 - P^D(s))f^{\Xi}(s)ds\right] \\
&= x^m \exp\left[(x-1)(1 - P^D(\phi))f^{\Xi}(\phi)\right] \\
&\quad \times \exp\left[(x-1)\int_{S}(1 - P^D(s))f^{\Xi}(s)ds\right].
\end{aligned} \tag{64}
$$

This PGF is the product of the PGFs of three mutually independent scattering processes. One is due to the endogenous measurement model and generates exactly m scatterers. The others are Poisson distributed and correspond to scatterers in state ϕ and to scatterers with states in S, with expected canonical numbers $(1 - P^D(\phi))f^{\Xi}(\phi)$ and $\int_S (1 - P^D(s))f^{\Xi}(s)ds$, respectively. The latter PGF is the first factor in (54).

The expected number of scatterers in the Bayes posterior process is

$$
\hat{N}_{\mathrm{iFilter}}(S^+) = \int_{S^+} f^{\Xi|\Upsilon}_{\mathrm{iFilter}}(x)dx = f^{\Xi|\Upsilon}_{\mathrm{iFilter}}(\phi) + \hat{N}_{\mathrm{iFilter}}(S) \tag{65}
$$

where

$$
\hat{N}_{\mathrm{iFilter}}(S) = \int_{S} f^{\Xi|\Upsilon}_{\mathrm{iFilter}}(x)dx \tag{66}
$$

is the expected number of scatterers with states in S. The iFilter estimate (66) is similar to the PHD estimate (53). The distribution of the number of scatterers in S (i.e., targets) is conditioned on the number of scatterers with state ϕ. This topic is the subject of on-going work [2] and is outside the scope of the present paper. Practical experience to date [12, 13] shows that the iFilter has excellent performance.

An alternative derivation of the iFilter can be based on the PGFLs of the detected and undetected scatterer processes. These processes are thinned versions of the parent process Ξ, where the thinning function is the probability of detection $P^D(s)$. Under the PPP assumption for Ξ, they are also mutually independent, not conditionally independent. The derivation is similar to the one just given and is omitted.

6. CONCLUDING REMARKS

Finite point process models are excellent models for sensor measurement sets in many traditional applications involving point targets whose measurements are superimposed with clutter. In contrast, they are only approximate models for multitarget state. Accepting the point process model for the multitarget state as a given, the PHD filter and iFilter are good applications of the PGFL approach.

The PGFL approach is seen to provide a common language to clarify the similarities and differences between the clutter and target models used in the PHD filter and the iFilter. Approximations and other issues seem to preclude using PGFLs directly as a basis for comparing tracking performance. In any event, such comparisons are outside the scope of the present paper.

Both the exogenous clutter model and the endogenous scattering model lead to enumerations of measurements to targets. Although the technical details differ somewhat between the PHD filter and iFilter, manipulating the required enumerations is facilitated by the PGFL approach.

By-passing explicit enumerations in the Bayes posterior process can sometimes obscure salient features of the problem, features that can make direct methods worthwhile. For example, the iFilter is derived in [15] for the scattering model by a direct Bayesian method without resorting to the PGFL. This alternative derivation—not given here—illuminates several aspects of the PGFL method and how it works.

Enumerations of measurement to target assignments are encoded in the PGFL of the measurement-target process. An excellent example is the way in which the PGFL of the general Bayes posterior process is written as a ratio of functional derivatives of the joint PGFL.

The functional derivatives of the PGFL of the Bayes posterior process decode the probability structure. When the functional derivatives are of high computational complexity, the utility of the PGFL approach is severely limited. Examples of this limitation—not discussed in this paper—are extended targets and the target-centric multisensor PHD filter and iFilter.

Finally, the PGFL approach may suggest alternative problems of independent interest. One example is a traffic process [18] that counts sensor target detections, not the targets themselves. Its computational complexity is linear in the number of sensors.

APPENDIX. ASSIGNMENT ENUMERATION IN THE PHD FILTER

For $n = 0$, no targets are present and (44) reduces to the PGFL of clutter. When the sensor reports no measurements, $\upsilon = (0,\emptyset)$ and

$$p^{\Upsilon|\Xi}(\upsilon = (0,\emptyset) \mid \xi = (n,\{s_1,\ldots,s_n\})) = G^{\Upsilon|\Xi}[0 \mid \xi] = e^{-\int_Y \lambda(y)dy} \prod_{i=1}^{n} (1 - P^D(s_i)). \tag{67}$$

This is the probability that the realizations of the clutter process and the n target processes contain no points.

Explicit expressions of $p^{\Upsilon|\Xi}(\upsilon \mid \xi)$ for events $\upsilon = (m,\{y_1,\ldots,y_m\})$, $m \geq 1$, are found by functional differentiation of (44) with respect to impulses at the (distinct) points $y_1,\ldots,y_m$. Functional differentiation masks enumerations of measurement to target. For $m = 1$,

$$\begin{aligned} p^{\Upsilon|\Xi}(\upsilon = (1,\{y_1\}) \mid \xi = (n,\{s_1,\ldots,s_n\})) &= \frac{\partial G^{\Upsilon|\Xi}}{\partial y_1}[0 \mid \xi] \\ &= e^{-\int_Y \lambda(y)dy} \lambda(y_1) \prod_{i=1}^{n} (1 - P^D(s_i)) \\ &\quad + e^{-\int_Y \lambda(y)dy} \sum_{i=1}^{n} P^D(s_i) p(y_1 \mid s_i) \prod_{k=1,k\neq i}^{n} (1 - P^D(s_i)). \end{aligned} \tag{68}$$

In words, (68) is the probability that y_1 is produced either by 1) the clutter process and all n target processes are undetected or by 2) exactly one of the n target processes and there is no clutter.

The functional derivatives of the PGFL for $m \geq 2$ are expressions that sum over all assignments of m measurements to n targets that are consistent with the constraint of "at most one measurement per target per scan." Details are omitted.

REFERENCES

[1] K. B. Athreya and P. E. Ney
Branching Processes.
Berlin: Springer-Verlag, 1972 (Republished by Mineola, NY: Dover, 2004).

[2] Ö. Bozdogan, R. Streit, and M. Efe
Palm intensity for track extraction.
In preparation.

[3] X. Chen, R. Tharamrasa, T. Kirubarajan, and M. Pelletier
Integrated clutter estimation and target tracking using Poisson point process.
SPIE vol. 7445, 2009.

[4] D. J. Daley and D. Vere-Jones
An Introduction to the Theory of Point Processes, Vol. I: Elementary Theory and Methods.
NY: Springer, 1988 (Second Ed., 2003).

[5] D. J. Daley and D. Vere-Jones
An Introduction to the Theory of Point Processes, Vol. II: General Theory and Structure.
NY: Springer, 1988 (Second Ed., 2008).

[6] T. E. Harris
The Theory of Branching Processes.
Berlin: Springer-Verlag, 1963.

[7] G. E. Kopec
Formant tracking using hidden Markov models and vector quantization.
IEEE Transactions on Acoustics, Speech, and Signal Processing, 34 (1986), 709–729.

[8] R. P. S. Mahler
A theoretical foundation for the Stein-Winter probability hypothesis density (PHD) multitarget tracking approach.
In *Proceedings of the MSS National Symposium on Sensor and Data Fusion*, vol. 1, 2000, 99–117.

[9] R. P. S. Mahler
Multitarget Bayes filtering via first-order multitarget moments.
IEEE Transactions Aerospace and Electronic Systems, 39 (2003), 1152–1178.

[10] R. P. S. Mahler
Statistical Multisource-Multitarget Information Fusion.
Boston: Artech House, 2007.

[11] J. E. Moyal
The general theory of stochastic population processes.
Acta Mathematica, 108 (1962), 1–31.

[12] M. Schikora, W. Koch, R. Streit, and D. Cremers
Sequential Monte Carlo method for the iFilter.
In *Proceedings of the Fourteenth International Conference on Information Fusion*, June 2011.

[13] M. Schikora, W. Koch, R. Streit, and D. Cremers
A sequential Monte Carlo method for multi-target tracking with the intensity filter.
Chapter 3 in: *Advances in Intelligent Signal Processing and Data Mining: Theory and Applications*, P. Georgieva, L. Mihaylova, L. Jain (Eds.): Springer, 2012.

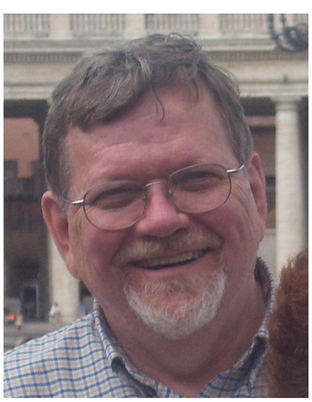
[14] M. C. Stein and C. L. Winter
An additive theory of probabilistic evidence accrual.
Los Alamos National Laboratory, Report LA-UR-93-3336, 1993.

[15] R. L. Streit and L. D. Stone
Bayes derivation of multitarget intensity filters.
In *Proceedings of the Eleventh International Conference on Information Fusion*, June 2008.

[16] R. L. Streit
Poisson Point Processes—Imaging, Tracking, and Sensing.
NY: Springer, 2010.

[17] R. L. Streit
PHD intensity filtering is one step of an EM algorithm for positron emission tomography.
In *Proceedings of the Twelfth International Conference on Information Fusion*, July 2009.

[18] R. L. Streit
Multisensor traffic mapping filters.
In *Proceedings of the Workshop on Sensor Data Fusion*, Sept. 4–6, 2012.



Roy Streit received a Ph.D. in mathematics in 1978 from the University of Rhode Island. He is currently a senior scientist at Metron in Reston, VA. His research interests include multi-target tracking, multi-sensor data fusion, distributed systems, medical imaging, and signal processing, as well as statistical methods for pharmacovigilance and business analytics. Prior to joining Metron in 2005, he was a senior scientist in the Senior Executive Service at the Naval Undersea Warfare Center in Newport, RI, working primarily on the development, evaluation and application of multi-sensor data fusion algorithms in support of submarine sonar and combat control automation. He was an exchange scientist at the Defence Science and Technology Organisation (DSTO) in Adelaide, Australia, from 1987–1989. In 1999, he received the Solberg Award from the American Society of Naval Engineers for contributions to naval engineering through personal research, and in 2001 the Department of the Navy Superior Civilian Achievement Award. From 1996 to 2005 he served on the Sonar Technology Panel (Panel 9) of The Technical Cooperation Program (TTCP), a multinational governmental organization supporting scientific information exchange between member nations. He was President of the International Society for Information Fusion (ISIF) in 2012, and he continues to serve on the ISIF Board of Directors.

Dr. Streit is a Fellow of the IEEE. He is the author of *Poisson Point Processes: Imaging, Tracking, and Sensing*, published by Springer in 2010. It was translated into Chinese and published by Science Press, Beijing, in 2013. He has published papers in over a dozen refereed technical journals, and given numerous invited and contributed papers at international conferences and workshops. He holds nine U.S. patents. He is a professor (adjunct) in the Department of Electrical and Computer Engineering at the University of Massachusetts–Dartmouth. He was a visiting scientist at Yale University from 1982–1984, and a visiting scholar at Stanford University from 1981–1982. He is currently an adjunct affiliate consultant with SUCCEED Educational Consultancy based in Rhode Island.