Basic concepts for multi-object estimation

Lecture notes

Daniel Clark, Emmanuel Delande[1], Jérémie Houssineau
Heriot-Watt University
{D.E.Clark, E.D.Delande, J.Houssineau}@hw.ac.uk

v.1 (July 5, 2016)

[1]Corresponding author, email: E.D.Delande@hw.ac.uk

Motivation

These lecture notes present fundamental concepts in point process theory for multi-object estimation problems, and includes practical derivation tools for the derivation of multi-target detection and tracking filters. Even though these notes aims at being as self-contained as possible, the reader is expected to have basic knowledge in probability theory. Some parts, notably in Chap. 3, may require additional knowledge in measure theory. In order to keep a natural flow in the development of the arguments exposed in these notes, simplicity is sometimes favoured over strict mathematical rigour in the presentation of some advanced concepts, notably pertaining to measure theory. Fortunately, the available literature on point processes propose some excellent books covering the topic in deeper details; some of them are provided in the next section.

A few useful references

A comprehensive study on point processes is given by Daley and Vere-Jones in [6, 7], digging deep in measure theory to present all the fundamental concepts related to point processes and many useful applications. Stoyan, Kendall, and Mecke follow a different approach in [14], casting the point processes in a more intuitive but perhaps less mathematically-involved framework, and provide an excellent complement to [6, 7]. Fundamental concepts in measure theory can be found in Bogachev's [2, 3], and their exploitation in the context of multi-object filtering is covered in more details in the authors' notes from the First International School on Finite Set Statistics [10], from which these lecture notes are inspired.

The exploitation of point processes for practical target tracking applications is to the credit of Goodman, Mahler, and Nguyen in [9], and Mahler in [13], where the Finite Set Statistics (FISST) framework is presented in detail. Mahler's seminal papers on the Probability Hypothesis Density (PHD) [11] and Cardinalized Probability Hypothesis Density (CPHD) [12] filters paved the way for most of the subsequent developments in multi-object filtering.

Introduction

In the context of multi-target tracking, multi-object estimation problems are the study of a population of objects or *targets*, whose number and individual states (e.g. position, velocity coordinates) are unknown. Cast in a Bayesian framework, the multi-object filters aim at describing the uncertainty on this population through a probabilistic description, and update that description across time whenever additional information on the population of targets are available – typically, through observations collected from some sensor system observing the surveillance scene.

A *point process* is a random variable whose realizations are sequences whose size and elements are both random; it is thus particularily adapted to the description of a *multi-object configuration*, i.e., a number of objects *and* their respective states. To a certain extent, a point process can be seen as the extension of an integer-valued random variable, describing the size of a population of objects, to a random variable describing the size *and* the states of a population of objects.

This key remark motivated the two-step pedagogical approach followed by these lecture notes, presenting first integer-valued random variables in Chap 1, and then point processes in Chap 2. The concepts pertaining to integer-valued random variables have (almost) always a counterpart for point processes, and the structures of Chaps 1, 2 present many remarkable similarities.

Organization of the lecture notes

The lecture notes are organized in three main chapters of increasing complexity:

- Chap. 1 serves as an introduction, and describes the study of a integer-valued random variable through its probability generating function (p.g.f.). The exploitation of integer-valued random variables is illustrated through the modelling and derivation of the "cardinality only" PHD filter.
- Chap. 2 contains the core notions presented in the lecture notes, and describes the study of a point process through its probability generating functional (p.g.fl.). The exploitation of point processes is illustrated through the modelling and derivation of the PHD filter [11].
- Chap. 3 explores the construction and exploitation of higher-order moments for point processes, and illustrates the concept for the Poisson point process.

Finally, a few exercises relating to the three chapters above are proposed in Chap. 4.

Note

Most of the recent developments in multi-object filtering, following the terminology proposed in the FISST methodology [12] pertaining to Random Finite Sets (RFSs), make use of *sets of points*, *multi-object densities*, and *set integrals*. The general terminology pertaining to point processes, on the other hand, make use of *sequences of points*, *probability measures*, *probability densities*, and *measure-theoretical integrals*.

These notions are largely equivalent, as a RFS can be seen as a (simple) point process (see Chap. 2). However the expression of higher-order moments for point processes, presented in Chap. 3, requires the construction of quantities described with *measures* but admitting no *densities*, and more easily described with *measure-theoretical* integrals than *set* integrals. For this reason, these lectures notes follow the general terminology of point processes.

Chapter 1

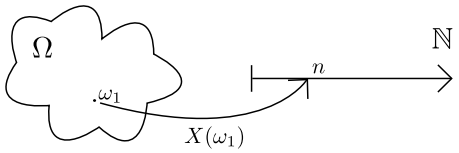
Integer-valued random variables

In this chapter, we shall focus on the estimation of the *number* of targets in the surveillance scene, and *not* their individual states.

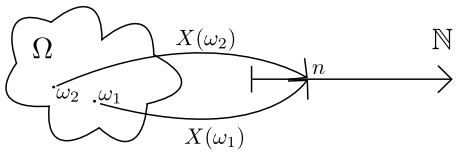
1.1 Integer-valued random variables: basic concepts

The number of targets in the scene is obviously an *integer*, but it is *unknown*; thus, it is aptly described by an *integer-valued random variable* X.

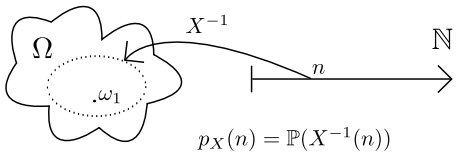
The random variable X is a mapping from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the set of non-negative integers $\mathbb{N}$.



Depending on the construction of the random variable X, several *outcomes* ω_i may be associated to the same *realization* k.



The quantity $X^{-1}(n)$ represents the collection of all the possible outcomes ω_i leading to the realization n. The probability space is endowed with a *probability measure* $\mathbb{P}$ which allows us to measure the "size" of $X^{-1}(n)$. The "larger" $X^{-1}(n)$ is, the more likely is the value n to be drawn when sampling from X.



The quantity

$$p_X(n) = \mathbb{P}(X^{-1}(n)) \tag{1.1}$$

denotes the likelihood that n is drawn when sampling from X, which we shall describe as the event $X = n$. The structure of the probability space is such that

$$\sum_{n\geq 0} p_X(n) = \sum_{n\geq 0} \mathbb{P}(X^{-1}(n)) \tag{1.2a}$$

$$= 1, \tag{1.2b}$$

which ensures that the elements $p_X(n)$ can be readily intepreted as cardinality *probabilities*, and the family $\{p_X(n)\}_{n\geq 0}$ as a cardinality *distribution*. In our context, $p_X(n)$ is the probability that the population described by X has *exactly* n objects.

The cardinality distribution $\{p_X(n)\}_{n\geq 0}$ fully characterizes the random variable X, but its full knowledge is seldom available in practical problems as it may be intractable to estimate and propagate across time. Random variables can also be described by their *moments*, which provide a limited but meaningful description of the behaviour of X. Given an integer $k \geq 0$, the kth order *non factorial* (respectively (resp.) *factorial*) moment $\mu_X^{(k)}$ (resp. $\alpha_X^{(k)}$) of X are defined as

$$\mu_X^{(k)} = \mathbb{E}\left[X^k\right] = \sum_{n\geq 0} p_X(n) n^k, \tag{1.3}$$

$$\alpha_X^{(k)} = \mathbb{E}\left[X(X-1)\dots(X-k+1)\right] = \sum_{n\geq k} p_X(n) n(n-1)\dots(n-k+1). \tag{1.4}$$

The non factorial moments are useful for the construction of the *central moments* such as the variance

$$\mathrm{var}_X = \mu_X^{(2)} - \left(\mu_X^{(1)}\right)^2, \tag{1.5}$$

a well-known statistic which describes the spread of the values taken by X around its mean value $\mu_X^{(1)}$. Also, the correlation between two random variables X, Y can be studied through the covariance

$$\mathrm{cov}_{X,Y} = \mu_{XY}^{(1)} - \mu_X^{(1)}\mu_Y^{(1)}. \tag{1.6}$$

The factorial moments have no easy physical interpretation and are seldomly used to produce meaningful statistics on random variables. The exception is the first-order factorial moment, which equals the first-order non factorial moment $\mu_X^{(1)}$ and provides the *mean value* of X, usually noted μ_X:

$$\mu_X = \sum_{n\geq 0} p_X(n) n (= \mu_X^{(1)} = \alpha_X^{(1)}). \tag{1.7}$$

The cardinality distribution is a convenient tool to study a single given random variable. In the multi-object Bayesian framework, however, *different* random variables are used to describe the evolution of our knowledge of the *same* concept across time – for example, our knowledge on the number of targets in the scene is enriched when the sensor system produces new measurements, and the random variable describing the number of targets is updated accordingly. It turns out

that the transition between these random variables is difficult to describe through their cardinality distributions, and that another representation of the random variable is necessary to be able to produce the filtering equations effectively.

For example, suppose that the random variables X_1 and X_2 are fully known through their respective cardinality distributions $\{p_{X_1}(n)\}_{n\geq 0}$, $\{p_{X_2}(n)\}_{n\geq 0}$, and that the random variable X is defined as the sum

$$X = X_1 + X_2. \tag{1.8}$$

What is the cardinality distribution $\{p_X(n)\}_{n\geq 0}$ of X? One way to find out is to enumerate every possible realization n of X and consider all the possible joint realizations of $X_1,\ X_2$ whose sum equals n:

$$\begin{aligned}
p_X(0) &= p_{X_1}(0)p_{X_2}(0), && (1.9)\\
p_X(1) &= p_{X_1}(1)p_{X_2}(0) + p_{X_1}(0)p_{X_2}(1), && (1.10)\\
p_X(2) &= p_{X_1}(2)p_{X_2}(0) + p_{X_1}(1)p_{X_2}(1) + p_{X_1}(0)p_{X_2}(2), && (1.11)\\
&\dots
\end{aligned}$$

We see on the example above that a simple operation on random variables - the sum - does not translate into a simple operation on the cardinality probabilities.

Just as the Fourier transform allows us to shift the study of time-varying signals from the time to the frequency domain in which simple operations on signals are easily transcribed, one would like to shift the study of random variables from the cardinality probabilities to a more adequate domain.

1.2 Probability generating function: definitions

A generating function is a function $G : \mathbb{R}^+ \to \mathbb{R}$ which is built upon (or "generated by") a (possibly infinite) sequence of real numbers[1]. Given a sequence of real numbers $(u_n)_{n\geq 0}$, its generating function G is defined as

$$G(s) = \sum_{n\geq 0} u_n s^n \tag{1.12}$$

for any $s \in \mathbb{R}^+$ such that the sum on the right hand side of (1.12) is finite.

Applied to a random variable X, one can substitute the cardinality probabilities $\{p_X(n)\}_{n\geq 0}$ in (1.12) to produce the *p.g.f.* G_X of X, defined as the expectation

$$\begin{aligned}
G_X(s) &= \mathbb{E}\left[s^X\right] && (1.13a)\\
&= \sum_{n\geq 0} p_X(n)s^n. && (1.13b)
\end{aligned}$$

[1] $\mathbb{R}^+$ is the set of non negative real numbers. More general definitions of the generating function exist, but it is out of the scope of this lecture.

Note that using the cardinality probabibilities as the generating sequence imposes some restrictions on the range of admissible values for the *test variable* s, and the p.g.f. is defined for $0 \leq s \leq 1$. From (1.13) it is easy to see that:

$$G_X(0) = \sum_{n \geq 0} p_X(n) 0^n = p_X(0) \tag{1.14}$$

$$G_X(1) = \sum_{n \geq 0} p_X(n) 1^n = \sum_{n \geq 0} p_X(n) = 1, \tag{1.15}$$

Setting s to 0 in (1.13) allowed us to *extract* the cardinality probability $p_X(0)$ from the p.g.f.; we will see in Section 1.3 that other cardinality probabilities can be extracted through differentiation of the p.g.f..

In multi-object filtering applications it will be necessary to study the joint behaviour of several random variables; for example, to describe the number of measurements produced by the sensor system given the number of targets in the scene. The joint p.g.f. $\mathcal{G}_{Z,X}$ of two (possibly dependent) random variables Z, X is defined as the expectation

$$\mathcal{G}_{Z,X}(t,s) = \mathbb{E}\left[t^Z s^X\right] \tag{1.16a}$$

$$= \sum_{m,n \geq 0} p_{Z,X}(m,n) t^m s^n, \tag{1.16b}$$

where $p_{Z,X}(m,n)$ is the *joint probability* that $Z = m$ *and* $X = n$. Note that, if Z and X are independent variables, then by definition $p_{Z,X}(m,n) = p_Z(m)p_X(n)$ and the joint p.g.f. becomes:

$$\mathcal{G}_{Z,X}(t,s) = \sum_{m,n \geq 0} p_Z(m) p_X(n) t^m s^n \tag{1.17a}$$

$$= \left(\sum_{m \geq 0} p_Z(m) t^m\right)\left(\sum_{n \geq 0} p_X(n) s^n\right) \tag{1.17b}$$

$$= G_Z(t) G_X(s). \tag{1.17c}$$

We now need to introduce the notion of derivative to further exploit the p.g.f..

1.3 Ordinary differentiation

1.3.1 Definition and basic rules

As p.g.f.s are real-valued functions taking a real number as argument, the "classic" derivative can be applied to the p.g.f.s. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is some function, we call "the derivative of f (evaluated) at $x \in \mathbb{R}$", and denote it by $f'(x)$, the limit

$$f'(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}, \tag{1.18}$$

where $\epsilon \in \mathbb{R}$, if it exists. The ordinary derivative comes with a few calculus rules that will be useful for the differentiation of p.g.f.s. Suppose that f and g are admissible functions, then:

$$\begin{aligned}
\text{sum:} \quad & (f+g)'(x) = f'(x) + g'(x) && (1.19)\\
\text{product:} \quad & (f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x) && (1.20)\\
\text{power:} \quad & (f^m)'(x) = m f^{m-1}(x) f'(x) && (1.21)\\
\text{chain (or composition):} \quad & (f \circ g)'(x) = g'(x) f'(g(x)) && (1.22)
\end{aligned}$$

1.3.2 A few advanced rules

A very important function that is used extensively in the construction of filter is the exponential function. Indeed, a common approximation in the design of multi-object filters is to assume that some p.g.f. can be written as an exponential as it often leads to tractable and easily implementable filtering equations. The following "tricks" involving the exponential function will be used later on:

$$\begin{aligned}
\text{ordinary differentiation:} \quad & \exp'(x) = \exp(x) && (1.23)\\
\text{composition:} \quad & (\exp \circ f)'(x) = f'(x)(\exp \circ f)(x) && (1.24)\\
\text{Taylor expansion:} \quad & \exp(x) = \sum_{n \geq 0} \frac{\exp^{(n)}(0)}{n!} x^n = \sum_{n \geq 0} \frac{x^n}{n!} && (1.25)
\end{aligned}$$

1.4 p.g.f.s and differentiation

We shall now apply the ordinary differentiation to the p.g.f. to see what kind of information can be extracted from it. Suppose that X is a random variable with known p.g.f. G_X and one wish to determine the cardinality distribution $\{p_X(n)\}_{n \geq 0}$. Let us have a look at the successive derivatives of G_X:

$$\begin{aligned}
G_X(s) &= \sum_{n \geq 0} p_X(n) s^n, && (1.26)\\
G'_X(s) &= \sum_{n \geq 0} p_X(n) (s^n)' = \sum_{n \geq 1} p_X(n) n s^{n-1}, && (1.27)\\
G^{(2)}_X(s) &= \sum_{n \geq 1} p_X(n) n (s^{n-1})' = \sum_{n \geq 2} p_X(n) n(n-1) s^{n-2}, && (1.28)\\
&\dots\\
G^{(k)}_X(s) &= \sum_{n \geq k} p_X(n) \underbrace{n(n-1)\cdots(n-k+1)}_{\substack{= \frac{n(n-1)\cdots(n-k+1)(n-k)\cdots 1}{(n-k)\cdots 1} \\ = \frac{n!}{(n-k)!}}} s^{n-k}. && (1.29)
\end{aligned}$$

Now, if we set $s = 0$ or $s = 1$ in (1.29) we get

$$G_X^{(k)}(0) = \sum_{n \geq k} p_X(n) \frac{n!}{(n-k)!} 0^{n-k} = p_X(k) \frac{k!}{(k-k)!} = k! p_X(k), \tag{1.30}$$

$$G_X^{(k)}(1) = \sum_{n \geq k} p_X(n) n(n-1) \cdots (n-k+1) = \alpha_X^{(k)}. \tag{1.31}$$

We have thus shown that any cardinality probability and any factorial moment can be extracted from the p.g.f.. Since the cardinality distribution fully characterizes the random variable, it follows from (1.30) that the p.g.f. *does as well*. In other words, the knowledge of a p.g.f. G_X is sufficient to provide a full description of the associated random variable X. From (1.30) and (1.31) we can draw the practical extraction rules:

$$\begin{cases} \dfrac{G_X^{(k)}(0)}{k!} = p_X(k), \\ G'_X(1) = \alpha_X^{(1)} = \mu_X. \end{cases} \tag{1.32}$$

Joint p.g.f.s, of course, can be derivated as well. Suppose, for example, that one wish to describe the joint behaviour of some random variables Z, X in the specific case where $Z = m$ (it will be very useful in Section 1.7). This is described by the univariate p.g.f. $\mathcal{G}_{Z=m,X}(s)$, which can be extracted from the joint p.g.f. $\mathcal{G}_{Z,X}$ as follows:

$$\mathcal{G}_{Z=m,X}(s) = \sum_{n \geq 0} p_{Z,X}(m,n) s^n \tag{1.33a}$$

$$= \frac{1}{m!} \frac{\mathrm{d}^m}{\mathrm{d}t^m} \mathcal{G}_{Z,X}(t,s) \Big|_{t=0}, \tag{1.33b}$$

where (1.33a) is drawn from the definition of the p.g.f. (1.13) and (1.33b) is obtained with a similar reasoning as shown in (1.29) and (1.30). Note that the test variable with respect to (w.r.t.) which the joint p.g.f. is differentiated appears explicitly in (1.33b) to avoid ambiguity. If necessary, $\frac{\mathrm{d}^m}{\mathrm{d}t^m} \mathcal{G}_{Z,X}(t,s)\big|_{t=0}$ in (1.33b) can then be differentiated w.r.t. the test variable s to produce the joint cardinality probabilities $\{p_{Z,X}(m,n)\}_{m,n \geq 0}$.

1.5 Operations on p.g.f.s

We will now explore how some simple operations on random variables translate into operations on p.g.f.s, just as some simple operations on time-varying signals translate into simple operations on their Fourier transforms. We will consider three operations on random variables which are very useful to model physical mechanisms in multi-target problems as illustrated in Section 1.7.

1.5.1 Marginalization

Marginalization occurs when two random variables Z, X have a known joint behaviour and one wish to "isolate" the behaviour of one of the random variable, say Z. One must marginalize the

joint behaviour over X, i.e. "integrate" the joint cardinality probabilities over all the possible realizations of X since

$$\forall m \in \mathbb{N},\ p_Z(m) = \sum_{n \geq 0} p_{Z,X}(m,n). \tag{1.34}$$

Suppose that the joint behaviour is known through the joint p.g.f. $\mathcal{G}_{Z,X}$. Using the definition of the joint p.g.f. (1.16) we can write:

$$\begin{aligned}
\mathcal{G}_{Z,X}(t,1) &= \sum_{m,n \geq 0} p_{Z,X}(m,n) t^m 1^n && \text{(1.35a)}\\
&= \sum_{m \geq 0} \left(\sum_{n \geq 0} p_{Z,X}(m,n) \right) t^m && \text{(1.35b)}\\
&= \sum_{m \geq 0} p_Z(m) t^m && \text{(1.35c)}\\
&= G_Z(t). && \text{(1.35d)}
\end{aligned}$$

That is, the marginalization of a random variable easily translates into a very simple operation on the joint p.g.f.:

$$G_Z(t) = \mathcal{G}_{Z,X}(t,1). \tag{1.36}$$

1.5.2 Sum (or superposition)

Superposition occurs when one is not interested in the individual realizations of two *independent* random variables X and Y, but only in the sum of the two realizations. If we denote by Z the sum of random variables X, Y with known p.g.f.s G_X, G_Y, then Z is also a random variable; using the definition of the p.g.f. (1.13) yields

$$\begin{aligned}
G_Z(s) &= \mathbb{E}\left[s^Z\right] && \text{(1.37a)}\\
&= \mathbb{E}\left[s^{X+Y}\right] && \text{(1.37b)}\\
&= \mathbb{E}\left[s^X s^Y\right] && \text{(1.37c)}\\
&= \mathbb{E}\left[s^X\right] \mathbb{E}\left[s^Y\right] && \text{(1.37d)}\\
&= G_X(s) G_Y(s), && \text{(1.37e)}
\end{aligned}$$

where (1.37c) is equivalent to (1.37d) because X and Y are independent.

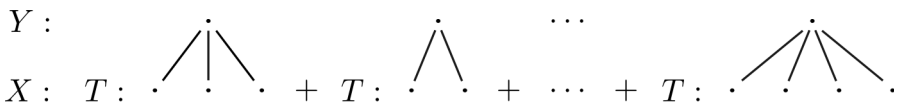
In other words, the sum of two *independent* random variables easily translates into the product of the associated p.g.f.s:

$$\mathcal{G}_{X+Y}(s) = G_X(s) G_Y(s). \tag{1.38}$$

1.5.3 Branching

Branching is a special kind of dependence between two random variables Y, X. Upon any realization m of the *parent random variable* Y, the *daughter random variable* X will be the superposition of m identical but independent random variables T, as if any object in the parent population

was "spawning" a number of objects in the daughter population following a common transition mechanism described by T.



Suppose that the parent random variable Y and the transition random variable T are known through the p.g.f.s G_Y, G_T, and that one wish to describe the daughter random variable X. The p.g.fl. describing the joint behaviour of the parent Y and daughter X random variables can be written as follows:

$$\mathcal{G}_{Y,X}(t,s) = \sum_{m,n\geq 0} p_{Y,X}(m,n)t^m s^n \tag{1.39a}$$

$$= \sum_{m,n\geq 0} p_Y(m)p_{X|Y}(n|m)t^m s^n \tag{1.39b}$$

$$= \sum_{m\geq 0} p_Y(m) \left(\sum_{n\geq 0} p_{X|Y}(n|m)s^n \right) t^m \tag{1.39c}$$

$$= \sum_{m\geq 0} p_Y(m)\mathcal{G}_{X|Y}(s|m)t^m, \tag{1.39d}$$

where $\mathcal{G}_{X|Y}(s|m)$ is the p.g.f. describing the daughter random variable X *conditioned* on the realization $Y = k$. If $Y = m$, then $X|Y$ is the superposition of m independent "copies" of the transition random variable T. Thus from (1.38) we have

$$\mathcal{G}_{X|Y}(s|m) = (G_T(s))^m. \tag{1.40}$$

Substituting (1.40) in (1.39d) gives

$$\mathcal{G}_{Y,X}(t,s) = \sum_{m\geq 0} p_Y(m)(G_T(s))^m t^m \tag{1.41a}$$

$$= \sum_{m\geq 0} p_Y(m)(tG_T(s))^m \tag{1.41b}$$

$$= G_Y(tG_T(s)). \tag{1.41c}$$

The result (1.41c) above is an important result which describes the joint behaviour of the parent and daughter random variables and that we shall use in Section 1.7. For now, since we are interested in the description of the daughter random variable X alone, we can simply marginalize this result over the parent random variable Y using (1.36) and we get

$$G_X(s) = \mathcal{G}_{Y,X}(1,s) \tag{1.42a}$$

$$= G_Y(G_T(s)). \tag{1.42b}$$

In other words, the branching of a parent random variable following a mechanism described by a transition random variable translates into the composition of the associated p.g.f.s.

1.6 A few examples of random variables and their p.g.f.s

We shall now present two specific classes of random variables which are often used in multi-object filtering and for which it is useful to learn beforehand the structure of the associated p.g.f.s.

1.6.1 Bernoulli random variable

A *Bernoulli* random variable X with parameter $0 \leq p \leq 1$ is a very simple integer-valued random variable defined as follows:

$$X = \begin{cases} 0, & \text{with probability } 1-p, \\ 1, & \text{with probability } p. \end{cases} \tag{1.43}$$

The construction of the p.g.f. G_X is straightforward using definition (1.13):

$$G_X(s) = \sum_{n \geq 0} p_X(n) s^n \tag{1.44a}$$

$$= \underbrace{p_X(0)}_{=1-p} + \underbrace{p_X(1)}_{=p} s + \sum_{n \geq 2} \underbrace{p_X(n)}_{=0} s^n \tag{1.44b}$$

$$= 1 - p + ps. \tag{1.44c}$$

The Bernoulli random variable is a "basic component" in the modelling of multi-object filters because it depicts the physical mechanisms of target survival and target detection (see Section 1.7 for more details).

1.6.2 Poisson random variable

A *Poisson* random variable X with rate $\lambda_X \geq 0$ is defined as follows:

$$\forall n \geq 0,\ X = n \quad \text{with probability } \exp(\lambda_X) \frac{\lambda_X^n}{n!}. \tag{1.45}$$

The construction of the p.g.f. G_X using definition (1.13) gives:

$$G_X(s) = \sum_{n \geq 0} p_X(n) s^n \tag{1.46a}$$

$$= \sum_{n \geq 0} \exp(-\lambda_X) \frac{\lambda_X^n}{n!} s^n \tag{1.46b}$$

$$= \exp(-\lambda_X) \sum_{n \geq 0} \frac{(\lambda_X s)^n}{n!} \tag{1.46c}$$

That is, using the Taylor expansion of the exponential (1.25):

$$G_X(s) = \exp(-\lambda_X) \exp(\lambda_X s) \tag{1.46d}$$

$$= \exp(\lambda_X (s-1)). \tag{1.46e}$$

It is formative to extract the mean value of a Poisson random variable using the differentiation of the p.g.f. (1.31):

$$\begin{aligned}
\mu_X &= G'_X(s)|_{s=1} && \text{(1.47a)}\\
&= (\exp(\lambda_X(s-1)))'|_{s=1} && \text{(1.47b)}\\
&= (\lambda_X(s-1))'\exp(\lambda_X(s-1))|_{s=1} && \text{(1.47c)}\\
&= \lambda_X \exp(\lambda_X(s-1))|_{s=1} && \text{(1.47d)}\\
&= \lambda_X \exp(\lambda_X(1-1)) && \text{(1.47e)}\\
&= \lambda_X. && \text{(1.47f)}
\end{aligned}$$

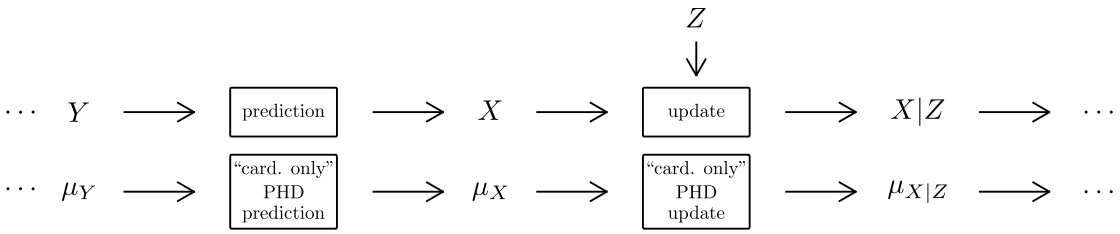
In other words, the mean value of Poisson random variable X equals its rate; since λ_X fully characterizes X through the definition (1.45), *so does the mean value* μ_X. Another important property of a Poisson random variable, left as exercise in Ex. 4.1.2, is that its variance var_X equals its mean μ_X.

Despite the simplicity of their structure, Poisson random variables provide a rather accurate description of a various number of natural phenomena (e.g. customer arrivals in queue lines). In multi-object filtering, Poisson random variables are appealing because of the exponential form of their p.g.f. (1.46e), easily differentiable; *assuming* some random variables to be Poisson allows the production of tractable and easily implementable filtering equations.

1.7 Application: the "cardinality only" PHD filter

We shall now apply the results we have seen in the previous sections to construct the "cardinality only" PHD filter. The purpose of this Bayesian filter is to estimate and propagate the mean number of target in the scene observed by some sensor with known characteristics. The modelling and filtering assumptions are identical to Mahler's PHD filter [11] – hence its name – and shall be detailed later. In other words, the "cardinality only" PHD filter can be seen as the reduction of the PHD filter to its *cardinality* component – we are interested in the *number* of targets *only*, *not* their state. A similar application for the "full" PHD filter will be the topic of Chap. 2.

The data flow of one iteration of the "cardinality only" PHD filter can be represented as follows:



where the random variables provide a description of the size of the following populations:

- Y: the targets before the prediction (prior knowledge from past iterations);
- X: the targets after the prediction;
- Z: the current measurements;
- $X|Z$: the targets after the data update (i.e. conditioned on some realization $Z = m$).

The construction of a filter follows two steps:

1. The *modelling* phase: we translate the physical phenomena of the tracking problem into relations between the random variables describing the populations of interest. In our case, we have to describe how to get X from Y, then how to get $X|Z$ from X. We have seen in Section 1.5 that operations on random variables are easily transcribed into operations on their p.g.f.s: for this reason, we will describe how to get G_X from G_Y, then how to get $\mathcal{G}_{X|Z}$ from G_X.
2. The *differentiation* phase: we extract the information that we wish to propagate from the appropriate differentiation of the p.g.f.s produced in the modelling phase. In our case, we have to describe how to get μ_X from μ_Y, and how to get $\mu_{X|Z}$ from μ_X.

The modelling phase relies on *modelling assumptions*, constituting a description of the physical phenomena that we wish to take into account and that we can afford to include in the design of the filter: it might be so, for example, that there is very slight chance that pairs of targets move in a correlated manner, but we may have to discard the modelling of correlated targets if the increasing complexity of the designed filter is not worth it and/or is unaffordable. Once completed, the modelling phase provides a full description of the sizes of the population of interest since random variables are completely described by their p.g.f.s (see Section 1.4). In other words, the modelling phase gives us *exactly what we are looking for* and the differentiation phase is, in theory at least, superfluous.

The differentiation phase aims at extracting a *reduced* information from the p.g.f.s produced by the modelling phase. It is of course necessary in the construction of a practical filter, as the storage of a p.g.f. requires, in the most general case, an *infinite* amount of memory (see definition (1.13)). The challenge of the differentiation phase is to extract the right amount of information, i.e. *meaningful enough* to the operator for tracking purposes, and resulting in filtering equations that are *tractable enough*. In our present case, for example, we aim at reducing the propagated information to the mean target number in the scene. In order to produce the filtering equations, it is often necessary to make *filtering approximations* on top of the modelling assumptions; it is the combination of both that characterizes the resulting filter - in our case, the "cardinality only" PHD filter.

1.7.1 Modelling phase

Prediction step

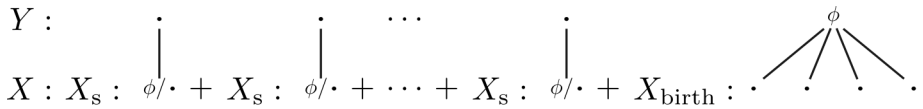
The modelling assumptions are as follows:

1. The targets are independent;

2. A target survives with probability p_{s}, dies (i.e. vanishes from the scene) otherwise;

3. A number of newborn targets enter the scene, independently of the number surviving targets, following a birth mechanism described by a random variable X_{birth} with known characteristics (p.g.f. $\mathcal{G}_{\mathrm{birth}}$).

The prediction step can be represented as follows:



Exploiting the results established in sections 1.5 and 1.6, we can then say that:

1. Since each target survives with probability p_{s}, the "survival" random variable X_{s} in the figure above is Bernoulli with parameter p_{s}:

$$\mathcal{G}_{\mathrm{s}}(s) = 1 - p_{\mathrm{s}} + p_{\mathrm{s}}s. \tag{1.48}$$

2. The number of surviving targets is described by a random variable $X_{\mathrm{sur.}}$ which is the result of a branching with parent random variable Y and transition random variable X_{s}:

$$\mathcal{G}_{\mathrm{sur.}}(s) = G_Y(\mathcal{G}_{\mathrm{s}}(s)). \tag{1.49}$$

3. The predicted number of targets, described by X, is the sum of the surviving targets and the newborn targets:

$$G_X(s) = \mathcal{G}_{\mathrm{sur.}}(s)\mathcal{G}_{\mathrm{birth}}(s). \tag{1.50}$$

In consequence, the p.g.f. form of the prediction step of the "cardinality only" PHD filter is given by:

$$G_X(s) = G_Y(1 - p_{\mathrm{s}} + p_{\mathrm{s}}s)\mathcal{G}_{\mathrm{birth}}(s). \tag{1.51}$$

Note that using the p.g.f.s allowed us to produce a full description of the predicted number of targets X *without enumerating and computing each cardinality probability* $p_X(n)$ for every target number $n \in \mathbb{N}$.

Data update step

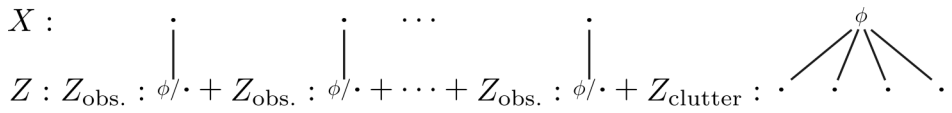
The modelling assumptions are as follows:

1. The measurements are produced independently;

2. A target is detected and produces a *single* measurement with probability p_{d}, is undetected otherwise;

3. A number of clutter measurements are produced, independently from the target measurements, following a clutter mechanism described by a random variable Z_{clutter} with known characteristics (p.g.f. $\mathcal{G}_{\mathrm{clutter}}$).

The update step can be represented as follows:



Exploiting the results established in sections 1.5 and 1.6, we can then say that:

1. Since each target is detected with probability p_{d}, the "observation" random variable $Z_{\text{obs.}}$ in the figure above is Bernoulli with parameter p_{d}:

$$\mathcal{G}_{\text{obs.}}(t) = 1 - p_{\mathrm{d}} + p_{\mathrm{d}}t. \tag{1.52}$$

2. The number of target measurements is described by a random variable Z_{target} which is the result of a branching with parent random variable X and transition random variable $Z_{\text{obs.}}$:

$$\mathcal{G}_{Z_{\text{target}},X}(t,s) = G_X(s\mathcal{G}_{\text{obs.}}(t)). \tag{1.53}$$

3. The number of measurement, described by Z, is the sum of the target measurements and the clutter measurements:

$$\mathcal{G}_{Z,X}(t,s) = \mathcal{G}_{Z_{\text{target}},X}(t,s)\mathcal{G}_{\text{clutter}}(t). \tag{1.54}$$

In consequence, the joint p.g.f. of the number of measurements and targets is given by:

$$\mathcal{G}_{Z,X}(t,s) = G_X(s(1 - p_{\mathrm{d}} + p_{\mathrm{d}}t))\mathcal{G}_{\text{clutter}}(t). \tag{1.55}$$

So far, the structures of the prediction and update steps have been remarkably similar and have led to identical results. The main difference is that we are not interested, at least as a final result, in marginalizing (1.55) over the predicted number of targets X in the same way as (1.51) is (implicitly) marginalized over the prior number of targets Y. Nor are we interested in marginalizing (1.55) over the number of measurements Z; we *know with certainty* that the sensor system produced m measurements and we wish to estimate the number of targets *conditioned on the realization* $Z = m$.

In order to do this, we will use the classic Bayes' rule for conditional probabilities which states that

$$p_{X|Z}(n|m) = \frac{p_{Z,X}(m,n)}{p_Z(m)}, \tag{1.56}$$

that is, the probability that there are $X = n$ targets in the scene, *given that* there $Z = m$ measurements, is the joint probability that there are $X = n$ targets *and* $Z = m$ measurements over the probability that there are $Z = m$ measurements.

If we multiply both sides of (1.56) by s^n and sum over all possible realizations of X we get

$$\sum_{n\geq 0} p_{X|Z}(n|m)s^n = \frac{\sum_{n\geq 0} p_{Z,X}(m,n)s^n}{p_Z(m)}. \tag{1.57a}$$

Using (1.13) and (1.33a), (1.57a) is equivalent to

$$\mathcal{G}_{X|Z}(s|m) = \frac{\mathcal{G}_{Z=m,X}(s)}{p_Z(m)}, \tag{1.57b}$$

where (1.33b) and (1.32) yield

$$\mathcal{G}_{X|Z}(s|m) = \frac{\frac{1}{m!}\,\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s)\big|_{t=0}}{\frac{1}{m!}G_Z^{(m)}(0)}. \tag{1.57c}$$

Finally, the denominator of Bayes' rule being the probability that there are $Z = m$ measurements marginalized over all the possible target numbers, $G_Z(t) = \mathcal{G}_{Z,X}(t,1)$ and thus $\frac{1}{m!}G_Z^{(m)}(0) = \frac{1}{m!}\,\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,1)\big|_{t=0}$. Thus (1.57c) becomes

$$\mathcal{G}_{X|Z}(s|m) = \frac{\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s)\big|_{t=0}}{\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,1)\big|_{t=0}}. \tag{1.57d}$$

With (1.55) and (1.57d), we have now produced the p.g.f. form of the data update step of the "cardinality only" PHD filter:

$$\mathcal{G}_{X|Z}(s|m) = \frac{\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s)\big|_{t=0}}{\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,1)\big|_{t=0}},$$
$$\text{where} \quad \mathcal{G}_{Z,X}(t,s) = G_X(s(1-p_\mathrm{d}+p_\mathrm{d}t))\mathcal{G}_\mathrm{clutter}(t). \tag{1.58}$$

As for the prediction step, working with the p.g.f.s allowed us to produce a full description of the updated number of targets $X|Z$ *without enumerating and computing each cardinality probability* $p_{X|Z}(n|m)$ for every target number $n \in \mathbb{N}$.

1.7.2 Differentiation phase

Prediction step

Exploiting (1.32) we can extract the mean value μ_X from the first derivative of the p.g.f. G_X:

$$\mu_X = G'_X(s)|_{s=1} \tag{1.59a}$$

Substituting the expression of the p.g.f. G_X (1.51) yields

$$\mu_X = (G_Y(1-p_\mathrm{s}+p_\mathrm{s}s)\mathcal{G}_\mathrm{birth}(s))'|_{s=1} \tag{1.59b}$$

Using the product rule (1.20) then gives

$$\mu_X = (G_Y(1-p_\mathrm{s}+p_\mathrm{s}s))'|_{s=1}\mathcal{G}_\mathrm{birth}(s)|_{s=1} + G_Y(1-p_\mathrm{s}+p_\mathrm{s}s)|_{s=1}G'_\mathrm{birth}(s)|_{s=1} \tag{1.59c}$$

With the chain rule (1.22) it becomes

$$\mu_X = (1-p_\mathrm{s}+p_\mathrm{s}s)'|_{s=1}G'_Y(1-p_\mathrm{s}+p_\mathrm{s}s)|_{s=1}\mathcal{G}_\mathrm{birth}(1) + G_Y(1)G'_\mathrm{birth}(1) \tag{1.59d}$$
$$= p_\mathrm{s}G'_Y(1)\mathcal{G}_\mathrm{birth}(1) + G_Y(1)G'_\mathrm{birth}(1) \tag{1.59e}$$

Recall from (1.15) that the p.g.f.s evaluated at $s = 1$ always yield 1, that is:

$$\mu_X = p_{\mathrm{s}} G'_Y(1) + G'_{\mathrm{birth}}(1) \tag{1.59f}$$

And finally, exploiting again the relation between the mean value and the first differentiation of the p.g.f. (1.32) yields the result

$$\mu_X = p_{\mathrm{s}} \mu_Y + \mu_{\mathrm{birth}}. \tag{1.59g}$$

Note that *no filtering approximations were necessary* to produce this result, which means that the validity of the prediction step is not limited to a particular model for the prior cardinality Y and/or the newborn cardinality X_{birth}.

Update step

Again, it is straightforward to write their expression of the posterior number of targets $\mu_{X|Z=m}$, given that the sensor system produced m observations, as the first order derivative of the p.g.f. $\mathcal{G}_{X|Z}(\cdot|m)$:

$$\mu_{X|Z=m} = G'_{X|Z}(s|m)|_{s=1} \tag{1.60a}$$

Substituting the expression of the p.g.f. $\mathcal{G}_{X|Z}(\cdot|m)$ (1.58) yields

$$\mu_{X|Z=m} = \frac{\frac{\mathrm{d}^{m+1}}{\mathrm{d}s\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s)\Big|_{t=0,s=1}}{\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,1)\Big|_{t=0}} \tag{1.60b}$$

The previous result (1.58) provides an expression of the joint p.g.f. $\mathcal{G}_{Z,X}$ w.r.t. the predicted p.g.f. G_X and the clutter p.g.f. $\mathcal{G}_{\mathrm{clutter}}$, but at this point we have not made any assumptions on the predicted cardinality X or the clutter cardinality and their respective p.g.f.s. If we attempt to proceed with the derivation in (1.60b) without assuming any particular forms for the p.g.f.s G_X and $\mathcal{G}_{\mathrm{clutter}}$, we will end up with a very general but intractable result. We will thus assume that:

1. The predicted number of targets X is Poisson;
2. The number of false alarms Z_{clutter} is Poisson.

Using the expression of a Poisson random variable w.r.t. its mean value (1.46e), we can rewrite the joint p.g.f. $\mathcal{G}_{Z,X}$ (1.55) as follows:

$$\mathcal{G}_{Z,X}(t,s) = G_X(s(1-p_{\mathrm{d}}+p_{\mathrm{d}}t))\mathcal{G}_{\mathrm{clutter}}(t) \tag{1.61a}$$

$$= e^{\mu_X(s(1-p_{\mathrm{d}}+p_{\mathrm{d}}t)-1)} e^{\mu_{\mathrm{clutter}}(t-1)} \tag{1.61b}$$

$$= e^{\mu_X(s(1-p_{\mathrm{d}}+p_{\mathrm{d}}t)-1)+\mu_{\mathrm{clutter}}(t-1)}. \tag{1.61c}$$

We can now proceed to the derivation of the joint p.g.f. in its new form (1.61c), for its exponential form makes the derivation easier exploiting the composition rule (1.24). Indeed, resolving the

first-order derivative yields immediately:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}_{Z,X}(t,s) = \frac{\mathrm{d}}{\mathrm{d}t}e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)} \tag{1.62a}$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)\right)e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)} \tag{1.62b}$$

$$= \left(\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}\right)e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)}. \tag{1.62c}$$

Since the multiplicative term in front of the exponential is independent of t, it is straightforward to write the m-th order derivative of the joint p.g.f. w.r.t. t:

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s) = \left(\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}\right)^m e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)}. \tag{1.63}$$

For the numerator in (1.58), we need to differentiate (1.63) once w.r.t. s. This is a simple task using first the product rule (1.20):

$$\begin{aligned}\frac{\mathrm{d}^{m+1}}{\mathrm{d}s\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s) = {} & \frac{\mathrm{d}}{\mathrm{d}s}\left(\left(\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}\right)^m\right)e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)} \\ & + \left(\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}\right)^m \frac{\mathrm{d}}{\mathrm{d}s}e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)}\end{aligned} \tag{1.64a}$$

We then resolve the first differentiation using the power rule (1.21), and the second one using the composition rule (1.24):

$$\begin{aligned}\frac{\mathrm{d}^{m+1}}{\mathrm{d}s\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s) = {} & m\left(\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}\right)^{m-1}\mu_X p_\mathrm{d}e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)} \\ & + \left(\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}\right)^m \mu_X(1-p_\mathrm{d}+p_\mathrm{d}t)e^{\mu_X(s(1-p_\mathrm{d}+p_\mathrm{d}t)-1)+\mu_\mathrm{clutter}(t-1)}\end{aligned} \tag{1.64b}$$

Dividing (1.64b) by (1.63) then yields:

$$\frac{\frac{\mathrm{d}^{m+1}}{\mathrm{d}s\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,s)}{\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathcal{G}_{Z,X}(t,1)} = m\frac{\mu_X p_\mathrm{d}}{\mu_X s p_\mathrm{d}+\mu_\mathrm{clutter}}+\mu_X(1-p_\mathrm{d}+p_\mathrm{d}t) \tag{1.64c}$$

At this point we just have to set $s=1$ and $t=0$ to produce the desired result (recall the general expression (1.60b)):

$$\mu_{X|Z=m} = m\frac{\mu_X p_\mathrm{d}}{\mu_X p_\mathrm{d}+\mu_\mathrm{clutter}}+\mu_X(1-p_\mathrm{d}). \tag{1.65}$$

1.7.3 Filtering equations

We have now succeeded in producing the filtering equations of the "cardinality only" PHD filter with equations (1.59g) and (1.65), repeated here:

$$\begin{cases} \mu_X = \mu_Y p_\mathrm{s}+\mu_\mathrm{birth}, \\ \mu_{X|Z=m} = \mu_X(1-p_\mathrm{d})+m\dfrac{\mu_X p_\mathrm{d}}{\mu_X p_\mathrm{d}+\mu_\mathrm{clutter}}. \end{cases} \tag{1.66}$$

Chapter 2

Point processes

Here we extend the estimation problem exposed in Chap. 1 to the full scope of multi-object filtering: we are now interested in the number *and* the spatial distribution of the objects. For this reason, we cover the description of the size and the spatial configuration of a population with *point processes* and their exploitation through p.g.fl.s.

We will see that results in Chaps. 1 and 2, and notably the exploitation of p.g.f.s and p.g.fl.s, are remarkably similar. In a broad sense, considering the spatial distribution of the objects in addition to the object number means that a lot of the quantities we defined in the previous chapter will appear in a similar form except that a dependency upon a particular multi-object configuration – a sequence of object states $(x_1, x_2, \ldots, x_n)$ – will be added. Whenever a new result is provided in this chapter, we shall reference the equivalent result in the previous chapter for pedagogical purpose.

2.1 Point processes: basic concepts

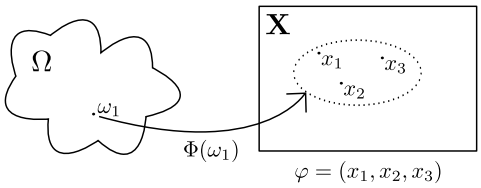
The number of targets in the scene is obviously an *integer* but it is *unknown*; besides, we suppose that each target has a state x in some *target state space* $\mathbf{X} \subseteq \mathbb{R}^{d_x}$ (e.g., position and velocity coordinates), but it is *unkwown as well*. For this reason, the description of a multi-target configuration is naturally provided by a *point process* Φ, a random variable whose realization is a sequence whose size and elements are both random.

Remark 1. *The target state space* $\mathbf{X}$ *is continuous, and we must proceed with care when defining random variables on* $\mathbf{X}$ *describing the state of a single target. Events of the form "the target has a state* equal *to some value* $x \in \mathbf{X}$*" have little practical interest, because they will occur with probability zero; rather, we wish to assess events of the form "the target has a state* within *some neighborhood* $\mathrm{d}x$ *of* $x \in \mathbf{X}$*". Intuitively speaking, if* $\mathbf{X}$ *is one dimensional and describes the target's coordinate on some axis, we wish to be able to determine the probability that the coordinate of the target lies within some "suitable" range of values* $\mathrm{d}x$ *(say,* $5\,\mathrm{m}$ *with a tolerance of* $2\,\mathrm{mm}$*) rather than a value* x *(say,* exactly $5\,\mathrm{m}$*). We shall call the set of all these "suitable" regions the Borel* σ*-algebra* $\mathcal{B}(\mathbf{X})$ *of* $\mathbf{X}$*, and whenever we shall select a (suitable) region* $B \subseteq \mathbf{X}$

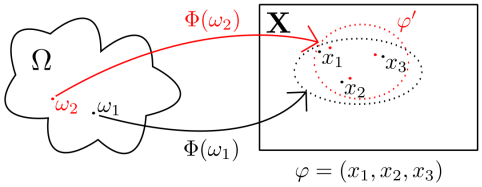
throughout the chapter it is to be understood that $B \in \mathcal{B}(\mathbf{X})$. The same concept shall apply to other continuous spaces on which probabilites are defined.

Likewise, the number of measurements produced by the sensor system between two time steps is an integer, and we suppose that each measurement has a state z in some state space $\mathbf{Z} \subseteq \mathbb{R}^{d_z}$ (e.g. polar and radial velocity coordinates).

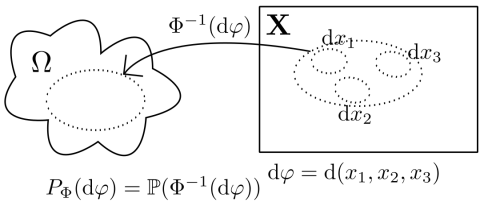
A point process Φ on the state space $\mathbf{X}$ is a mapping from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the space $\mathcal{X}$ of all the sequences of points in $\mathbf{X}$, i.e. $\mathcal{X} = \bigcup_{k \geq 0} \mathbf{X}^k$.



Depending on the construction of the point process Φ, several *outcomes* ω_i may be associated to "close" *realizations* φ, φ'.



The quantity $\Phi^{-1}(\mathrm{d}\varphi)$ represents the collection of all the possible outcomes ω_i leading to a realization within the neighborhood $\mathrm{d}\varphi$ around φ. The probability space is endowed with a *probability measure* $\mathbb{P}$ which allows us to measure the "size" of $\Phi^{-1}(\mathrm{d}\varphi)$. The "larger" $\Phi^{-1}(\mathrm{d}\varphi)$ is, the more likely is a realization to be drawn within $\mathrm{d}\varphi$ when sampling from Φ.



The *probability distribution* of the point process Φ, given by

$$P_\Phi(\mathrm{d}\varphi) = \mathbb{P}(\Phi^{-1}(\mathrm{d}\varphi)) \tag{2.1}$$

denotes the likelihood that a realization is drawn within $\mathrm{d}\varphi$ when sampling from Φ. The structure of the probability space is such that

$$\int_{\mathcal{X}} P_\Phi(\mathrm{d}\varphi) = \int_{\mathcal{X}} \mathbb{P}(\Phi^{-1}(\mathrm{d}\varphi)) \tag{2.2a}$$

$$= 1, \tag{2.2b}$$

which ensures that P_Φ can be readily interpreted as a *probability measure*. In our context, $P_\Phi(\mathrm{d}(x_1, \ldots, x_k))$ is the probability that the population described by Φ has *exactly* n objects *and* that the ith object is localized in the neighbourhood $\mathrm{d}x_i$, $1 \leq i \leq n$.

An important property of point processes is that their probability distributions are always defined as symmetric functions, so that permutations of a given realization occur with *equal probability* – e.g., $P_\Phi(\mathrm{d}(x_1, x_2)) = P_\Phi(\mathrm{d}(x_2, x_1))$. In addition, if the realizations of a point process are sequences of points that are always pairwise distinct, then the point process is called *simple*. In the context of multi-target tracking problems, the point processes are (almost) always considered simple, and this will be the case throughout this lecture.

Remark 2. *An alternative construction of simple point processes as random objects whose realizations are* sets *of points $\varphi = \{x_1, \ldots, x_n\}$, in which the elements are by construction unordered, is more common in the literature relating to the FISST framework [13]. In this context, a point process is called a RFS.*

The probability distribution P_Φ is characterized by its projection measures $P_\Phi^{(n)}$, for any $n \geq 0$. The nth-order projection measure $P_\Phi^{(n)}$, for any $n \geq 1$, is defined on $\mathbf{X}^n$; it gives the probability for the point process to be composed of n points, and the probability distribution of these points. By extension, $P_\Phi^{(0)}$ is the probability for the point process to be empty. It is important to note that the projection measures $P_\Phi^{(n)}$ are *not* probability measures as they do not integrate to one. For any $n \geq 0$, we indeed have

$$\int_{\mathbf{X}^n} P_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)) = \rho_\Phi(n), \tag{2.3}$$

where ρ_Φ is the *cardinality distribution* of the point process, describing the size of its realizations: that is, $\rho_\Phi(n)$ is the probability that a realization φ of the point process Φ is a sequence of n points.

Since the probability distribution P_Φ is symmetrical, so are the projection measures $P_\Phi^{(n)}$. For this reason, point processes are often described through their *Janossy* measures, for they "aggregate" the information provided by the projection measures over all the possible permutations of points. More precisely, for any $n \geq 0$, $J_\Phi^{(n)}$ denotes the nth-order Janossy measure of the point process Φ and is defined as

$$J_\Phi^{(n)}(B_1 \times \ldots \times B_n) = \sum_{\sigma(n)} P_\Phi^{(n)}(B_{\sigma_1} \times \ldots \times B_{\sigma_n}) \tag{2.4a}$$

$$= n! P_\Phi^{(n)}(B_1 \times \ldots \times B_n), \tag{2.4b}$$

where B_i, $1 \leq i \leq n$, is a region of $\mathbf{X}$, and $\sigma(n)$ denotes the set of all permutations $(\sigma_1, \ldots, \sigma_n)$ of $(1, \ldots, n)$.

In many practical multi-object estimation problems, the probability distribution P_Φ admits a *density* p_Φ, which quantifies the rate of change of the probability measure P_Φ per unit volume of the state space. The quantity $p_\Phi^{(n)}(x_1, \ldots, x_n)$ is thus the density of probability, per unit volume, of the point process Φ evaluated at the sequence $(x_1, \ldots, x_n)$; loosely speaking, we may describe it as the "probability that $\Phi = (x_1, \ldots, x_n)$". The projection measures $P_\Phi^{(n)}$ and the Janossy measures $J_\Phi^{(n)}$ admit densities as well, denoted $p_\Phi^{(n)}$ and $j_\Phi^{(n)}$, respectively.

We have now several tools allowing for an equivalent description of a point process: assuming that f is suitable function on $\mathcal{X}$, then the integral of f w.r.t. to the measure P_Φ can be written

$$P_\Phi(f) = \int_{\mathcal{X}} f(\varphi) P_\Phi(\mathrm{d}\varphi) \tag{2.5a}$$

$$= \int_{\mathcal{X}} f(\varphi) p_\Phi(\varphi) \mathrm{d}\varphi \tag{2.5b}$$

$$= \sum_{n\geq 0} \int_{\mathbf{X}^n} f(x_1, \ldots, x_n) P_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)) \tag{2.5c}$$

$$= \sum_{n\geq 0} \int_{\mathbf{X}^n} f(x_1, \ldots, x_n) p_\Phi^{(n)}(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{2.5d}$$

$$= \sum_{n\geq 0} \frac{1}{n!} \int_{\mathbf{X}^n} f(x_1, \ldots, x_n) J_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)) \tag{2.5e}$$

$$= \sum_{n\geq 0} \frac{1}{n!} \int_{\mathbf{X}^n} f(x_1, \ldots, x_n) j_\Phi^{(n)}(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n. \tag{2.5f}$$

A measure-theoretical formulation provides a more general framework that is required to construct certain statistical properties on point processes that can be exploited for practical applications, such as seen in Chap. 3, but is not necessary to obtain the more common results of this chapter. Throughout this chapter we shall favour expression exploiting densities rather than measures, as they are probably more common to the reader, but keep in mind that equivalent results can be obtained with a measure-theoretic formulation as well. We shall also favour probability densities over Janossy densities, as the former spare the handling of factorial terms of the form $\frac{1}{n!}$ are a more convenient tools in the context of functional differentiation.

Remark 3. *Recall that in the FISST litterature, point processes are RFSs whose realizations are sets of points [13]. It is common to define the* set integral*, for any suitable function f and region $B \subseteq \mathbf{X}$, as*

$$\int_B f(X)\delta X = \sum_{n\geq 0} \frac{1}{n!} \int_{B^n} f(\{x_1, \ldots, x_n\}) \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{2.6}$$

Set integrals are practical tools in the derivation of multi-object filtering solutions such as the PHD filter, and are convenient because of their compact expression. They are not, however, measure-theoretic integrals; for example, they are non additive as

$$\int_{B\cup B'} f(X)\delta X \neq \int_B f(X)\delta X + \int_{B'} f(X)\delta X, \tag{2.7}$$

in the general case, even if B and B' are disjoint regions of the target state space $\mathbf{X}$.

Similarly to random variables (see Chap. 1), the full knowledge of the multi-object density is seldom available in practical problems and a limited description of a point process Φ is provided

by its *moment measures* or its *moment densities*. Factorial and non factorial moment measures can be defined for any order, but their construction is more involved than for random variables and their will be the topic of a specific chapter (see Chap. 3). In this chapter we shall focus on the *first-order moment density* or *intensity* or *Probability Hypothesis Density* μ_Φ, the equivalent of the mean value of a random variable μ_X defined in Chap. 1.

The quantity $\mu_\Phi(x)$ is the density, per unit volume, of the average number of objects evaluated at x or, loosely speaking, the "average number of objects with state x". In order to compute it, one must count all the possible realizations φ of Φ with an object with state x, i.e.

$$\mu_\Phi(x) = \int_{\mathcal{X}} \left(\sum_{x_i \in \varphi} \delta_x(x_i) \right) p_\Phi(\varphi) \mathrm{d}\varphi \tag{2.8a}$$

$$= \sum_{n \geq 1} \int_{\mathbf{X}^n} \left(\sum_{i=1}^{n} \delta_x(x_i) \right) p_\Phi^{(n)}(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n, \tag{2.8b}$$

where $\delta_x(\cdot) = \delta(\cdot - x)$ is the Dirac delta function. Thus (2.8b) equals to:

$$\mu_\Phi(x) = \sum_{k \geq 1} \int_{\mathbf{X}^{k-1}} \left(\sum_{i=1}^{n} p_\Phi^{(n)}(\underbrace{x_1, \ldots, x, \ldots, x_{n-1}}_{\text{x as ith variable}}) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \tag{2.8c}$$

$$= \sum_{n \geq 1} \int_{\mathbf{X}^{n-1}} n p_\Phi^{(n)}(x, x_1, \ldots, x_{n-1}) \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \tag{2.8d}$$

$$= \sum_{n \geq 0} (n+1) \int_{\mathbf{X}^n} p_\Phi^{(n+1)}(x, x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{2.8e}$$

The last result (2.8e) shows explicitly that the first moment density is constructed by considering all the possible realizations of Φ which contains x, and marginalizing over all the possible cardinalities and over all the possible states of the remaining elements.

Just as for the integer-valued random variables, the probability and the multi-object densities are not convenient to deal with when one wish to describe simple operations on point processes (see discussion in Section 1.1). We thus need to shift the study of the point processes from the probability density to another domain.

2.2 Probability generating functional: definitions

A generating functional on $\mathbf{X}$ is a mapping G from the functions $h : \mathbf{X} \to \mathbb{R}^+$ to $\mathbb{R}$; it is built upon (or "generated by") a (possibly infinite) sequence of functions $(u_n)_{n \geq 0}$, where $u_n : \mathbf{X}^n \to \mathbb{R}^+$. The generating functional G of the sequence $(u_n)_{n \geq 0}$ is defined as

$$G(h) = \sum_{n \geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} h(x_i) \right) u_k(x_1, \ldots, x_n) \mathrm{d}x_1 \ldots \mathrm{d}x_n, \tag{2.9}$$

for any $h : \mathbf{X} \to \mathbb{R}^+$ such that the right hand side of (2.9) is finite.

Applied to a point process Φ, one can substitute the probability density in (2.9) in order to produce the *p.g.fl.* $\mathcal{G}_\Phi$ of Φ, defined as

$$\mathcal{G}_\Phi(h) = \mathbb{E}\left[\prod_{x\in\Phi} h(x)\right] \tag{2.10a}$$

$$= \int_{\mathcal{X}} \left(\prod_{x\in\varphi} h(x)\right) p_\Phi(\varphi)\mathrm{d}\varphi \tag{2.10b}$$

$$= \sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} h(x_i)\right) p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n. \tag{2.10c}$$

Note that using the probability density as the generating sequence imposes some restrictions on the range of admissible values for the *test function* h, and the p.g.fl. is defined for $h : \mathbf{X} \to [0\ 1]$.

Note the similarities between the p.g.f. of a random variable (1.13) and the p.g.fl. of a point process (2.10). The test variable s of a p.g.f. is a real number in $[0\ 1]$, while the test function of a p.g.fl. is a mapping from the target space $\mathbf{X}$ into $[0\ 1]$: the p.g.fl. "adds" the spatial component to the p.g.f., and the sum over all the possible cardinalities in the p.g.f. (1.13) becomes a sum over all the possible cardinalities *and*, for a given cardinality, an integral over all the possible object states in the p.g.fl. (2.10). From (2.10) it is easy to see that

$$\mathcal{G}_\Phi(0) = \sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} 0\right) p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n = \rho_\Phi(0), \tag{2.11}$$

$$\mathcal{G}_\Phi(1) = \sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} 1\right) p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n = 1. \tag{2.12}$$

Setting h to the mapping $h : x \in \mathbf{X} \mapsto 0$ in (2.10) allowed us to *extract* the scalar $\rho_\Phi(0)$ from the p.g.fl., i.e. the probability that there are no objects in the scene; we will see in Section 2.3 that the probability density evaluated in any number of points can be extracted through differentiation of the p.g.fl..

In multi-object filtering applications it will be necessary to study the joint behaviour of several point processes; for example, to describe the multi-measurement configuration produced by the sensor system given the multi-target configuration in the scene. The joint p.g.fl. $\mathcal{G}_{\Xi,\Phi}$ of two (possibly dependent) point processes Ξ (on $\mathbf{Z}$), Φ (on $\mathbf{X}$) is defined as the expectation

$$\mathcal{G}_{\Xi,\Phi}(g,h) = \mathbb{E}\left[\left(\prod_{z\in\Xi} g(z)\right)\left(\prod_{x\in\Phi} h(x)\right)\right] \tag{2.13a}$$

$$= \int_{\mathcal{Z}}\int_{\mathcal{X}} \left(\prod_{z\in\xi} g(z)\right)\left(\prod_{x\in\varphi} h(x)\right) p_{\Xi,\Phi}(\xi,\varphi)\mathrm{d}\xi\mathrm{d}\varphi, \tag{2.13b}$$

where $p_{\Xi,\Phi}(\xi,\varphi)$ is the joint probability density, per unit volume, evaluated at ξ *and* φ or, loosely speaking, the "probability that $\Xi = \xi$ *and* $\Phi = \varphi$". Note that, if Ξ and Φ are independent processes, then by definition $p_{\Xi,\Phi}(\xi,\varphi) = p_\Xi(\xi)p_\Phi(\varphi)$ and the joint p.g.fl. becomes:

$$\mathcal{G}_{\Xi,\Phi}(g,h) = \int_{\mathcal{Z}}\int_{\mathcal{X}} \left(\prod_{z\in\xi} g(z)\right)\left(\prod_{x\in\varphi} h(x)\right) p_\Xi(\xi)p_\Phi(\varphi)\mathrm{d}\xi\mathrm{d}\varphi \tag{2.14a}$$

$$= \left(\int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z)\right) p_\Xi(\xi)\mathrm{d}\xi\right)\left(\int_{\mathcal{X}} \left(\prod_{x\in\varphi} h(x)\right) p_\Phi(\varphi)\mathrm{d}\varphi\right) \tag{2.14b}$$

$$= \mathcal{G}_\Xi(g)\mathcal{G}_\Phi(h). \tag{2.14c}$$

We now need to introduce the notion of *functional* derivative to further exploit the p.g.fl..

2.3 Functional differentiation

2.3.1 Definition and basic rules

The first step is to check if the "classic" derivative can be applied to functionals as well as functions (see Section 1.3 in Chap. 1). Following the definition (1.18), the ordinary derivative of some functional F evaluated at h would look like:

$$F'(h) = \lim_{\eta\to 0} \frac{F(h+\eta) - F(h)}{\eta}, \tag{2.15}$$

where η would be some function of the same nature as h, i.e. $\eta : \mathbf{X} \to [0\ 1]$. Two problems arise in the definition (2.15), as neither the convergence $\eta \to 0$ nor the division by a function η are well defined – recall that the argument of a functional F is a function h – and so is η in (2.15) – *not* a real number $h(x)$.

Fortunately, other differentiation tools adapted to functionals do exist: the *functional derivatives*. Different functional derivatives have been defined by different authors for different applications, the most popular being perhaps the Fréchet and the Gâteaux derivatives. The Fréchet derivative is more restrictive, but comes with calculus rules similar to the ordinary derivative ; the Gâteaux is more general, but does not a admit a chain rule similar to the ordinary derivative given in (1.22). Quite recently, the chain derivative has been proposed as an intermediary between Fréchet and Gâteaux for which a chain rule is available; since the chain rule will be important for the derivation of filtering equations, we will use the chain derivative.

Given a functional F and two functions $h, \eta : \mathbf{X} \to \mathbb{R}^+$, we call "the (chain) derivative of F (evaluated) at h in the direction (or increment) η", and denote it by $\delta F(h;\eta)$, the limit

$$\delta F(h;\eta) = \lim_{n\to\infty} \frac{F(h+\epsilon_n\eta_n) - F(h)}{\epsilon_n}, \tag{2.16}$$

where $\{\eta_n\}_{n\geq 0}$ is a sequence of functions $\eta_n : \mathbf{X} \to \mathbb{R}^+$ converging (pointwise) to η and $\{\epsilon_n\}_{n\geq 0}$ is a sequence of positive real numbers converging to zero, if it exists and is identical for any

admissible sequences $\{\eta_n\}_{n\geq 0}$ and $\{\epsilon_n\}_{n\geq 0}$. Note that the derivative $\delta F(h;\eta)$ is a function on the object space $\mathbf{X}$; the variation of F around h in direction η being still dependent on the point x where η is evaluated, the direction appears explicitly in the functional derivative while this is not the case in the notation $f'(x)$ of the ordinary derivative.

It is formative to see how the functional derivative can be interpreted as an "extension" of the ordinary derivative. If we consider in definition (2.16) the special case where h is the constant function equal to some point $x \in \mathbf{X}$, η another constant function equal to some point to be specified later, and f is a functional on constant functions on $\mathbf{X}$, and therefore can be seen as a function on $\mathbf{X}$, we can write

$$\delta f(x;\eta) = \lim_{n\to\infty} \frac{f(x+\epsilon_n\eta_n) - f(x)}{\epsilon_n} \tag{2.17a}$$

$$= \lim_{n\to\infty} \eta_n \frac{f(x+\epsilon_n\eta_n) - f(x)}{\epsilon_n\eta_n} \tag{2.17b}$$

$$= \eta \lim_{\epsilon\to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon} \tag{2.17c}$$

$$= \eta f'(x) \tag{2.17d}$$

And thus, by setting η to the constant function equal to 1:

$$\delta f(x;1) = f'(x) \tag{2.17e}$$

The functional derivative comes with a few calculus rules that will be useful for the differentiation of p.g.fl.s. Suppose that F and G are admissible functionals, then:

$$\text{sum:} \quad \delta(F+G)(h;\eta) = \delta F(h;\eta) + \delta G(h;\eta), \tag{2.18}$$

$$\text{product:} \quad \delta(F\cdot G)(h;\eta) = \delta F(h;\eta)G(h) + F(h)\delta G(h;\eta), \tag{2.19}$$

$$\text{chain (or composition):} \quad \delta(F\circ G)(h;\eta) = \delta F(G(h);\delta G(h;\eta)). \tag{2.20}$$

Note that the sum and product rules (2.18), (2.19) are similar to those pertaining to the ordinary differentiation (1.19), (1.20).

The chain rule (2.20), on the other hand, is no longer a product as in the ordinary case (1.22). Higher-order derivations of composite functionals can be established through the Faà di Bruno's formula for chain differentials [4, 5]. The 2nd order shall be used in the next chapter, it states that

$$\delta^2(F\circ G)(h;\eta_1,\eta_2) = \delta F\left(G(h);\delta^2 G(h;\eta_1,\eta_2)\right) + \delta^2 F\left(G(h);\delta G(h;\eta_1),\delta G(h;\eta_2)\right). \tag{2.21}$$

2.3.2 A few advanced rules

The derivation of filtering equations for multi-object filters will involve the differentiation of a number of p.g.fl.s or more general functionals of various forms. Some functionals with an identical structure need to be derivated in the design of a specific filter; for this reason, it is interesting to detail here the differentiation of the most common functionals and consider the results as

"advanced rules" later on.

Let us consider a functional F such that $F(h) = h(x)$ for some fixed point x in the state space $\mathbf{X}$. F could be the p.g.fl. of a point process Φ that describes the trivial situation where there is single target in the state, and that this target has state x, with probability one. Then from the definition (2.16) we draw

$$\begin{aligned}
\delta F(h;\eta) &= \lim_{n\to\infty} \frac{F(h+\epsilon_n\eta_n) - F(h)}{\epsilon_n} && \text{(2.22a)}\\
&= \lim_{n\to\infty} \frac{h(x)+\epsilon_n\eta_n(x) - h(x)}{\epsilon_n} && \text{(2.22b)}\\
&= \lim_{n\to\infty} \eta_n(x) && \text{(2.22c)}\\
&= \eta(x) && \text{(2.22d)}
\end{aligned}$$

That is:

$$\delta(h(x);\eta) = \eta(x). \tag{2.23}$$

It is important to note that while "$\delta(h(x);\eta)$" is a very convenient notation to use, it is improper because the functional w.r.t. which we differentiate, namely F, does not appear. It can be written with the more rigorous but cumbersome form

$$\delta\left(\cdot \to \cdot(x)\right)(h;\eta) = \eta(x). \tag{2.24}$$

You will probably favour the more cumbersome form (2.24) when you start dealing with rather intricate functionals, because it helps you remembering the three basic elements of the diffentiation process: the functional, the function where it is evaluated, and the direction in which it is differentiated. With a little experience, you will probably switch to the lighter notation (2.26). Exploiting the definition of the differentiation (2.16) we can expand (2.23) to the more general result:

$$\delta((h(x))^k;\eta) = k\eta(x)(h(x))^{k-1}, \tag{2.25}$$

or, in the more exact form:

$$\delta\left(\cdot \to (\cdot(x))^k\right)(h;\eta) = k\eta(x)(h(x))^{k-1}. \tag{2.26}$$

Let us now consider a functional F such that $F(h) = \int h(x)f(x)\mathrm{d}x$ for some function f on the state space $\mathbf{X}$. If $\int f(x)\mathrm{d}x = 1$, F could be the p.g.fl. of a point process Φ that describes the trivial situation where there is single target in the state with probability one, and that the target

is distributed in space according to f. Then from the definition (2.16) we draw

$$\begin{aligned}
\delta F(h;\eta) &= \lim_{n\to\infty} \frac{F(h+\epsilon_n\eta_n) - F(h)}{\epsilon_n} && \text{(2.27a)}\\
&= \lim_{n\to\infty} \frac{\int (h(x)+\epsilon_n\eta_n(x))f(x)\mathrm{d}x - \int h(x)f(x)\mathrm{d}x}{\epsilon_n} && \text{(2.27b)}\\
&= \lim_{n\to\infty} \frac{\int h(x)f(x)\mathrm{d}x + \epsilon_n \int \eta_n(x)f(x)\mathrm{d}x - \int h(x)f(x)\mathrm{d}x}{\epsilon_n} && \text{(2.27c)}\\
&= \lim_{n\to\infty} \int \eta_n(x)f(x)\mathrm{d}x && \text{(2.27d)}\\
&= \int \eta(x)f(x)\mathrm{d}x && \text{(2.27e)}
\end{aligned}$$

That is:

$$\delta\left(\int h(x)f(x)\mathrm{d}x;\eta\right) = \int \eta(x)f(x)\mathrm{d}x \tag{2.28}$$

Similarly to (2.23), the result above uses a lighter notation where the functional does not appear exlicitly. We need to be even more cautious in this case, for h and f seem to play the same role and there is an ambiguity regarding the function in which the functional is evaluated. The more exact but cumbersome notation would be

$$\delta\left(\cdot \to \int \cdot(x)f(x)\mathrm{d}x\right)(h;\eta) = \int \eta(x)f(x)\mathrm{d}x, \tag{2.29}$$

where the functional to be differentiated, i.e. $F : h \mapsto \int h(x)f(x)\mathrm{d}x$, appears explicitly.

Let us now consider a functional F such that $F(h) = \int G(h|x)f(x)\mathrm{d}x$ for some function f on the state space $\mathbf{X}$ and some functional G. If $\int f(x)\mathrm{d}x = 1$, G could be the p.g.fl. of a point process Φ whose behaviour depends on the state of some target, and the F the p.g.fl. of the point process marginalized over all the possible values x of the said target (this will be encountered in the branching for point processes discussed in Section 2.5). Then from the definition (2.16) we draw

$$\begin{aligned}
\delta F(h;\eta) &= \lim_{n\to\infty} \frac{F(h+\epsilon_n\eta_n) - F(h)}{\epsilon_n} && \text{(2.30a)}\\
&= \lim_{n\to\infty} \frac{\int G(h+\epsilon_n\eta_n|x)f(x)\mathrm{d}x - \int G(h|x)f(x)\mathrm{d}x}{\epsilon_n} && \text{(2.30b)}\\
&= \int \lim_{n\to\infty} \frac{G(h+\epsilon_n\eta_n|x) - G(h|x)}{\epsilon_n} f(x)\mathrm{d}x && \text{(2.30c)}\\
&= \int \delta G(h|x;\eta)f(x)\mathrm{d}x && \text{(2.30d)}\\
&= F(\delta G(h|\cdot;\eta)) && \text{(2.30e)}
\end{aligned}$$

That is:

$$\delta\left(\int G(h|x)f(x)\mathrm{d}x;\eta\right) = \int \delta G(h|x;\eta)f(x)\mathrm{d}x, \tag{2.31}$$

or, with the more exact notation where the functional F appears explicitly in the left-hand side:

$$\delta\left(\cdot \to \int G(\cdot|x)f(x)\mathrm{d}x\right)(h;\eta) = \int \delta G(h|x;\eta)f(x)\mathrm{d}x, \tag{2.32}$$

Loosely speaking, we can "swap" the integral and the derivative in (2.31) or (2.32) because the functional $G(h|.)$ is encapsulated in an integral which does not depend on the test function h where the differentiation takes place. This result is easily expendable to integrals over a arbitrary number of variables $x_i \in \mathbf{X}$:

$$\delta\left(\int_{\mathcal{X}} G(h|\varphi)f(\varphi)\mathrm{d}\varphi;\eta\right) = \int_{\mathcal{X}} \delta G(h|\varphi;\eta)f(\varphi)\mathrm{d}\varphi, \tag{2.33}$$

or, with the more exact notation where the outer functional F appears explicitly in the left-hand side:

$$\delta\left(\cdot \to \int_{\mathcal{X}} G(\cdot|\varphi)f(\varphi)\mathrm{d}\varphi\right)(h;\eta) = \int_{\mathcal{X}} \delta G(h|\varphi;\eta)f(\varphi)\mathrm{d}\varphi. \tag{2.34}$$

As you may imagine, the ability to swap integrals and derivatives will be extremely handy in practical derivations. The key element that led us to the general result (2.33) is that the integral is a linear continuous operator that allowed us to proceed from (2.30b) to (2.30c). This result can be extended to other functionals than the integral, as seen in Ex. 4.2.1.

Finally, as explained in Section 1.3, the exponential form will be used extensively in the derivation of multi-object filters because the derivative of the exponential functional are easy to produce. The following results will be particularly important in the scope of this lecture. First of all, the ordinary differentiation rule (1.23) tells us that $\exp'(x) = \exp(x)$; therefore if we exploit the relationship between ordinary and functional differentiations (2.17d) we get

$$\delta \exp(x;\eta) = \eta \exp(x). \tag{2.35}$$

This result is extremely simple, but it might seem of little use in the context of point processes since we shall be dealing primarily with functionals – recall that in this context the exponential is evaluated in x and differentiated in the direction η, which are real-valued numbers. It will be, in fact, very useful in the resolution of composition of functionals where the outer functional is an exponential (more on that in Chap. 3).

Let us move on to the chain rule (2.20). When the outer functional is the exponential, it simplifies as follows:

$$\delta(\exp \circ F)(h;\eta) = \delta F(h;\eta)(\exp \circ F)(h). \tag{2.36}$$

Establishing this result is left as exercise (see Ex. 4.2.2). Note that the chain rule for exponentials for ordinary (1.24) and functional (2.36) derivatives are remarkably close.

Another useful result can be drawn from (2.36). Suppose that F is the functional $F(h) = h(x)$ for some fixed point $x \in \mathbf{X}$ (i.e., the functional F evaluates the test function h at some point of the target state space). Then, using (2.36) we can write

$$\begin{aligned} \delta \exp(h(x);\eta) &= \delta(\exp \circ F)(h;\eta) && (2.37)\\ &= \delta F(h;\eta)(\exp \circ F)(h) && (2.38)\\ &= \eta(x)\exp(h(x)), && (2.39) \end{aligned}$$

where we have used the property (2.24) to proceed from (2.38) to the final result (2.39). The development above is very formative, because it follows a typical derivation process when dealing with functional differentiation. We start from an initial expression that is a slight abuse of notation, but is convenient to write: the differentiation in the left-hand side is made in the direction of the *function* η and evaluated at the *function* h, *not* at the *real-valued number* $h(x)$ (the cases (2.35) and (2.37) are very different!). We then rewrite the initial in its "cumbersome" but rigorous expression, then we apply previously established derivation rules (2.36), (2.24), and then we revert back to a lighter and more convenient expression (2.39). Of course, with some experience, one might go straight from the left-hand side of (2.37) to the final result (2.39), having in mind the elementary rules involved in the derivation process.

2.4 p.g.fl.s and differentiation

We shall now apply the functional differentiation to the p.g.fl. to see what kind of information can be extracted from it. Suppose that Φ is a point process with known p.g.fl. $\mathcal{G}_\Phi$ and one wish to determine the probability density $p_\Phi(x)$, i.e. the probability that 1) there is a single target in the state space, and 2) that target has state x. Suppose that one wish to determine the first moment density $\mu_\Phi(x)$ as well. The results (1.32) we obtained for the p.g.f.s provide some insight on the method to produce the desired result: we should differentiate the p.g.fl. *once*. Since the spatial component is now relevant and we want to evaluate the probability density and the first moment density *in a given point* $x \in \mathbf{X}$, we shall differentiate $\mathcal{G}_\Phi$ *in the direction* δ_x:

$$\delta\mathcal{G}_\Phi(h;\delta_x) = \delta\left(\sum_{n\geq 0}\int_{\mathbf{X}^n}\left(\prod_{i=1}^{n}h(x_i)\right)p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n;\delta_x\right). \tag{2.40a}$$

This is a typical example where we can apply the swapping rule (2.33) and we get

$$\delta\mathcal{G}_\Phi(h;\delta_x) = \sum_{n\geq 1}\int_{\mathbf{X}^n}\delta\left(\left(\prod_{i=1}^{n}h(x_i)\right);\delta_x\right)p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n, \tag{2.40b}$$

where proceeding with the product rule (2.19) yields

$$\delta\mathcal{G}_\Phi(h;\delta_x) = \sum_{n\geq 1}\int_{\mathbf{X}^n}\left(\sum_{i=1}^{n}\delta(h(x_i);\delta_x)\left[\prod_{j\neq i}h(x_j)\right]\right)p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n, \tag{2.40c}$$

which simplifies with the advanced rule (2.23) and gives

$$\delta\mathcal{G}_\Phi(h;\delta_x) = \sum_{n\geq 1}\int_{\mathbf{X}^n}\left(\sum_{i=1}^n \delta_x(x_i)\left[\prod_{j\neq i} h(x_j)\right]\right) p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{2.40d}$$

$$= \sum_{n\geq 1}\int_{\mathbf{X}^{n-1}}\sum_{i=1}^n\left(\prod_{j\neq i} h(x_j)\right) p_\Phi^{(n)}(\underbrace{x_1,\ldots,x,\ldots,x_n}_{\text{x as ith variable}})\mathrm{d}x_1\ldots\mathrm{d}x_{i-1}\mathrm{d}x_{i+1}\ldots\mathrm{d}x_n \tag{2.40e}$$

$$= \sum_{n\geq 1}\int_{\mathbf{X}^{n-1}} n\left(\prod_{i=1}^{n-1} h(x_i)\right) p_\Phi^{(n)}(x,x_1,\ldots,x_{n-1})\mathrm{d}x_1\ldots\mathrm{d}x_{n-1} \tag{2.40f}$$

$$= \sum_{n\geq 1} n\int_{\mathbf{X}^{n-1}}\left(\prod_{i=1}^{n-1} h(x_i)\right) p_\Phi^{(n)}(x,x_1,\ldots,x_{n-1})\mathrm{d}x_1\ldots\mathrm{d}x_{n-1} \tag{2.40g}$$

$$= \sum_{n\geq 0}(n+1)\int_{\mathbf{X}^n}\left(\prod_{i=1}^{n} h(x_i)\right) p_\Phi^{(n+1)}(x,x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{2.40h}$$

Now, if we set $h=0$ or $h=1$ in (2.40h) we get

$$\begin{aligned}\delta\mathcal{G}_\Phi(h;\delta_x)|_{h=0} &= \sum_{n\geq 0}(n+1)\int_{\mathbf{X}^n}\left(\prod_{i=1}^{n} 0\right) p_\Phi^{(n)}(x,x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \\ &= p_\Phi^{(1)}(x),\end{aligned} \tag{2.41}$$

$$\begin{aligned}\delta\mathcal{G}_\Phi(h;\delta_x)|_{h=1} &= \sum_{n\geq 0}(n+1)\int_{\mathbf{X}^n}\left(\prod_{i=1}^{n} 1\right) p_\Phi^{(n+1)}(x,x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \\ &= \sum_{n\geq 0}(n+1)\int_{\mathbf{X}^n} p_\Phi^{(n+1)}(x,x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \\ &= \mu_\Phi(x).\end{aligned} \tag{2.42}$$

Further differentiating (2.40h) before setting $h=0$ produces the probability density evaluated at a set of any desired size. Alternatively, further differentiating (2.40h) before setting $h=1$ produces higher order factorial moment densities, which are out of the scope of this lecture. From (2.41) and (2.42) we can draw the practical extraction rules:

$$\begin{cases}\dfrac{1}{k!}\,\delta^k\mathcal{G}_\Phi(h;\delta_{x_1},\ldots,\delta_{x_k})\big|_{h=0} = p_\Phi^{(k)}(x_1,\ldots,x_k)\left(=\dfrac{j_\Phi^{(k)}(x_1,\ldots,x_n)}{k!}\right), \\ \delta\mathcal{G}_\Phi(h;\delta_x)|_{h=1} = \mu_\Phi(x).\end{cases} \tag{2.43}$$

Again, the extraction rules for p.g.f.s (1.32) and p.g.fl.s (2.43) are very similar. Since its probability density fully characterizes a point process, it follows from (2.43) that its p.g.fl. *does as well*. In other words. the knowledge of a p.g.fl. $\mathcal{G}_\Phi$ is sufficient to provide a full description of the associated point process Φ.

Before moving on to joint p.g.fl.s, we shall write a "technical result" that is known as Campbell's theorem in point process theory. Assume that one wish to evaluate some real-valued function f, defined on the state space $\mathbf{X}$, on each point $x \in \varphi$, where φ takes all the possible realizations of some point process Φ; in other words, one wish to compute the expected value of f w.r.t. to Φ – for example, f could be a function such that $f(x)$ evaluates the level of threat of a target with state $x \in \mathbf{X}$, and one wish to evaluate the average global level of threat of the population of targets described by Φ. Then we have:

$$\mathbb{E}\left[\sum_{x\in\varphi} f(x)\right] = \sum_{n\geq 1} \int_{\mathbf{X}^n} \left(\sum_{i=1}^{n} f(x_i)\right) p_{\Phi}^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{2.44a}$$

$$= \sum_{n\geq 1} \left(\sum_{i=1}^{n} \int_{\mathbf{X}^n} f(x_i) p_{\Phi}^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n\right) \tag{2.44b}$$

But, since the probability density p_{Φ} is symmetrical:

$$\mathbb{E}\left[\sum_{x\in\varphi} f(x)\right] = \sum_{n\geq 1} n \int_{\mathbf{X}^n} f(x_1) p_{\Phi}^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{2.44c}$$

$$= \int f(x) \left(\sum_{n\geq 1} n \int_{\mathbf{X}^{n-1}} p_{\Phi}^{(n)}(x,x_1,\ldots,x_{n-1})\mathrm{d}x_1\ldots\mathrm{d}x_{n-1}\right)\mathrm{d}x \tag{2.44d}$$

$$= \int f(x) \left(\sum_{n\geq 0} (n+1) \int_{\mathbf{X}^n} p_{\Phi}^{(n+1)}(x,x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n\right)\mathrm{d}x \tag{2.44e}$$

And finally, using the expression of the first-order moment density (2.8e):

$$\mathbb{E}\left[\sum_{x\in\varphi} f(x)\right] = \int f(x)\mu_{\Phi}(x)\mathrm{d}x. \tag{2.44f}$$

In other words, rather than evaluating f at each object $x \in \varphi$ averaged over all the possible multi-object realizations φ of Φ, it is equivalent to evaluate f at each point x of the state space $\mathbf{X}$, weighted by the scalar $\mu_{\Phi}(x)$, i.e. the "average number of objects with state x". It is a very important result, because it shifts the study of f from the space $\mathcal{X}$ of all the finite sequences of points of $\mathbf{X}$ to the "much smaller" space $\mathbf{X}$. Furthermore, it is valid *regardless* of the point process Φ, since we have assumed no particular form to produce the result (2.44f). The Campbell's theorem will be very handy in the derivation of the PHD filter in Section 2.7. We can rewrite Campbell's theorem (2.44) under the equivalent form

$$\int_{\mathcal{X}} \left(\sum_{x\in\varphi} f(x)\right) p_{\Phi}(\varphi)\mathrm{d}\varphi = \int f(x)\mu_{\Phi}(x)\mathrm{d}x. \tag{2.45}$$

Joint p.g.fl.s, of course, can be differentiated as well. Suppose, for example, that one wish to describe the joint behaviour of some point processes Ξ, Φ in the specific case where Ξ yields the

realization $\xi = (z_1, \ldots, z_m)$. This is described by the single variate p.g.fl. $\mathcal{G}_{\Xi=\xi,\Phi}(h)$, which can be extracted from the joint p.g.fl. $\mathcal{G}_{\Xi,\Phi}$ as follows:

$$\mathcal{G}_{\Xi=\xi,\Phi}(h) = \int_{\mathcal{X}} \left(\prod_{x\in\varphi} h(x) \right) p_{\Xi,\Phi}(\xi,\varphi)\mathrm{d}\varphi \tag{2.46a}$$

$$= \frac{1}{m!}\, \delta^m \mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}\,, \tag{2.46b}$$

where (2.46a) is drawn from the definition of the p.g.fl. (2.10) and (2.46b) is obtained with a similar reasoning as shown in (2.40) and (2.41). Note that one must keep track of test function w.r.t. which the joint p.g.fl. is differentiated in (2.46b) as it does not appear explicitly. Other notations can be adopted to avoid ambiguity, for example

$$\delta^m \mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m};\emptyset), \tag{2.47}$$

which means that the directions $\delta_{z_1},\ldots,\delta_{z_m}$ pertain to a differentiation w.r.t the first function (i.e. g), while $\emptyset$ means that no differentiation has taken place (yet) w.r.t. the second function (i.e. h). For the remainder of this lecture, points z will relate to the g function defined on the observation space $\mathbf{Z}$, while points x, y will relate to the h function defined on the target state space $\mathbf{X}$. For this reason, there will be no ambiguity on the function to which the directions relate in the various differentials, and we will use the lighter notation (2.46b) rather than the more cumbersome (2.47).

If necessary, $\delta^m \mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}$ in (2.46b) can then be differentiated w.r.t. the test function h to produce the joint probability density $p_{\Xi,\Phi}(\xi,\varphi)$ for any desired realization φ.

2.5 Operations on p.g.fl.s

We will now explore how some simple operations on point processes translate into operations on p.g.fl.s, just as some simple operations on random variables translate into simple operations on their p.g.f.s. We will consider three operations on point processes which are very useful to model physical mechanisms in multi-target filtering problems.

2.5.1 Marginalization

Marginalization occurs when two point processes Ξ, Φ have a known joint behaviour and one wish to "isolate" the behaviour of one of the point process, say Ξ. One must marginalize the joint behaviour over Φ, i.e. "integrate" the joint probability density over all the possible realizations of Φ since

$$\forall \xi \in \mathbf{Z},\ p_\Xi(\xi) = \int_{\mathcal{X}} p_{\Xi,\Phi}(\xi,\varphi)\mathrm{d}\varphi. \tag{2.48}$$

Suppose that the joint behaviour is known through the joint p.g.fl. $\mathcal{G}_{\Xi,\Phi}$. Using the definition of a joint p.g.fl. (2.13) we can write:

$$\mathcal{G}_{\Xi,\Phi}(g,1) = \int_{\mathcal{Z}}\int_{\mathcal{X}} \left(\prod_{z\in\xi} g(z)\right)\left(\prod_{x\in\varphi} 1\right) p_{\Xi,\Phi}(\xi,\varphi)d\xi\mathrm{d}\varphi \tag{2.49a}$$

$$= \int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z)\right)\left(\int_{\mathcal{X}} p_{\Xi,\Phi}(\xi,\varphi)\mathrm{d}\varphi\right) d\xi \tag{2.49b}$$

$$= \int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z)\right) p_{\Xi}(\xi)d\xi \tag{2.49c}$$

$$= \mathcal{G}_{\Xi}(g) \tag{2.49d}$$

Exactly as for the random variables (1.36), the marginalization of a point process easily translates into a very simple operation on the joint p.g.fl.:

$$\mathcal{G}_{\Xi}(g) = \mathcal{G}_{\Xi,\Phi}(g,1). \tag{2.50}$$

2.5.2 Superposition

Superposition occurs when one is not interested in the individual realizations of two *independent* point processes Φ_1 and Φ_2, but only in the union of the two realizations. If we denote by Ξ the union of two point processes Φ_1, Φ_2 with known p.g.f.s $\mathcal{G}_{\Phi_1}$, $\mathcal{G}_{\Phi_2}$, then Ξ is also a point process; using the definition of the p.g.f. (2.10) yields

$$\mathcal{G}_{\Xi}(h) = \mathbb{E}\left[\prod_{x\in\Xi} h(x)\right] \tag{2.51a}$$

$$= \mathbb{E}\left[\prod_{x\in\Phi_1\cup\Phi_2} h(x)\right] \tag{2.51b}$$

$$= \mathbb{E}\left[\prod_{x_1\in\Phi_1} h(x) \prod_{x_2\in\Phi_2} h(x)\right] \tag{2.51c}$$

$$= \mathbb{E}\left[\prod_{x_1\in\Phi_1} h(x)\right]\mathbb{E}\left[\prod_{x_2\in\Phi_2} h(x)\right] \tag{2.51d}$$

$$= \mathcal{G}_{\Phi_1}(h)\mathcal{G}_{\Phi_2}(h), \tag{2.51e}$$

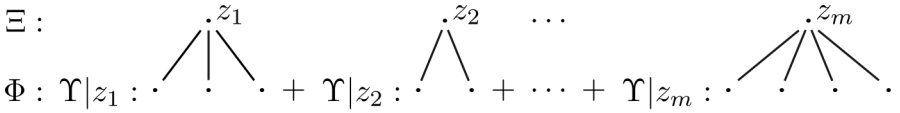
where (2.51c) is equivalent to (2.51d) because Φ_1 and Φ_2 are independent.

Similarly to the random variables (1.38), the superposition of two *independent* point processess easily translates into the product of the associated p.g.fl.s:

$$\mathcal{G}_{\Phi_1\cup\Phi_2}(s) = \mathcal{G}_{\Phi_1}(s)\mathcal{G}_{\Phi_2}(s). \tag{2.52}$$

2.5.3 Branching

Branching is a special kind of dependence between two point processes Ξ, Φ. Upon any realization $\xi = (z_1, \ldots, z_m)$ of the *parent process* Ξ, the *daughter process* Φ will be the superposition of m independent point processes $\Upsilon|z_i$, each one depending on the state of a different element z_i, as if any object z_i in the parent population was "spawning" a number of objects in the daughter population.



Suppose that the parent process Φ and the transition process $\Upsilon|\cdot$ are known through their p.g.fl.s, and that one wish to describe the daughter process Φ. The p.g.fl. describing the joint behaviour of the parent Ξ and daughter Φ processes can be written as follows:

$$\mathcal{G}_{\Xi,\Phi}(g,h) = \int_{\mathcal{Z}} \int_{\mathcal{X}} \left(\prod_{z\in\xi} g(z) \right) \left(\prod_{x\in\varphi} h(x) \right) p_{\Xi,\Phi}(\xi,\varphi) d\xi \mathrm{d}\varphi \tag{2.53a}$$

$$= \int_{\mathcal{Z}} \int_{\mathcal{X}} \left(\prod_{z\in\xi} g(z) \right) \left(\prod_{x\in\varphi} h(x) \right) p_{\Xi}(\xi) p_{\Phi|\Xi}(\varphi|\xi) d\xi \mathrm{d}\varphi \tag{2.53b}$$

$$= \int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z) \right) \left(\int_{\mathcal{X}} \left(\prod_{x\in\varphi} h(x) \right) p_{\Phi|\Xi}(\varphi|\xi) \mathrm{d}\varphi \right) p_{\Xi}(\xi) d\xi \tag{2.53c}$$

$$= \int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z) \right) \mathcal{G}_{\Phi|\Xi}(h|\xi) p_{\Xi}(\xi) d\xi, \tag{2.53d}$$

where $\mathcal{G}_{\Phi|\Xi}(h|\xi)$ is the p.g.fl. describing the daughter process Φ *conditioned* on the realization $\Xi = \xi$. If $\Xi = \xi$, then $\Phi|\Xi$ is the superposition of $|\xi|$ independent transition processes $\Upsilon|z_i$, where $z_i \in \xi$. Thus from (2.52) we have

$$\mathcal{G}_{\Phi|\Xi}(h|\xi) = \prod_{z\in\xi} \mathcal{G}_{\Upsilon}(h|z). \tag{2.54}$$

Substituting (2.54) in (2.53d) gives

$$\mathcal{G}_{\Xi,\Phi}(g,h) = \int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z) \right) \left(\prod_{z\in\xi} \mathcal{G}_{\Upsilon}(h|z) \right) p_{\Xi}(\xi) d\xi \tag{2.55a}$$

$$= \int_{\mathcal{Z}} \left(\prod_{z\in\xi} g(z) \mathcal{G}_{\Upsilon}(h|z) \right) p_{\Xi}(\xi) d\xi \tag{2.55b}$$

$$= \mathcal{G}_{\Xi}(g \mathcal{G}_{\Upsilon}(h|\cdot)). \tag{2.55c}$$

The result (2.55c) above describes the joint behaviour of the parent and daughter processes and is used in the construction of the data update equation of the multi-object Bayes filter. As before, the result is very close to its p.g.f. counterpart (1.41c). The notable difference is that the inner p.g.fl. in (2.55c) depends on the states of the elements in the parent population, which appears explicitly in the notation $\mathcal{G}_\Upsilon(h|\cdot)$, while the inner p.g.f. in (1.41c) does not. For now, since we are interested in the description of the daughter process Φ alone, we can simply marginalize (2.55c) over the parent process Ξ using (2.50) and we get

$$\mathcal{G}_\Phi(h) = \mathcal{G}_{\Xi,\Phi}(1,h) \tag{2.56a}$$
$$= \mathcal{G}_\Xi(\mathcal{G}_\Upsilon(h|\cdot)). \tag{2.56b}$$

In other words, the branching of a parent point process following a mechanism described by a transition point process translates into the composition of the associated p.g.fl.s.

2.6 A few examples of point processes and their p.g.fl.s

We shall now present two specific classes of point processes which are often used in multi-object filtering and for which it is useful to learn beforehand the structure of the associated p.g.fl.s.

2.6.1 Bernoulli point process

A *Bernoulli* point process Φ with parameter $0 \leq p \leq 1$ and spatial distribution s – that is, $\int s(x)\mathrm{d}x = 1$ – is a very simple point process defined as follows:

$$\Phi = \begin{cases} \emptyset, & \text{with probability } 1-p, \\ x, & \text{with probability } ps(x). \end{cases} \tag{2.57}$$

The Bernoulli point process describes a simple situation where 1) either there is no objects in the scene, or 2) there is a single object, with state distributed according to s. It is a "basic component" in the modelling of multi-object filters in order to describe the behaviour of a single target or a single measurement.

The construction of the p.g.fl. $\mathcal{G}_\Phi$ is straightforward using definition (2.10):

$$\mathcal{G}_\Phi(h) = \sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^n h(x_i)\right) p_\Phi^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{2.58a}$$
$$= \underbrace{p_\Phi^{(0)}(\emptyset)}_{=1-p} + \int h(x)\underbrace{p_\Phi^{(1)}(x)}_{=ps(x)}\mathrm{d}x + \sum_{n\geq 2}\int_{\mathbf{X}^n}\left(\prod_{i=1}^n h(x_i)\right)\underbrace{p_\Phi^{(n)}(x_1,\ldots,x_n)}_{=0}\mathrm{d}x_1\ldots\mathrm{d}x_n \tag{2.58b}$$
$$= 1 - p + p\int h(x)s(x)\mathrm{d}x. \tag{2.58c}$$

2.6.2 Poisson point process

A *Poisson* point process Φ with rate $\lambda \geq 0$ and spatial distribution s is defined as follows:

$$\begin{cases} \forall n \geq 0,\ |\Phi| = n \quad \text{with probability } e^{-\lambda}\dfrac{\lambda^n}{n!}, \\ \text{The object states are i.i.d. according to } s. \end{cases} \tag{2.59}$$

A Poisson point process describes a population whose number of element follows a Poisson distribution, and whose element states are independently identically distributed (i.i.d.) in space. The point patterns produced by a Poisson process, especially when the spatial distribution s is chosen as uniform over the state space, epitomize the notion of spatial randomness. It can be used to describe some natural phenomena, such as the distribution of a certain tree species in a forest.

The construction of the p.g.fl. $\mathcal{G}_\Phi$ using definition (2.10) gives:

$$\mathcal{G}_\Phi(h) = \int_{\mathcal{X}} \left(\prod_{x \in \varphi} h(x) \right) p_\Phi(\varphi) \mathrm{d}\varphi \tag{2.60a}$$

$$= \sum_{n \geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} h(x_i) \right) e^{-\lambda} \frac{\lambda^n}{n!} \left(\prod_{i=1}^{n} s(x_i) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{2.60b}$$

$$= e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} h(x_i) s(x_i) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{2.60c}$$

$$= e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} \left(\int h(x) s(x) \mathrm{d}x \right)^n \tag{2.60d}$$

$$= e^{-\lambda} \sum_{n \geq 0} \frac{\left(\lambda \int h(x) s(x) \mathrm{d}x \right)^n}{n!}. \tag{2.60e}$$

That is, using the Taylor expansion of the exponential (1.25):

$$\mathcal{G}_\Phi(h) = e^{-\lambda} e^{\lambda \int h(x) s(x) \mathrm{d}x} \tag{2.60f}$$

$$= e^{\lambda (\int h(x) s(x) \mathrm{d}x - 1)}. \tag{2.60g}$$

In multi-object filtering, Poisson random variables are appealing because of the exponential form of their p.g.fl. (2.60g), easily differentiable; *assuming* some point processes to be Poisson allows the production of tractable and easily implementable filtering equations. It is, in essence, the principle behind the derivation of the PHD filter (see Section 2.7).

It is formative to extract the first moment density of a Poisson point process in some point, say $y \in \mathbf{X}$, using the differentiation of the p.g.fl. (2.42):

$$\mu_\Phi(y) = \delta \mathcal{G}_\Phi(h; \delta_y)|_{h=1} \tag{2.61a}$$

$$= \delta(e^{\lambda (\int h(x) s(x) \mathrm{d}x - 1)}; \delta_y)\Big|_{h=1}, \tag{2.61b}$$

where using the derivation rule for exponential functionals (2.36) yields

$$\mu_\Phi(y) = \delta\left(\lambda\left(\int h(x)s(x)\mathrm{d}x - 1\right); \delta_y\right) e^{\lambda(\int h(x)s(x)\mathrm{d}x - 1)}\Big|_{h=1} \tag{2.61c}$$

$$= \lambda\delta\left(\int h(x)s(x)\mathrm{d}x; \delta_y\right)\Big|_{h=1} \underbrace{e^{\lambda(\int s(x)\mathrm{d}x - 1)}}_{=e^{\lambda(1-1)}=e^0=1}, \tag{2.61d}$$

$$\tag{2.61e}$$

where using the derivation rule (2.28) gives

$$\mu_\Phi(y) = \lambda \int \delta_y(x)s(x)\mathrm{d}x \tag{2.61f}$$

$$= \lambda s(y) \tag{2.61g}$$

In other words, the first moment density of a Poisson process Φ equals its rate multiplied by its spatial distribution; since λ and s fully characterize Φ through the definition (2.59), *so does the first moment density* μ_Φ. It is indeed customary to describe a Poisson process through its first moment density μ_Φ from which the rate λ and the spatial distribution s are easily retrieved:

$$\begin{cases} \lambda = \int \mu_\Phi(x)\mathrm{d}x \\ s(\cdot) = \lambda^{-1}\mu_\Phi(\cdot) \end{cases} \tag{2.62}$$

From (2.60g) and (2.62) we can write another expression of the p.g.fl. of a Poisson process Φ that is common in the point process litterature:

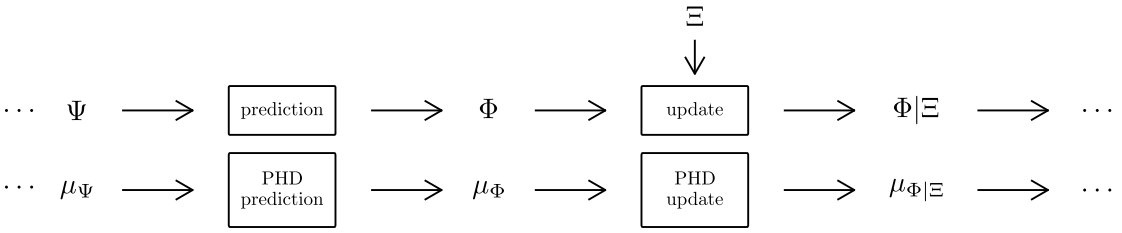
$$\mathcal{G}_\Phi(h) = e^{\int (h(x)-1)\mu_\Phi(x)\mathrm{d}x}. \tag{2.63}$$

Since it allows the expression of the p.g.fl. w.r.t. a single quantity μ_Φ which is propagated by the PHD filter, we will favour this last expression in the construction of the filter in Section 2.7.

2.7 Application: equations of the PHD filter

We shall now apply the results we have seen in the previous sections to derive the filtering equations, in their p.g.fl. form, of the PHD filter. The purpose of this Bayesian filter is to estimate and propagate the mean number of target *in any subregion of the scene*, observed by some sensor with known characteristics. Roughly speaking, the PHD filter adds a spatial component to the "cardinality only" PHD filter presented in Chap. 1, i.e., we are now interested in the number of targets *and* their state. It is formative to compare the current section to the corresponding one in the previous chapter (Section 1.7), because the construction of the filtering equations share remarkable similarities.

The data flow of one iteration of the PHD filter can be represented as follows:



where the point processes provide a description of the configuration of the following populations:

- Ψ: the targets before the prediction (prior knowledge from past iterations);
- Φ: the targets after the prediction;
- Ξ: the current measurements;
- $\Phi|\Xi$: the targets after the data update (i.e. conditioned on some realization $\Xi = (z_1, ..., z_m)$).

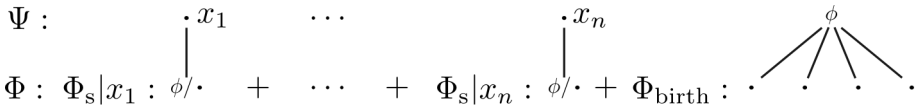
2.7.1 Modelling phase

Prediction step

The modelling assumptions are as follows:

1. The targets are independent;
2. A target with state $x \in \mathbf{X}$ survives with probability $p_{\mathrm{s}}(x)$ and moves to some new state $y \in \mathbf{X}$ distributed acc. to $m(y|x)$, dies (i.e. vanishes from the scene) otherwise;
3. A number of newborn targets enter the scene, independently of the number surviving targets, following a birth mechanism described by a point process Φ_{birth} with known characteristics (p.g.fl. $\mathcal{G}_{\mathrm{birth}}$).

The prediction step can be represented as follows:



Exploiting the results established in sections 2.5 and 2.6, we can then say that:

1. Since a target with state $x \in \mathbf{X}$ survives with probability $p_{\mathrm{s}}(x)$ and moves to some state $y \in \mathbf{X}$ distributed acc. to $m(y|x)$,, the "survival" point process $\Phi_{\mathrm{s}}|x$ in the figure above is Bernoulli with parameter $p_{\mathrm{s}}(x)$ and spatial distribution $m(\cdot|x)$:

$$\mathcal{G}_{\mathrm{s}}(h|\cdot) = 1 - p_{\mathrm{s}}(\cdot) + p_{\mathrm{s}}(\cdot) \int h(y) m(y|\cdot) \mathrm{d}y. \tag{2.64}$$

Note the difference between the p.g.f. (1.48) and the p.g.fl. above (2.64). The p.g.fl. is essentially in "augmented version" of the p.g.f. in which the spatial component is taken into account, because the state y of the (eventually) surviving target does matter *and* does depend on the previous state x of the target. This dependency upon the prior state is important, because it means that the different survival point processes $\mathcal{G}_{\mathrm{s}}$ will be dependent on different prior states, and will behave differently. This is the reason why we mark the dependency clearly and write "$\mathcal{G}_{\mathrm{s}}(h|\cdot)$" rather than "$\mathcal{G}_{\mathrm{s}}(h)$".

2. The surviving targets are described by a point process $\Phi_{\mathrm{sur.}}$ which is the result of a branching with parent point process Ψ and transition point process Φ_{s}:

$$\mathcal{G}_{\mathrm{sur.}}(h) = \mathcal{G}_{\Psi}(\mathcal{G}_{\mathrm{s}}(h|\cdot)). \tag{2.65}$$

3. The predicted targets, described by Φ, is the superposition of the surviving targets and the newborn targets:

$$\mathcal{G}_{\Phi}(h) = \mathcal{G}_{\mathrm{sur.}}(h)\mathcal{G}_{\mathrm{birth}}(h). \tag{2.66}$$

In consequence, the p.g.fl. form of the prediction step of the PHD filter is given by:

$$\mathcal{G}_{\Phi}(h) = \mathcal{G}_{\Psi}\left(1 - p_{\mathrm{s}}(\cdot) + p_{\mathrm{s}}(\cdot)\int h(y)m(y|\cdot)\mathrm{d}y\right)\mathcal{G}_{\mathrm{birth}}(h). \tag{2.67}$$

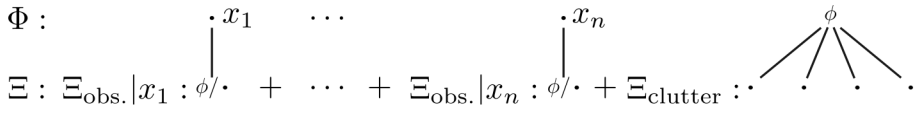
Similarly the "cardinality only" PHD, the p.g.fl.s allowed us to produce a full description of the predicted targets Φ *without computing its probability density* $p_{\Phi}(\varphi)$ for every possible sequence $\varphi \in \mathcal{X}$.

Data update step

The modelling assumptions are as follows:

1. The measurements are produced independently;
2. A target with state $x \in \mathbf{X}$ is detected with probability $p_{\mathrm{d}}(x)$ and produces a *single* measurement $z \in Z$ distributed acc. to $\ell(z|x)$, is undetected otherwise;
3. A number of clutter measurements "enter the scene", independently from the target measurements, following a clutter mechanism described by a point process Ξ_{clutter} with known characteristics (p.g.fl. $\mathcal{G}_{\mathrm{clutter}}$).

The update step can be represented as follows:



Exploiting the results established in sections 2.5 and 2.6, we can then say that:

1. Since a target with state $x \in \mathbf{X}$ is detected with probability $p_{\mathrm{d}}(x)$ and produces a measurement $z \in \mathbf{Z}$ dsitributed acc. to $\ell(z|x)$, the "observation" point process $\Xi_{\text{obs.}}|x$ in the figure above is Bernoulli with parameter $p_{\mathrm{d}}(x)$ and spatial distribution $\ell(\cdot|x)$:

$$\mathcal{G}_{\text{obs.}}(g|\cdot) = 1 - p_{\mathrm{d}}(\cdot) + p_{\mathrm{d}}(\cdot) \int g(z)\ell(z|\cdot)\mathrm{d}z. \tag{2.68}$$

2. The target measurements are described by a point process Ξ_{target} which is the result of a branching with parent process Φ and transition process $\Xi_{\text{obs.}}$:

$$\mathcal{G}_{\Xi_{\text{target}},\Phi}(g,h) = \mathcal{G}_{\Phi}(h\mathcal{G}_{\text{obs.}}(g|\cdot)). \tag{2.69}$$

3. The collection of all measurements, described by Ξ, is the superposition of the target measurements and the clutter measurements:

$$\mathcal{G}_{\Xi,\Phi}(g,h) = \mathcal{G}_{\Xi_{\text{target}},\Phi}(g,h)\mathcal{G}_{\text{clutter}}(g). \tag{2.70}$$

In consequence, the joint p.g.fl. of the measurements and targets is given by:

$$\mathcal{G}_{\Xi,\Phi}(g,h) = \mathcal{G}_{\Phi}\left(h(1 - p_{\mathrm{d}}(\cdot) + p_{\mathrm{d}}(\cdot) \int g(z)\ell(z|\cdot)\mathrm{d}z)\right)\mathcal{G}_{\text{clutter}}(g). \tag{2.71}$$

So far, the structures of the prediction and update steps have been remarkably similar and have led to identical results. The main difference is that we are not interested, at least as a final result, in marginalizing (2.71) over the predicted targets Ξ in the same way as (2.67) is (implicitly) marginalized over the prior targets Ψ. Nor are we interested in marginalizing (2.71) over the measurement process Ξ; we *know with certainty* that the sensor system produced the measurement set $Z = (z_1, \ldots, z_m)$ and we wish to estimate the multi-target configuration *conditioned on the realization* $\Xi = Z$.

In order to do this, we will use the multi-object Bayes' rule for conditional probabilities which states that

$$p_{\Xi|\Phi}(\varphi|Z) = \frac{p_{\Xi,\Phi}(Z,\varphi)}{p_{\Xi}(Z)}, \tag{2.72}$$

that is, the probability that the multi-target configuration in the scene is $\Xi = \varphi$, *given that* the collected measurement set is $\Xi = Z$, is the joint probability that the multi-target configuration is $\Phi = \varphi$ *and* the multi-measurement configuration is $\Xi = Z$, over the probability that the multi-measurement configuration is $\Xi = Z$.

If we multiply both sides of (2.72) by $\prod_{x\in\varphi} h(x)$ and integrate over all possible realizations of Φ we get

$$\int_{\mathcal{X}} \left(\prod_{x\in\varphi} h(x)\right) p_{\Phi|\Xi}(\varphi)\mathrm{d}\varphi = \frac{\int_{\mathcal{X}} \left(\prod_{x\in\varphi} h(x)\right) p_{\Xi,\Phi}(Z,\varphi)\mathrm{d}\varphi}{p_{\Xi}(Z)}. \tag{2.73a}$$

Using (2.10) and (2.46a), (2.73a) is equivalent to

$$\mathcal{G}_{\Phi|\Xi}(h|Z) = \frac{\mathcal{G}_{\Xi=Z,\Phi}(s)}{p_{\Xi}(Z)}, \tag{2.73b}$$

where (2.46b) and (2.43) yield

$$\mathcal{G}_{\Phi|\Xi}(h|Z) = \frac{\frac{1}{m!}\delta^m \, \mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}{\frac{1}{m!}\delta^m \, \mathcal{G}_{\Xi}(g;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}. \tag{2.73c}$$

Finally, the denominator of Bayes' rule being the probability that the collected measurements are $\Xi = Z$ marginalized over all the possible multi-target configurations, $\mathcal{G}_{\Xi}(g) = \mathcal{G}_{\Xi,\Phi}(g,1)$ and thus $\frac{1}{m!}\delta^m \, \mathcal{G}_{\Xi}(g;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0} = \frac{1}{m!}\delta^m \, \mathcal{G}_{\Xi,\Phi}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}$. Thus (2.73c) becomes

$$\mathcal{G}_{\Phi|\Xi}(h|Z) = \frac{\delta^m \, \mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}{\delta^m \, \mathcal{G}_{\Xi,\Phi}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}. \tag{2.73d}$$

With (2.71) and (2.73d), we have now produced the p.g.fl. form of the data update step of the PHD filter:

$$\mathcal{G}_{\Phi|\Xi}(h|Z) = \frac{\delta^m \, \mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}}{\delta^m \, \mathcal{G}_{\Xi,\Phi}(g,1;\delta_{z_1},\ldots,\delta_{z_m})|_{g=0}},$$
$$\text{where} \quad \mathcal{G}_{\Xi,\Phi}(g,h) = \mathcal{G}_{\Phi}\left(h(1-p_{\mathrm{d}}(\cdot) + p_{\mathrm{d}}(\cdot)\int g(z)\ell(z|\cdot)\mathrm{d}z\right)\mathcal{G}_{\mathrm{clutter}}(g). \tag{2.74}$$

As for the prediction step, working with the p.g.fl.s allowed us to produce a full description of the updated targets $\Phi|\Xi$ *without computing the probability density* $p_{\Phi|\Xi}(\varphi|Z)$ for every possible sequence $\varphi \in \mathcal{X}$.

2.7.2 Differentiation phase

Prediction step

Exploiting (2.43) we can extract the first-order moment density μ_Φ from the first derivative of the p.g.fl. $\mathcal{G}_\Phi$:

$$\mu_\Phi(x) = \delta\mathcal{G}_\Phi(h;\delta_x)|_{h=1} \tag{2.75a}$$

Substituting the expression of the p.g.fl. $\mathcal{G}_\Phi$ (2.67) yields

$$\mu_\Phi(x) = \delta\left(\mathcal{G}_\Psi(\mathcal{G}_{\mathrm{s}}(h|\cdot))\mathcal{G}_{\mathrm{birth}}(h);\delta_x\right)|_{h=1} \tag{2.75b}$$

Using the product rule (2.19) then gives

$$\mu_\Phi(x) = \delta\left(\mathcal{G}_\Psi(\mathcal{G}_{\mathrm{s}}(h|\cdot))\,;\delta_x\right)|_{h=1}\underbrace{\mathcal{G}_{\mathrm{birth}}(h)|_{h=1}}_{=\mathcal{G}_{\mathrm{birth}}(1)=1} + \underbrace{\mathcal{G}_\Psi(\mathcal{G}_{\mathrm{s}}(h|\cdot))|_{h=1}}_{=\mathcal{G}_\Psi(\mathcal{G}_{\mathrm{s}}(1|\cdot))=\mathcal{G}_\Psi(1)=1}\delta\mathcal{G}_{\mathrm{birth}}(h;\delta_x)|_{h=1} \tag{2.75c}$$

Exploiting again the relation between the first-order moment density and the first differentiation of the p.g.fl. (2.43) then yields

$$\mu_\Phi(x) = \delta\left(\mathcal{G}_\Psi(\mathcal{G}_{\mathrm{s}}(h|\cdot))\,;\delta_x\right)|_{h=1} + \mu_{\mathrm{birth}}(x). \tag{2.75d}$$

We now have to solve the composition $\delta\left(\mathcal{G}_\Psi(\mathcal{G}_{\mathrm{s}}(h|\cdot))\,;\delta_x\right)|_{h=1}$ in (2.75d). This is the challenging part, because we have *not* assumed any particular form for the prior process Ψ. If we were to

assume that Ψ was Poisson, for example, we could exploit the special composition rule (2.36); however, the PHD filter does *not* assume such thing and we have to resolve the composition in the general case. Quite fortunately, Campbell's theorem (2.45) will help us in this task. Starting from the definition of the p.g.fl. (2.10) we can write down explicitly the outer functional:

$$\delta\left(\mathcal{G}_\Psi(\mathcal{G}_\mathrm{s}(h|\cdot))\,;\delta_x\right)|_{h=1} = \delta\left(\int_{\mathcal{X}}\left(\prod_{\bar{x}\in\varphi}\mathcal{G}_\mathrm{s}(h|\bar{x})\right)p_\Phi(\varphi)\mathrm{d}\varphi;\delta_x\right)\Bigg|_{h=1} \tag{2.76a}$$

Using (2.33) we can swap the differentiation and integral:

$$\delta\left(\mathcal{G}_\Psi(\mathcal{G}_\mathrm{s}(h|\cdot))\,;\delta_x\right)|_{h=1} = \int_{\mathcal{X}}\delta\left(\prod_{\bar{x}\in\varphi}\mathcal{G}_\mathrm{s}(h|\bar{x});\delta_x\right)\Bigg|_{h=1}p_\Phi(\varphi)\mathrm{d}\varphi \tag{2.76b}$$

The derivative can now be expanded using the product rule:

$$\delta\left(\mathcal{G}_\Psi(\mathcal{G}_\mathrm{s}(h|\cdot))\,;\delta_x\right)|_{h=1} = \int_{\mathcal{X}}\sum_{\bar{x}\in\varphi}\left(\delta\mathcal{G}_\mathrm{s}(h|\bar{x};\delta_x)|_{h=1}\prod_{\bar{\bar{x}}\in\varphi\setminus\{\bar{x}\}}\underbrace{\mathcal{G}_\mathrm{s}(h|\bar{\bar{x}})|_{h=1}}_{=\mathcal{G}_\mathrm{s}(1|\bar{\bar{x}})=1}\right)p_\Phi(\varphi)\mathrm{d}\varphi \tag{2.76c}$$

$$= \int_{\mathcal{X}}\left(\sum_{\bar{x}\in\varphi}\delta\mathcal{G}_\mathrm{s}(h|\bar{x};\delta_x)|_{h=1}\right)p_\Phi(\varphi)\mathrm{d}\varphi \tag{2.76d}$$

Given the simple structure (2.64) of the inner functional $\mathcal{G}_\mathrm{s}$, the derivation in (2.76d) can now be solved:

$$\delta\mathcal{G}_\mathrm{s}(h|\bar{x};\delta_x)|_{h=1} = \delta\left(1-p_\mathrm{s}(\bar{x})+p_\mathrm{s}(\bar{x})\int h(y)m(y|\bar{x})\mathrm{d}y;\delta_x\right)\Bigg|_{h=1} \tag{2.76e}$$

$$= p_\mathrm{s}(\bar{x})\,\delta\left(\int h(y)m(y|\bar{x})\mathrm{d}y;\delta_x\right)\Bigg|_{h=1} \tag{2.76f}$$

Where (2.28) gives

$$\delta\mathcal{G}_\mathrm{s}(h|\bar{x};\delta_x)|_{h=1} = p_\mathrm{s}(\bar{x})\int\delta_x(y)m(y|\bar{x})\mathrm{d}y \tag{2.76g}$$

$$= p_\mathrm{s}(\bar{x})m(x|\bar{x}). \tag{2.76h}$$

Substituting (2.76h) into (2.76d) gives

$$\delta\left(\mathcal{G}_\Psi(\mathcal{G}_\mathrm{s}(h|\cdot))\,;\delta_x\right)|_{h=1} = \int_{\mathcal{X}}\left(\sum_{\bar{x}\in\varphi}p_\mathrm{s}(\bar{x})m(x|\bar{x})\right)p_\Psi(\varphi)\mathrm{d}\varphi \tag{2.76i}$$

And, finally, Campbell's theorem (2.45) yields the desired result

$$\delta\left(\mathcal{G}_\Psi(\mathcal{G}_\mathrm{s}(h|\cdot))\,;\delta_x\right)|_{h=1} = \int p_\mathrm{s}(\bar{x})m(x|\bar{x})\mu_\Psi(\bar{x})\mathrm{d}\bar{x}. \tag{2.76j}$$

Substituting (2.76j) into (2.75d) yields the final result for the predictions step:

$$\mu_\Phi(x) = \int p_\mathrm{s}(\bar{x}) m(x|\bar{x}) \mu_\Psi(\bar{x}) \mathrm{d}\bar{x} + \mu_\mathrm{birth}(x). \tag{2.77}$$

Similarly to the "cardinality only" PHD filter, *no filtering approximations were necessary* to produce this result, which means that the validity of the prediction step is not limited to a particular model for the prior targets Ψ and/or the newborn targets Φ_birth. We have seen that the general expression had a cost as the construction of the prediction equation (2.77) was significantly more involved than its "cardinality only" counterpart (1.59g). On the other hand, we will see in the next section that the construction of the update equation for the PHD filter is almost identical to its "cardinality only" counterpart.

Update step

Again, it is straightforward to write the first-order moment of the posterior targets $\mu_{\Phi|\Xi=Z}$, given that the sensor system produced the observation set $Z = \{z_1, \ldots, z_m\}$, as the first order derivative of the p.g.fl. $\mathcal{G}_{\Phi|\Xi}(\cdot|Z)$:

$$\mu_{\Phi|\Xi}(x|Z) = \delta \mathcal{G}_{\Phi|\Xi}(h|Z)|_{h=1} \tag{2.78a}$$

Substituting the expression of the p.g.fl. $\mathcal{G}_{\Phi|\Xi}(h|Z)$ (2.74) yields

$$\mu_{\Phi|\Xi}(x|Z) = \frac{\delta^{m+1} \mathcal{G}_{\Xi,\Phi}(g, h; \delta_{z_1}, \ldots, \delta_{z_m}, \delta_x)\big|_{g=0,h=1}}{\delta^m \mathcal{G}_{\Xi,\Phi}(g, 1; \delta_{z_1}, \ldots, \delta_{z_m})\big|_{g=0}} \tag{2.78b}$$

The previous result (2.74) provides an expression of the joint p.g.f. $\mathcal{G}_{\Xi,\Phi}$ w.r.t. the predicted p.g.fl. $\mathcal{G}_\Phi$ and the clutter p.g.fl. $\mathcal{G}_\mathrm{clutter}$, but at this point we have not made any assumptions on the predicted targets Φ or the clutter process Ξ_clutter and their respective p.g.fl.s. If we attempt to proceed with the derivation in (2.78b) without assuming any particular forms for the p.g.fl.s $\mathcal{G}_\Xi$ and $\mathcal{G}_\mathrm{clutter}$, we will end up with a very general but intractable result. We will thus assume that:

1. The predicted targets Φ is a Poisson process;
2. The clutter Ξ_clutter is a Poisson process.

Using the expression of a Poisson process w.r.t. its first-order moment density (2.63), we can rewrite the joint p.g.fl. $\mathcal{G}_{\Xi,\Phi}$ (2.71) as follows:

$$\begin{aligned}
\mathcal{G}_{\Xi,\Phi}(g, h) &= \mathcal{G}_\Phi \left(h(1 - p_\mathrm{d}(\cdot) + p_\mathrm{d}(\cdot) \int g(z) \ell(z|\cdot) dz) \right) \mathcal{G}_\mathrm{clutter}(g) && (2.79a) \\
&= e^{\int (h(y)(1 - p_\mathrm{d}(y) + p_\mathrm{d}(y) \int g(z) \ell(z|y) \mathrm{d}z) - 1) \mu_\Phi(y) \mathrm{d}y} e^{\int (g(z) - 1) \mu_\mathrm{clutter}(z) \mathrm{d}z} && (2.79b) \\
&= e^{\int [h(x)(1 - p_\mathrm{d}(y) + p_\mathrm{d}(y) \int g(z) \ell(z|y) \mathrm{d}z) - 1] \mu_\Phi(x) \mathrm{d}x + \int [g(z) - 1] \mu_\mathrm{clutter}(z) \mathrm{d}z} && (2.79c) \\
&= e^{F(g,h)}, && (2.79d)
\end{aligned}$$

where we defined the inner bivariate functional

$$F(g, h) = \int \left[h(y)(1 - p_\mathrm{d}(y) + p_\mathrm{d}(y) \int g(z) \ell(z|y) \mathrm{d}z) - 1 \right] \mu_\Phi(y) \mathrm{d}y + \int [g(z) - 1] \mu_\mathrm{clutter}(z) \mathrm{d}z.$$

We can now proceed to the derivation of the joint p.g.fl. in its new form (2.79d), for its exponential form makes the derivation easier exploiting the composition rule (2.36). Indeed, resolving the first-order derivative yields immediately:

$$\delta\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1}) = \delta(e^{F(g,h)};\delta_{z_1}) \tag{2.80a}$$
$$= \delta F(g,h;\delta_{z_1})e^{F(g,h)}. \tag{2.80b}$$

Now, the resolution of $\delta F(g,h;\delta_{z_1})$ is straightforward if we first swap the integral and the differentation using (2.31), then resolve the differentiation using (2.28):

$$\delta F(g,h;\delta_{z_1}) = \int h(y)p_{\mathrm{d}}(y)\delta\left(\int g(z)\ell(z|y)\mathrm{d}z;\delta_{z_1}\right)\mu_\Phi(y)\mathrm{d}y + \delta\left(\int g(z)\mu_{\mathrm{clutter}}(z)\mathrm{d}z;\delta_{z_1}\right) \tag{2.81a}$$
$$= \int h(y)p_{\mathrm{d}}(y)\left(\int \delta_{z_1}(z)\ell(z|y)\mathrm{d}z\right)\mu_\Phi(y)\mathrm{d}y + \int \delta_{z_1}(z)\mu_{\mathrm{clutter}}(z)\mathrm{d}z \tag{2.81b}$$
$$= \int h(y)p_{\mathrm{d}}(y)\ell(z_1|y)\mu_\Phi(y)\mathrm{d}y + \mu_{\mathrm{clutter}}(z_1). \tag{2.81c}$$

And thus the first-order derivative (2.80b) reads:

$$\delta\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1}) = \left[\int h(y)p_{\mathrm{d}}(y)\ell(z_1|y)\mu_\Phi(y)\mathrm{d}y + \mu_{\mathrm{clutter}}(z_1)\right]e^{F(g,h)}. \tag{2.82a}$$

Since the multiplicative term in front of the exponential is independent of g, it is straightforward to write the m-th order derivative of the joint p.g.fl. w.r.t. the function g:

$$\delta^m\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m}) = \prod_{i=1}^{m}\left[\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y + \mu_{\mathrm{clutter}}(z_i)\right]e^{F(g,h)}. \tag{2.82b}$$

For the numerator in (2.74), we need to differentiate (2.82b) once w.r.t. h. This is a simple task using first the product rule (2.19):

$$\begin{aligned}
&\delta^{m+1}\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\delta_x)\\
&= \sum_{i=1}^{m}\left[\delta\left(\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y + \mu_{\mathrm{clutter}}(z_i);\delta_x\right)\prod_{\substack{j=1\\j\neq i}}^{m}\left[\int h(y)p_{\mathrm{d}}(y)\ell(z_j|y)\mu_\Phi(y)\mathrm{d}y + \mu_{\mathrm{clutter}}(z_j)\right]\right]e^{F(g,h)}\\
&+ \prod_{i=1}^{m}\left[\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y + \mu_{\mathrm{clutter}}(z_i)\right]\delta(e^{F(g,h)};\delta_x)
\end{aligned} \tag{2.83a}$$

Where (2.28) in the first summand and the composition rule (2.36) in the second summand give

$$\begin{aligned}
&\delta^{m+1}\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\delta_x)\\
&=\sum_{i=1}^{m}\left[\int \delta_x(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y\prod_{\substack{j=1\\j\neq i}}^{m}\Big[\int h(y)p_{\mathrm{d}}(y)\ell(z_j|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_j)\Big]\right]e^{F(g,h)}\\
&+\prod_{i=1}^{m}\left[\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_i)\right]\delta F(g,h;\delta_x)e^{F(g,h)} \qquad (2.83b)\\
&=\sum_{i=1}^{m}\left[p_{\mathrm{d}}(x)\ell(z_i|x)\mu_\Phi(x)\prod_{\substack{j=1\\j\neq i}}^{m}\Big[\int h(y)p_{\mathrm{d}}(y)\ell(z_j|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_j)\Big]\right]e^{F(g,h)}\\
&+\prod_{i=1}^{m}\left[\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_i)\right]\delta F(g,h;\delta_x)e^{F(g,h)} \qquad (2.83c)
\end{aligned}$$

Again, the resolution of $\delta F(g,h;\delta_x)$ is straightforward using (2.28):

$$\begin{aligned}
\delta F(g,h;\delta_x)&=\int\delta_x(y)\left(1-p_{\mathrm{d}}(y)+p_{\mathrm{d}}(y)\int g(z)\ell(z|y)\mathrm{d}z\right)\mu_\Phi(y)\mathrm{d}y \qquad (2.84a)\\
&=\left(1-p_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int g(z)\ell(z|x)\mathrm{d}z\right)\mu_\Phi(x). \qquad (2.84b)
\end{aligned}$$

Substituting (2.84b) in (2.83c) yields the desired numerator

$$\begin{aligned}
&\delta^{m+1}\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\delta_x)\\
&\quad=\sum_{i=1}^{m}\left[p_{\mathrm{d}}(x)\ell(z_i|x)\mu_\Phi(x)\prod_{\substack{j=1\\j\neq i}}^{m}\Big[\int h(y)p_{\mathrm{d}}(y)\ell(z_j|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_j)\Big]\right]e^{F(g,h)}\\
&\quad+\prod_{i=1}^{m}\left[\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_i)\right]\left(1-p_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int g(z)\ell(z|x)\mathrm{d}z\right)\mu_\Phi(x)e^{F(g,h)} \qquad (2.85)
\end{aligned}$$

Dividing the numerator (2.85) by the denominator (2.82b) finally yields:

$$\begin{aligned}
\frac{\delta^{m+1}\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m},\delta_x)}{\delta^{m}\mathcal{G}_{\Xi,\Phi}(g,h;\delta_{z_1},\ldots,\delta_{z_m})}&=\sum_{i=1}^{n}\frac{p_{\mathrm{d}}(x)\ell(z_i|x)\mu_\Phi(x)}{\int h(y)p_{\mathrm{d}}(y)\ell(z_i|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z_i)}\\
&\quad+\left(1-p_{\mathrm{d}}(x)+p_{\mathrm{d}}(x)\int g(z)\ell(z|x)\mathrm{d}z\right)\mu_\Phi(x) \qquad (2.86a)
\end{aligned}$$

At this point we just have to set $h=1$ and $g=0$ to produce the desired result (recall the general expression (2.78b)):

$$\mu_{\Phi|\Xi}(x|Z)=\sum_{z\in Z}\frac{p_{\mathrm{d}}(x)\ell(z|x)\mu_\Phi(x)}{\int p_{\mathrm{d}}(y)\ell(z|y)\mu_\Phi(y)\mathrm{d}y+\mu_{\mathrm{clutter}}(z)}+(1-p_{\mathrm{d}}(x))\mu_\Phi(x). \qquad (2.86b)$$

2.7.3 Filtering equations

We have now succeeded in producing the filtering equations of the PHD filter [11] with equations (2.77) and (2.86b), repeated here:

$$\begin{cases} \mu_\Phi(x) = \displaystyle\int p_\mathrm{s}(y) m(x|y) \mu_\Psi(y) \mathrm{d}y + \mu_\mathrm{birth}(x), \\ \mu_{\Phi|\Xi}(x|Z) = (1 - p_\mathrm{d}(x)) \mu_\Phi(x) + \displaystyle\sum_{z \in Z} \frac{p_\mathrm{d}(x) \ell(z|x) \mu_\Phi(x)}{\int p_\mathrm{d}(y) \ell(z|y) \mu_\Phi(y) \mathrm{d}y + \mu_\mathrm{clutter}(z)}. \end{cases} \tag{2.87}$$

If we compare the equations of the "cardinality only" PHD filter (1.66) with those above, we can remark that they have a remarkably similar structure.

Chapter 3

Point processes and higher order moments

This chapter explores the concept of higher-order moments for point processes, and provides the tools for their practical derivation using the functional derivatives introduced in Chap. 2. The construction of the PHD filter with variance in target number, an example of application of higher-order moments for target detection and tracking problems, is introduced. A more detailed construction is given for the PHD and the CPHD filters in [8].

3.1 Defining moments for point processes: first attempt

We have seen in Chap. 1 that statistical moments are readily available for integer-valued random variables. Inspired by the expression of the kth order moment in (1.3), we may be want to produce the kth order moment $\mu_{\Phi}^{(k)}$ of a point process Φ as

$$\mu_{\Phi}^{(k)} = \mathbb{E}\left[\Phi^k\right] = \int_{\mathcal{X}} p_{\Phi}(\varphi)\varphi^k \mathrm{d}\varphi. \tag{3.1}$$

Then, what is wrong with the result above? Suppose, for example, that Φ is a very simple point process on $\mathbf{X}$ such that $p_{\Phi}^{(2)}(x_1, x_2) = p_{\Phi}^{(2)}(x_2, x_1) = p_{\Phi}^{(1)}(x_3) = \frac{1}{3}$, where $x_1, x_2, x_3 \in \mathbf{X}$. If we exploit (3.1) then the 1th order moment μ_{Φ} is

$$\mu_{\Phi}^{(1)} = \frac{1}{3}(x_1, x_2) + \frac{1}{3}(x_1, x_2) + \frac{1}{3}(x_3). \tag{3.2}$$

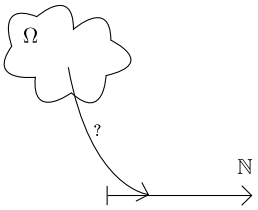
Does the result above denote some kind of "average behaviour" of the point process, as we would expect from the 1st order moment of a random variable? We may say that "on average, the population of targets has a size of $\frac{5}{3}$", but what can we say about the "average state" of those targets?

In turns out that summing sequences of points makes little sense, neither would multliplying them if we were to consider higher-order moments through our first attempt of definition (3.1). Integer-valued random variables take, by definition, values in the set of non-negative integers $\mathbb{N}$;

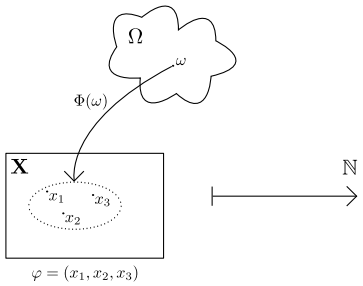
this a very convenient space as the operations of sum and multiplication are well-defined and we can thus produce expressions for statistical moments (1.3), (1.4). The point process Φ, however, is a random variable with values in the space $\mathcal{X}$ of finite sequences of points on $\mathbf{X}$; this space has a much more complicated structure on which the operations of sum and multiplication make little sense. In order to define statstical moments on a point process, then, we shall try *first* to characterize a point process with integer-valued random variables, *then* produce the moments of the said random variables.

3.2 Point processes and counting measures

Statistical moments are readily available for integer-valued random variables. In our case, we need to build an integer-valued random variable, providing a description of the target population, from a point process Φ.



Point processes are random variables whose realizations are sequences of points in the target state space $\mathcal{X}$. A point process is not a real (or complex) valued random variable and moments cannot be directly defined from point processes – the expression $\mathbb{E}[\Phi]$ has no mathematical sense, since realizations can be sequences of different sizes for which no sum operator is defined.

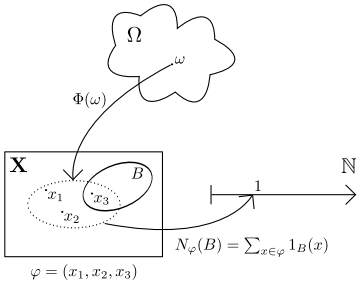


Let us fix a suitable region $B \subseteq \mathbf{X}$ [1]. One can map any realization φ of the point process to the number of elements in φ belonging to B, that is:

$$N_\varphi(B) = \sum_{x \in \varphi} 1_B(x), \tag{3.3}$$

where 1_B is the indicator function on B, i.e.

$$1_B(x) = \begin{cases} 1, & x \in B \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

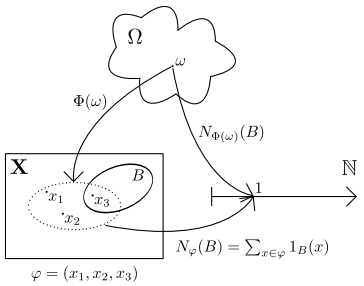


[1] Recall from Chap. 2 that B belongs to $\mathcal{B}(\mathbf{X})$, the Borel σ-algebra on the target space $\mathbf{X}$.

If we compose the point process with the mapping defined above in (3.3), we get the integer-valued random variable

$$N_\Phi(B) = \sum_{x \in \Phi} 1_B(x), \tag{3.5}$$

which provides a description of the number of targets *within* B acc. to the point process Φ.



Let us have a look at the nature of the different mathematical objects involved in the construction we have just illustrated.

1. $N_\varphi(B)$ is the number of elements of the *realization* φ belonging to the *fixed* region B; it is an *integer*;
2. $N_\Phi(B)$ maps an outcome $\omega \in \Omega$ to the number of elements of the realization $\Phi(\omega)$ belonging to the *fixed* region B; it is an *integer-valued random variable*;
3. $N_\varphi(\cdot)$ maps a suitable region $B \subseteq \mathbf{X}$ to the number of elements of the *realization* φ that it contains; it is an *integer-valued measure* called a *counting measure*;
4. $N_\Phi(\cdot)$ maps an outcome $\omega \in \Omega$ to the counting measure $N_{\Phi(\omega)}(\cdot)$; it is a *integer-valued random measure*.

We have now built a integer-valued random variable $N_\Phi(B)$ for any suitable region $B \subseteq \mathbf{X}$, and it can be shown that these random variables *characterize* the point process Φ when all the suitable regions of the target state space are considered. We can now focus on the construction of the moments of these random variables.

3.3 Defining moments for point processes: second attempt

Since $N_\Phi(B)$ is a random variable for any suitable region $B \subseteq \mathbf{X}$, inspired from (1.3), we can build the kth order *non factorial* moment of the point process Φ as the joint expectation

$$\mu_\Phi^{(k)}(B_1 \times \ldots \times B_k) = \mathbb{E}\left[N_\Phi(B_1) \cdots N_\Phi(B_k)\right] \tag{3.6a}$$

$$= \mathbb{E}\left[\sum_{x \in \Phi} 1_{B_1}(x) \cdots \sum_{x \in \Phi} 1_{B_k}(x)\right] \tag{3.6b}$$

$$= \mathbb{E}\left[\sum_{x_1, \ldots, x_k \in \Phi} 1_{B_1}(x_1) \cdots 1_{B_k}(x_k)\right], \tag{3.6c}$$

for any suitable regions $B_i \subseteq \mathbf{X}$, $1 \leq i \leq k$. We can make two important remarks about the structure of the kth order non factorial moment $\mu_\Phi^{(k)}$ above. First, we see that it takes as argument a suitable region of $\mathbf{X}^k$; it is actually a *measure* on $\mathbf{X}^k$ and its full name is the kth order

non factorial moment *measure* of the point process Φ. Second, equation (3.6b) provides some insight on the physical meaning on the non factorial moment measure: it assesses the number of targets falling *jointly* in regions B_i of the state space, and includes events where a *single* target belongs to *several* of these regions. If we wish to exclude the latter possibility, we obtain the kth order *factorial* moment measure of the point process Φ, defined as the joint expectation

$$\alpha_{\Phi}^{(k)}(B_1 \times \ldots \times B_k) = \mathbb{E}\left[\sum_{x_1,\ldots,x_k \in \Phi}^{\neq} 1_{B_1}(x_1) \cdots 1_{B_k}(x_k)\right], \tag{3.7}$$

where the $\neq$ sign indicates that the selected points in the sequence are all distinct. We may see from (3.6) and (3.7) that the 1st order moment measures, factorial and non factorial, are equal; the 1st order moment measure is also called the *intensity measure* of the point process and simply denoted by μ_{Φ}.

Remark 4. *For the sake of simplicity, we have adopted the same notation for the moment measures in this chapter and the moment densities in Chap. 2. In particular, we use the notation* μ_{Φ} *for* both *the intensity* measure *of* Φ *and the* density *of the intensity measure, also called the* Probability Hypothesis Density *or* intensity *of the point process* Φ*.*

Similarly as we have seen in (1.5) for integer-valued random variables, it is interesting to define the *central second moment* or *variance* of the point process Φ as

$$\mathrm{var}_{\Phi}(B) = \mu_{\Phi}^{(2)}(B \times B) - [\mu_{\Phi}(B)]^2, \tag{3.8}$$

for any suitable region $B \subseteq \mathbf{X}$. The regional statistics $(\mu_{\Phi}(B), \mathrm{var}_{\Phi}(B))$ of the point process Φ can then be interpreted as follows [8]

- $\mu_{\Phi}(B)$ is the mean value of the random variable $N_{\Phi}(B)$, i.e., the average number of targets within B;
- $\mathrm{var}_{\Phi}(B)$ quantifies the spread of the random variable $N_{\Phi}(B)$ around its mean value, i.e., the spread of the estimated number of targets within B around its mean value.

In detection and tracking problems, these statistics allow us to estimate the average number of target in any suitable region $B \subseteq \mathbf{X}$ of the state space, with associated uncertainty. They can be used, for example, in a sensor scheduling policy focussing on the regions where the uncertainty in the target number is the highest [1].

Remark 5. *Even if the variance is a function defined on the Borel* σ*-algebra* $\mathcal{B}(\mathbf{X})$ *associated to the state space* $\mathbf{X}$*, it is* not *a measure on* $\mathbf{X}$*. In particular, it can be shown that the variance is* not *additive, i.e.* $B_1 \subseteq B_2$ *does* not *imply that* $\mathrm{var}_{\Phi}(B_1) \leq \mathrm{var}_{\Phi}(B_2)$ *(this will be the topic of an exercise in Ex 4.3.3). An important consequence is that the variance does* not *admit a density, in the general case; i.e., there does* not *exist a density* v_{Φ} *on* $\mathbf{X}$*, such that*

$$\mathrm{var}_{\Phi}(B) = \int_B \mathrm{v}_{\Phi}(x)\mathrm{d}x, \tag{3.9}$$

for any suitable region $B \subseteq \mathbf{X}$*.*

Other meaningful statistical quantities about a point process can be defined from its moment measures, though they are out of the scope of this lecture. Suppose that one wish to study the correlation between the number of targets in suitable regions $B_1, B_2 \subseteq \mathbf{X}$ according to some point process, then similarly to (3.8) one can define the *covariance*

$$\mathrm{cov}_\Phi(B_1, B_2) = \mu_\Phi^{(2)}(B_1 \times B_2) - \mu_\Phi(B_1)\mu_\Phi(B_2). \tag{3.10}$$

3.4 Computing the moment measures

At this stage, we can make two important remarks regarding the moment measures for point processes, through a direct analogy with the moments for integer-valued random variables studied in Chap. 1. First, the *central moments* are the most interesting to us as we can provide a meaningful physical interpretation for them – at least for the lower orders! – and exploit them to study the point process, as we have just seen with the variance in (3.8). Second, the central moments can be easily written through the *non factorial* moment measures – again, we have just illutrated this point with the variance in (3.8).

Our next task, then, it to find a convenient way to produce the kth order moment measure $\mu_\Phi^{(k)}$ of a point process Φ. We have already seen in Chap. 2 that the 1th order moment density μ_Φ can be retrieved in any $x \in \mathbf{X}$ with one functional differentation in (2.43); in fact, all the derivation rules presented for densities in Chap. 2 extend naturally to measures when indicator functions are substituted to Dirac delta functions, i.e.

$$\begin{cases} \dfrac{1}{k!}\, \delta^k \mathcal{G}_\Phi(h; 1_{B_1}, \ldots, 1_{B_k})\big|_{h=0} = P_\Phi^{(k)}(B_1 \times \cdots \times B_k) \left(= \dfrac{J_\Phi^{(k)}(B_1 \times \cdots \times B_k)}{k!} \right), \\ \delta \mathcal{G}_\Phi(h; 1_B)\big|_{h=1} = \mu_\Phi(B). \end{cases} \tag{3.11}$$

The question is, can we differentiate k times the p.g.fl. $\mathcal{G}_\Phi$ to produce the kth order moment measures $\mu_\Phi^{(k)}$? Unfortunately for us, as suggested by the similar result for integer-valued random variables in (1.31), this operation yields the *factorial* moment measure $\alpha_\Phi^{(k)}$, i.e.

$$\delta^k \mathcal{G}_\Phi(h; 1_{B_1}, \ldots, 1_{B_k})\big|_{h=1} = \alpha_\Phi^{(k)}(B_1 \times \ldots \times B_k). \tag{3.12}$$

Can we, then, find a simple expression of the non factorial moment measures exploiting the factorial ones? Let us have a look at the case $k = 2$. Starting from the definition (3.6) we can write:

$$\mu_\Phi^{(2)}(B_1 \times B_2) = \mathbb{E}\left[N_\Phi(B_1) N_\Phi(B_2)\right] \tag{3.13a}$$

$$= \mathbb{E}\left[\sum_{x_1, x_2 \in \Phi} 1_{B_1}(x_1) 1_{B_2}(x_2) \right] \tag{3.13b}$$

$$= \underbrace{\mathbb{E}\left[\sum_{x_1, x_2 \in \Phi}^{\neq} 1_{B_1}(x_1) 1_{B_2}(x_2) \right]}_{= \alpha_\Phi^{(2)}(B_1 \times B_2)} + \underbrace{\mathbb{E}\left[\sum_{x \in \Phi} 1_{B_1}(x) 1_{B_2}(x) \right]}_{= \mu_\Phi(B_1 \cap B_2)} \tag{3.13c}$$

That is, the 2nd order moment measure $\mu_\Phi^{(2)}$ can be retrieved from the 2nd order factorial moment measure $\alpha_\Phi^{(2)}$ and the intensity μ_Φ. Since $\alpha_\Phi^{(2)}$ can be computed with a second-order derivative of the p.g.fl. with (3.12), $\mu_\Phi^{(2)}$ can be expressed as a combination of differentiated p.g.fl.s. While a relation between factorial and non factorial moments such as (3.13c) exists for higher orders, it becomes increasingly complicated and tedious to write in order to produce the expression of non factorial moment measures of increasing order.

So why can't we write $\mu_\Phi^{(2)}(B_1 \times B_2)$ directly as the derivative of a *single* p.g.fl.? Since we need to consider the occurrence of a single point of the point process falling in the intersection of the two regions, the quantity

$$1_{B_1}(x)1_{B_2}(x), \tag{3.14}$$

among others, must appear somewhere during the derivation process. Let us differentiate twice the p.g.fl. in the directions 1_{B_1} and 1_{B_2} and see what we get. Recall from the previous chapter in (2.10) that the p.g.fl. of a process Φ is given by

$$\mathcal{G}_\Phi(h) = \sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} h(x_i) \right) P_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)). \tag{3.15}$$

Thus, the differentiated p.g.fl. reads

$$\delta^2 \mathcal{G}_\Phi(h; 1_{B_1}, 1_{B_2})\big|_{h=1} = \delta^2 \left(\sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} h(x_i) \right) P_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)); 1_{B_1}, 1_{B_2} \right) \Bigg|_{h=1}. \tag{3.16a}$$

$$= \sum_{n\geq 0} \int_{\mathbf{X}^n} \delta^2 \left(\prod_{i=1}^{n} h(x_i); 1_{B_1}, 1_{B_2} \right) \Bigg|_{h=1} P_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)). \tag{3.16b}$$

Let us expand a derivation term in (3.16b), say $n = 3$. If we differentiate $h(x_1)h(x_2)h(x_3)$ once in the direction 1_{B_1} using the product rule (2.19) we get:

$$\begin{aligned}
&\delta(h(x_1)h(x_2)h(x_3); 1_{B_1}) \\
&= \delta(h(x_1); 1_{B_1})h(x_2)h(x_3) + h(x_1)\delta(h(x_2); 1_{B_1})h(x_3) + h(x_1)h(x_2)\delta(h(x_3); 1_{B_1}) && \text{(3.17a)} \\
&= 1_{B_1}(x_1)h(x_2)h(x_3) + h(x_1)1_{B_1}(x_2)h(x_3) + h(x_1)h(x_2)1_{B_1}(x_3) && \text{(3.17b)}
\end{aligned}$$

We see in (3.17b) that h is not "available" at a given point x_i once it has been differentiated – for example, the first term does not contain $h(x_1)$ anymore. If we differentiate a second time in the direction 1_{B_2}, then set $h = 1$, we get:

$$\begin{aligned}
&\delta^2(h(x_1)h(x_2)h(x_3); 1_{B_1}, 1_{B_2})\big|_{h=1} \\
&= 1_{B_1}(x_1)1_{B_2}(x_2) + 1_{B_1}(x_1)1_{B_2}(x_3) + 1_{B_2}(x_1)1_{B_1}(x_2) \\
&+ 1_{B_1}(x_2)1_{B_2}(x_3) + 1_{B_2}(x_1)1_{B_1}(x_3) + 1_{B_2}(x_2)1_{B_1}(x_3)
\end{aligned} \tag{3.18}$$

Because the test function h "disappeared" in the simple derivation terms such as $\delta(h(x_1); 1_{B_1})$, the desired products such as $1_{B_1}(x_1)1_{B_2}(x_1)$ do not appear in (3.18): in other words, the p.g.fl. is not adapted to the production of *non factorial* moment measures.

Let us consider the transformation $h \to e^{-f}$ and consider f as the new test function. Using the property (2.36), one can show that

$$\delta(e^{-f(x_1)}; 1_{B_1}) = -1_{B_1}(x_1)e^{-f(x_1)}. \tag{3.19}$$

This result is left as an exercise in Ex. 4.3.1.

Remark 6. *The result above may look familiar to the reader; indeed, if we consider functions rather than functional, we have the well-known result*

$$\left(e^{-f(x_1)}\right)' = -f'(x_1)e^{-f(x_1)}. \tag{3.20}$$

This time, the test function f did *not* disappear in the derivation process (3.19); differentiating a second time in direction 1_{B_2} yields

$$\begin{aligned} \delta^2(e^{-f(x_1)}; 1_{B_1}, 1_{B_2}) &= -1_{B_1}(x_1)\delta(e^{-f(x_1)}; 1_{B_2}) && \text{(3.21a)} \\ &= 1_{B_1}(x_1)1_{B_2}(x_1)e^{-f(x_1)}, && \text{(3.21b)} \end{aligned}$$

and the desired product $1_{B_1}(x_1)1_{B_2}(x_1)$ *do* appear.

The transformation $h \to e^{-f}$ is a promising lead to solve our problem at hand; if we apply it to the p.g.fl., we obtain the *Laplace functional*

$$\begin{aligned} \mathcal{L}_\Phi(f) &= \mathcal{G}_\Phi(e^{-f}) && \text{(3.22a)} \\ &= \mathbb{E}\left[\prod_{x\in\Phi} e^{-f(x)}\right] && \text{(3.22b)} \\ &= \int_{\mathcal{X}} \left(\prod_{x\in\varphi} e^{-f(x)}\right) P_\Phi(\mathrm{d}\varphi) && \text{(3.22c)} \\ &= \sum_{n\geq 0} \int_{\mathbf{X}^n} \left(\prod_{i=1}^{n} e^{-f(x_i)}\right) P_\Phi^{(n)}(\mathrm{d}(x_1, \ldots, x_n)). && \text{(3.22d)} \end{aligned}$$

As the developments above suggest, the Laplace functional is well-adapted to the production of *non factorial* moment measures; one can show that

$$(-1)^k\, \delta^k \mathcal{L}_\Phi(f; 1_{B_1}, \ldots, 1_{B_k})\big|_{f=0} = \mu_\Phi^{(k)}(B_1 \times \ldots \times B_k). \tag{3.23}$$

In particular, the quantity $\mu_\Phi^{(2)}(B \times B)$ in the expression of the variance (3.8) is given by the second-order derivative $\delta^2 \mathcal{L}_\Phi(f; 1_B, 1_B)\big|_{f=0}$.

We now have all the tools to produce the various statistical quantities of a point process, repeated here:

$$\begin{cases} P_\Phi^{(k)}(B_1 \times \cdots \times B_k) = \dfrac{1}{k!}\, \delta^k \mathcal{G}_\Phi(h; 1_{B_1}, \ldots, 1_{B_k})\big|_{h=0}\,, \\ \alpha_\Phi^{(k)}(B_1 \times \cdots \times B_k) = \delta^k \mathcal{G}_\Phi(h; 1_{B_1}, \ldots, 1_{B_k})\big|_{h=1}\,, \\ \mu_\Phi^{(k)}(B_1 \times \cdots \times B_k) = (-1)^k\, \delta^k \mathcal{L}_\Phi(f; 1_{B_1}, \ldots, 1_{B_k})\big|_{f=0}\,. \end{cases} \tag{3.24}$$

We know from the definition of the moment measures in Sec. 3.3, that the intensity measure μ_Φ is the 1st order non factorial moment measure $\mu_\Phi^{(1)}$ by definition, but it also equals the 1st order factorial moment measure $\alpha_\Phi^{(1)}$ by construction. From the result above, it follows that we should find be able to find the intensity measure μ_Φ from the 1st order derivative of either the p.g.fl. or the Laplace functional; this is left as exercise in Ex. 4.3.2.

3.5 Example: Poisson process and regional statistics

We have seen in Chap. 2 that Poisson point processes are very important in the modelling and derivation of multi-object filters, because their p.g.fl. has a very simple structure that is particularily easy to differentiate (see Section 2.6.2). Let us recall the definition (2.59) of a Poisson point process with rate $\lambda_\Phi \geq 0$ and spatial distribution s_Φ:

$$\begin{cases} \forall n \geq 0,\ |\Phi| = n \quad \text{with probability } e^{-\lambda_\Phi} \dfrac{\lambda_\Phi^n}{n!}, \\ \text{The object states are i.i.d. according to } s_\Phi. \end{cases} \tag{3.25}$$

We have also established in (2.61g) that $\mu_\Phi(x) = \lambda_\Phi s_\Phi(x)$; as all the results expressed with densities in Chap. 2 it extends to the corresponding measures, i.e.

$$\mu_\Phi(B) = \lambda_\Phi s_\Phi(B), \tag{3.26}$$

for any suitable region $B \subseteq \mathbf{X}$. Likewise, the p.g.fl. of Φ that we obtained for densities in (2.63) becomes

$$\mathcal{G}_\Phi(h) = \exp\left[\int (h(x) - 1)\mu_\Phi(\mathrm{d}x)\right]. \tag{3.27}$$

As seen in (3.25), a key element in the construction of a Poisson point process is to assume that the cardinality distribution of the process is itself distributed according to a *scalar* Poisson distribution, i.e.

$$\rho_\Phi(n) = e^{-\lambda_\Phi} \frac{\lambda_\Phi^n}{n!}. \tag{3.28}$$

As you may know, the variance of a Poisson variable equals its mean (left as exercise in Ex. 4.1.2). It suggests that a Poisson random variable is "flexible" and allows a great variability along its mean – the greater the estimated mean is, though, the greater is the uncertainty associated to the estimation.

Assume that the information we maintain about the population of targets in the scene is described by a Poisson point process Φ. Since the size of the population is described by a Poisson distribution (3.28), we conclude that the estimated target number *in the whole surveillance scene* has a mean $\mu_\Phi(\mathbf{X})$ equal to its variance $\mathrm{var}_\Phi(\mathbf{X})$. We may wonder, however, whether this property holds locally, i.e., whether

$$\mu_\Phi(B) = \mathrm{var}_\Phi(B), \tag{3.29}$$

for *any* suitable region $B \subseteq \mathbf{X}$. The fact that the targets' states are i.i.d. suggests that the property (3.29) might be true, and we shall now proceed to compute the regional statistics

$(\mu_\Phi(B), \mathrm{var}_\Phi(B))$ in order to check if it is the case.

Since our goal is to find the expression of the variance var_Φ, following the definition (1.5) we need to find the expression of the 2nd order moment measure $\mu_\Phi^{(2)}$ first, and we have just seen (Section 3.4) that we can compute this quantity from the Laplace functional $\mathcal{L}_\Phi$ of the point process. From the general expression of the Laplace functional (3.22) and the p.g.fl. of a Poisson point process (3.27) we get

$$\mathcal{L}_\Phi(f) = \exp\left[\int (e^{-f(x)} - 1)\mu_\Phi(\mathrm{d}x)\right]. \tag{3.30}$$

Let us fix some suitable region $B \subseteq \mathbf{X}$. Using the relation between moment measures and derivatives of the Laplace functional (3.23) we can write:

$$\begin{aligned}
\mu_\Phi^{(2)}(B \times B) &= \delta^2 \mathcal{L}_\Phi(f; 1_B, 1_B)\big|_{f=0} && \text{(3.31a)}\\
&= \delta^2 \left(\exp\left[\int (e^{-f(x)} - 1)\mu_\Phi(\mathrm{d}x)\right]; 1_B, 1_B\right)\Big|_{f=0} && \text{(3.31b)}\\
&= \delta^2 \left(\exp \circ F\right)(f; 1_B, 1_B)\big|_{f=0}, && \text{(3.31c)}
\end{aligned}$$

where the inner functional F is defined as $F(f) = \int (e^{-f(x)} - 1)\mu_\Phi(\mathrm{d}x)$. The derivation in (3.31c) can be easily expanded with Faà di Bruno's formula (2.21), and it yields

$$\mu_\Phi^{(2)}(B \times B) = \delta^2 \exp(F(f); \delta F(f; 1_B), \delta F(f; 1_B))\big|_{f=0} + \delta \exp(F(f); \delta^2 F(f; 1_B, 1_B))\big|_{f=0}. \tag{3.31d}$$

Let us have a look at the increment $\delta F(f; 1_B)$ in (3.31d). We can write

$$\begin{aligned}
\delta F(f; 1_B) &= \delta\left(\int (e^{-f(x)} - 1)\mu_\Phi(\mathrm{d}x); 1_B\right) && \text{(3.32a)}\\
&= \int \delta(e^{-f(x)} - 1; 1_B)\mu_\Phi(\mathrm{d}x) && \text{(3.32b)}\\
&= \int \delta(e^{-f(x)}; 1_B)\mu_\Phi(\mathrm{d}x) && \text{(3.32c)}
\end{aligned}$$

Then, using the previously established rule (3.19):

$$\delta F(f; 1_B) = \int (-1_B(x))e^{-f(x)}\mu_\Phi(\mathrm{d}x). \tag{3.33}$$

As expected, the test function f has not "disappeared" and thus $\delta F(f; 1_B)$ can be differentiated another time without vanishing. Following the same reasoning that led to (3.33), we can write the expression of the second-order derivative of the inner function F:

$$\begin{aligned}
\delta^2 F(f; 1_B, 1_B) &= \int (-1_B(x))^2 e^{-f(x)}\mu_\Phi(\mathrm{d}x) && \text{(3.34a)}\\
&= \int 1_B(x) e^{-f(x)}\mu_\Phi(\mathrm{d}x). && \text{(3.34b)}
\end{aligned}$$

Now that the increments $\delta F(f;1_B)$ and $\delta^2 F(f;1_B,1_B)$ are known, let us resolve the outer differentiation in (3.31d). At this stage, the expression (3.31d) looks rather involved; however, the only derivatives left involve exponentials differentiated in the directions $\delta F(f;1_B), \delta^2 F(f;1_B,1_B)$ which are *real-valued numbers*, and evaluated at the *real-valued number* $F(f)$. Resolving this type of differentiation is particularly easy, as we have seen in Chap. 2; using (2.35), the expression (3.31d) then becomes

$$\mu_\Phi^{(2)}(B\times B) = \left(\delta F(f;1_B)\right)^2 \exp(F(f))\Big|_{f=0} + \delta^2 F(f;1_B,1_B))\exp(F(f))\big|_{f=0}\,. \tag{3.35}$$

We can then substitute in (3.35) the values of the increments which have been determined in (3.33) and (3.34b):

$$\begin{aligned}
&\mu_\Phi^{(2)}(B\times B)\\
&= \left(\int(-1_B(x))e^{-f(x)}\mu_\Phi(\mathrm{d}x)\right)^2\Bigg|_{f=0} \exp(F(f))|_{f=0} + \left(\int 1_B(x)e^{-f(x)}\mu_\Phi(\mathrm{d}x)\right)\Bigg|_{f=0} \exp(F(f))|_{f=0} && \text{(3.36a)}\\
&= (-\mu_\Phi(B))^2\times 1 + (\mu_\Phi(B))\times 1 && \text{(3.36b)}\\
&= (\mu_\Phi(B))^2 + \mu_\Phi(B). && \text{(3.36c)}
\end{aligned}$$

Now that the second moment measure has been determined in (3.36c), we can produce the variance var_Φ of the Poisson process Φ from the definition (1.5):

$$\begin{aligned}
\mathrm{var}_\Phi(B) &= \mu_\Phi^{(2)}(B\times B) - (\mu_\Phi(B))^2 && \text{(3.37a)}\\
&= (\mu_\Phi(B))^2 + \mu_\Phi(B) - (\mu_\Phi(B))^2 && \text{(3.37b)}\\
&= \mu_\Phi(B). && \text{(3.37c)}
\end{aligned}$$

We have thus proved that property (3.29) holds: the mean and the variance in the number of targets in any suitable region $B\subseteq\mathbf{X}$ of the state space are always equal *when the population of targets is estimated with a Poisson point process.* This result, of course, does *not* hold in the general case!

An interesting consequence to (3.37c) is that the variance of a Poisson process in any region B is bounded by the mean target number in the state space since

$$\mathrm{var}_\Phi(B) = \mu_\Phi(B) \le \mu_\Phi(\mathbf{X}). \tag{3.38}$$

This means that the local behaviour of a Poisson process with "reasonable global average behaviour" – i.e. with a finite mean target number in the whole state space $\mu_\Phi(\mathbf{X})$ – can be estimated in any region B with "some accuracy" since the variance of the target number in B is finite as well according to (3.38). We will see in an exercise 4.3.5 that this is not necessarily true for other point processes.

Chapter 4

Exercises

4.1 Integer-valued random variables

4.1.1 Chain rule for ordinary differentiation

Prove the chain rule when the outer function is the exponential (1.24), i.e.

$$(\exp \circ f)'(x) = f'(x)(\exp \circ f)(x). \tag{4.1}$$

(Hint) Use the Taylor expansion of the exponential (1.25) *and the power rule* (1.21).

4.1.2 Poisson random variable

Assume X is a Poisson random variable with parameter λ_X, i.e.

$$p_X(n) = \exp(-\lambda_X)\frac{\lambda_X^n}{n!}. \tag{4.2}$$

Prove that its variance var_X equals its mean μ_X.

(Hint) Use the differentiation rule (1.31) *to extract the factorial moment* $\alpha_X^{(2)}$ *from the p.g.f.* G_X*, then use the definitions* (1.3), (1.4) *to find the expression of the non factorial moment* $\mu_X^{(2)}$ *w.r.t. to* $\alpha_X^{(2)}$ *and* μ_X*. Conclude using the definition of the variance* (1.5).

Remark 7. *Alternatively, cecause a Poisson random variable has a simple structure, the non factorial moment* $\mu_X^{(2)}$ *can be computed directly from its definition* (1.3).

4.2 Point processes

4.2.1 Functional differentiation of linear functionals

A linear functional is a functional L such that for any functions f, g, and for any scalars α, β:

$$L(\alpha f + \beta g) = \alpha L(f) + \beta L(g). \tag{4.3}$$

Let L be a continuous linear functional. Prove that, for any functional F, any test function h and any admissible direction η:

$$\delta(L \circ F)(h;\eta) = L(\delta F(h;\eta)). \tag{4.4}$$

(Hint) Use the definition of the chain differential (2.16).

This result tells us, loosely speaking, that we can "swap" a linear continuous functional and a derivative. It is very useful for practical derivations, and we sometimes apply the rule on expressions where the linear continuous functional does not appear explicity. The most common example is perhaps the rule (2.33) allowing us to swap the integral and derivative, which is a particular case of the rule (4.4) above.

4.2.2 Functional differentiation of exponential

Prove the chain rule when the outer function is the exponential (2.36), i.e.

$$\delta(\exp \circ F)(h;\eta) = \delta F(h;\eta)(\exp \circ F)(h). \tag{4.5}$$

(Hint) Use the chain rule (2.20)*, then write the outer differentation using the definition of the chain differential* (2.16)*. Use the Taylor expansion of the exponential* (1.25) *to conclude.*

4.3 Higher-order moments for point processes

4.3.1 Functional differentiation of exponential (cont.)

Let f be some suitable function on $\mathbf{X}$, and B some suitable region of $\mathbf{X}$, and x some point of $\mathbf{X}$. Prove the equality (3.19), i.e.

$$\delta\left(\exp(-f(x)); 1_B\right) = -1_B(x)\exp(-f(x)). \tag{4.6}$$

(Hint) Find the "cumbersome" expression of the left-hand side, i.e., find the functional F such that $\delta(\exp(-f(x)); 1_B) = \delta(\exp \circ F)(f; 1_B)$. Then, use the result (2.36) *(repeated in* (4.5) *in Ex. 4.2.2) to resolve the composition. Use the definition of the chain differential* (2.16) *on $\delta F(f; 1_B)$ to conclude.*

4.3.2 p.g.fl. and Laplace functional

Let Φ be a point process with p.g.fl. $\mathcal{G}_\Phi$ and Laplace functional $\mathcal{L}_\Phi$. Prove that:

$$\delta\mathcal{G}_\Phi(h;1_B)|_{h=1} = -\,\delta\mathcal{L}_\Phi(f;1_B)|_{f=0}\,, \tag{4.7}$$

for any suitable region $B \subseteq \mathbf{X}$.

What can we conclude on the first moment $\mu_\Phi^{(1)}$ and first factorial moment $\alpha_\Phi^{(1)}$ of Φ?

4.3.3 Variance and additivity

Let $B_1, B_2 \subseteq \mathbf{X}$ be two suitable regions of the state space such that $B_1 \cap B_2 = \emptyset$. Let Φ be a point process whose target number is decribed by the cardinality distribution

$$\rho_\Phi(n) = \begin{cases} 1, & n = 1, \\ 0, & \text{otherwise}, \end{cases} \tag{4.8}$$

and whose targets' states are i.i.d. according to some spatial distribution s_Φ such that $s_\Phi(B_1) = s_\Phi(B_2) = \frac{1}{2}$. The point process Φ describes a very simple situation where the number of target is the scene is 1 with certainty, and this target is within B_1 or B_2 with equal probability.

a) Prove that $\mu_\Phi(B_1) = \mu_\Phi(B_2) = \frac{1}{2}$, and $\mu_\Phi(B_1 \cup B_2) = 1$.

b) Prove that $\mathrm{var}_\Phi(B_1) = \mathrm{var}_\Phi(B_2) = \frac{1}{4}$, and $\mathrm{var}_\Phi(B_1 \cup B_2) = 0$.

Since, according to Φ, there is one and only one target in the scene, and it is either in B_1 or B_2, the results above make sense. There is uncertainty whether the target is in B_1 or B_2, and thus both $\mathrm{var}_\Phi(B_1)$ and $\mathrm{var}_\Phi(B_2)$ are non-zero. However the target is in $B_1 \cup B_2$ with certainty, and thus $\mu_\Phi(B_1 \cup B_2) = 1$ and $\mathrm{var}_\Phi(B_1 \cup B_2) = 0$.

This simple example, however, show that the variance is not additive, since $\mathrm{var}_\Phi(B_1 \cup B_2) \leq \mathrm{var}_\Phi(B_1)$ and yet $B_1 \subset B_1 \cup B_2$. Therefore, the variance is not a measure, and does not admit a density.

4.3.4 PHD filter and variance

We have seen in Chap. 2 that the intensity measure of the updated process $\Phi|\Xi$, propagated by the PHD filter, is given by (see (2.87)):

$$\mu_{\Phi|\Xi}(B|Z) = \int_B (1 - p_\mathrm{d}(x))\mu_\Phi(\mathrm{d}x) + \sum_{z\in Z} \frac{\int_B p_\mathrm{d}(x)\ell(z|x)\mu_\Phi(\mathrm{d}x)}{\int_\mathbf{X} p_\mathrm{d}(y)\ell(z|y)\mu_\Phi(\mathrm{d}y) + \mu_\mathrm{clutter}(z)}. \tag{4.9}$$

Show that the variance of the updated process is given by

$$\begin{aligned}\mathrm{var}_{\Phi|\Xi}(B|Z) &= \int_B (1-p_{\mathrm{d}}(x))\mu_\Phi(\mathrm{d}x) \\ &+ \sum_{z\in Z} \frac{\int_B p_{\mathrm{d}}(x)\ell(z|x)\mu_\Phi(\mathrm{d}x)}{\int_{\mathbf{X}} p_{\mathrm{d}}(y)\ell(z|y)\mu_\Phi(\mathrm{d}y)+\mu_{\mathrm{clutter}}(z)}\left(1-\frac{\int_B p_{\mathrm{d}}(x)\ell(z|x)\mu_\Phi(\mathrm{d}x)}{\int_{\mathbf{X}} p_{\mathrm{d}}(y)\ell(z|y)\mu_\Phi(\mathrm{d}y)+\mu_{\mathrm{clutter}}(z)}\right). \quad (4.10)\end{aligned}$$

We know that the *predicted* process Φ is assumed Poisson in the case of the PHD filter, and thus $\mu_\Phi(B) = \mathrm{var}_\Phi(B)$. What does the comparaison of (4.9) and (4.10) tell us about the *updated* process $\Phi|\Xi$?

4.3.5 i.i.d. cluster processes

A i.i.d. (cluster) process Φ is a generalization of a Poisson process which plays an important role in the construction of the CPHD filter [12], characterized by a spatial distribution s_Φ and a cardinality distribution ρ_Φ. It is defined as follows:

$$\begin{cases} \forall n \geq 0,\ |\Phi| = n \quad \text{with probability } \rho_\Phi(n) \\ \text{The object states are i.i.d. according to } s_\Phi. \end{cases} \quad (4.11)$$

Just as for the Poisson point process, the spatial distribution is the normalized intensity measure $s_\Phi(\cdot) = \frac{\mu_\Phi(\cdot)}{\mu_\Phi(\mathbf{X})}$. Also, it holds that $\sum_{n\geq 0} n\rho_\Phi(n) = \mu_\Phi(\mathbf{X})$.

Let Φ be a i.i.d. process, with intensity measure μ_Φ and cardinality distribution ρ_Φ.

a) Find the expression of the p.g.fl. of the point process.
(Hint) You should find

$$\mathcal{G}_\Phi(h) = \sum_{n\geq 0} \rho_\Phi(n)\left(\frac{\int h(x)\mu_\Phi(\mathrm{d}x)}{\mu_\Phi(\mathbf{X})}\right)^n. \quad (4.12)$$

b) In the special case where the cardinality distribution is a Poisson random variable with rate $\lambda_\Phi = \mu_\Phi(\mathbf{X})$, prove that Φ is a Poisson process.

(Hint) Show that if $\rho_\Phi(n) = \exp(-\lambda_\Phi)\frac{(\lambda_\Phi)^n}{n!}$, *then the p.g.fl.* (4.12) *reduces to the p.g.fl. of a Poisson process* (2.60).

c) Show that the variance of Φ in any suitable region $B \subseteq \mathbf{X}$ is equal to

$$\mathrm{var}_\Phi(B) = \mu_\Phi(B) + (\mu_\Phi(B))^2\left[\frac{\sum_{n\geq 2} n(n-1)\rho_\Phi(n)}{\left[\sum_{n\geq 1} n\rho_\Phi(n)\right]^2} - 1\right]. \quad (4.13)$$

What happens if the cardinality distribution is Poisson?

d) (Advanced) i.i.d. cluster processes have a less intuitive behaviour than Poisson processes and can yield surprising results. Prove that, for any mean target number $0 < \mu < \infty$, any constant $0 < C < \infty$, there exists a i.i.d. point process Φ_C such that:

- $\int \mu_{\Phi_C}(dx) = \mu$;
- $\forall B \in \mathcal{B}(\mathbf{X}), B \neq \emptyset,\ \mathrm{var}_{\Phi_C}(B) > \mu_{\Phi_C}(B) + C[\lambda(B)]^2$;

where λ is the dimensionless Lebesgue measure on $\mathbf{X}$.

(Hint) Produce a sequence of i.i.d. point processes $\{\Phi_m\}_{m\geq 0}$ such that:

- $\forall m \in \mathbb{N},\ \int \mu_{\Phi_m}(dx) = \mu$;
- $\forall B \in \mathcal{B}(\mathbf{X}),\ \lim_{m\to\infty} \frac{\mathrm{var}_{\Phi_m}(B) - \mu_{\Phi_m}(B)}{\lambda(B)^2} = \infty$.

That is, contrary to a Poisson process (see Section 3.5), one can always find a i.i.d. process Φ with a "reasonable average behaviour" – i.e. a finite *global* mean target number $\mu_\Phi(\mathbf{X})$ – and yet an *arbitrary high* variance in *any* non-empty region B – the information on the *local* target number in B is "completely unreliable".

Bibliography

[1] M. Andrecki, E. D. Delande, J. Houssineau, and D. E. Clark. Sensor management with regional statistics for the PHD filter. In *Sensor Signal Processing for Defence (SSPD)*, September 2015.

[2] V. I. Bogachev. *Measure Theory*, volume 1. Springer, 2007.

[3] V. I. Bogachev. *Measure Theory*, volume 2. Springer, 2007.

[4] D. E. Clark and J. Houssineau. Faà di Bruno's formula for chain differentials. *arXiv:1310.2833*, 2013.

[5] D. E. Clark, J. Houssineau, and E. D. Delande. A few calculus rules for chain differentials. *arXiv:1506.08626*, 2015.

[6] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*, volume 1 of *Statistical Theory and Methods*. Springer Series in Statistics, 2 edition, 2003.

[7] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*, volume 2 of *Statistical Theory and Methods*. Springer Series in Statistics, 2 edition, 2008.

[8] E. D. Delande, M. Uney, J. Houssineau, and D. E. Clark. Regional Variance for Multi-Object Filtering. *Signal Processing, IEEE Transactions on*, 13(62):3415–3428, 2014.

[9] I. R. Goodman, R. P. S. Mahler, and H.-T. Nguyen. *Mathematics of Data Fusion*. Kluwer Academic Publishers, 1997.

[10] J. Houssineau, E. D. Delande, and D. E. Clark. Notes of the Summer School on Finite Set Statistics. Technical report, Heriot-Watt University, 2013. arXiv:1308.2586.

[11] R. P. S. Mahler. Multitarget Bayes Filtering via First-Order Multitarget Moments. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1152–1178, 2003.

[12] R. P. S. Mahler. PHD Filters of Higher Order in Target Number. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1523–1543, 2007.

[13] R. P. S. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, 2007.

[14] D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. Wiley, 2 edition, 1995.