Random-Set Theory and Wireless Communications

By Ezio Biglieri, Emanuele Grossi and Marco Lops

Contents

1 Solving Estimation Problems Where You Do Not Know the Number of Things You Do Not Know 319

2 An Engineering Introduction 323

2.1 Random Sets, Set Integral, and Set Derivative 323
2.2 Bayesian and Maximum-likelihood Estimates 332
2.3 Dynamic Models and Bayesian Recursions 341

3 The Complexity Problem and Some Solutions 346

3.1 Solving the Bayesian Recursion 347
3.2 Particle Filtering/Sequential Monte-Carlo 350
3.3 The Probability Hypothesis Density 352
3.4 Zero-Order Approximations 371

4 Multiuser Detection 375

4.1 Introduction and Motivation 375
4.2 Detection of Identities and Data 377
4.3 Data Detection and Parameter Estimation 394
5 Channel Estimation 404
5.1 Introduction and Motivation 404
5.2 A Case Study: SISO Wide-Band Transmission 406
5.3 Channel Dynamics 409
5.4 Extension to MIMO Channels 412

6 Concluding Remarks 416
6.1 Topics Not Covered Here 417

Acknowledgments 420

A Mathematical Aspects of Random-Set Theory 421
A.1 Topologies and all that 421
A.2 Random Closed Sets 423
A.3 Random Finite Sets 425

B Relation to Point Processes 436
B.1 Basic Definitions 437
B.2 Simple, Finite Point Processes 439
B.3 Point Processes and Random Sets 444

C Relation to Dempster–Shafer Theory 446
C.1 Theory of Evidence vs. Probability Theory 446
C.2 D–S Theory 447
C.3 Connection with Random-Set Theory 450

Acronyms and notations 451

References 456
Random-Set Theory and Wireless Communications

Ezio Biglieri\textsuperscript{1}, Emanuele Grossi\textsuperscript{2} and Marco Lops\textsuperscript{3}

\textsuperscript{1} UPF, Barcelona, Spain and King Saud University, KSA, e.biglieri@ieee.org
\textsuperscript{2} UNICAS, DIEI, Cassino, Italy, e.grossi@unicas.it
\textsuperscript{3} UNICAS, DIEI, Cassino, Italy, lops@unicas.it

Abstract

This monograph is devoted to random-set theory, and allows unordered collections of random elements, drawn from an arbitrary space, to be handled. After illustrating its foundations, we focus on Random Finite Sets, i.e., unordered collections of random cardinality of points from an arbitrary space, and show how this theory can be applied to a number of problems arising in wireless communication systems. Three of these problems are: (1) neighbor discovery in wireless networks, (2) multiuser detection in which the number of active users is unknown and time-varying, and (3) estimation of multipath channels where the number of paths is not known \textit{a priori} and which are possibly time-varying. Standard solutions to these problems are intrinsically suboptimum as they proceed either by assuming a fixed number
of vector components, or by first estimating this number and then the values taken on by the components. It is shown how random-set theory provides optimum solutions to all these problems. The complexity issue is also examined, and suboptimum solutions are presented and discussed.
Solving Estimation Problems Where You Do Not Know the Number of Things You Do Not Know

In this section we briefly glance at random-set theory (RST) and the motivations for its application to problems in digital communication, which will be described in detail in subsequent sections.

Roughly speaking, random sets are random entities whose realizations are subsets of a given space. In RST, the outcome of a random experiment is a set, or a vector whose number of components is unknown. RST can be applied to modeling observed phenomena which are sets rather than vectors, hence it generalizes the concept of random vectors. For a simple example, consider the description of random polygons on a plane. Each polygon can be described by a vector whose components are the coordinates of its vertices. Hence, if the number of vertices can vary from 3 to a finite number \( N \), a random polygon is described by a vector whose randomness is in its components as well as in the number of its components. The first systematic exposition of random-set theory was developed in 1975 by Matheron [68]. More recent books investigating RST from a mathematical viewpoint are [72, 73].
The first engineering application of RST was found in the multisensor–multitarget data-fusion area. Multisensor–multitarget systems include “randomly varying numbers of randomly varying objects of various kinds: randomly varying collections of targets, randomly varying collections of sensors and sensor-carrying platforms, and randomly varying observation sets collected by those sensors.” [60, 66, p. 8] while data fusion is “the process of directing the right data sources on the right platform to the right targets at the right times, with the goal of detecting, localizing, identifying, and determining the threat potential of as many targets of interest as possible.” [60]. While a rigorous mathematical tool for solving stochastic multiobject problems is point process theory (see, e.g., [51] and Appendix B, *infra*), a more “engineering-friendly” framework has been advocated by Ronald Mahler in the form of a statistical theory directly based on RST and called *finite-set statistics* (FISST) [66]. The basic idea of FISST is “to transform [a] multisource–multitarget problem into a mathematically equivalent single-sensor, single-target problem. All the sensors are mathematically bundled into a single ‘meta-sensor’ that retains all of the characteristics of the original sensors. (...) The targets are likewise bundled into a single ‘meta-target’ that retains all of the characteristics of the individual targets.” [60] Quoting again from [60], FISST

is engineering-friendly in that it is geometric (i.e., treats multitarget systems as visualizable images); and directly generalizes the Bayes “Statistics 101” formalism that most signal processing engineers already understand — including formal Bayes-statistical modeling methods.

Since its introduction in 1994, FISST has attracted much interest from a number of research areas. Its first applications to digital communication problems were examined in [5, 6, 22].

The main focus of this monograph is on estimation problems where the quantities to be estimated are in a random number. In addition, a model of their evolution may be available with time. A typical problem is that one observes a superposition of a random number $n$ of random signals in additive noise and one wants to estimate this number as
well as certain parameters of the individual signals. In the “classical” framework (see, e.g., [90]), the problem is solved in two steps: (1), determine whether or not signals exist and, if so, to what number, and (2) estimate the parameters of the signals under the assumption that their true number is its estimate \( \hat{n} \). As we shall see in Section 2.2, this approach may not provide the best solution, which depends on the cost function selected for the estimation problem.

The monograph has been organized in six sections and three appendixes. More precisely, Section 2 focuses on the statistical characterization of Random Sets, and in particular of Random Finite Sets (RFS): first the concept of integrating “finite-set functions”, i.e., functions whose arguments are finite sets, as well as the inverse operation, the so-called ”set derivative” amounting to differentiating a set function with respect to a finite set, are introduced and interpreted in an engineering-friendly way. The remainder of Section 2 is concerned with the definition of RFS probability density functions (pdfs), and with the extensions of such concepts as Bayesian recursions and statistical inference to the point where the object of interest is an RFS.

Section 3 is devoted to the problem of reduced-complexity implementation of RFS estimators, and summarizes the major techniques proposed so far in order to scale down the computational burden from combinatorial to algebraic: special attention is paid to Probability Hypothesis Density (PHD) filtering and to its “Cardinalized” form (CPHD), which have been the object of great interest in multiobject tracking for more than a decade now.

Sections 4 and 5 illustrate the application of RFS theory to two relevant problems of Communication Theory, i.e., multiuser detection and channel estimation in dynamic environments. In particular, Section 4 considers the situation where the set of active users varies over time according to a known transition pdf, while the object of interest in Section 5 is the set of active paths in a frequency-selective wireless channel. Finally, Section 6 contains some concluding remarks and a brief list of the topics concerning RST that were not included in this monograph.

Appendix A complements Section 2 by illustrating the mathematical foundations of Random-Set Theory in a more formal way: it is
however important to underline that Section 2 is self-contained, and therefore a thorough understanding of Appendix A is not strictly necessary in order to “operate” with RFSs. Likewise, Appendices B and C are reserved to readers who want to have a deeper understanding of the connections of RFS to well-established “classical” theories, such as Point Processes (Appendix B) and Dempster–Shafer Theory (Appendix C).
In this section we introduce the basic concepts of random-set theory (RST). We skip the mathematical details, which the interested reader can find in Appendix A, and list only the main facts needed to understand the main applications of RST.

2.1 Random Sets, Set Integral, and Set Derivative

Imagine we have a random experiment of which the outcome is a set of which cardinality is a random integer and its elements are random quantities. In the simple example mentioned in Section 1, each random polygon in the plane is modeled by a random number \( n \) of vertices and by an \( n \)-tuple of pairs of real random variables specifying the location of the vertices. More formally, a random set is a random entity that draws its realizations from the space of all subsets of some underlying space \( S \). Thus, an operational (albeit mathematically inaccurate) definition of a random set \( X \) is that of a map between a sample space \( \Omega \) (the collection of all the outcomes of a random experiment) and a family of subsets of \( S \):

\[
X : \Omega \rightarrow \mathcal{F}(S) \tag{2.1}
\]
A central role in the applications of RST is played by hybrid spaces $S$. A hybrid space is a Cartesian product $W \times \mathbb{R}^d$, consisting of all pairs $(w, x)$ with $x \in \mathbb{R}^d$, $d$ a fixed integer, and $w \in W$, where $W$ is a finite set (a simple example has $W = \{1, 2, \ldots, K\}$). In the balance of this monograph we shall mostly deal with hybrid spaces, and use the following terminology: $x$ is called the state, and $w$ the type of the element $(w, x)$ of the random set. As an example, $X$ may denote the set of users in a multiuser transmission system, with vector $x$ including power and geographic location, and $w$ describing the identity of the user and the data it transmits. See Figure 2.1 for an example of a hybrid set, modeling a point randomly located on one of three squares: here $d = 2$, $|W| = 3$. In the rest of the text we shall use the following notation: $\pi(X)$ denotes the discrete part of $X$, that is, the set $\{w_1, \ldots, w_{|\pi(X)|}\}$, while $\pi'(X)$ denotes its continuous part.\footnote{Notice that, in general, $|X| = |\pi'(X)| \geq |\pi(X)|$, almost surely. Actually, the states of two elements of a random set differ almost surely (unless the probability measure is singular), while their types may be equal with nonzero probability.}

Fig. 2.1 Illustration of a hybrid set.
the cumulative distribution function of $X$, with argument $x$, is defined as the probability that $X$ does not exceed value $x$. With random sets, instead of inequality we use inclusion, and we define, as a counterpart to cumulative distribution, a set function called belief function. This is defined as

$$\beta_X(C) \triangleq \mathbb{P}(X \subseteq C).$$

(2.2)

Given a set $C \subseteq \mathcal{F}(\mathbb{S})$, (2.2) defines the probability that the outcome of the random experiment lies in region $C$, i.e., is a subset of $C$.$^2$

The belief function is strongly connected to the standard probability measure. Suppose in fact that the random set we are dealing with is the singleton $X = \{x\}$, $x$ a real vector with a fixed number of components. Applying the definition of belief function (2.2) to this special case, we obtain

$$\beta_X(C) \triangleq \mathbb{P}(X \subseteq C)$$

$$= \mathbb{P}(x \in C)$$

$$= P_x(C)$$

where $P_x(C)$ is the probability measure of $C$ induced by the probability distribution of vector $x$, as we would directly obtain using standard probability theory. Notice, however, that belief functions are not measures (although sometimes they are referred to as such): this can be easily seen by observing that $\beta_X(\emptyset)$ may not be zero (this occurs whenever the empty set has a finite probability of occurrence), or that, while for probability measures $P_x$ we have

$$C_1 \cap C_2 = \emptyset \quad \Rightarrow \quad \mathbb{P}(x \in C_1 \cup C_2) = P_x(C_1) + P_x(C_2)$$

for belief functions we have instead

$$C_1 \cap C_2 = \emptyset \quad \Rightarrow \quad \beta_X(C_1 \cup C_2) \geq \beta_X(C_1) + \beta_X(C_2).$$

$^2$It may be appropriate to observe here that the belief function (2.2) allows one to express the uncertain knowledge of $X$, when the only valid statement about it is that its realization belongs to a given region. This observation is at the root of the fact that RST is a proper tool when it comes to dealing with imperfectly characterized models. More on this can be found, e.g., in [61, 66].
Example 2.1. Consider the set \( X = \{x_1, x_2\} \), with \( x_1 \) and \( x_2 \) two independent random vectors with a fixed number of components. We have, with obvious notations,

\[
\beta_X(C) \triangleq \mathbb{P}(\{x_1, x_2\} \subseteq C) = \mathbb{P}(\{x_1\} \cup \{x_2\} \subseteq C) = \mathbb{P}(\{x_1\} \subseteq C, \{x_2\} \subseteq C) = \mathbb{P}(x_1 \in C, x_2 \in C) = P_{x_1}(C)P_{x_1}(C).
\]

We quickly observe that the calculation of (2.2) does not require an extension of the conceptual framework of probability theory: in fact, we can compute \( \beta_X(C) \) using the decomposition

\[
\beta_X(C) \triangleq \mathbb{P}(X \subseteq C) = \sum_{k=0}^{\infty} \mathbb{P}(|X| = k) \mathbb{P}(X \subseteq C \mid |X| = k)
\]

where in a number of cases the infinite series will be replaced by a finite summation, extended to all possible cardinalities of set \( X \).

Example 2.2. Assume the simple, yet important, case of a random set whose realizations are either the empty set \( \emptyset \) (with probability \( q \)) or the singleton \( \{x\} \), \( x \in \mathbb{R}^d \) (with probability \( 1 - q \)). The corresponding belief function is

\[
\beta_X(C) = \mathbb{P}\{X = \emptyset\} + \mathbb{P}\{x \in C, X \neq \emptyset\} = q + (1 - q)P_X(C)
\]

where \( P_X(C) \) denotes again the probability measure of \( C \) induced by the distribution of \( x \). Now, let us assume \( S \) to be a hybrid set with \( W = \{0,1\} \) and \( d = 1 \), and \( C = \{0\} \times I \), \( I \) an interval of the real line. Assume furthermore that continuous part, say \( x' \), and discrete part, say \( w \), of \( x \) are independent. We have

\[
P_X(C) = \mathbb{P}(w = 0) \int_I p_{x'}(z) dz
\]
or, in a more formal notation,

\[ P_x(C) = \int_C p_{x'}(z)p_w(v) \kappa(v) \, dz \]

\[ = \int_C p_{x'}(z)p_w(v) \lambda(w, z) \]

where \( \kappa(\cdot) \) denotes a counting measure and \( \lambda \) is the product measure.

Just as probability density functions are obtained, in standard probability theory, by differentiating cumulative distributions, in RST we derive set densities by taking a suitably defined derivative of the belief function.\(^3\)

A key point here is that since we are dealing with set functions, we cannot use ordinary differentiation, and hence we need suitable definitions for set derivatives and set integrals. These are two entities related by a version of the fundamental theorem of calculus saying that one is the inverse of the other.

2.1.1 Set Integral

We start with set integrals. Deferring to Appendix A a more rigorous discussion, we define here a set function as a map from \( \mathcal{F}(S) \) to \( \mathbb{R} \), and a finite-set function as a similar map from the collection of finite sets with elements in \( S \) to \( \mathbb{R} \). (Notice that, by its definition, the belief function is in general a set function but not a finite-set function.) Here we confine our attention to the problem of defining integrals of finite-set functions, i.e., of functions defined by

\[ \varphi: Z = \{z_1, \ldots, z_n\} \in \mathcal{F}(S) \rightarrow \varphi(Z) \in \mathbb{R}. \]

Also observe that finite-set functions are inherently symmetric, i.e., invariant to argument permutations, which reflects the unordered nature of sets. Consider first the simple case where \( S = \mathbb{R}^d \), whereby

---

\(^3\)The parallelism between standard probability theory and RST is so close that “general statistical methodologies can, with a bit of prudence, be directly translated [to RST]. Consequently, any sentence (any concept or algorithm) phrased in the random vector language can, in principle, be directly translated into a corresponding sentence (corresponding concept or algorithm) in the random set language. The correspondence between dictionaries is, of course, not precisely one-to-one. For example, vectors can be added an subtracted, whereas finite sets cannot.” [61, p. 377].
$Z = \{x_1, \ldots, x_n\}$ is any unordered collection (of arbitrary cardinality) of $d$-dimensional vectors. Thus, if $S \subseteq \mathbb{R}^d$, the set integral of the function $\varphi(\cdot)$ on the set $S$ is defined as

$$
\int_S \varphi(Z) \delta Z \equiv \sum_{k=0}^{\infty} \frac{1}{k!} \int_{S^k} \varphi(\{x_1, \ldots, x_k\}) \, dx_1 \cdots dx_k
$$

$$
= \varphi(\emptyset) + \int_S \varphi(\{x\}) \, dx + \frac{1}{2} \int_{S \times S} \varphi(\{x_1, x_2\}) \, dx_1 \, dx_2 + \cdots \tag{2.4}
$$

The combinatorial term $1/k!$ in (2.4) requires some explanation. Its presence is caused by the difference between sets and vectors. In a set, the order of its elements has no relevance, while with vectors it counts. Thus, the term $k!$ reflects the number of permutations of the elements of a set which leaves it invariant. Observe also that the infinite series in the definition of set integral is actually a finite summation whenever $\varphi = 0$ identically over sets with cardinality larger than a given finite integer.

The concept of set integral is now generalized to the case of abstract measure spaces. Referring, again, to Appendix A for a more rigorous explanation, here we observe that, after endowing the measurable space $S$ with measure $\lambda$, the set integral of a finite-set function becomes

$$
\int_S \varphi(Z) \delta Z = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{S^k} \varphi(\{x_1, \ldots, x_k\}) \, d\lambda(x_1) \cdots d\lambda(x_k)
$$

where each term of the summation is a Lebesgue integral with respect to the measure $\lambda$.

### 2.1.2 Set Derivative

The intuition behind the concept of a “derivative” is the evaluation of “how fast” a certain function changes when its argument is perturbed. Referring, again, to Appendix A for a more rigorous treatment, we just give an operational definition of set derivative here. Consider a point $x \in S$ and define its “neighborhood” $B_{x, \rho}$, a closed ball centered at $x$ with radius $\rho$. As $\rho$ approaches zero, we can write, for a function $\varphi : S \to \mathbb{R}$,

$$
\Phi(B_{x, \rho}) \equiv \int_{B_{x, \rho}} \varphi(y) \, d\lambda(y) \approx \varphi(x) \, \lambda(B_{x, \rho})
$$
which yields
\[ \varphi(x) \approx \frac{\Phi(B_{x,\rho})}{\lambda(B_{x,\rho})} \]
and justifies the definition
\[ \frac{d\Phi}{d\lambda}(x) \triangleq \lim_{\rho \to 0} \frac{\Phi(B_{x,\rho})}{\lambda(B_{x,\rho})} = \varphi(x). \]
Assuming now that \( C \subseteq S \) and \( \lim_{\rho \to 0} C \cap B_{x,\rho} = \emptyset \), we define the set derivative at point \( x \) the set function given by
\[ \frac{\delta \Phi}{\delta x}(C) \triangleq \lim_{\rho \to 0} \frac{\Phi(C \cup B_{x,\rho}) - \Phi(C)}{\lambda(B_{x,\rho})} \quad (2.5) \]
whenever the limit exists. More generally, if \( X = \{x_1, \ldots, x_m\} \) is a finite subset of \( S \), repeated application of (2.5) allows defining
\[ \frac{\delta \Phi}{\delta X}(C) \triangleq \frac{\delta^m \Phi}{\delta x_1 \cdots \delta x_m}(C). \quad (2.6) \]
Observe that, as shown in Appendix A, the definition in (2.5) can be generalized so that the set derivative can be defined on every subset \( C \). However, it is worthwhile mentioning that the set derivative is generally independent of \( C \) so that (2.6) is a function of the finite set \( X \) only.

Finally, we remark that the limit (2.5) always exists in the relevant case that the set function to be derived is the belief function (2.2) of a RFS (see Appendix A).

### 2.1.3 Integration and Differentiation Rules

To evaluate a set derivative or integral, the direct use of their definitions can be avoided by resorting to some elementary rules of calculus. A few of them are listed in the following [66, Chapter 11]:
\[ \frac{\delta}{\delta Z} \left[ a_1 \varphi_1(C) + a_2 \varphi_2(C) \right] = a_1 \frac{\delta \varphi_1}{\delta Z}(C) + a_2 \frac{\delta \varphi_2}{\delta Z}(C) \quad (2.7a) \]
\[ \int \left[ a_1 \varphi_1(Z) + a_2 \varphi_2(Z) \right] \delta Z = a_1 \int \varphi_1(Z) \delta Z + a_2 \int \varphi_2(Z) \delta Z \quad (2.7b) \]
2.1.4 Belief Density

A central entity in applications of RST theory is the belief density of the random set \( X \). This is the function \( f_X(Z) \) of which the set integral yields the belief function:

\[
\int_C f_X(Z) \delta Z = \mathbb{P}(X \subseteq C) \triangleq \beta_X(C)
\]

or, equivalently, the set derivative of the belief function:

\[
f_X(Z) \triangleq \left. \frac{\delta \beta_X(C)}{\delta Z} \right|_{C=\varnothing}.
\]

A key point here is that the belief density plays the same role as the probability density function does in ordinary detection/estimation theory.

**Example 2.3.** Continuing with Example 2.1, we have the following belief density:

\[
\begin{align*}
f_X(\varnothing) &= \left. \frac{\delta \beta_X}{\delta \varnothing} \right|_{C=\varnothing} = P_{X_1}(\varnothing)P_{X_1}(\varnothing) = 0 \\
f_X(\{x_1\}) &= \left. \frac{\delta \beta_X}{\delta x_1} \right|_{C=\varnothing} = p(x_1)P_{X_2}(\varnothing) + P_{X_1}(\varnothing)p(x_2) = 0 \\
f_X(\{x_2\}) &= \left. \frac{\delta \beta_X}{\delta x_2} \right|_{C=\varnothing} = p(x_2)P_{X_1}(\varnothing) + P_{X_2}(\varnothing)p(x_1) = 0
\end{align*}
\]
2.1 Random Sets, Set Integral, and Set Derivative

$$f_X(\{x_1, x_2\}) = \frac{\delta^2\beta_X}{\delta x_1 \delta x_2} \bigg|_{C=\emptyset} = p(x_2)p(x_1) + p(x_2)p(x_1)$$

$$f_X(\{x_1, x_2, x_3\}) = \frac{\delta^3\beta_X}{\delta x_1 \delta x_2 \delta x_3} \bigg|_{C=\emptyset} = 0$$

and, in general, $f_X(Z)$ is identically zero if $Z$ does not contain exactly two elements.

2.1.5 Independent Sets

Although a rigorous definition of independent random sets will be provided in Appendix A, one can intuitively grasp the independence condition of two random sets $X_1$ and $X_2$ leading to the property

$$\beta_{X_1 \cup X_2}(C) = P(X_1 \cup X_2 \subseteq C)$$

$$= P(X_1 \subseteq C, X_2 \subseteq C)$$

$$= P(X_1 \subseteq C)P(X_2 \subseteq C)$$

$$= \beta_{X_1}(C)\beta_{X_2}(C). \quad (2.8)$$

More generally, if $X = \bigcup_{i=1}^{n} X_i$ is a union of independent random sets, then

$$\beta_X(C) = \prod_{i=1}^{n} \beta_{X_i}(C). \quad (2.9)$$

Using differentiation rule (2.7d), we obtain, for the corresponding density,

$$f_X(Z) = \sum_{Y_1 \cup \cdots \cup Y_n = Z} \prod_{i=1}^{n} f_{X_i}(Y_i) \quad (2.10)$$

where the summation is taken over all mutually disjoint sets $Y_i$, $i = 1, \ldots, n$, of which the union is $Z$. When $n = 2$, we have explicitly

$$f_{X_1 \cup X_2}(Z) = \sum_{Y \subseteq Z} f_{X_1}(Y)f_{X_2}(Z \setminus Y). \quad (2.11)$$

The previous equation defines what is called the generalized convolution of two belief densities.
2.2 Bayesian and Maximum-likelihood Estimates

Consider a vector observation \( y \), which we assume to be statistically connected to the random set \( X \) by the conditional density \( f_{y|X} \). Notice that this is an ordinary pdf, parametrized by a realization of the RFS \( X \).

It is not difficult to verify (see [44]) that the marginal and a posteriori distributions, defined in analogy with standard probability theory on random vectors, as

\[
f_y(y) = \int f_{y|X}(y | S) f_X(S) \delta S
\]

(2.12a)

\[
f_{X|y}(S | y) = \frac{f_{y|X}(y | S) f_X(S)}{\int f_{y|X}(y | C) f_X(C) \delta C}
\]

(2.12b)

respectively, are set densities. Therefore, the density in (2.12b) can be employed for the statistical inference on the RFS \( X \) based on the observation of the random vector \( y \).

Now denote by \( \hat{X} \) the estimate of the RFS \( X \), and consider a nonnegative function \( C(S,B) \) of two finite set variables \( S \) and \( B \): specifically, \( C(S,B) \) is the cost incurred when \( X = S \) and \( \hat{X} = B \). A Bayesian estimator \( \hat{X} = \hat{X}(y) \) is an RFS (i.e., a set-valued function of the random vector \( y \)) obtained by minimizing the average risk \( R \) defined as the expected value of \( C(X, \hat{X}(y)) \):

\[
R = \int\int C(S, \hat{X}(y)) f_{X|y}(S | y) f_y(y) \delta S dy
\]

Thus, the optimum Bayes estimator is the one minimizing the conditional risk given the observations, namely

\[
\hat{X}(y) = \arg\min_Z \int C(S, Z) f_{X|y}(S | y) \delta S.
\]

\footnote{In most RFS literature, the observation consists of a set \( Y = \{y^{(1)},\ldots,y^{(N)}\} \), \( y^{(i)} \subset W \), where \( W \) is the space containing \( N \) observations from as many sensors. In order to keep the discussion as simple as possible, and in view of the relevance, to the applications considered in this monograph, of the so-called “superpositional sensors” model of [66, 67], we confine here our attention to the case of observations consisting of a single random vector, say \( y \), typically belonging to \( \mathbb{R}^N \).}
2.2 Bayesian and Maximum-likelihood Estimates

2.2.1 Choice of the Cost Function, and Set Estimators

A sensible choice of the cost function $C$ must satisfy two criteria, which might be conflicting:

1. The cost function must be relevant for the problem at hand, i.e., significantly express the loss in system performance caused by inaccurate estimates.
2. The resulting minimization of the expected cost should lead to an estimator structure which is simple to implement.

In random set theory, choosing a cost function may not be an easy task: think, for example, of a situation in which we should determine the cost of a wrong choice of the cardinality of set $X$ along with that of an incorrect estimate of its continuous parameters. In practice, we may settle for a “reasonable” set estimator, i.e., one yielding a controlled complexity, and verify how it performs when the most significant performance parameters are evaluated. Three set estimators are especially relevant here.

Again, we focus our attention on hybrid spaces $S = W \times \mathbb{R}^d$, that is, to random sets with elements $(w_i, x_i)$, with $x_i$ a random vector, $w_i$ a random variable taking on a finite number of values, and $i = 1, \ldots, |X|$. We shall use the notation introduced supra in Section 2.1 to denote the discrete and the continuous part of $X$: $\pi(X)$ and $\pi'(X)$, respectively.

A suitable cost may be generated by combining that of a wrong choice of the cardinality with that of an inaccurate estimate of the continuous parameters. Consider first a cost function $C_0$ penalizing a wrong estimate of the cardinality of $X$ while assuming that every wrong choice has the same cost:

$$C_0(X, \hat{X}) \triangleq \begin{cases} 0, & \text{if } |\hat{X}| = |X| \\ 1, & \text{otherwise.} \end{cases}$$

Introduce next a cost function penalizing a wrong estimate of the continuous part of $X$:

$$C_1(X, \hat{X}) \triangleq \begin{cases} 0, & \text{if } |\hat{X}| = |X|, \hat{w}_i = w_{\tau(i)}, \text{ and } \|\hat{x}_i - x_{\tau(i)}\| < \epsilon \\ i = 1, \ldots, |X|, \text{ for some permutation } \tau \\ 1, & \text{otherwise.} \end{cases}$$
The choice of this function implies that no cost is incurred whenever the number of elements of $X$ and their types are estimated correctly, and in addition the estimate of the states of $X$ are “good enough.” Formally, there exists an index permutation $\tau$ such that the estimated state $\hat{x}_i$ belongs to a $d$-dimensional hypersphere of radius $\epsilon$ centered in $x_{\tau(i)}$, reflecting the definition of “closeness,” i.e., the amount of tolerable difference between two state vectors. If any of these conditions is not satisfied, then the cost is 1.

### 2.2.1.1 Set estimator I

With this estimator (often referred to in the literature as GMAP-I [44]), the cardinality of $X$ is estimated first as $\hat{n}$. Next, under the assumption that $\hat{n}$ is the correct cardinality, the remaining parameters of $X$ are estimated. In formulas, this two-step estimator does first

\[
\hat{n} \triangleq \arg\max_{n} P(|X| = n \mid y) = \arg\max_{n} \int_{|Z| = n} f_{X|Y}(Z \mid y) \delta Z \tag{2.13}
\]

and then

\[
\hat{X}_I = \begin{cases} 
\arg\max_{Z : |Z| = \hat{n}} f_{X|Y}(Z \mid y), & \text{if } \hat{n} \geq 1 \\
\emptyset, & \text{if } \hat{n} = 0.
\end{cases} \tag{2.14}
\]

This two-step estimator looks quite natural, as it estimates $n$ first, and then, assuming that this is the correct cardinality of $X$, transforms the estimation problem in a standard, i.e., non-random-set-theoretical, one. However, in terms of risk, this estimator turns out to yield a minimum of the expected cost function $C_0$ [44, pp. 192–194], one which associates no cost to the inaccuracy in the estimation of the continuous parameters of $X$.

### 2.2.1.2 Set estimator II

A single-step estimator has

\[
\hat{X}_{II} = \max_{n, Z : |Z| = n} f_{X|Y}(Z \mid y) \frac{e^n}{n!}.
\]
This minimizes the risk associated with the cost function $C_0 + C_1$, as shown in [44, pp. 192–194].

This set estimator, called GMAP-II, can also be implemented in two steps, running as follows: for all $n$, compute

$$
\hat{X}_n \triangleq \begin{cases} 
\arg \max_{Z:|Z|=n} f_{X|y}(Z \mid y), & \text{if } n \geq 1 \\
\emptyset, & \text{if } n = 0.
\end{cases} 
$$

(2.15)

and then

$$
\hat{n} \triangleq \arg \max_n f_{X|y}(\hat{X}_n \mid y) \frac{\epsilon^n}{n!}.
$$

(2.16)

Then,

$$
\hat{X}_{II} = \hat{X}_{\hat{n}}.
$$

(2.17)

### 2.2.1.3 Set estimator III

The two previous estimators were developed under no assumption on $|\pi(X)|$, which does not allow separating the costs for the estimation of state and type of a random set. Now, in the applications we are considering in this monograph (see especially Sections 4 and 5), each element of the random set has, almost surely, a different type,\(^5\) and hence the following condition holds:

$$
|\pi(X)| = |\pi'(X)| = |X|.
$$

(2.18)

We introduce a third estimator under (2.18) called GMAP-III that is aimed at minimizing the risk associated with the cost function

$$
C(X, \hat{X}) \triangleq \begin{cases} 
\sum_{i=1}^{X} g(x_{\tau(i)}, \hat{x}_i), & \text{if } |\hat{X}| = |X| \text{ and } \hat{w}_i = w_{\tau(i)} \\
Q, & \text{otherwise}.
\end{cases}
$$

This assigns a cost $Q$ to any wrong estimate of the type of the random set. If this estimate is correct, the cost is a function of the discrepancies

\(^5\)For further persuasion, consider the example of a random set of users in a multiuser transmission system of which the types are the individual user identities and the states their geographic locations.
between estimated and true states. The actual selection of \( Q \) and of \( g(\cdot, \cdot) \) reflects the relative weight of an error in the discrete and the continuous parameters. The risk associated with this cost function is

\[
E[C(X, \hat{X})] = Q \mathbb{P}\left( \{ \pi(\hat{X}) \neq \pi(X) \} \right) \\
+ \mathbb{E}\left[ \sum_{i=1}^{\|X\|} g(\tau_i, \hat{x}_i) \big| \pi(\hat{X}) = \pi(X) \right] \times \mathbb{P}(\pi(\hat{X}) = \pi(X)) \tag{2.19}
\]

where, in the second line, \( \tau \) is the permutation which ensures \( \hat{w}_i = w_{\tau(i)} \) for all \( i \). The resulting Bayesian estimator requires the minimization of the sum in (2.19). Besides the necessity of appropriately selecting \( Q \) and \( g(\cdot, \cdot) \) (which might not be a straightforward task), computation of the minimum may involve an unacceptable estimator complexity. For this reason, we define a suboptimum, simpler version of the estimator, one that minimizes separately the two terms and is realized by a two-step procedure. The first step minimizes the probability \( \mathbb{P}(\pi(\hat{X}) \neq \pi(X)|y) \). This is obtained through a maximum a posteriori probability (MAP) estimator of \( \pi(X) \), which maximizes the conditional pdf of \( X \) by assuming the continuous parameters of \( X \) as nuisance. In the second step, the term involving \( \sum_{i=1}^{\|X\|} g(\tau_i, \hat{x}_i) \) is minimized. If \( g(\cdot, \cdot) \) is a quadratic function, this second minimization yields the conditional expectation

\[ \hat{x}_i = \mathbb{E}[\tau_i | y], \quad i = 1, \ldots, \|X\|. \]

We denote the resulting estimator by \( \hat{X}_{III} \).

**Maximum-likelihood estimation:** maximum-likelihood (ML) estimates (which are not Bayesian) can also be defined in random-set theory: we have\(^7\)

\[ \hat{X}_{ML} = \arg\max_C f_{y|X}(y|C). \]

\(^6\)Note that estimator \( \hat{X}_{III} \), unlike those previously described, relies upon the condition \( \|X\| = \|\pi(X)\| \), i.e., on the requirement that all elements of \( X \) have different types.

\(^7\)Unlike ML estimates, maximum a posteriori (MAP) estimates may be not defined in random-set theory [66, Section 14.5]. Indeed, set densities may have units of measurement (if so, so does the state space), whereby a posteriori densities corresponding to different set cardinalities may be incommensurable (for a detailed discussion of this point, see [66, 87, Section 14.5]). However, if the state space has no units of measurement (which occurs in the applications considered here), then MAP estimators can be defined in the usual way.
2.2 Bayesian and Maximum-likelihood Estimates

2.2.2 A Simple Example

In this example, we examine a random finite set $X$ whose realizations are the empty set, a singleton, and a doubleton. This is defined on $S = \mathbb{R}$ as

$$X = (\{x_1\} \cap D_1) \cup (\{x_2\} \cap D_2)$$

where:

- $x_1, x_2$ are independent, Gaussian random variables, with unit variance and mean $\mu_i, i = 1, 2$;
- $D_1, D_2$ are independent random sets, independent of $x_1$ and $x_2$, defined as

$$D_i = \begin{cases} \mathbb{R}, & \text{with probability } p \\ \emptyset, & \text{with probability } 1 - p \end{cases}, \quad i = 1, 2.$$

The belief function of $X$ takes the form

$$\beta_X(C) = \mathbb{P}(\{x_1\} \cap D_1 \subseteq C)\mathbb{P}(\{x_2\} \cap D_2 \subseteq C)$$

$$= \left(1 - p + \frac{p}{\sqrt{2\pi}} \int_C e^{-\frac{(z - \mu_1)^2}{2}} \, dz\right)$$

$$\times \left(1 - p + \frac{p}{\sqrt{2\pi}} \int_C e^{-\frac{(z - \mu_2)^2}{2}} \, dz\right)$$

$$= (1 - p)^2$$

$$+ \frac{p(1 - p)}{\sqrt{2\pi}} \int_C \left(e^{-\frac{(z - \mu_1)^2}{2}} + e^{-\frac{(z - \mu_2)^2}{2}}\right) \, dz$$

$$+ \frac{p^2}{2\pi} \int_{C^2} e^{-\frac{(z_1 - \mu_1)^2 + (z_2 - \mu_2)^2}{2}} \, dz_1 \, dz_2$$

where (2.8) has been exploited in the first line. The belief function can also be expressed as a set integral, in which case we have

$$\beta_X(C) = \int_C f_X(Z) \delta Z$$

$$= f_X(\emptyset) + \int_C f_X(\{z\}) \, dz + \frac{1}{2} \int_{C^2} f_X(\{z_1, z_2\}) \, dz_1 \, dz_2$$
where the set density $f_X$ is given by

$$f_X(Z) = \begin{cases} 
(1-p)^2, & \text{if } Z = \emptyset \\
\frac{p(1-p)}{\sqrt{2\pi}} \left( e^{-\frac{(z_1-\mu_1)^2}{2}} + e^{-\frac{(z_2-\mu_2)^2}{2}} \right), & \text{if } Z = \{z\} \\
\frac{p^2}{2\pi} \left( e^{-\frac{(z_1-\mu_1)^2+(z_2-\mu_2)^2}{2}} + e^{-\frac{(z_2-\mu_1)^2+(z_1-\mu_2)^2}{2}} \right), & \text{if } Z = \{z_1, z_2\} \\
0, & \text{otherwise.}
\end{cases}$$

Suppose now that the following noisy observation of $X$ is taken

$$y = h(X) + n$$

where $n$ is a two-dimensional vector of independent, standard, Gaussian random variables, independent of $X$, and

$$h(Z) = \begin{cases} 
z(\theta 1 - \theta)^T, & \text{if } Z = \{z\} \\
z_1(\theta 1 - \theta)^T + z_2(1 - \theta \theta)^T, & \text{if } Z = \{z_1, z_2\} \text{ with } z_1 \neq z_2 \\
(0 0)^T, & \text{otherwise}
\end{cases}$$

$\theta$ being a binary, uniform random variable taking on values in $\{0, 1\}$, independent of $n$ and $X$. In other words, the elements of $X$ can be observed indifferently on the first or on the second channel of the measurement $y$. In order to extract an estimate of $X$ from $y$ the likelihood function $f_{y|X}$ or the a posteriori density $f_{X|y}$ are needed. In this scenario, the likelihood function takes the form

$$f_{y|X}(y|Z) = \begin{cases} 
\frac{1}{2\pi} e^{-\frac{y_1^2+y_2^2}{2}}, & \text{if } Z = \emptyset \\
\frac{1}{4\pi} e^{-\frac{(y_1-z_1)^2+y_2^2}{2}} + \frac{1}{4\pi} e^{-\frac{(y_2-z_1)^2}{2}}, & \text{if } Z = \{z\} \\
\frac{1}{4\pi} e^{-\frac{(y_1-z_1)^2+(y_2-z_2)^2}{2}} \\
+ \frac{1}{4\pi} e^{-\frac{(y_1-z_2)^2+(y_2-z_1)^2}{2}}, & \text{if } Z = \{z_1, z_2\} \text{ with } z_1 \neq z_2 \\
0, & \text{otherwise}
\end{cases}$$
and, then, the a posteriori density is

\[
\begin{align*}
\text{if } Z = \emptyset &\quad \frac{(1-p)^2}{2\pi} e^{-\frac{y_1^2+y_2^2}{2}}, \\
\text{if } Z = \{z\} &\quad \frac{p(1-p)}{\sqrt{2\pi}} \left( e^{-\frac{(z-\mu_1)^2}{2}} + e^{-\frac{(z-\mu_2)^2}{2}} \right) \\
&\quad \times \frac{1}{4\pi} \left( e^{-\frac{(y_1-z)^2+y_2^2}{2}} + e^{-\frac{y_1^2+(y_2-z)^2}{2}} \right), \\
\text{if } Z = \{z_1, z_2\}, \text{ with } z_1 \neq z_2 &\quad 0,
\end{align*}
\]

where the term \( f_y(y) \) in (2.12b) has been skipped. The a posteriori distribution of the cardinality of \( X \) can be found as

\[
\mathbb{P}(|X| = n \mid y) = \int_{|Z|=n} f_{X|Y}(Z \mid y) \delta Z
\]

whereby we have

\[
\mathbb{P}(|X| = n \mid y) \propto \begin{cases} 
\frac{(1-p)^2}{\sqrt{2\pi}} e^{-\frac{y_1^2+y_2^2}{2}}, & \text{if } n = 0 \\
\frac{p(1-p)}{4\pi\sqrt{2}} \left( e^{-\frac{(y_1-\mu_1)^2+2y_2^2}{4}} + e^{-\frac{(y_1-\mu_2)^2+2y_2^2}{4}} \\
+ e^{-\frac{2y_1^2+(y_2-\mu_1)^2}{4}} + e^{-\frac{2y_1^2+(y_2-\mu_2)^2}{4}} \right), & \text{if } n = 1 \\
\frac{p^2}{8\pi} \left( e^{-\frac{(y_1-\mu_1)^2+(y_2-\mu_2)^2}{4}} \\
+ e^{-\frac{(y_1-\mu_2)^2+(y_2-\mu_1)^2}{4}} \right), & \text{if } n = 2 \\
0, & \text{otherwise.}
\end{cases}
\]
At this point, the family of estimators in (2.15) takes the form

\[
\hat{X}_0 = \emptyset
\]

\[
\hat{X}_1 = \arg\max_z \left( e^{-\frac{(z-\mu_1)^2}{2}} + e^{-\frac{(z-\mu_2)^2}{2}} \right)
\]

\[
\times \left( e^{-\frac{(y_1-z)^2+y_2^2}{2}} + e^{-\frac{(y_1-\mu_1)^2+(y_2-\mu_2)^2}{2}} \right)
\]

\[
\hat{X}_2 = \arg\max_{z_1, z_2: z_1 \neq z_2} \left( e^{-\frac{(z_1-\mu_1)^2+(z_2-\mu_2)^2}{2}} + e^{-\frac{(z_2-\mu_1)^2+(z_1-\mu_2)^2}{2}} \right)
\]

\[
\times \left( e^{-\frac{(y_1-z_1)^2+(y_2-z_2)^2}{2}} + e^{-\frac{(y_1-\mu_1)^2+(y_2-\mu_2)^2}{2}} \right)
\]

so that the set estimator I in (2.13) and (2.14) becomes

\[
\hat{n}_I = \arg\max_{n \in \{0, 1, 2\}} \mathbb{P}(|X| = n | y)
\]

\[
\hat{X}_I = \hat{X}_{\hat{n}_I}
\]

while the set estimator II in (2.16) and (2.17) becomes

\[
\hat{n}_{II} = \arg\max_{n \in \{0, 1, 2\}} f_{X|y}(\hat{X}_n | y) \left( \frac{\epsilon^n}{n!} \right)
\]

\[
\hat{X}_{II} = \hat{X}_{\hat{n}_{II}}.
\]

Suppose that, for example, \(\mu_1 = 1, \mu_2 = 4\) and \(p = 1/2\), the observation \(y = (1.4 \ 3.4)^T\) has been made, and that the set estimator I is being used. Then the a posteriori distribution of the cardinality of \(X\) is

\[
\mathbb{P}(|X| = n | y) = \begin{cases} 0.75\% & \text{if } n = 0 \\ 39.75\% & \text{if } n = 1 \\ 59.50\% & \text{if } n = 2 \end{cases}
\]

so that \(\hat{n}_I = 2\). Finally, \(\hat{X}_I = \hat{X}_2\) and, from Figure 2.2, we get

\[
\hat{X}_I = \{1.27, 3.71\}.
\]
2.3 Dynamic Models and Bayesian Recursions

In causal dynamic set estimation, a random set at time $t$, denoted $X_t$, is modeled using its a posteriori belief density given the observations up to time $t$, which we denote $f_{X_t|y_1:t}$. A common method for the evaluation of the above density relies on Bayesian recursions. To derive these, we make the Markovian assumption on the evolution of $X_t$, i.e., the assumption that $X_t$ depends on its past only through $X_{t-1}$. Formally, we have

$$f_{X_t|X_{t-1}, X_{t-2}, \ldots, X_1} = f_{X_t|X_{t-1}} \quad (2.21)$$

where we have omitted the arguments for notational simplicity, and we have defined the conditional belief densities in (2.21) as the set derivatives of appropriate conditional belief functions.

The Bayesian recursions take the form

$$f_{X_t|y_1:t-1}(C \mid y_{1:t-1}) = \int f_{X_t|X_{t-1}}(C \mid B) \times f_{X_{t-1}|y_1:t-1}(B \mid y_{1:t-1}) \delta B \quad (2.22a)$$

$$f_{X_t|y_1:t}(C \mid y_{1:t}) \propto f_{y_t|X_t}(y_t \mid C) f_{X_t|y_1:t-1}(C \mid y_{1:t-1}) \quad (2.22b)$$

Equations (2.22a) and (2.22b) describe the “predict” and “update” step, respectively. The proportionality relationship (2.22b) can be
transformed in an equality, if needed, by dividing the Right-Hand Side (RHS) by the “Bayesian normalization factor”

\[ f_{y_t|y_{1:t-1}}(y_t \mid y_{1:t-1}) = \int f_{y_t|B}(y_t \mid X_t) f_{X_t|y_{1:t-1}}(B \mid y_{1:t-1}) \delta B. \]

### 2.3.1 Useful Dynamic Models

Simple, yet useful, dynamic models can be constructed under the assumption that the random set \( X_t \) evolves by having some of its elements disappear and some appear independently. Specifically, assume that at time \( t \) the random set \( X_t \) is the union of two conditionally independent sets, viz., \( S_t \), the “survivors’ set” containing the new states of elements that were already present in \( X_{t-1} \), and \( N_t \), the “new” elements that were not in \( X_{t-1} \). “Deaths” and “births” are assumed independent, and are subject to the constraint that an element of \( X_{t-1} \) perishing between \( (t - 1) \) and \( t \) cannot reenter \( X_t \) as a newborn. Formally, we have the conditions

\[ X_t = S_t \cup N_t, \quad S_t = \bigcup_{x \in X_{t-1}} S_{t|t-1}(x), \quad \text{and} \]

\[ \beta_{X_t|X_{t-1}}(C \mid X_{t-1}) = \beta_{S_t|X_{t-1}}(C \mid X_{t-1}) \beta_{N_t|X_{t-1}}(C \mid X_{t-1}) \]

where \( S_{t|t-1}(x) \) is the singleton incorporating the state at \( t \) of the survivor whose state at \( t - 1 \) was \( x \) (see Figure 2.3 for an illustration).

With this model we obtain, from (2.10)–(2.11),

\[ f_{X_t|X_{t-1}}(C \mid B) = \sum_{Z \subseteq C} f_{S_t|X_{t-1}}(Z \mid B) f_{N_t|X_{t-1}}(C \setminus Z \mid B) \]

which shows how the dynamic system can be modeled once the random sets \( S_t \) and \( N_t \) are characterized.

### 2.3.2 Modeling \( S_t \) and \( N_t \)

Suppose that \( X_{t-1} \) contains \( n \) elements: \( X_{t-1} = \{x_{t-1}^{(1)}, \ldots, x_{t-1}^{(n)}\} \), all statistically independent. Then the random set \( S_t \) of elements of \( X_{t-1} \) still present in \( X_t \) can be represented in the form

\[ S_t = \bigcup_{i=1}^{n} X_t^{(i)} \quad (2.23) \]
2.3 Dynamic Models and Bayesian Recursions

Fig. 2.3 Illustration of a dynamic model: at epoch $t$ new elements appear in $N_t$, and some elements in $X_{t-1}$ survive and evolve in $S_t$, while the rest die.

where the components $X_t^{(i)} \triangleq \{x_t\} \cap D_t^{(i)}$ are either empty sets or singletons, as determined by the random set $D_t^{(i)}$ defined as follows:

$$D_t^{(i)} \triangleq \begin{cases} S, & \text{with probability } q \\ \emptyset, & \text{with probability } 1 - q \end{cases} \quad (2.24)$$

Here $D_t^{(i)}$ and $\{x_t\}$ are independent, and $q$ is the probability of the survival of an element (we have assumed that $q$ is the same for all elements — if this assumption does not hold, one should write $q(x_t^{(i)})$ instead).

The conditional belief function of $X_t$ has the form, which can be derived from previous examples in this section,

$$\beta_{X_t^{(i)} | X_{t-1}^{(i)}} (C \mid b) = \mathbb{P}((x_t^{(i)}) \cap D_t^{(i)} \subseteq C \mid x_{t-1}^{(i)} = b)$$

$$= \mathbb{P}(D_t^{(i)} = \emptyset \mid x_{t-1}^{(i)} = b)$$

$$+ \mathbb{P}(D_t^{(i)} = S, x_{t-1}^{(i)} \subseteq C \mid x_{t-1}^{(i)} = b)$$

$$= (1 - q) + q \mathbb{P}(x_t^{(i)} \in C \mid x_{t-1}^{(i)} = b)$$

where $\mathbb{P}(x_t^{(i)} \in C \mid x_{t-1}^{(i)} = b)$ is the single-element dynamic model. The corresponding belief density is

$$f_{X_t^{(i)} | X_{t-1}^{(i)}} (Z \mid b) = \begin{cases} 1 - q, & \text{if } Z = \emptyset \\ q f_{x_t^{(i)} | x_{t-1}^{(i)}} (z \mid b), & \text{if } Z = \{z\} \\ 0, & \text{if } |Z| > 1 \end{cases}$$
where \( f_{x_t|i|x_{t-1}^i} (Z|b) \) is the (standard) density of the single-element dynamic model \( \mathbb{P}(x_t^i \in C|x_{t-1}^i = b) \). The conditional belief density of \( S_t \) is obtained by first observing that, due to the assumption of independence of the random sets \( X_t^i \), from (2.9) and (2.23) we have

\[
\beta_{S_t|x_{t-1}}(C | \{ b^{(1)}, \ldots, b^{(n)} \}) = \prod_{i=1}^{n} \beta_{X_t^i|x_{t-1}^i}(C | x_{t-1}^i = b^{(i)}).
\]

The corresponding belief density can be obtained as shown in Appendix A:

\[
f_{S_t|x_{t-1}}(Z | \{ b^{(1)}, \ldots, b^{(n)} \}) = \begin{cases} 
q|Z| (1-q)^{n-|Z|} \sum_{\pi} \prod_{i=1}^{n} f_{x_t^i|x_{t-1}^i}(z_i|b^{(\pi(i))}) & \text{if } |Z| \leq n \\
0, & \text{otherwise}
\end{cases}
\]

where the summation runs through all the permutations \( \pi \) of the integers 1, \ldots, n.

As for the birth process \( N_t \), several models have been proposed. In tracking applications, where the single objects composing the multiobjects are not individually identifiable and trackable, the types are often modeled as statistically independent, and so are the respective states, which results in \( N_t \) being statistically independent of \( X_{t-1} \). In particular, for bounded random sets with \( |X_t| \leq K \) and independent, identically distributed elements with common pdf \( f_n(\cdot) \), a model is the Bernoulli random set

\[
f_{N_t}(C) = f_{N_t|x_{t-1}}(C|B) = \frac{K!}{(K-|C|)!} \alpha^{|C|} (1-\alpha)^{K-|C|} \prod_{c \in C} f_n(c) \tag{2.25}
\]

for \( |C| \leq K \), where \( \alpha \) denotes the “reproduction rate.” Notice that, if \( K \to \infty \) and \( \alpha = \mu/K \) (and hence \( N_t \) is an unbounded random set) we obtain the “Poisson model” (see Appendix A)

\[
f_{N_t}(C) = \mu^{|C|} e^{-\mu} \prod_{c \in C} f_n(c). \tag{2.26}
\]
2.3.3 Identifiable objects

We say that the elements of a multiple object are identifiable if each one of them has a unique type, so that condition (2.18) applies, and

\[ \pi(X_{t-1}) \cap \pi(N_t) = \emptyset \]

almost surely. In this situation, only conditional independence between \( S_t \) and \( N_t \) (given \( X_{t-1} \)) can be claimed. However, \( \pi(X_{t-1}) \) determines a partition of \( S \) into two subsets, say \( S_1[\pi(X_{t-1})] \) and \( S_2[\pi(X_{t-1})] \), whereby, conditioned upon \( X_{t-1} \), \( S_t \) and \( N_t \) are disjoint and independent RFSs on \( S_1[\pi(X_{t-1})] \) and \( S_2[\pi(X_{t-1})] \), respectively. The conclusions of Example A.19 apply, see Appendix A, and the transition density simplifies to

\[
f_{X_t|x_{t-1}}(C \mid B) = f_{S_t|x_{t-1}}(C \cap S_1[\pi(B)] \mid B) \\
\times f_{N_t|x_{t-1}}(C \setminus C \cap S_1[\pi(B)] \mid B).
\]
The implementation of Bayes recursions for a “standard” dynamical system, whose state is a vector evolving according to a known transition density, is *per se* computationally intensive, except in very few instances, like the one involving a linear-Gaussian model and leading to the well-known Kalman-filtering equations. The task becomes prohibitive when the object of interest is actually a multiobject consisting of a randomly time-varying number of elementary objects, i.e., a dynamical system whose state is an RFS. In fact, RFS theory has been advocated and applied so far primarily in multtarget, multisensor tracking, wherein an unknown number of manoeuvering targets should be tracked based on a set of observations coming from several sensors, which might in turn not be identifiable and may operate at very low signal-to-noise ratios. These sensors, whose number may be unknown, may decide to forward a (possibly quantized) version of their observations to a fusion center, whose task is to estimate the “multitarget state,” i.e., the realization of the random set containing the information of interest. In order to keep the discussion as simple as possible, and in view of the relevance, to wireless communications and information theory applications of the so-called “superpositional sensor” model [66, 67]
into account, the applications discussed in this monograph assume that the observations consist, at time $t$, of a simple random vector $y_t$ (typically, $y_t \in \mathbb{R}^N$). However, most of the procedures discussed in this section allow the approximation of Bayesian recursions in a more general case, viz., when the observations coming from all sensors at epoch $t$ are collected in an observation set $Y_t$. Therefore, with some notational abuse, in this section we will denote by $y_t^{(i)} \in \mathbb{W}$ the measurement from sensor $i$ at time $t$, with $\mathbb{W}$ the observation space (typically $\mathbb{R}^M$), so that $Y_t = \{y_t^{(1)}, \ldots, y_t^{(N_t)}\}$ is the observation random set where $N_t$ is the number of active sensors at time $t$. In this framework, the general entity of interest is the a posteriori density $f_{X_t|Y_{1:t}}$, which can be evaluated through the Bayesian recursions if $X_t$ is a Markovian random set.

### 3.1 Solving the Bayesian Recursion

For the sake of clarity, we briefly recall here the model we shall be focusing upon. The random set of interest $X_t$ is assumed to be defined upon the hybrid space $S = W \times \mathbb{R}^d$, and the reference measure is the product of the Lebesgue and of the counting measure. Throughout this section, we use the same notation for a random vector/set and its realization, whenever this does not generate confusion. This convention will keep the number of different symbols we use to a minimum. The RFS $X_t$ may be represented as $X_t = S_t \cup N_t$, where $S_t$ is the set of the elements surviving from $X_{t-1}$ and $N_t$ the set of new elements in $X_t$. To simplify the exposition, no spawning from elements in $X_{t-1}$ takes place, although, in principle, it could be taken into account rather straightforwardly [64, 66, 86]. Every element $x \in X_{t-1}$ can survive or die with probability $q_t(x)$ and $1 - q_t(x)$ respectively\(^1\): $f_{x_t|x_{t-1}}$ denotes the conditional probability density function that an element, active at time $t - 1$, experiences a state transition from $x_{t-1}$ to $x_t$. Notice that Markovity is ensured in any case, because the survivors set $S_t$ is modeled as conditionally independent of $X_{t-2}$ given $X_{t-1}$, and so is the set $N_t$. As for the observation set, we assume, as anticipated, that the RFS $Y_t$ is observed, and denote $f_{Y_t|x_t}$ as its conditional density.

\(^1\)Here $\{q(x_{t-1}), 1 - q(x_{t-1})\}$ is a probability distribution parameterized by $x_{t-1}$. 
Under these assumptions, the Bayesian recursions to be solved for inference are

\[ f_{t|t-1}(C \mid Y_{1:t-1}) = \int f_{X_t|X_{t-1}}(C \mid B) f_{t|t-1}(B \mid Y_{1:t-1}) \delta B \] (3.1a)

\[ f_{t|t}(C \mid Y_{1:t}) = \frac{f_{Y_t|X_t}(Y_t \mid C) f_{t|t-1}(C \mid Y_{1:t-1})}{\int f_{Y_t|X_t}(Y_t \mid B) f_{t|t-1}(B \mid Y_{1:t-1}) \delta B} \] (3.1b)

where the shorthand notation \( f_{t|k} \) is used for the posterior density \( f_{X_t \mid Y_{1:k}} \). Once this posterior density is obtained, one of the procedures described in Section 2.2 can be used to estimate the set \( X_t \).

Looking back at the general form of Bayesian recursions, the prediction step in (3.1a) requires on-line computation of a set integral, which can be done using the rules outlined in Section 2 and Appendix A. In principle, this amounts to the numerical evaluation of infinitely many integrals, which is out of the question. On the other hand, even for bounded random sets, whose cardinality admits a known upper bound, say \( |X_t| \leq M, \forall t \), and such that (3.1a) simplifies to

\[ f_{t|t-1}(C \mid Y_{1:t-1}) = \sum_{m=0}^{M} \frac{1}{m!} \int_{\mathcal{S}^m} f_{X_t|X_{t-1}}(C \mid \{b_1, \ldots, b_m\}) \times f_{t-1|t-1}(\{b_1, \ldots, b_m\} \mid Y_{1:t-1}) d\lambda(b_1) \cdots d\lambda(b_m) \]

direct computation of the above recursions requires a massive computational effort. In general, indeed, the elements of \( X_t \) cannot be individually tracked, whereby the event \( \{X_t = C\} \) simply means that the ordered sequence \( x_t^{(1)}, \ldots, x_t^{(|C|)} \) is equal (modulo an arbitrary permutation) to the elements of the set \( C \). Likewise, the event

\[ \{X_{t-1} = B\} \rightarrow \{X_t = C\} \]

means that an unordered collection of \( |B| \) vectors has experienced a transition to an unordered collection of \( |C| \) vectors. Yet, we do not know:

a. Which elements of \( X_t \) are surviving from \( X_{t-1} \), and which are newly born.
b. How to associate each single-object initial state to each single-object final state.

For example, following [83], we have for independent object trajectories and independent individual death and birth events:

\[
f_{S_{t}}|X_{t-1}(\{c_{1}, \ldots, c_{|C|}\} | \{b_{1}, \ldots, b_{|B|}\}) = \begin{cases} 
\frac{1}{(|B| - |C|)!} \sum_{\pi} \left( \prod_{i=1}^{|C|} q_{t}(b_{\pi(i)}) f_{X_{t}}|X_{t-1}(c_{i} | b_{\pi(i)}) \right) 
\times \prod_{i=|C|+1}^{|B|} (1 - q_{t}(b_{\pi(i)})) \right), & \text{if } |C| \leq |B| \\
0, & \text{if } |C| > |B| 
\end{cases}
\]

(3.2)

where \( \pi \) runs over all of the possible permutations of the integers 1, 2, ..., |B|. Since \( X_{t} = S_{t} \cup N_{t} \), under conditional independence between \( S_{t} \) and \( N_{t} \), the transition density to be inserted in (3.1a) results from the generalized convolution between (3.2) and the assumed conditional density of \( N_{t} \). Even very simple models for the multi-object set result into the need for (numerical) evaluation of a number of integrals. A slight simplification of the computational burden occurs when each individual object has a unique label, since this allows us to exploit the condition \( \pi(X_{t-1}) \cap \pi(N_{t}) = \emptyset \). In fact, this situation would ultimately be reduced to assuming that \( X_{t-1} \in \mathcal{F}(S_{1}) \) and \( N_{t} \in \mathcal{F}(S_{2}) \), where \( S_{1} \) and \( S_{2} \) are disjoint subsets of \( S \), and hence the results of Example A.19, in Appendix A can be used, and the transition density simplifies. Situations like this occur with the applications studied in Sections 4 and 5.

A number of techniques, available from the literature and to be described in the rest of this section, allow the problem complexity to be reduced, leading to implementable versions of (3.1). A popular, “turn-of-the-crank” approach consists of using sequential Monte-Carlo filtering, which was originally conceived for traditional non-linear filtering problems and may be adapted to ease numerical evaluation of the integrals in (3.1). A theoretically more sophisticated, yet definitely simpler approach relies on the concept of Probability Hypothesis Density (PHD) filtering, which takes into account the fact that both
the object of interest and the observations are RFSs. Finally, under certain circumstances (particularly when each single object is identifiable through a unique label), the complexity problem can be solved by combining Sphere-Decoding algorithms with gridding algorithms. These three approaches will be illustrated in the sequel.

### 3.2 Particle Filtering/Sequential Monte-Carlo

The main problem with Bayesian recursive filtering (3.1) is the lack of a closed form. In fact, for stochastic systems, not even the linear–Gaussian case admits a closed form [16], which calls for approximate methods. The simplest among these is to discretize the continuous parameters into a fixed grid. The main difficulty with this approach is that, if a reasonable accuracy is sought, dense gridding is needed, which may easily yield an unmanageable computational burden. A more efficient approach relies on Sequential Monte-Carlo (SMC), or “particle filtering,” methods (see, e.g., [37] and references therein). Suppose that we have a set of random samples \( \{X_i \} \), each associated with a weight \( \{w_i \} \), such that the sample–weight pairs (or “particles”) represent the density \( f_{t|t} \) in the form

\[
f_{t|t}(C | Y_{1:t}) \approx \sum_{i=1}^{N} w_i \delta_{X_i}(C)
\]

where \( \delta_X \) is the Dirac measure centered at \( X \). SMC methods provide rules for propagating and updating these samples and weights to obtain a set of new particles \( \{X_i, w_i \} \) representing posterior density \( f_{t+1|t+1} \) in the form

\[
f_{t+1|t+1}(C | Y_{1:t+1}) \approx \sum_{i=1}^{N} w_i \delta_{X_i}(C).
\]

Among the various SMC methods, the bootstrap filter [45] has commanded considerable attention for its simplicity. With it, particles

---

2The only case of hybrid systems admitting an optimum closed-form expression seems to be the jump-Markov linear-Gaussian system [39]. Yet, this expression is not a recursive one, and the optimum rule has a complexity which grows exponentially with the frame length.
are recursively updated as:

\begin{align}
X_{t+1}^{(i)} &\sim f_{X_{t+1}|X_t}(C \mid X_t^{(i)}), \quad i = 1, 2, \ldots, N \quad (3.4a) \\
\hat{w}^{(i)}_{t+1} &= w^{(i)}_t f_{Y_{t+1}|X_{t+1}}(Y_{t+1} \mid X_{t+1}^{(i)}) \quad (3.4b) \\
w^{(i)}_{t+1} &= \frac{\hat{w}^{(i)}_{t+1}}{\sum_{j=1}^{N} \hat{w}^{(j)}_{t+1}} \quad (3.4c)
\end{align}

where \( \sim \) means that \( X_{t+1}^{(i)} \) is sampled according to the specified distribution. A common problem with (3.4) stems from the degeneracy phenomenon, which occurs when all but one particle have negligible weights after few iterations [38]: a brute-force approach to reducing its effect consists of using a very large \( N \), which increases the computational effort. A more refined technique relies upon adaptive resampling, which consists of eliminating particles that have small weights, and concentrating on particles with large weights whenever the degeneracy becomes relevant. An indicator of this relevance is the effective sample size, \( N_{\text{eff}} \triangleq \frac{1}{\sum_{i=1}^{N} (w^{(i)}_t)^2} \). From (3.4), \( N_{\text{eff}} \in [1, N] \), with a small \( N_{\text{eff}} \) suggesting degeneracy (in practice, smaller than a threshold \( N_{\text{th}} \) typically chosen as \( N/4 \) or \( N/2 \) in most applications). The resampling step involves generating a new set of particles \( \{\tilde{X}_{t+1}^{(i)}, \tilde{w}^{(i)}_{t+1}\}_{i=1}^{N} \) by resampling (with replacement) \( N \) times from the approximate posterior density in (3.3), computed by (3.4), in such a way that

\[ P(\{\tilde{X}_{t+1}^{(i)} = X_{t+1}^{(i)}\}) = w^{(i)}_{t+1}. \]

The new weights are set to be uniform, i.e., \( \tilde{w}^{(i)}_{t+1} = 1/N \). Finally, if resampling is performed, the posterior density is set to be

\[ f_{t+1|t+1}(C \mid Y_{1:t}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{\tilde{X}_{t+1}^{(i)}}(C). \]

With a representation of \( f_{t+1|t+1} \) in terms of particles, the estimators described in Section 4.3.2 can be implemented in a straightforward manner. The interested reader is referred to [87] for a detailed treatment of this topic.
3.3 The Probability Hypothesis Density

The PHD filter is an approximation developed to reduce the computational complexity of the Bayesian recursions (3.1). Its basic idea consists of \textit{(lossily) compressing} the posterior distribution into a reduced-size set of statistics, and propagating these statistics instead of the posterior itself. This is the same philosophy leading to the Kalman Filter (KF), which propagates the first- and second-order moments in the standard vector case. In order to fully understand the PHD procedure, the principles and equations of the standard KF are reported next.

3.3.1 Single-Object Filtering: Kalman Filter

With Kalman filtering, the statistics propagated in lieu of the posteriors are mean and variance [3]. In the Gauss–Markov case, these statistics are sufficient (in the sense that they carry all of the information about the unknown parameter, see [31]), and the KF evaluates the actual posterior. In general, they yield an approximation which becomes tighter as the SNR increases, since higher-order moments may be neglected. Denoting by
\[
\hat{x}_{t|k} = \int x f_{t|k}(x | y_{1:k}) d\lambda(x)
\]
\[
\Sigma_{t|k} = \int x x^H f_{t|k}(x | y_{1:k}) d\lambda(x) - \hat{x}_{t|k} \hat{x}_{t|k}^H
\]
the posterior mean and variance, respectively, the procedures follows the scheme
\[
\cdots \rightarrow f_{t|t} \xrightarrow{\text{predict}} f_{t+1|t} \xrightarrow{\text{update}} f_{t+1|t+1} \rightarrow \cdots
\]
\[
\downarrow \quad \downarrow \quad \downarrow
\]
\[
\cdots \rightarrow \hat{x}_{t|t} \xrightarrow{\text{KF predict}} \hat{x}_{t+1|t} \xrightarrow{\text{KF update}} \hat{x}_{t+1|t+1} \rightarrow \cdots
\]
\[
\Sigma_{t|t} \xrightarrow{\text{KF predict}} \Sigma_{t+1|t} \xrightarrow{\text{KF update}} \Sigma_{t+1|t+1}
\]
That is, the posterior mean and variances are propagated (bottom row) instead of the posterior densities (top row).

The constant-gain KF (called the alpha–beta filter) further simplifies the recursive equations, by propagating only the posterior mean. This approximation works well if the SNR is large enough so that the second moment can be neglected.
3.3.2 The Intensity Function

For finite random sets lacking a meaningful notion of set addition,\(^3\) a first-order moment equivalent of the mean of a random vector should be defined. To this end, the equivalence between finite random sets and point processes (illustrated in Appendix B) can be exploited, after observing that the expectation (or mean, first moment, intensity) measure of a point process is the analogue of the expected value of a random variable. Consequently, we can define the expectation measure (first-moment measure) \(M\) of an RFS \(X\) as

\[
M_X(B) = \mathbb{E}[|X \cap B|]
\]

for any \(B \subset \mathbb{S}\). If \(M_X\) is absolutely continuous with respect to the reference measure \(\lambda\), then its density \(m_X\) is called intensity function (or first moment density), and

\[
M_X(B) = \int_B m_X(x) d\lambda(x)
\]

for any \(B \subset \mathbb{S}\). That is, the expectation measure of a set \(B\) yields the average number of points of \(X\) in \(B\). Therefore, the intensity function at \(x\) yields the density of the expected number of points of \(X\) at \(x\).

**Example 3.1 (Binomial process).** Suppose that \(n\) points are chosen, randomly and independently, in a bounded region \(A \subset \mathbb{R}^2\). Let \(C \subset A\), and consider the RFS \(X\) of the points in \(C\). In this case, \(|X|\) is a binomial random variable, and

\[
M_X(B) = n \frac{\lambda(B \cap C)}{\lambda(A)} \quad (3.5a)
\]

\[
m_X(x) = \begin{cases} 
  n/\lambda(A), & \text{if } x \in C \\
  0, & \text{otherwise}
\end{cases} \quad (3.5b)
\]

for any \(B \subset A\) (see the illustration of Figure 3.1).

---

\(^3\) Various works developed a concept of expectation for random compact closed subset and for general random closed sets [11, 12, 14, 71, 72], but these definitions do not seem to be immediately and usefully applicable to finite random sets.
Fig. 3.1 Here, \( n \) points are drawn at random from \( A \), and \( X \) is the RFS of points in \( C \). The expectation measure of \( X \) is given by (3.5a), where the shaded area is the set \( B \cap C \).

**Example 3.2 (Poisson process).** A Poisson RFS is completely characterized by the expectation measure \( M_X \). Specifically, its cardinality is a Poisson random variable with mean \( M_X(S) \), and, for a given cardinality \( n \), the elements of \( X \) are independent and identically distributed with distribution \( n^{-1} M_X(\cdot) \) [34].

**Intensity function and FISST:** exploiting the direct connection between RFS and point processes (see Theorem B.12 in Appendix B), an alternative representation of the intensity function is

\[
m_X(x) = \sum_{n=0}^{\infty} \frac{1}{n!} \Pr \{|X| = n + 1\} (n + 1)! \int_{S^n} f_{x_1, \ldots, x_n, x}(\mathbf{c}_1, \ldots, \mathbf{c}_n, x) \, d\lambda(\mathbf{c}_1) \cdots d\lambda(\mathbf{c}_n).
\]

The previous equation shows that the intensity function can be linked directly to the density of the RFS \( X \) by the equation

\[
m_X(x) = \int_{S} f_X(C \cup \{x\}) \delta C. \tag{3.6}
\]

Equation (3.6) implies that the intensity function can also be expressed in terms of the belief measure as follows:

\[
m_X(x) = \frac{\delta \beta_X}{\delta x}(S).
\]
3.3.3 Information-Theoretic Motivation of the PHD

As anticipated, the concept of PHD filtering is inherently tied to the idea of approximating the unknown pdf of a random set. In particular, let $X$ be an RFS with density $f_X$, and let $Z$ be a Poissonian random set, i.e., an RFS of which the cardinality follows a Poisson law with parameter $\mu$ (we write $|Z| \sim \mathcal{P}(\mu)$), and the elements are independent and identically distributed, i.e.:

$$f_Z(C) = \mu^{|C|} e^{-\mu} \prod_{c \in C} f_z(c).$$

A measure of the divergence between the densities $f_X(C)$ and $f_Z(B)$ is the pair of Kullback–Leibler divergences

$$D_1 = D(f_X(C)\|f_Z(C)) \triangleq \int_S f_X(C) \log \frac{f_X(C)}{f_Z(C)} \delta C$$

$$D_2 = D(f_Z(C)\|f_X(C)) \triangleq \int_S f_Z(C) \log \frac{f_Z(C)}{f_X(C)} \delta C.$$

Applying FISST, and duplicating the proof in [64], we obtain:

$$D_1 = K + \mu - \mathbb{E}[|X| \log \mu] - \sum_{n=0}^{\infty} \Pr \{|X| = n\}$$

$$\times \sum_{i=1}^{n} \int_{S^n} f_{x_1 \ldots x_n}(c_1, \ldots, c_n) \log f_z(c_i) d\lambda(c_1) \ldots d\lambda(c_n)$$

$$D_2 = D(\mathcal{P}(\mu)\|f_X(\cdot))$$

$$+ \mathbb{E}_N \left[ D \left( \prod_{i=1}^{N} f_z(c_i) \| f_{x_1, \ldots, x_N}(c_1, \ldots, c_N) \right) \right].$$

In the latter equation, $N$ is a random variable distributed as $\mathcal{P}(\mu)$, and $\mathbb{E}_N$ denotes expectation with respect to $N$. Elementary calculations reveal that the $D_1$-optimal Poissonian set approximating $X$, i.e., the Poissonian RFS whose $D_1$ divergence from the true set pdf is minimum,
is defined by

$$
\mu_1 = \mathbb{E}[|X|] \tag{3.7a}
$$

$$
f_{z}(c) = \int_{\mathbb{S}^{n-1}} f_{x_1,\ldots,x_{n-1},x_n}(c_1,\ldots,c_{n-1},c) d\lambda(c_1)\cdots d\lambda(c_{n-1})
= f_{X_{i}}(c). \tag{3.7b}
$$

No closed-form solution seems to exist for the $D_2$-optimum Poissonian set except for the relevant case of independent elements of $X$, which produces

$$
\mu_2 = \arg\min_{\mu > 0} D\left(\mathcal{P}(\mu)\|f_{X_{i}}(\cdot)\right)
$$

$$
f_{z}(c) = f_{X_{i}}(c)
$$

meaning that the parameter of the optimum Poissonian set is in general different from the one in (3.7), but the optimum density of the elements is the same.

### 3.3.4 Estimation Through PHD

The probability hypothesis density is the intensity function defined with reference to a posterior set density [44, 63], i.e.,

$$
D_{t|k}(x) = \int f_{t|k}(\{x\} \cup Z \mid Y_{1:k}) \delta Z
$$

$$
= \sum_{n=0}^{\infty} \frac{1}{n!} \int f_{t|k}(\{x,x_1,\ldots,x_n\} \mid Y_{1:k}) \, dx_1 \cdots dx_n.
$$

Therefore, the integral of the PHD over a set $B$ is the average number of objects in the set given the observations $Y_{1:k}$. Moreover, since the PHD gives the density of the expected number of objects at a certain point $x$, the peaks of the PHD indicate the locations with relatively high concentration of the expected number of objects, i.e., the locations with high probability of occurrence.

The previous remark suggests the following estimation procedure. Once the PHD $D_{t|t}(x)$ is computed, estimate the number of points present in $X_t$, given that the observations $Y_{1:t}$ are available, by using,
3.3 The Probability Hypothesis Density

for example, the procedure highlighted in Section 2.2, or through

\[ \hat{N}_t = \int S D_{lt}(x) d\lambda(x). \]

At this point, round \( \hat{N}_t \) off to the nearest integer, say \( \hat{n}_t \), and extract the \( \hat{n}_t \) points in \( S \) corresponding to local maxima of \( D_{lt}(x) \).

**Example 3.3.** Suppose that the posterior distribution of the RFS \( X_1 \subset \mathbb{R} \) is

\[
 f_{1|1}(\{x_1, x_2\} | \{y_1, y_2\}) = \frac{1}{\sigma} g\left(\frac{x_1 - y_1}{\sigma}\right) \frac{1}{\sigma} g\left(\frac{x_2 - y_2}{\sigma}\right) \\
+ \frac{1}{\sigma} g\left(\frac{x_1 - y_2}{\sigma}\right) \frac{1}{\sigma} g\left(\frac{x_2 - y_1}{\sigma}\right)
\]

where \( \{y_1, y_2\} \subset \mathbb{R} \) is the set of measurements at epoch 1, \( \sigma \in \mathbb{R} \), and

\[
g(x) \triangleq \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

is the standard Gaussian density function. The corresponding PHD is

\[
 D_{1|1}(x) = \int_{\mathbb{R}} f_{1|1}(\{x, w\} | \{y_1, y_2\}) dw \\
= \frac{1}{\sigma} g\left(\frac{x - y_1}{\sigma}\right) + \frac{1}{\sigma} g\left(\frac{x - y_2}{\sigma}\right) 
\]

(3.8)

of which the integral

\[
\int_{\mathbb{R}} D_{1|1}(x) dx = 2
\]

gives the expected number of points in \( X_1 \). In general, \( D_{1|1} \) in (3.8) is bimodal, and its two distinct peaks correspond to the locations of the points of \( X_1 \), as shown in Figure 3.2. However, (3.8) shows that \( D_{1|1} \) is unimodal whenever \( |y_1 - y_2| < 2\sigma \): if this is the case, two closely spaced points are indistinguishable, as shown in Figure 3.3.
The Complexity Problem and Some Solutions

Fig. 3.2 Plot of the PHD in (3.8) for \(\{y_1, y_2\} = \{1, 3\}\) and \(\sigma = 0.6\): \(D_{1|1}\) is bimodal, and its peaks occur near the locations of the two points in \(X_t\).

Fig. 3.3 Plot of the PHD in (3.8) for \(\{y_1, y_2\} = \{1.5, 2.5\}\), and \(\sigma = 0.6\): \(D_{1|1}\) is unimodal and its peak occur near the locations of the two closely spaced points in \(X_t\).
3.3.5 The PHD Filter

We recall once again that our goal is to approximate the prediction equation (3.1a). In this framework, PHD tracking amounts to updating (and thus transmitting) the conditional PHD only, i.e., the quantity $D_{t|t}(x)$, instead of implementing the whole prediction equation, extending the philosophy underlying Section 3.3.1, i.e., single-object filtering, to the multiobject tracking framework. Updating and transmitting the posterior PHD amounts to a *lossless* operation if the posterior density of the RFS $X_t$ given the past observations $Y_{1:t-1}$ is Poissonian with parameter $\mu_{t|t-1} = E[|\mathbf{x}_t||Y_{1:t-1}]$ and has independent entries. Summing up, if

1. the posterior density is, at least approximately, Poissonian, and
2. the elements of $X_t$ are conditionally independent (or nearly so) given the past observations,

then $D_{t|t}$ is an “approximately sufficient” statistic for $f_{t|t}$, and the losses incurred by the operation summarized in the diagram that follows will be negligible:

\[
\cdots \rightarrow f_{t|t} \xrightarrow{\text{predict}} f_{t+1|t} \xrightarrow{\text{update}} f_{t+1|t+1} \rightarrow \cdots
\]
\[
\downarrow \quad \text{PHD} \quad \downarrow \quad \text{PHD} \quad \downarrow
\]
\[
\cdots \rightarrow D_{t|t} \xrightarrow{\text{predict}} D_{t+1|t} \xrightarrow{\text{update}} D_{t+1|t+1} \rightarrow \cdots
\]

Here, the top row represents the Bayesian recursions (3.1) while the bottom row refers to the PHD predict-and-update steps: these form a set of recursive filtering operations on PHDs that should always result in the PHD directly derived from the corresponding posterior probability.

If certain assumptions are met, it can be shown that the PHD can be propagated using suitable extensions of the single-object filtering equations given in Section 3.3.1. In the following, the conditions under which the PHD can be propagated, and its recursive equations, will be presented. The mathematical derivations can be found in [64, 66] from a FISST point of view, and in [80] from a point-process perspective.
3.3.5.1 PHD filter initialization

To initialize the PHD filter, an a priori density must be assigned, i.e.,

$$D_{0|0}(x) = N_0 f_0(x)$$

where $N_0$ is an initial estimate of the number of points in the RFS, and $f_0(x)$ is a probability density, whose peaks correspond to the a priori points location. For example, $D_{0|0}(x)$ can be taken to be the sum of $N_0$ Gaussian densities, where the variances are related to the a priori accuracy on point positions.

3.3.5.2 PHD filter predictor

In order to have a closed-form expression for the prediction step $D_{t|t} \rightarrow D_{t+1|t}$, the following conditions must be met:

1. The motions of the objects to be tracked are independent, with transition densities $f_{x_{t+1}|x_t}$.
2. Each point survival is independent of the other points; the probability that point $x \in X_t$ survives, and hence is present in $X_{t+1}$, is $q_{t+1}(x)$.
3. The appearance of new points is independent of the existing points. The intensity function of the RFS $N_t$ is $m_{N_t}$.

If all previous assumptions hold, then the PHD predictor equation becomes [64]

$$D_{t+1|t}(x) = m_{N_t}(x) + \int_S q_{t+1}(w)f_{x_{t+1}|x_t}(x | w)D_{t|t}(x) d\lambda(w) \quad (3.9)$$

Again, spawning from existing points has been excluded for simplicity, but can be easily accounted for [64]. Notice that if there is no missing observation and no false measurement, then (3.9) reduces to the standard predictor in Bayesian recursion for single-object tracking, namely,

$$D_{t+1|t}(x) = \int_S f_{x_{t+1}|x_t}(x | w)D_{t|t}(x) d\lambda(w).$$
3.3.5.3 PHD Filter Corrector

In order to obtain a closed-form expression for the correction step \( D_{t+1|t} \rightarrow D_{t+1|t+1} \), the following conditions must be met:

1. The probability that each point is observed is independent of other points. The probability that the point \( x \in X_{t+1} \) is observed at epoch \( t + 1 \) is \( p_{t+1}(x) \).
2. Observed points in \( Y_{t+1} \) have conditional probability density \( f_{y_{t+1}|x_t} \).
3. The false measurement RFS \( \Gamma_{t+1} \) is Poisson, with intensity function \( m_{\Gamma_{t+1}} \).
4. The posterior probability \( f_{t+1|t} \) is (at least approximately) Poisson.

Notice that, except for the last one, all the assumptions in the prediction and correction steps are standard in most tracking applications (see, e.g., [17, 24]). As to the last condition, it represents a reasonable assumption whenever the interactions among points of \( X_t \) are negligible. In particular, it is satisfied if there is no spawning, and \( X_t \) is Poisson [86].

If all previous assumptions hold, then the PHD corrector equation becomes \[64]\)

\[
D_{t+1|t+1}(x) = (1 - p_{t+1}(x))D_{t+1|t}(x) + p_{t+1}(x)D_{t+1|t}(x)
\times \sum_{y \in Y_{t+1}} \frac{f_{x_{t+1}|y_{t+1}}(x | y)}{m_{\Gamma_{t+1}}(x) + \int_S p_{t+1}(w)f_{x_{t+1}|y_{t+1}}(w | y)d\lambda(w)}
\]

(3.10)

An interesting interpretation of the recursive equations (3.9) and (3.10) can be found in [42].

3.3.6 Implementation of the PHD Filter

From (3.9) and (3.10), it can be seen that the PHD filter does not suffer from the combinatorial complexity caused by the unknown measurement–data association. Moreover, the computational complexity of the PHD recursion is drastically smaller than that required by
the Bayesian recursions in (3.1), since the latter operate on $\mathcal{B}(\mathcal{S})$, (the collection of all finite subsets of arbitrary cardinality) while the former operates on $\mathcal{S}$. However, (3.9) and (3.10) still involve multidimensional integrals, which may not admit closed-form solutions. If this is the case, they must be approximated. Two popular approximation techniques are

1. Sequential Monte-Carlo (SMC) approximation, also known as particle filtering.
2. Gaussian-Mixture (GM) approximation.

3.3.6.1 SMC Approximation

Starting from [79, 95], SMC methods have been used to approximate PHD recursions. A particle system approximation of $D_{t|t}$ is a collection of points $\{x_{t|i}^{(i)}\}_{i=1}^{N}$ and importance weights $\{w_{t|i}^{(i)}\}_{i=1}^{N}$ such that

$$D_{t|t} \approx \sum_{i=1}^{N} w_{t|i}^{(i)} \delta_{x_{t|i}^{(i)}}(x).$$

The main difference with standard SMC techniques — originally devised to approximate probability densities — is that the important weights do not sum up to one, but rather to the expected number of points in the RFS $X_t$:

$$\sum_{i=1}^{N} w_{t|i}^{(i)} \approx \hat{N}_t.$$ 

In spite of these differences, the extension of SMC approximations to the PHD filter follows quite easily. A detailed discussion can be found in [87].

3.3.6.2 GM Approximation

The GM approximation for the PHD recursion has been proposed in [86]. It assumes that the posteriors can be modeled as Gaussian mixtures, and it propagates them by propagating the weighing coefficients, the means, and the covariances of their constituent Gaussian components. It is a generalization of the GM method in [2, 81] to the case of intensity functions, and provides a closed-form solution to the
PHD filter equations whenever certain conditions (as specified next) are satisfied.

Specifically, in addition to the assumptions required for the recursions (3.9) and (3.10), the following conditions should be met:

1. Each object in the RFS $X_t$ follows a linear, Gaussian dynamical model, i.e.,
   $$ f_{x_{t+1}|x_t}(x | x') = g(x, F_t x', Q_t) $$
   where $g(\cdot, m, P)$ is the density of a jointly Gaussian random vector with mean $m$ and covariance matrix $P$, $F_t$ is the state transition matrix, and $Q_t$ is the process noise covariance matrix.

2. The measurements in $Y_{t+1}$ follow a linear-Gaussian model, i.e.,
   $$ f_{y_{t+1}|x_{t+1}}(y | x) = g(y, H_{t+1} x, R_{t+1}) $$
   where $H_{t+1}$ is the observation matrix, and $R_{t+1}$ is the observation noise covariance matrix.

3. The probability of survival is independent of the point, i.e.,
   $$ q_{t+1}(x) = q_{t+1}. $$

4. The probability of being observed is independent of the point, i.e.,
   $$ p_{t+1}(x) = p_{t+1}. $$

5. The intensity function of the birth RFS $N_{t+1}$ is a Gaussian mixture of the form
   $$ m_{N_{t+1}}(x) = \sum_{i=1}^{\alpha_{t+1}} \beta_{t+1}^{(i)} g(x, b_{t+1}^{(i)}, B_{t+1}^{(i)}). $$  

(3.11)

Again spawning is omitted here for brevity: if it is included, its intensity function has to be a Gaussian mixture.

If conditions (1)–(5) are met, and if the updated PHD is a Gaussian mixture of the form
$$ D_{t|t}(x) = \sum_{i=1}^{n_{t|t}} w_{t|t}^{(i)} g(x, m_{t|t}^{(i)}, P_{t|t}^{(i)}) $$
then the predicted PHD is
$$ D_{t+1|t}(x) = q_{t+1} \sum_{i=1}^{n_{t|t}} w_{t|t}^{(i)} g(x, F_t m_{t|t}^{(i)}, Q_t + F_t P_{t|t}^{(i)} F_t^T) + m_{N_{t+1}}(x). $$
If conditions (1)–(5) are met, and if the predicted PHD is a Gaussian mixture of the form

\[ D_{t+1|t}(x) = \sum_{i=1}^{n_{t+1|t}} w_{t+1|t}^{(i)} g(x, m_{t+1|t}^{(i)}, P_{t+1|t}^{(i)}) \]

then the corrected PHD is

\[ D_{t+1|t+1}(x) = (1 - p_{t+1}) D_{t+1|t}(x) + \sum_{y \in Y_{t+1}} \sum_{i=1}^{n_{t+1|t}} w_{t+1|t+1}^{(i)} g(x, m_{t+1|t+1}^{(i)}(y), P_{t+1|t+1}^{(i)}) \]

where

\[ w_{t+1|t+1}^{(i)}(y) \triangleq \frac{p_{t+1} w_{t+1|t}^{(i)} g_{t+1}(y)}{m r_{t+1}(y) + p_{t+1} \sum_{\ell=1}^{n_{t+1|t}} w_{t+1|t}^{(\ell)} g_{t+1}(y)} \]

\[ g_{t+1}(y) \triangleq g(y, H_{t+1} m_{t+1|t}^{(i)}, R_{t+1} H_{t+1} P_{t+1|t}^{(i)} H_{t+1}^T) \]

\[ m_{t+1|t+1}^{(i)}(y) \triangleq m_{t+1|t}^{(i)}(y) + K_{t+1}(y - H_{t+1} m_{t+1|t}^{(i)}) \]

\[ P_{t+1|t+1}^{(i)} \triangleq (I - K_{t+1}(H_{t+1} P_{t+1|t}^{(i)} H_{t+1}^T + R_{t+1})^{-1} \]

\[ K_{t+1} \triangleq P_{t+1|t+1}^{(i)} H_{t+1}^T (H_{t+1} P_{t+1|t}^{(i)} H_{t+1}^T + R_{t+1})^{-1} \]

and I is the identity matrix. The proof can be found in [86].

In conclusion, if the PHD filter is initialized with a Gaussian mixture, all predicted and corrected PHDs will be Gaussian mixtures. As to the expected number of objects, they can be obtained upon summing up the appropriate weights, i.e.,

\[ \hat{N}_{t+1|t} = q_{t+1} \hat{N}_{t|t} + \sum_{i=1}^{n_{t+1|t}} \beta_{t+1}^{(i)} \]

\[ \hat{N}_{t+1|t+1} = \hat{N}_{t+1|t}(1 - p_{t+1}) + \sum_{y \in Y_{t+1}} \sum_{i=1}^{n_{t+1|t}} w_{t+1|t+1}^{(i)}(y). \]

The GM-PHD filter, like the Gaussian sum filter, suffers from computational problems associated with the increasing number of
3.3 The Probability Hypothesis Density

Gaussian components as it evolves. This effect can be controlled through appropriate merging and pruning procedures [86]. Finally, it is worthwhile mentioning that, under a linear-Gaussian assumption for the dynamical model, the GM-PHD filter can approximate the true PHD with an error that converges to zero (as the number of Gaussian terms in the mixture tends to infinity) in $L_1$ [89].

3.3.7 Higher order approximation: Cardinalized PHD

The PHD filter has a number of advantages, as it

• reduces the computational complexity required by the full Bayesian recursions (3.1),
• can be implemented through SMC or GM approximations and admit a closed-form solution in the linear-Gaussian case, and
• does not require measurement-to-data association.

At the same time, it exhibits some drawbacks [41, 66]. In fact, since it approximates the cardinality distribution by a Poisson with matching mean, and since mean and variance are equal for a Poisson distribution, it happens that, when the number of points is large, the PHD filter tends to give high variance estimates of the number of points. Moreover, it is known that estimating the number of objects through a posterior mean (the integral of the PHD, which is the intensity function with respect to the posterior density) gives unreliable values in the low-SNR region because of minor modes induced by noise.

To address these problems, a second-moment approximation of the Bayesian recursions may be adopted, which in principle guarantees a better performance than the first-moment-only approximation provided by the PHD filter in the same way that the (second-order) KF outperforms its constant-gain version [66]. The idea here is to propagate both the first-moment density, $D_{t|t}$, and the second factorial moment density, which in turn is defined as

$$C_{t|t}(w, z) = \int f_{X_t}(\{w, z\} \cup B \mid Y_{1:t}) \delta B.$$  

Note that the intensity function of an RFS $X$ is the density of the measure $M(A) = E[N(A)]$, where $N(A) \triangleq |X \cap A|$. Similarly, the second factorial moment density is the
This approach, however, seems to be computationally too demanding, and in [65] a solution that is a compromise between the information loss of the PHD filter and the intractability of the second-order filter is proposed. Specifically, the Poisson assumption on the number of points is relaxed, and a generalization of the PHD filter, known as Cardinalized PHD filter (CPHD), is derived. The recursion jointly propagates the PHD $D_{t|t}$ and the posterior cardinality distribution. The CPHD recursions are computationally less demanding (as that in the PHD filter), and, propagating the entire cardinality distribution, improve the accuracy in estimating the set cardinality. Denoting by

$$\rho_{t|k}(n) = \mathbb{P}\{|X_t| = n \mid Y_{1:k}\}, \quad \forall n \in \mathbb{N}$$

the posterior cardinality distribution of $X_t$, the recursion follows the diagram

$$\cdots \rightarrow f_{t|t} \xrightarrow{\text{predict}} f_{t+1|t} \xrightarrow{\text{update}} f_{t+1|t+1} \rightarrow \cdots$$

$$\cdots \rightarrow \rho_{t|t} \xrightarrow{\text{CPHD predict}} \rho_{t+1|t} \xrightarrow{\text{CPHD update}} \rho_{t+1|t+1} \rightarrow \cdots$$

$$\cdots \rightarrow D_{t|t} \xrightarrow{\text{CPHD predict}} D_{t+1|t} \xrightarrow{\text{CPHD update}} D_{t+1|t+1} \rightarrow \cdots$$

We now proceed to derive the equations for the prediction and correction steps.

**3.3.7.1 CPHD Filter**

As with the PHD filter, which is based on a Poisson approximation of the RFS, the CPHD recursion is based on i.i.d. cluster RFSs. Specifically, $X$ is called an *i.i.d. cluster RFS* if, given $|X| = n$, the elements of $X$ are i.i.d. with distribution $n^{-1} M_X$, where $M_X$ is the expectation measure of $X$ [34].

---

*Density of the measure $M_{[2]}(A \times B) = \mathbb{E}[N(A)N(B)] - \mathbb{E}[N(A \cap B)]$. The term subtracted assures no points repetition, while the term *factorial* comes from the fact that $M_{[2]}(A \times A) = \mathbb{E}[N(A)(N(A) - 1)]$ is called the *second factorial moment* of the random variable $N(A)$. See Appendix B for further details.*
To initialize the CPHD filter, two \textit{a priori} densities must be assigned, i.e.,

\[
\rho_{0|0}(n) = s_0(n) \\
D_{0|0}(x) = N_0 f_0(x)
\]

where \(s_0\) is an initial estimate of the cardinality distribution, \(N_0 = \sum_{n=0}^{\infty} ns_0(n)\) is the corresponding expected number of points in the RFS, and \(f_0(x)\) is a probability density of which the peaks correspond to the \textit{a priori} points location. For example, \(D_{0|0}(x)\) can be taken to be the sum of \(N_0\) Gaussian densities, where the variances are related to the \textit{a priori} accuracy on points positions, and \(\rho_{0|0}\) can be taken to be binomial.

As for the prediction step, the CPHD filter stems from the same assumptions required in the PHD filter (see Section 3.3.5.2), with the exception that no point spawning is allowed. Furthermore, it is also required that the posterior density \(f_t|t\) be that of an i.i.d. cluster RFS. In this case, the CPHD prediction step produces \([65, 66, 89]\)

\[
\rho_{t+1|t}(n) = \sum_{i=1}^{n} b_{t+1}(n - i) \Pi_{t+1|t}[D_{t|t}, \rho_{t|t}](i)
\]

\[
D_{t+1|t}(x) = \int q_{t+1}(w)f_{x_{t+1}|x_{t}}(x | w)D_{t|t}(w)d\lambda(w) + m_{N_{t+1}}(x)
\]

where

\[
\Pi_{t+1|t}[D, \rho](i) \triangleq \sum_{\ell=1}^{\infty} \binom{\ell}{i} \langle q_{t+1}, D \rangle^i \langle 1 - q_{t+1}, D \rangle^{\ell - i} \rho(\ell)
\]

\[
m_{N_{t+1}} \triangleq \text{intensity function of } N_{t+1}
\]

\[
b_{t+1} \triangleq \text{cardinality distribution of } N_{t+1}
\]

and \(\langle \cdot, \cdot \rangle\) denotes inner product of two real-valued functions \(\alpha\) and \(\beta\), i.e., \(\int \alpha(x)\beta(x)dx\) (or \(\sum_{\ell=0}^{\infty} \alpha(\ell)\beta(\ell)\) if \(\alpha\) and \(\beta\) are real-valued sequences).

Finally, the correction step necessitates the same assumptions needed for the PHD filter (see Section 3.3.5.3). Furthermore, it is required that the false measurement set \(\Gamma_t\) be an i.i.d. cluster RFS,
The Complexity Problem and Some Solutions

and that the posterior density \( f_{t+1|t} \) be itself that of an i.i.d. cluster RFS. Under these hypotheses, the CPHD prediction step runs as follows [65, 66, 89]:

\[
\rho_{t+1|t+1}(n) = \frac{\Upsilon^0_{t+1}[D_{t+1|t}, Y_{t+1}](n) \rho_{t+1|t}(n)}{(\Upsilon^0_{t+1}[D_{t+1|t}, Y_{t+1}], \rho_{t+1|t})}
\]

\[
D_{t+1|t}(x) = \frac{\langle \Upsilon^1_{t+1}[D_{t+1|t}, Y_{t+1}], \rho_{t+1|t} \rangle}{(\Upsilon^0_{t+1}[D_{t+1|t}, Y_{t+1}], \rho_{t+1|t})} (1 - p_{t+1}(x)) D_{t+1|t}(x)
\]

\[
+ \sum_{y \in Y_{t+1}} \frac{\langle \Upsilon^1_{t+1}[D_{t+1|t}, Y_{t+1} \setminus \{y\}], \rho_{t+1|t} \rangle}{(\Upsilon^0_{t+1}[D_{t+1|t}, Y_{t+1}], \rho_{t+1|t})} \psi_{t+1,y}(x) D_{t+1|t}(x)
\]

where

\[
\Upsilon^k_{t+1}[D,Y](n) \triangleq \min\{|Y|, n\} \sum_{i=0}^{\min\{|Y|, n\}} (|Y| - i) \gamma_{t+1}(|Y| - i) \frac{n!}{(n-i)!} \times \frac{(1 - p_{t+1}, D)^{n-(i+k)}}{(1, D)^n} e_i(\Xi_{t+1}(D,Y))
\]

\[
\psi_{t+1,y}(x) \triangleq \frac{\langle 1, m_{\Upsilon_{t+1}} \rangle}{m_{\Upsilon_{t+1}}(y)} f_{\Upsilon_{t+1}|x_{t+1}}(y \mid x) p_{t+1}(x)
\]

\[
\Xi_{t+1}(D,Y) \triangleq \{ \langle D, \psi_{t+1,y} \rangle : y \in Y \}
\]

\[
m_{\Upsilon} \triangleq \text{intensity function of } \Upsilon_{t+1}
\]

\[
\gamma_{t+1} \triangleq \text{cardinality distribution of } \Upsilon_{t+1}
\]

and \( e_i \) is the elementary symmetric function (see [26], Section 1.1.E.2) of order \( i \), defined, for a finite set \( Y \) of real numbers, by

\[
e_i(Y) \triangleq \sum_{S \subseteq Y, |Y| = i} \prod_{y \in S} y
\]

with \( e_0(Y) = 1 \) by convention.

The original CPHD recursion of [65, 66] has been modified here according to [88, 89].
3.3 The Probability Hypothesis Density

3.3.7.2 SMC CPHD approximation

The basic concepts of particle system approximation of the CPHD filter are similar to those needed for the PHD filter of Section 3.3.6.1. The details are omitted here for brevity, and the interested reader is referred to [66] for a thorough discussion.

3.3.7.3 GM CPHD approximation

The Gaussian-mixture CPHD filter was derived in [89]. It requires the same assumptions needed for the GM PHD filter in Section 3.3.6.2, and yields a closed-form expression for the CPHD recursion whenever those assumptions are met.

Specifically, if conditions (1)–(5) of Section 3.3.6.2 are met, and if the updated PHD is a Gaussian mixture of the form

\[ D_t|t(x) = \sum_{i=1}^{n_t|t} w_t^{(i)} g(x, m_t^{(i)}, P_t^{(i)}) \]

where \( g(\cdot, m, P) \) is the density of a jointly Gaussian random vector with mean \( m \) and covariance matrix \( P \), then the predicted PHD is also a Gaussian mixture, and the CPHD prediction step is simplified to

\[
\rho_{t+1|t}(n) = \sum_{i=0}^{n} b_{t+1}(n - i) \sum_{\ell=i}^{\infty} \binom{\ell}{i} \rho_{t|t}(\ell) q_{t+1}(1 - q_{t+1})^{\ell-i}
\]

\[
D_{t+1|t}(x) = q_{t+1} \sum_{i=1}^{n_{t+1}} u_{t+1}^{(i)} g(x, F_t m_t^{(i)} + Q_t, P_t^{(i)} P_t^{T}) + m_{N_{t+1}}(x).
\]

where \( m_{N_{t+1}} \) has the form given in (3.11).

If conditions (1)–(5) of Section 3.3.6.2 are met, and if the predicted PHD is a Gaussian mixture of the form

\[
D_{t+1|t}(x) = \sum_{i=1}^{n_{t+1}} w_{t+1|t}^{(i)} g(x, m_{t+1}^{(i)}, P_{t+1|t}^{(i)})
\]
then the correction step in the CPHD recursions simplifies to

\[
\rho_{t+1|t+1}(n) = \frac{\Psi_{t+1}^0[w_{t+1|t},Y_{t+1|t}]^{(n)} \rho_{t+1|t}^{(n)}}{\langle \Psi_{t+1}^0[w_{t+1|t},Y_{t+1|t}], \rho_{t+1|t} \rangle}
\]

\[
D_{t+1|t+1}(x) = \frac{\langle \Psi_{t+1}^1[w_{t+1|t}, Y_{t+1|t}], \rho_{t+1|t} \rangle}{\langle \Psi_{t+1}^0[w_{t+1|t}, Y_{t+1|t}], \rho_{t+1|t} \rangle} \left(1 - p_{t+1}(x)\right) D_{t+1|t}(x)
+ \sum_{y \in Y_{t+1}} \sum_{i=1}^{n_{t+1|t}} w_{i+1|t+1}(y)g(x, m_{i+1|t+1}(y), P_{i+1|t+1})
\]

where

\[
\Psi_{t+1}^k[w,Y]^{(n)}(n) \triangleq \min\{|Y|, n\} \sum_{i=0}^{\min\{|Y|, n\}} \left(|Y| - i\right) \gamma_{t+1}(|Y| - i) \frac{n!}{(n-i)!} \times \frac{(1 - p_{t+1})^{n-(i+k)}}{(1, w)^{i+k}} e_i(A_{t+1}(D, Y))
\]

\[
A_{t+1}(w, y) \triangleq \left\{ \frac{1}{m_{t+1}(y)} p_{t+1} w_{t+1|t}^T d_{t+1}(y) : y \in Y \right\}
\]

\[
W_{t+1|t} \triangleq \left( w_{1|t}^{(1)} \cdots w_{n_{t+1|t}}^{(n_{t+1|t})} \right)^T
\]

\[
d_{t+1}(y) \triangleq \left( d_{1|t}^{(1)}(y) \cdots d_{n_{t+1|t}}^{(n_{t+1|t})}(y) \right)^T
\]

\[
d_{t+1|t}^{(i)}(y) \triangleq g \left( y, H_{t+1} m_{t+1|t}, H_{t+1} P_{t+1|t} H_{t+1}^T + R_{t+1} \right)
\]

\[
w_{t+1|t}^{(i)}(y) \triangleq p_{t+1} w_{i+1|t}^{(i)} d_{t+1}(y)
\]

\[
m_{t+1|t}^{(i)}(y) = m_{t+1|t}^{(i)}(y) + K_{t+1}^{(i)}(y - H_{t+1} m_{t+1|t})
\]

\[
P_{t+1|t}^{(i)} = (I - K_{t+1}^{(i)} H_{t+1}) P_{t+1|t}
\]

\[
K_{t+1}^{(i)} = P_{t+1|t}^{(i)} H_{t+1}^T (H_{t+1} P_{t+1|t} H_{t+1}^T + R_{t+1})^{-1}
\]

The proof can be found in [89].
3.4 Zero-Order Approximations

Note that if the cardinality distribution of $X_t$ is infinite-tailed, then it may be truncated at some relatively large integer. As for the evaluation of the elementary symmetric function, following [89], it can be done through a simple procedure [26], which requires $O(|Y|^2)$ operations for a finite set $Y$, or $O(|Y| \ln^2 |Y|)$ operations using suitable decompositions and recursions (see [1], Theorem 8.14). Finally, as in the GM-PHD filter, the problems associated with the increasing number of Gaussian components can be managed through suitable merging and pruning procedures [89].

3.4 Zero-Order Approximations

The situation we consider in this section is one in which a single superpositional sensor should detect and track a multiobject consisting of an unknown, bounded number of elementary objects. Each object is characterized by a known signature and a set of unitless parameters inserted in a scalar and possibly random coefficient, which leads to the observation model

$$y_t = Sd_t(X_t) + n_t$$  \hspace{1cm} (3.12)

where $n_t$ is a sequence of mutually independent, zero-mean, Gaussian random vectors with covariance matrix $E[n_t^T n_t] = \frac{N_0}{2} I_N$, $I_N$ denoting the identity matrix of order $N$. We assume, for the moment, that the columns of $S$ in (3.12) contain the signatures of the constituent objects and are linearly independent (the case of dependent signatures is deferred to the sequel). The vector $d_t(X_t)$ has $|X_t|$ non-zero entries, the $i$th one summarizing all the unknown parameters of the $i$th object, while $X_t$ is a random set of which the elements are vectors containing the unknown parameters of the active objects. Notice that model (3.12) assumes that the multiobject has, at most, $N$ constituents, where $N$ is the size of the observation vector and that each signature may be present once at most.

Revisiting Bayesian recursions, and in particular at the prediction equation (3.1a), we can use some concepts from the linearization techniques producing Extended Kalman Filtering (EKF). Indeed, under a low-disturbance condition, it happens that the posterior $f_{t-1|t-1}$
The Complexity Problem and Some Solutions

exhibits a sharp peak around its maximum, say \( \hat{X}_{t-1} \), which also represents the best causal MAP estimate available at time \( t - 1 \). Consequently, the prediction step can be approximated in the form

\[
f_{t|t-1}(C | y_{1:t-1}) \simeq w_t f_{X_t|X_{t-1}}(C | \hat{X}_{t-1}) \times f_{t-1|t-1}(\hat{X}_{t-1} | y_{1:t-1}). \tag{3.13}
\]

Defining the log-likelihood:

\[
\Lambda_{t|t}(C) \triangleq - \ln f_{X_t|y_{1:t}}(C | y_{1:t})
\]

we obtain

\[
\Lambda_{t|t}(C) = \Lambda_{t-1|t-1}(\hat{X}_{t-1}) + \ln w_t - \ln f_{X_t|X_{t-1}}(C | \hat{X}_{t-1}) - \ln f_{y_t|X_t}(y_t | C)
\]

and the approximated optimum causal estimator becomes

\[
\hat{X}_t = \arg \max_{C \in 2^S} \left[ \ln w_t - \ln f_{C|X_{t-1}}(C | \hat{X}_{t-1}) - \ln f_{y_t|X_t}(y_t | C) \right]. \tag{3.14}
\]

### 3.4.1 Identification

We start with the simple, yet important, case in which \( S = \{1, \ldots, N\} \). This occurs when estimating \( X_t \) amounts to determining which signatures are present in the observations, out of a set of \( N \) possible (different and linearly independent) signatures. Under these circumstances, (3.14) simplifies to:

\[
\hat{X}_t = \arg \min_{C \in 2^S} \left[ - \ln f_{X_t|X_{t-1}}(C | \hat{X}_{t-1}) - \ln f_{y_t|X_t}(y_t | C) \right] \tag{3.15}
\]

where \( 2^S \) is the power set of \( S \). Notice that by construction \( d(X_t) \) is nonzero at entry \( i \) if and only if \( \{i\} \subseteq X_t \). Denoting by \( q_t(i) \) the probability that object \( i \) survives from \( t-1 \) to \( t \), and by \( p_t(i) \) the probability that object \( i \), absent from \( X_{t-1} \), enters \( X_t \) as a newborn, statistical independence of deaths, births, and survivals is tantamount to having

\[
\ln f_{X_t|X_{t-1}}(C | \hat{X}_{t-1}) = \sum_{i=1}^{N} \ln f_{X_t(i)|X_{t-1}^{(i)}}(X_t^{(i)} | \hat{X}_{t-1}^{(i)}) \tag{3.16}
\]
3.4 Zero-Order Approximations

where \( \mathbf{X}_t^{(i)} \) is the sequence of singleton-or-empty random sets

\[
\mathbf{X}_t^{(i)} = \begin{cases} 
\{i\}, & \text{if } \{i\} \text{ is present} \\
\emptyset, & \text{otherwise.} 
\end{cases}
\]  

(3.17)

whereby

\[
f_{\mathbf{X}_t^{(i)} | \mathbf{X}_{t-1}^{(i)}} (\mathbf{X}_t^{(i)} | \hat{\mathbf{X}}_{t-1}^{(i)}) = \begin{cases} 
q_t(i), & \text{if } \mathbf{X}_t^{(i)} = \hat{\mathbf{X}}_{t-1}^{(i)} = \{i\} \\
1 - q_t(i), & \text{if } \mathbf{X}_t^{(i)} = \emptyset & \& \hat{\mathbf{X}}_{t-1}^{(i)} = \{i\} \\
p_t(i), & \text{if } \mathbf{X}_t^{(i)} = \{i\} & \& \hat{\mathbf{X}}_{t-1}^{(i)} = \emptyset. 
\end{cases}
\]

If the matrix \( \mathbf{S} \) has full rank, it is amenable to the QR-decomposition \( \mathbf{S} = \mathbf{QR} \), with \( \mathbf{Q} \) unitary and \( \mathbf{R} \) upper-triangular, whereby (3.15), upon transforming the observation through

\[
z_t = \mathbf{Q}^\dagger \mathbf{y}_t, \quad (\cdot)^\dagger \text{ denoting pseudo-inverse,}
\]

produces

\[
\hat{\mathbf{X}}_t = \arg\min \left\{ \sum_{i=1}^{N} \left| z_{t,i} - \sum_{j=1}^{N} R_{i,j} \mathbb{I}_{\{X_t^{(i)} \neq \emptyset\}} \right| - N_0 \ln f_{\mathbf{X}_t^{(i)} | \mathbf{X}_{t-1}^{(i)}} (\mathbf{X}_t^{(i)} | \hat{\mathbf{X}}_{t-1}^{(i)}) \right\} 
\]

(3.18)

where

- \( z_{t,i} \) is the \( i \)th entry of the vector \( \mathbf{z}_t \).
- \( R_{i,j} \) is the entry \((i,j)\) of the triangular matrix \( \mathbf{R} \).
- \( \mathbb{I}_E \) is the indicator of event \( E \).
- The minimization is over singleton-or-empty sets of the form (3.17).

The minimization problem appearing in (3.18) can be solved via Sphere Decoding (SD) [85], which results in scaling down the complexity from exponential to algebraic, or even to linear, for increasingly large SNRs [48]. This algorithm

(1) is single-pass and causal, which makes it prone to error propagation. Moreover, its initial state, say \( \mathbf{X}_0 \), should be either known or accurately estimated, and the SNR should be large enough to guarantee accuracy of estimation at each step;
(2) is applicable whenever $X_t$ can be written as a disjoint union of conditionally independent countable RFSs. In particular, it can be easily generalized to the case that the space whereupon the sets are built is any Cartesian product of the type $\{1,\ldots,N\} \times U$, with $U$ countable;

(3) can be generalized to the case of a matrix $S$ with linearly dependent columns. In this case, the QR factorization produces a trapezoidal $R$ matrix, which means that the search for the minimum should be reformulated through Generalized Sphere Decoding (GSD) [33]. This entails a complexity exponential in the difference between $N$ and the matrix rank.

### 3.4.2 Hybrid Spaces

The generalization of the previous ideas to the estimation of an RFS built on a hybrid space, i.e., when the elements of the RFSs contain a continuous part, is simple but not obvious due to the fact that the constant $w_t$ in (3.14) needs to be chosen with suitable accuracy. One possible approach here is to resort to a gridding algorithm. In grid-based approximation methods, continuous parameters are approximated by a finite and fixed grid of cells centered at $L$ points $\mathcal{R} = \{R_1,\ldots,R_L\}$. Thus, continuous parameters are approximated by their quantized versions, and the probability densities, whether unconditional or conditional, by corresponding probability mass functions obtained by integration [13]. By doing this, the problem of estimating the sequence of hybrid RFSs is reduced to that of estimating a sequence of countable RFSs, with accuracy being tied to $L$. Thus, the algorithms described in previous subsection can be directly applied. For an example of application of this methodology, see [7], where the object to be estimated includes the identities of active users and their relevant parameters in a multiuser CDMA system.
4

Multiuser Detection

4.1 Introduction and Motivation

In random-access communication systems, the number of active users varies with time, and has considerable bearing on system performance. In typical multiuser detection (MUD) analyses [82], detectors operate on the assumption that the number of active users is constant, known to the receiver, and equal to the maximum number of users entitled to access the system. The last assumption is often overly pessimistic, since many users might be inactive at any given time, and detection under the assumption of a number of users larger than the real one may impair performance. Optimum MUD, in this context, should identify the active users and estimate their parameters and data, which in turn may vary over time. This is a typical coupled detection–estimation problem, or, equivalently, a state estimation problem of a complex multiobject system under uncertainty as to the number of active subsystems. If the number of active users admits a known and conveniently low upper bound, say $K$, then the multiobject can be modeled as a dynamical system with state-space $S^K$, with $S$ the state-space of the constituent subsystems endowed with an “idle state,” modeling inactivity, and a conventional probability space can be constructed.
adopting $\mathcal{S}^k$ as sample space. As, instead, $K$ is unknown, the sample space should be $\times_{k=0}^{\infty} \mathcal{S}^k$, with $\mathcal{S}^0 = \emptyset$, i.e., the collection of all sets containing a finite number of points of $\mathcal{S}$, and there is no need for defining an idle state in $\mathcal{S}$. Thus, MUD amounts to estimating a simple point-process on $\mathcal{S}$ or, equivalently, a random set on $\mathcal{F}(\mathcal{S})$ (see Section 2.1). The aim of this section is to illustrate this new setup in the framework of classical code-division multiple access (CDMA) systems: in particular, Section 4.2 considers active user identification and data detection, which corresponds to a discrete single-user state space $\mathcal{S}$, while Section 4.3 examines the more complex problem of estimating user parameters, which requires consideration of a hybrid state space $\mathcal{S}$.

In MUD, the fact that the set of active users at any time may not be known to the receiver has long been recognized as one of the important issues. For conventional (i.e., matched-filter) receivers, this dearth of information does not affect performance, because they are based on the assumption that all signals other than the useful one can be modeled as additive noise. However, for other receivers, the assumption that all users are active may cause significant performance degradation.\(^1\)

In addition, certain detectors based on interference cancellation must know the strongest active users in order to perform satisfactorily \cite{47}. Moreover, as observed in \cite{93}, “identification of active users will help the system to promptly process requests and efficiently allocate channels. In such a way, system capacity can be increased.”

The problem of detecting active users in a multiuser system has been addressed by several authors in a CDMA context \cite{29, 47, 69, 70, 74, 91}. Typically, the resulting multiuser receiver is a combination of two separate modules, namely, the active-user identifier and the multiuser detector. The treatment in \cite{47, 69, 70} focuses on the problem of detecting a single user entering or leaving the system. Reference \cite{91} advocates a subspace-based method (MUSIC algorithm) for identifying the active users, assuming that the receiver knows the pool of all spreading codes possibly used in transmission. In \cite{55}, the active-user-identification

\(^1\)See, e.g., \cite{30, 69, 91}. As an example, if a decorrelator detector \cite[Section 5]{82} does more nulling than needed, its performance is impaired. Ref. \cite{49} describes a case where a multiuser detector suffers from catastrophic error if a new user becomes active.
algorithm is subspace-based, as in [91]; the receiver is not interested in decoding all active users, but only those transmitting a message addressed to it. Reference [93] solves the problem of estimating the number of active users when synchronous and asynchronous users coexist in the system.

Further performance improvement can be expected if the receiver can use \textit{a priori} information about users entering and exiting the system. The knowledge of a traffic model is exploited in [30] in which the authors employ it to improve the detection of active users. They model bursty traffic for an individual source as a two-state Markov chain.

Other systems in which user identification is necessary include spatial multiplexing schemes, where the total system throughput can be optimized by properly selecting a subset of users to which the power is allocated [58, 94]. Thus, optimum power-control strategy requires the identification of this best set of users. In \textit{ad hoc} networks, optimal transmission strategies require the identification and localization of active nodes in the neighborhood of the transmitter.

More recently, the vibrant interest towards compressed sensing has brought up the issue of exploiting the sparsity of the data vector in underloaded CDMA systems [9, 10, 92]. The main drawback of such an approach, which may yield substantial performance improvement if the design hypotheses are met, is that it only works for underloaded systems, nor does it lend itself to a generalization to the case in which intercell interference must be taken into account, which causes an upper bound to the number of interferers to be unpredictable.\footnote{We may also observe that, in general, standard sparse signal processing can only be applied to situations where the state of a multiobject is represented by an ordered set of points, a “zero” in a specific position denoting the idleness of the corresponding subsystem of the multiobject under consideration. In other terms, techniques exploiting sparsity, and hence using conventional probability theory, are only applicable to multiobjects of which constituent subsystems are individually trackable, so that there is no need to deal with a sample space being a collection of sets (with no element ordering and arbitrary cardinality).}

\section*{4.2 Detection of Identities and Data}

We first consider the general problem of detecting the data and identities of a set of users whose cardinality is constant but unknown. Next,
we shall see how the receiver performance can be enhanced by using some information about traffic dynamics, i.e., a model for the evolution of the number of active users and their parameters. If this model is available, information from the past of the parameters may carry a considerable amount of extra information whenever their changes are not overly abrupt (for example, when the number of active users does not change too quickly from frame to frame).

4.2.1 Channel Model and Statement of the Problem

We assume that there is a reference user transmitting digital data over a common channel, also accessed by \( K \) interfering users. Let \( s(x_t^{(i)}) \) denote the signal transmitted at discrete time \( t, t = 1,2,\ldots \), by the reference user (identified by \( i = 0 \)) and by the \( K \) interferers, denoted by \( i = 1,\ldots,K \) (thus, index \( i \) reflects the identity of the user, and is typically associated with its signature). Each signal has a number of known parameters in it, reflected by the deterministic function \( s(\cdot) \), and a number of random parameters, summarized by vector \( x_t^{(i)} \). The observed signal at time \( t \) is a sum of \( s(x_t^{(0)}) \), the signals generated by the interfering users active at time \( t \), which are in a random number, and stationary random noise \( z_t \). We write

\[
y_t = \sum_{x_t^{(i)} \in X_t} s(x_t^{(i)}) + z_t
\]

where \( X_t \) is a random set, encapsulating what is unknown about all active users. We assume that the reference user is always active, that is, \( y_t \) is observed only when there is a useful signal.

The reader should notice here that the random set \( X_t \) has a less general structure than allowed by RST. In fact, we are in the situation, mentioned in the previous sections, in which every individual object (user) has a unique label (identity, signature). In addition, the observations consist of the random vectors \( y_t, t = 1,\ldots,T \), i.e., we consider a superpositional sensor. Here \( y_t \) has a conditional probability density function

\[
f_{y_t|X_t}(y_t \mid B) = f(x_t - \sigma(B))
\]
where \( B = \{b_1, \ldots, b_k\} \) is a realization of \( X_t \), that is, a realization of a random set of users and their data, \( f_Z(\cdot) \) is the probability density function (pdf) of the additive noise, and

\[
\sigma(B) \triangleq \sum_{b_i \in B} s(b_i).
\]

We now proceed to formulate the problem of detecting user 0’s data. The set of assumptions made leads to the following three scenarios, which we shall treat separately, in an order of increasing complexity:

1. The receiver has no information about the a priori probabilities that the individual interferers are active. Two alternative options we consider here are maximum-likelihood (ML) detection of the reference user’s data under the assumption that all potential users are active, and joint ML detection of the number of active users and of the reference user’s data. Consider binary transmission for simplicity. In the first case, detection implies choosing among \( 2 \times 2^K \) hypotheses. In the second case, the choice is among \( 2 \times 3^K \) hypotheses, as every interferer may transmit one out of two binary symbols, or be inactive. The difference in performance between the two situations is illustrated in Figure 4.1, which compares the two detectors previously described in terms of Bit Error Probability, or Bit Error Rate (BER) of the reference user. The ordinate shows the BER of the reference user in a multiuser system with two independent interferers transmitting binary antipodal signals over an additive white Gaussian noise (AWGN) channel with the same a priori probability of activity \( \alpha \), Kasami spreading sequences with length 15 [50, p. 240], and perfect power control. The single-user bound is also shown as a baseline. It is seen that RST yields a detector much more robust than classic MUD to variations in the (unknown) user activity factor. We also observe that classic MUD can outperform RST for high values of \( \alpha \), as this situation corresponds to having reliable side information about the number of active users.
Multiuser Detection

Fig. 4.1 Bit error probability of the reference user in a multiuser system with two interferers, independently active with probability \( \alpha \).

(2) The receiver knows the a priori probabilities that the individual interferers to be active. System performance can be further improved if the receiver is able to exploit additional side information in the form of a priori probabilities of user activity. By assuming that the activity factor \( \alpha \) is known, maximum a posteriori detection yields the results shown in Figure 4.2.

(3) The receiver has a dynamic model of users’ activity. The receiver performance can be further improved by using additional information about the interferers, in the form of a model of their dynamic behavior. This information can be generated once a model of users’ mobility is available, as discussed in the rest of this section.
4.2 Detection of Identities and Data

4.2.2 Defining Estimators

Development of estimators within our model must take into account the peculiarities of RST. In particular, one should recall that estimators based on a posteriori expectations may not exist, as mentioned in Section 2.2.

A possible estimator maximizes the a posteriori probability (APP) of $X_t$ given $y_{1:T}$, the latter denoting the whole set of observations corresponding to a data frame transmitted from $t = 1$ to $t = T$. Another possibility is to restrict oneself to a causal estimator, which searches for the maximum probability of $X_t$ given $y_{1:t}$. In a delay-constrained system, one may estimate $X_t$ on the basis of the observations $y_{t-\Delta:t+\Delta}$, with $\Delta$ a fixed interval duration (sliding-window estimator).
4.2.3 Consideration of a Dynamic Environment

Since \( \{X_t\}_{t=1}^{\infty} \) forms a random set sequence, the statistical characterization of \( X_t \) is needed for all discrete time instants \( t \). If a dynamic model of the transmission system is available, then the APP of \( X_t \) given \( y_{1:T} \) can be updated recursively, thus allowing one to take advantage of the information gathered from the past evolution of the system.\(^3\)

We make the assumption that \( \{X_t\}_{t=1}^{\infty} \) forms a Markov set sequence, i.e., that \( X_t \) depends on its past only through \( X_{t-1} \). This allows us to use Bayesian-filter recursions for \( \hat{X}_t \), as described in Section 2.3.

Thus, the causal maximum-a-posteriori estimate of \( X_t \) is obtained by maximizing, over \( B \), the APP \( f_{X_t|y_{1:t}}(B|y_{1:t}) \), which is tantamount to minimizing

\[
m(B) \triangleq (y_t - \sigma(B))^2 - \varepsilon(B)
\]

where \( \varepsilon(B) \triangleq N_0 \ln f_{X_t|y_{1:t-1}}(B|y_{1:t-1}) \), and \( N_0/2 \) is the power spectral density of additive Gaussian noise. The first term in the RHS of definition above is the Euclidean distance between the observation and the sum of the interfering signals at time \( t \). Its minimization yields ML estimates of \( X_t \). The second term in the RHS, generated by the upper-most step of iterations, reflects the influence on \( X_t \) of its past history, and its consideration yields MAP estimates.

From now on we restrict ourselves to the detection of the number and identity of active interferers, and of the data they carry, under the assumption that the remaining parameters, which were previously estimated by the receiver in a training phase, do not change in any appreciable way during the tracking phase. Estimation of these parameters will be described in Section 4.3.

4.2.4 Detection of Active Users

We assume first that we are only interested in detecting which interferers, out of a universe of \( K \) potential system users, are present

\(^3\) The concept of a time-adaptive detector was examined previously by several authors (see, e.g., [77] and references therein), while the effects on analysis of a dynamic model were touched upon by, among others, the authors of [28, 47, 69, 70].
4.2 Detection of Identities and Data

at time \( t \). This information may, for example, be used to do decorrelation detection, under the standard assumption that the signatures of all users are known to the receiver. In our theory, \( X_t \) takes values in \( 2^K \). Since this set is finite, a probability measure for \( X_t \) can be defined by assigning all probabilities \( P(A), A \in 2^K \).

4.2.4.1 Static model

At any fixed time \( t \), suppose that the probability of interferer \( x_t^{(i)} \) to be active is \( \alpha \), independent of \( t \) and \( i \). In this case the probability of the interferer set \( X_t \) depends only on its cardinality \( |X_t| \), and we can write

\[
f_{X_t}(B) = \alpha^{|B|}(1 - \alpha)^{K-|B|}. \tag{4.1}
\]

Although this result, in the simplified setting considered here, can be derived through classical probability theory, it is instructive at this stage to use RST by first computing the belief function

\[
\beta_X(S) \equiv P(X \subseteq S)
\]

\[
= \sum_{j=0}^{|S|} \sum_{B: B \subseteq S, |B|=j} P(X = B)
\]

\[
= \sum_{j=0}^{|S|} \binom{|S|}{j} \alpha^j (1 - \alpha)^{K-j}
\]

and subsequently computing its set derivative, which, in the discrete case, is computed using the Möbius transformation of (C.1b).

4.2.4.2 Dynamic Model

Consider now the evolution of \( X_t \). We assume that from \( t-1 \) to \( t \) some new users become active, while some old users become inactive. We write

\[
X_t = S_t \cup N_t \tag{4.2}
\]
where \( S_t \) is the set containing the identities of surviving users still active from \( t - 1 \), and \( N_t \) is the set of new users becoming active at \( t \). This model was described in Section 2.3.1.

Consider first \( S_t \). Suppose that there are \( k \) active users at \( t - 1 \), the elements of the random set \( X_{t-1} = \{x^{(1)}_{t-1}, \ldots, x^{(k)}_{t-1}\} \). Then we may write, for the set of surviving users,

\[
S_t = \bigcup_{i=1}^{k} X^{(i)}_{t}
\]

where \( X^{(i)}_{t} \) denotes either an empty set (if user \( i \) has become inactive) or the singleton \( \{x^{(i)}_{t}\} \) (user \( i \) is still active). Let \( \mu \) denote the “persistence” probability, i.e., the probability that a user survives from \( t - 1 \) to \( t \). We obtain, for the conditional probability of \( S_t \) given that \( X_{t-1} = B \),

\[
f_{S_t | X_{t-1}}(C \mid B) = \begin{cases} 
\mu^{|C|} (1 - \mu)^{|B|-|C|}, & \text{if } C \subseteq B \\
0, & \text{if } C \not\subseteq B.
\end{cases} \tag{4.3}
\]

A model for new users has

\[
f_{N_t | X_{t-1}}(C \mid B) = \begin{cases} 
\alpha^{|C|} (1 - \alpha)^{|K| - |B| - |C|}, & \text{if } C \cap B = \emptyset \\
0, & \text{if } C \cap B \neq \emptyset.
\end{cases} \tag{4.4}
\]

Finally, by assuming that births and deaths of users are conditionally independent given \( X_{t-1} = B \), the pdf of the union of the independent random sets \( S_t \) and \( N_t \) is obtained from the generalized convolution (2.11)

\[
f_{X_t | X_{t-1}}(C \mid B) = f_{S_t | X_{t-1}}(C \cap B) f_{N_t | X_{t-1}}(C \setminus (C \cap B)).
\]

### 4.2.4.3 Bayesian-filter Recursions

In our context, recursions (2.22a) and (2.22b) can be implemented as follows:

1. Assign the a priori probability distribution of \( X_0 \) at the beginning of the detection process. Description of this distribution consists in assigning probabilities to all the elements of \( 2^K \). This can be done by, for example, assuming independent users with the same stationary activity factor.
(2) Determine the conditional pdf’s \( f_{y_t|x_t} \), depending on the channel model.

(3) Determine the “evolution” pdf’s \( f_{x_t|x_{t-1}} \), depending on the dynamic model.

(4) Acquire the observations \( y_t \) and evaluate \( f_{x_t|y_1:t}(\cdot | y_{1:t}) \).

We note that here the RFS are defined on the (discrete) users identity set \( \mathbb{K} \), whereby the set integrals in (2.22a) and (2.22b) involve Lebesgue integrals with respect to the counting measure, and thus boil down to pure summations.

### 4.2.5 Active Users and their Data

Assume binary information data, independent in time and across users, and a discrete-time unit such that from \( t \) to \( t+1 \) each user transmits \( N \) binary symbols. In this case \( X_t \) is an RFS on \( \mathcal{S} = \mathbb{K} \times \{0,1\}^N \) of which the power set, which we denote as \( (1+2^N)^\mathbb{K} \), has cardinality

\[
\sum_{k=0}^{K} \binom{K}{k} 2^{kN} = (1+2^N)^K.
\]

Equation (4.1) becomes

\[
f_{X_t}(B) = 2^{-N|B|} \alpha^{|B|} (1 - \alpha)^{K-|B|}
\]

where the new factor \( 2^{-N|B|} \) accounts for the fact that there are \( N|B| \) equally likely binary symbols transmitted at time \( t \) by \( |B| \) users.

Similarly, (4.3) becomes

\[
f_{S_t|x_{t-1}}(C | B) = \begin{cases} 2^{-N|C|} \mu^{|C|} (1 - \mu)^{|B|-|C|}, & \text{if } i(C) \subseteq i(B) \\ 0, & \text{if } i(C) \not\subseteq i(B) \end{cases}
\]

and (4.4) into

\[
f_{N_t|x_{t-1}}(C | B) = \begin{cases} 2^{-N|C|} \alpha^{|C|} (1 - \alpha)^{K-|B|-|C|}, & \text{if } i(C) \cap i(B) = \emptyset \\ 0, & \text{if } i(C) \cap i(B) \neq \emptyset. \end{cases}
\]

In the previous equations we denoted \( i(B) \in 2^\mathbb{K} \) and \( i(C) \in 2^\mathbb{K} \) the sets containing the identities of the active users, i.e., the projections of the sets \( B \) and \( C \) on \( \mathbb{K} \), respectively.
4.2.6 Possible scenarios

We note that we assume that the only unknown signal quantities may be the identities of the users and their data. Specifically, we may distinguish four scenarios in our context:

1. Static channel, unknown identities, known data. This corresponds to a training phase targeted at identifying users, and assumes that the user identities do not change during transmission. In this case we again write $X$ in lieu of $X_t$.

2. Static channel, unknown identities, unknown data. This may correspond to a tracking phase following (1) above. We again write $X$ instead of $X_t$ and assume that $X$ contains the whole transmitted data sequence.

3. Dynamic channel, unknown identities, known data. This corresponds to the identification of users preliminary to data detection (which, for example, may be based on decorrelation).

4. Dynamic channel, unknown identities, unknown data. This corresponds to simultaneous user identification and data detection in a time-varying environment.

4.2.7 An Example of Application

Assume now the specific situation of a direct-sequence CDMA (DS-CDMA) system with signature sequences of length $L$ and additive white Gaussian noise. At discrete time $t$, we may write for the sufficient statistics of the received signal,

$$y_t = R A b_t(X_t) + z_t, \quad t = 1, \ldots, T$$

where $X_t$ is now the random set of all active users, $R$ is the $L \times L$ correlation matrix of the signature sequences (assumed to have unit norm), $A$ is the diagonal matrix of the users’ signal amplitudes, the vector $b_t(X_t)$ has nonzero entries in the locations corresponding to the active-user identities described by the components of $X_t$, and $z_t \sim N(0, (N_0/2)R)$ is the noise vector, with $N_0/2$ the power spectral density of the received noise.\footnote{Here the short-hand notation $N(\mu, C)$ is used to denote a Gaussian distribution with mean $\mu$ and covariance matrix $C$.} We further assume, for simplicity, $N = 1$, i.e.,
that at every discrete time instant only one binary antipodal symbol is transmitted.

### 4.2.7.1 Static Channel

The \textit{a posteriori} probability of $X$, given the whole received sequence (we omit the time subscript for simplicity), is\footnote{In order to keep the notation as simple as possible, here and in the following, we interchangeably use the notations $f_X$ and $f(X)$, both for random sets and for random vectors, whenever this does not generate confusion.}

$$f(X \mid y_1, \ldots, y_T) \propto f(X) f(y_1, \ldots, y_T \mid X) = e^{-\frac{1}{2} \sum_{t=1}^{T} (y_t - RA_b(X))^T R^{-1} (y_t - RA_b(X))} f(X).$$

Thus, the MAP estimator of user identities is

$$\hat{X} = \arg \max_{X \in 2^K} f(X \mid y_{1:T})$$

where, as usual, $y_{1:T} \triangleq (y_1^T \ldots y_T^T)^T$. The MAP estimator of user identities and data is

$$\hat{X} = \arg \max_{X} f(X \mid y_{1:T})$$

where the set of possible realizations of $X$ includes $\binom{K}{|X|} 2^{|X|T}$ elements: in fact, in $T$ time interval the number of transmitted binary symbols is $2^{|X|T}$, and

$$\sum_{|X|=0}^{K} \binom{K}{|X|} 2^{|X|T} = (1 + 2^T)^K.$$
The introduction of this “fine-grain” notation for the random set suggests that the MAP detector may be implemented in the form of a sequential detector, thus simplifying its operation (more on this subsequently).

4.2.7.2 Dynamic Channel

Now consider a dynamic channel, and examine first the case of known data. We have, accounting for the Markov property of our channel model,

\[
f(X_1, \ldots, X_T | y_{1:T}) \propto f(y_{1:T} | X_1, \ldots, X_T) f(X_1) \prod_{t=2}^{T} f(X_t | X_{t-1})
\]

with \( f(X_1) \) a density for which assignment is based upon prior knowledge of the channel state at the beginning of the transmission. The MAP estimator here maximizes the RHS of the above (or its logarithm) with respect to the values taken on by the sequence \( (X_1, \ldots, X_T) \). Even in this case we may think of a sequential detector, which searches for the maximum-APP path traversing a trellis [20, p. 126] having \( T \) stages and a number of states at stage \( i \) equal to the number of realizations of the random set \( X_i \).

**Implementing a sequential detector.** Implementation of the sequential detector through a version of the Viterbi algorithm has the following consequences:

1. The decision on the whole sequence of users’ identities and their data should be taken only after the whole sequence of observations \( y_1, \ldots, y_T \) has been recorded.
2. The decision on the user identities and their data at time \( t \) depends not only on the past observations, but also on observations that have not been recorded yet at time \( t \).
3. A suboptimum version of the optimum sequential algorithm, the *sliding-window Viterbi* algorithm (see, e.g., [20, p. 133 ff.]) can be implemented. This consists of forcing a decision on \( X_t, b_t(X_t) \) based on a sliding window of observations that includes \( y_t \), but of which the length is smaller than \( T \).
4.2.8 Numerical Results

In this section, some numerical examples illustrate the theory developed before. Figure 4.3 shows how the knowledge of the channel dynamics can improve the performance of a multiuser detector. There are a maximum of $K = 3$ users, whereby the “reference user” may experience, at most, two interferers, with $\alpha = 0.2$ and $\mu = 0.8$. We consider the following situations:

1. all of the three users are considered active;
2. only the activity factor $\alpha$ is taken into account;
3. the entire channel dynamics are taken into account, and non-causal Viterbi-based demodulation is undertaken;
4. the entire channel dynamics are taken into account, and Bayesian recursions are implemented, i.e., the causality constraint is assumed.

![Figure 4.3](image)

Fig. 4.3 Bit error probability of the reference user in a multiuser system with two interferers, following the dynamic model described above with $\alpha = 2$ and $\mu = 0.8$. 
As expected, even this simplified scenario confirms the advantages of a full exploitation of the prior knowledge as to the channel statistics in an RFS framework, suggesting a marginal loss if the causality constraint is assumed.

Figure 4.4 refers to a trained system with the same configuration as in Figure 4.3, tracked across an interval of length $T = 10$. The ordinate shows the set error probability (SEP), i.e., the probability of an erroneous estimate of the active-user set. Here, a comparison is made between a noncausal Viterbi-based set estimate and a causal estimate, obtained through Bayesian-filter recursions. To illustrate the impact of the causality constraint, we represent the SEP for the set $X_1$, where the causality constraint prevents sequence detection, and for the set $X_{10}$, where such a constraint has no effect: as expected, the performances of the Viterbi algorithm and of the Bayesian recursions coincide when estimating $X_{10}$ while the causality constraint has a perceivable effect on the performance when $X_1$ is estimated.

Fig. 4.4 Trained acquisition of the set of active users through the Viterbi Algorithm and Bayesian recursion: effect of the causality constraint.
4.2 Detection of Identities and Data

Before proceeding to our result presentation, it may be worthwhile to underline that all curves in the previous figures were derived from computer simulation, which becomes computationally heavy for long transmitted frames and/or large $K$. Because of this, semianalytical methods were developed for the optimum Viterbi-based estimators in order to obtain a computable bound to the achievable performance (see [22] for details).

An example of the validity of such an approximation is illustrated in Figure 4.5, referring to a static channel and to the case in which the active users transmit a known sequence of bits in order to be identified. We assume here that all users (including the reference user) are active with probability $\alpha = 0.5$. In this situation $K = 6$, the transmitted signals are binary antipodal, spreading is done through $m$-sequences with length 7, the power control is perfect (hence, $A$ is a scalar matrix) and the data-frame length varies from $T = 1$ to $T = 3$. Since the channel

![Graph](image_url)

Fig. 4.5 Set-error probability with ML detection based on RST with $K = 6$, $L = 7$. Comparison among “exact” probability (obtained by simulation), union-bound to it (denoted $P(e)$), and an approximation to the union bound (see [22, Eq. (41)]).
is static, the measure of interest here is the SEP and its behavior in relation to $T$, the length of the training phase. Not surprisingly, larger values of $T$ result in more reliable estimates of the RFS of the active users, as shown by the simulation results and by the semianalytical approximation (the latter follows closely the “experimental” points in the large-SNR region, which is where the union bound becomes tight).

A global figure of merit of both trained and untrained systems in a dynamic environment is provided by the Set Sequence Error Probability (SSEP), namely, the probability that for some $t$, $0 \leq t \leq T$, the estimated set $\hat{X}_t$ differs from the true set $X_t$ either in its cardinality or in its elements. For trained systems, this is the probability that, for some $t$, the number and/or the identities of the active users are not correctly identified, while, for untrained systems, it is the probability that, for some $t$, estimated and true set differ in the cardinality and/or in the identities of the active users and/or in the transmitted data. Plots of the SSEP are shown in Figure 4.6 for a trained system with

![Fig. 4.6 Set sequence error probability with $K = 6, L = 7$. Also, the curve corresponding to a semi-analytical performance evaluation under trained acquisition of the set of the active users is included.](image-url)
a maximum number of active users $K = 6$. The figure also shows the curve obtained through the semianalytical approximation mentioned in previous section, which follows the numerical results quite closely.

The case in which not only the identities, but also the data of the active users are to be estimated is shown in Figure 4.7, assuming a maximum of $K = 3$ active users and, again, $L = 7$; the data-frame length is $T = 10$. Here we compare a Viterbi-algorithm receiver with one based on Bayesian recursions for estimating the set of interferers and the transmitted bits. The ordinate shows the Bit-Sequence Error Probability (BSEP), at time $t = 1$ and at time $t = T = 10$, defined as the probability that the estimated and the true set do not coincide: the term “bit sequence error probability” is tied here to the fact that an error in estimating the identities of the active users automatically implies an error in estimating the stream $b(X_t)$, while the converse is not true. Once again, the effect of the causality constraint on the

![Graph showing BSEP vs. $\frac{E_b}{N_0}$ for different methods.](image)
performance is elicited, and the results are in accordance with the intuition as well as with the curves of Figure 4.4.

We finally observe that all the above receivers have a complexity growing exponentially with $K$ and $T$, which may make their implementation prohibitive. However, in [7] it is shown how a zero-order approximation of the Bayesian recursions (see Section 3.4) allows the application of the Sphere Decoding algorithm to estimate the (causal) optimal set sequence in both trained and untrained systems.

## 4.3 Data Detection and Parameter Estimation

As anticipated in previous sections, if the channels linking the users to the access point are only partially known (and hence the received signals contain a number of continuous unknown parameters), the object of interest is an RFS on a hybrid space, i.e., MUD amounts to user identification and/or data detection and channel parameter estimation.

### 4.3.1 Channel Model and Statement of the Problem

In this section, we implement a slight change of notations that is more suitable. Specifically, we assume $K$ users transmitting digital data over a common channel and denote again $s(x_t^{(i)})$ the signal from user $i$, $i = 1, \ldots, K$, at discrete time $t$, $t = 1, 2, \ldots$. Each signal has in it a number of known parameters, included in the deterministic function $s(\cdot)$, and a number of random parameters, summarized by the vector $x_t^{(i)}$. The observed signal at time $t$ is a sum of the signals generated by all users active at time $t$, which are in a random number, and of stationary random noise $z_t$. We write again

$$y_t = \sum_{x_t^{(i)} \in X_t} s(x_t^{(i)}) + z_t$$  \hspace{1cm} (4.5)

where, as before, $X_t \triangleq \{x_t^{(1)}, \ldots, x_t^{(k)}\}$, $k$ a random integer, is the random set encapsulating what is unknown about the active users, which needs to be estimated. $X_t$ is thus defined over a hybrid space $S = W \times \mathbb{R}^d$ as defined in Section 2. This consists of all pairs $(u, d)$, with $d \in \mathbb{R}^d$ and $u \in W$. $\mathbb{R}^d$ contains the values of the $d$ real parameters.
4.3 Data Detection and Parameter Estimation

of the users, and $W$ is a discrete set. In our context, we may have, for example, $W = \{1, \ldots, K\} \triangleq \mathbb{K}$, the set of the identities of the users, or $W = \mathbb{K} \times \{+1, -1\}$, the set of users transmitting binary antipodal data. We use again the notation, defined in Section 2, where $\pi(G)$ denotes the projection of $G$ onto $\mathbb{K}$. Similarly, $\pi'(G)$ is the projection of $G$ onto $\mathbb{R}^d$.

Formally, at time $t$ we have

$$X_t = \bigcup_{k \in \mathbb{K}} X_t^{(k)}$$

with $X_t^{(k)}$ a singleton-or-empty set:

$$X_t^{(k)} = \begin{cases} \{x_t^{(k)}\}, & \text{if user } k \text{ is active at time } t \\ \emptyset, & \text{otherwise.} \end{cases}$$

We may consider two distinct situations, namely:

1. **Known-data ("trained") systems**, where $W = \mathbb{K}$, $S = \mathbb{K} \times \mathbb{R}^d$, and $x_t^{(k)} = (k, a_t^{(k)})^T$, with $a_t^{(k)}$ a $d$-dimensional random vector.

2. **Unknown-data ("untrained") systems**, where $W = \mathbb{K} \times \mathcal{M}$, $S = \mathbb{K} \times \mathcal{M} \times \mathbb{R}^d$, with $\mathcal{M}$ an $M$-ary symbol alphabet, and $x_t^{(k)} = (k, d_t^{(k)}, a_t^{(k)})^T$, $d_t^{(k)}$ the transmitted symbol.

Observe that, in a trained system, if $g \in G$, then $g = (\pi(g), \pi'(g))$.

### 4.3.1.1 Measurement Model

With measurement model (4.5), the receiver detects only a superposition of signals, whereby the observation vector at time $t$ has conditional pdf given by

$$f(y_t | X_t) = f_Z\left(y_t - \sum_{x_t^{(k)} \in X_t} s(x_t^{(k)})\right)$$

with $f_Z(\cdot)$ the pdf of the additive noise.

### 4.3.1.2 Dynamic Model

The multiuser set can be expressed in a fairly general form using model (4.2). In addition, we make the assumption that $\{X_t\}_{t=1}^\infty$ forms
Multiuser Detection

a Markov set sequence, i.e., that $X_t$ depends on its past only through $X_{t-1}$, and that the death-and-birth process follows a binomial law. More precisely, we denote by $\alpha$ the probability that a user silent at time $t-1$ becomes active at time $t$, and by $\mu$ the probability that a user active at epoch $t-1$ survives into epoch $t$. We assume a trained system first. The conditional belief measures (given $X_{t-1} = B$) of surviving and newly born user sets are written as

$$
\beta_{S_t|X_{t-1}}(Z | B) = \sum_{j=0}^{|\pi(Z)|} \sum_{\substack{G: |\pi(G)| = j \land \pi(G) \subseteq \pi(Z)}} f_{\pi(S_t)|X_{t-1}}(\pi(G) | B) \times P\{S_t \subseteq G | X_{t-1} = B, \pi(S_t) = \pi(G)\}
$$

$$
\beta_{N_t|X_{t-1}}(Z | B) = \sum_{j=0}^{|\pi(Z)|} \sum_{\substack{G: |\pi(G)| = j \land \pi(G) \subseteq \pi(Z)}} f_{\pi(N_t)|X_{t-1}}(\pi(G) | B) \times P\{N_t \subseteq G | X_{t-1} = B, \pi(N_t) = \pi(G)\}
$$

with, as supra,

$$
f_{\pi(S_t)|X_{t-1}}(\pi(G) | B) \triangleq P\{\pi(S_t) = \pi(G) | X_{t-1} = B\} = \begin{cases} 
\mu^{|\pi(G)|}(1 - \mu)^{|B| - |\pi(G)|}, & \text{if } \pi(G) \subseteq \pi(B) \\
0, & \text{otherwise}
\end{cases} \quad (4.6)
$$

and

$$
f_{\pi(N_t)|X_{t-1}}(\pi(G) | B) \triangleq P\{\pi(N_t) = \pi(G) | X_{t-1} = B\} = \begin{cases} 
\alpha^{|\pi(G)|}(1 - \alpha)^{|B| - |\pi(G)|}, & \text{if } \pi(G) \cap \pi(B) = \emptyset \\
0, & \text{otherwise.}
\end{cases} \quad (4.7)
$$

Likewise, assuming that $a_t^{(k)}$ and $a_t^{(m)}$ are independent for $k \neq m$, we have:

$$
P\{S_t \subseteq G | X_{t-1} = B, \pi(S_t) = \pi(G)\} = \int_{\pi'(G)} \prod_{k \in \pi(G)} f_{a_t^{(k)} | a_{t-1}^{(k)}}(a_t^{(k)} | b^{(k)}) da_t^{(k)}
$$
\[ \mathbb{P}(\{N_t \subseteq G \mid X_{t-1} = B, \pi(N_t) = \pi(G)\}) = \int_{\pi'(G) \cap \pi(G)} \prod_{k \in \pi(G)} f_{\alpha_t}^{(k)}(a_t^{(k)}) da_t^{(k)}. \]

Computing the set derivative of the above, we obtain, with some notational abuse:

\[ f_{X_t|X_{t-1}}(C \mid B) = f_{\pi(S_t)|X_{t-1}}(\pi(C) \cap \pi(B) \mid B) \times \prod_{k \in \pi(C) \cap \pi(B)} f_{a_t^{(k)}|a_{t-1}^{(k)}}(a_t^{(k)} \mid b^{(k)}) \times f_{\pi(N_t)|X_{t-1}}(\pi(C) \setminus \pi(B) \mid \pi(B)) \times \prod_{k \in \pi(C) \setminus \pi(B)} f_{a_t^{(k)}|a_{t-1}^{(k)}}(a_t^{(k)}) \]

\[ = \mu^{\pi(C) \cap \pi(B)}(1 - \mu)^{|B| - |\pi(C) \cap \pi(B)|} \times \alpha^{\pi(C) \setminus \pi(B)}(1 - \alpha)^{K - |B| - |\pi(C) \setminus \pi(B)|} \times \prod_{k \in \pi(C) \setminus \pi(B)} f_{a_t^{(k)}|a_{t-1}^{(k)}}(a_t^{(k)} \mid b^{(k)}) \times \prod_{k \in \pi(C) \setminus \pi(B)} f_{a_t^{(k)}}(a_t^{(k)}). \quad (4.8) \]

The companion case of untrained systems may be dealt with in a very similar way. Indeed, if we denote by \( d_t^{(k)} \) the data transmitted by the \( k \)th user, the Markov assumption on \( X_t \) is not violated under either one of the following conditions:

1. The data sources are memoryless.
2. The data sources are Markovian; in this case, the densities \( f_{d_t^{(k)}}(\cdot) \) and \( f_{d_t^{(k)}|d_{t-1}^{(k)}}(\cdot | \cdot) \) are assumed to be known.

Under condition (2), which trivially subsumes (1), (4.8) generalizes to:

\[ f_{X_t|X_{t-1}}(C \mid B) = \mu^{\pi(C) \cap \pi(B)}(1 - \mu)^{|B| - |\pi(C) \cap \pi(B)|} \times \alpha^{\pi(C) \setminus \pi(B)}(1 - \alpha)^{K - |B| - |\pi(C) \setminus \pi(B)|} \times \prod_{k \in \pi(C) \setminus \pi(B)} f_{a_t^{(k)}|a_{t-1}^{(k)}}(a_t^{(k)} \mid b^{(k)}) \times \prod_{k \in \pi(C) \setminus \pi(B)} f_{a_t^{(k)}}(a_t^{(k)}). \]
\[
\times \prod_{k \in i(C) \cap i(B)} f_{d_t^{(k)}}(d_{t-1}^{(k)} | d_{t-1}^{(k)}) f_{a_t^{(k)}}(a_{t-1}^{(k)} | b^{(k)})
\times \prod_{k \in i(C) \setminus i(B)} f_{d_t^{(k)}}(d_t^{(k)}) f_{a_t^{(k)}}(a_t^{(k)})
\]

where, once again, \(i(B) \in 2^K\) and \(i(C) \in 2^K\) denote sets containing the user identities only.

### 4.3.2 Optimal Bayesian Sets Estimators

In causal dynamic set estimation, a random set at time \(t\) is described by its \textit{a posteriori} belief density given the observations up to time \(t\), which we denote \(f_{X_t | Y_1:t}(B | y_1:t)\). A common method for the evaluation of the above density relies on Bayesian recursions (2.22a) and (2.22b).\(^6\)

It is important to notice that each active user has a unique identifier, whereby set estimation may be performed by resorting either to the two classical GMAP-I and GMAP-II RFS estimators or to the GMAP-III estimator (see Section 2.2.1). As explained in Section 2.2.1.3, the GMAP-III estimator relies on estimating first the discrete part \(\pi(X_t)\) of the RFS of interest, next implementing an “optimal” estimator for the continuous parts of the elements detected as “active.” In the context of optimal MUD, this suggests introducing two performance measures, i.e., the Discrete Set Error Probability (DSEP), and the Root Mean Square Error (RMSE) between the true and the estimated continuous parameters of the users detected as active. This is discussed below.

### 4.3.3 An Example of Application

Following [22], in this section we apply the above methodologies to the scenario of an uncoded, synchronous, single-rate DS-CDMA system with a maximum number \(K\) of users, processing gain \(L\), and additive white Gaussian noise. Let us assume coherent detection, while the amplitudes are unknown. The chipwise matched-filtered received signal

---

\(^6\) We could also look for a fixed-lag smoothing estimation, but we neglect this attractive alternative here.
at time $t$ is

$$y_t = Sv(X_t) + z_t, \quad t = 1, 2, \ldots, T$$

where:

1. $S$ is an $L \times K$ matrix of which the columns are the spreading codes of the users.
2. $v(X_t)$ is a $K$-dimensional vector with non-zero entries in the locations dictated by $\pi(X_t)$. In particular, we have
   $$v_m(X_t) = \begin{cases} d^{(m)}_t a^{(m)}_t, & \text{if } m \in i(X_t) \\ 0, & \text{otherwise}. \end{cases}$$
3. $z_t \sim N(0, (N_0/2)I_L)$ is an $L$-dimensional white Gaussian noise vector, where $N_0/2$ is the power spectral density of the received noise and $I_L$ denotes the $L \times L$ identity matrix.

Consequently, the set $X_t$ consists of a random number of elements, each element containing the active-user identity, the transmitted data, and the received power (in random-set theoretical parlance, the hybrid space is now $\mathbb{S} = \mathbb{K} \times \mathbb{M} \times \mathbb{R}^+$, with $\mathbb{M}$ the symbol alphabet). Thus, the measurement model is

$$f_{y_t \mid X_t}(y_t \mid C) = f_z(y_t - Sv(C)). \quad (4.9)$$

We consider a binomial birth-and-death process with parameters $\alpha$ and $\mu$ as in (4.6) and (4.7). Assuming independent, equally likely symbols and memoryless modulation, we have

$$f(d^{(k)}_t \mid d^{(k)}_{t-1}) = f(d^{(k)}_t) = \frac{1}{|\mathbb{M}|} = \frac{1}{M}, \quad \forall k \in \pi(X_t)$$

(in a training phase, employed for user identification and channel estimation, the symbols $d^{(k)}_t$ would form a known sequence).

4.3.4 Numerical Results

Some numerical examples will now illustrate the theory developed above.
**Trained CDMA system.** A synchronous, uncoded, single-rate DS-CDMA system with processing gain $L = 7$ is employed in this simulation. The maximum number of active users is $K = 3$. Each user is assigned an $m$-sequence of length 7, and the transmission duration is $T = 10$. The measurement model is (4.9), while the dynamic model is the one described in (4.8). The birth rate and the persistence probability are $\alpha = 0.3$ and $\mu = 0.9$, respectively. The dynamic power model follows that defined in [5, Appendix], where $\rho = 0.9999$, $\varepsilon = 1.2$, and $\beta = 0.5$ for all users. We assume that the users transmit known symbols (trained system), whereby only the identities and the power of active users are to be estimated. In approximating Bayesian recursions, a bootstrap filter has been used with $N = 1000$ particles and threshold $N_{\text{th}} = N/4$. As anticipated, the DSEP is
\[
\mathbb{P}(\exists t \in \{1, \ldots, T\} : \hat{\pi}(X_t) \neq \pi(X_t))
\]
and only measures the system ability to identify the active users, whereby a further performance measure is needed to assess the accuracy of the power estimates. To this end, consider the standard RMSE
\[
\sqrt{\sum_{k \in \hat{\pi}(X_t) \cap \pi(X_t)} \left( \hat{\alpha}_t^{(k)} - \alpha_t^{(k)} \right)^2}
\]
(4.10)
Figure 4.8 shows the DSEP for the set estimator I, set estimator II, and set estimator III of Section 2.2.1, hereafter called GMAP-I, GMAP-II, and GMAP-III. As a baseline for performance comparisons, we also report the DSEP corresponding to Complete Channel-State Information (labeled CCSI), i.e., perfect knowledge of the channel parameters (in this case, the power). Figure 4.9 shows the distance (4.10) for the GMAP-I and GMAP-II estimators.

**Blind CDMA system.** The case of blind systems very much resembles that of trained systems, and hence we only illustrate the simulation results. The system parameters are the same as in Section 4.3.4, and binary transmission is assumed ($M = \{\pm 1\}$). Of course DSEP now measures the probability that the estimated set sequence containing the identities and the data of active users differs, for some $t$, from the true
Fig. 4.8 Discrete Set Error Probability for various receivers (trained CMDA system).

Fig. 4.9 Root-mean-square error for various receivers (trained CMDA system).
Fig. 4.10 Discrete Set Error Probability for various receivers (blind CMDA System).

Fig. 4.11 Root-mean-square error for various receivers (blind CMDA system).
set in either its cardinality, or its elements, or both. Figure 4.10 shows such a DSEP for GMAP-I, GMAP-II, GMAP-III, and CCSI receivers, while Figure 4.11 shows the set distance for GMAP-I and GMAP-II.

As a final remark, we observe that even in this case the complexity of optimal set estimators may prove to be a major bottleneck. However, coupling a zero-order approximation of the Bayesian recursions (see, again, Section 3.4) with properly designed gridding algorithms leads to reduced-complexity causal receivers, of which the complexity scales down to algebraic, and eventually to linear for increasingly large signal-to-noise ratios, as described in [7].
5

Channel Estimation

5.1 Introduction and Motivation

Wireless communication through “wideband” signals, i.e., signals of which the bandwidth largely exceeds the channel coherence bandwidth, has now, possibly in conjunction with Multiple-Input, become Multiple-Output (MIMO) architectures [21], a standard of transmission systems for virtually any application. Coherent transmission requires acquisition of Channel State Information (CSI), whereby transmitted data are interleaved with training signals, spaced in time by less than the channel coherence time, in order to keep the receiver updated about the channel evolution. This forces the channel to be modeled as a time-varying, frequency-selective system [23]. In particular, the transmitted signal suffers from multipath distortion, generated by transmission paths of which number, location, and gain are varying over time. Such a scenario becomes even more challenging when MIMO architectures are considered, since in that case CSI should be acquired and possibly tracked for each spatial channel.

In general, if no prior information is available about the channel evolution, not much can be done other than extracting the CSI at the end of each training phase through standard estimation procedures, in
which the channel state is modeled by a nonrandom vector. This is, for example, the approach taken in [18] for a MIMO system with Orthogonal Frequency Division Multiplexing (OFDM), wherein a Least-Square (LS) estimate of the channel is formed *ex-novo* after a training phase, and employed to demodulate the next information-bearing symbols. A more general approach to non-Bayesian MIMO–OFDM channel estimation is the one described in [54], establishing (parameter) Cramér–Rao bounds on the achievable accuracy.

If, instead, prior information on the channel evolution is available, through, e.g., a conditional probability density function of the channel state — now modeled as a random process — at a given time, given its past history, then more sophisticated estimators can be conceived. A standard approach may consist of importing to the new scenario well-established methodologies from neighboring disciplines — primarily, target tracking. This is the approach taken, for example, in [56], in order to track the state of a time-selective MIMO channel, or, in a slightly more general framework, in [53], where Kalman filtering is applied to the problem of joint channel tracking and equalization. More recent papers have advocated the use of EKF or distribution-free sequential estimation [52]. Even in the presence of a realistic channel evolution model, however, undertaking conventional sequential estimation amounts to neglecting a number of aspects of the problem, namely:

1. The number of active paths may be time-varying, with some paths perishing and new paths being born.
2. The path location itself can be time-varying, and the corresponding delays are not necessarily integer multiples of the inverse of the transmitted bandwidth.
3. The evolution model should be realistic with regard to the conditional densities of the taps and in addition able to reproduce the death-and-birth process in a tractable way.

In a conventional approach, item (1) is generally addressed by assuming that all paths are simultaneously active and estimating the corresponding gains. This way, an inactive-path gain will eventually be estimated as nearly zero, while a newly born path starts a transient until a reliable estimate of its gain is available. Items (2) and (3) have not received
much attention, mainly because of the lack of a flexible tool to handle the variability of the number of objects of which the state is to be estimated. This section illustrates how RST can logically address the previous three above, while yielding effective and relatively simple solutions to the channel estimation problem. It is worth emphasizing that RST has an inherent added value with respect to conventional methods, as discussed later on: indeed, estimating the set of active paths means jointly detecting their number and estimating their “states.” Path deaths or births are immediately detected, thus avoiding the transient phases typical of a conventional approach in which a path is declared active or inactive only when its state is “conveniently far from” or “conveniently close to” idleness, respectively.

5.2 A Case Study: SISO Wide-Band Transmission

Let us at first consider a Single-Input, Single-Output (SISO) system with transmission rate \(1/T_s\) and occupying a bandwidth \(W \gg (\Delta f)_c\), where \(T_s\) is the symbol interval and \((\Delta f)_c\) denotes the coherence bandwidth of the wireless channel. Assume that a training phase occurs at \(I\) time instants multiple of \(T_s\), say \((i-1)PT_s, i = 1, \ldots, I\), with \(PT_s\) not exceeding the channel coherence time \((\Delta t)_c\). The training phase consists of the transmission of a known pilot signal, say \(S(t)\), with energy \(E_s\), occupying the whole signal bandwidth \(W\). During a single training period the channel behaves as a time-invariant, frequency-selective system, but becomes time-varying when observed over different training epochs. Consequently, upon transmission of the pilot, the signal observed in the \(i\)th training phase has the form

\[
y(t) = \sum_{\ell \in \mathcal{L}_i} h[(i-1)PT_s; \tau_{\ell,i}]S(t - \tau_{\ell,i}) + w(t)
\]  

where \(\mathcal{L}_i \subseteq \mathbb{N}\) represents the set indexing the paths that are active in such a phase of which the number is approximately upper-bounded by \(L_{\text{max}} = W/(\Delta f)_c\), \(0 \leq \tau_{\ell,i} \leq 1/(\Delta f)_c\) the corresponding delays, while \(h[(i-1)PT_s; \tau_{\ell,i}] = h_{i,\ell}\) is the channel gain, and \(w(t)\) denotes white noise.

In practice, channel estimation requires observations that are matched-filtered and sampled at a rate which is an integer multiple
of $W$, say $rW$, where $r$ is a system parameter dictating the channel estimation accuracy. Therefore, the discrete-time version of (5.1) reads:

$$y_i = \frac{1}{\sqrt{E_s}} R_s h_i + w_i, \quad i = 1, \ldots, I. \quad (5.2)$$

The dimension of vector $y_i$ is $L_{\text{max}} = \lceil rWT \rceil$, with $r$ an integer and $T$ the duration of the training phase, while the matrix $R_s \in \mathbb{C}^{L_{\text{max}} \times L_{\text{max}}}$ has, as its $(j, \ell)$ entry, the correlation of the pilot signal $S(t)$ evaluated at $j/rW - \tau_{\ell,i}$, and the vector $h_i$ has dimension $L_{\text{max}}$. The conventional approach assumes that all the entries of vector $h_i$ are active, albeit possibly of zero value. This requires assigning a conditional density to the vector sequence $h_i$ (assumed Markovian). The conventional Bayesian recursions are implemented in order to sequentially estimate $h_i$ based on the observations available up to epoch $i$.

The object of interest here is the posterior density

$$f(h_i \mid y_{1:i}) \quad (5.3)$$

where, as usual, $y_{1:i} \triangleq (y_1^T \cdots y_i^T)^T$ represents the observations sequence. Maximization of (5.3) produces the MAP estimate. If this density is Gaussian, the estimation problem is solved through the conventional KF, producing the optimum channel tracking procedure. Notice that, as anticipated, this approach relies on tracking the evolution of all of the $L_{\text{max}}$ potential paths, even when some of them have gain zero. This may imply a waste of resources, as the computational resources of the estimator can be used more productively if only the nonzero path gains are estimated. Moreover, since some paths may die and new paths may be born, implementing the previous procedure inevitably leads to transient phases with potentially deleterious effects on performance. In other words, the problem of channel tracking in a time-varying environment entails detection of active paths and estimation of their relevant parameters, while resorting to (5.3) amounts to just passing over the prior uncertainty as to the presence of the paths and performing the detection ex-post, i.e., concluding that a path is

---

1In this section, in order to keep the notation as simple as possible, whenever this does not generate confusion, we interchangeably use the notations $f_X$ and $f(X)$, both for random sets and for random vectors.
absent if its amplitude is “small enough.” Indeed, by going back to (5.2) it is seen that the original problem entails estimating the support of the vector $h_i$ and, for its nonzero entries, the relevant parameters.\textsuperscript{2} To illustrate this concept, we rewrite (5.2) in a more general form. First of all, define a collection of singleton-or-empty random sets in the form:

$$H_i^{(\ell)} = \begin{cases} \{ (\tau_{\ell,i}, \Re \{ h_{i,\ell} \}, \Im \{ h_{i,\ell} \})^T \} , & \text{if path } \ell \text{ is active at epoch } i \\ \emptyset , & \text{otherwise} \end{cases}$$

$(\cdot)^T$ denoting transpose. As a consequence, the channel state at epoch $i$ is encapsulated in the realization of the disjoint union of these singleton-or-empty random sets, i.e.,

$$H_i = \bigcup_{\ell=1}^{L_{\max}} H_i^{(\ell)}$$

namely the channel state, when observed over the training epochs, is a sequence of finite random sets built upon $\mathbb{R}^3$. It is probably worth observing here that the assumption that $\tau_{\ell,i}$ is an integer multiple of $1/(rW)$ would lead to an RFS built upon the hybrid space $S = \{1, \ldots, L_{\max}\} \times \mathbb{R}^2$. Similarly, should more parameters be needed (like the Doppler frequency for highly time-varying channels), this notation can be easily generalized by changing the space on which the random set is built.

In conclusion, in a random-set-theoretic framework the signal model (5.2) is conveniently recast — with a slight notational abuse — in the form

$$y_i = s(H_i) + w_i, \quad i = 1, \ldots, I.$$  \hfill (5.4)

Notice that estimating the random set $H_i$ is tantamount to estimating its cardinality (i.e., the number of paths active at epoch $i$) \textit{and} its elements, i.e., the indexes of the nonzero entries of vector $h_i$ — its support, the delay, the channel gains, and so on: no transient is started as

\textsuperscript{2}As an \textit{obiter dictum}, observe that if the vector has only few non-zero entries, then sparse estimation procedures can be used equivalently.
an effect of path births and deaths, no resource is wasted in estimating
the amplitudes of absent paths, and, above all, no decoupling between
detection and estimation takes place.

To apply the theory outlined in the previous sections, we now need
to define a viable model for the channel dynamics, leading to a channel-
tracking solution based on Bayesian recursions, and a suitable sequential
estimator of the channel state.

### 5.3 Channel Dynamics

In this section we keep on considering the SISO channel of Sec-
tion 5.2, outlining a possible Bayesian model for the random set
sequence \( \{ \mathbf{H}_i \} \). As anticipated, and using the model introduced in
Section 2.3.1, the time-varying fading channel will be characterized by
a death-and-birth model for the active paths of which the gains vary
over time with a known statistic. For the sake of simplicity, we start
assuming that the delays may only be integer multiples of \( 1/(rW) \),
while the case of noninteger delays will briefly be dealt with at the end
of this section. The hybrid space here is \( S = \{1, \ldots, L \} \times \mathbb{R}^2 \),
where \( L \) is either \( L_{\text{max}} \) or a suitably defined upper bound to the largest delay.
Consequently, a general realization of \( \mathbf{H}_i \), of which cardinality is the
random variable \( |\mathbf{H}_i| = m_i \), is

\[
\mathbf{H}_i = \left\{ \left( \ell_1 \ a_i^{(\ell_1)} \ b_i^{(\ell_1)} \right)^T, \ldots, \left( \ell_{m_i} \ a_i^{(\ell_{m_i})} \ b_i^{(\ell_{m_i})} \right)^T \right\}
\]

where \( a_i^{(\ell_k)} \) and \( b_i^{(\ell_k)} \) represent real and imaginary part of the replica
starting at \( \ell_k/(rW) \), if present. Notice that in any case the nature of the
problem allows defining the two projections of \( \mathbf{H}_i \), \( \pi(\mathbf{H}_i) \) and \( \pi'(\mathbf{H}_i) \),
onto the countable and the continuous parts of \( S \), respectively, so that
we have:

\[
|\mathbf{H}_i| = |\pi(\mathbf{H}_i)| = |\pi'(\mathbf{H}_i)|. \tag{5.5}
\]

This is once again a direct consequence of the fact that all the elements
in \( \mathbf{H}_i \) have distinct integer parts, a key difference with other applications,
such as multitarget tracking, wherein the discrete parts typically
denote target types and/or threat levels, and hence equalities (5.5) may
not hold.
Similar to the considerations in Section 2.3.1, we may write, for the channel evolution,

\[ H_i = S_i \cup N_i \]

\[ \pi(S_i) \subseteq \pi(H_{i-1}) \]

\[ \pi(N_i) \cap \pi(H_{i-1}) = \emptyset \]

which reflects the fact that the set of active paths at time \( i \) is the union of the sets surviving from the previous training epoch and of the set of newly born paths. Further, no path becoming inactive at the end of the previous epoch can re-enter the set as a newly born path. Notice first that we have

\[
f(S_i \mid H_{i-1}) = f(\pi(S_i), \pi'(S_i) \mid H_{i-1}) = f(\pi(S_i) \mid \pi(H_{i-1})) \times f(\pi'(S_i) \mid \pi(H_{i-1}), \pi'(H_{i-1}), \pi(S_i)).
\]

Assuming that the paths may disappear or survive independent of each other, and denoting \( P_{\text{death}} \) the probability of disappearing, we find:

\[
f(\pi(S_i) \mid \pi(H_{i-1})) = \begin{cases} P_{\text{death}}^{\pi(H_{i-1})-\pi(S_i)}(1 - P_{\text{death}})^{\pi(S_i)}, & \text{if } \pi(S_i) \subseteq \pi(H_{i-1}) \\ 0, & \text{otherwise.} \end{cases}
\]

Likewise, if we assume a Markov model for the evolution of the channel tap gains, forcing statistical independence between the gains of different paths, we obtain

\[
f(\pi'(S_i) \mid \pi(H_{i-1}), \pi'(H_{i-1}), \pi(S_i)) = \prod_{\ell \in \pi(S_i)} f(a_{\ell}^{(i)} \mid a_{\ell-1}^{(i)}, b_{\ell-1}^{(i)}) \times f(b_{\ell}^{(i)} \mid b_{\ell-1}^{(i)})
\]

where the last factorization is due to the (usual) assumption of circular symmetry of the complex process modeling the path gains. These steps can be reproduced to derive the conditional density of the set \( N_i \) of
the newly born paths, the only difference being that the conditional densities of the gains of the active paths should now be replaced by their unconditional counterparts, since no “past history” can be incorporated. Denoting by $P_{\text{birth}}$ the a priori probability that a path is born, and assuming again independence in the path birth process, we obtain

$$f(\pi(N_i) \mid \pi(H_{i-1})) = \begin{cases} 
P_{\text{birth}}|N_i|(1 - P_{\text{birth}})^{|H_{i-1}|-|N_i|}, & \text{if } \pi(N_i) \cap \pi(H_{i-1}) = \emptyset \\
0, & \text{otherwise}
\end{cases}$$

and

$$f(N_i \mid H_{i-1}) = \begin{cases} 
f(\pi(N_i) \mid \pi(H_{i-1})) \\
\prod_{\ell \in \pi(N_i)} f(a^{(\ell)}_i) f(b^{(\ell)}_i), & \text{if } \pi(N_i) \cap \pi(H_{i-1}) = \emptyset \\
0, & \text{otherwise.}
\end{cases}$$

Notice that the two random sets $S_i$ and $N_i$ are conditionally independent given $H_{i-1}$, whereby we have that $H_i$ is Markovian, and that its conditional density is derived by noticing that

$$f(H_i \mid H_{i-1}) = f(\pi(H_i) \mid \pi(H_{i-1})) f(\pi'(H_i) \mid \pi(H_{i-1}), \pi'(H_{i-1})).$$

Since $\pi(H_i) = \pi(S_i) \cup \pi(N_i)$ and $\pi'(H_i) = \pi'(S_i) \cup \pi'(N_i)$, and due to the conditional independence previously stated, generalized convolution can be applied to evaluate the transition densities of the random set $H_i$, resulting in

$$f(\pi(H_i) \mid \pi(H_{i-1})) = P_{\text{death}}^{|\pi(H_i)|-|\pi(H_{i-1})\cap\pi(H_i)|} \\
\times (1 - P_{\text{death}})^{|\pi(H_{i-1})\cap\pi(H_i)|} \\
\times P_{\text{birth}}^{|\pi(H_i)\setminus\pi(H_{i-1})\cap\pi(H_i)|} \\
\times (1 - P_{\text{birth}})^{|\pi(H_{i-1})\setminus|\pi(H_i)\cap\pi(H_i)|}$$

$$f(\pi'(H_i) \mid H_{i-1}) = \prod_{\ell \in \pi(H_{i-1})\cap\pi(H_i)} f(a^{(\ell)}_i \mid a^{(\ell)}_{i-1}) f(b^{(\ell)}_i \mid b^{(\ell)}_{i-1}) \\
\times \prod_{\ell \in \pi(H_i)\setminus\pi(H_{i-1})\cap\pi(H_i)} f(a^{(\ell)}_i) f(b^{(\ell)}_i) . \quad (5.6a)$$
These expressions are far less intricate than it appears at first sight. Indeed, the death-and-birth mechanism predicts that, given $H_{i-1}$, the identities of the survivors at epoch $i$ must necessarily be a subset of those in $H_i$, and thus are simply given by the intersection $\pi(H_{i-1}) \cap \pi(H_i)$. Likewise, the newly born paths are those present in $\pi(H_i)$, but not in $H_{i-1}$, and are thus the elements of the random set $\pi(H_i) \setminus \pi(H_{i-1}) \cap \pi(H_i)$. Once the paths are correctly identified as either “survivors” or “newly born,” it should be obvious that the continuous part follows either a conditional or an unconditional law as reflected by (5.6b).

Once the transition density is determined, the Bayesian recursions can be easily implemented exactly as in the case of multiuser detection. Also note that the separability condition (5.5) allows the application of the set estimator $\Pi$ of Section 2.2.1 to the prediction of $\pi(H_i)$, i.e., of the identities of the active paths — or, equivalently, the support of the vector $h_i$ in (5.2) — and of the channel gains $\pi'(H_i)$ of the active paths (corresponding to the nonzero entries of the vector $h_i$).

Before proceeding to the extension to the case of MIMO channels, it may be worthwhile to observe that the assumption of delays integer multiples of $1/(rW)$ can be easily relaxed. In particular, the case of arbitrary delays amounts to building the random set modeling the channel state on $\mathbb{S} = \{1, \ldots, L_{\text{max}}\} \times \mathbb{R}^3$, and assigning to the delays $\tau_{\ell,i}$ a prior and a conditional density. This can be done in principle with no limitation, although a reasonable choice is the one considered in [6], where the delay evolution over different training phases has been modeled as a Markovian process with reflective boundaries.

### 5.4 Extension to MIMO Channels

The extension of this concept to more complex scenarios, such as MIMO channels, is straightforward. Indeed, dispersive MIMO channels with $N$ transmit and $M$ receiver antennas can be dealt with quite similarly to SISO channels. Indeed, denoting by $H_{m,n,i}$ the state of the channel linking the $m$th transmit to the $n$th receive antenna at the $i$th training epoch, the model (5.4) can be directly applied to the signal observed.
at the $m$th receive antenna:

$$y_{m;i} = \sum_{n=1}^{N} s_{n} (H_{m,n;i}) + w_{m;i}$$

$$= s(H_{m;i}) + w_{m;i}$$

where $s_{n} (H_{m,n;i})$ represents the pilot signal transmitted from the $n$th transmit antenna, and received by the $m$th receiver antenna after crossing a channel of which the state is encapsulated in $H_{m,n;i}$, with

$$H_{m;i} = \bigcup_{n=1}^{N} H_{m,n;i}$$

representing the state of the sets of the $N$ channels seen by the $m$th receiver antenna. In particular, letting

$$H_{m,n;i}^{(\ell)} = \bigcup_{\ell=1}^{L_{\text{max}}} H_{m,n;i}(\ell)$$  \hspace{1cm} (5.7)$$

with $H_{m,n;i}^{(\ell)}$ the singleton-or-empty random set defined as

$$H_{m,n;i}^{(\ell)} = \begin{cases} 
\left\{ \begin{pmatrix} m & n & \ell & \Re \{ h_{i,\ell,m,n} \} & \Im \{ h_{i,\ell,m,n} \} \end{pmatrix}^T \right\}, & \text{if path } \ell \text{ is active at epoch } i \text{ on channel } (m,n) \\
\emptyset, & \text{otherwise} 
\end{cases}$$

an evolution model similar to that considered for a SISO channel can be defined for the random sets $H_{m,n;i}$. 

### 5.4.1 A Case Study: MIMO–OFDM Systems

Let us consider a MIMO–OFDM system with $N$ transmit and $M$ receive antennas, and whose signaling scheme is outlined in Figure 5.1. Antenna $n$ at time $i$ transmits a stream of $K$ symbols, consisting of $K - P$ information-bearing symbols and $P$ known symbols producing the pilot signal. The symbol stream undergoes a $K$-point inverse discrete Fourier transform and the insertion of a cyclic prefix of
length larger than or equal to $L - 1$, the maximum length of all of the spatial channels. Of course, we have $P \geq LN$, and we denote by $(d_{k_1,n}; \cdots; d_{k_P,n})^T$ the $P$-dimensional vector containing the training symbols transmitted by antenna $n$.

With these notations, the signal received at antenna $m$ can be written in the form

\[
\text{Pilot subcarriers: } y_{m,t} = \sum_{n=1}^{N} D_{n,t} F h_{m,n,t} + z_{m,t} \ \ (5.8a)
\]

\[
\text{Information subcarriers: } y_{m,t}^I = \sum_{n=1}^{N} D_{n,t}^I F I h_{m,n,t} + z_{m,t}^I \ \ (5.8b)
\]

where

\[
D_{n,t} \triangleq \text{diag}(d_{k_1,n}; t, \ldots, d_{k_P,n}; t)
\]

\[
D_{n,t}^I \triangleq \text{diag}(d_{k_1,n}; t, \ldots, d_{k_K-n}; t)
\]

\[
z_{m,t} \triangleq (z_{k_1,m}; t \cdots z_{k_P,m}; t)^T
\]

\[
z_{m,t}^I \triangleq (z_{k_1,m}; t \cdots z_{k_K-n,m}; t)^T
\]
while \( \mathbf{F} \) and \( \mathbf{F}^I \) are submatrices, of order \( P \times L_{\max} \) and \((K - P) \times L_{\max}\), respectively, of the matrix \( \mathbf{T} \), whose entry \((k, \ell)\) is

\[
T_{k,\ell} = e^{-j2\pi(\ell-1)\frac{k-1}{K}}
\]

\( k = 1, \ldots, K, \ell = 1, \ldots, L_{\max} \). For the rest,

\[
\mathbf{h}_{m,n;i} = \left( a_{m,n;i}^{(1)} \cdots a_{m,n;i}^{(L_{\max})} \right)^T
\]

is an \( L_{\max} \)-dimensional column vector representing the channel impulse response from the \( n \)th transmit to the \( m \)th receive antenna, while \( \tilde{\mathbf{z}}_{m;i} \), represents additive noise: \( \tilde{\mathbf{z}}_{m;i} \sim \mathcal{N}_{c}(\mathbf{0}_{K \times 1}, \mathbf{N}_0 \mathbf{I}_K) \).

As anticipated in the previous section, the state of the channel linking transmit antenna \( n \) to receive antenna \( m \) at time \( i \) is captured by the random set (5.7), which fully defines channel vector \( \mathbf{h}_{m,n;i} \) through the set function

\[
[\mathbf{h}_{m,n;i}](\mathbf{H}_{m,n;i})_\ell = \begin{cases} 
  a_{m,n;i}^{(\ell)} + jb_{m,n;i}^{(\ell)}, & \text{if } (m,n,\ell) \in \pi(\mathbf{H}_{m,n;i}) \\
  0, & \text{otherwise}
\end{cases}
\]

It may be useful to observe here that the format of the pilot signals in (5.8a), while generalizing the SISO model (5.1) to the MIMO case, is fully consistent with the format assumed therein, after noticing that now the training and transmission epochs coincide, i.e., the interleaving between pilots and information-bearing signals takes place in the frequency domain rather than in the time domain.

With these notations, and extending the concepts outlined in Section 5.3, all the equations derived therein with reference to a SISO channel carry over to model each of the \( NM \) channels defined by the MIMO architecture. In particular, one can use the set estimator III of Section 2.2.1, which predicts with minimum error probabilities the identities of the paths present in each spatial channel. Subsequently, a minimum mean-square error estimator can be used, aimed at estimating the values of the tap gains corresponding to active channels [6].
The basic motivation that encouraged the authors to write this monograph is twofold. On the one hand, to popularize in the Communications and Information Theory community a tool that has attracted much interest from experts active in neighboring disciplines, primarily multitarget tracking. On the other hand, the authors wanted to answer a basic question concerning whether RFS theory — and RST in general — allows one to model and solve problems that conventional probability theory does not allow to handle, at least in a simple and engineer-friendly way. This question is definitely challenging, and hence requires a detailed response. Indeed, from the beginning of this monograph the authors have made it clear that the context in which RFSs have been developed is one in which the object of interest is an unordered collection of an arbitrary number of points drawn from an arbitrary space: this is a multiobject consisting of individual subsystems, each possibly describing a trajectory in its own state-space, which we have denoted by $S$. Even in the simplified scenario — considered here — where the observations are
random vectors, there are two basic conditions for the problem to be manageable in the context of ordinary probability. These are:

(1) The number of elementary objects composing the multiobject admits a known finite upper $K$.

(2) The elementary objects should be distinguishable: this means that, given the multiobject states in two different epochs, we can predict the initial and the final state of each elementary object, i.e., the single objects are individually trackable.

If conditions (1) and (2) above are met, the multiobject state is a realization of an ordered collection of $K$ points of $S$, and $S^K$ is the sample space of a conventional probability space. In all other cases, the state space is a collection of sets of $S$ of random cardinality, and the object of interest is inherently an RFS on $S$. We reiterate here that RFS-based models prove advantageous in at least two points, i.e.:

(1) The formalism of RFS is very close to that of conventional random vectors.

(2) Bayesian recursions for Markovian RFSs can be solved through bounded-complexity approximations, especially Probability Hypothesis Density (PHD) filtering and its Cardinalized form (CPHD).

For these reasons, although the basic question about RFS indispensability cannot be answered with a clear Yes/No, the authors are confident that this tool will prove to be useful in more and more applications.

### 6.1 Topics Not Covered Here

A number of topics, related to theory, extensions, and applications of random sets, could not find their way into this monograph due to space limitations. In the following, we provide a brief list of applications and their basic bibliographic references.

---

1 In a general multisensor multitarget scenario the observations themselves are an unordered collection of a random number of points.
Concluding Remarks

Stochastic geometry and modeling problems. It has been observed [46] that random shapes could be modeled as random sets, so that images can be conceived as realizations of a two-dimensional random set model. More generally, random-set models can be used to analyze objects in the $d$-dimensional Euclidean space [32]. A specific application, considered in [32], is the geometrical characterization of the growth of a tumor, characterized as a geometric object.

Radar applications. The use of random-set theory for radar analyses (the context in which Finite-Set Statistics was originally developed by Ronald Mahler) gives rise to a variety of applications, described extensively in [66] and summarized in [60], to which the interested reader is addressed for details. Among these applications, we may mention multi-sensor/multitarget algorithms, robust target identification, simultaneous optimal estimation of the numbers, identities, and geokinematics of targets, a systematic approach for detection, tracking, and ID of multiple group targets, and robust automatic target recognition using Synthetic Aperture Radar data.

Modeling and simulation of cellular systems. A cellular system can often be regarded as a complex dynamical system ruled by birth–death-migration mechanisms concerning the number and the composition of the subsystems forming it. For example, in [27] a biological system describing the evolution of cancer stem cells into tumors is modeled and simulated. A stem cell may either die or reproduce, giving rise to a pair of stem cells or to a stem cell and a first-generation daughter cell. Daughter cells have in turn a much more limited lifetime and may either die or generate two second-generation daughter cells, with the understanding that the number of generations is limited and known. Thus, a set of stem cells in an in vitro laboratory gives rise to agglomerates of cells randomly evolving in the experiment domain, and having stem cells in their cores and daughter cells of different generations in the outer layers. The lab as a whole appears as a set of agglomerates of random cardinality, and each agglomerate has a random number of cells of different nature (i.e., stem cells and daughters cells of various generations). In addition, the agglomerates may describe, over time,
trajectories which cannot be individually tracked. This is a typical scenario where Markovian RFSs appear a key tool for modeling the system evolution to the end of replacing costly \textit{in vitro} experiments with cheap and reproducible \textit{in silico} emulators. This paves the way to a larger scale application of Random-Set Theory in the field of biological system modeling and simulation.

\textbf{An open problem: Neighbor discovery in sensor networks.} Neighbor discovery (ND) in a sensor network consists of the determination of all sensors that may directly communicate to a reference one. An approach to ND, advocated in [25], assumes that sensors transmit, at random instants and in a finite period, copies of the signature waveform associated with their identities. One can write the signal observed by the reference sensor at discrete time $n$ in the form

\begin{equation}
y_n = \sum_{k \in K} \alpha_k s_{k,n} + z_n \tag{6.1}\end{equation}

where $\alpha_k$ denotes the gain of the path from sensor $k$ to the reference sensor (assumed to be constant during all the discovery session), $s_{k,n}$ is the signature of sensor $k$ at time $n$, $z_n$ is noise, and $K$ is the random set of nearest neighbors. A sensor is declared to be a neighbor of the reference sensor if its amplitude, as received by the latter, exceeds a pre-assigned activity threshold. Hence, the ND problem is now reduced to determining the set $K$ of which the elements exceed an activity threshold. The determination of the nearest-neighbor set from the observation of $y_n$, $N = 1, \ldots, N$, is a problem that was solved in [8] using techniques borrowed for multiuser detection, but the formulation of the problem itself can lead to a “natural” solution based on random-set theory. The difficulty here stems from the complexity of the implementation of an RST neighbor discoverer. The search for a suboptimum solution, possibly based on PHD filtering, is an open problem.
Acknowledgments

The work of Ezio Biglieri was supported by the Project CONSOLIDER-INGENIO 2010 CSD2008-00010 “COMONSENS”. The authors wish to thank the Editor Sergio Verdú for his careful handling of this manuscript, and the reviewers for their insightful comments.
In this appendix we recall the basic mathematical concepts needed to understand the fundamentals of Random Set Theory.

A.1 Topologies and all that

Given a nonempty set $S$, a *topology* on $S$ is a collection $T$ of its subsets which satisfies the following conditions:

1. The empty set and $S$ belong to $T$;
2. $T$ is closed under arbitrary unions and finite intersections.

The pair $\Theta = (S,T)$ is called a *topological space*, the sets in $T$ are called *open sets*, and their complements *closed sets*. A *basis* of a topological space is a collection of sets in $T$ such that each element of $T$ can be represented as union of basis sets. A sub-base of a topology is a family of sets such that their finite intersections form the basis of the topology. An assumption that will be made from now on is that $(S,T)$ is a Hausdorff (H) space, i.e., that every two disjoint points of $S$ have disjoint open neighborhoods. Also, we assume that the topological space
Mathematical Aspects of Random-Set Theory

is Locally Compact (LC), i.e., that every point of $S$ has a neighborhood with a compact closure, and Second countable (S), i.e., that the topology on $T$ has a countable base. These four properties are summarized by the acronym LCHS. The Euclidean space $\mathbb{R}^n$ is an example of such space.

The discrete topology on $S$ is its power set $2^S$, so that every subset of $S$ is open (and hence also closed), while the trivial topology (also called indiscrete topology) is $T = \{S, \emptyset\}$: these represent two extreme cases of “extremely rich” and “overly poor” topologies, and every other topology lying in between. If $S$ is discrete, however, the discrete topology is a fairly natural choice. The real line $\mathbb{R}$ can be endowed with a “standard topology,” induced by the metric given by the absolute difference: the open sets are $\mathbb{R}$, $\emptyset$, and the open intervals on $\mathbb{R}$, which define the Euclidean topology. Likewise, the Euclidean topology on $\mathbb{R}^n$ is based on open balls. What is interesting is that these open sets can in turn be seen as set products of $n$ open intervals, namely the open sets in $\mathbb{R}$ are products of as many open sets on $\mathbb{R}$: this naturally leads to the concept of product topology on two topological spaces $(S_1, T_1)$, $(S_2, T_2)$, that is to the definition of the topological space $(S_1 \times S_2, T_1 \times T_2)$, which admits as a basis the set product between the respective bases.

There are occasions when we have to deal with topological spaces whose elements are subsets of another topological space, i.e., with hyper-spaces. In particular, considering the topological space $(S, T)$, we focus our attention on the collection $\mathcal{F}$ of all closed subsets of $S$. $\mathcal{F}$ is one of the standard objects in set theory and general topology. There are a number of ways to endow the family $\mathcal{F}$ with the structure of a topological space. In the following we describe one of the most common topologies on $\mathcal{F}$, the Fell topology (also known as hit-and-miss topology). Denote $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{K}$ the collection of closed, open, and compact subsets of $S$, respectively. Given an arbitrary compact set $K \in \mathcal{K}$, we define the collections of closed subsets of $S$ “missing” $K$ and “hitting” $K$, $\mathcal{F}^K$ and $\mathcal{F}_K$, respectively, as:

$\mathcal{F}^K = \{F \in \mathcal{F} : F \cap K = \emptyset\}$

$\mathcal{F}_K = \{F \in \mathcal{F} : F \cap K \neq \emptyset\}$
The Fell topology has a sub-base which consists of \( F^G \) for all \( G \in \mathcal{G} \) and \( F^K \) for all \( K \in \mathcal{K} \) [43, 68, 72], i.e.:

\[
\{ F^K \cap F_{G_1} \cap \cdots \cap F_{G_n} : n \geq 0, K \in \mathcal{K}, G_k \in \mathcal{G}, G_i \cap G_j = \emptyset, \forall i \neq j \}.
\]

An important result, is that if \( \mathcal{S} \) is LCHS, then \( \mathcal{F} \) endowed with the Fell topology is also LCHS [68, 72].

Before proceeding, we recall here that a family of sets is a \( \sigma \)-algebra if it contains the empty set and is closed under complements and countable unions. If \( \mathcal{M} \) is any family of sets, then \( \sigma(\mathcal{M}) \) denotes the smallest \( \sigma \)-algebra generated by \( \mathcal{M} \). In a topological space \( (\mathcal{S}, \mathcal{F}) \), an important case of this construction is the Borel \( \sigma \)-algebra, denoted with \( \mathcal{B}(\mathcal{S}) \), which is the \( \sigma \)-algebra generated by the open sets.

### A.2 Random Closed Sets

Let \( \mathcal{S} \) be an LCHS topological space, and \( \mathcal{F} \) the family of closed subsets of \( \mathcal{S} \). Endow \( \mathcal{F} \) with the Fell topology, and consider the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathcal{F}) \) on \( \mathcal{F} \). Let \( (\Omega, \Sigma, \mathbb{P}) \) be a standard probability space, which will be used to define random elements. Intuitively, a Random Closed Set (RACS) is just an \( \mathcal{F}(\mathcal{S}) \)-valued random element. To be more specific,

**Definition A.1.** A RACS is a measurable mapping \( X \) from \( (\Omega, \Sigma) \) to \( (\mathcal{F}, \mathcal{B}(\mathcal{F})) \).

To gain a deeper insight into this definition, observe that, as shown in [72], \( \mathcal{B}(\mathcal{F}) \) coincides with the \( \sigma \)-algebra generated by \( \{ F^K \}_{K \in \mathcal{K}} \). Therefore, Definition A.1 implies that, for every compact set \( K \) in \( \mathcal{S} \),

\[
\{ \omega \in \Omega : X(\omega) \cap K \neq \emptyset \} \in \Sigma
\]

which simply means that, observing \( X \), one can always tell if \( X \) hits or misses any given compact set \( K \).

The distribution of an RACS \( X \) can be determined by \( P_X(\mathcal{X}) = \mathbb{P}(\{ X \in \mathcal{X} \}), \forall \mathcal{X} \in \mathcal{B}(\mathcal{F}). \) The particular choice \( \mathcal{X} = F^K \) is useful, since the family \( \{ F^K \}_{K \in \mathcal{K}} \) generate the \( \sigma \)-algebra \( \mathcal{B}(\mathcal{F}) \).
Definition A.2. The functional $T_X : \mathcal{K} \to [0,1]$ defined as
\[
T_X(K) \triangleq \mathbb{P} \left( \{ X \in \mathcal{F}_K \} \right) = \mathbb{P} \left( \{ X \cap K \neq \emptyset \} \right), \quad \forall K \in \mathcal{K}
\]
is called the capacity functional of the RACS $X$.

The Choquet theorem (see, e.g., [68, 72]) ensures that a capacity functional on $\mathcal{K}$ uniquely determines the distribution of a RACS. The capacity functional can also be extended to open sets as follows:
\[
T_X^*(G) = \sup \{ T(K) : K \subset G, K \in \mathcal{K} \}, \quad \forall G \in \mathcal{G}.
\]

Theorem A.3 ([72, Th. 1.12]). $T_X^*(K) = T_X(K)$, for all $K \in \mathcal{K}$, and $T_X^*(B) = \sup \{ T(K) : K \subset B, K \in \mathcal{K} \}$, for all Borel sets $B$.

In the following, we use the same notation $T_X$ to denote the extension $T_X^*$. Theorem A.3 and the continuity property of probability measures imply that $T_X(B) = \mathbb{P}(\{ X \cap B \neq \emptyset \})$, for all Borel $B$.

Definition A.4. A functional $C_X : \mathcal{F} \to [0,1]$ defined by
\[
C_X(F) \triangleq \mathbb{P} \left( \{ X \subseteq F \} \right), \quad \forall F \in \mathcal{F}
\]
is called the containment functional of the RACS $X$.

From this definition, it immediately follows that $C_X(F) = 1 - T_X(\overline{F})$, $\forall F \in \mathcal{F}$. This functional can be extended onto the family of open sets and all sets in the same way as it has been done for the capacity functional, and, in the following, $C_X$ will denote this extension. A simple reformulation of the Choquet theorem shows that $C_X$ determines uniquely the distribution of $X$.

Definition A.5. RACS $X_1, \ldots, X_n$ are said to be independent if
\[
\mathbb{P}(\{ X_1 \in \mathcal{X}_1, \ldots, X_n \in \mathcal{X}_n \}) = \prod_{i=1}^{n} \mathbb{P}(\{ X_i \in \mathcal{X}_i \})
\]
The Choquet theorem can be used to characterize independent RACS, i.e.,

**Proposition A.6 ([72, Prop. 1.19]).** RACS $X_1, \ldots, X_n$ are independent if and only if

$$
P(\{X_1 \cap K_1 \neq \emptyset, \ldots, X_n \cap K_1 \neq \emptyset\}) = \prod_{i=1}^{n} T_{X_i}(K_i)
$$

for all $K_1, \ldots, K_n \in \mathcal{K}$.

Conditional distributions of RACS can be derived in the same way as conditional distributions of random elements in an abstract measurable space.

### A.3 Random Finite Sets

In this section we restrict our attention to random finite sets (RFS), defined as follows:

**Definition A.7.** An RFS is an RACS $X$ such that $|X| < \infty$, $\mathbb{P}$-a.e.

In RFSs, it is customary to refer to the containment functional of Definition A.4 as the belief function (or, less precisely, belief measure) $\beta_X$, i.e.,

$$
\beta_X(A) = C_X(A) = \mathbb{P}(\{X \subseteq A\}), \quad \forall A \in \mathcal{B}(\mathcal{S})
$$

where $\mathcal{B}(\mathcal{S})$ is the Borel $\sigma$-algebra of subsets of $\mathcal{S}$.

**Example A.8.** Let $X_1$ and $X_2$ be two independent RFSs, then $X_1 \cup X_2$ is an RFS, and

$$
\beta_{X_1 \cup X_2}(C) = \mathbb{P}(\{X_1 \cup X_2 \subseteq C\})
= \mathbb{P}(\{(X_1 \cup X_2) \cap \bar{C} = \emptyset\})
= \mathbb{P}(\{X_1 \cap \bar{C} = \emptyset, X_2 \cap \bar{C} = \emptyset\})
= (1 - T_{X_1}(\bar{C}))(1 - T_{X_2}(\bar{C}))
= \beta_{X_1}(C)\beta_{X_2}(C)
$$
for any $C \in \mathcal{B}(S)$. This relationship yields an immediate proof of (2.8) concerning the union of independent random sets, and can be generalized as

$$\beta_{\cup, X_i}(C) = \prod \beta_{X_i}(C), \quad C \subseteq S$$

The belief function admits a simple expression, as it can be seen from the following.

**Theorem A.9 ([44, Ch. 4, Th. 17]).** Let $X$ be an RFS with belief function $\beta_X$, then

$$\beta_X(C) = p_0 + \sum_{k=1}^{\infty} p_k Q_k(C^k), \quad \forall C \in \mathcal{B}(S) \quad (A.1)$$

where $p_k$ is a probability measure on $\mathbb{N}$ and $\{Q_k\}_{k=1}^{\infty}$ are probability measures on $S^k$. In addition

$$p_k = \mathbb{P}(\{|X| = k\})$$

$$Q_k(C^k) = \mathbb{P}(\{X \subseteq C \mid \{|X| = k\})$$

Therefore, we can always write

$$\beta_X(C) = \mathbb{P}(\{X \subseteq A\})$$

$$= \mathbb{P}(\{|X| = 0\})$$

$$+ \sum_{k=1}^{\infty} \mathbb{P}(\{|X| = k\}) \mathbb{P}(\{X \subseteq C \mid \{|X| = k\})$$

Notice that the belief function is not a measure since it is not additive, as it can be seen from (A.1).

**A.3.1 Calculus of set functions**

From now on, we restrict our attention to the case $S = W \times \mathbb{R}^d$, where $W$ is a countable set, but most results hold for a general LCHS space. $S$ is endowed with the product topology of the Euclidean topology.
in $\mathbb{R}^d$ and the discrete topology in $W$. Let $\lambda$ be a measure on $(S, \mathcal{B}(S))$, e.g., the product measure between the Lebesgue measure on $\mathbb{R}^d$ and the counting measure on $W$. We want to introduce a measure on $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$.

To this end, it will be shown that the topological spaces $S^k$ and $\mathcal{F}_{=k}$, the family of closed subsets of $S$ with exactly $k$ elements, are homeomorphic, and then a measure on $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ will be constructed using the measures $\lambda^k$ on $(S^k, \mathcal{B}(S^k))$. Let $\mathcal{F}_{\geq k}, \mathcal{F}_{\leq k},$ and $\mathcal{F}_{=k}$ be the families of closed subsets of $S$ with at least, at most, and exactly $k$ elements, respectively. The following proposition states that these collections are measurable.

**Proposition A.10** ([44, Ch. 4, Prop. 1]). $\mathcal{F}_{\leq k}$ is a closed subset of the Fell topology on $\mathcal{F}$. Therefore, $\mathcal{F}_{=k}$ is Borel, and hence measurable.

Let $\gamma_k : S^k \to \mathcal{F}_{=k}$ be defined as

$$\gamma_k(x_1, \ldots, x_k) = \{x_1, \ldots, x_k\}$$

$\forall x_1, \ldots, x_k \in S$, and let $\hat{\gamma}_k$ be the restriction of $\gamma_k$ to $[S]^k$, the subset of $S^k$ lexicographically ordered, for all $k \in \mathbb{N}$. Then

**Proposition A.11** ([44, Ch. 4, Prop. 2]). $\hat{\gamma}_k$ is a homeomorphism between $[S]^k$ and $\mathcal{F}_{=k}$.

The key differences between RFSs and random vectors is that the former have random cardinality and no ordering. However, the previous proposition allows to interpret finite sets as points in a suitable space, namely $[S]^k$. To give an idea, the subsets $\{x_1, x_2\}$ of the interval $[0, 1]$ can be represented as points $(x_1, x_2)$ of the square $[0, 1]^2$ with $x_1 < x_2$ (i.e., half of the square).

Given this correspondence, we can construct a measure on $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ using the measures $\lambda^k$ on $(S^k, \mathcal{B}(S^k))$. Specifically, we can define a set measure as

$$\mu(Z) = \delta_{\varnothing}(Z) + \sum_{k=1}^{\infty} \frac{1}{k!} \lambda^k(\gamma_k^{-1}(Z \cap \mathcal{F}_{=k})) \quad (A.2)$$

for all $Z \in \mathcal{B}(\mathcal{F})$. Observe that $\delta_{\varnothing}(Z) = 1$ if $\varnothing \in Z$ and not if $\varnothing \subseteq Z$, i.e., $\delta_{\varnothing}(\varnothing) = 0$ and $\delta_{\varnothing}(\{\varnothing\}) = 1$. 


Remark A.12. The rationale underlying the definition of measure (A.2) can be easily understood by considering that any $Z \in \mathcal{B}(\mathcal{F})$ can be represented as the disjoint union

$$Z = \bigcup_{k=0}^{\infty} (Z \cap \mathcal{F} = k)$$

where $\mathcal{F}_{=0} = \{\emptyset\}$. Therefore,

$$\mu(Z) = \sum_{k=0}^{\infty} \mu(Z \cap \mathcal{F} = k).$$

We would like to associate a nonzero measure to $\{\emptyset\}$, as in RFS $X$ there can be a nonzero probability that $X = \{\emptyset\}$, and hence

$$\mu(Z \cap \{\emptyset\}) = \delta_{\emptyset}(Z).$$

Since there is a bijection $\hat{\gamma}_k$ between $Z \cap \mathcal{F} = k$ and $\mathcal{F}^k$, we would like $\mu$ to preserve the correspondence between measures, i.e.,

$$\mu(Z \cap \mathcal{F} = k) = \lambda^k \left( \hat{\gamma}_k^{-1}(Z \cap \mathcal{F} = k) \right)$$

$$= \frac{1}{k!} \lambda^k \left( \gamma_k^{-1}(Z \cap \mathcal{F} = k) \right).$$

The reason for the factorial term is that the choice of $\hat{\gamma}$ is not unique, and that there are $k!$ equivalent restrictions of $\gamma$, i.e.,

$$\lambda^k \left( \gamma_k^{-1}(Z \cap \mathcal{F} = k) \right) = \lambda^k \left( \bigcup_{i=1}^{k!} \left\{ \hat{\gamma}_{k,i}^{-1}(Z \cap \mathcal{F} = k) \right\} \right)$$

$$= \sum_{i=1}^{k!} \lambda^k \left( \hat{\gamma}_{k,i}^{-1}(Z \cap \mathcal{F} = k) \right)$$

$$= k! \lambda^k \left( \hat{\gamma}_k^{-1}(Z \cap \mathcal{F} = k) \right).$$

To give an example, the subsets $\{x_1, x_2\}$ of the interval $[0,1]$ can be represented as points $(x_1, x_2)$ of $[0,1]^2$ with $x_1 < x_2$ (upper triangle) or, similarly, with $x_1 > x_2$ (lower triangle), and its measure is half the measure of the square.
A.3 Random Finite Sets

A.3.1.1 Set Integral

A result of Proposition A.11 is that finite-set functions defined on $\mathcal{F}_k$ can be used interchangeably with symmetric functions on $\mathcal{S}^k$ [44, Ch. 4, Prop. 3]. A set function is a function $\varphi(S)$ of which the arguments are sets, and a finite-set function is one of which the arguments are finite sets. We can now define a set integral.

**Definition A.13.** Let $f : \mathcal{F} \to \mathbb{R}$ be a finite set function. The set integral of $f$ is defined as an integral with respect to the measure $\mu$ in (A.2), i.e.,

$$\int_S f(Z) \delta Z = \int_S f(Z) d\mu(Z) = f(\emptyset) \delta(\emptyset)(S) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\gamma^{-1}(S \cap \mathcal{F}_k)} f(\{x_1, \ldots, x_k\}) d\lambda(x_1) \cdots d\lambda(x_k)$$

for all $S \in \mathcal{B}(\mathcal{F})$.

Notice that, if $A \in \mathcal{F}$, denoting $\mathcal{C}_A$ the collection of all of the closed subsets of $A$ (observe that $\mathcal{C}_S = \mathcal{F}$), then $\gamma^{-1}(\mathcal{C}_A \cap \mathcal{F}_k) = A_k$ and

$$\int_{\mathcal{C}_A} f(Z) \delta Z = f(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{A_k} f(\{x_1, \ldots, x_k\}) d\lambda(x_1) \cdots d\lambda(x_k).$$

A.3.1.2 Set Density and Set Derivative

The concepts of set density and set derivatives are explained next.

**Definition A.14.** A finite set function $f$ is a set density if it is non-negative, integrable, and $\int_{\mathcal{F}} f(Z) \delta Z = 1$.

Let $B_{x,r}$ be the open ball centered at $x$ with radius $r$, and let $\bar{B}_{x,r}$ be its closure.
\textbf{Definition A.15 ([44, Ch. 4, Def. 13])}. Let $\varphi : \mathcal{F} \to \mathbb{R}$ be a set function and $s \in \mathcal{S}$. If it exists, the derivative of $\varphi$ at $s$ is defined by

$$
\frac{\delta \varphi}{\delta s}(Z) = \lim_{j \to \infty} \lim_{i \to \infty} \frac{\varphi((Z \setminus B_{s,1/j}) \cup \tilde{B}_{s,1/i}) - \varphi(Z \setminus B_{s,1/j})}{\lambda(\tilde{B}_{s,1/i})}. \quad (A.3)
$$

The reason for such a cumbersome definition is that, whenever $s \in Z$, we have that $\varphi(Z \cup \tilde{B}_{s,1/i})$ eventually (i.e., for $i$ large enough) coincides with $\varphi(Z)$, for which the more intuitive definition

$$
\frac{\delta \varphi}{\delta s}(Z) = \lim_{i \to \infty} \frac{\varphi(Z \cup \tilde{B}_{s,1/i}) - \varphi(Z)}{\lambda(\tilde{B}_{s,1/i})} \quad (A.4)
$$

only applies to the case that $s \notin Z$, and in fact under these circumstances (A.3) coincides with (A.4).

Since the set derivative of a set function is itself a set function, it can be derived, as shown in the definition that follows:

\textbf{Definition A.16}. Let $\varphi : \mathcal{F} \to \mathbb{R}$ be a set function and $s_1, \ldots, s_k \in \mathcal{S}$. If it exists, the derivative of order $k$ of $\Phi$ is defined iteratively by

$$
\frac{\delta^k \varphi}{\delta s_k \cdots \delta s_1}(Z) = \frac{\delta}{\delta s_k} \frac{\delta^{k-1} \varphi}{\delta s_{k-1} \cdots s_1}(Z).
$$

Since the iterated derivative in Definition A.16 does not depend on the order of differentiation [44], we can define a set density as follows.

\textbf{Definition A.17}. Let $\varphi : \mathcal{F} \to \mathbb{R}$ be a set function and let $C = \{s_1, \ldots, s_k\} \subseteq \mathcal{S}$ be a finite subset with $k$ distinct elements. If all the derivatives of order $k$ of $\varphi$ exist, then the set derivative of $\varphi$ at $Z$ is defined by

$$
\frac{\delta \varphi}{\delta C}(Z) = \begin{cases} 
\frac{\delta^k \varphi}{\delta s_1 \cdots \delta s_k}(Z), & \text{if } C \neq \emptyset \\
\varphi(Z), & \text{if } C = \emptyset.
\end{cases}
$$
Example A.18. Direct application of (A.3) allows one to prove all of the rules listed in (2.1). For example, consider the product of two set functions, \( \varphi_1(Z) \) and \( \varphi_2(Z) \), and compute its set derivative with respect to \( s \). We have:

\[
\frac{\delta \varphi_1(Z) \varphi_2(Z)}{\delta s} = \lim_{j \to \infty} \lim_{i \to \infty} \left\{ \frac{\varphi_1(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) \varphi_2(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i})}{\lambda(B_{s,1/i})} \right. \\
\left. - \frac{\varphi_1(Z \setminus B_{s,1/j}) \varphi_2(Z \setminus B_{s,1/j})}{\lambda(B_{s,1/i})} \right\} 
\]

(A.5)

Summing and subtracting the term

\[
\varphi_1(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) \varphi_2(Z \setminus B_{s,1/j})
\]

in the numerator we have

\[
\varphi_1(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) \varphi_2(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) \\
- \varphi_1(Z \setminus B_{s,1/j}) \varphi_2(Z \setminus B_{s,1/j}) \\
= \varphi_1(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) \left[ \varphi_2(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) \\
- \varphi_2(Z \setminus B_{s,1/j}) \right] + \varphi_2(Z \setminus B_{s,1/j}) \\
\times \left[ \varphi_1(Z \setminus B_{s,1/j} \cup \tilde{B}_{s,1/i}) - \varphi_1(Z \setminus B_{s,1/j}) \right]
\]

(A.6)

Substituting (A.6) into (A.5), and elaborating, yields

\[
\frac{\delta \varphi_1(Z) \varphi_2(Z)}{\delta s} = \frac{\delta \varphi_1(Z)}{\delta s} \varphi_2(Z) + \frac{\delta \varphi_2(Z)}{\delta s} \varphi_1(Z)
\]

(A.7)

Further derivation of (A.7) with respect to the singleton \( s' \) yields the derivative with respect to the set \( C = \{s,s'\} \) in the form

\[
\frac{\delta \varphi_1(Z) \varphi_2(Z)}{\delta C} = \frac{\delta \varphi_1(Z)}{\delta \{s,s'\}} \varphi_2(Z) + \frac{\delta \varphi_1(Z) \delta \varphi_2(Z)}{\delta s \delta s'} \\
+ \frac{\delta \varphi_2(Z)}{\delta \{s,s'\}} \varphi_1(Z) + \frac{\delta \varphi_2(Z) \delta \varphi_1(Z)}{\delta s \delta s'}
\]
\[
\delta \phi_1 (Z) \frac{\partial}{\partial C} \phi_2 (Z) + \delta \phi_2 (Z) \frac{\partial}{\partial s} \delta (C \setminus s) + \delta \phi_2 (Z) \frac{\partial}{\partial C} \phi_1 (Z) + \delta \phi_1 (Z) \frac{\partial}{\partial s'} \delta (C \setminus s')
\]
\[
= \sum_{W \subseteq C} \frac{\delta \phi_1 (Z)}{\delta W} \frac{\delta \phi_2 (Z)}{\delta (C \setminus W)}
\]

which is a special case of rule (2.7d).

**Example A.19.** As an example of application, let \((S_1, S_2)\) be a partition of \(S\), and let us consider two independent RFSs \(X_1\) on \(S_1\) and \(X_2\) on \(S_2\). We want to calculate the set density of the RFS \(Z = X_1 \cup X_2\), which is an RFS on \(S\). Since the two RFSs are independent and disjoint,

\[
\beta_{Z}(C) = \beta_{X_1}(C \cap S_1) \beta_{X_2}(C \cap S_2)
\]

so that, applying the generalized convolution rule, we obtain

\[
f_{Z}(Z) = \sum_{Y \subseteq Z} \frac{\delta \beta_{X_1}(C \cap S_1)}{\delta Y} \frac{\delta \beta_{X_2}(C \setminus (C \cap S_1))}{\delta (Z \setminus Y)} \bigg|_{C = \emptyset}.
\]

Suppose that \(Z = \{z_1, \ldots, z_{m_1}, z_{m_1+1}, \ldots, z_m\}\), with \(\{z_1, \ldots, z_{m_1}\} \in S_1\) and \(\{z_{m_1+1}, \ldots, z_m\} \in S_2\). Hence, any subset of \(Z\) can be split up in a subset of \(S_1\) and a subset of \(S_2\); moreover, \(\beta_{X_1}(C \cap S_1)\) is functionally independent of \(\{z_{m_1+1}, \ldots, z_m\}\) and of any subset thereof, and so is \(\beta_{X_2}(C \setminus (C \cap S_1))\) with respect to \(\{z_1, \ldots, z_{m_1}\}\) and its subsets. Consequently, due to property (2.7e), there is one and only one surviving term in the sum (A.8), the one corresponding to \(Y = \{z_1, \ldots, z_{m_1}\}\) and \(Z \setminus Y = \{z_{m_1+1}, \ldots, z_m\}\), which yields

\[
f_{Z}(Z) = f_{X_1}(Z \cap S_1) f_{X_2}(Z \setminus (Z \cap S_1))
\]

\[
= f_{X_1}(\{z_1, \ldots, z_{m_1}\}) f_{X_2}(\{z_{m_1+1}, \ldots, z_m\}).
\]

Set density and belief function are related to each other through operations of set derivative and set integral when the RFS is absolutely continuous.
Definition A.20. The RFS $X$ with belief measure $\beta_X(C) = p_0 + \sum_{k=1}^{\infty} p_k Q_k(C^k)$ is said to be absolutely continuous if $Q_k$ is absolutely continuous with respect to $\lambda$ for all $k \in \mathbb{N}$.

For an absolutely continuous RFS we have the following:

Proposition A.21 ([44, Ch. 4, Prop. 19 and Cor. 21]). Let $X$ be an absolutely continuous RFS. Then, the finite set function defined, for all finite $Z \in \mathcal{F}$, by

$$f_X(Z) \triangleq \frac{\delta \beta_X}{\delta Z}(\emptyset)$$

is a set density, and

$$\mathbb{P}\left(\{X \in S\}\right) = \int_S f_X(Z) \delta Z, \quad \forall S \in \mathcal{B}(\mathcal{F}).$$

In particular, setting $S = C_A$,

$$\beta_X(A) = \int_{C_A} f_X(Z) \delta Z.$$

From the previous proposition, we immediately have that, if $X$ is an absolutely continuous RFS with belief measure $\beta_X(C) = p_0 + \sum_{k=1}^{\infty} p_k Q_k(C^k)$, then

$$f_X(\{x_1, \ldots, x_k\}) = k! p_k q_k(x_1, \ldots, x_k)$$

for all $k \in \mathbb{N}$, and for all distinct $x_1, \ldots, x_k$, where $q_k = dQ_k/d\lambda$ is the density of the probability measure $Q_k$.

A.3.2 Probability Generating Functional

Given an RFS $X$ and a real function $h$ defined on $S$, we define, for any finite set $Y$, the finite-set function

$$h^Y \triangleq \begin{cases} 1, & \text{if } Y = \emptyset \\ \prod_{i=1}^{\left|Y\right|} h(y_i), & \text{if } Y = \{y_1, \ldots, y_{\left|Y\right|}\} \end{cases}$$
**Definition A.22.** Let $X$ be an RFS with set density $f_X$, then the Probability Generating Functional of $X$ is the functional $G_X$ defined as

$$G_X[h] \triangleq \int h_Y f_X(Y) \delta Y$$

for any finite set function $h$.

This functional shares a number of interesting properties with the belief measure. Indeed we have that

$$\beta_X(C) = G_X[\mathbb{1}_C]$$

where $\mathbb{1}_C$ is the indicator function of $C$. Moreover, it allows interpreting the set derivative of the belief measure as a Fréchet derivative. Specifically, the set derivative of $\beta_X$ at $x$ evaluated in $C$ is the Fréchet derivative of $G_X$ along the direction $\delta x$ evaluated in $\mathbb{1}_C$, i.e.,

$$\frac{\delta \beta_X}{\delta x}(C) = \frac{\partial G_X}{\partial \delta x}[\mathbb{1}_C]$$

and, for any finite set $Z = \{z_1, \ldots, z_n\}$,

$$\frac{\delta \beta_X}{\delta Z}(C) = \frac{\partial^n G_X}{\partial \delta z_n \cdots \partial \delta z_1}[\mathbb{1}_C]$$

**Example A.23.** Here we show that the Bernoulli model (2.25) for a bounded random set with maximum number of elements $K$ tends to the Poissonian model (2.26) as $K \to \infty$. Notice first that, if $P$ is Poissonian with parameter $\mu$, then

$$G_P[h] = \int_S h_Y e^{-\mu_Y} \prod_{p \in Y} f_p(y) \delta Y$$

$$= e^{-\mu} \sum_{n=0}^{\infty} \frac{\mu^n}{n!} \prod_{i=1}^{n} \int_S h(y_i) f_p(y_i) \lambda(dy_i)$$

$$= e^{-\mu} \sum_{n=0}^{\infty} \frac{(\mu L[h])^n}{n!}$$

$$= e^{\mu(L[h]-1)}$$
where

\[ L[h] = \int_{\mathbb{S}} h(y)f(p(y)\lambda(dy) \quad (A.9) \]

Consider a Bernoullian random set \( Z \) with set densities as in (2.25). We have

\[
G_Z[h;K] = \int_{\mathbb{S}} y^{|Y|} \left(1 - \alpha\right)^{K-|Y|} \prod_{z \in Y} f(z) \delta Y
\]

\[
= \sum_{n=0}^{K} \binom{K}{n} \alpha^n (1 - \alpha)^{K-n} \prod_{i=1}^{n} \int_{\mathbb{S}} h(y_i)f(z_i)\lambda(dy_i)
\]

\[
= \sum_{n=0}^{K} \binom{K}{n} (\alpha L[h])^n (1 - \alpha)^{K-n}
\]

\[
= 
\left[1 - \alpha + \alpha L[h]\right]^K
\]

where \( L[h] \) is equal to (A.9) if the elements of the two random sets \( P \) and \( Z \) have the same probability density function. Setting \( \alpha = \mu/K \) we have:

\[
\lim_{K \to \infty} G_Z[h;K] = \lim_{K \to \infty} \left[1 - \frac{\mu - \mu L[h]}{K}\right]^K
\]

\[
= e^{\mu(L[h]-1)}
\]

\[
= G_P[h].
\]
The purpose of this appendix is to highlight the basic connections between finite random sets and point processes. The reader may find further details in textbooks [34, 35, 72].

A point process is a stochastic process in which realizations consist of collections of points in a suitable space: the points are indistinguishable except for their locations. Such a process can be specified mathematically in several ways, the most common being through the joint distributions of the counts of points in arbitrary sets, i.e., through a random counting measure. Random measures and random sets are intimately related. In a nutshell, since every realization of a point process is a closed set (i.e., the set of its constituent points), point processes can be viewed as special cases of random closed sets. On the other hand, almost every realization of a simple, finite point process is a finite collection of distinct points, i.e., a random finite set. Therefore, finite random sets can be viewed as special cases of point processes (more specifically, of *simple, finite point processes*).
B.1 Basic Definitions

Let \( M \) be a complete separable metric space and \( \mathcal{B}(M) \) the \( \sigma \)-algebra of its Borel sets. All the measures considered here on \((M, \mathcal{B}(M))\) are requested to satisfy the following boundedness condition:

**Definition B.1.** A Borel measure \( \mu \) on \((M, \mathcal{B}(M))\) is **boundedly finite** if \( \mu(B) < \infty \) for every bounded \( B \subset \mathcal{B}(M) \).

Let \( M \) be the space of all boundedly finite measures on \((M, \mathcal{B}(M))\), \( M_c \) the subset of \( M \) of all integer-valued measures (called here counting measures), and \( M_s \) the subset of \( M_c \) of all simple measures, i.e., \( N \in M_c \) such that \( N(\{x\}) = 0 \) or \( 1 \) for all \( x \in M \). With \( \delta_x \) denoting the Dirac measure centered at \( x \), the following holds:

**Proposition B.2 ([35, Prop. 9.1.III]).** Let \( N \in M \), then \( N \in M_c \) if and only if \( N = \sum_{i \in I} k_i \delta_{x_i} \), where \( k_i \) are positive integers, and \( \{x_i\}_{i \in I} \) is a countable set of distinct points in \( M \) with at most finitely many points in any bounded Borel set. In particular, \( N \in M_s \) if and only if \( N = \sum_{i \in I} \delta_{x_i} \).

**Proposition B.3 ([35, Prop. 9.1.IV]).** \( M \) (\( M_c \)) is a complete separable metric space, and the corresponding Borel \( \sigma \)-algebra, \( \mathcal{B}(M) \) (\( \mathcal{B}(M_c) \)), is the smallest \( \sigma \)-algebra on \( M \) (\( M_c \)) such that the mappings \( \mu \to \mu(A) \) are measurable for all \( A \in \mathcal{B}(M) \).

Proposition B.3 allows defining random measures and leads to a simple characterization of point processes.

**Definition B.4.** Let \( (\Omega, \Sigma, \mathbb{P}) \) be a probability space.

(i) A **random measure** on \( M \) is a measurable mapping \( \xi \) from \((\Omega, \Sigma)\) to \((M, \mathcal{B}(M))\).

(ii) A **point process** on \( M \) is a measurable mapping \( N \) from \((\Omega, \Sigma)\) to \((M_c, \mathcal{B}(M_c))\).
(iii) A point process $N$ is simple if $N \in \mathcal{M}_s$, $\mathbb{P}$-a.s.
(iv) A point process $N$ is finite if $N(\mathcal{M}) < \infty$, $\mathbb{P}$-a.s.

The following proposition highlights the relationship between a random measure and the random variables indexed by the Borel sets of $\mathcal{M}$.

**Proposition B.5 ([35, Cor. 9.1.IX]).** $\xi$ is a random measure ($N$ is a point process) if and only if $\xi(A)$ ($N(A)$) is a random variable for each bounded $A \in \mathcal{B}_M$.

Therefore, from Propositions B.2 and B.5, a point process $N$ is a random distribution of indistinguishable points in $\mathcal{M}$, and $N(A)$ is the number of points in $A$, for each $A \in \mathcal{B}(\mathcal{M})$ (see Figure B.1).

**Definition B.6.** The distribution of a random measure $\xi$ or point process $N$ is the probability measure it induces on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ or $(\mathcal{M}_c, \mathcal{B}(\mathcal{M}_c))$, respectively. The finite-dimensional distributions are the joint distributions, for all finite families of bounded Borel sets $\{A_i\}_{i=1}^k$, of the random variables $\{\xi(A_i)\}_{i=1}^k$ or $\{N(A_i)\}_{i=1}^k$.

![Fig. B.1 Counting variables for a realization of a simple point process.](image-url)
Proposition B.7 ([35, Cor. 9.2.IV]). The distribution of a random measure or point process is completely determined by its finite-dimensional distributions.

B.2 Simple, Finite Point Processes

A simple, finite point process is described by specifying:

(i) a probability distribution \( \{p_n\}_{n=0}^{\infty} \), which determines the total number of points in the population; and

(ii) for each \( n \in \mathbb{N} \), a symmetric measure \( Q_n \) on the Borel sets of \( \mathcal{M}^n \), which determines the joint distribution of the positions of the points of the processes, given that their total number is \( n \).

The measures \( Q_n \) need to be symmetric — i.e., \( Q_n \) must give equal weight to all \( n! \) permutations of the coordinates \( (x_1, \ldots, x_n) \) — to allow treating point processes using unordered sets. Should nonsymmetric measures \( Q'_n \) be given, then the symmetry needed in condition ((ii) can be easily achieved by using

\[
Q_n(A_1, \ldots, A_n) = \frac{1}{n!} \sum_{\pi \in \Pi} Q'_n(A_{\pi(1)}, \ldots, A_{\pi(n)})
\]

for any partition \( A_1, \ldots, A_n \) of \( \mathcal{M} \), where \( \Pi \) is the set containing all of the \( n! \) permutation of the integers \( 1, \ldots, n \). An alternative notation, which has some advantages in simplifying combinatorial formulas, uses the family of (nonprobability) measures

\[
J_n(A_1 \times \cdots \times A_n) = p_n \sum_{\pi \in \Pi} Q'_n(A_{\pi(1)}, \ldots, A_{\pi(n)})
\]

\[
= n! p_n Q_n(A_1 \times \cdots \times A_n)
\]

which are known as Janossy measures.
Example B.8. Let \( \{ J_n \}_{n=0}^{\infty} \) be the Janossy measures of a simple, finite point process \( N \), and let \( A, B \in \mathcal{B}(\mathbb{M}) \) be disjoint. Then

\[
J_2(A \times B) = 2p_2Q_2(A \times B)
\]

\[
= p_2(Q_2(A \times B) + Q_2(B \times A))
\]

is the probability that \( N \) has exactly 2 points, one located in \( A \) and one in \( B \).

Example B.8 can be extended to an arbitrary number of disjoint sets. Let \( \{ A_i \}_{i=1}^{k} \) be \( k \) disjoint sets in \( \mathcal{B}(\mathbb{M}) \), and \( C = (A_1 \cup \cdots \cup A_k)^c \), then the probability of finding exactly \( n_i \) points in \( A_i \), \( i = 1, \ldots, k \), is

\[
\mathbb{P}\left( \bigcap_{i=1}^{k} \{ N(A_i) = n_i \} \right) = p_n\left( \begin{array}{c} n \\ n_1, \ldots, n_k \end{array} \right) Q_n(A_1^{n_1} \times \cdots \times A_k^{n_k})
\]

\[
= \frac{J_n(A_1^{n_1} \times \cdots \times A_k^{n_k})}{n! \cdots n_k!}
\]

(B.1)

where \( n = \sum_{i=1}^{k} n_i \), and the multinomial coefficient

\[
\left( \begin{array}{c} n \\ n_1, \ldots, n_k \end{array} \right) = \frac{n!}{n_1! \cdots n_k!}
\]

accounts for all the distributions of \( n \) indistinguishable points into \( k \) groups of \( n_1, \ldots, n_k \) elements. Therefore, we have the following

Proposition B.9 ([35, Ex. 9.2(c)]). The distribution of a simple, finite point process is completely determined by the family of Janossy measures.

B.2.1 Moment measures and densities

From Proposition B.5, the family of random variables \( N(A) \), for each bounded \( A \in \mathcal{B}(\mathbb{M}) \), completely specifies the point process \( N \). Therefore, it is reasonable to consider the moments of \( N(A) \) when defining the moments of a point process. For integers \( n \) and \( k \), with \( n \geq k \), \( n[k] \triangleq n(n-1) \cdots (n-k+1) \) denotes the \( k \)th factorial power of \( n \).
**Definition B.10.** If $N$ has finite $k$th moment $\mathbb{E}[(N(M))^k]$, then

(i) the *expectation measure* is defined by

$$M(A) = \mathbb{E}[N(A)], \quad \forall A \in \mathcal{B}(M)$$

(ii) the *$k$th moment measure* as $M_k = M$, if $k = 1$, and, for $k \geq 2$,

$$M_k(A_1^{k_1} \times \cdots \times A_r^{k_r}) = \mathbb{E}[(N(A_1))^{k_1} \cdots (N(A_r))^{k_r}]$$

where $A_i$ are disjoint subset of $\mathcal{B}(M)$, and $k_i$ are positive integers such that $k_1 + \cdots + k_r = k$.

(iii) the *$k$th factorial moment measure* as $M[k] = M$, if $k = 1$, and, for $k \geq 2$,

$$M[k](A_1^{k_1} \times \cdots \times A_r^{k_r}) = \mathbb{E}[(N(A_1))^{[k_1]} \cdots (N(A_r))^{[k_r]}]$$

where $A_i$ are disjoint subset of $\mathcal{B}(M)$, and $k_i$ are positive integers such that $k_1 + \cdots + k_r = k$.

---

**Proposition B.11 ([34, Th. 5.4.I]).** If $\mathbb{E}[(N(M))^k] < \infty$, the set function $M_k$ and $M[k]$ are countably additive on rectangular sets, and have unique extensions to symmetric measures on $\mathcal{B}(M^k)$.

In fact, $M_k$ and $M[k]$ can be interpreted as the expectation measure of a certain point process in $M^k$: the point process consists of all $k$-tuples (distinguishing the order and allowing repetitions in $M_k$, not allowing repetitions in $M[k]$) of points from the original realization.

In order to explore the relation between factorial moments and Janossy measures, consider the following expression for the expectation measure. For all $A \in \mathcal{B}(M)$,

$$M(A) = \sum_{k=1}^{\infty} k \mathbb{P}(\{N(A) = k\})$$

$$= \sum_{k=1}^{\infty} k \sum_{m=0}^{\infty} \mathbb{P}(\{N(A) = k, N(A^c) = m\})$$
Relation to Point Processes

\[
\sum_{k=1}^{\infty} \sum_{m=0}^{\infty} J_{k+m} \left( A^k \times (A^c)^m \right) \frac{1}{k!m!} = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} J_{k+m+1} \left( A^{k+1} \times (A^c)^m \right) \frac{1}{k!m!}
\]

where the third equality follows from (B.1). Grouping together the terms for which \( m + k = n \)

\[
M(A) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\substack{m, n \geq 0: \atop m + k = n}} \left( \binom{n}{k, m} J_{n+1} \left( A \times A^k \times (A^c)^m \right) \right).
\]

Partitioning \( M \) in two subsets, \( A \) and \( A^c \), induces a partition into \( M^n \) in \( 2^n \) subsets indexed by \( k, m \) such that \( k + m = n \). For fixed \( k \) and \( m \), there are \( n!/(k!m!) \) such subsets, all with measure \( J_{n+1} \left( A \times A^k \times (A^c)^m \right) \), for the Janossy measures are symmetric. Therefore, the inner summation is the measure of the whole \( M^n \), and

\[
M(A) = \sum_{n=0}^{\infty} \frac{1}{n!} J_{n+1} \left( A \times M^n \right).
\]

This result can be extended to factorial moments as follows.

**Theorem B.12.** ([34, Th. 5.4.II]) If \( \mathbb{E} \left[ (N(M))^k \right] < \infty \), then

\[
M_{[k]}(B) = \sum_{n=0}^{\infty} \frac{1}{n!} J_{n+k} \left( B \times M^n \right), \quad \forall B \in \mathcal{B}(M^n). \tag{B.2}
\]

Conversely, if \( \mathbb{E} \left[ (N(M))^k \right] < \infty \) for all \( k \), and, for some \( \epsilon > 0 \), \( M \left[ z^{N(M)} \right] \) is analytic in \( |z| < 2 + \epsilon \), then, for all \( n \),

\[
J_n(B) = \sum_{k=0}^{\infty} \left( \frac{-1}{k!} \right)^k M_{[n+k]} \left( B \times M^k \right), \quad \forall B \in \mathcal{B}(M^n). \tag{B.3}
\]

Consider now the case of \( M \) a real Euclidean space. Let \( \lambda \) be a reference measure on \( (M, \mathcal{B}(M)) \), then

**Lemma B.13** ([34, Lem. 5.4.III]). If \( M_{[k]} \) exists, then it is absolutely continuous with respect to \( \lambda \) if and only if \( J_n \) is absolutely
continuous with respect to \( \lambda \) for all \( n \geq k \), in which case \( m_{[k]} = dM_{[k]}/d\lambda \) and \( j_n = dJ_n/d\lambda \) are related by

\[
m_{[k]}(x_1, \ldots, x_k) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{M^n} j_{k+n}(x_1, \ldots, x_k, y_1, \ldots, y_n) dy_1 \cdots dy_n
\]

for all \((x_1, \ldots, x_k) \in M^k\). Conversely, if \( M_{[k]} \) exists and is absolutely continuous with respect to \( \lambda \), for all \( k \), and if the series in (B.3) is absolutely convergent, then the corresponding \( J_n \) is absolutely continuous with respect to \( \lambda \), and

\[
j_n(x_1, \ldots, x_n) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{M^k} m_{[n+k]}(x_1, \ldots, x_n, y_1, \ldots, y_k) dy_1 \cdots dy_k
\]

for all \((x_1, \ldots, x_n) \in M^n\).

The Janossy densities and the factorial moment densities (also called product densities) are the Radon–Nikodym derivatives of the corresponding measure with respect to \( \lambda \).

**Remark B.14.** It may be interesting to express the factorial moment measures in terms of the distributions \( p_n \) and \( Q_n \). Equation (B.2) can be developed as

\[
M_{[k]}(B) = \sum_{n=0}^{\infty} \frac{(n + k)!}{n!} p_{n+k} Q_{k+n}(B \times M^n)
\]

for all \((x_1, \ldots, x_n) \in M^n\). If \( M_{[k]} \) is absolutely continuous with respect to \( \lambda \), we also have

\[
m_{[k]}(x_1, \ldots, x_k) = q_k(x_1, \ldots, x_k) \sum_{n=k}^{\infty} n^{[k]} p_n
\]

Therefore, if \( M_{[k]} \) is absolutely continuous with respect to \( \lambda \), we also have

\[
m_{[k]}(x_1, \ldots, x_k) = q_k(x_1, \ldots, x_k) \sum_{n=k}^{\infty} n^{[k]} p_n
\]

for all \((x_1, \ldots, x_k) \in M^k\).
where $q_k = dQ_k/d\lambda$. In particular, taking $k = 1$,

$$M(B) = Q_1(B)M(\mathcal{M})$$

$$m(x) = q_1(x)M(\mathcal{M})$$

so that the average number of points in $B \in \mathcal{B}_{\mathcal{M}}$ is equal to the average number of the population times the probability measure of the set $B$.

### B.3 Point Processes and Random Sets

In the intimate relationship between random sets and random measures, the notion of support of a measure plays a key role. Here $\mathcal{M}$ is supposed to be also locally compact.

**Definition B.15.** The *support* of a measure $\mu$ on $(\mathcal{M}, \mathcal{B}_\mathcal{M})$, denoted $\text{supp } \mu$, is the set of all $x \in \mathcal{M}$ such that $x \in G$ for an open set $G$ implies $\mu(G) > 0$.

The following proposition answers the question as whether a random measure determines the distribution of a random closed set.

**Proposition B.16 ([72, Props. 8.1 and 8.16]).** If $\mu \in \mathcal{M}$, then $\text{supp } \mu$ is closed. Moreover, if $\xi$ is a random measure, then $\text{supp } \xi$ is a random closed set of which the distribution is uniquely determined by the distribution of $\xi$.

The converse is not always true, and the conditions under which it holds have been studied in [15, 19]. The connection between random measures and point processes is stronger in the case of simple point processes. Indeed, from Definition B.4, almost all realizations of a simple point process are simple counting measures, and, from Proposition B.2, they are uniquely determined by their support, which is a locally finite subset of $\mathcal{M}$ (i.e., with at most a finite number of points in any compact set). Therefore,

**Proposition B.17 ([72, Cor. 8.2]).** $N$ is a simple point process in $\mathcal{M}$ if and only if its support is a locally finite random closed set.
Since the distribution of a random closed set \( X \) is uniquely determined by its capacity functional \( T_X \) (see Appendix A), an immediate consequence of Proposition B.17 is the following

**Theorem B.18 ([72, Th. 8.3]).** The distribution of a simple point process \( N \) in \( \mathbb{M} \) is uniquely determined by the probabilities \( \mathbb{P}(\{N(K) = 0\}) \), for any compact subset \( K \) of \( \mathbb{K} \).

Indeed, if \( N \) is a a simple point process in \( \mathbb{M} \), then \( X = \text{supp} \, N \) is a locally compact random closed set, and

\[
\mathbb{P}(\{N(K) = 0\}) = 1 - \mathbb{P}(\{X \cap K = \emptyset\})
\]

\[
= 1 - T_X(K).
\]

From Proposition B.17 we also have the following

**Corollary B.19.** If \( X \) is a finite random set on \( \mathbb{M} \), then \( N = \sum_{x \in X} \delta_x \) is a simple, finite point process. Conversely, if \( N \) is a simple, finite point process on \( \mathbb{M} \), then \( X = \text{supp} \, N \) is a finite random set.

As a simple example, consider the realization of the point-process/random-set of Figure B.2. Then, \( N(w) = \delta_{x_1} + \cdots + \delta_{x_5} \) and \( X(w) = \{x_1, \ldots, x_5\} \). Finally, observe that the set densities and the PHD in [44, 64, 66] are just the Janossy densities \( j_n(x_1, \ldots, x_n) \) and the expectation density \( m(x) \), respectively, as also pointed out in [63].

---

**Fig. B.2** A realization of a simple, finite point process or, equivalently, of a finite random set: in this case \( N(w) = \delta_{x_1} + \cdots + \delta_{x_5} \), and \( X(w) = \{x_1, \ldots, x_5\} \).
C

Relation to Dempster–Shafer Theory

It was pointed out in [73] that the mathematical theory of evidence, usually referred to as Dempster–Shafer (D–S) theory [78], can be described rigorously in terms of random sets. See also [83, p. 43–44], [66]. In this appendix we summarize Dempster–Shafer “theory of evidence,” its connection with standard probability theory, and its connection with random-set theory.

C.1 Theory of Evidence vs. Probability Theory

Standard probability theory deals with numbers that reflect how often an event will happen if an experiment is repeated a large number of times. Now, when one performs a random experiment, the probability density governing it is, at first, unknown. If one examines the “belief” in a given proposition rather than its probability of being true, two opposite views have been advocated. One assumes that the numerical degree of support of a given proposition is objectively determined by a body of evidence. The competing one claims that the degrees of belief are psychological facts, to be discovered by observing an individual’s preferences among bets. D–S theory distinguishes between probabilities
and beliefs: the former must be conceived as a feature of the world, and hence “are not necessarily features of our knowledge or belief,” while the latter do not reflect “what Nature chooses, but rather the state of our knowledge after making a measurement.” “If we know [the probabilities], then we should adopt them as our degrees of belief. But if we do not know [the probabilities], then it will be an extraordinary coincidence for our degrees of belief to be equal to them.” [78]

Probability theory takes it as given that a statement is either true or false. Thus, since “standard” probability theory “cannot distinguish between lack of belief and disbelief, it does not allow one to withhold belief from a proposition without according that belief to the negation of the proposition.” [78]. D–S theory admits plausible propositions, those either supported by evidence or by uncertainty, and allows for more nebulous statements, such as “I don’t know,” and hence can be used in problems in which probabilities are unknown, or only partially known.

C.2 D–S Theory

Let $\mathcal{X}$ be a finite set, and $2^{\mathcal{X}}$ its power set. Assign a belief mass to all elements of $2^{\mathcal{X}}$:

$$ m : 2^{\mathcal{X}} \mapsto [0, 1] $$

such that

$$(a) \quad m(\emptyset) = 0$$

$$(b) \quad \sum_{A \in 2^{\mathcal{X}}} m(A) = 1$$

The belief mass assigned to the set $A \in 2^{\mathcal{X}}$ describes the portion of belief assigned to $A$ (but to no particular subset of $A$ — subsets of $A$ have their own belief masses).\(^1\)

\(^1\)For a different interpretation, J. Pearl [75, pp. 420 ff.] defines the value of $m(A)$ as a measure of the strength of the argument in favor of hypothesis $A$. 
Given $A \in 2^X$, we define the belief of $A$ (or the degree of assurance in $A$ [75, p. 418]) as the total mass of all sets implying $A$:

$$\text{bel}(A) \triangleq \sum_{B \mid B \subseteq A} m(B)$$

and the plausibility of $A$ as the total mass of all sets not inconsistent with $A$ (or, the probability that $A$ is compatible with the available evidence, and hence possible [75, p. 41]) as

$$\text{pl}(A) \triangleq \sum_{B \mid B \cap A \neq \emptyset} m(B)$$

The belief is not generally a probability measure: in fact

$$\text{bel}(\emptyset) = 0$$
$$\text{bel}(2^X) = 1$$
$$\text{bel}(A) \leq \text{bel}(B), \text{ if } A \subseteq B,$$
and

$$\text{bel}(A \cup B) \geq \text{bel}(A) + \text{bel}(B), \text{ if } A \cap B = \emptyset.$$ If

$$\text{bel}(A \cup B) = \text{bel}(A) + \text{bel}(B), \text{ if } A \cap B = \emptyset$$
then bel($A$) is called a Bayesian belief function.

Some authors also define a “commonality” function as

$$Q(A) \triangleq \sum_{B \mid A \subseteq B} m(B)$$

Notice that bel($A$) does not reveal to what extent one doubts $A$ — i.e., to what extents one believes in its negation $A^c$. ($\cdot)^c$ denoting set complementation. We may define the degree of doubt about $A$ as

$$\text{dou}(A) \triangleq \text{bel}(A^c)$$

The relations between belief mass, belief, and plausibility are expressed by

$$\text{pl}(A) = 1 - \text{bel}(A^c) \quad \text{(C.1a)}$$
$$m(a) = \sum_{B \mid B \subseteq A} (-1)^{|A-B|} \text{bel}(B) \quad \text{(C.1b)}$$
where (C.1b) is usually referred to as the M"obius Transform of \( \text{bel}(\cdot) \). Notice that beliefs are not probabilities: we have in particular

\[
\begin{align*}
\text{bel}(\emptyset) &= 0 \\
\text{bel}(2^X) &= 1 \\
\text{bel}(A \cup B) &\geq \text{bel}(A) + \text{bel}(B), \quad A \cap B = \emptyset
\end{align*}
\]

A probability of an event may be defined through the belief and plausibility of that event by the inequalities [36] \(^2\)

\[
\text{bel}(A) \leq \mathbb{P}(A) \leq \text{pl}(A).
\]

A fundamental tool of D–S theory is Dempster’s rule of combination, which “can be interpreted as a method for changing prior opinions in the light of new evidence” [78], and hence viewed as an information fusion tool (see [40] for an example of application). Given two distributions \( m_1 \) and \( m_2 \) of belief masses, Dempster’s rule of combination yields

\[
\begin{align*}
m_{1,2}(\emptyset) &= 0 \\
m_{1,2}(A) &= (m_1 \oplus m_2)(A) \\
&= \frac{1}{1 - K} \sum_{B \cap C = A \neq \emptyset} m_1(B)m_2(C)
\end{align*}
\]

where

\[
K \triangleq \sum_{B \cap C = \emptyset} m_1(B)m_2(C)
\]

is a measure of the amount of conflict between the two mass distributions. This combination rule for evidence can produce counterintuitive results when there is significant conflict: on the other hand, as observed in [66, p. 141], “the fact that two information sources are greatly in conflict indicates that there is something wrong or incomplete in our modeling.”

\(^2\)Quoting Shafer [78, p. 16], “If we know [the probabilities], then we should adopt them as our degrees of belief. But if we do not know [the probabilities], then it will be an extraordinary coincidence for our degrees of belief to be equal to them. (⋯) [Probabilities] must be conceived as features of the world. They are not necessarily features of our knowledge or belief.”
As a simple example, consider the statement “source is active,” whose belief is 0.5 and whose plausibility is 0.8. The above means that we have evidence allowing us to state strongly that the source is active with a confidence of 0.5. The evidence contrary to that proposition (“source is inactive”) has confidence 0.2. The remaining mass (0.3 = 0.5 – (1 – 0.8)) is “indeterminate” (“source either active or inactive”). The plausibility is either supported by evidence or by uncertainty. See Figure C.1.

C.3 Connection with Random-Set Theory

In [66, pp. 144 ff.] Dempster–Shafer theory is reformulated in terms of RST. The key equality is obtained by interpreting beliefs as probabilities:

\[ m(A) = P(X = A) \]

where \( X \) denotes a random set called a representation of \( m \). It follows that plausibility, belief, and commonality can be redefined probabilistically:

\[ p_l(A) = P(X \cap A \neq \emptyset) \]
\[ b_e(A) = P(X \subseteq A) \]
\[ Q(A) = P(X \supseteq A) \]

and

\[ (m_1 \oplus m_2)(A) = P(X_1 \cap X_2 = A | X_1 \cap X_2 \neq \emptyset) \]

provided that \( X_1 \) and \( X_2 \) are statistically independent and \( P(X_1 \cap X_2 \neq \emptyset) \neq 0 \).

It should be kept in mind that RST, unlike Dempster–Shafer theory, allows \( m(\emptyset) \neq 0 \).
# Acronyms and notations

## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP</td>
<td>A posteriori probability</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive white Gaussian noise</td>
</tr>
<tr>
<td>BER</td>
<td>Bit-error rate</td>
</tr>
<tr>
<td>BSEP</td>
<td>Bit-sequence error probability</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code-division multiple access</td>
</tr>
<tr>
<td>CPHD</td>
<td>Cardinalized probability hypothesis density</td>
</tr>
<tr>
<td>CCSI</td>
<td>Complete channel state information</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel state information</td>
</tr>
<tr>
<td>D–S</td>
<td>Dempster–Shafer</td>
</tr>
<tr>
<td>DS-CDMA</td>
<td>Direct sequence code-division multiple access</td>
</tr>
<tr>
<td>DSEP</td>
<td>Discrete set error probability</td>
</tr>
<tr>
<td>EKF</td>
<td>Extended Kalman filter</td>
</tr>
<tr>
<td>FISST</td>
<td>Finite-set statistics</td>
</tr>
<tr>
<td>GM</td>
<td>Gaussian mixture</td>
</tr>
<tr>
<td>GMAP</td>
<td>Global maximum a posteriori</td>
</tr>
<tr>
<td>GSD</td>
<td>Generalized sphere decoding</td>
</tr>
</tbody>
</table>
### Notations and acronyms

- **i.i.d.** Independently and identically distributed
- **KF** Kalman filter
- **LCHS** Locally compact, Haussdorf, and second countable
- **LS** Least square
- **MAP** Maximum a posteriori
- **MIMO** Multiple-input, multiple-output
- **ML** Maximum likelihood
- **MUD** Multiuser detection
- **ND** Neighbor discovery
- **OFDM** Orthogonal frequency division multiplexing
- **pdf** Probability density function
- **PHD** Probability hypothesis density
- **RACS** Random closed set
- **RFS** Random finite set
- **RHS** Right-hand side
- **RMSE** Root mean-square error
- **RST** Random-set theory
- **SD** Sphere decoding
- **SEP** Set error probability
- **SISO** Single-input, single-output
- **SMC** Sequential Monte Carlo
- **SSEP** Set sequence error probability

### General Notations

- **\( \mathbb{R} \)** Space of real numbers
- **\( \mathbb{R}^+ \)** Space of positive real numbers
- **\( \mathbb{R}^d \)** Space of real \( d \)-dimensional vectors
- **\( \mathbb{N} \)** Space of natural numbers, i.e., \( \{1, 2, \ldots\} \)
- **\( \times \)** Cartesian product
- **\( a, b, c, \ldots \)** Elements (possibly random) of some set
- **\( \mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots \)** Vectors (possibly random)
- **\( \mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots \)** Matrices (possibly random)
- **\( \emptyset \)** Empty set
- **\( A, B, C, \ldots \)** Sets
- **\( 2^A \)** Power set of the set \( A \), i.e., the collection of all subsets of \( A \)
<table>
<thead>
<tr>
<th>Notations and acronyms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>A</td>
</tr>
<tr>
<td>$B_{x,r}$</td>
<td>Open ball centered at $x$ with radius $r$</td>
</tr>
<tr>
<td>$\hat{B}_{x,r}$</td>
<td>Closure of $B_{x,r}$</td>
</tr>
<tr>
<td>$\subseteq$</td>
<td>Set inclusion</td>
</tr>
<tr>
<td>$\subset$</td>
<td>Strict set inclusion</td>
</tr>
<tr>
<td>\setminus</td>
<td>Set difference</td>
</tr>
<tr>
<td>$\cup$</td>
<td>Set union</td>
</tr>
<tr>
<td>$\cap$</td>
<td>Set intersection</td>
</tr>
<tr>
<td>$\cdot^c$</td>
<td>Set complementation</td>
</tr>
<tr>
<td>$\oplus$</td>
<td>Direct sum</td>
</tr>
<tr>
<td>$\triangleq$</td>
<td>Equal by definition</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>Asymptotic notation for the growth rate of a function</td>
</tr>
<tr>
<td>$(\cdot)^T$</td>
<td>Transpose</td>
</tr>
<tr>
<td>$(\cdot)^\dagger$</td>
<td>Pseudoinverse (Moore–Penrose generalized inverse)</td>
</tr>
<tr>
<td>$(\cdot)^{-1}$</td>
<td>Inverse</td>
</tr>
<tr>
<td>$\mathbf{I}_L$</td>
<td>$L \times L$ identity matrix</td>
</tr>
<tr>
<td>$\text{diag}(a_1, \ldots, a_n)$</td>
<td>Diagonal matrix with $a_1, \ldots, a_n$ on the main diagonal</td>
</tr>
<tr>
<td>$n!$</td>
<td>Factorial, $n(n-1)\cdots1$</td>
</tr>
<tr>
<td>$\binom{n}{k}$</td>
<td>Binomial coefficient, $\frac{n!}{k!(n-k)!}$</td>
</tr>
<tr>
<td>$\binom{n}{n_1 \ldots n_k}$</td>
<td>Multinomial coefficient, $\frac{n!}{n_1! \cdots n_k!}$</td>
</tr>
<tr>
<td>$| \cdot |$</td>
<td>Euclidean norm</td>
</tr>
<tr>
<td>$\mathbb{I}_A$</td>
<td>Indicator function of the set $A$, i.e., $\mathbb{I}_A(x) = 1$, if $x \in A$, and $\mathbb{I}_A(x) = 0$, otherwise</td>
</tr>
<tr>
<td>$\delta_x$</td>
<td>Dirac measure centered at $x$, i.e., $\delta_x(A) = 1$, if $x \in A$, and $\delta_x(A) = 0$, otherwise</td>
</tr>
<tr>
<td>$(\Omega, \Sigma, \mathbb{P})$</td>
<td>Underlying probability space</td>
</tr>
<tr>
<td>$\mathbb{P}$ler, $\mathbb{P}$-a.s.</td>
<td>Almost everywhere, almost surely with respect $\mathbb{P}$</td>
</tr>
<tr>
<td>$E$</td>
<td>Expected value</td>
</tr>
<tr>
<td>$\mathcal{N}(\mu, \mathbf{C})$</td>
<td>Gaussian distribution with mean $\mu$ and covariance matrix $\mathbf{C}$</td>
</tr>
<tr>
<td>$\mathcal{N}_c(\mu, \mathbf{C})$</td>
<td>Gaussian, circularly symmetric distribution with mean $\mu$ and covariance matrix $\mathbf{C}$</td>
</tr>
<tr>
<td>$\mathcal{P}(\mu)$</td>
<td>Poisson distribution with mean $\mu$</td>
</tr>
</tbody>
</table>
Random-Set Notations

☞ $S$ Space underlying a random set
☞ $\mathcal{F}(S)$ Family of subsets of $S$
☞ $[S]^k$ Lexicographically ordered subset of $S^k$
☞ $\mathcal{F}, \mathcal{G}, \mathcal{K}$ The collection of closed, open, compact subsets of $S$
☞ $\mathcal{F}^K, \mathcal{F}_K$ The collections of closed subsets of $S$ “missing” $K$ and “hitting” $K$
☞ $\mathcal{F}_{\geq k}, \mathcal{F}_{\leq k}, \mathcal{F}=k$ Families of closed subsets of $S$ with at least, at most, exactly $k$ elements
☞ $\mathcal{C}_A$ The collection of all closed subsets of the set $A$
☞ $\mathcal{B}(S)$ Borel $\sigma$-algebra of subsets of $S$
☞ $\frac{\delta \Phi}{\delta x}$ Set derivative at point $x$ of the set function $\Phi$
☞ $\frac{\delta \Phi}{\delta X}$ Set derivative at set $X$ of the set function $\Phi$
☞ $X$ Random set $X$
☞ $\hat{X}$ Estimate of random set $X$
☞ $C(S,B)$ Cost incurred when $X = S$ and $\hat{X} = B$
☞ $\pi(X)$ Discrete part of $X$
☞ $\pi'(X)$ Continuous part of $X$
☞ $\beta_X$ Belief function of $X$
☞ $f_X$ Belief density of $X$
☞ $G_X$ Probability generating functional of $X$
☞ $T_X$ Capacity functional of $X$
☞ $C_X$ Containment functional of $X$
☞ $\delta_A$ Dirac measure centered at set $A$, i.e.,
\[ \delta_A(Z) = 1, \text{ if } A \in Z, \text{ and } \delta_A(A) = 0, \]
\[ \text{otherwise, where } Z \in \mathcal{B}(\mathcal{F}) \]
☞ $S_t$ Set of survivors at epoch $t$
☞ $N_t$ Set of new elements born at epoch $t$
☞ $f_{X_t | X_{t-1}}(X_t | X_{t-1})$ Transition density of the RFS $X_t$
<table>
<thead>
<tr>
<th>Notations and acronyms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{Y_t</td>
<td>X_t}$</td>
</tr>
<tr>
<td>$f_{X_t</td>
<td>Y_{1:t}}$</td>
</tr>
<tr>
<td>$D_{t</td>
<td>t}$</td>
</tr>
<tr>
<td>$M_X$</td>
<td>First-order moment measure of RFS $X$</td>
</tr>
<tr>
<td>$m_X$</td>
<td>Intensity function of RFS $X$</td>
</tr>
<tr>
<td>$\mathcal{M}, \mathcal{M}_c, \mathcal{M}_s$</td>
<td>Space of all boundedly finite measures, counting measures, simple measures on a given measurable space</td>
</tr>
<tr>
<td>supp</td>
<td>Support</td>
</tr>
<tr>
<td>$J_n$</td>
<td>Janossy measure</td>
</tr>
<tr>
<td>bel($A$)</td>
<td>Belief of the set $A$</td>
</tr>
<tr>
<td>pl($A$)</td>
<td>Plausibility of the set $A$</td>
</tr>
<tr>
<td>Q($A$)</td>
<td>Commonality of the set $A$</td>
</tr>
<tr>
<td>dou($A$)</td>
<td>Degree of doubt of the set $A$</td>
</tr>
</tbody>
</table>
References


459


References

