Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

Diplomarbeit

Random Finite Sets: Theory and an Image Processing Application

ausgeführt zur Erlangung des akademischen Grades
eines Diplom-Ingenieurs unter der Leitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Franz Hlawatsch
Dipl.-Ing. Dr.techn. Günther Koliander

am

Institute of Telecommunications

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Rene Repp
0725397

Wien, im März 2016

© Copyright 2016 Rene Repp

Abstract

Starting in the early 2000s, there has been a steadily growing interest in random finite sets in the research community worldwide. Since then, the theory of random finite sets has been successfully used in a wide variety of scientific fields and in applications such as radar and sonar systems, air traffic control and navigation, telecommunications, medicine, audio and image processing, visual tracking, robotics, agriculture, and forestry. The first main contribution of this thesis is to provide a systematic, detailed, and rigorous introduction to the theory of random finite sets that can serve as an entry point for readers with no prior exposure to this field. In the second main contribution, we apply the theory of random finite sets to the problem of estimating the states of an unknown and random number of objects from image observations. We investigate a scenario where two independent sensors record partly overlapping images of a bigger scene and derive an estimator based on the posterior probability hypothesis density. We show how the estimation performance can be improved by exchanging information between the sensors over the case where each sensor calculates the estimates separately. Furthermore, we propose a novel algorithm solving this estimation problem and demonstrate its performance in simulated scenarios.

Kurzfassung

Mit Beginn der 2000er Jahre zeichnete sich ein stetig wachsendes Interesse der internationalen Forschungsgemeinschaft an endlichen Zufallsmengen ab. Seitdem wurde die Theorie der endlichen Zufallsmengen erfolgreich in vielen Wissenschaftszweigen und in Anwendungen wie Radar- und Sonarsystemen, Flugsicherung und Navigation, Telekommunikation, Medizin, Audio- und Bildverarbeitung, visuelles Tracking, Robotik sowie Land- und Forstwirtschaft eingesetzt. Der erste Hauptbeitrag dieser Arbeit besteht in einer systematischen, detaillierten und mathematisch rigorosen Einführung in die Theorie der endlichen Zufallsmengen, die als erster Einstiegspunkt für Leser und Leserinnen ohne Vorkenntnisse auf diesem Gebiet dienen kann. Im zweiten Hauptbeitrag dieser Arbeit wird die Theorie der endlichen Zufallsmengen auf ein Schätzproblem angewandt, in dem die Zustände einer unbekannten und zufälligen Anzahl von Objekten basierend auf Bilddaten geschätzt werden sollen. Wir untersuchen ein Szenario, in dem zwei voneinander unabhängige Sensoren teilweise überlappende Bilder aufnehmen und entwickeln einen Schätzer basierend auf der a-posteriori *probability hypothesis density*. Wir zeigen, wie die Schätzqualität durch Austausch von Informationen zwischen den Sensoren gegenüber dem Fall verbessert werden kann, in dem jeder Sensor Schätzwerte getrennt berechnet. Schließlich entwickeln wir einen neuartigen Algorithmus zur Lösung dieses Schätzproblems und demonstrieren dessen Ergebnisse in simulierten Szenarien.

Acknowledgments

I want to thank my supervisors, Prof. Franz Hlawatsch and Dr. Günther Koliander, for their continual support and the many interesting discussions and valuable feedback during my studies. I would also like to thank Dr. Florian Meyer for his helpful and insightful inputs, which are very much appreciated. Last but not least, I would like to express my gratitude to my family and friends for their constant support and encouragement. Without them, none of this would have been possible.

Contents

1 Introduction **1**

1.1 Motivation . . . 1

1.2 State of the Art . . . 2

1.3 Thesis Outline . . . 3

2 Fundamentals of Measure Theoretic Probability Theory **5**

2.1 σ-Algebras . . . 5

2.2 Topologies and Borel Algebras . . . 6

2.3 Product σ-Algebras . . . 8

2.4 Measure Spaces . . . 9

2.5 Measurable Functions . . . 12

2.6 Probability Spaces and Random Elements . . . 13

2.7 Integration . . . 14

2.7.1 Definition of the Lebesgue Integral . . . 14

2.7.2 Properties of the Lebesgue Integral . . . 17

3 Random Finite Sets **21**

3.1 Definition of Random Finite Sets . . . 22

3.2 Cardinality Distribution . . . 23

3.3 Belief Mass Function . . . 24

3.4 Relation between Finite Sets and Vectors . . . 25

3.4.1 Set Measure for k-ary Sets . . . 27

3.4.2 Set Measure for General Finite Sets . . . 30

3.5 Probability Density Function for RFS . . . 32

3.6 Relation between RFSs and Random Vectors . . . 34

3.6.1 k-ary RFS . . . 34

3.6.2 General RFS . . . 36

3.7 Joint Distributions 38
3.7.1 Joint Probability Measure 38
3.7.2 Joint Cardinality Distribution 39
3.7.3 Joint Belief Mass Function 40
3.7.4 Joint Probability Density Function 40
3.8 Conditional Distributions 45
3.8.1 Conditional Probability Density Functions 45
3.8.2 Conditional Probability Measure 45
3.8.3 Conditional Cardinality Distribution 46
3.8.4 Conditional Belief Mass Function 46
3.8.5 Statistical Independence 46
3.8.6 Conditional Independence 47
3.9 Expectation for RFS 47

4 Finite Set Statistics (FISST) 49
4.1 Set Integral 49
4.2 Probability Generating Functionals 50
4.2.1 Joint Probability Generating Functionals 53
4.2.2 Conditional Probability Generating Functionals 54
4.3 Functional Derivatives 54
4.4 Probability Hypothesis Density 57
4.5 Elementary Distributions 59
4.5.1 Independent and Identically Distributed Cluster RFS 59
4.5.2 Poisson RFS 60
4.5.3 Bernoulli RFS 60
4.5.4 Multi-Bernoulli RFS 61

5 State Estimation from Image Observations 63
5.1 State Estimation from a Single Image 63
5.1.1 Data Model 63
5.1.2 Estimator 64
5.2 State Estimation from Two Images 65
5.2.1 Data Model 65
5.2.2 Estimator 67
5.3 Numerical Study 70

5.3.1 Single Image Scenario . . . 70
5.3.2 Two Images Scenario . . . 74
5.3.3 Performance Evaluation . . . 76

6 Conclusion **81**

A Mathematical Proofs **83**
A.1 Proof of Lemma 3.3 . . . 83
A.2 Proof of Lemma 3.5 . . . 84
A.3 Proof of Lemma 3.7 . . . 85
A.4 Proof of Lemma 3.9 . . . 86
A.5 Proof of Theorem 3.15 . . . 87
A.6 Proof of Theorem 4.6 . . . 88
A.7 Proof of Proposition 5.1 . . . 89
A.8 Proof of Lemma 5.3 . . . 90
A.9 Proof of Proposition 5.5 . . . 90
A.10 Proof of Proposition 5.6 . . . 93

Bibliography **98**

List of Figures

2.1 Illustration of the Lebesgue measure λ_2 on $\mathbb{R}^2$. The left part shows a simple open cover of the set A consisting of only one two-dimensional (2D) interval I. The right part shows a tighter open cover consisting of four 2D intervals $I_1, \ldots, I_4$, which has a smaller total area than I. One can repeat this refinement by searching for another, tighter open cover of A, containing more 2D intervals. The Lebesgue measure can be thought of as the area of the open cover in the limit of this refinement process. 11

2.2 The function $f(x)$ is approximated by a simple function $f_i(x)$, for which the Lebesgue integral with respect to λ can be calculated. One can find another simple function with more steps that fits the curve $f(x)$ more tightly. The Lebesgue integral of $f(x)$ with respect to λ is the area in the limit of this refinement process. 15

3.1 Mapping of 2-dimensional vectors ($N = 1$, $k = 2$) to finite subsets of $\mathbb{R}$ via the function χ_2. Vectors mirrored along the line $(\mathbb{X}_2)^c$ (in our case, (a, b) and (b, a)) are mapped to the same set with two elements. 26

3.2 Setup for generating a general RFS with real elements X. A sequence of random vectors $\mathbf{x}_k$ is converted to a sequence of random sets X_k. The discrete random variable n selects an element of this set sequence as the outcome of X. 36

4.1 Graphical summary of the various descriptors of RFSs and their interrelations. The dashed edges indicate a nonconstructive relation. 58

5.1 Example of two overlapping state spaces $R_1 = R_{1\setminus 2} \cup R_{12}$ associated with image s_1, and $R_2 = R_{12} \cup R_{2\setminus 1}$ associated with image s_2. 66

5.2 The relation between a grayscale image (solid grid) and the 2-dimensional Euclidean space. Shaded arrays represent the 4×4 arrays of pixels that are illuminated by objects located at the crosses. The dark shaded pixels are the ones that are part of the image. As can be seen, an object may influence less than 16 pixels. We require that the objects are located in the subspace $R \subseteq \mathbb{R}^2$, indicated by the dashed rectangle, so that at least one pixel is illuminated by every object. 71

5.3 Different image realzations and their corresponding cost functions $F(\boldsymbol{z}) = \log(\eta(\boldsymbol{z}, \boldsymbol{s}))$. The source intensities I of the point spread function are from top to bottom $I = 1, 30, 100$. Actual object positions are indicated by purple markers. 73

5.4 Typical image realizations of $\boldsymbol{s}_1$ (top left) and $\boldsymbol{s}_2$ (top right) for $I = 30$. The red corners indicate the shared region of both images. The corresponding cost functions $F(\boldsymbol{x}|\boldsymbol{s}_1)$ (utilizing just the single image $\boldsymbol{s}_1$) and $F(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ (utilizing both images) are shown in the middle and bottom part, respectively. The purple markers indicate actual object positions. 77

5.5 Monte Carlo averaged OSPA metric (top), cardinality error (middle), and localization error (bottom) for $p = 1$ and cutoff $c = 30$. 79

List of Abbreviations

BMF .... Belief Mass Function
CBMeMBer .. Cardinality Balanced Multi-Target Multi-Bernoulli
CD .... Cardinality Distribution
CPHD .... Cardinalized Probability Hypothesis Density
EVD .... Equivalent Vector Densities
EVF .... Equivalent Vector Function
FISST .... Finite Set Statistics
GM .... Gaussian Mixture
JPDA .... Joint Probabilistic Data Association
MeMBer ..... Multi-Target Multi-Bernoulli
MHT .... Multiple Hypothesis Tracking
OSPA .... Optimal Subpattern Assignment
PDF .... Probability Density Function
PGFL .... Probability Generating Functional
PHD .... Probability Hypothesis Density
PMF .... Probability Mass Function
RFS .... Random Finite Set
RN .... Radon-Nikodym
RND .... Radon-Nikodym Derivative
SMC .... Sequential Monte Carlo
SNR .... Signal-to-Noise Ratio
VST .... Vector-to-Set Transformation

Chapter 1

Introduction

1.1 Motivation

To illustrate the utility of random finite sets (RFSs), let us consider an intuitive application, namely tracking of multiple targets in a radar system. At each discrete time step, a sensor collects a sequence of noisy and possibly otherwise distorted measurements of the current target states (e.g., position, velocity, types, etc.). The task is to obtain an estimate of the target states based on all collected measurements up to the current time step. Given a suitable state space model, this problem could be solved by employing a standard Bayes filter or a variation thereof (see [1]). However, in real-world radar systems, the situation is more complicated since there are additional aspects that are not incorporated in this simple model [2]:

1. The number of targets is a priori unknown and random. Furthermore, the number of targets can (randomly) change between time steps since new targets may enter the observation area (target birth) and old targets may leave it (target death).

2. Similarly, the number of collected measurements is a priori unknown and random. A target present in the observation area does not necessarily generate a measurement (missed detection). Conversely, measurements may originate from the environment (clutter); such measurements are undesired since they do not convey any information about the targets. The central problem here is that, given a set of measurements, it is not clear whether an individual measurement is clutter or was generated by a target. Furthermore, even if this distinction could be made, it is not clear which measurement corresponds to which target. This problem, which is known as the *data association problem*, is of fundamental importance in target tracking applications.

Various classical approaches to solve this problem such as the joint probabilistic data association (JPDA) filter and multiple hypothesis tracking (MHT) can be found in [2]. A conceptually different approach was first proposed by Mahler in [3]. In this approach, target states and measurements are modeled as RFSs. An RFS is essentially a finite set consisting of a random number of random elements. Therefore, employing RFSs naturally captures the random number of targets and measurements. Moreover, since a finite set is an unordered collection of elements, modeling the target states and measurements as RFSs inherently implies the lack of an association between these quantities. The RFS approach to our target tracking example results in the formulation of an optimum *multi-object Bayes filter* [3], which can be thought of as the RFS counterpart of the conventional Bayes filter and additionally includes all the aspects discussed above.

1.2 State of the Art

The first appearance of RFSs can be traced back to [4], building upon the theory of point processes [5]. However, it was Mahler who popularized RFS theory in the context of multi-target tracking and data fusion problems with his original work in [3], introducing the multi-object Bayes filter. A direct implementation of the multi-object Bayes filter is in general infeasible and therefore a considerable amount of research has been carried out since then on finding tractable approximations. The first of these is the so-called *probability hypothesis density (PHD) filter* proposed in [6], which propagates only the first-order moment (the PHD) of the posterior RFS distribution and has been shown to outperform classical approaches in [7] and [8]. Sequential Monte Carlo (SMC) and Gaussian Mixture (GM) implementations of the PHD filter have been introduced in [9] and [10], and their convergence has been studied in [11–13]. An extension of the PHD filter, known as the *cardinalized PHD (CPHD) filter*, that propagates also the posterior cardinality distribution has been proposed in [14]. Closed-form solutions of the CPHD recursions have been established in [15]. Another approximation to the multi-object Bayes filter, called the *multi-target multi-Bernoulli (MeMBer) filter*, is presented in [16]. In contrast to the PHD and CPHD filters, which propagate first-order moments and cardinality distributions of the posterior distributions, the MeMBer filter approximates the posteriors as multi-Bernoulli distributions and propagates their parameters. An unbiased version of the MemBer filter, known as the *cardinality balanced MeMBer (CBMeMBer) filter*, has been introduced in [17]. An in-depth study of the CPHD and CBMeMBer filters is provided in [18].

RFS theory has been employed in a variety of scientific fields and in applications such as tracking in sonar images [19]; visual tracking and image processing [20–25]; audio processing [26]; multiuser detection and channel estimation in communications systems [27]; RFS-based simultaneous localization and mapping (SLAM) [28], [29]; tracking in indoor and urban environments [30–32]; pedestrian tracking [33].

Recently, the concept of *labeled RFSs* has attracted considerable interest. First established in [34], labeled RFSs are an extension of conventional RFSs that includes discrete tags. Based on this approach, labeled versions of the multi-Bernoulli filter that allow the output of target tracks have been proposed in [35–37]. Another recent trend is the distributed implementation of multi-target filters in sensor networks using various data fusion approaches [38–41].

1.3 Thesis Outline

One contribution of this thesis is to provide a systematic, detailed, and rigorous introduction to the theory of RFSs that can be followed easily by readers with no prior knowledge of the material. To this end, we begin in Chapter 2 by establishing the most important measure-theoretic concepts of probability theory that will be required later on. Based on this knowledge, in Chapter 3 we rigorously define RFSs in the measure-theoretic framework of probability theory. Furthermore, we introduce various descriptions of RFSs such as the cardinality distribution, belief mass function, and probability density function for RFSs. We discuss the relation between finite sets and vectors, and present an intuitive interpretation of RFSs in terms of a sequence of random vectors. Concluding this chapter, we extend our results to accommodate multiple RFSs and introduce joint and conditional versions of the descriptors used for single RFSs, as well as the expectation operator for functions of RFSs.

Chapter 4 discusses the basic concepts of the FISST framework [16] such as set integrals, probability generating functionals, functional derivatives, and the PHD. At the end of this chapter, four common types of RFSs are presented.

In Chapter 5, we apply the theory of RFSs to the problem of estimating the states of an unknown and random number of objects from image observations. As a reference, we begin by analyzing the situation where only one image observation is available and derive an estimator based on the posterior PHD. Next, we investigate a scenario where two independent sensors acquire partly overlapping images of a bigger scene and show how by exchanging information between the sensors the estimation performance can be improved over the case where each sensor calculates the estimates separately. Furthermore, we propose a novel RFS-based algorithm solving this latter estimation problem and demonstrate its performance in simulated scenarios. Chapter 6 summarizes the obtained results and suggests future directions of research.

Chapter 2

Fundamentals of Measure Theoretic Probability Theory

This chapter gives a short introduction to the basic concepts of measure theoretic probability theory, which will be needed later on for the treatment of random finite sets. An excellent treatment of this subject, which is also accessible to non-mathematicians, is [42]. Further good texts amongst others are [43–45], although these require a more advanced mathematical background.

In Sections 2.1 to 2.5, we define the basic mathematical objects considered in measure theory and summarize their properties, culminating in the definition of probability spaces and general random elements in Section 2.6. As we will see in Chapter 3, a random finite set is just a special case of such a general random element. Section 2.7 establishes the Lebesgue integral and its main properties, which will later on allow us to define probability density functions for random finite sets. Throughout this chapter, we will also explain how the generally abstract definitions and concepts translate to the familiar case of real random variables or random vectors.

2.1 σ-Algebras

Definition 2.1: *Let Ω be a set. The **power set** of Ω is the collection of all subsets of Ω, denoted by $\mathcal{P}(\Omega) \triangleq \{A : A \subseteq \Omega\}$. A **$\sigma$-algebra** Σ_Ω on Ω is a collection of subsets of Ω, i.e., $\Sigma_\Omega \subseteq \mathcal{P}(\Omega)$, that satisfies the following conditions:*

(a) $\Omega \in \Sigma_\Omega$.

(b) If $A \in \Sigma_\Omega$, then $A^c \in \Sigma_\Omega$.

(c) For every countable collection $\{A_i\}$ with $A_i \in \Sigma_\Omega$, $i \in \mathbb{N}$,

$$\bigcup_{i=1}^{\infty} A_i \in \Sigma_\Omega.$$

The elements of the σ-algebra are called ***measurable sets*** *and the pair* (Ω, Σ_Ω) *is called a* ***measurable space****.*

Note that due to (a) and (b), Ω and the empty set $\emptyset = \Omega^c$ are always contained in any σ-algebra. Indeed, the smallest σ-algebra that can be constructed is $\Sigma_\Omega = \{\Omega, \emptyset\}$, which is also called the **trivial σ-algebra**. Because of (c), Σ_Ω is closed under countable union. Applying De Morgan's law $\bigcap_{i \in I} A_i^c = (\bigcup_{i \in I} A_i)^c$, where I is any index set, it follows from (b) and (c) that Σ_Ω is also closed under countable intersection, i.e.,

$$A_i \in \Sigma_\Omega \quad \Rightarrow \quad \bigcap_{i=1}^{\infty} A_i \in \Sigma_\Omega. \tag{2.1}$$

Another example of a σ-algebra is the power set $\mathcal{P}(\Omega)$ itself. This specific σ-algebra is also called the **discrete σ-algebra**, and can be considered the most inclusive of all σ-algebras.

In a probability model, we assign probabilities to the elements of a σ-algebra (the events). At this point, one could ask why we should bother about σ-algebras, instead of just always using the power set. The reason is that, whenever Ω contains infinitely many elements, we cannot consistently assign probabilities to each element of $\mathcal{P}(\Omega)$ (e.g., think of $\mathcal{P}(\mathbb{R})$ or $\mathcal{P}(\mathbb{R}^n)$). To remedy this, we restrict ourselves to a σ-algebra, which is small enough to be tractable and at the same time big enough to contain events of interest. The next definition is key for actually constructing σ-algebras.

Definition 2.2: *Let Ω be a set and $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ an arbitrary collection of subsets of Ω. The σ-algebra on Ω* ***generated*** *by $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$, is the minimal σ-algebra that contains all elements of $\mathcal{C}$. That is,*

(a) $\mathcal{C} \subseteq \sigma(\mathcal{C})$*;*

(b) $\sigma(\mathcal{C}) \subseteq \Sigma_\Omega$ *for every σ-algebra Σ_Ω with $\mathcal{C} \subseteq \Sigma_\Omega$.*

Note that $\sigma(\mathcal{C})$ is well-defined (i.e., unique for a given $\mathcal{C}$) [42, p. 16]. Usually, one starts with the specification of a collection $\mathcal{C}$ and then works with $\sigma(\mathcal{C})$, thereby insisting on the ability to perform any standard set operation (complement, countable union, countable intersection) on elements of $\mathcal{C}$ while still staying in $\sigma(\mathcal{C})$.

2.2 Topologies and Borel Algebras

Before introducing the most important σ-algebra, the Borel algebra, we need to define open sets [46, pp. 116–117]:

Definition 2.3: *Let Ω be a set, and $\mathcal{T} \subseteq \mathcal{P}(\Omega)$ a collection of subsets of Ω with the following properties:*

(a) $\emptyset \in \mathcal{T}$ *and* $\Omega \in \mathcal{T}$.

(b) If $\mathcal{C} \subseteq \mathcal{T}$, *then*

$$\bigcup_{A \in \mathcal{C}} A \in \mathcal{T}.$$

(c) For any finite collection $\{A_1, \ldots, A_n\} \subseteq \mathcal{T}$ *with* $A_i \in \mathcal{T}$ *and* $n \in \mathbb{N}$,

$$\bigcap_{i=1}^{n} A_i \in \mathcal{T}.$$

Then the collection $\mathcal{T}$ is called a ***topology*** *on Ω and the pair $(\Omega, \mathcal{T})$ a* ***topological space****. Furthermore, the elements A of $\mathcal{T}$ are called* ***open sets*** *and their complements A^c* ***closed sets****.*

Similarly to σ-algebras, there are many topologies that can be defined for a given set Ω. For instance, the **trivial topology** is $\mathcal{T} = \{\emptyset, \Omega\}$ and the **discrete topology** is $\mathcal{T} = \mathcal{P}(\Omega)$.

In the important case of the Euclidean space $\Omega = \mathbb{R}^d$, open sets are defined as follows [47, Chapter 2]: $A \subseteq \mathbb{R}^d$ is open if for every $\boldsymbol{x} \in A$ there exists a neighborhood of $\boldsymbol{x}$

$$N_r(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{y}\| < r\}, \tag{2.2}$$

with some $r > 0$, such that $N_r(\boldsymbol{x}) \subseteq A$. The collection $\mathcal{T} = \mathcal{T}_{\mathbb{R}^d}$ of all those open sets A is then a topology as in Definition 2.3, and is called the **standard topology** on $\mathbb{R}^d$. In the standard topology, the closed sets in $\mathbb{R}^d$ are those that contain all their limit points, where $\boldsymbol{x}$ is a limit point of a set $A \subseteq \mathbb{R}^d$ if in every neighborhood of $\boldsymbol{x}$ there is another point $\boldsymbol{y} \neq \boldsymbol{x}$ with $\boldsymbol{y} \in A$. For instance, in the case of $\Omega = \mathbb{R}$, examples of open sets are $(-\infty, a)$, (a, b), $(a, +\infty)$; and examples of closed sets are $(-\infty, a]$, $[a, b]$, $[a, +\infty)$ and finite sets $\{a_1, \ldots, a_n\}$.

Definition 2.4: *Let $(\Omega, \mathcal{T})$ be a topological space and $\mathcal{D} \subseteq \mathcal{T}$ a collection of open sets. If every $T \in \mathcal{T}$ can be represented as*

$$T = \bigcup_{i \in I} D_i,$$

for some index set I and $D_i \in \mathcal{D}$, then $\mathcal{D}$ is called a ***base*** *for the topology $\mathcal{T}$. We also say that $\mathcal{T}$ is* ***generated*** *by $\mathcal{D}$.*

This notion is similar to a σ-algebra generated by a collection as in Definition 2.2. A base for the standard topology on $\mathbb{R}^d$ is the collection of all neighborhoods as in (2.2).

Definition 2.5: *Let $(\Omega, \mathcal{T})$ be a topological space. The σ-algebra generated by the topology $\mathcal{T}$ is called the* ***Borel algebra*** *on $(\Omega, \mathcal{T})$, denoted by*

$$\mathcal{B}(\Omega, \mathcal{T}) \triangleq \sigma(\mathcal{T}).$$

That is, the Borel algebra of a topological space is the smallest σ-algebra that contains all open sets. Therefore, by Definition 2.1 and the fact that the complement of an open set is a closed set, the Borel algebra additionally contains all closed sets, and all countable unions and intersections of open or closed sets. Talking about the Borel algebra of the Euclidean space $\Omega = \mathbb{R}^d$, it is always tacitly assumed that the standard topology $\mathcal{T} = \mathcal{T}_{\mathbb{R}^d}$ is used. We will thus briefly write

$$\mathcal{B}(\mathbb{R}^d) \triangleq \mathcal{B}(\mathbb{R}^d, \mathcal{T}_{\mathbb{R}^d}). \tag{2.3}$$

It is difficult to find subsets of $\mathbb{R}$ (or $\mathbb{R}^d$) that are not elements of the Borel algebra, but such subsets do exist (e.g., Vitali sets).

An important fact in the case of $\mathbb{R}$ is that $\mathcal{B}(\mathbb{R})$ can be generated in a multitude of ways. For instance, let $\mathcal{I}_{(a,b)}$ denote the collection of all open intervals, i.e.,

$$\mathcal{I}_{(a,b)} \triangleq \{(a, b) : a, b \in \mathbb{R}, a \leq b\}. \tag{2.4}$$

Likewise denote the collection of other types of intervals $\mathcal{I}_{(a,b]}$, $\mathcal{I}_{[a,b]}$, $\mathcal{I}_{(-\infty,a)}$, etc. Then it can be shown [42, pp. 16–18] that $\mathcal{B}(\mathbb{R})$ can be generated by any kind of intervals, e.g.,

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{T}_{\mathbb{R}}) = \sigma(\mathcal{I}_{(a,b)}) = \sigma(\mathcal{I}_{(a,b]}) = \sigma(\mathcal{I}_{[a,b]}) = \sigma(\mathcal{I}_{(-\infty,a)}). \tag{2.5}$$

2.3 Product σ-Algebras

The following definition allows us to combine measurable spaces [42, pp. 143–145]:

Definition 2.6: *Let $(\Omega_1, \Sigma_{\Omega_1})$ and $(\Omega_2, \Sigma_{\Omega_2})$ be two measurable spaces. Let*

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

be the Cartesian product of Ω_1 and Ω_2. Furthermore, let

$$\Sigma_{\Omega_1} \otimes \Sigma_{\Omega_2} \triangleq \sigma(\{A_1 \times A_2 : A_1 \in \Sigma_{\Omega_1}, A_2 \in \Sigma_{\Omega_2}\}).$$

The measurable space $(\Omega_1 \times \Omega_2, \Sigma_{\Omega_1} \otimes \Sigma_{\Omega_2})$ is called the ***product space****, and $\Sigma_{\Omega_1} \otimes \Sigma_{\Omega_2}$ the* ***product σ-algebra****.*

This definition is extended straightforwardly to the product of more than two measurable spaces: if $(\Omega_1, \Sigma_{\Omega_1}), \dots, (\Omega_n, \Sigma_{\Omega_n})$ are measurable spaces, the product σ-algebra on $\Omega_1 \times \cdots \times \Omega_n$ is

$$\Sigma_{\Omega_1} \otimes \cdots \otimes \Sigma_{\Omega_n} \triangleq \sigma(\{A_1 \times \cdots \times A_n : A_i \in \Sigma_{\Omega_i}, i = 1, \dots, n\}). \tag{2.6}$$

For $\mathbb{R}^d$, the product of d copies of the Borel algebra $\mathcal{B}(\mathbb{R})$ is equal to the Borel algebra $\mathcal{B}(\mathbb{R}^d)$, i.e.,

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R}). \tag{2.7}$$

Specifically, note that all d-dimensional intervals, their complements, countable unions, and countable intersections are contained in $\mathcal{B}(\mathbb{R}^d)$.

2.4 Measure Spaces

Definition 2.7: *Let (Ω, Σ_Ω) be a measurable space. A function $\mu\colon \Sigma_\Omega \to [0, +\infty]$ is called a **measure** on Ω if the following holds:*

(a) $\mu(\emptyset) = 0$.

(b) For every countable collection $\{A_i, i \in \mathbb{N}\} \subseteq \Sigma_\Omega$ of pairwise disjoint sets ($A_i \cap A_j = \emptyset$ whenever $i \neq j$):

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

*The triple $(\Omega, \Sigma_\Omega, \mu)$ is called a **measure space** and the elements of Σ_Ω are called μ-**measurable** sets.*

*A measure μ is called **finite** if $\mu(\Omega) < \infty$.*

*If there exists a countable collection of measurable sets $A_i \in \Sigma_\Omega$ such that $\bigcup_{i=1}^{\infty} A_i = \Omega$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$, then the measure μ is called σ-**finite**.*

The basic idea behind a measure is to establish a notion of "size" (length, area, volume, ...) associated with the elements of a σ-algebra. Two simple consequences of Definition 2.7 are [42, p. 31]:

1. If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

2. For every countable collection of measurable sets $A_i \in \Sigma_\Omega$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i). \tag{2.8}$$

Definition 2.8: *Let $(\Omega, \Sigma_\Omega, \mu)$ be a measure space. A set $A \in \Sigma_\Omega$ for which $\mu(A) = 0$ is called a **null set**. A statement $S(\omega)$, $\omega \in \Omega$, holds **almost everywhere** (abbreviated a.e.) if the set of elements ω of Ω for which it does not hold is a null set.*

Often this notion is encountered in the context of equality of two functions. For instance, let $(\Omega_1, \Sigma_{\Omega_1}, \mu)$ be a measure space, and let $f(\omega_1)$ and $g(\omega_1)$ be two functions with the same domain Ω_1 and codomain Ω_2. Then

$$f(\omega_1) = g(\omega_1) \quad \text{a.e.}$$

means that $f(\omega_1) = g(\omega_1)$ for all $\omega_1 \in \Omega_1$ except on a null set.

One fundamental measure that can be defined on any measurable space is the **counting measure** $\mu_C(A) = |A|$, where $|A|$ denotes the cardinality of A. More precisely, the counting measure assigns to each finite set $A = \{a_1, \ldots, a_n\}$ the number of elements in A, i.e., $\mu_C(A) = n$. If A is (countably or uncountably) infinite, $\mu_C(A) = +\infty$.

An important measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the Lebesgue measure $\lambda(A)$ constructed as follows [47, pp. 302–310]. The **length** of a closed interval $I = [a, b]$, with $a \leq b$, is defined as

$$l(I) \triangleq b - a. \tag{2.9}$$

Let $A \in \mathcal{B}(\mathbb{R})$, and let $\mathcal{C} \subseteq \mathcal{P}(\mathbb{R})$ be a countable collection of closed intervals. We say that $\mathcal{C}$ covers A if

$$A \subseteq \bigcup_{I \in \mathcal{C}} I. \tag{2.10}$$

The **Lebesgue measure** $\lambda(A)$ is defined by

$$\lambda(A) \triangleq \inf_{\mathcal{C}} \left(\sum_{I \in \mathcal{C}} l(I) \right), \tag{2.11}$$

where the infimum is taken over all collections $\mathcal{C}$ that cover A.

Intuitively the Lebesgue measure is the total length of the collection of intervals that cover the set A most tightly. For instance, sets containing only one point of $\mathbb{R}$ have Lebesgue measure zero (i.e., they are null sets). It follows that all countable subsets of $\mathbb{R}$ are null sets as well. This specifically means that adding or subtracting countably many points to/from a given set does not change its Lebesgue measure. The intervals (a, b), $(a, b]$, $[a, b)$, and $[a, b]$ all have Lebesgue measure $b - a$. Infinite intervals (as well as $\mathbb{R}$) have Lebesgue measure ∞. Also note that every set in $\mathcal{B}(\mathbb{R})$ is λ-measurable.[1]

[1] The converse is not true. That is, there are Lebesgue measurable sets that are not contained in the Borel algebra.

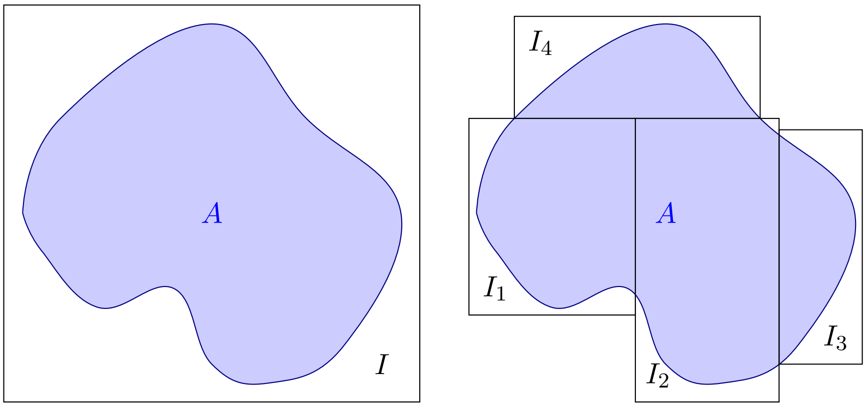

Fig. 2.1: *Illustration of the Lebesgue measure λ_2 on $\mathbb{R}^2$. The left part shows a simple open cover of the set A consisting of only one two-dimensional (2D) interval I. The right part shows a tighter open cover consisting of four 2D intervals $I_1, \ldots, I_4$, which has a smaller total area than I. One can repeat this refinement by searching for another, tighter open cover of A, containing more 2D intervals. The Lebesgue measure can be thought of as the area of the open cover in the limit of this refinement process.*

The definition of the Lebesgue measure can be extended to the Euclidean space $\mathbb{R}^d$, if one replaces the closed intervals with d-dimensional closed intervals of the form

$$I \triangleq [a_1, b_1] \times \cdots \times [a_d, b_d], \tag{2.12}$$

and the length with the hypervolume

$$v(I) \triangleq \prod_{i=1}^{d} (b_i - a_i). \tag{2.13}$$

In the d-dimensional case, the definition of the Lebesgue measure (cf. (2.11)) becomes

$$\lambda_d(A) \triangleq \inf_{\mathcal{C}} \left(\sum_{I \in \mathcal{C}} v(I) \right), \tag{2.14}$$

where the infimum is taken over all collections $\mathcal{C}$ that cover A. Figure 2.1 shows an example of this concept for $d = 2$.

The next theorem presents a construction of a measure on the product space of two measure spaces [43, p. 232].

Theorem 2.9: Let $(\Omega_1, \Sigma_{\Omega_1}, \mu_1)$ and $(\Omega_2, \Sigma_{\Omega_2}, \mu_2)$ be two measure spaces, with μ_1 and μ_2 σ-finite. Then there exists a unique measure $\mu_1 \times \mu_2 \colon \Sigma_{\Omega_1} \otimes \Sigma_{\Omega_2} \to [0, +\infty]$ on the product space $(\Omega_1 \times \Omega_2, \Sigma_{\Omega_1} \otimes \Sigma_{\Omega_2})$ that satisfies

$$\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

for all $A_1 \in \Sigma_{\Omega_1}$, $A_2 \in \Sigma_{\Omega_2}$. The measure $\mu_1 \times \mu_2$ is called the **product measure**.

Analogous to (2.6), this can be extended to more than two measure spaces. Specifically, the product Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, given by

$$\lambda_d(A_1 \times \cdots \times A_d) = \prod_{i=1}^{d} \lambda(A_i), \tag{2.15}$$

is consistent with the original definition in (2.14).

2.5 Measurable Functions

Definition 2.10: *Let Ω and $\mathcal{X}$ be two sets and let $f\colon \Omega \to \mathcal{X}$ be a function. The* ***inverse image*** *of $A \subseteq \mathcal{X}$ under f is defined as*

$$f^{-1}(A) \triangleq \{\omega \in \Omega : f(\omega) \in A\}.$$

Note that $f^{-1}\colon \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\Omega)$. The inverse image f^{-1} has the following properties [42, pp. 71–72]:

1. $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(\mathcal{X}) = \Omega$.

2. For all $A \in \mathcal{P}(\mathcal{X})$,
$$f^{-1}(A^c) = (f^{-1}(A))^c. \tag{2.16}$$

3. For $A_i \in \mathcal{P}(\mathcal{X})$, $i \in I$, where I is an arbitrary index set,
$$f^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f^{-1}(A_i), \tag{2.17}$$
$$f^{-1}\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f^{-1}(A_i). \tag{2.18}$$

For $\mathcal{C} \subseteq \mathcal{P}(\mathcal{X})$ a collection of subsets of $\mathcal{X}$, we define

$$f^{-1}(\mathcal{C}) \triangleq \{f^{-1}(A) : A \in \mathcal{C}\}. \tag{2.19}$$

The following important fact can be shown [42, p. 73]: If $\Sigma_{\mathcal{X}}$ is a σ-algebra on $\mathcal{X}$, then $f^{-1}(\Sigma_{\mathcal{X}})$ is a σ-algebra on Ω.

Definition 2.11: *Let (Ω, Σ_Ω) and $(\mathcal{X}, \Sigma_{\mathcal{X}})$ be two measurable spaces. A function $f\colon \Omega \to \mathcal{X}$ is called $\Sigma_\Omega/\Sigma_{\mathcal{X}}$-**measurable** (or briefly measurable) if*

$$f^{-1}(\Sigma_{\mathcal{X}}) \subseteq \Sigma_\Omega.$$

The concept of a measurable function is important because it makes sure that the inverse image of any set $A \in \Sigma_{\mathcal{X}}$ under a measurable function is contained in Σ_Ω. In other words, with the inverse image operation one never leaves Σ_Ω.

Definition 2.12: *Let $(\Omega, \Sigma_\Omega, \mu)$ be a measure space, $(\mathcal{X}, \Sigma_{\mathcal{X}})$ a measurable space, and $f : \Omega \to \mathcal{X}$ a measurable function. The measure μ_f on $\Sigma_{\mathcal{X}}$ defined by*

$$\mu_f(A) = \mu(f^{-1}(A)),$$

*for all $A \in \Sigma_{\mathcal{X}}$, is called the **push-forward measure** of μ. We also say μ_f is **induced** by f.*

2.6 Probability Spaces and Random Elements

Definition 2.13: *A **probability space** is a measure space $(\Omega, \Sigma_\Omega, P)$ with $P(\Omega) = 1$.*

*The set Ω that contains all the possible outcomes of a random experiment is called the **sample space**. The elements of Σ_Ω are called **events** and the measure P is called the **probability measure** (also probability distribution).*

The following are some simple properties of a probability measure P [42, pp. 29–31].

1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
2. For every $A \in \Sigma_\Omega$,
$$P(A) \in [0, 1]. \tag{2.20}$$
3. For every $A \in \Sigma_\Omega$,
$$P(A^c) = 1 - P(A). \tag{2.21}$$
4. For every $A, B \in \Sigma_\Omega$,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{2.22}$$
5. For every $A, B \in \Sigma_\Omega$ with $A \subseteq B$,
$$P(A) \leq P(B). \tag{2.23}$$
6. For every countable collection of events $A_i \in \Sigma_\Omega$,
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i). \tag{2.24}$$

Definition 2.14: *Let $(\Omega, \Sigma_\Omega, P)$ be a probability space, $(\mathcal{X}, \Sigma_\mathcal{X})$ a measurable space, and $\mathfrak{X}\colon \Omega \to \mathcal{X}$ a measurable function. Then $\mathfrak{X}$ is called a* ***random element*** *of $\mathcal{X}$.*

Since $\mathfrak{X}$ is measurable, it induces a probability measure $P_\mathfrak{X}\colon \Sigma_\mathcal{X} \to [0,1]$ on $(\mathcal{X}, \Sigma_\mathcal{X})$ by (cf. Definition 2.12)

$$P_\mathfrak{X}(A) = P(\mathfrak{X}^{-1}(A)), \tag{2.25}$$

for all $A \in \Sigma_\mathcal{X}$. Consequently, the triple $(\mathcal{X}, \Sigma_\mathcal{X}, P_\mathfrak{X})$ is a probability space.

At this point we make some notational conventions:

- If $(\mathcal{X}, \Sigma_\mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we write $\mathfrak{X} = \mathsf{x}$ and call $\mathsf{x}\colon \Omega \to \mathbb{R}$ a **real random variable**. Real random variables will always be denoted by sans-serif type, such as $\mathsf{x}, \mathsf{y}, \mathsf{z}$.
- If $(\mathcal{X}, \Sigma_\mathcal{X}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we write $\mathfrak{X} = \boldsymbol{\mathsf{x}}$ and call $\boldsymbol{\mathsf{x}}\colon \Omega \to \mathbb{R}^d$ a **real random vector**. Real random vectors will always be denoted by bold sans-serif type, such as $\boldsymbol{\mathsf{x}}, \boldsymbol{\mathsf{y}}, \boldsymbol{\mathsf{z}}$.
- In Chapter 3, we take define $\mathcal{X}$ to be the collection of all finite subsets of $\mathbb{R}^d$ and introduce a Borel algebra $\mathcal{B}(\mathcal{X})$ on $\mathcal{X}$. In this case, we will write $\mathfrak{X} = \mathsf{X}$ and call $\mathsf{X}\colon \Omega \to \mathcal{X}$ a **random finite set**. Random finite sets will always be denoted by capital sans-serif type, such as $\mathsf{X}, \mathsf{Y}, \mathsf{Z}$.

2.7 Integration

In this section, we define a generalized integral, the so-called Lebesgue integral, which allows us to integrate real-valued functions defined on measure spaces.

2.7.1 Definition of the Lebesgue Integral

Let $(\mathcal{X}, \Sigma_\mathcal{X}, \mu)$ be a measure space and let $f\colon \mathcal{X} \to \mathbb{R}$ be a measurable function with finite range, i.e.,

$$f(\mathcal{X}) = \{a_1, \ldots, a_k\}, \tag{2.26}$$

for some $k \in \mathbb{N}$ and $a_1, \ldots, a_k \in \mathbb{R}$. Such a function is called **simple** and can always be written as

$$f(x) = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}(x), \tag{2.27}$$

where $A_i \in \Sigma_\mathcal{X}$ are disjoint subsets of $\mathcal{X}$ with $\bigcup_{i=1}^k A_i = \mathcal{X}$, and $\mathbf{1}_{A_i}(x)$ is the indicator function of A_i defined by

$$\mathbf{1}_{A_i}(x) \triangleq \begin{cases} 1 & \text{if } x \in A_i, \\ 0 & \text{if } x \notin A_i. \end{cases} \tag{2.28}$$

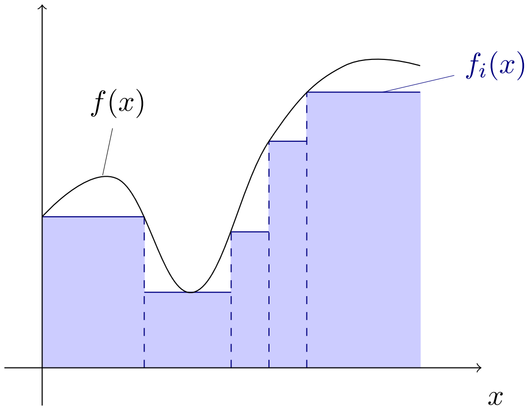

Fig. 2.2: *The function $f(x)$ is approximated by a simple function $f_i(x)$, for which the Lebesgue integral with respect to λ can be calculated. One can find another simple function with more steps that fits the curve $f(x)$ more tightly. The Lebesgue integral of $f(x)$ with respect to λ is the area in the limit of this refinement process.*

The **Lebesgue integral** of a simple function $f(x)$ with respect to the measure μ is defined as[2]

$$\int f(x)\mathrm{d}\mu(x) \triangleq \sum_{i=1}^{k} a_i \mu(A_i). \tag{2.29}$$

It can be shown [42, p. 118] that a general nonnegative function $f\colon \mathcal{X} \to \mathbb{R}$ is $\Sigma_{\mathcal{X}}/\mathcal{B}(\mathbb{R})$-measurable if and only if there exists an increasing[3] sequence of nonnegative simple functions $f_i(x)$, $i = 1, 2, \dots$ such that

$$f(x) = \lim_{i\to\infty} f_i(x) \quad \text{a.e.} \tag{2.30}$$

The sequence f_i is called an approximating sequence.

This result allows us to extend the definition of the Lebesgue integral to all nonnegative measurable functions $f(x) \geq 0$ as

$$\int f(x)\mathrm{d}\mu(x) \triangleq \lim_{i\to\infty} \int f_i(x)\mathrm{d}\mu(x), \tag{2.31}$$

where f_i is an approximating sequence of simple functions as in (2.30) (see Figure 2.2 for an illustration). Finally, we extend this integral again to all measurable functions $f(x)$ by first

[2]Other common notations for the Lebesgue integral are $\int f(x)\mu(\mathrm{d}x)$ or briefly $\int f \mathrm{d}\mu$.
[3]An increasing sequence of functions means that $f_i(x) \leq f_j(x) \leq f(x)$ for all $x \in \mathcal{X}$ if $i \leq j$.

defining

$$f^+(x) \triangleq \begin{cases} f(x) & \text{if } f(x) \geq 0, \\ 0 & \text{if } f(x) < 0; \end{cases} \tag{2.32}$$

$$f^-(x) \triangleq \begin{cases} -f(x) & \text{if } f(x) \leq 0, \\ 0 & \text{if } f(x) > 0. \end{cases} \tag{2.33}$$

Definition 2.15: *Let $(\mathcal{X}, \Sigma_{\mathcal{X}}, \mu)$ be a measure space and $f\colon \mathcal{X} \to \mathbb{R}$ a measurable function. If $\int f^+(x)\mathrm{d}\mu(x) < \infty$ or $\int f^-(x)\mathrm{d}\mu(x) < \infty$, then the **Lebesgue integral** of $f(x)$ with respect to μ is defined as*

$$\int f(x)\mathrm{d}\mu(x) \triangleq \int f^+(x)\mathrm{d}\mu(x) - \int f^-(x)\mathrm{d}\mu(x). \tag{2.34}$$

*The measurable function $f(x)$ is called μ-**integrable** if $\int f(x)\mathrm{d}\mu(x) \neq \pm\infty$.*

Often we will be interested in integrals of functions with domain $\mathbb{R}$, with respect to the Lebesgue measure λ as defined in (2.11). One important fact is the following [43, pp. 221–222]: if $f(x)$ is Riemann integrable, then $f(x)$ is Lebesgue integrable (with respect to λ), and

$$\int f(x)\mathrm{d}x = \int f(x)\mathrm{d}\lambda(x). \tag{2.35}$$

Note that the converse is not true. In this sense, the Lebesgue integral is an extension of the familiar Riemann integral. A similar statement holds for multivariate functions with domain $\mathbb{R}^d$ and the Lebesgue measure λ_d.

The Lebesgue integral of $f(x)$ with respect to μ over a measurable set $A \in \Sigma_{\mathcal{X}}$ is defined by

$$\int_A f(x)\mathrm{d}\mu(x) \triangleq \int \mathbf{1}_A(x) f(x)\mathrm{d}\mu(x). \tag{2.36}$$

Note that the Lebesgue integral $\int f(x)\mathrm{d}\mu(x)$ as defined in (2.34) is equal to $\int_{\mathcal{X}} f(x)\mathrm{d}\mu(x)$. Furthermore, we have

$$\mu(A) = \int_A \mathrm{d}\mu(x), \tag{2.37}$$

for all $A \in \Sigma_{\mathcal{X}}$. For a general random element $\mathfrak{X}$ with associated probability space $(\mathcal{X}, \Sigma_{\mathcal{X}}, P_{\mathfrak{X}})$ this becomes

$$P_{\mathfrak{X}}(A) = \int_A \mathrm{d}P_{\mathfrak{X}}(x), \tag{2.38}$$

for all $A \in \Sigma_{\mathcal{X}}$. This is the probability of the event A, i.e., $\Pr(\mathfrak{X}(\omega) \in A)$. Moreover, for a measurable real-valued function $g\colon \mathcal{X} \to \mathbb{R}$, the **expectation operator** is defined as

$$\mathsf{E}[g(\mathfrak{X})] \triangleq \int g(x)\mathrm{d}P_{\mathfrak{X}}(x). \tag{2.39}$$

Note that in the special case of a random variable x, this yields the k-th moments $\mathsf{E}[\mathsf{x}^k]$ if $g(x) = x^k$, and the k-th central moments $\mathsf{E}[(\mathsf{x} - \mathsf{E}[\mathsf{x}])^k]$ if $g(x) = (x - \mathsf{E}(\mathsf{x}))^k$.

2.7.2 Properties of the Lebesgue Integral

In this section, we present some important properties of the Lebesgue integral. Let $f, g\colon \mathcal{X} \to \mathbb{R}$ be two integrable functions and $a, b \in \mathbb{R}$, then [46, pp. 126–128]

$$\text{(Linearity)} \quad \int (af(x) + bg(x))\mathrm{d}\mu(x) = a\int f(x)\mathrm{d}\mu(x) + b\int g(x)\mathrm{d}\mu(x), \tag{2.40}$$

$$\text{(Modulus inequality)} \quad \left|\int f(x)\mathrm{d}\mu(x)\right| \leq \int |f(x)|\mathrm{d}\mu(x). \tag{2.41}$$

If $f(x) \leq g(x)$ a.e., then

$$\int f(x)\mathrm{d}\mu(x) \leq \int g(x)\mathrm{d}\mu(x). \tag{2.42}$$

The following theorem states the essential property of the push-forward measure (which we will use extensively in Section 3.5 in developing a measure for the set space) [43, pp. 215–216]:

Theorem 2.16: Let $(\Omega, \Sigma_\Omega, \mu)$ be a measure space, $(\mathcal{X}, \Sigma_\mathcal{X})$ a measurable space, and $f\colon \Omega \to \mathcal{X}$ a measurable function. Furthermore, let $(\mathcal{X}, \Sigma_\mathcal{X})$ be equipped with the push-forward measure μ_f of μ (cf. Definition 2.12). Then, a function $g : \mathcal{X} \to \mathbb{R}$ is integrable with respect to μ_f if and only if the composition $g \circ f$ is integrable with respect to μ. In that case

$$\int_A g(x)\mathrm{d}\mu_f(x) = \int_{f^{-1}(A)} g(f(\omega))\mathrm{d}\mu(\omega), \tag{2.43}$$

for all $A \in \Sigma_\mathcal{X}$.

Definition 2.17: *Let $(\mathcal{X}, \Sigma_\mathcal{X})$ be a measurable space and let μ_1, μ_2 be two measures on $(\mathcal{X}, \Sigma_\mathcal{X})$. Then μ_1 is **absolutely continuous** with respect to μ_2, denoted by $\mu_1 \ll \mu_2$, if*

$$\mu_2(A) = 0 \quad \Rightarrow \quad \mu_1(A) = 0,$$

*for all $A \in \Sigma_\mathcal{X}$. One also says μ_1 is **dominated** by μ_2, or simply that μ_2 is the **dominant** measure.*

The next theorem will be frequently invoked in Chapter 3 to establish probability density functions for random finite sets (for a proof see [42, Section 10.1]).

Theorem 2.18 (Radon-Nikodym (RN)): Let $(\mathcal{X}, \Sigma_\mathcal{X})$ be a measurable space, μ_1 and μ_2 two σ-finite measures on $(\mathcal{X}, \Sigma_\mathcal{X})$, and $\mu_1 \ll \mu_2$. Then there exists a (a.e.) unique measurable function $f\colon \mathcal{X} \to [0, \infty)$ such that

$$\mu_1(A) = \int_A f(x)\mathrm{d}\mu_2(x),$$

for all $A \in \Sigma_{\mathcal{X}}$. This function is denoted by

$$f(x) = \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}(x),$$

and called the **Radon-Nikodym derivative (RND)** of μ_1 with respect to μ_2.

Applied to a random element, the RN theorem states the following. Consider a random element $\mathfrak{X}$ with its associated probability space $(\mathcal{X}, \Sigma_{\mathcal{X}}, P_{\mathfrak{X}})$, and a second measure μ on $\Sigma_{\mathcal{X}}$ with $P_{\mathfrak{X}} \ll \mu$. Then there exists a unique measurable function $f_{\mathfrak{X}} \colon \mathcal{X} \to [0, \infty)$ such that

$$P_{\mathfrak{X}}(A) = \int_A f_{\mathfrak{X}}(x)\mathrm{d}\mu(x), \tag{2.44}$$

for all $A \in \Sigma_{\mathcal{X}}$. That is, the probability measure $P_{\mathfrak{X}}$ is completely characterized by the nonnegative function $f_{\mathfrak{X}}$ defined on $\mathcal{X}$. This function is the **probability density function (PDF)** of the random element $\mathfrak{X}$. Note that

$$\int f_{\mathfrak{X}}(x)\mathrm{d}\mu(x) = \int_{\mathcal{X}} f_{\mathfrak{X}}(x)\mathrm{d}\mu(x) = P_{\mathfrak{X}}(\mathcal{X}) = 1. \tag{2.45}$$

Special case: random variable. If we have a random variable x with associated probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\mathsf{x}})$ and dominant Lebesgue measure $\mu = \lambda$, i.e., $P_{\mathsf{x}} \ll \lambda$, then (2.44) becomes the familiar equation

$$P_{\mathsf{x}}(A) = \int_A f_{\mathsf{x}}(x)\mathrm{d}\lambda(x), \tag{2.46}$$

for all $A \in \mathcal{B}(\mathbb{R})$. Here, $f_{\mathsf{x}} \colon \mathbb{R} \to [0, \infty)$ is the RND

$$f_{\mathsf{x}}(x) = \frac{\mathrm{d}P_{\mathsf{x}}}{\mathrm{d}\lambda}(x). \tag{2.47}$$

Special case: random vector. If we have a d-dimensional random vector $\boldsymbol{\mathsf{x}} = (\mathsf{x}_1, \ldots, \mathsf{x}_d)$ with associated probability space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_{\boldsymbol{\mathsf{x}}})$ and dominant Lebesgue measure $\mu = \lambda_d$, i.e., $P_{\boldsymbol{\mathsf{x}}} \ll \lambda_d$, then (2.44) becomes

$$P_{\boldsymbol{\mathsf{x}}}(A) = \int_A f_{\boldsymbol{\mathsf{x}}}(\boldsymbol{x})\mathrm{d}\lambda_d(\boldsymbol{x}), \tag{2.48}$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. Here, $f_{\boldsymbol{\mathsf{x}}} \colon \mathbb{R}^d \to [0, \infty)$ is the RND

$$f_{\mathsf{x}}(\boldsymbol{x}) = \frac{\mathrm{d}P_{\boldsymbol{\mathsf{x}}}}{\mathrm{d}\lambda_d}(\boldsymbol{x}). \tag{2.49}$$

Of course, one can always start from the other side (which is usually done in practice), by specifying a nonnegative, μ-integrable function $f_{\mathfrak{X}}(x)$ on $\mathcal{X}$ that normalizes to one, i.e.,

$$\int f_{\mathfrak{X}}(x)\mathrm{d}\mu(x) = 1. \tag{2.50}$$

This function then defines a probability measure $P_{\mathfrak{X}}$ on $\Sigma_{\mathcal{X}}$ according to (2.44). Note that in this case $P_{\mathfrak{X}} \ll \mu$ by construction.

It should be noted that the RN theorem is non-constructive, that is, it only guarantees the existence of a function, but it does not tell us how to calculate it given a probability measure. One important property of the RND is the following [46, pp. 127–128]: If $\mu_1 \ll \mu_2$ and $g(x)$ is a μ_1-integrable function, then

$$\int g(x)\mathrm{d}\mu_1(x) = \int g(x)\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}(x)\mathrm{d}\mu_2(x). \tag{2.51}$$

Applied to the definition of expectation (2.39), this yields (here, $\mu_1 = P_{\mathfrak{X}}$ and $\mu_2 = \mu$)

$$\mathsf{E}[g(\mathfrak{X})] = \int g(x)\mathrm{d}P_{\mathfrak{X}}(x) = \int g(x)f_{\mathfrak{X}}(x)\mathrm{d}\mu(x), \tag{2.52}$$

which is the familiar definition of expectation. In the special case of a random variable x, this becomes the familiar equation

$$\mathsf{E}[g(\mathsf{x})] = \int g(x)\mathrm{d}P_{\mathsf{x}}(x) = \int g(x)f_{\mathsf{x}}(x)\mathrm{d}\lambda(x). \tag{2.53}$$

The next theorem will be instrumental in treating joint distributions of random finite sets in Section 3.7. It is stated here in a general form (for a proof see [42, Section 5.9]).

Theorem 2.19 (Fubini): Let $(\mathcal{X}, \Sigma_{\mathcal{X}}, \mu_{\mathcal{X}})$ and $(\mathcal{Y}, \Sigma_{\mathcal{Y}}, \mu_{\mathcal{Y}})$ be measure spaces, and let $(\mathcal{X} \times \mathcal{Y}, \Sigma_{\mathcal{X}} \otimes \Sigma_{\mathcal{Y}}, \mu_{\mathcal{X}} \times \mu_{\mathcal{Y}})$ be the product space equipped with the product measure as in Definition 2.6 and Theorem 2.9. If $f\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is $\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$-integrable, then

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\mathrm{d}(\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}})(x,y) &= \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} f(x,y)\mathrm{d}\mu_{\mathcal{Y}}(y)\right)\mathrm{d}\mu_{\mathcal{X}}(x)\\ &= \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} f(x,y)\mathrm{d}\mu_{\mathcal{X}}(x)\right)\mathrm{d}\mu_{\mathcal{Y}}(y).\end{aligned}$$

Besides stating that the order of integration can be changed in an iterated integral, Fubini's theorem also implicitly states that the expressions in parentheses are well defined and integrable. Specifically, it states that $f(x, y)$ is $\mu_{\mathcal{Y}}$-integrable a.e. on $\mathcal{X}$, and the resulting function $g(x) = \int_{\mathcal{Y}} f(x, y)\mathrm{d}\mu_{\mathcal{Y}}(y)$ is $\mu_{\mathcal{X}}$-integrable. The second line states the counterpart for $\mu_{\mathcal{X}}$.

A special version of Fubini's theorem arises if $f(x, y)$ is nonnegative. In this case, the requirement that $f(x, y)$ is $\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$-integrable can be dropped [43, p. 233]. This is especially useful when dealing with PDFs (which are always nonnegative), because the integrability does not need to be checked before applying the theorem.

Chapter 3

Random Finite Sets

A random finite set (RFS) X is a random element taking as realizations finite subsets from an underlying set U, i.e., $\mathsf{X} \subseteq U$. In this work, we will consider only RFSs that take on finite subsets of the Euclidean space $U = \mathbb{R}^N$, with some $N \in \mathbb{N}$.

RFSs are useful tools in modeling random experiments, because they incorporate two aspects not incorporated by random vectors:

1. A random vector can be thought of as an ordered list of random variables, e.g., $\mathbf{x} = (\mathsf{x}_1, \mathsf{x}_2) \neq (\mathsf{x}_2, \mathsf{x}_1)$. In contrast, an RFS is unordered, e.g., $\mathsf{X} = \{\mathsf{x}_1, \mathsf{x}_2\} = \{\mathsf{x}_2, \mathsf{x}_1\}$.
2. Whereas the number of components of a random vector is usually fixed, RFSs generally allow varying cardinalities. For instance, X may take realizations $\{x_1\}$, $\{x_2, x_3\}$, ..., for any $x_i \in \mathbb{R}$. Particularly, the outcome of X may also be the empty set $\emptyset$ (with cardinality 0).

Another distinguishing feature – and major mathematical inconvenience – of RFSs is that they do not form a vector space, since addition of two sets cannot be defined in a useful way. One consequence of this is that moments cannot be defined for RFSs as for random vectors. At this point, it should already be clear that RFSs are not a straightforward generalization of random vectors.

In Section 3.1, we start by defining RFSs rigorously as a special kind of random elements, establishing the underlying sample space and Borel algebra. As with random vectors, the full probability measure is often inconvenient to work with. Sections 3.2 and 3.3 introduce two characterizations of RFSs – the cardinality distribution and the belief mass function – that determine the probability measure only on a restricted class of events but are more tractable. In Section 3.4, we discuss how finite sets can be equivalently represented in vector spaces and introduce a second measure (besides the probability measure) on the measurable set space. This allows us to invoke the RN theorem to define a PDF for RFSs in Section 3.5. Section 3.6 shows how RFSs can be related to a sequence of random vectors. Finally, Sections 3.7 and 3.8 discuss joint and conditional distributions of two or more random finite sets, respectively.

3.1 Definition of Random Finite Sets

Following [46, Chapter 5], we will now rigorously define RFSs as random elements in the measure-theoretic framework of Chapter 2. Recall from Section 2.6 that any random element is a measurable mapping from an abstract probability space $(\Omega, \Sigma_\Omega, P)$ to a measurable space $(\mathcal{X}, \Sigma_\mathcal{X})$.

For an RFS, the set $\mathcal{X}$ is the collection of all finite subsets of $\mathbb{R}^N$, i.e.,

$$\mathcal{X} \triangleq \{X \subseteq \mathbb{R}^N : |X| \in \mathbb{N}_0\}, \tag{3.1}$$

where $|X| = 0$ iff $X = \emptyset$. A σ-algebra $\Sigma_\mathcal{X}$ is a collection of subsets[1] of $\mathcal{X}$ that fulfills the conditions in Definition 2.1. Following Section 2.2, we would like to generate $\Sigma_\mathcal{X}$ by the collection of open subsets of $\mathcal{X}$, that is, we are looking for a Borel algebra $\mathcal{B}(\mathcal{X})$. In order to do so, according to Definition 2.5, we need to specify a topology on $\mathcal{X}$. We take the **hit-or-miss topology** from [3, p. 94], restricted to[2] $\mathcal{X}$.

Definition 3.1: *Let $\mathcal{X}$ be the collection of all finite subsets of $\mathbb{R}^N$ as defined in (3.1), let*

$$\mathcal{K} \triangleq \{K \subseteq \mathbb{R}^N : K \text{ is compact}\}$$

be the collection of all compact subsets of $\mathbb{R}^N$, and let

$$\mathcal{O} \triangleq \{O \subseteq \mathbb{R}^N : O \text{ is open}\}$$

be the collection of all open subsets of $\mathbb{R}^N$. Furthermore, for $A \subseteq \mathbb{R}^N$, let

$$\mathcal{X}_A \triangleq \{X \in \mathcal{X} : X \cap A \neq \emptyset\}$$

*be the collection of all finite subsets of $\mathbb{R}^N$ that **hit** A, and let*

$$\mathcal{X}^A \triangleq (\mathcal{X}_A)^c = \{X \in \mathcal{X} : X \cap A = \emptyset\}$$

*be the collection of all finite subsets of $\mathbb{R}^N$ that **miss** A. Then, the collection of collections of finite sets $\mathcal{D}$, defined as*

$$\mathcal{D} \triangleq \left\{\mathcal{X}^K \cap \mathcal{X}_{O_1} \cap \cdots \cap \mathcal{X}_{O_n} : K \in \mathcal{K}, O_i \in \mathcal{O}, n \geq 1\right\}, \tag{3.2}$$

[1] Since the elements of $\mathcal{X}$ are finite subsets of $\mathbb{R}^N$, a subset $\mathcal{A} \subseteq \mathcal{X}$ is a collection of finite subsets. Therefore, a σ-algebra is a collection of collections of finite subsets of $\mathbb{R}^N$.

[2] In [3], the hit-or-miss topology $\mathcal{T}'$ is defined on the collection $\mathcal{C}$ of all *closed* subsets of $\mathbb{R}^N$, rather than the collection of all *finite* subsets $\mathcal{X} \subseteq \mathcal{C}$. In this work, by hit-or-miss topology we mean the "subspace" topology $\mathcal{T}$ of $\mathcal{T}'$ on $\mathcal{X} \subseteq \mathcal{C}$, given as $\mathcal{T} = \{\mathcal{A} \cap \mathcal{X} : \mathcal{A} \in \mathcal{T}'\}$.

is a base for a topology $\mathcal{T}$. The topology $\mathcal{T}$ generated by this base $\mathcal{D}$ is called the ***hit-or-miss topology*** *on $\mathcal{X}$.*

In other words, a collection of finite sets $\mathcal{A} \subseteq \mathcal{X}$ is a base member $\mathcal{A} \in \mathcal{D}$ iff there exists a compact subset K of $\mathbb{R}^N$ and a finite number of open subsets $O_1, \dots, O_n$ of $\mathbb{R}^N$, such that

$$\mathcal{A} = \mathcal{X}^K \cap \mathcal{X}_{O_1} \cap \dots \cap \mathcal{X}_{O_n}. \tag{3.3}$$

The whole topology $\mathcal{T}$ is then generated by this base (see Definition 2.4), i.e., it is constituted by all unions of these base members.

With the hit-or-miss topology $\mathcal{T}$ in place, we can now define the Borel algebra on $\mathcal{X}$ as the σ-algebra generated by $\mathcal{T}$ (cf. Definition 2.5), i.e.,

$$\mathcal{B}(\mathcal{X}) \triangleq \sigma(\mathcal{T}). \tag{3.4}$$

Hence, $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a measurable space. Of course, since the definition of the hit-or-miss topology $\mathcal{T}$ is quite abstract, $\mathcal{B}(\mathcal{X})$ is abstract too. In the following two sections, we will look at two more practical types of collections of finite sets and show that they are contained in $\mathcal{B}(\mathcal{X})$. However, at this point, we can formally define an RFS.

Definition 3.2: *Let $(\Omega, \Sigma_\Omega, P)$ be a probability space, and let $\mathcal{X}$ be the collection of all finite subsets of $\mathbb{R}^N$ as defined in* (3.1). *Moreover let $\mathcal{B}(\mathcal{X})$ be the Borel algebra on $\mathcal{X}$ generated by the hit-or-miss topology as in* (3.4). *A* ***random finite set (RFS)*** *is a $\Sigma_\Omega/\mathcal{B}(\mathcal{X})$-measurable mapping $\mathsf{X}\colon \Omega \to \mathcal{X}$.*

Particularly, as in Section 2.6, X induces a probability measure $P_\mathsf{X}\colon \mathcal{B}(\mathcal{X}) \to [0,1]$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ by

$$P_\mathsf{X}(\mathcal{A}) \triangleq P(\mathsf{X}^{-1}(\mathcal{A})) = \Pr(\mathsf{X} \in \mathcal{A}), \tag{3.5}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. Consequently, $(\mathcal{X}, \mathcal{B}(\mathcal{X}), P_\mathsf{X})$ is a probability space.

3.2 Cardinality Distribution

It would be desirable to make statements about the cardinality of an RFS X such as: "what is the probability that $|\mathsf{X}| = k$, for some $k \in \mathbb{N}_0$." Let us define the collections of all finite subsets of $\mathbb{R}^N$ with exactly k elements as

$$\mathcal{X}_k \triangleq \{X \in \mathcal{X} : |X| = k\}, \tag{3.6}$$

for all $k \in \mathbb{N}_0$, where $\mathcal{X}_0 = \{\emptyset\}$. With this definition, $|\mathsf{X}| = k$ is equivalent to $\mathsf{X} \in \mathcal{X}_k$. It is clear that these collections $\mathcal{X}_k$ are disjoint and form a partition of $\mathcal{X}$, i.e.,

$$\mathcal{X} = \bigcup_{k=0}^{\infty} \mathcal{X}_k, \qquad \mathcal{X}_k \cap \mathcal{X}_l = \emptyset \text{ for } k \neq l. \tag{3.7}$$

Additionally, we define the collections containing all finite sets with no more than k elements,

$$\mathcal{X}_{\leq k} \triangleq \{X \in \mathcal{X} : |X| \leq k\} = \bigcup_{i=0}^{k} \mathcal{X}_i, \tag{3.8}$$

and the collections containing all finite sets with at least k elements,

$$\mathcal{X}_{\geq k} \triangleq \{X \in \mathcal{X} : |X| \geq k\} = \bigcup_{i=k}^{\infty} \mathcal{X}_i. \tag{3.9}$$

The proof of the following result can be found in Appendix A.1.

Lemma 3.3: For all $k \in \mathbb{N}_0$, the collections $\mathcal{X}_k$, $\mathcal{X}_{\leq k}$, and $\mathcal{X}_{\geq k}$ are elements of $\mathcal{B}(\mathcal{X})$.

After checking that the collections $\mathcal{X}_k$ are indeed events in our σ-algebra $\mathcal{B}(\mathcal{X})$, we define the cardinality distribution of X as the restriction of the probability measure P_{X} on the events $\mathcal{X}_k$:

Definition 3.4: *The* ***cardinality distribution (CD)*** *of an RFS* X *with associated probability space* $(\mathcal{X}, \mathcal{B}(\mathcal{X}), P_{\mathsf{X}})$ *is a function* $p_{|\mathsf{X}|} \colon \mathbb{N}_0 \to [0, 1]$ *defined by*

$$p_{|\mathsf{X}|}(k) \triangleq P_{\mathsf{X}}(\mathcal{X}_k) = \Pr(\mathsf{X} \in \mathcal{X}_k) = \Pr(|\mathsf{X}| = k).$$

That is, the cardinality distribution is the probability mass function (PMF) of the discrete random variable $|\mathsf{X}| : \mathcal{X} \to \mathbb{N}_0$.

It should be stressed that the CD characterizes an RFS only in part, since the probability measure is only specified for the particular events $\mathcal{X}_k$.

3.3 Belief Mass Function

Another statement about an RFS that would be interesting to make is: "what is the probability that $X \subseteq A$, for some $A \subseteq \mathbb{R}^N$." Let $\mathcal{C}$ be the collection of all closed subsets of $\mathbb{R}^N$. We define the collection of all finite subsets of a closed set $A \in \mathcal{C}$ as

$$\mathcal{X}(A) \triangleq \{X \in \mathcal{X} : X \subseteq A\}. \tag{3.10}$$

With this definition, $\mathsf{X} \subseteq A$ is equivalent to $\mathsf{X} \in \mathcal{X}(A)$. A proof of the following lemma is provided in Appendix A.2.

Lemma 3.5: For any closed subset $A \in \mathcal{C}$, $\mathcal{X}(A) \in \mathcal{B}(\mathcal{X})$.

After establishing that the collections $\mathcal{X}(A)$ are indeed events, we make the following definition.

Definition 3.6: *Let* X *be an RFS with associated probability space* $(\mathcal{X}, \mathcal{B}(\mathcal{X}), P_{\mathsf{X}})$*, and let* $\mathcal{C}$ *the collection of all closed subsets of* $\mathbb{R}^N$ *and* $A \in \mathcal{C}$*. The function* $\beta_{\mathsf{X}} \colon \mathcal{C} \to [0,1]$ *defined by*

$$\beta_{\mathsf{X}}(A) \triangleq P_{\mathsf{X}}\left(\mathcal{X}(A)\right) = \Pr(\mathsf{X} \in \mathcal{X}(A)) = \Pr(\mathsf{X} \subseteq A),$$

is called the ***belief mass function (BMF)*** *of* X*.*

For $A = \mathbb{R}^N$, we have $\mathcal{X}(\mathbb{R}^N) = \mathcal{X}$ and obtain

$$\beta_{\mathsf{X}}(\mathbb{R}^N) = P_{\mathsf{X}}(\mathcal{X}) = 1. \tag{3.11}$$

For $A = \emptyset$, we have $\mathcal{X}(\emptyset) = \{\emptyset\} = \mathcal{X}_0$ where the last equality follows from (3.6). Therefore, the BMF evaluated at $A = \emptyset$ becomes

$$\beta_{\mathsf{X}}(\emptyset) = P_{\mathsf{X}}(\{\emptyset\}) = P_{\mathsf{X}}(\mathcal{X}_0) = p_{|\mathsf{X}|}(0). \tag{3.12}$$

Because $\beta_{\mathsf{X}}(\emptyset) = p_{|\mathsf{X}|}(0) \neq 0$ in general, β_{X} is not a measure as in Definition 2.7. Note the difference between the events $\mathsf{X} \in \{\emptyset\}$ (which means $\mathsf{X} = \emptyset$, i.e., X does not contain any elements) and $\mathsf{X} \in \emptyset$ (which means there is no realization of X). Of course $P_{\mathsf{X}}(\emptyset) = 0$.

A remarkable fact about the BMF is that, although it explicitly assigns a probability only to the particular events $\mathcal{X}(A)$, it completely characterizes the whole probability measure P_{X} – in the sense that, given β_{X}, there exists a unique P_{X} satisfying $P_{\mathsf{X}}(\mathcal{X}(A)) = \beta_{\mathsf{X}}(A)$. This is a consequence of the Choquet theorem in [46, pp. 44–45].

3.4 Relation between Finite Sets and Vectors

If we take $\mu_1 = P_{\mathsf{X}}$ in the RN theorem (Theorem 2.18) and establish a second σ-finite measure $\mu_{\mathcal{X}}$ (corresponding to μ_2 in Theorem 2.18) on $\mathcal{B}(\mathcal{X})$ such that $P_{\mathsf{X}} \ll \mu_{\mathcal{X}}$, we can formally define a PDF $f_{\mathsf{X}} \colon \mathcal{X} \to [0, \infty)$ for RFSs as the RND

$$f_{\mathsf{X}}(X) = \frac{\mathrm{d}P_{\mathsf{X}}}{\mathrm{d}\mu_{\mathcal{X}}}(X). \tag{3.13}$$

The probability of any event $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ can then be calculated by integrating the PDF over the event with respect to the measure $\mu_{\mathcal{X}}$, i.e.,

$$P_{\mathsf{X}}(\mathcal{A}) = \int_{\mathcal{A}} f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.14}$$

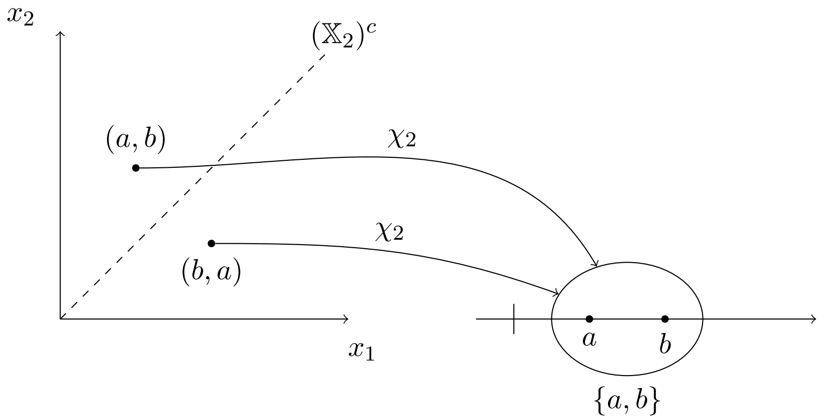

Fig. 3.1: *Mapping of 2-dimensional vectors ($N = 1$, $k = 2$) to finite subsets of $\mathbb{R}$ via the function χ_2. Vectors mirrored along the line $(\mathbb{X}_2)^c$ (in our case, (a, b) and (b, a)) are mapped to the same set with two elements.*

Having defined a measure $\mu_{\mathcal{X}}$, we usually start from the other side by specifying a nonnegative, $\mu_{\mathcal{X}}$-integrable function $f_{\mathsf{X}}(X)$ on $\mathcal{X}$ that normalizes to one, i.e.,

$$\int f_{\mathsf{X}}(X) \mathrm{d}\mu(x) = 1. \tag{3.15}$$

This function (the PDF) then defines a probability measure P_{X} on $\mathcal{B}(\mathcal{X})$ according to (3.14). In this case, $P_{\mathsf{X}} \ll \mu_{\mathcal{X}}$ by construction.

The measure $\mu_{\mathcal{X}}$, which we call **set measure**, is not unique, i.e., we are free to choose any σ-finite measure on $\mathcal{B}(\mathcal{X})$. Depending on our choice we will end up with different classes of RFSs for which PDFs $f_{\mathsf{X}}(X)$ exist. Naturally then, the question arises which specific set measure shall be used so that PDFs exist for a sufficiently large and useful class of RFSs.

The key concept for introducing the set measure $\mu_{\mathcal{X}}$ in [3, 46] is to represent a finite set as an equivalent set of vectors. For $k \in \mathbb{N}$, let $\mathbb{X}_k \subseteq \mathbb{R}^{kN}$ be the subset of the Euclidean space $\mathbb{R}^{kN}$ containing all vectors with k different components, i.e.,

$$\mathbb{X}_k \triangleq \big\{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{R}^{kN} : \boldsymbol{x}_i \neq \boldsymbol{x}_j \text{ for all } i \neq j\big\}. \tag{3.16}$$

Furthermore, let $\chi_k \colon \mathbb{X}_k \to \mathcal{X}_k$ be functions converting vectors to finite sets, defined by

$$\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \triangleq \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}, \tag{3.17}$$

where $\boldsymbol{x}_i \in \mathbb{R}^N$, for $i = 1, \ldots, k$ (the case $N = 1$ and $k = 2$ is visualized in Figure 3.1). We will call these functions **vector-to-set transformations (VST)**.

Note that for every finite set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \in \mathcal{X}_k$ there are exactly $k!$ vectors in $\mathbb{X}_k$ that are mapped to it. These vectors correspond to the $k!$ permutations of the elements of the finite set, i.e.,

$$\chi_k(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}) = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}. \tag{3.18}$$

for all permutations σ over $1, \ldots, k$. It follows that VSTs are surjective and symmetric functions.

The inverse image of a collection of finite sets $\mathcal{A} \subseteq \mathcal{X}_k$ under χ_k gives the set of all vectors that map to the finite sets in $\mathcal{A}$, i.e.,

$$\chi_k^{-1}(\mathcal{A}) = \{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{X}_k : \chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathcal{A}\}. \tag{3.19}$$

In the special case where $\mathcal{A} = \{X\}$ contains only one k-ary set with

$$X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \in \mathcal{X}_k, \tag{3.20}$$

the inverse image consists of the $k!$ permutations of the elements of X as discussed above in (3.18), i.e.,

$$\chi_k^{-1}(\{X\}) = \{(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}) : \sigma \text{ is a permuation over } 1, \ldots, k\}. \tag{3.21}$$

Thus, we can equivalently think about any finite set $X \subseteq \mathbb{R}^N$ with k elements as a set of $k!$ points in the Euclidean domain $\mathbb{X}_k$.

3.4.1 Set Measure for k-ary Sets

The basic idea for establishing a set measure $\mu_{\mathcal{X}}$ on $\mathcal{X}$ in [3,46] is to use the VSTs introduced in the preceding section to induce measures $\mu_{\mathcal{X}_k}$ on the subspaces $\mathcal{X}_k \subseteq \mathcal{X}$. Then, these measures are combined to construct the set measure $\mu_{\mathcal{X}}$ for the whole set space $\mathcal{X}$.

We begin by equipping $\mathbb{X}_k$ with a σ-algebra $\mathcal{B}(\mathbb{X}_k)$. Here, we choose the Borel algebra $\mathcal{B}(\mathbb{R}^{kN})$ restricted to $\mathbb{X}_k \subseteq \mathbb{R}^{kN}$ given by [42, Theorem 1.8.1]

$$\mathcal{B}(\mathbb{X}_k) = \{A \cap \mathbb{X}_k : A \in \mathcal{B}(\mathbb{R}^{kN})\}. \tag{3.22}$$

Furthermore, we equip the measurable space $(\mathbb{X}_k, \mathcal{B}(\mathbb{X}_k))$ with the kN-th product Lebesgue measure λ_{kN} (cf. (2.15)) so that $(\mathbb{X}_k, \mathcal{B}(\mathbb{X}_k), \lambda_{kN})$ is a measure space. Similarly, we equip $\mathcal{X}_k$ with the Borel algebra $\mathcal{B}(\mathcal{X})$ (cf. (3.4)) restricted to $\mathcal{X}_k \subseteq \mathcal{X}$, i.e.,

$$\mathcal{B}(\mathcal{X}_k) = \{\mathcal{A} \cap \mathcal{X}_k : \mathcal{A} \in \mathcal{B}(\mathcal{X})\}, \tag{3.23}$$

resulting in the measurable space $(\mathcal{X}_k, \mathcal{B}(\mathcal{X}_k))$. A proof of the following lemma can be found in Appendix A.3.

Lemma 3.7: The VST χ_k is $\mathcal{B}(\mathbb{X}_k)/\mathcal{B}(\mathcal{X}_k)$ measurable.

After verifying the measurability of χ_k we define the measure $\mu_{\mathcal{X}_k}$ on $\mathcal{B}(\mathcal{X}_k)$ as the push-forward measure (Definition 2.12) of λ_{kN} weighted by $\frac{1}{k!}$, i.e.,

$$\mu_{\mathcal{X}_k}(\mathcal{A}) \triangleq \frac{1}{k!}\lambda_{kN}(\chi_k^{-1}(\mathcal{A})), \qquad \mathcal{A} \in \mathcal{B}(\mathcal{X}_k). \tag{3.24}$$

The factor $\frac{1}{k!}$ accounts for the fact that $k!$ points are mapped to each set and, thus, the full Lebesgue measure would "overvalue" the sets. For $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$ we will call the event $\chi_k^{-1}(\mathcal{A}) \in \mathcal{B}(\mathbb{X}_k)$ the **vector event** corresponding to the (set) event $\mathcal{A}$. That is, we are measuring an event $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$ indirectly by first transforming it to the corresponding vector event $\chi_k^{-1}(\mathcal{A}) \in \mathcal{B}(\mathbb{X}_k)$, and then measuring the vector event with the Lebesgue measure λ_{kN} and divide by $k!$.

Example 3.8: Before proceeding, let us consider some basic events to better understand the measure defined in (3.24).

(a) Let $\mathcal{A} = \mathcal{X}_k$. Here, we have

$$\chi_k^{-1}(\mathcal{X}_k) = \mathbb{X}_k. \tag{3.25}$$

Recall that $\mathbb{X}_k \subseteq \mathbb{R}^{kN}$ consists of all vectors $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in \mathbb{R}^{kN}$ with k different components (cf. (3.16)). Accordingly, its complement $(\mathbb{X}_k)^c$ contains all vectors with at least two identical components and is the union of proper (lower-dimensional) subspaces of $\mathbb{R}^{kN}$. Since proper subspaces of $\mathbb{R}^{kN}$ are null sets with respect to the Lebesgue measure λ_{kN}, $(\mathbb{X}_k)^c$ is a null set and (3.24) becomes

$$\mu_{\mathcal{X}_k}(\mathcal{X}_k) = \frac{1}{k!}\lambda_{kN}(\mathbb{X}_k) = \frac{1}{k!}\lambda_{kN}(\mathbb{R}^{kN} \setminus (\mathbb{X}_k)^c) = \frac{1}{k!}\lambda_{kN}(\mathbb{R}^{kN}) = \infty. \tag{3.26}$$

(b) Let $\mathcal{A} = \mathcal{X}_k(A)$ be the collection of all k-ary subsets of the closed set $A \subseteq \mathbb{R}^N$ (cf. (3.10)). Note that $\chi_k(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in \mathcal{X}_k(A)$ iff all components $\boldsymbol{x}_i$ are different and $\boldsymbol{x}_i \in A$, for $i = 1, \dots, k$, or equivalently iff $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in (A \times \cdots \times A) \cap \mathbb{X}_k = A^k \cap \mathbb{X}_k$. Hence,

$$\chi_k^{-1}(\mathcal{X}_k(A)) = A^k \cap \mathbb{X}_k = A^k \setminus (\mathbb{X}_k)^c. \tag{3.27}$$

As discussed above in (a), $(\mathbb{X}_k)^c$ is a null set with respect to the Lebesgue measure λ_{kN}. Thus,

$$\mu_{\mathcal{X}_k}(\mathcal{X}_k(A)) = \frac{1}{k!}\lambda_{kN}(A^k \setminus (\mathbb{X}_k)^c) = \frac{1}{k!}\lambda_{kN}(A^k) = \frac{1}{k!}(\lambda_N(A))^k. \tag{3.28}$$

The essential benefit of choosing the measure $\mu_{\mathcal{X}_k}$ as the scaled push-forward measure of λ_{kN} under χ_k is that we can use Theorem 2.16 to integrate functions defined on $\mathcal{X}_k$. Let $g\colon \mathcal{X}_k \to \mathbb{R}$ be an integrable function with respect to $\mu_{\mathcal{X}_k}$. Then, according to (2.43) (with $f = \chi_k$, $\mu_{\mathcal{X}_k} = \frac{1}{k!}\mu_f$, and $\mu = \lambda_{kN}$) we have

$$\int_{\mathcal{A}} g(X)\mathrm{d}\mu_{\mathcal{X}_k}(X) = \int_{\mathcal{A}} g(X)\mathrm{d}\left(\frac{1}{k!}\mu_f(X)\right) \tag{3.29}$$

$$= \frac{1}{k!}\int_{\mathcal{A}} g(X)\mathrm{d}\mu_f(X) \tag{3.30}$$

$$= \frac{1}{k!}\int_{\chi_k^{-1}(\mathcal{A})} g(\chi_k(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k))\mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k), \tag{3.31}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$. That is, the integral of $g(X)$ with respect to $\mu_{\mathcal{X}_k}$ can be calculated as a "conventional" integral of the composite function $g(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k))$ with respect to λ_{kN}. We define the function $\tilde{g}\colon \mathbb{R}^{kN} \to \mathbb{R}$ as

$$\tilde{g}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \triangleq \begin{cases} g(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)) & \text{if } (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{X}_k, \\ 0 & \text{if } (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \notin \mathbb{X}_k, \end{cases} \tag{3.32}$$

which we call the **equivalent vector function (EVF)** corresponding to the set function $g(X)$. Since the VST χ_k is a symmetric function (cf. (3.18)), $\tilde{g}$ is symmetric as well, i.e.,

$$\tilde{g}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \tilde{g}(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}), \tag{3.33}$$

for all permutations σ over $1, \ldots, k$. Note that there is a one-to-one correspondence between $\tilde{g}$ and g. Clearly, given a set function g we can calculate the EVF $\tilde{g}$ as given by (3.32). Conversely, given an EVF $\tilde{g}$ we can calculate $g(\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\})$ by evaluating $\tilde{g}$ at any of the $k!$ permutations $(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k})$. One way to formulate this more precisely is to define $\theta_k\colon \mathcal{X}_k \to \mathbb{X}_k$ given by

$$\theta_k(X) \triangleq (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \qquad \boldsymbol{x}_i \in X, \text{ such that } \boldsymbol{x}_1 < \boldsymbol{x}_2 < \cdots < \boldsymbol{x}_k, \tag{3.34}$$

where $<$ is the lexicographical order on $\mathbb{R}^N$. That is, $\theta_k(X)$ sorts the k elements of X in an ascending order and returns the corresponding vector. Note that $\chi_k(\theta_k(X)) = X$. Using the function θ_k, the set function g can then be obtained from the EVF $\tilde{g}$ by

$$g(X) = \tilde{g}(\theta_k(X)). \tag{3.35}$$

With the definition of the EVF in (3.32) we can write the integral in (3.31) as

$$\int_{\mathcal{A}} g(X) \mathrm{d}\mu_{\mathcal{X}_k}(X) = \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A})} \tilde{g}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \tag{3.36}$$

Again, the point made here is that by using the VST χ_k to induce the measure $\mu_{\mathcal{X}_k}$ we can represent a set function $g(X)$ by an EVF $\tilde{g}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$. Integrating the set function $g(X)$ over an event $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$ with respect to $\mu_{\mathcal{X}_k}$ is then equal to the weighted "conventional" integral of the EVF over the corresponding vector event $\chi_k^{-1}(\mathcal{A}) \in \mathcal{B}(\mathbb{X}_k)$.

Special case $k = 0$. In our discussion so far we have introduced the measure spaces $(\mathcal{X}_k, \mathcal{B}(\mathcal{X}_k), \mu_{\mathcal{X}_k})$ for $k \geq 1$. The set space $\mathcal{X}_0 = \{\emptyset\}$ consists of just one element, the empty set. Hence, the only σ-algebra is the trivial σ-algebra $\mathcal{B}(\mathcal{X}_0) = \{\emptyset, \mathcal{X}_0\} = \{\emptyset, \{\emptyset\}\}$. We define the measure $\mu_{\mathcal{X}_0}$ on $\mathcal{B}(\mathcal{X}_0)$ as

$$\mu_{\mathcal{X}_0}(\mathcal{A}) \triangleq \mathbf{1}_{\mathcal{A}}(\emptyset) = \begin{cases} 1 & \text{if } \emptyset \in \mathcal{A}, \\ 0 & \text{if } \emptyset \notin \mathcal{A}, \end{cases} \tag{3.37}$$

for $\mathcal{A} \in \mathcal{B}(\mathcal{X}_0)$. Note that a function $g\colon \mathcal{X}_0 \to \mathbb{R}$ is defined on a single element only, i.e., it is just a single number $g(\emptyset) \in \mathbb{R}$. Furthermore, we have

$$\int_{\mathcal{A}} g(X)\mathrm{d}\mu_{\mathcal{X}_0}(X) = g(\emptyset)\mathbf{1}_{\mathcal{A}}(\emptyset). \tag{3.38}$$

For convenience we make here the convention that

$$\tilde{g}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_0) \triangleq g(\emptyset), \tag{3.39}$$

and

$$\frac{1}{0!}\int_{\chi_0^{-1}(\mathcal{A})} \tilde{g}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_0)\mathrm{d}\lambda_{0N}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_0) \triangleq g(\emptyset)\mathbf{1}_{\mathcal{A}}(\emptyset), \tag{3.40}$$

so we can use the same notation for $k = 0$ as for $k \geq 1$ in (3.36).

3.4.2 Set Measure for General Finite Sets

After introducing the measure spaces $(\mathcal{X}_k, \mathcal{B}(\mathcal{X}_k), \mu_{\mathcal{X}_k})$, $k \in \mathbb{N}_0$, for k-ary sets in the preceding section, we will now establish the set measure $\mu_{\mathcal{X}}$ for general finite sets $X \in \mathcal{X}$ with the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Since the $\mathcal{X}_k$, $k \in \mathbb{N}_0$, form a disjoint partition of $\mathcal{X} = \bigcup_{k=0}^{\infty} \mathcal{X}_k$ (cf. (3.7)), the idea for constructing the set measure $\mu_{\mathcal{X}}$ is as follows:

1. The event $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ is partitioned into disjoint subevents $\mathcal{A}_k$, $k \in \mathbb{N}_0$, consisting of all k-ary sets in $\mathcal{A}$, i.e.,

$$\mathcal{A}_k \triangleq \mathcal{A} \cap \mathcal{X}_k, \qquad k \in \mathbb{N}_0, \tag{3.41}$$

$$\mathcal{A} = \bigcup_{k=0}^{\infty} \mathcal{A}_k. \tag{3.42}$$

   Note that each subevent $\mathcal{A}_k$ is included in the corresponding σ-algebra $\mathcal{B}(\mathcal{X}_k)$ (cf. (3.23)).

2. Each subevent $\mathcal{A}_k \in \mathcal{B}(\mathcal{X}_k)$ is then measured with the k-ary set measures $\mu_{\mathcal{X}_k}(\mathcal{A}_k)$ introduced in Section 3.4.1 and the results summed up, i.e.,

$$\mu_{\mathcal{X}}(\mathcal{A}) \triangleq \sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}(\mathcal{A}_k), \qquad \mathcal{A} \in \mathcal{B}(\mathcal{X}). \tag{3.43}$$

A proof of the following Lemma is provided in Appendix A.4.

Lemma 3.9: The function $\mu_{\mathcal{X}}$ as defined in (3.43) is a measure on $\mathcal{B}(\mathcal{X})$.

Inserting the measures $\mu_{\mathcal{X}_k}$ from (3.24) and (3.37) we can express the set measure in (3.43) above as

$$\mu_{\mathcal{X}}(\mathcal{A}) = \mathbf{1}_{\mathcal{A}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!}\lambda_{kN}(\chi_k^{-1}(\mathcal{A}_k)), \tag{3.44}$$

where $\mathbf{1}_{\mathcal{A}_0}(\emptyset) = \mathbf{1}_{\mathcal{A}}(\emptyset)$ has been used.

Example 3.10: Before proceeding, let us consider some basic events to better understand the set measure defined in (3.43).

(a) Let $\mathcal{A} = \mathcal{X}_0 = \{\emptyset\}$. Here $\mathbf{1}_{\mathcal{A}}(\emptyset) = 1$. Furthermore, for $k \in \mathbb{N}$, since $\emptyset \notin \mathcal{X}_k$, we have $\mathcal{A}_k = \{\emptyset\} \cap \mathcal{X}_k = \emptyset$. Hence, $\chi_k^{-1}(\mathcal{A}_k) = \chi_k^{-1}(\emptyset) = \emptyset$ and the sum in (3.44) vanishes since $\lambda_{kN}(\emptyset) = 0$ for all $k \in \mathbb{N}$. Consequently,

$$\mu_{\mathcal{X}}(\{\emptyset\}) = 1. \tag{3.45}$$

(b) Let $\mathcal{A} = \mathcal{X}_n$, with some $n \in \mathbb{N}$. In this case, $\mathbf{1}_{\mathcal{A}}(\emptyset) = 0$ since $\emptyset \notin \mathcal{X}_n$. Moreover, because $\mathcal{X}_n \cap \mathcal{X}_k = \emptyset$ if $k \neq n$, only the term with $k = n$ does not vanish in the sum, and thus

$$\mu_{\mathcal{X}}(\mathcal{X}_n) = \mu_{\mathcal{X}_n}(\mathcal{X}_n) = \frac{1}{n!}\lambda_{nN}\big(\chi_n^{-1}(\mathcal{X}_n)\big) = \infty, \tag{3.46}$$

where the last result follows from (3.26) from Example 3.8(a).

(c) Let $\mathcal{A} = \mathcal{X}(A)$, with some closed subset $A \subseteq \mathbb{R}^N$, i.e., $\mathcal{A}$ is the collection of all finite subsets of A. Because $\emptyset \subseteq A$, we have $\emptyset \in \mathcal{A}$ and $\mathbf{1}_{\mathcal{A}}(\emptyset) = 1$. Furthermore, the subevent $\mathcal{A}_k = \mathcal{X}(A) \cap \mathcal{X}_k = \mathcal{X}_k(A)$ is the collection of all k-ary subsets of A. As shown in Example 3.8(b) we have

$$\mu_{\mathcal{X}_k}(\mathcal{X}_k(A)) = \frac{1}{k!}(\lambda_N(A))^k. \tag{3.47}$$

Therefore, the set measure (3.43) becomes (where we use the convention $0^0 = 1$)

$$\mu_{\mathcal{X}}(\mathcal{X}(A)) = 1 + \sum_{k=1}^{\infty} \frac{1}{k!}(\lambda_N(A))^k = \sum_{k=0}^{\infty} \frac{1}{k!}(\lambda_N(A))^k = e^{\lambda_N(A)}. \tag{3.48}$$

After introducing the set measure $\mu_{\mathcal{X}}$, our next aim is to obtain a formula for integrating a general set function $g\colon \mathcal{X} \to \mathbb{R}$ similar to the one we derived in (3.36) for integrating k-ary set functions. To this end, note that every set function $g\colon \mathcal{X} \to \mathbb{R}$ can be written in a piecewise fashion as

$$g(X) = \begin{cases} g^{(0)}(X) & \text{if } |X| = 0, \\ g^{(1)}(X) & \text{if } |X| = 1, \\ g^{(2)}(X) & \text{if } |X| = 2, \\ \quad\vdots & \quad\quad\vdots \end{cases} \tag{3.49}$$

where $g^{(k)}\colon \mathcal{X}_k \to \mathbb{R}$ are k-ary set functions, for $k \in \mathbb{N}_0$. With this notation we obtain upon inserting the set measure (3.43),

$$\int_{\mathcal{A}} g(X)\mathrm{d}\mu_{\mathcal{X}}(X) = \int_{\mathcal{A}} g(X)\mathrm{d}\left(\sum_{k=0}^{\infty}\mu_{\mathcal{X}_k}(X\cap\mathcal{X}_k)\right) \tag{3.50}$$

$$= \sum_{k=0}^{\infty}\int_{\mathcal{A}\cap\mathcal{X}_k} g(X)\mathrm{d}\mu_{\mathcal{X}_k}(X) \tag{3.51}$$

$$= \sum_{k=0}^{\infty}\int_{\mathcal{A}_k} g^{(k)}(X)\mathrm{d}\mu_{\mathcal{X}_k}(X). \tag{3.52}$$

Since $g^{(k)}(X)$ are k-ary set functions, we can use (3.36) to express the integrals in the sum as

$$\int_{\mathcal{A}_k} g^{(k)}(X)\mathrm{d}\mu_{\mathcal{X}_k}(X) = \frac{1}{k!}\int_{\chi_k^{-1}(\mathcal{A}_k)} \tilde{g}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)\mathrm{d}\lambda_{kN}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k). \tag{3.53}$$

where $\tilde{g}^{(k)}\colon \mathbb{R}^{kN} \to \mathbb{R}$ are EVFs corresponding to the set functions $g^{(k)}$ given by (cf. (3.32))

$$\tilde{g}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) \triangleq \begin{cases} g^{(k)}(\chi_k(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)) & \text{if } (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)\in\mathbb{X}_k, \\ 0 & \text{if } (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)\notin\mathbb{X}_k, \end{cases} \tag{3.54}$$

$$= \begin{cases} g(\chi_k(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)) & \text{if } (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)\in\mathbb{X}_k, \\ 0 & \text{if } (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)\notin\mathbb{X}_k, \end{cases} \tag{3.55}$$

The set function can then be obtained from the EVFs by

$$g(X) = \tilde{g}^{(k)}(\theta_k(X)), \qquad |X| = k, \tag{3.56}$$

where θ_k are the functions defined in (3.34). Hence, there is a one-to-one correspondence between a general set function $g(X)$ and a sequence of symmetric EVFs $\tilde{g}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)$, $k\in\mathbb{N}_0$. Inserting (3.53) in (3.52), we finally obtain

$$\int_{\mathcal{A}} g(X)\mathrm{d}\mu_{\mathcal{X}}(X) = \sum_{k=0}^{\infty}\frac{1}{k!}\int_{\chi_k^{-1}(\mathcal{A}_k)} \tilde{g}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)\mathrm{d}\lambda_{kN}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k). \tag{3.57}$$

Integrating the set function $g(X)$ over an event $\mathcal{A}\in\mathcal{B}(\mathcal{X})$ with respect to $\mu_{\mathcal{X}}$ is then equal to the weighted sum of "conventional" integrals of the EVFs. We first partition the event $\mathcal{A}$ into k-ary subevents $\mathcal{A}_k$. Then, for each $k\in\mathbb{N}_0$, the EVFs $\tilde{g}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)$ are integrated over the corresponding vector events $\chi_k^{-1}(\mathcal{A}_k)$. Finally, the results are weighted by $\frac{1}{k!}$ and summed up.

3.5 Probability Density Function for RFS

Recall that our motivation for introducing the set measure $\mu_{\mathcal{X}}$ on $\mathcal{B}(\mathcal{X})$ in the previous section was to be able to define a PDF $f_{\mathsf{X}}\colon \mathcal{X}\to[0,\infty)$ for RFSs as the RND

$$f_{\mathsf{X}}(X) = \frac{\mathrm{d}P_{\mathsf{X}}}{\mathrm{d}\mu_{\mathcal{X}}}(X). \tag{3.58}$$

We can then calculate the probability of any event $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ by

$$P_{\mathsf{X}}(\mathcal{A}) = \int_{\mathcal{A}} f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.59}$$

Using (3.57) (with $g(X) = f_{\mathsf{X}}(X)$) this becomes

$$P_{\mathsf{X}}(\mathcal{A}) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A}_k)} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k), \tag{3.60}$$

where $\tilde{f}_{\mathsf{X}}^{(k)} \colon \mathbb{R}^{kN} \to [0, \infty)$, $k \in \mathbb{N}_0$ are EVFs corresponding to f_{X} (cf. (3.55)) given by

$$\tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \begin{cases} f_{\mathsf{X}}(\chi_k(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)) & \text{if } (\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in \mathbb{X}_k, \\ 0 & \text{if } (\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \notin \mathbb{X}_k. \end{cases} \tag{3.61}$$

We will call EVFs corresponding to a PDF **equivalent vector densities (EVDs)**. According to (3.56) the set PDF can then be obtained from the EVDs by

$$f_{\mathsf{X}}(X) = \tilde{f}_{\mathsf{X}}^{(k)}(\theta_k(X)), \qquad |X| = k. \tag{3.62}$$

Example 3.11: Let us evaluate (3.60) for the events discussed in Example 3.10. To do so, we first rewrite (3.60) as (cf. (3.40))

$$P_{\mathsf{X}}(\mathcal{A}) = f_{\mathsf{X}}(\emptyset)\mathbf{1}_{\mathcal{A}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A}_k)} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k). \tag{3.63}$$

(a) For $\mathcal{A} = \mathcal{X}_0 = \{\emptyset\}$, we have $\mathbf{1}_{\mathcal{A}}(\emptyset) = 1$ and $\chi_k^{-1}(\mathcal{A}_k) = \emptyset$, i.e., the integrals in the sum in (3.63) are all equal to zero. Hence,

$$P_{\mathsf{X}}(\{\emptyset\}) = f_{\mathsf{X}}(\emptyset). \tag{3.64}$$

Note that because $P_{\mathsf{X}}(\{\emptyset\}) = \Pr(\mathsf{X} \in \{\emptyset\}) = \Pr(\mathsf{X} = \emptyset) = p_{|\mathsf{X}|}(0)$, this implies that

$$f_{\mathsf{X}}(\emptyset) = p_{|\mathsf{X}|}(0). \tag{3.65}$$

(b) In the case $\mathcal{A} = \mathcal{X}_n$, $n \in \mathbb{N}$, we have $\mathbf{1}_{\mathcal{A}}(\emptyset) = 0$ and $\mathcal{A}_k = \mathcal{X}_n \cap \mathcal{X}_k = \emptyset$ if $k \neq n$. Thus, only the integral term with $k = n$ does not vanish in the sum:

$$\begin{aligned} P_{\mathsf{X}}(\mathcal{X}_n) &= \frac{1}{n!} \int_{\chi_n^{-1}(\mathcal{X}_n)} \tilde{f}_{\mathsf{X}}^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mathrm{d}\lambda_{nN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) &(3.66)\\ &= \frac{1}{n!} \int_{\mathbb{X}_n} \tilde{f}_{\mathsf{X}}^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mathrm{d}\lambda_{nN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) &(3.67)\\ &= \frac{1}{n!} \int_{\mathbb{R}^{nN}} \tilde{f}_{\mathsf{X}}^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mathrm{d}\lambda_{nN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n), &(3.68) \end{aligned}$$

where the last expression holds since $(\mathbb{X}_n)^c$ is a null set. Note that because $P_{\mathsf{X}}(\mathcal{X}_n) = \Pr(|\mathsf{X}| = n) = p_{|\mathsf{X}|}(n)$, (3.68) states that the EVD $\tilde{f}_{\mathsf{X}}^{(n)}$ normalizes to $n! p_{|\mathsf{X}|}(n)$, i.e.,

$$\int_{\mathbb{R}^{nN}} \tilde{f}_{\mathsf{X}}^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mathrm{d}\lambda_{nN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = n! p_{|\mathsf{X}|}(n), \qquad n \in \mathbb{N}_0. \tag{3.69}$$

That is, any sequence of symmetric functions $\tilde{f}_{\mathsf{X}}^{(k)}: \mathbb{R}^{kN} \to [0,\infty)$, $k \in \mathbb{N}_0$, that satisfy (3.69) for an arbitrary CD constitute a valid set PDF f_{X} according to (3.62). In particular, since $\sum_{n=0}^{\infty} p_{|\mathsf{X}|}(n) = 1$, this implies that

$$f_{\mathsf{X}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathbb{R}^{kN}} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) = 1. \tag{3.70}$$

(c) For $\mathcal{A} = \mathcal{X}(A)$, with some closed $A \subseteq \mathbb{R}^N$, we have $\mathbf{1}_{\mathcal{A}}(\emptyset) = 1$ and $\chi_k^{-1}(\mathcal{A}_k) = A^k \cap \mathbb{X}_k$. Hence, (3.63) becomes

$$P_{\mathsf{X}}(\mathcal{X}(A)) = f_{\mathsf{X}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{A^k \cap \mathbb{X}_k} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) \tag{3.71}$$

$$= f_{\mathsf{X}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{A^k} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k), \tag{3.72}$$

where the last expression holds since $(\mathbb{X}_k)^c$ is a null set. Because $P_{\mathsf{X}}(\mathcal{X}(A)) = \Pr(\mathsf{X} \in \mathcal{X}(A)) = \Pr(\mathsf{X} \subseteq A) = \beta_{\mathsf{X}}(A)$, the BMF can therefore be calculated from the EVDs by

$$\beta_{\mathsf{X}}(A) = f_{\mathsf{X}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{A^k} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k). \tag{3.73}$$

3.6 Relation between RFSs and Random Vectors

In this section we show how any RFS can be related to a sequence of random vectors and determine the EVDs corresponding to the RFS (and thereby the set PDF) given the PDFs of the random vectors. We begin by introducing a **k-ary RFS** in Section 3.6.1, which is the result of converting a single random vector to a set and has exactly k elements. In Section 3.6.2 we then construct a general RFS (with a random number of elements) using a mixture model based on a sequence of k-ary RFSs.

3.6.1 k-ary RFS

Let $k, N \in \mathbb{N}$ arbitrary but fixed. Consider a kN-dimensional random vector

$$\mathbf{x} = (\mathbf{x}_1,\ldots,\mathbf{x}_k) \in \mathbb{X}_k, \qquad \mathbf{x}_i \in \mathbb{R}^N, i = 1,\ldots,k; \tag{3.74}$$

with corresponding probability space $(\mathbb{X}_k, \mathcal{B}(\mathbb{X}_k), P_{\mathbf{x}})$, dominating Lebesgue measure λ_{kN}, and PDF $f_{\mathbf{x}}(\boldsymbol{x}) = f_{\mathbf{x}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k)$. The realizations of the random vector $\mathbf{x}$ are then transformed to k-ary subsets of $\mathbb{R}^N$ using the VST χ_k (cf. (3.17)) given by

$$\chi_k(\boldsymbol{x}) = \chi_k(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k\}. \tag{3.75}$$

Since x is a random vector, the finite set

$$\mathsf{X} \triangleq \chi_k(\mathsf{x}) \in \mathcal{X}_k, \tag{3.76}$$

is random as well. We will refer to X as a **k-ary RFS**. Note that in contrast to a general RFS (cf. Definition 3.2), which takes realizations in $\mathcal{X}$, a k-ary RFS takes realizations only in the subspace $\mathcal{X}_k \subseteq \mathcal{X}$. That is, the outcome of X has exactly k elements. We follow our derivations in Section 3.4.1 and equip $\mathcal{X}_k$ with the σ-algebra $\mathcal{B}(\mathcal{X}_k)$ and the measure $\mu_{\mathcal{X}_k}$. Since, as shown in Lemma 3.7, χ_k is a measurable function, the probability measure P_X on $\mathcal{B}(\mathcal{X}_k)$ is induced by

$$P_\mathsf{X}(\mathcal{A}) = P_\mathsf{x}(\chi_k^{-1}(\mathcal{A})) = \int_{\chi_k^{-1}(\mathcal{A})} f_\mathsf{x}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \tag{3.77}$$

for any $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$.

Per definition (3.24), $\mu_{\mathcal{X}_k}(\mathcal{A}) = 0$ implies $\lambda_{kN}(\chi_k^{-1}(\mathcal{A})) = 0$, which due to the right-hand side of (3.77), implies $P_\mathsf{X}(\mathcal{A}) = 0$. Hence, $P_\mathsf{X} \ll \mu_\mathcal{X}$. Therefore, the RN theorem (Theorem 2.18) states that there exists an (a.e.) unique function $f_\mathsf{X} : \mathcal{X}_k \to [0, \infty)$, such that

$$P_\mathsf{X}(\mathcal{A}) = \int_\mathcal{A} f_\mathsf{X}(X) \mathrm{d}\mu_{\mathcal{X}_k}(X), \tag{3.78}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$. Using (3.36), this can be expressed as

$$P_\mathsf{X}(\mathcal{A}) = \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A})} \tilde{f}_\mathsf{X}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \tag{3.79}$$

where $\tilde{f}_\mathsf{X}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ is the EVD corresponding to $f_\mathsf{X}(X)$. Combining (3.77) and (3.79), we see that the EVD $\tilde{f}_\mathsf{X}$ is related to the vector PDF f_x by

$$\frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A})} \tilde{f}_\mathsf{X}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \int_{\chi_k^{-1}(\mathcal{A})} f_\mathsf{x}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \tag{3.80}$$

Note that $\frac{1}{k!}\tilde{f}_\mathsf{X}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \neq f_\mathsf{x}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ in general, since $\tilde{f}_\mathsf{X}$ is symmetric whereas f_x may be not. The key observation here is that the corresponding vector event $\chi_k^{-1}(\mathcal{A})$ of any event $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$ is symmetric, in the sense that,

$$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \chi_k^{-1}(\mathcal{A}) \Rightarrow (\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}) \in \chi_k^{-1}(\mathcal{A}), \tag{3.81}$$

for all permutations σ over $1, \ldots, k$. Because of this symmetry we have

$$\begin{aligned} P_\mathsf{X}(\mathcal{A}) &= \int_{\chi_k^{-1}(\mathcal{A})} f_\mathsf{x}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \\ &= \int_{\chi_k^{-1}(\mathcal{A})} f_\mathsf{x}(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \end{aligned} \tag{3.82}$$

We define the symmetric function

$$f_\mathsf{x}^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \triangleq \sum_\sigma f_\mathsf{x}(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}). \tag{3.83}$$

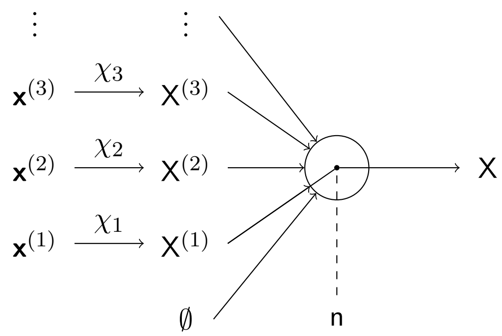

Fig. 3.2: *Setup for generating a general RFS with real elements X. A sequence of random vectors x_k is converted to a sequence of random sets X_k. The discrete random variable n selects an element of this set sequence as the outcome of X.*

where the sum is taken over all $k!$ permutations σ over $1, \dots, k$. Because of (3.82), $\frac{1}{k!} f_{\mathsf{x}}^*$ integrates to the probability measure:

$$\frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A})} f_{\mathsf{x}}^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = P_{\mathsf{X}}(\mathcal{A}). \tag{3.84}$$

Combining (3.84) with (3.79) we finally obtain

$$\tilde{f}_{\mathsf{X}}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = f_{\mathsf{x}}^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \sum_{\sigma} f_{\mathsf{x}}(\boldsymbol{x}_{\sigma_1}, \dots, \boldsymbol{x}_{\sigma_k}). \tag{3.85}$$

3.6.2 General RFS

In this section we will use the k-ary RFSs discussed in Section 3.6.1 to construct a general RFS $\mathsf{X} \subseteq \mathbb{R}^N$ with a random number of elements and determine the corresponding sequence of EVDs. Consider for this the setup depicted in Figure 3.2.

Let $\mathsf{n} \in \mathbb{N}_0$ be a discrete random variable, described by its PMF

$$p_n \triangleq p_{\mathsf{n}}(n) = P_{\mathsf{n}}(\mathsf{n} = n), \qquad n \in \mathbb{N}_0. \tag{3.86}$$

Furthermore, let $\mathsf{x}^{(k)} \in \mathbb{X}_k$, $k \in \mathbb{N}$, be a sequence of random vectors of increasing dimensionality kN, i.e.,

$$\mathsf{x}^{(k)} = (\mathsf{x}_1^{(k)}, \dots, \mathsf{x}_k^{(k)}), \qquad \mathsf{x}_i^{(k)} \in \mathbb{R}^N, \tag{3.87}$$

with associated PDFs

$$f_{\mathsf{x}^{(k)}}(\boldsymbol{x}) = f_{\mathsf{x}^{(k)}}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k). \tag{3.88}$$

As in Section 3.6.1, these random vectors are converted to k-ary RFSs $\mathsf{X}^{(k)}$ via VSTs $\chi_k \colon \mathbb{X}_k \to \mathcal{X}_k$, i.e.,

$$\mathsf{X}^{(k)} \triangleq \chi_k(\mathsf{x}^{(k)}) = \{\mathsf{x}_1^{(k)}, \dots, \mathsf{x}_k^{(k)}\}, \tag{3.89}$$

yielding a sequence of k-ary RFSs $\mathsf{X}^{(k)}$, $k \in \mathbb{N}$. The probability space corresponding to each $\mathsf{X}^{(k)}$ is $(\mathcal{X}_k, \mathcal{B}(\mathcal{X}_k), P_{\mathsf{X}^{(k)}})$. According to (3.85), the EVD of each $\mathsf{X}^{(k)}$ is

$$\tilde{f}_{\mathsf{X}^{(k)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \sum_{\sigma} f_{\mathsf{x}^{(k)}}(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}). \tag{3.90}$$

and the probability measures $P_{\mathsf{X}^{(k)}}$ can be calculated from the EVDs by

$$P_{\mathsf{X}^{(k)}}(\mathcal{A}) = \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A})} \tilde{f}_{\mathsf{X}^{(k)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \tag{3.91}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X}_k)$.

Moreover, for $k = 0$ we define the trivial RFS

$$\mathsf{X}^{(0)} \equiv \emptyset, \tag{3.92}$$

with sample space $\mathcal{X}_0 = \{\emptyset\}$, and trivial σ-algebra $\mathcal{B}(\mathcal{X}_0) = \{\emptyset, \{\emptyset\}\}$.

Using the discrete random variable n as a random index, the RFS X is then defined to be a randomly selected element of the sequence $\mathsf{X}^{(k)}$, i.e.,

$$\mathsf{X} \triangleq \mathsf{X}^{(\mathsf{n})}. \tag{3.93}$$

Clearly $\mathsf{X} \in \mathcal{X} = \bigcup_{k=0}^{\infty} \mathcal{X}_k$. We equip $\mathcal{X}$ with the Borel algebra $\mathcal{B}(\mathcal{X})$ generated by the hit-or-miss topology (Definition 3.1) and the set measure $\mu_{\mathcal{X}}$ as defined in (3.43). Since n is used as a selector, the probability measure associated with X, conditioned on $\mathsf{n} = n$, is

$$P_{\mathsf{X}}(\mathcal{A}|\mathsf{n} = n) = P_{\mathsf{X}^{(n)}}(\mathcal{A} \cap \mathcal{X}_n) = P_{\mathsf{X}^{(n)}}(\mathcal{A}_n), \tag{3.94}$$

for $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. For the special case $\mathsf{n} = 0$, (3.94) becomes

$$P_{\mathsf{X}}(\mathcal{A}|\mathsf{n} = 0) = P_{\mathsf{X}^{(0)}}(\mathcal{A}_0) = \mathbf{1}_{\mathcal{A}}(\emptyset). \tag{3.95}$$

Consequently, the total (unconditional) probability measure P_{X} is given by

$$P_{\mathsf{X}}(\mathcal{A}) = \sum_{n=0}^{\infty} p_n P_{\mathsf{X}}(\mathcal{A}|\mathsf{n} = n) \tag{3.96}$$

$$= p_0 \mathbf{1}_{\mathcal{A}}(\emptyset) + \sum_{n=1}^{\infty} p_n P_{\mathsf{X}^{(n)}}(\mathcal{A}_n). \tag{3.97}$$

Using (3.91), we can write (3.97) as

$$P_{\mathsf{X}}(\mathcal{A}) = p_0 \mathbf{1}_{\mathcal{A}}(\emptyset) + \sum_{k=1}^{\infty} p_k \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A}_k)} \tilde{f}_{\mathsf{X}^{(k)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \tag{3.98}$$

Note that, $\mu_{\mathcal{X}}(\mathcal{A}) = 0$ implies $\mathbf{1}_{\mathcal{A}}(\emptyset) = 0$ and $\lambda_{kN}(\chi_k^{-1}(\mathcal{A}_k)) = 0$ for all $k \in \mathbb{N}$, which due to (3.98) implies $P_{\mathsf{X}}(\mathcal{A}) = 0$. Hence, $P_{\mathsf{X}} \ll \mu_{\mathcal{X}}$. The RN theorem (Theorem 2.18) states that there exists an a.e. unique function $f_{\mathsf{X}} \colon \mathcal{X} \to [0, \infty)$, such that

$$P_{\mathsf{X}}(\mathcal{A}) = \int_{\mathcal{A}} f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}}(X) \tag{3.99}$$

for any $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. Using (3.63), this becomes

$$P_{\mathsf{X}}(\mathcal{A}) = f_{\mathsf{X}}(\emptyset)\mathbf{1}_{\mathcal{A}}(\emptyset) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A}_k)} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \tag{3.100}$$

Comparing with (3.98) we obtain

$$\tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = p_k \tilde{f}_{\mathsf{X}^{(k)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = p_k \sum_{\sigma} f_{\mathbf{x}^{(k)}}(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}). \tag{3.101}$$

Note that for the special case $k = 0$ this is

$$\tilde{f}_{\mathsf{X}}^{(0)} = f_{\mathsf{X}}(\emptyset) = p_0. \tag{3.102}$$

3.7 Joint Distributions

In this section, we will introduce and discuss joint versions of the statistical descriptors of RFSs we encountered so far – i.e., joint probability measures, CDs, BMFs, and PDFs – for the case of multiple RFSs. For ease of exposition, we will restrict ourselves to the case of two RFSs, although our results can be generalized in a straightforward manner to more than two RFSs.

3.7.1 Joint Probability Measure

Let (Ω, Σ, P) be a probability space, and let $\mathsf{X} \colon \Omega \to \mathcal{X}$ and $\mathsf{Y} \colon \Omega \to \mathcal{Y}$ be two RFSs, where

$$\mathcal{X} \triangleq \{X \subseteq \mathbb{R}^N : |X| \in \mathbb{N}_0\}, \tag{3.103}$$

$$\mathcal{Y} \triangleq \{Y \subseteq \mathbb{R}^M : |Y| \in \mathbb{N}_0\}, \tag{3.104}$$

with some $N, M \in \mathbb{N}$. Furthermore, let $\mathcal{B}(\mathcal{X})$ and $\mathcal{B}(\mathcal{Y})$ be the Borel algebras generated by the hit-or-miss topology (Definition 3.1). The probability measures are induced by (cf. (3.5))

$$P_{\mathsf{X}}(\mathcal{A}) = P(\mathsf{X}^{-1}(\mathcal{A})), \tag{3.105}$$

$$P_{\mathsf{Y}}(\mathcal{E}) = P(\mathsf{Y}^{-1}(\mathcal{E})), \tag{3.106}$$

for any $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$.

We consider the mapping $\mathbf{Z} \colon \Omega \to \mathcal{X} \times \mathcal{Y}$ defined by

$$\mathbf{Z}(\omega) \triangleq (\mathsf{X}(\omega), \mathsf{Y}(\omega)), \tag{3.107}$$

for any $\omega \in \Omega$. As the σ-algebra on $\mathcal{X} \times \mathcal{Y}$, we take the product σ-algebra $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ (Definition 2.6). Note that for any $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$,

$$\mathbf{Z}^{-1}(\mathcal{A} \times \mathcal{E}) = \{\omega \in \Omega : \mathsf{X}(\omega) \in \mathcal{A} \wedge \mathsf{Y}(\omega) \in \mathcal{E}\} = \mathsf{X}^{-1}(\mathcal{A}) \cap \mathsf{Y}^{-1}(\mathcal{E}). \tag{3.108}$$

Since $\mathsf{X}^{-1}(\mathcal{A}) \in \Sigma$ and $\mathsf{Y}^{-1}(\mathcal{E}) \in \Sigma$, $\mathsf{X}^{-1}(\mathcal{A}) \cap \mathsf{Y}^{-1}(\mathcal{E}) \in \Sigma$, and therefore $\mathbf{Z}^{-1}(\mathcal{A} \times \mathcal{E}) \in \Sigma$. Thus, $\mathbf{Z}$ is a $\Sigma/\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$-measurable function, and the **joint probability measure** $P_{\mathsf{X},\mathsf{Y}} \triangleq P_{\mathbf{Z}}$ on $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ is defined as the measure induced by $\mathbf{Z}$, i.e.,

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) \triangleq P(\mathbf{Z}^{-1}(\mathcal{A} \times \mathcal{E})) \tag{3.109}$$

$$= P(\mathsf{X}^{-1}(\mathcal{A}) \cap \mathsf{Y}^{-1}(\mathcal{E})) \tag{3.110}$$

$$= \Pr(\mathsf{X} \in \mathcal{A} \wedge \mathsf{Y} \in \mathcal{E}), \tag{3.111}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$. Note that $P_{\mathsf{X}}(\mathcal{A})$ and $P_{\mathsf{Y}}(\mathcal{E})$ are the marginal probability measures of $P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E})$, in the sense that

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{Y}) = P\left(\mathsf{X}^{-1}(\mathcal{A}) \cap \mathsf{Y}^{-1}(\mathcal{Y})\right) \tag{3.112}$$

$$= P\left(\mathsf{X}^{-1}(\mathcal{A}) \cap \Omega\right) \tag{3.113}$$

$$= P\left(\mathsf{X}^{-1}(\mathcal{A})\right) \tag{3.114}$$

$$= P_{\mathsf{X}}(\mathcal{A}), \tag{3.115}$$

and similarly

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{X} \times \mathcal{E}) = P_{\mathsf{Y}}(\mathcal{E}). \tag{3.116}$$

Definition 3.12: *The RFSs* X *and* Y *are called* ***statistically independent*** *if*

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) = P_{\mathsf{X}}(\mathcal{A}) P_{\mathsf{Y}}(\mathcal{E}), \tag{3.117}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ *and* $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$.

According to Theorem 2.9, the product measure $P_{\mathsf{X}} \times P_{\mathsf{Y}}$ is the unique measure satisfying (3.117). Therefore, X and Y are statistically independent if

$$P_{\mathsf{X},\mathsf{Y}} = P_{\mathsf{X}} \times P_{\mathsf{Y}}. \tag{3.118}$$

3.7.2 Joint Cardinality Distribution

Analogously to the CD for a single RFS (see Section 3.2), we define the joint CD for two RFSs as the restriction of the joint probability measure $P_{\mathsf{X},\mathsf{Y}}$ on the following events. For $k \in \mathbb{N}_0$, let $\mathcal{X}_k \in \mathcal{B}(\mathcal{X})$ and $\mathcal{Y}_k \in \mathcal{B}(\mathcal{Y})$ (cf. Lemma 3.3) be the events consisting of all finite subsets with exactly k elements, i.e. (cf. (3.6)),

$$\mathcal{X}_k \triangleq \{X \subseteq \mathbb{R}^N : |X| = k\}, \tag{3.119}$$

$$\mathcal{Y}_k \triangleq \{Y \subseteq \mathbb{R}^M : |Y| = k\}. \tag{3.120}$$

Note that the events $\mathcal{X}_k$ and $\mathcal{Y}_k$ form disjoint partitions of $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Definition 3.13: *The function* $p_{|\mathsf{X}|,|\mathsf{Y}|} \colon \mathbb{N}_0 \times \mathbb{N}_0 \to [0,1]$ *defined by*

$$p_{|\mathsf{X}|,|\mathsf{Y}|}(k,l) \triangleq P_{\mathsf{X},\mathsf{Y}}(\mathcal{X}_k \times \mathcal{Y}_l) = \Pr(|\mathsf{X}| = k \wedge |\mathsf{Y}| = l)$$

is called the ***joint CD*** *of the RFSs* X *and* Y. *That is, the joint CD is the joint PMF of the two discrete random variables* $|\mathsf{X}|$ *and* $|\mathsf{Y}|$.

If X and Y are statistically independent, it follows immediately from Definition 3.12 that the joint CD factorizes as

$$p_{|\mathsf{X}|,|\mathsf{Y}|}(k,l) = p_{|\mathsf{X}|}(k) p_{|\mathsf{Y}|}(l), \tag{3.121}$$

where $p_{|\mathsf{X}|}(k)$ and $p_{|\mathsf{Y}|}(l)$ are the CDs of X and Y, respectively.

3.7.3 Joint Belief Mass Function

Let $\mathcal{C}_X$ and $\mathcal{C}_Y$ be the collections of all closed subsets of $\mathbb{R}^N$ and $\mathbb{R}^M$, respectively. For any closed subsets $A \in \mathcal{C}_X$ and $B \in \mathcal{C}_Y$, we define the events $\mathcal{X}(A) \in \mathcal{B}(\mathcal{X})$ and $\mathcal{Y}(B) \in \mathcal{B}(\mathcal{Y})$ (cf. Lemma 3.5) as (cf. (3.10))

$$\mathcal{X}(A) \triangleq \{X \in \mathcal{X} : X \subseteq A\}, \tag{3.122}$$

$$\mathcal{Y}(B) \triangleq \{Y \in \mathcal{Y} : Y \subseteq B\}. \tag{3.123}$$

Thus, $\mathcal{X}(A)$ is the collection of all finite subsets of the closed set $A \in \mathcal{C}_X$, and similarly for $\mathcal{Y}(B)$.

Definition 3.14: *The function* $\beta_{\mathsf{X},\mathsf{Y}} \colon \mathcal{C}_X \times \mathcal{C}_Y \to [0,1]$ *defined by*

$$\beta_{\mathsf{X},\mathsf{Y}}(A,B) \triangleq P_{\mathsf{X},\mathsf{Y}}(\mathcal{X}(A) \times \mathcal{Y}(B)) = \Pr(\mathsf{X} \subseteq A \wedge \mathsf{Y} \subseteq B) \tag{3.124}$$

is called the ***joint BMF*** *of* X *and* Y.

If X and Y are statistically independent, it follows from Definition 3.12 that the joint BMF becomes

$$\beta_{\mathsf{X},\mathsf{Y}}(A,B) = \beta_{\mathsf{X}}(A)\beta_{\mathsf{Y}}(B), \tag{3.125}$$

where $\beta_{\mathsf{X}}(A)$ and $\beta_{\mathsf{Y}}(B)$ are the BMFs of X and Y respectively, as defined in Section 3.3, .

3.7.4 Joint Probability Density Function

For $k \in \mathbb{N}$, $l \in \mathbb{N}$, let $\mathbb{X}_k \subseteq \mathbb{R}^{kN}$ and $\mathbb{Y}_l \subseteq \mathbb{R}^{lM}$ be (cf. (3.16))

$$\mathbb{X}_k = \{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{R}^{kN} : \boldsymbol{x}_i \neq \boldsymbol{x}_j \text{ for all } i \neq j\} \tag{3.126}$$

$$\mathbb{Y}_l = \{(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) \in \mathbb{R}^{lM} : \boldsymbol{y}_i \neq \boldsymbol{y}_j \text{ for all } i \neq j\}. \tag{3.127}$$

Furthermore, let $\chi_k\colon \mathbb{X}_k \to \mathcal{X}_k$ and $\xi_l\colon \mathbb{Y}_l \to \mathcal{Y}_l$ be VSTs as defined in (3.17), i.e.,

$$\chi_k(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k) = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k\}, \tag{3.128}$$

$$\xi_l(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_l) = \{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_l\}, \tag{3.129}$$

where $\boldsymbol{x}_i \in \mathbb{R}^N$ and $\boldsymbol{y}_i \in \mathbb{R}^M$. Finally, let $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ be set measures as defined in (3.43) on $\mathcal{B}(\mathcal{X})$ and $\mathcal{B}(\mathcal{Y})$, i.e.,

$$\mu_{\mathcal{X}}(\mathcal{A}) = \sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}(\mathcal{A}_k), \qquad \mu_{\mathcal{X}_k}(\mathcal{A}_k) = \begin{cases} \mathbf{1}_{\mathcal{A}_0}(\emptyset) & \text{if } k=0, \\ \frac{1}{k!}\lambda_{kN}(\chi_k^{-1}(\mathcal{A}_k)) & \text{if } k\geq 1, \end{cases} \tag{3.130}$$

$$\mu_{\mathcal{Y}}(\mathcal{E}) = \sum_{k=0}^{\infty} \mu_{\mathcal{Y}_l}(\mathcal{E}_l), \qquad \mu_{\mathcal{Y}_l}(\mathcal{E}_l) = \begin{cases} \mathbf{1}_{\mathcal{E}_0}(\emptyset) & \text{if } l=0, \\ \frac{1}{l!}\lambda_{lM}(\xi_l^{-1}(\mathcal{E}_l)) & \text{if } l\geq 1, \end{cases} \tag{3.131}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$.

As the set measure on the product σ-algebra $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$ we take the product measure $\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$, i.e., the unique measure satisfying (cf. Theorem 2.9)

$$\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}(\mathcal{A} \times \mathcal{E}) = \mu_{\mathcal{X}}(\mathcal{A})\mu_{\mathcal{Y}}(\mathcal{E}), \tag{3.132}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$. Provided that $P_{\mathsf{X},\mathsf{Y}} \ll \mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$, the RND theorem (Theorem 2.18) states that there exists an (a.e.) unique measurable function $f_{\mathsf{X},\mathsf{Y}}\colon \mathcal{X} \times \mathcal{Y} \to [0,\infty)$ such that

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) = \int_{\mathcal{A}\times\mathcal{E}} f_{\mathsf{X},\mathsf{Y}}(X,Y)\mathrm{d}(\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}})(X,Y), \tag{3.133}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$. The function $f_{\mathsf{X},\mathsf{Y}}(X,Y)$ is the RND of $P_{\mathsf{X},\mathsf{Y}}$ with respect to $\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$, i.e.,

$$f_{\mathsf{X},\mathsf{Y}}(X,Y) = \frac{\mathrm{d}P_{\mathsf{X},\mathsf{Y}}}{\mathrm{d}(\mu_{\mathcal{X}} \times \mu_{\mathcal{Y}})}(X,Y), \tag{3.134}$$

and is called the **joint PDF** of the two RFSs X and Y. Fubini's theorem (Theorem 2.19) allows us to write (3.133) as

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) = \int_{\mathcal{E}}\int_{\mathcal{A}} f_{\mathsf{X},\mathsf{Y}}(X,Y)\mathrm{d}\mu_{\mathcal{X}}(X)\mathrm{d}\mu_{\mathcal{Y}}(Y), \tag{3.135}$$

where the order of integration does not matter.

Similarly to (3.49) we write the joint PDF in a piecewise fashion as

$$f_{\mathsf{X},\mathsf{Y}}(X,Y) = f_{\mathsf{X},\mathsf{Y}}^{(k,l)}(X,Y) \qquad \text{if } |X| = k \text{ and } |Y| = l, \tag{3.136}$$

for $k \in \mathbb{N}_0$, $l \in \mathbb{N}_0$, and $f^{(k,l)}_{\mathsf{X},\mathsf{Y}} : \mathcal{X}_k \times \mathcal{Y}_l \to [0, \infty)$. Analogously to (3.50)–(3.52), by inserting the set measures $\mu_{\mathcal{X}}$ (3.130) and $\mu_{\mathcal{Y}}$ (3.131) in (3.135) we obtain

$$
\begin{aligned}
P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) &= \int_{\mathcal{E}} \int_{\mathcal{A}} f_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{X}}(X) \mathrm{d}\mu_{\mathcal{Y}}(Y) && (3.137)\\
&= \int_{\mathcal{E}} \Big(\sum_{k=0}^{\infty} \int_{\mathcal{A}_k} f_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{X}_k}(X) \Big) \mathrm{d}\mu_{\mathcal{Y}}(Y) && (3.138)\\
&= \sum_{k=0}^{\infty} \int_{\mathcal{A}_k} \Big(\int_{\mathcal{E}} f_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{Y}}(Y) \Big) \mathrm{d}\mu_{\mathcal{X}_k}(X) && (3.139)\\
&= \sum_{k=0}^{\infty} \int_{\mathcal{A}_k} \Big(\sum_{l=0}^{\infty} \int_{\mathcal{E}_l} f_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{Y}_l}(Y) \Big) \mathrm{d}\mu_{\mathcal{X}_k}(X) && (3.140)\\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \int_{\mathcal{E}_l} \int_{\mathcal{A}_k} f_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{X}_k}(X) \mathrm{d}\mu_{\mathcal{Y}_l}(Y) && (3.141)\\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \int_{\mathcal{E}_l} \int_{\mathcal{A}_k} f^{(k,l)}_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{X}_k}(X) \mathrm{d}\mu_{\mathcal{Y}_l}(Y), && (3.142)
\end{aligned}
$$

where in the last step (3.136) has been used. Continuing, we use the push-forward property (3.31) for X and Y separately, resulting in

$$
\begin{aligned}
P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \int_{\mathcal{E}_l} \int_{\mathcal{A}_k} f^{(k,l)}_{\mathsf{X},\mathsf{Y}}(X, Y) \mathrm{d}\mu_{\mathcal{X}_k}(X) \mathrm{d}\mu_{\mathcal{Y}_l}(Y) && (3.143)\\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \int_{\mathcal{E}_l} \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A}_k)} f^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), Y) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\mu_{\mathcal{Y}_l}(Y) && (3.144)\\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{l!} \int_{\xi_l^{-1}(\mathcal{E}_l)} \frac{1}{k!} \int_{\chi_k^{-1}(\mathcal{A}_k)} f^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \xi(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l))\\
&\qquad \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{lM}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l). && (3.145)
\end{aligned}
$$

Similary to the EVDs of a single RFS (cf. (3.61) and (3.62)), we define the **joint EVDs** $\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}} : \mathbb{R}^{kN} \times \mathbb{R}^{lM} \to [0, \infty)$, for $k, l \in \mathbb{N}_0$, corresponding to the joint PDF $f_{\mathsf{X},\mathsf{Y}}(X, Y)$ by

$$
\begin{aligned}
\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) &\triangleq f^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \xi(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l)) && (3.146)\\
&= f_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \xi(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l)), && (3.147)
\end{aligned}
$$

for $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{X}_k$ and $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) \in \mathbb{Y}_l$. Note that due to the symmetry of the VSTs χ_k and ξ_l, the joint EVDs $\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}$ are partially symmetric in the following sense: For all $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in \mathbb{X}_k$ and $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) \in \mathbb{Y}_l$,

$$
\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) = \tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_{\sigma_1}, \ldots, \boldsymbol{x}_{\sigma_k}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l), \qquad (3.148)
$$

for all permutations σ over $1, \ldots, k$, and

$$
\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) = \tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_{\sigma_1}, \ldots, \boldsymbol{y}_{\sigma_l}) \qquad (3.149)
$$

for all permutations σ over $1, \ldots, l$.

The joint PDF can then be calculated from the joint EVDs by

$$f_{\mathsf{X},\mathsf{Y}}(X, Y) = \tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\theta_k(X), \psi_l(Y)), \qquad |X| = k, |Y| = l, \tag{3.150}$$

where $\theta_k\colon \mathcal{X}_k \to \mathbb{X}_k$ as in (3.34), and likewise $\psi_l\colon \mathcal{Y}_l \to \mathbb{Y}_l$ given by

$$\psi_l(Y) \triangleq (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l), \qquad \boldsymbol{y}_i \in Y, \text{ such that } \boldsymbol{y}_1 < \boldsymbol{y}_2 < \cdots < \boldsymbol{y}_k, \tag{3.151}$$

where $<$ is the lexicographical order on $\mathbb{R}^M$. That is, there is a one-to-one correspondence between the joint set PDF $f_{\mathsf{X},\mathsf{Y}}(X, Y)$ and a family (indexed by k and l) of joint EVDs $\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}$.

Inserting the joint EVDs in (3.145) we finally obtain

$$\begin{aligned} P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!l!} \int_{\xi_l^{-1}(\mathcal{E}_l)} \int_{\chi_k^{-1}(\mathcal{A}_k)} &\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) \\ &\mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)\mathrm{d}\lambda_{lM}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l). \end{aligned} \tag{3.152}$$

This is the analogous expression to (3.60) for the case of a single RFS.

Marginalization

Provided that $P_{\mathsf{X}} \ll \mu_{\mathcal{X}}$, the PDF $f_{\mathsf{X}}(X) = \frac{\mathrm{d}P_{\mathsf{X}}}{\mathrm{d}\mu_{\mathcal{X}}}(X)$ exists and therefore

$$P_{\mathsf{X}}(\mathcal{A}) = \int_{\mathcal{A}} f_{\mathsf{X}}(X)\mathrm{d}\mu_{\mathcal{X}}(X), \tag{3.153}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. Because of (3.115) and (3.135), we obtain on the other hand

$$P_{\mathsf{X}}(\mathcal{A}) = P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{Y}) = \int_{\mathcal{A}} \left(\int_{\mathcal{Y}} f_{\mathsf{X},\mathsf{Y}}(X, Y)\mathrm{d}\mu_{\mathcal{Y}}(Y) \right) \mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.154}$$

By comparing with (3.153), it immediately follows that

$$f_{\mathsf{X}}(X) = \int_{\mathcal{Y}} f_{\mathsf{X},\mathsf{Y}}(X, Y)\mathrm{d}\mu_{\mathcal{Y}}(Y). \tag{3.155}$$

By an analogous argument, we obtain

$$f_{\mathsf{Y}}(Y) = \int_{\mathcal{X}} f_{\mathsf{X},\mathsf{Y}}(X, Y)\mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.156}$$

That is, $f_{\mathsf{X}}(X)$ and $f_{\mathsf{Y}}(Y)$ are the **marginal PDFs** of the joint PDF $f_{\mathsf{X},\mathsf{Y}}(X, Y)$.

Using (3.155), the EVDs $\tilde{f}^{(k)}_{\mathsf{X}}$ of the RFS X can be calculated from the joint EVDs $\tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}$ as follows:

$$\tilde{f}^{(k)}_{\mathsf{X}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = f_{\mathsf{X}}(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)) = \int_{\mathcal{Y}} f_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), Y)\mathrm{d}\mu_{\mathcal{Y}}(Y). \tag{3.157}$$

Inserting $\mu_{\mathcal{Y}}$ from (3.131) and using the push-forward property (3.31) for Y we further obtain

$$\tilde{f}^{(k)}_{\mathsf{X}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) = \sum_{l=0}^{\infty} \int_{\mathcal{Y}_l} f_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k), Y)\mathrm{d}\mu_{\mathcal{Y}_l}(Y) \tag{3.158}$$

$$= \sum_{l=0}^{\infty} \frac{1}{l!} \int_{\xi_l^{-1}(\mathcal{Y}_l)} f_{\mathsf{X},\mathsf{Y}}(\chi_k(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k), \xi(\boldsymbol{y}_1,\dots,\boldsymbol{y}_l))\mathrm{d}\lambda_{lM}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_l) \tag{3.159}$$

$$= \sum_{l=0}^{\infty} \frac{1}{l!} \int_{\mathbb{R}^{lM}} \tilde{f}^{(k,l)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{y}_1,\dots,\boldsymbol{y}_l)\mathrm{d}\lambda_{lM}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_l). \tag{3.160}$$

Statistical Independence

A proof of the following theorem can be found in Appendix A.5.

Theorem 3.15: Let X and Y be two RFSs with joint PDF $f_{\mathsf{X},\mathsf{Y}}(X,Y)$ and marginal PDFs $f_{\mathsf{X}}(X)$ and $f_{\mathsf{Y}}(Y)$. Then, X and Y are statistically independent (see Definition 3.12) iff

$$f_{\mathsf{X},\mathsf{Y}}(X,Y) = f_{\mathsf{X}}(X) f_{\mathsf{Y}}(Y) \qquad \text{a.e.}$$

Relation to Joint Cardinality Distribution

The joint cardinality distribution defined in Section 3.7.2 can be calculated from the joint EVDs by inserting $\mathcal{A} = \mathcal{X}_n = \{X \subseteq \mathbb{R}^N : |X| = n\}$ and $\mathcal{E} = \mathcal{Y}_m = \{Y \subseteq \mathbb{R}^M : |Y| = m\}$ in (3.152). Here, the regions of integration become

$$\chi_k^{-1}(\mathcal{A}_k) = \chi_k^{-1}(\mathcal{X}_n \cap \mathcal{X}_k) = \begin{cases} \mathbb{X}_n & \text{if } k = n, \\ \emptyset & \text{if } k \neq n, \end{cases} \tag{3.161}$$

and similarly

$$\xi_l^{-1}(\mathcal{E}_l) = \xi_l^{-1}(\mathcal{Y}_m \cap \mathcal{Y}_l) = \begin{cases} \mathbb{Y}_m & \text{if } l = m, \\ \emptyset & \text{if } l \neq m. \end{cases} \tag{3.162}$$

Thus, only the integral terms in the sum with $k = n$ and $l = m$ do not vanish and we obtain

$$p_{|\mathsf{X}|,|\mathsf{Y}|}(n,m) = P_{\mathsf{X},\mathsf{Y}}(\mathcal{X}_n \times \mathcal{Y}_m) \tag{3.163}$$

$$= \frac{1}{n!m!} \int_{\mathbb{Y}_m} \int_{\mathbb{X}_n} \tilde{f}^{(n,m)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{y}_1,\dots,\boldsymbol{y}_m)$$
$$\lambda_{nN}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)\mathrm{d}\lambda_{mM}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_m) \tag{3.164}$$

$$= \frac{1}{n!m!} \int_{\mathbb{R}^{mM}} \int_{\mathbb{R}^{nN}} \tilde{f}^{(n,m)}_{\mathsf{X},\mathsf{Y}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{y}_1,\dots,\boldsymbol{y}_m)$$
$$\lambda_{nN}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)\mathrm{d}\lambda_{mM}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_m), \tag{3.165}$$

where the last result follows since $(\mathbb{X}_n)^c$ and $(\mathbb{Y}_m)^c$ are null sets.

Relation to Joint Belief Mass Function

The joint belief mass function defined in Section 3.7.3 can be obtained from the joint EVDs by inserting $\mathcal{A} = \mathcal{X}(A) = \{X \in \mathcal{X} : X \subseteq A\}$ and $\mathcal{E} = \mathcal{Y}(B) = \{Y \in \mathcal{Y} : Y \subseteq B\}$ in (3.152), where A and B are closed subsets of $\mathbb{R}^N$ and $\mathbb{R}^M$, respectively. Here, the regions of integration become $\chi_k^{-1}(\mathcal{A}_k) = A^k \cap \mathbb{X}_k$ and $\xi_l^{-1}(\mathcal{E}_l) = B^l \cap \mathbb{Y}_l$ (cf. Example 3.8(b)) and we obtain

$$\beta_{\mathsf{X},\mathsf{Y}}(A, B) = P_{\mathsf{X},\mathsf{Y}}(\mathcal{X}(A) \times \mathcal{Y}(B)) \tag{3.166}$$

$$= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!l!} \int_{B^l \cap \mathbb{Y}_l} \int_{A^k \cap \mathbb{X}_k} \tilde{f}_{\mathsf{X},\mathsf{Y}}^{(k,l)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l)$$

$$\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{lM}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l) \tag{3.167}$$

$$= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!l!} \int_{B^l} \int_{A^k} \tilde{f}_{\mathsf{X},\mathsf{Y}}^{(k,l)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_l)$$

$$\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{lM}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l), \tag{3.168}$$

where the last result follows since $(\mathbb{X}_k)^c$ and $(\mathbb{Y}_l)^c$ are null sets.

3.8 Conditional Distributions

3.8.1 Conditional Probability Density Functions

In order to avoid the involved theory of conditional probability measures [42, Chapter 10], we begin by defining conditional PDFs. Let $f_{\mathsf{X},\mathsf{Y}}(X, Y)$ be a continuous joint PDF of two RFSs X and Y. Then, the **conditional PDF** of X, given the outcome $\mathsf{Y} = Y$, is defined for $f_{\mathsf{Y}}(Y) \neq 0$ by

$$f_{\mathsf{X}|\mathsf{Y}}(X|Y) \triangleq \frac{f_{\mathsf{X},\mathsf{Y}}(X, Y)}{f_{\mathsf{Y}}(Y)}, \tag{3.169}$$

and similarly for $f_{\mathsf{Y}|\mathsf{X}}(Y|X)$. Note that in general the conditional PDF is only defined for almost all $Y \in \mathcal{Y}$ (or $X \in \mathcal{X}$, respectively). Combining (3.169) and the analogous expression of $f_{\mathsf{Y}|\mathsf{X}}(Y|X)$ yields

$$f_{\mathsf{X}|\mathsf{Y}}(X|Y) = f_{\mathsf{Y}|\mathsf{X}}(Y|X) \frac{f_{\mathsf{X}}(X)}{f_{\mathsf{Y}}(Y)}. \tag{3.170}$$

This is *Bayes' theorem for RFSs.*

3.8.2 Conditional Probability Measure

The conditional PDF $f_{\mathsf{X}|\mathsf{Y}}(X|Y)$ establishes a **conditional probability measure** for X by

$$P_{\mathsf{X}|\mathsf{Y}}(\mathcal{A}|Y) = \int_{\mathcal{A}} f_{\mathsf{X}|\mathsf{Y}}(X|Y) \mathrm{d}\mu_{\mathcal{X}}(X), \tag{3.171}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. This conditional probability measure is related to the joint probability measure (cf. (3.135)) by

$$\int_{\mathcal{E}} P_{\mathsf{X}|\mathsf{Y}}(\mathcal{A}|Y) f_{\mathsf{Y}}(Y) \mathrm{d}\mu_{\mathcal{Y}}(Y) = \int_{\mathcal{E}} \left(\int_{\mathcal{A}} f_{\mathsf{X}|\mathsf{Y}}(X|Y) \mathrm{d}\mu_{\mathcal{X}}(X) \right) f_{\mathsf{Y}}(Y) \mathrm{d}\mu_{\mathcal{Y}}(Y) \tag{3.172}$$

$$= \int_{\mathcal{E}} \int_{\mathcal{A}} f_{\mathsf{X},\mathsf{Y}}(X,Y) \mathrm{d}\mu_{\mathcal{X}}(X) \mathrm{d}\mu_{\mathcal{Y}}(Y) \tag{3.173}$$

$$= P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}), \tag{3.174}$$

where in (3.173) Fubini's theorem (Theorem 2.19) and (3.169) have been used. Furthermore, we have

$$\int_{\mathcal{Y}} f_{\mathsf{X}|\mathsf{Y}}(X|Y) f_{\mathsf{Y}}(Y) \mathrm{d}\mu_{\mathcal{Y}}(Y) = \int_{\mathcal{Y}} f_{\mathsf{X},\mathsf{Y}}(X,Y) \mathrm{d}\mu_{\mathcal{Y}}(Y) = f_{\mathsf{X}}(X), \tag{3.175}$$

where (3.169) and the marginalization equation (3.155) have been used. This is the *total probability theorem for RFSs.*

3.8.3 Conditional Cardinality Distribution

Restricting the conditional probability measure in (3.171) to the events $\mathcal{A} = \mathcal{X}_n$, we define the conditional CD of X given $\mathsf{Y} = Y$ by

$$p_{|\mathsf{X}||\mathsf{Y}}(n|Y) \triangleq P_{\mathsf{X}|\mathsf{Y}}(\mathcal{X}_n|Y) = \int_{\mathcal{X}_n} f_{\mathsf{X}|\mathsf{Y}}(X|Y) \mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.176}$$

3.8.4 Conditional Belief Mass Function

Similarly, restricting the conditional probability measure in (3.171) to the events $\mathcal{A} = \mathcal{X}(A)$, the conditional BMF of X given $\mathsf{Y} = Y$ is defined as

$$\beta_{\mathsf{X}|\mathsf{Y}}(A|Y) \triangleq P_{\mathsf{X}|\mathsf{Y}}(\mathcal{X}(A)|Y) = \int_{\mathcal{X}(A)} f_{\mathsf{X}|\mathsf{Y}}(X|Y) \mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.177}$$

3.8.5 Statistical Independence

From Theorem 3.15 and (3.169), it immediately follows that two RFSs X and Y are statistically independent iff

$$f_{\mathsf{X}|\mathsf{Y}}(X|Y) = f_{\mathsf{X}}(X), \tag{3.178}$$

or, equivalently, iff

$$f_{\mathsf{Y}|\mathsf{X}}(Y|X) = f_{\mathsf{Y}}(Y). \tag{3.179}$$

3.8.6 Conditional Independence

In the case of joint PDFs of more than two RFSs, one can condition on any of these RFSs by straightforward generalization of (3.169). As an example, for three RFSs X, Y, Z with joint PDF $f_{\mathsf{X},\mathsf{Y},\mathsf{Z}}(X,Y,Z)$, the conditional joint PDF of X and Y given the observation $\mathsf{Z} = Z$ is

$$f_{\mathsf{X},\mathsf{Y}|\mathsf{Z}}(X,Y|Z) = \frac{f_{\mathsf{X},\mathsf{Y},\mathsf{Z}}(X,Y,Z)}{f_{\mathsf{Z}}(Z)}. \tag{3.180}$$

The RFSs X and Y are called **conditionally independent given** $\mathsf{Z} = Z$ if (cf. Theorem 3.15)

$$f_{\mathsf{X},\mathsf{Y}|\mathsf{Z}}(X,Y|Z) = f_{\mathsf{X}|\mathsf{Z}}(X|Z) f_{\mathsf{Y}|\mathsf{Z}}(Y|Z), \tag{3.181}$$

or equivalently (cf. (3.178))

$$f_{\mathsf{X}|\mathsf{Y},\mathsf{Z}}(X|Y,Z) = f_{\mathsf{X}|\mathsf{Z}}(X|Z), \tag{3.182}$$

or equivalently (cf. (3.179))

$$f_{\mathsf{Y}|\mathsf{X},\mathsf{Z}}(Y|X,Z) = f_{\mathsf{Y}|\mathsf{Z}}(Y|Z). \tag{3.183}$$

3.9 Expectation for RFS

Let X be a RFS with PDF $f_{\mathsf{X}}(X)$ and let $g\colon \mathcal{X} \to \mathbb{R}$ be a measurable function. The expectation of $g(\mathsf{X})$ is defined by (cf. (2.39) and (2.52))

$$\mathsf{E}[g(\mathsf{X})] \triangleq \int_{\mathcal{X}} g(X) \mathrm{d}P_{\mathsf{X}}(X) = \int_{\mathcal{X}} g(X) f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.184}$$

Note that for RFSs there is no direct analog to the 1-st order moment of a random vector since $g(X) = X$ is not a real-valued function. In Section 4.4 we introduce a 1-st order moment, called the probability hypothesis density.

Example 3.16: Let $g(X) = |X|$. Since the EVFs of $|X|$ are $|\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)| = k$, (3.184) yields upon using (3.57)

$$\mathsf{E}[|\mathsf{X}|] = \int_{\mathcal{X}} |X| f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}}(X) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{\mathbb{R}^{kN}} k \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k). \tag{3.185}$$

Continuing, by using (3.69) we obtain

$$\mathsf{E}[|\mathsf{X}|] = \sum_{k=0}^{\infty} \frac{k}{k!} k! p_{|\mathsf{X}|}(k) = \sum_{k=0}^{\infty} k p_{|\mathsf{X}|}(k). \tag{3.186}$$

Analogously to (3.184), we define the **joint expectation** of a measurable function $f\colon \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ with respect to the joint PDF $f_{\mathsf{X},\mathsf{Y}}(X,Y)$ as

$$\mathsf{E}[f(\mathsf{X},\mathsf{Y})] \triangleq \int_{\mathcal{Y}}\int_{\mathcal{X}} f(X,Y) f_{\mathsf{X},\mathsf{Y}}(X,Y)\mathrm{d}\mu_{\mathcal{X}}(X)\mathrm{d}\mu_{\mathcal{Y}}(Y). \tag{3.187}$$

Let X and Y be two statistically independent RFSs, i.e. (cf. Theorem 3.15), $f_{\mathsf{X},\mathsf{Y}}(X,Y) = f_{\mathsf{X}}(X)f_{\mathsf{Y}}(Y)$. Furthermore, let $f(X,Y) = g(X)h(Y)$, where $g\colon \mathcal{X} \to \mathbb{R}$ and $h\colon \mathcal{Y} \to \mathbb{R}$ are two measurable functions. Then, (3.187) becomes

$$\begin{aligned}
\mathsf{E}[g(\mathsf{X})h(\mathsf{Y})] &= \int_{\mathcal{Y}}\int_{\mathcal{X}} g(X)h(Y)f_{\mathsf{X}}(X)f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{X}}(X)\mathrm{d}\mu_{\mathcal{Y}}(Y) &(3.188)\\
&= \int_{\mathcal{X}} g(X)f_{\mathsf{X}}(X)\mathrm{d}\mu_{\mathcal{X}}(X)\int_{\mathcal{Y}} h(Y)f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{Y}}(Y) &(3.189)\\
&= \mathsf{E}[g(\mathsf{X})]\mathsf{E}[h(\mathsf{Y})]. &(3.190)
\end{aligned}$$

That is, the expectation of the product $g(\mathsf{X})h(\mathsf{Y})$ of two statistically independent RFSs X and Y is the product of the individual expectations $\mathsf{E}[g(\mathsf{X})]$ and $\mathsf{E}[h(\mathsf{Y})]$.

The **conditional expectation** of a measurable function $g\colon \mathcal{X} \to \mathbb{R}$ with respect to a continuous conditional PDF $f_{\mathsf{X}|\mathsf{Y}}(X|Y)$ is defined by

$$\mathsf{E}[g(\mathsf{X})|Y] \triangleq \int_{\mathcal{X}} g(X)f_{\mathsf{X}|\mathsf{Y}}(X|Y)\mathrm{d}\mu_{\mathcal{X}}(X). \tag{3.191}$$

Let X and Y be two statistically independent RFSs. Then, according to (3.178), $f_{\mathsf{X}|\mathsf{Y}}(X|Y) = f_{\mathsf{X}}(X)$ and (3.191) becomes

$$\mathsf{E}[g(\mathsf{X})|Y] = \int_{\mathcal{X}} g(X)f_{\mathsf{X}}(X)\mathrm{d}\mu_{\mathcal{X}}(X) = \mathsf{E}[g(\mathsf{X})]. \tag{3.192}$$

That is, the conditional expectation reduces to the unconditional expectation. Finally, by inserting $f(X,Y) = g(X)$ and $f_{\mathsf{X},\mathsf{Y}}(X,Y) = f_{\mathsf{X}|\mathsf{Y}}(X|Y)f_{\mathsf{Y}}(Y)$ (cf. (3.169)) in (3.187) we can calculate the expectation of $g(\mathsf{X})$ from the conditional expectation $\mathsf{E}[g(\mathsf{X})|Y]$ as follows:

$$\begin{aligned}
\mathsf{E}[g(\mathsf{X})] &= \int_{\mathcal{Y}}\int_{\mathcal{X}} g(X)f_{\mathsf{X}|\mathsf{Y}}(X|Y)f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{X}}(X)\mathrm{d}\mu_{\mathcal{Y}}(Y) &(3.193)\\
&= \int_{\mathcal{Y}}\left(\int_{\mathcal{X}} g(X)f_{\mathsf{X}|\mathsf{Y}}(X|Y)\mathrm{d}\mu_{\mathcal{X}}(X)\right)f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{Y}}(Y) &(3.194)\\
&= \int_{\mathcal{Y}} \mathsf{E}[g(\mathsf{X})|Y]f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{Y}}(Y). &(3.195)
\end{aligned}$$

This is the total probability theorem for the expectations of RFSs (cf. (3.175)).

Chapter 4

Finite Set Statistics (FISST)

This chapter provides an introduction to a mathematical framework for treating RFSs, called **Finite Set Statistics** (**FISST**), which was developed by Mahler in [3,6,16,48–50]. In particular a very readable introduction is [16]. Unless stated otherwise, the material of this chapter is taken from [16, Chapter 11].

The FISST framework builds upon and extends the measure-theoretic formalism for RFSs considered in Chapter 3. Section 4.1 introduces the set integral, which is essentially just a notation for the Lebesgue integral with respect to the measure $\mu_{\mathcal{X}}$ on $\mathcal{B}(\mathcal{X})$ as defined in (3.43). In Section 4.2, a new descriptor of RFSs, the probability generating functional, is defined and discussed. In Section 4.3, the concept of functional derivatives is established. Section 4.4 defines a first order-moment of an RFS known as the probability hypothesis density. Furthermore, at the end of that section, a pictorial summary of all the descriptors of RFSs considered in this work and their relationships is provided. Finally, in Section 4.5, we present and discuss four common types of RFSs.

4.1 Set Integral

Let $\mathcal{X}$ be the collection of all finite subsets of $\mathbb{R}^N$ (3.1), equipped with the Borel algebra $\mathcal{B}(\mathcal{X})$ (3.4) and the set measure $\mu_{\mathcal{X}}$ given by (3.43)

$$\mu_{\mathcal{X}}(\mathcal{A}) = \sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}(\mathcal{A}_k), \qquad \mu_{\mathcal{X}_k}(\mathcal{A}_k) = \begin{cases} \mathbf{1}_{\mathcal{A}_0}(\emptyset) & \text{if } k = 0, \\ \frac{1}{k!}\lambda_{kN}(\chi_k^{-1}(\mathcal{A}_k)) & \text{if } k \geq 1. \end{cases} \tag{4.1}$$

The **set integral** of a $\mu_{\mathcal{X}}$-integrable function $f\colon \mathcal{X} \to \mathbb{R}$ over a closed subset $A \subseteq \mathbb{R}^N$ is defined as

$$\int_A f(X)\delta X \triangleq \int_{\mathcal{X}(A)} f(X)\mathrm{d}\mu_{\mathcal{X}}(X), \tag{4.2}$$

where $\mathcal{X}(A) = \{X \in \mathcal{X} : X \subseteq A\}$ (cf. (3.10)). That is, the set integral is just a different notation for the Lebesgue integral with respect to $\mu_{\mathcal{X}}$ over $\mathcal{X}(A)$ (the collection of all finite subsets of a closed set A). Using (3.57), we can express the set integral as

$$\int_A f(X)\delta X = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{A^k} \tilde{f}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k), \tag{4.3}$$

where $\tilde{f}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$, $k \in \mathbb{N}_0$ are the EVFs corresponding to $f(X)$ (cf. (3.55)) and

$$\lambda_{kN}[\chi_k^{-1}(\mathcal{X}(A))] = \lambda_{kN}(A^k \cap \mathbb{X}_k) = \lambda_{kN}(A^k), \tag{4.4}$$

has been used.

It follows that set integrals are linear, i.e.,

$$\int_A (a_1 f_1(X) + a_2 f_2(X))\, \delta X = a_1 \int_A f_1(X)\delta X + a_2 \int_A f_2(X)\delta X, \tag{4.5}$$

for any $a_1, a_2 \in \mathbb{R}$ and any $\mu_{\mathcal{X}}$-integrable functions $f_1, f_2 \colon \mathcal{X} \to \mathbb{R}$. However, since $\mathcal{X}(A_1 \cup A_2) \neq \mathcal{X}(A_1) \cup \mathcal{X}(A_2)$, the set integral is not additive in A, i.e., if $A_1 \cap A_2 = \emptyset$ then

$$\int_{A_1 \cup A_2} f(X)\delta X \neq \int_{A_1} f(X)\delta X + \int_{A_2} f(X)\delta X. \tag{4.6}$$

Iterated set integrals of functions with two finite sets as argument $f \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ are defined by (cf. (3.135))

$$\int_B \int_A f(X, Y)\delta X \delta Y \triangleq \int_{\mathcal{Y}(B)} \int_{\mathcal{X}(A)} f(X, Y) \mathrm{d}\mu_{\mathcal{X}}(X) \mathrm{d}\mu_{\mathcal{Y}}(Y), \tag{4.7}$$

and similarly for functions with more than two finite sets as arguments.

Note that if $f(X) = f_{\mathsf{X}}(X)$ is the PDF of a RFS X, (4.2) becomes

$$\int_A f_{\mathsf{X}}(X)\delta X = P_{\mathsf{X}}(\mathcal{X}(A)) = \Pr(\mathsf{X} \subseteq A) = \beta_{\mathsf{X}}(A). \tag{4.8}$$

Similarly, if $f(X, Y) = f_{\mathsf{X},\mathsf{Y}}(X, Y)$ is the joint PDF of the RFSs X and Y, (4.7) becomes

$$\int_B \int_A f_{\mathsf{X},\mathsf{Y}}(X, Y)\delta X \delta Y = P_{\mathsf{X},\mathsf{Y}}(\mathcal{X}(A) \times \mathcal{Y}(B)) = \Pr(\mathsf{X} \subseteq A \wedge \mathsf{Y} \subseteq B) = \beta_{\mathsf{X},\mathsf{Y}}(A, B). \tag{4.9}$$

4.2 Probability Generating Functionals

A functional $F[h]$ is a mapping of functions $h \colon \mathbb{R}^N \to \mathbb{R}$ to real numbers. In most cases, we will consider functions $h \colon \mathbb{R}^N \to [0, 1]$, which we will refer to as **test functions**.

Example 4.1: One important class of functionals are **linear functionals** $f[h]$ corresponding to some function $f \colon \mathbb{R}^N \to \mathbb{R}$, defined as

$$f[h] \triangleq \int_{\mathbb{R}^N} h(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x}). \tag{4.10}$$

As the name already suggests, these functionals are linear, i.e.,

$$f[a_1h_1 + a_2h_2] = a_1f[h_1] + a_2f[h_2], \tag{4.11}$$

for any functions h_1, h_2 and scalars $a_1, a_2 \in \mathbb{R}$, provided the corresponding integrals exist. Note that for $h = \mathbf{1}_A$, where $\mathbf{1}_A$ is the indicator function of a closed set $A \subseteq \mathbb{R}^N$, the linear functional is the integral of $f(\boldsymbol{x})$ over A, i.e. (cf. (2.36)),

$$f[\mathbf{1}_A] = \int_{\mathbb{R}^N} \mathbf{1}_A(\boldsymbol{x})f(\boldsymbol{x})\mathrm{d}\lambda_N(\boldsymbol{x}) = \int_A f(\boldsymbol{x})\mathrm{d}\lambda_N(\boldsymbol{x}). \tag{4.12}$$

Furthermore, if the function $f(\boldsymbol{x}) = f_{\mathsf{x}}(\boldsymbol{x})$ is the PDF of a random vector $\mathsf{x} \in \mathbb{R}^N$, we have

$$f_{\mathsf{x}}[h] = \int_{\mathbb{R}^N} h(\boldsymbol{x})f_{\mathsf{x}}(\boldsymbol{x})\mathrm{d}\lambda_N(\boldsymbol{x}) = \mathsf{E}[h(\mathsf{x})], \tag{4.13}$$

and according to (4.12)

$$f_{\mathsf{x}}[\mathbf{1}_A] = \int_A f_{\mathsf{x}}(\boldsymbol{x})\mathrm{d}\lambda_N(\boldsymbol{x}) = P_{\mathsf{x}}(A) = \Pr(\mathsf{x} \in A). \tag{4.14}$$

Let h be a real valued-function on $\mathbb{R}^N$, and let $X \in \mathcal{X}$ be a finite subset of $\mathbb{R}^N$. The power of h with respect to X is defined as

$$h^X \triangleq \begin{cases} 1 & \text{if } X = \emptyset, \\ \prod_{i=1}^k h(\boldsymbol{x}_i) & \text{if } |X| = k, X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}. \end{cases} \tag{4.15}$$

Note that for h fixed, h^X is a function from $\mathcal{X}$ to $\mathbb{R}$.

Example 4.2: As a simple example, consider the constant function $h(\boldsymbol{x}) \equiv c$, with some $c \in \mathbb{R}$. Then $h^X = c^{|X|}$. In particular, for $h(\boldsymbol{x}) \equiv 0$ we have

$$0^X = \begin{cases} 1 & \text{if } X = \emptyset, \\ 0 & \text{if } X \neq \emptyset, \end{cases} \tag{4.16}$$

$$= \mathbf{1}_{\mathcal{X}_0}(X), \tag{4.17}$$

where $\mathcal{X}_0 = \{\emptyset\}$. For $h(\boldsymbol{x}) \equiv 1$ we have

$$1^X = 1. \tag{4.18}$$

Furthermore, the power of the indicator function $\mathbf{1}_A(\boldsymbol{x})$ for $A \subseteq \mathbb{R}^N$ is

$$\mathbf{1}_A^X = \begin{cases} 1 & \text{if } X = \emptyset, \\ \prod_{i=1}^k \mathbf{1}_A(\boldsymbol{x}_i) & \text{if } |X| = k, X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}, \end{cases} \tag{4.19}$$

$$= \begin{cases} 1 & \text{if } X \subseteq A, \\ 0 & \text{if } X \nsubseteq A, \end{cases} \tag{4.20}$$

$$= \mathbf{1}_{\mathcal{X}(A)}(X), \tag{4.21}$$

where $\mathcal{X}(A) = \{X \in \mathcal{X} : X \subseteq A\}$ is the collection of all finite subsets of A.

Let $X_i \in \mathcal{X}$, $i = 1, \ldots, k$ be finite sets. Then, it follows immediately from (4.15) that

$$h^{\bigcup_{i=1}^{k} X_i} = \prod_{i=1}^{k} h^{X_i}. \tag{4.22}$$

The **probability generating functional (PGFL)** of an RFS $\mathsf{X} \in \mathcal{X}$ with PDF $f_{\mathsf{X}}(X)$ is defined as the expectation of h^{X} (cf. (3.184))

$$G_{\mathsf{X}}[h] \triangleq \mathsf{E}[h^{\mathsf{X}}] = \int_{\mathbb{R}^N} h^X f_{\mathsf{X}}(X) \delta X. \tag{4.23}$$

Using (4.2) and (4.3), the PGFL can be written as

$$\begin{aligned} G_{\mathsf{X}}[h] &= \int_{\mathcal{X}} h^X f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}}(X) && (4.24) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \int_{\mathbb{R}^{kN}} \left(\prod_{i=1}^{k} h(\boldsymbol{x}_i) \right) \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) && (4.25) \end{aligned}$$

where by convention $\prod_{i=1}^{0} h(\boldsymbol{x}_i) = 1$.

The following properties of the PGFL immediately follow from (4.25):

- $G_{\mathsf{X}}[0] = f_{\mathsf{X}}(\emptyset)$.
- $G_{\mathsf{X}}[1] = 1$.
- For nonnegative functions h_1 and h_2 satisfying $h_1(\boldsymbol{x}) \leq h_2(\boldsymbol{x})$ a.e., we obtain $G_{\mathsf{X}}[h_1] \leq G_{\mathsf{X}}[h_2]$.

Furthermore, the PGFL at $h = \mathbf{1}_A$, where A is a closed subset of $\mathbb{R}^N$, is equal to the BMF:

$$\begin{aligned} G_{\mathsf{X}}[\mathbf{1}_A] &= \int_{\mathcal{X}} \mathbf{1}_A^X f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}} && (4.26) \\ &= \int_{\mathcal{X}} \mathbf{1}_{\mathcal{X}(A)}(X) f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}} && (4.27) \\ &= \int_{\mathcal{X}(A)} f_{\mathsf{X}}(X) \mathrm{d}\mu_{\mathcal{X}} && (4.28) \\ &= P_{\mathsf{X}}(\mathcal{X}(A)) = \beta_{\mathsf{X}}(A), && (4.29) \end{aligned}$$

where (4.21) has been used.

The most important property of PGFLs is the following [16, p. 803].

Theorem 4.3: Let $\mathsf{X}_i \in \mathcal{X}$, $i = 1, \ldots, n$ be statistically independent RFSs with PGFLs $G_{\mathsf{X}_i}[h]$, and let

$$\mathsf{X} = \bigcup_{i=1}^{n} \mathsf{X}_i. \tag{4.30}$$

Then

$$G_{\mathsf{X}}[h] = \prod_{i=1}^{n} G_{\mathsf{X}_i}[h]. \tag{4.31}$$

That is, the PGFL of the union of statistically independent RFSs is the product of the PGFLs of the individual RFSs. Inserting $h = \mathbf{1}_A$ in (4.31) and using (4.29), we obtain an analogous factorization for the BMF of a union of independent RFSs, i.e.,

$$\beta_{\mathsf{X}}(A) = \prod_{i=1}^{n} \beta_{\mathsf{X}_i}(A). \tag{4.32}$$

4.2.1 Joint Probability Generating Functionals

Let $\mathsf{X} \in \mathcal{X}$ and $\mathsf{Y} \in \mathcal{Y}$ be two RFS with joint PDF $f_{\mathsf{X},\mathsf{Y}}(X,Y)$. The **joint PGFL** of X and Y is defined as the expectation of $h^{\mathsf{X}} g^{\mathsf{Y}}$ with respect to $f_{\mathsf{X},\mathsf{Y}}(X,Y)$ (cf. (3.187)), i.e.,

$$G_{\mathsf{X},\mathsf{Y}}[h,g] \triangleq \mathsf{E}[h^{\mathsf{X}} g^{\mathsf{Y}}] = \int_{\mathbb{R}^M} \int_{\mathbb{R}^N} h^X g^Y f_{\mathsf{X},\mathsf{Y}}(X,Y) \delta X \delta Y, \tag{4.33}$$

where h and g are test functions on $\mathbb{R}^N$ and $\mathbb{R}^M$, respectively. It is easily shown that the joint PGFL has the following properties:

- $G_{\mathsf{X},\mathsf{Y}}[0,0] = f_{\mathsf{X},\mathsf{Y}}(\emptyset,\emptyset)$.
- $G_{\mathsf{X},\mathsf{Y}}[0,1] = f_{\mathsf{X}}(\emptyset)$.
- $G_{\mathsf{X},\mathsf{Y}}[1,0] = f_{\mathsf{Y}}(\emptyset)$.
- $G_{\mathsf{X},\mathsf{Y}}[1,1] = 1$.

Analogous to (4.26)–(4.29), the joint BMF can be obtained from the joint PGFL by setting $h = \mathbf{1}_A$ and $g = \mathbf{1}_B$, where A and B are closed subsets of $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively, i.e.,

$$G_{\mathsf{X},\mathsf{Y}}[\mathbf{1}_A, \mathbf{1}_B] = \beta_{\mathsf{X},\mathsf{Y}}(A,B). \tag{4.34}$$

Furthermore, the marginal PGFLs $G_{\mathsf{X}}[h]$ and $G_{\mathsf{Y}}[g]$ can be obtained from the joint PGFL by setting either h or g to 1, i.e.,

$$G_{\mathsf{X},\mathsf{Y}}[h,1] = G_{\mathsf{X}}[h], \tag{4.35}$$

$$G_{\mathsf{X},\mathsf{Y}}[1,g] = G_{\mathsf{Y}}[g]. \tag{4.36}$$

Finally, if X and Y are statistically independent, the joint PGFL factorizes as

$$G_{\mathsf{X},\mathsf{Y}}[h,g] = G_{\mathsf{X}}[h] G_{\mathsf{Y}}[g]. \tag{4.37}$$

4.2.2 Conditional Probability Generating Functionals

Let $\mathsf{X} \in \mathcal{X}$ and $\mathsf{Y} \in \mathcal{Y}$ be two RFSs with conditional PDF $f_{\mathsf{X}|\mathsf{Y}}(X|Y)$. We define the **conditional PGFL** of X given $\mathsf{Y} = Y$ as the expectation of h^{X} with respect to $f_{\mathsf{X}|\mathsf{Y}}(X|Y)$ (cf. (3.191)), i.e.,

$$G_{\mathsf{X}|\mathsf{Y}}[h|Y] \triangleq \mathsf{E}[h^{\mathsf{X}}|Y] = \int_{\mathbb{R}^N} h^X f_{\mathsf{X}|\mathsf{Y}}(X|Y)\delta X. \tag{4.38}$$

The following properties follow from this definition:

- $G_{\mathsf{X}|\mathsf{Y}}[0|Y] = f_{\mathsf{X}|\mathsf{Y}}(\emptyset|Y)$.
- $G_{\mathsf{X}|\mathsf{Y}}[1|Y] = 1$.
- $G_{\mathsf{X}|\mathsf{Y}}[\mathbf{1}_A|Y] = \beta_{\mathsf{X}|\mathsf{Y}}(A|Y)$.

The conditional BMF can be obtained from the condtional PGFL by setting $h = \mathbf{1}_A$, where A is a closed subset of $\mathbb{R}^N$, i.e.,

$$G_{\mathsf{X}|\mathsf{Y}}[\mathbf{1}_A|Y] = \beta_{\mathsf{X}|\mathsf{Y}}(A|Y). \tag{4.39}$$

Furthermore, if X and Y are statistically independent, the condtional PGFL equals the marginal PGFL, i.e.,

$$G_{\mathsf{X}|\mathsf{Y}}[h|Y] = G_{\mathsf{X}}[h]. \tag{4.40}$$

Finally, using (4.33) and (3.169), the joint PGFL can be calculated from the conditional PGFL by

$$\begin{aligned} G_{\mathsf{X},\mathsf{Y}}[h,g] &= \int_{\mathbb{R}^M}\int_{\mathbb{R}^N} h^X g^Y f_{\mathsf{X},\mathsf{Y}}(X,Y)\delta X\delta Y && (4.41)\\ &= \int_{\mathbb{R}^M} g^Y \left(\int_{\mathbb{R}^N} h^X f_{\mathsf{X}|\mathsf{Y}}(X|Y)\delta X\right) f_{\mathsf{Y}}(Y)\delta Y && (4.42)\\ &= \int_{\mathbb{R}^M} g^Y G_{\mathsf{X}|\mathsf{Y}}[h|Y] f_{\mathsf{Y}}(Y)\delta Y. && (4.43) \end{aligned}$$

This is the PGFL version of the total probability theorem.

4.3 Functional Derivatives

The **Fréchet derivative** (also known as directional or gradient derivative) of a functional $F[h]$ in direction g is defined as [6]

$$\frac{\delta F}{\delta g}[h] \triangleq \lim_{\varepsilon \searrow 0} \frac{F[h+\varepsilon g] - F[h]}{\varepsilon}. \tag{4.44}$$

Some functionals also allow distributions, such as the Dirac delta function $\delta_{\boldsymbol{x}}(\boldsymbol{y}) = \delta(\boldsymbol{y} - \boldsymbol{x})$, as arguments.

Example 4.4: For the linear functional corresponding to a continuous function f as defined in (4.10), we have

$$f[\delta_{\boldsymbol{x}}] = \int_{\mathbb{R}^N} \delta(\boldsymbol{y}-\boldsymbol{x}) f(\boldsymbol{y}) \mathrm{d}\lambda_N(\boldsymbol{y}) = f(\boldsymbol{x}). \tag{4.45}$$

In this case, the **functional derivative** of a functional $F[h]$ at $\boldsymbol{x} \in \mathbb{R}^N$, if it exists, is defined as the Fréchet derivative of $F[h]$ in direction $g = \delta_{\boldsymbol{x}}$ [6], i.e.,

$$\frac{\delta F}{\delta \boldsymbol{x}}[h] \triangleq \lim_{\varepsilon \searrow 0} \frac{F[h+\varepsilon \delta_{\boldsymbol{x}}] - F[h]}{\varepsilon}. \tag{4.46}$$

Note that the functional derivative $\frac{\delta F}{\delta \boldsymbol{x}}[h]$ depends on the function h and the vector $\boldsymbol{x}$, i.e., it can be considered as a functional with respect to h parameterized by $\boldsymbol{x}$. Higher-order functional derivatives of $F[h]$ at $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^N$ are defined recursively by

$$\frac{\delta^k F}{\delta \boldsymbol{x}_1 \cdots \delta \boldsymbol{x}_k}[h] \triangleq \frac{\delta}{\delta \boldsymbol{x}_1} \frac{\delta^{k-1} F}{\delta \boldsymbol{x}_2 \cdots \delta \boldsymbol{x}_k}[h], \tag{4.47}$$

provided that all derivatives exist. If X is a finite subset of $\mathbb{R}^N$, then the functional derivative of $F[h]$ at X is defined as

$$\frac{\delta F}{\delta X}[h] \triangleq \begin{cases} F[h] & \text{if } X = \emptyset, \\ \frac{\delta^k F}{\delta \boldsymbol{x}_1 \cdots \delta \boldsymbol{x}_k}[h] & \text{if } |X| = k \text{ and } X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}. \end{cases} \tag{4.48}$$

Note that for $\frac{\delta F}{\delta X}[h]$ to be well defined it is implicity assumed that for any given h the order of the derivatives does not matter. (This is necessary because in $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ the elements $\boldsymbol{x}_i$ are not ordered.)

Example 4.5: Consider the special case where the PDF of X takes the form

$$f_{\mathsf{X}}(X) = \begin{cases} f(\boldsymbol{x}) & \text{if } X = \{\boldsymbol{x}\}, \\ 0 & \text{if } |X| \neq 1, \end{cases} \tag{4.49}$$

where $f(\boldsymbol{x})$ is an arbitrary continuous vector PDF. Note that in this case all EVDs with $k \neq 1$ are zero and $\tilde{f}_{\mathsf{X}}^{(1)}(\boldsymbol{x}) = f(\boldsymbol{x})$. Using (4.25), the PGFL of X is obtained as

$$G_{\mathsf{X}}[h] = \int_{\mathbb{R}^N} h(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x}) = f[h]. \tag{4.50}$$

That is, the PGFL is the linear functional $f[h]$ (cf. (4.10)) corresponding to the vector PDF $f(\boldsymbol{x})$. The functional derivative of $G_{\mathsf{X}}[h] = f[h]$ at some point $\boldsymbol{x} \in \mathbb{R}^N$ is then given by

$$\frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[h] = \lim_{\varepsilon \searrow 0} \frac{f[h+\varepsilon \delta_{\boldsymbol{x}}] - f[h]}{\varepsilon} \tag{4.51}$$

$$= \lim_{\varepsilon \searrow 0} \frac{f[h] + \varepsilon f[\delta_{\boldsymbol{x}}] - f[h]}{\varepsilon} \tag{4.52}$$

$$= f[\delta_{\boldsymbol{x}}] \tag{4.53}$$

$$= f(\boldsymbol{x}), \tag{4.54}$$

where the last result follows from (4.45). Since $\frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[h]$ does not depend on h anymore, differentiating it again yields

$$\frac{\delta}{\delta \boldsymbol{y}} \frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[h] = \lim_{\varepsilon \searrow 0} \frac{f(\boldsymbol{x}) - f(\boldsymbol{x})}{\varepsilon} = 0. \tag{4.55}$$

Consequently, the functional derivative of $G_{\mathsf{X}}[h]$ at some finite set $X \subseteq \mathbb{R}^N$ is obtained from (4.48) as

$$\frac{\delta G_{\mathsf{X}}}{\delta X}[h] = \begin{cases} f[h] & \text{if } X = \emptyset, \\ f(\boldsymbol{x}) & \text{if } X = \{\boldsymbol{x}\}, \\ 0 & \text{if } |X| \geq 2. \end{cases} \tag{4.56}$$

We finally list some basic properties of functional derivatives (for a more complete list and proofs see [16, Section 11.6]). For any differentiable functionals $F_1[h], F_2[h]$, any scalars $c, a_1, a_2 \in \mathbb{R}$ and $n \in \mathbb{N}$, any vector $\boldsymbol{x} \in \mathbb{R}^N$, any finite set $X \subseteq \mathbb{R}^N$, and any differentiable function $g\colon \mathbb{R} \to \mathbb{R}$, the following holds:

1. Constant Rule:

$$\frac{\delta}{\delta X} c = \begin{cases} c & \text{if } X = \emptyset, \\ 0 & \text{if } X \neq \emptyset. \end{cases} \tag{4.57}$$

2. Linearity:

$$\frac{\delta}{\delta X} \left(a_1 F_1[h] + a_2 F_2[h] \right) = a_1 \frac{\delta F_1}{\delta X}[h] + a_2 \frac{\delta F_2}{\delta X}[h]. \tag{4.58}$$

3. Power Rule:

$$\frac{\delta}{\delta \boldsymbol{x}} \left(F_1[h]^n \right) = n F_1[h]^{n-1} \frac{\delta F_1}{\delta \boldsymbol{x}}[h]. \tag{4.59}$$

4. Product Rule I:

$$\frac{\delta}{\delta \boldsymbol{x}} \left(F_1[h] F_2[h] \right) = \frac{\delta F_1}{\delta \boldsymbol{x}}[h] F_2[h] + \frac{\delta F_2}{\delta \boldsymbol{x}}[h] F_1[h]. \tag{4.60}$$

5. Product Rule II:

$$\frac{\delta}{\delta X} \left(F_1[h] F_2[h] \right) = \sum_{Y \subseteq X} \frac{\delta F_1}{\delta Y}[h] \frac{\delta F_2}{\delta (X \setminus Y)}[h]. \tag{4.61}$$

6. Chain Rule:

$$\frac{\delta}{\delta \boldsymbol{x}} g(F_1[h]) = \frac{\mathrm{d}g}{\mathrm{d}y}(F_1[h]) \frac{\delta F_1}{\delta \boldsymbol{x}}[h]. \tag{4.62}$$

Here, $\frac{\mathrm{d}g}{\mathrm{d}y}(F_1[h])$ denotes the derivative of the function $g(y)$ evaluated at $y = F_1[h]$.

4.4 Probability Hypothesis Density

In this section, we introduce a first-order moment of an RFS, called the **probability hypothesis density (PHD)** [16, Section 16.2]. Let $\mathsf{X} \subseteq \mathbb{R}^N$ be an RFS with PDF $f_{\mathsf{X}}(X)$. We define the function $\nu_{\mathsf{X}} \colon \mathcal{B}(\mathbb{R}^N) \to [0, \infty)$ by

$$\nu_{\mathsf{X}}(S) \triangleq \mathsf{E}[|\mathsf{X} \cap S|] = \int_{\mathbb{R}^N} |X \cap S| f_{\mathsf{X}}(X) \delta X, \tag{4.63}$$

which is known as the **intensity measure** of the RFS X. That is, $\nu_{\mathsf{X}}(S)$ is equal to the expected number of elements of X that are contained in the region $S \subseteq \mathbb{R}^N$. A proof of the following theorem is provided in Appendix A.6.

Theorem 4.6: The function ν_{X} as defined in (4.63) is a measure on $\mathcal{B}(\mathbb{R}^N)$. Furthermore, ν_{X} is absolutely continuous with respect to the product Lebesgue measure λ_N, i.e.,

$$\nu_{\mathsf{X}} \ll \lambda_N. \tag{4.64}$$

Since ν_{X} and λ_N are measures on $\mathcal{B}(\mathbb{R}^N)$ with $\nu_{\mathsf{X}} \ll \lambda_N$, the RN theorem (Theorem 2.18) states that there exists an (a.e.) unique measurable function $D_{\mathsf{X}}(\boldsymbol{x}) \colon \mathbb{R}^N \to [0, \infty)$ such that

$$\nu_{\mathsf{X}}(S) = \int_S D_{\mathsf{X}}(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x}). \tag{4.65}$$

This function is the PHD (also known as the intensity density) of the RFS X. That is, the PHD is a unique function associated to an RFS whose integral over a region $S \in \mathcal{B}(\mathbb{R}^N)$ is equal to the expected number of elements of X contained in that region. In particular, the integral of the PHD over $\mathbb{R}^N$ yields the expected number of elements of X since

$$\int_{\mathbb{R}^N} D_{\mathsf{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \mathsf{E}[|\mathsf{X} \cap \mathbb{R}^N|] = \mathsf{E}[|\mathsf{X}|] = \sum_{n=0}^{\infty} n p_{|\mathsf{X}|}(n). \tag{4.66}$$

As shown in [16, Chapter 16], the PHD can be calculated from the PDF $f_{\mathsf{X}}(X)$ by

$$D_{\mathsf{X}}(\boldsymbol{x}) = \mathsf{E}[\delta_X(\boldsymbol{x})] = \int_{\mathbb{R}^N} \delta_X(\boldsymbol{x}) f_{\mathsf{X}}(X) \delta X, \tag{4.67}$$

where

$$\delta_X(\boldsymbol{x}) \triangleq \begin{cases} 0 & \text{if } X = \emptyset, \\ \sum_{i=1}^k \delta_{\boldsymbol{x}_i}(\boldsymbol{x}) & \text{if } |X| = k, X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}, \end{cases} \tag{4.68}$$

and $\delta_{\boldsymbol{x}_i}(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_i)$ are Dirac delta functions located at $\boldsymbol{x}_i$.

The PHD can also be derived from the PGFL by [16, Chapter 16]

$$D_{\mathsf{X}}(\boldsymbol{x}) = \frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[1]. \tag{4.69}$$

Another important property of the PHD is the following [5, p. 163].

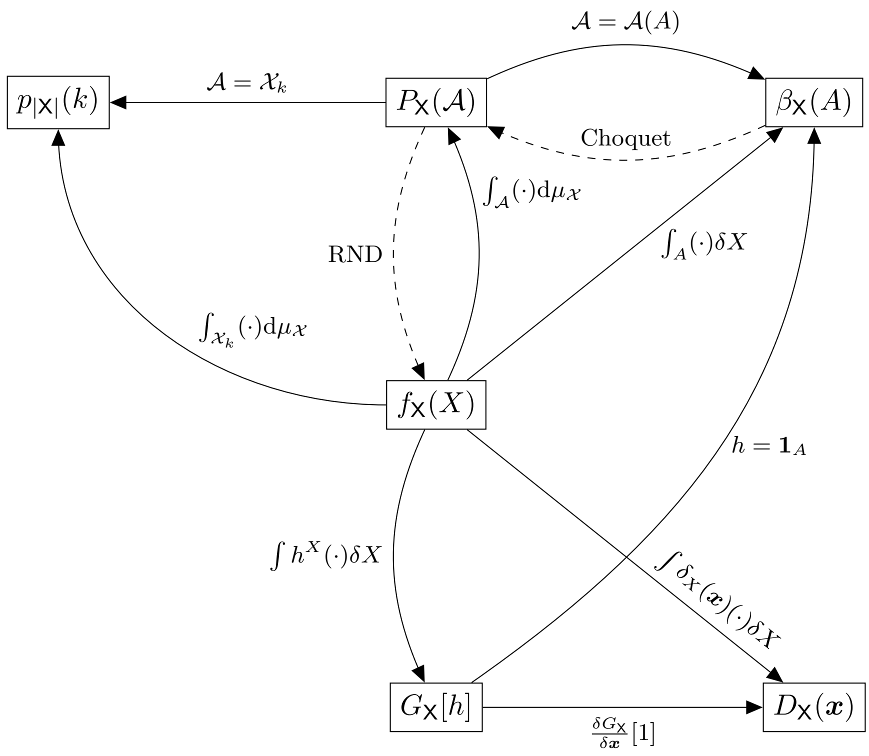

Fig. 4.1: *Graphical summary of the various descriptors of RFSs and their interrelations. The dashed edges indicate a nonconstructive relation.*

Theorem 4.7 (Campbell): Let $D_{\mathsf{X}}(\boldsymbol{x})$ be the PHD of the RFS X and let $g\colon \mathbb{R}^N \to \mathbb{R}$. Then

$$\mathsf{E}\left[\sum_{\mathbf{x}\in\mathsf{X}} g(\mathbf{x})\right] = \int_{\mathbb{R}^N} g(\boldsymbol{x}) D_{\mathsf{X}}(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x}). \tag{4.70}$$

At this point, we have introduced the main descriptors of an RFS X:

- Probability measure $P_{\mathsf{X}}(\mathcal{A})$.
- Cardinality distribution (CD) $p_{|\mathsf{X}|}(n)$.
- Belief mass function (BMF) $\beta_{\mathsf{X}}(A)$.
- Probability density function (PDF) $f_{\mathsf{X}}(X)$.
- Probability generating functional (PGFL) $G_{\mathsf{X}}[h]$.
- Probability hypothesis density (PHD) $D_{\mathsf{X}}(\boldsymbol{x})$.

A graphical summary of these descriptors and their interrelations is provided in Figure 4.1.

4.5 Elementary Distributions

In this section we discuss four common types of RFSs and derive their PDFs, BMFs, PGFLs, and PHDs [32, Section 2.9].

4.5.1 Independent and Identically Distributed Cluster RFS

An independent, identically distributed (i.i.d.) cluster RFS arises if all component random vectors $\mathsf{x}_i^{(k)}$ in our general construction in Section 3.6.2 are i.i.d. with a common PDF $f_{\mathsf{x}_i^{(k)}}(\boldsymbol{x}) = f(\boldsymbol{x})$, i.e., $f_{\mathsf{x}^{(k)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \prod_{i=1}^k f(\boldsymbol{x}_i)$. Hence, the EVDs of an i.i.d. cluster RFS are given by (cf. (3.101))

$$\tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = p_k \sum_{\sigma} \prod_{i=1}^{k} f(\boldsymbol{x}_{\sigma_i}) = p_k k! \prod_{i=1}^{k} f(\boldsymbol{x}_i), \tag{4.71}$$

where $p_k = p_{|\mathsf{X}|}(k)$, $k \in \mathbb{N}_0$, is an arbitrary CD. Note that the EVDs (and thereby the PDF) are fully characterized by the PDF $f(\boldsymbol{x})$ and the CD p_k. The PGFL of this RFS is, from (4.25),

$$G_{\mathsf{X}}[h] = \sum_{k=0}^{\infty} p_k \int_{\mathbb{R}^{kN}} \left(\prod_{i=1}^{k} h(\boldsymbol{x}_i) f(\boldsymbol{x}_i) \right) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \tag{4.72}$$

$$= \sum_{k=0}^{\infty} p_k \prod_{i=1}^{k} \int_{\mathbb{R}^N} h(\boldsymbol{x}_i) f(\boldsymbol{x}_i) \mathrm{d}\lambda_N(\boldsymbol{x}_i) \tag{4.73}$$

$$= \sum_{k=0}^{\infty} p_k (f[h])^k, \tag{4.74}$$

where $f[h] = \int_{\mathbb{X}_1} h(\boldsymbol{x}) f(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x})$. Using (4.29) and $f[\mathbf{1}_A] = \int_A f(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x}) = P_{\mathsf{x}}(A)$, the BMF can be easily calculated from the PGFL as

$$\beta_{\mathsf{X}}(A) = G_{\mathsf{X}}[\mathbf{1}_A] = \sum_{k=0}^{\infty} p_k (P_{\mathsf{x}}(A))^k. \tag{4.75}$$

For the PHD we get from (4.69)

$$D_{\mathsf{X}}(\boldsymbol{x}) = \frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[1] = \sum_{k=0}^{\infty} p_k \left[\frac{\delta}{\delta \boldsymbol{x}} (f[h]^k) \right]_{h=1}, \tag{4.76}$$

where the linearity of the functional derivative (4.58) has been used. Using the power rule (4.59) and $\frac{\delta}{\delta \boldsymbol{x}} f[h] = f(\boldsymbol{x})$ (cf. (4.54)), the derivative becomes

$$\frac{\delta}{\delta \boldsymbol{x}} (f[h]^k) = k f[h]^{k-1} \frac{\delta}{\delta \boldsymbol{x}} f[h] = k f[h]^{k-1} f(\boldsymbol{x}), \tag{4.77}$$

and therefore, since $f[1] = \int_{R^N} f(\boldsymbol{x}) \mathrm{d}\lambda_N(\boldsymbol{x}) = 1$, we obtain

$$D_{\mathsf{X}}(\boldsymbol{x}) = \sum_{k=1}^{\infty} p_k k f[1]^{k-1} f(\boldsymbol{x}) = \sum_{k=0}^{\infty} p_k k f(\boldsymbol{x}) = \mathsf{E}[|\mathsf{X}|] f(\boldsymbol{x}). \tag{4.78}$$

Note that the i.i.d. cluster RFS is fully characterized by the CD and PHD.

4.5.2 Poisson RFS

Poisson RFSs are a special case of i.i.d. cluster RFSs where the cardinality distribution is Poisson, i.e.,

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \tag{4.79}$$

with mean $\lambda > 0$. Inserting this into (4.71), the EVDs of a Poisson RFS are

$$\tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = e^{-\lambda}\lambda^k \prod_{i=1}^{k} f(\boldsymbol{x}_i), \tag{4.80}$$

where $f(\boldsymbol{x})$ is again an arbitrary vector PDF. Note that a Poisson RFS is completely specified by λ and $f(\boldsymbol{x})$. Similarly, the PGFL, BMF, and PHD can be calculated by inserting (4.79) into (4.74), (4.75), and (4.78), respectively. We thus obtain for the PGFL

$$G_{\mathsf{X}}[h] = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} f^k[h] = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda f[h])^k}{k!} = e^{-\lambda} e^{\lambda f[h]} = e^{\lambda(f[h]-1)}. \tag{4.81}$$

The BMF is hence given by

$$\beta_{\mathsf{X}}(A) = G_{\mathsf{X}}[\mathbf{1}_A] = e^{\lambda(P_{\mathsf{x}}(A)-1)}. \tag{4.82}$$

Finally, since $\mathsf{E}[|\mathsf{X}|] = \lambda$, the PHD is

$$D_{\mathsf{X}}(\boldsymbol{x}) = \lambda f(\boldsymbol{x}). \tag{4.83}$$

Note that the Poisson RFS is fully characterized by the PHD.

4.5.3 Bernoulli RFS

Bernoulli RFSs are very simple RFSs that are either empty with probability $1-p$, or contain a single element with probability p. Consequently, the PDF of a Bernoulli RFS is of the form

$$f_{\mathsf{X}}(X) = \begin{cases} 1-p & \text{if } X = \emptyset, \\ pf(\boldsymbol{x}) & \text{if } X = \{\boldsymbol{x}\}, \\ 0 & \text{if } |X| \geq 2, \end{cases} \tag{4.84}$$

with some $p \in [0,1]$ and vector PDF $f(\boldsymbol{x})$. The PGFL, BMF, and PHD are easily derived as

$$G_{\mathsf{X}}[h] = f_{\mathsf{X}}(\emptyset) + \int h(\boldsymbol{x}) p f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1 - p + pf[h], \tag{4.85}$$

$$\beta_{\mathsf{X}}(A) = G_{\mathsf{X}}[\mathbf{1}_A] = 1 - p + pP_{\mathsf{x}}(A), \tag{4.86}$$

$$D_{\mathsf{X}}(\boldsymbol{x}) = \frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[1] = p\frac{\delta f}{\delta \boldsymbol{x}}[1] = pf(\boldsymbol{x}). \tag{4.87}$$

4.5.4 Multi-Bernoulli RFS

A multi-Bernoulli RFS is an RFS X that is the union of a finite number of statistically independent Bernoulli RFSs X_k, i.e.,

$$\mathsf{X} = \mathsf{X}_1 \cup \cdots \cup \mathsf{X}_K. \tag{4.88}$$

Here, $K \geq 2$, and each X_i has a PDF f_{X_i} of the form (4.84) with parameters p_i and $f_i(\boldsymbol{x})$. Due to the independence assumption, the PGFL and BMF are given by the product of the individual PGFLs and BMFs, respectively (cf. (4.31) and (4.32)):

$$G_{\mathsf{X}}[h] = \prod_{i=1}^{K} G_{\mathsf{X}_i}[h] = \prod_{i=1}^{K} (1 - p_i + p_i f_i[h]) \tag{4.89}$$

and

$$\beta_{\mathsf{X}}(A) = \prod_{i=1}^{K} \beta_{\mathsf{X}_i}(A) = \prod_{i=1}^{K} (1 - p_i + p_i P_{\mathbf{x}_i}(A)). \tag{4.90}$$

The PHD can be calculated by using the product rule (4.60)

$$\begin{aligned}
D_{\mathsf{X}}(\boldsymbol{x}) &= \frac{\delta G_{\mathsf{X}}}{\delta \boldsymbol{x}}[1] &(4.91)\\
&= \left[\frac{\delta}{\delta \boldsymbol{x}} \prod_{i=1}^{K} G_{\mathsf{X}_i}[h]\right]_{h=1} &(4.92)\\
&= \sum_{i=1}^{K} G_{\mathsf{X}_1}[h] \cdots \frac{\delta G_{\mathsf{X}_i}}{\delta \boldsymbol{x}}[h] \cdots G_{\mathsf{X}_K}[h]\Big|_{h=1} &(4.93)\\
&= \sum_{i=1}^{K} G_{\mathsf{X}_1}[1] \cdots p_i f_i(\boldsymbol{x}) \cdots G_{\mathsf{X}_K}[1] &(4.94)\\
&= \sum_{i=1}^{K} p_i f_i(\boldsymbol{x}), &(4.95)
\end{aligned}$$

where $G_{\mathsf{X}_i}[1] = 1$ has been used. The EVDs are given by [16, pp. 368–370]

$$\tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq K} Q_{i_1,\ldots,i_k} \prod_{j=1}^{k} f_{i_j}(\boldsymbol{x}_j), \tag{4.96}$$

where

$$Q_{i_1,\ldots,i_k} = \frac{\prod_{i=1}^{K}(1 - p_i)}{\prod_{j=1}^{k}(1 - p_{i_j})} \prod_{j=1}^{k} p_{i_j}. \tag{4.97}$$

Chapter 5

State Estimation from Image Observations

5.1 State Estimation from a Single Image

In this section we discuss the problem of jointly estimating the number of objects and their corresponding states based on a noisy image observation.

5.1.1 Data Model

Let $\mathsf{s} = (\mathsf{s}_1, \ldots, \mathsf{s}_m) \in \mathbb{R}^m$, with $m = m_1 m_2$ be a random vector consisting of the pixel values of an $m_1 \times m_2$ grayscale image and let $\boldsymbol{z} \in \mathbb{R}^N$ denote the state of an object. Furthermore, let $T(\boldsymbol{z})$ be a function that gives the set of pixel-indices in the image that are influenced by an object with state $\boldsymbol{z}$. We will refer to the pixels influenced by an object as its *pixel set*. An object affects the pixels of its pixel set by changing their distributions. The conditional PDF of pixel s_i given the state $\boldsymbol{z}$ can therefore be written as [22]

$$p(s_i|\boldsymbol{z}) = \begin{cases} \varphi(s_i, \boldsymbol{z}) & \text{if } i \in T(\boldsymbol{z}), \\ \phi(s_i) & \text{if } i \notin T(\boldsymbol{z}), \end{cases} \tag{5.1}$$

with arbitrary but known PDFs $\varphi(s_i, \boldsymbol{z})$ and $\phi(s_i)$.

We are only interested in objects that actually affect our image, i.e., objects located in the state space

$$R \triangleq \{\boldsymbol{z} \in \mathbb{R}^N : T(\boldsymbol{z}) \neq \emptyset\} \subseteq \mathbb{R}^N. \tag{5.2}$$

The number of objects as well as their states are unknown and random. Since the objects do not possess an inherent ordering, we therefore model them as an RFS $\mathsf{Z} \subseteq R$ with given (prior) PDF $f_{\mathsf{Z}}(Z)$. Furthermore, we assume that the likelihood function $g(\boldsymbol{s}|Z) \triangleq f_{\mathsf{s}|\mathsf{Z}}(\boldsymbol{s}|Z)$ is *separable*, i.e., it can be written in a factorized form as [22]

$$g(\boldsymbol{s}|Z) = \pi(\boldsymbol{s}) \prod_{\boldsymbol{z} \in Z} \eta(\boldsymbol{z}, \boldsymbol{s}), \tag{5.3}$$

where

$$\pi(\boldsymbol{s}) = \prod_{i=1}^{m} \phi(s_i), \tag{5.4}$$

$$\eta(\boldsymbol{z}, \boldsymbol{s}) = \prod_{i \in T(\boldsymbol{z})} \frac{\varphi(s_i, \boldsymbol{z})}{\phi(s_i)}. \tag{5.5}$$

5.1.2 Estimator

Our task is to obtain an estimate $\hat{Z}$ of the realization $\mathsf{Z} = Z$ based on the image observation $\mathbf{s} = \boldsymbol{s}$. We will use an estimator based on the posterior PHD (cf. (4.67))

$$D(\boldsymbol{z}|\boldsymbol{s}) \triangleq \int_R \delta_Z(\boldsymbol{z}) f(Z|\boldsymbol{s}) \delta Z, \tag{5.6}$$

where the posterior $f(Z|\boldsymbol{s}) \triangleq f_{\mathsf{Z}|\mathbf{s}}(Z|\boldsymbol{s})$ can be calculated from the likelihood function $g(\boldsymbol{s}|Z)$ and prior $f_{\mathsf{Z}}(Z)$ by using Bayes' rule, i.e.,

$$f(Z|\boldsymbol{s}) = \frac{g(\boldsymbol{s}|Z) f_{\mathsf{Z}}(Z)}{\int_R g(\boldsymbol{s}|Z) f_{\mathsf{Z}}(Z) \delta Z}. \tag{5.7}$$

An estimation procedure based on the posterior PHD is outlined in [16, pp. 504–505] and consists of the following two steps.

Algorithm 1 PHD Estimator for One Image

1: Calculate the expected number of objects given the image observation (cf. (4.66))

$$\mathsf{E}[|\mathsf{Z}||\boldsymbol{s}] = \int_R D(\boldsymbol{z}|\boldsymbol{s}) \mathrm{d}\lambda_N(\boldsymbol{z}) \tag{5.8}$$

and round it to the nearest integer to obtain an estimate for the number of objects

$$\hat{K} = \mathrm{round}\big(\mathsf{E}[|\mathsf{Z}||\boldsymbol{s}]\big). \tag{5.9}$$

2: Determine the position of the $\hat{K}$ highest local maxima $\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_{\hat{K}}$ of the posterior PHD $D(\boldsymbol{z}|\boldsymbol{s})$. The estimate $\hat{Z}$ is then given by

$$\hat{Z} = \{\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_{\hat{K}}\}. \tag{5.10}$$

In general, this estimation problem does not admit closed form solutions since the posterior PDF $f(Z|\boldsymbol{s})$ (and hence the posterior PHD and the mean number of objects) cannot be evaluated explicitly. However, as we will show subsequently, if the prior is Poisson distributed (cf. Section 4.5.2) the posterior PHD can be determined up to a constant factor. A proof of the following proposition is provided in Appendix A.7.

Proposition 5.1: Let the RFS Z be a Poisson RFS (cf. Section 4.5.2) with CD $p_{|\mathsf{Z}|}(k) = \frac{e^{-\mu}\mu^k}{k!}$, $\mu > 0$ and elementary vector PDF $\psi(\boldsymbol{z})$, i.e., the EVDs of Z are given by

$$\tilde{f}_{\mathsf{Z}}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) = e^{-\mu}\mu^k \prod_{i=1}^{k} \psi(\boldsymbol{z}_i), \qquad k \in \mathbb{N}_0. \tag{5.11}$$

Then the posterior $f(Z|\boldsymbol{s})$ is Poisson distributed as well, i.e.,

$$\tilde{f}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k|\boldsymbol{s}) = e^{-\nu(\boldsymbol{s})}\nu(\boldsymbol{s})^k \prod_{i=1}^{k} f(\boldsymbol{z}_i|\boldsymbol{s}), \qquad k \in \mathbb{N}_0, \tag{5.12}$$

with mean

$$\nu(\boldsymbol{s}) = \mu\alpha(\boldsymbol{s}), \tag{5.13}$$

$$\alpha(\boldsymbol{s}) \triangleq \int_R \psi(\boldsymbol{z})\eta(\boldsymbol{z}, \boldsymbol{s})\mathrm{d}\lambda_N(\boldsymbol{z}), \tag{5.14}$$

and elementary vector PDF

$$f(\boldsymbol{z}|\boldsymbol{s}) = \frac{\psi(\boldsymbol{z})\eta(\boldsymbol{z}, \boldsymbol{s})}{\alpha(\boldsymbol{s})}. \tag{5.15}$$

Corollary 5.2: Under the premise of Proposition 5.1, the expected number of objects $\mathsf{E}[|\mathsf{Z}||\boldsymbol{s}]$ and the posterior PHD $D(\boldsymbol{z}|\boldsymbol{s})$ are given by

$$\mathsf{E}[|\mathsf{Z}||\boldsymbol{s}] = \mu\alpha(\boldsymbol{s}), \tag{5.16}$$

$$D(\boldsymbol{z}|\boldsymbol{s}) = \mu\psi(\boldsymbol{z})\eta(\boldsymbol{z}, \boldsymbol{s}). \tag{5.17}$$

Proof: This follows immediately from (4.83) by inserting $\nu(\boldsymbol{s})$ and $f(\boldsymbol{z}|\boldsymbol{s})$ from Proposition 5.1. □

5.2 State Estimation from Two Images

We are now extending the estimation problem from the previous section to the case where two partly overlapping grayscale images are observed.

5.2.1 Data Model

As before, let $\boldsymbol{z} \in \mathbb{R}^N$ be a state vector and $\mathbf{s}_i = (\mathsf{s}_i^{(1)}, \ldots, \mathsf{s}_i^{(m)}) \in \mathbb{R}^m$, $i = 1, 2$ with $m = m_1 m_2$ be two random vectors consisting of the pixel values of two $m_1 \times m_2$ grayscale images. The functions $T_i(\boldsymbol{z})$ give the set of pixel-indices in image observation i that are illuminated by an object with state $\boldsymbol{z}$. The conditional PDF of pixel $s_i^{(j)}$ given the state $\boldsymbol{z}$ is (cf. (5.1))

$$p(s_i^{(j)}|\boldsymbol{z}) = \begin{cases} \varphi_i(s_i^{(j)}, \boldsymbol{z}) & \text{if } j \in T_i(\boldsymbol{z}), \\ \phi_i(s_i^{(j)}) & \text{if } j \notin T_i(\boldsymbol{z}), \end{cases} \tag{5.18}$$

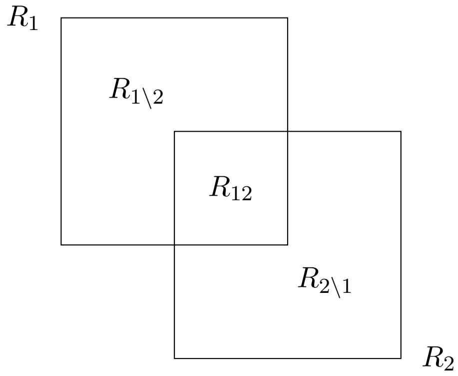

Fig. 5.1: *Example of two overlapping state spaces $R_1 = R_{1\backslash 2} \cup R_{12}$ associated with image s_1, and $R_2 = R_{12} \cup R_{2\backslash 1}$ associated with image s_2.*

with arbitrary but known PDFs $\varphi_i(s_i^{(j)}, \boldsymbol{z})$ and $\phi_i(s_i^{(j)})$.

We define the local state spaces R_i consisting of all state vectors $\boldsymbol{z} \in \mathbb{R}^N$ that illuminate at least one pixel in image $\boldsymbol{s}_i$, i.e.,

$$R_i \triangleq \{\boldsymbol{z} \in \mathbb{R}^N : T_i(\boldsymbol{z}) \neq \emptyset\}, \qquad i = 1, 2. \tag{5.19}$$

The global state space R is the union of the local state spaces,

$$R \triangleq R_1 \cup R_2. \tag{5.20}$$

We explicitly assume that $R_{12} \triangleq R_1 \cap R_2 \neq \emptyset$, i.e., there are states $\boldsymbol{z} \in R$ that illuminate pixels both in $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ (see Figure 5.1).

As before, we model the random objects as a RFS $\mathsf{Z} \subseteq R$ with given (prior) PDF $f_{\mathsf{Z}}(Z)$. For each image, we are interested in estimating the local RFSs

$$\mathsf{X} \triangleq \mathsf{Z} \cap R_1 \subseteq R_1, \tag{5.21}$$

$$\mathsf{Y} \triangleq \mathsf{Z} \cap R_2 \subseteq R_2. \tag{5.22}$$

That is, X contains only the elements of Z that lie in R_1 and Y contains only the elements of Z that lie in R_2. Since R_1 and R_2 overlap, there might of course be some states that are part of both X and Y (this happens if the state is in the region R_{12}). Hence, we expect to obtain a better estimate of X given both observations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ rather than using only the local observation $\boldsymbol{s}_1$ (and similary for Y).

Furthermore, we assume again that the local likelihood functions $g_1(\boldsymbol{s}_1|X)$ and $g_2(\boldsymbol{s}_2|Y)$ are separable and can be written as (cf. (5.3))

$$g_1(\boldsymbol{s}_1|X) = \pi_1(\boldsymbol{s}_1) \prod_{\boldsymbol{x}\in X} \eta_1(\boldsymbol{x}, \boldsymbol{s}_1), \tag{5.23}$$

$$g_2(\boldsymbol{s}_2|Y) = \pi_2(\boldsymbol{s}_2) \prod_{\boldsymbol{y}\in Y} \eta_2(\boldsymbol{y}, \boldsymbol{s}_2), \tag{5.24}$$

where for $i = 1, 2$ (cf. (5.4) and (5.5))

$$\pi_i(\boldsymbol{s}_i) = \prod_{j=1}^{m} \phi_i(\boldsymbol{s}_i^{(j)}), \tag{5.25}$$

$$\eta_i(\boldsymbol{z}, \boldsymbol{s}_i) = \prod_{j\in T_i(\boldsymbol{z})} \frac{\varphi_i(\boldsymbol{s}_i^{(j)}, \boldsymbol{z})}{\phi_i(\boldsymbol{s}_i^{(j)})}. \tag{5.26}$$

Since by construction X contains all state vectors that influence $\mathbf{s}_1$, the image observation $\mathbf{s}_1$ is conditionally independent of Y given X. By a similar argument $\mathbf{s}_2$ is conditionally independent of X given Y. Therefore, the conditional PDFs of $\mathbf{s}_i$ given X and Y are

$$g(\boldsymbol{s}_1|X, Y) = g_1(\boldsymbol{s}_1|X), \tag{5.27}$$

$$g(\boldsymbol{s}_2|X, Y) = g_2(\boldsymbol{s}_2|Y). \tag{5.28}$$

Because of (5.18) $\mathbf{s}_1$ is conditionally independent of $\mathbf{s}_2$ given X and Y. Thus, the global likelihood function $g(\boldsymbol{s}_1, \boldsymbol{s}_2|X, Y)$ factorizes as

$$g(\boldsymbol{s}_1, \boldsymbol{s}_2|X, Y) = g(\boldsymbol{s}_1|X, Y)g(\boldsymbol{s}_2|X, Y) = g_1(\boldsymbol{s}_1|X)g_2(\boldsymbol{s}_2|Y). \tag{5.29}$$

Using Bayes' rule we obtain the global posterior $f(X, Y|\boldsymbol{s}_1, \boldsymbol{s}_2)$ as

$$f(X, Y|\boldsymbol{s}_1, \boldsymbol{s}_2) = C g_1(\boldsymbol{s}_1|X)g_2(\boldsymbol{s}_2|Y)f(X, Y), \tag{5.30}$$

where $C = C(\boldsymbol{s}_1, \boldsymbol{s}_2)$ is a normalization constant given by

$$C = \left(\int_{R_2}\int_{R_1} g_1(\boldsymbol{s}_1|X)g_2(\boldsymbol{s}_2|Y)f(X, Y)\delta X\delta Y\right)^{-1} \tag{5.31}$$

and $f(X, Y)$ is the joint prior PDF of X and Y (which has yet to be determined from the given global prior PDF $f_{\mathsf{Z}}(Z)$).

5.2.2 Estimator

Our task is to obtain an estimate $\hat{X}$ of the realization $\mathsf{X} = X$ based on the image observations $\mathbf{s}_1 = \boldsymbol{s}_1$ and $\mathbf{s}_2 = \boldsymbol{s}_2$. (Due to the symmetry of the problem it is straightforward to modify the derived results for estimation of $\mathsf{Y} = Y$). Similarly to the case with only one image in Section 5.1.2 we use an estimator based on the posterior PHD (cf. (5.6))

$$D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2) \triangleq \int_{R_1} \delta_X(\boldsymbol{x}) f(X|\boldsymbol{s}_1, \boldsymbol{s}_2)\delta X, \tag{5.32}$$

where the marginal posterior $f(X|\boldsymbol{s}_1, \boldsymbol{s}_2)$ can be calculated from the global posterior given in (5.30) by integrating out Y, i.e.,

$$f(X|\boldsymbol{s}_1, \boldsymbol{s}_2) = \int_{R_2} f(X, Y|\boldsymbol{s}_1, \boldsymbol{s}_2)\delta Y = C \int_{R_2} g_1(\boldsymbol{s}_1|X) g_2(\boldsymbol{s}_2|Y) f(X, Y)\delta Y. \tag{5.33}$$

In analogy to Algorithm 1, an estimation procedure based on the posterior PHD is as follows.

Algorithm 2 PHD Estimator for Two Images

1: Calculate the expected number of objects given the image observations

$$\mathsf{E}[|\mathsf{X}||\boldsymbol{s}_1, \boldsymbol{s}_2] = \int_{R_1} D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)\mathrm{d}\lambda_N(\boldsymbol{x}) \tag{5.34}$$

and round it to the nearest integer to obtain an estimate for the number of objects

$$\hat{K} = \mathrm{round}\big(\mathsf{E}[|\mathsf{X}||\boldsymbol{s}_1, \boldsymbol{s}_2]\big). \tag{5.35}$$

2: Determine the position of the $\hat{K}$ highest local maxima $\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\hat{K}}$ of the posterior PHD $D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$. The estimate $\hat{X}$ is then given by

$$\hat{X} = \{\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\hat{K}}\}. \tag{5.36}$$

Since $g_1(\boldsymbol{s}_1|X)$ and $g_2(\boldsymbol{s}_2|Y)$ are given by (5.23) and (5.24) it remains to determine the joint PDF $f(X, Y)$ from the global prior $f_{\mathsf{Z}}(Z)$ and calculate the integral in (5.33). In general, (5.33) does not admit closed form solutions. However, as we will see, if the global prior $f_{\mathsf{Z}}(Z)$ is Poisson distributed then the marginal posterior $f(X|\boldsymbol{s}_1, \boldsymbol{s}_2)$ is Poisson distributed as well and the posterior PHD $D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ can be determined up to a constant factor.

We start by showing the following important result (a proof is provided in Appendix A.8).

Lemma 5.3: Let $f(Z)$ be a PDF on $\mathcal{Z}(R)$ (cf. (3.10)) with some closed $R \subseteq \mathbb{R}^N$ and let $R_1 \subseteq R$ and $R_2 \subseteq R$. Then

$$\int_{R_1 \cup R_2} f(Z)\delta Z = \int_{R_2 \setminus R_1} \int_{R_1} f(X \cup Y)\delta X \delta Y.$$

Before proceeding, we define the Dirac measure on set spaces.

Definition 5.4: *Let $\mathcal{X}$ be a set space as in* (3.1) *and let $\Sigma_{\mathcal{X}}$ be a σ-algebra on $\mathcal{X}$. The* ***Dirac measure*** *on $\Sigma_{\mathcal{X}}$ for an arbitrary $X_0 \in \mathcal{X}$ is defined by*

$$\delta_{X_0}(\mathcal{A}) \triangleq \mathbf{1}_{\mathcal{A}}(X_0) = \begin{cases} 1 & \text{if } X_0 \in \mathcal{A}, \\ 0 & \text{if } X_0 \notin \mathcal{A}, \end{cases} \tag{5.37}$$

for all $\mathcal{A} \in \Sigma_{\mathcal{X}}$.

It can be easily shown that for a measurable real-valued function $f(X)$ the following holds [51]

$$\int_{\mathcal{X}} f(X)\mathrm{d}\delta_{X_0}(X) = f(X_0). \tag{5.38}$$

We will follow the common abuse of notation and write this integral as

$$\int_{\mathcal{X}} f(X)\delta_{X_0}(X)\delta X \triangleq \int_{\mathcal{X}} f(X)\mathrm{d}\delta_{X_0}(X) = f(X_0), \tag{5.39}$$

and refer to $\delta_{X_0}(X)$ as the *Dirac delta function* centered at X_0. The equation above is also known as the *sifting property* of the Dirac delta function. For a proof of the following proposition see Appendix A.9.

Proposition 5.5: Let $f_{\mathsf{Z}}(Z)$ be the PDF of the global RFS Z. Then the joint PDF of the local RFSs $\mathsf{X} = \mathsf{Z} \cap R_1$ and $\mathsf{Y} = \mathsf{Z} \cap R_2$ is given by

$$f(X, Y) = f_{\mathsf{Z}}(X \cup Y)\delta_{X_{12}}(Y_{12}),$$

where $\delta_{X_{12}}(Y_{12})$ is the Dirac delta function as in (5.39), $X_{12} \triangleq X \cap R_{12}$ and $Y_{12} \triangleq Y \cap R_{12}$.

Finally, utilizing Lemma 5.3 and Proposition 5.5, the following can be shown (a proof can be found in Appendix A.10).

Proposition 5.6: Let the RFS Z be a Poisson RFS (cf. Section 4.5.2) with CD $p_{|\mathsf{Z}|}(k) = \frac{e^{-\mu}\mu^k}{k!}$, $\mu > 0$ and elementary vector PDF $\psi(\boldsymbol{z})$, i.e., the EVDs of Z are given by

$$\tilde{f}_{\mathsf{Z}}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) = e^{-\mu}\mu^k \prod_{i=1}^{k} \psi(\boldsymbol{z}_i), \qquad k \in \mathbb{N}_0. \tag{5.40}$$

Then the marginal posterior $f(X|\boldsymbol{s}_1, \boldsymbol{s}_2)$ is Poisson distributed as well, i.e.,

$$\tilde{f}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k|\boldsymbol{s}_1, \boldsymbol{s}_2) = e^{-\nu}\nu^k \prod_{i=1}^{k} f(\boldsymbol{x}_i|\boldsymbol{s}_1, \boldsymbol{s}_2), \qquad k \in \mathbb{N}_0, \tag{5.41}$$

with mean

$$\nu \triangleq \mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2), \tag{5.42}$$

$$\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2) \triangleq \int_{R_1} \psi(\boldsymbol{x})\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2)\mathrm{d}\lambda_N(\boldsymbol{x}), \tag{5.43}$$

$$\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2) \triangleq \begin{cases} \eta_2(\boldsymbol{x}, \boldsymbol{s}_2) & \text{if } \boldsymbol{x} \in R_{12} \\ 1 & \text{otherwise.} \end{cases} \tag{5.44}$$

and elementary vector PDF

$$f(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\psi(\boldsymbol{x})\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2)}{\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)}. \tag{5.45}$$

Corollary 5.7: Under the premise of Proposition 5.6, the expected number of objects $\mathsf{E}[|\mathsf{X}||\boldsymbol{s}_1, \boldsymbol{s}_2]$ and the posterior PHD $D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ are given by

$$\mathsf{E}[|\mathsf{X}||\boldsymbol{s}_1, \boldsymbol{s}_2] = \mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2), \tag{5.46}$$

$$D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2) = \mu\psi(\boldsymbol{x})\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2). \tag{5.47}$$

Proof: This follows immediately from (4.83) by inserting ν and $f(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ from Proposition 5.6. □

Note that for non-overlapping images, i.e., $R_{12} = \emptyset$, we have $\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2) \equiv 1$ and $\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2) = \int_{R_1} \psi(\boldsymbol{x})\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)\mathrm{d}\lambda_N(\boldsymbol{x}) = \alpha(\boldsymbol{s}_1)$ from (5.14). In this case, the image observation $\boldsymbol{s}_2$ does not contain any information about the objects in X and the expected number of objects $\mathsf{E}[|\mathsf{X}||\boldsymbol{s}_1, \boldsymbol{s}_2]$ and the posterior PHD $D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ given above reduce to (5.16) and (5.17) for the case of a single image.

5.3 Numerical Study

In this section, we implement the PHD estimators derived in Sections 5.1 and 5.2 in a 2-dimensional scenario and demonstrate the performance gain that can be achieved by using both image observations instead of just one.

5.3.1 Single Image Scenario

We will investigate grayscale images consisting of 45×45 pixels (i.e., $m_1 = m_2 = 45$). Each pixel covers a square area in $\mathbb{R}^2$ with side lengths Δ. The whole image therefore occupies the rectangle $[0, m_1\Delta] \times [0, m_2\Delta]$, as illustrated in Figure 5.2. An object (if present) is completely characterized by its state $\boldsymbol{z} = (z_1, z_2) \in \mathbb{R}^2$, which is simply the objects position in the 2-dimensional space. If an object with state $\boldsymbol{z}$ is present it influences a 4×4 array of pixels around it, where the center of the array is chosen as the nearest point in the pixel grid $(a\Delta, b\Delta)$. As indicated in Figure 5.2, there might be fewer than 16 pixels affected by an object (in the extreme case there might be just one). Given an object with state $\boldsymbol{z}$ we model the random pixel values s_i by

$$\mathsf{s}_i = \begin{cases} \mathsf{n}_i + h(i, \boldsymbol{z}) & \text{if } i \in T(\boldsymbol{z}), \\ \mathsf{n}_i & \text{if } i \notin T(\boldsymbol{z}), \end{cases} \qquad i = 1, \ldots, m. \tag{5.48}$$

Here, n_i is zero mean white Gaussian noise with variance σ^2 and $h(i, \boldsymbol{z})$ is the point spread function given by [22]

$$h(i, \boldsymbol{z}) = \frac{\Delta^2 I}{2\pi\sigma_h^2} \exp\left(-\frac{((a-0.5)\Delta - z_1)^2 + ((b-0.5)\Delta - z_2)^2}{2\sigma_h^2}\right), \tag{5.49}$$

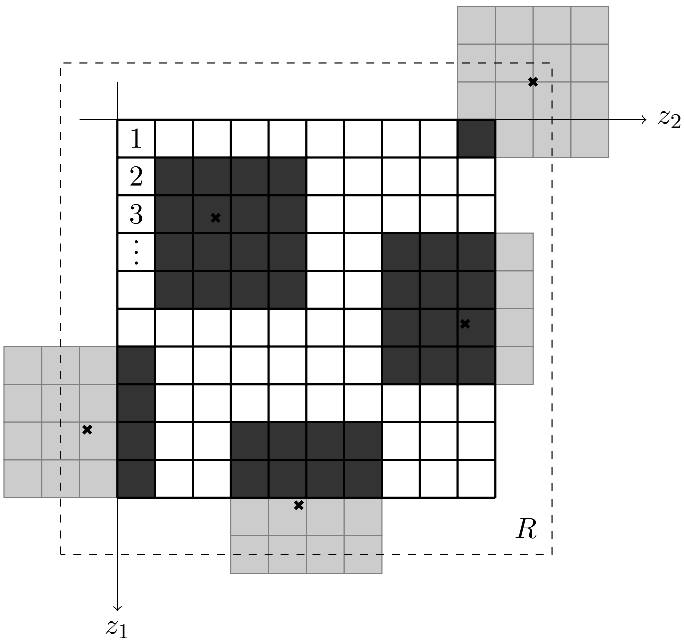

Fig. 5.2: *The relation between a grayscale image (solid grid) and the 2-dimensional Euclidean space. Shaded arrays represent the 4×4 arrays of pixels that are illuminated by objects located at the crosses. The dark shaded pixels are the ones that are part of the image. As can be seen, an object may influence less than 16 pixels. We require that the objects are located in the subspace $R \subseteq \mathbb{R}^2$, indicated by the dashed rectangle, so that at least one pixel is illuminated by every object.*

where I is the source intensity, σ_h^2 is the blurring factor, and $i = (a-1)m_2+b$ with $a \in \{1, \dots, m_1\}$ and $b \in \{1, \dots, m_2\}$. Note that the point spread function is essentially a Gaussian distribution with mean $\boldsymbol{z}$ and variance σ_h^2 that is additionally scaled by the factor $\Delta^2 I$.

The conditional PDF of pixel s_i given the state $\boldsymbol{z}$ can therefore be written as (cf. (5.1))

$$p(s_i|\boldsymbol{z}) = \begin{cases} \varphi(s_i, \boldsymbol{z}) & \text{if } i \in T(\boldsymbol{z}), \\ \phi(s_i) & \text{if } i \notin T(\boldsymbol{z}), \end{cases} \tag{5.50}$$

where

$$\varphi(s_i, \boldsymbol{z}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(s_i - h(i, \boldsymbol{z}))^2}{2\sigma^2}\right), \tag{5.51}$$

$$\phi(s_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s_i^2}{2\sigma^2}\right). \tag{5.52}$$

That is, an object with state $\boldsymbol{z}$ introduces a mean $h(i, \boldsymbol{z})$ to the noisy pixel values in its vicinity that declines exponentially with the distance from the pixel to the objects position. If multiple objects are present, there may be several objects influencing the same pixel. For example, if $Z(i) = \{\boldsymbol{z} \in Z : i \in T(\boldsymbol{z})\} \subseteq Z$, then $\mathsf{s}_i = \mathsf{n}_i + \sum_{\boldsymbol{z} \in Z(i)} h(i, \boldsymbol{z})$. We will however neglect this case and approximate the likelihood function to be separable as (cf. (5.3)) [22]

$$g(\boldsymbol{s}|Z) = \pi(\boldsymbol{s}) \prod_{\boldsymbol{z}\in Z} \eta(\boldsymbol{z}, \boldsymbol{s}), \tag{5.53}$$

where

$$\pi(\boldsymbol{s}) = \prod_{i=1}^{m} \phi(s_i) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s_i^2}{2\sigma^2}\right), \tag{5.54}$$

$$\eta(\boldsymbol{z}, \boldsymbol{s}) = \prod_{i\in T(\boldsymbol{z})} \frac{\varphi(s_i, \boldsymbol{z})}{\phi(s_i)} \tag{5.55}$$

$$= \prod_{i\in T(\boldsymbol{z})} \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(s_i - h(i,\boldsymbol{z}))^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s_i^2}{2\sigma^2}\right)} \tag{5.56}$$

$$= \exp\left(\sum_{i\in T(\boldsymbol{z})} \frac{2s_i h(i,\boldsymbol{z}) - h^2(i,\boldsymbol{z})}{2\sigma^2}\right). \tag{5.57}$$

Of course, with this approximation we will incur an additional error whenever two or more objects are located close enough to each other to affect the same pixel. Under this premise, our results from Section 5.1 hold approximately and the posterior PHD is given by (cf. (5.17))

$$D(\boldsymbol{z}|\boldsymbol{s}) = \mu\psi(\boldsymbol{z})\eta(\boldsymbol{z}, \boldsymbol{s}). \tag{5.58}$$

The RFS Z is taken as a Poisson RFS with mean μ and uniformly distributed elementary vector PDF $\psi(\boldsymbol{z}) \equiv \frac{1}{\lambda_2(R)}$. With this assumption, the peak localization of the posterior PHD $D(\boldsymbol{z}|\boldsymbol{s})$ is equivalent to finding the local maxima of

$$F(\boldsymbol{z}|\boldsymbol{s}) \triangleq \log(\eta(\boldsymbol{z}, \boldsymbol{s})) = \sum_{i\in T(\boldsymbol{z})} \frac{2s_i h(i,\boldsymbol{z}) - h^2(i,\boldsymbol{z})}{2\sigma^2}, \tag{5.59}$$

where (5.57) has been used. Note that $\mathsf{E}[|\mathsf{Z}||\boldsymbol{s}]$ (cf. (5.16)) cannot be determined explicitly since there is no closed form solution to $\alpha(\boldsymbol{s}) = \int_R \psi(\boldsymbol{z})\eta(\boldsymbol{z}, \boldsymbol{s}) \mathrm{d}\lambda_2(\boldsymbol{z})$. A good approximation of $\alpha(\boldsymbol{s})$ is numerically extremely challenging due to a high dynamic range of $\eta(\boldsymbol{z}, \boldsymbol{s})$. The problem can be appreciated by looking at Figure 5.3 where three images and the corresponding $\log(\eta(\boldsymbol{z}, \boldsymbol{s}))$ functions are shown for three different source intensities with the parameters $\Delta = 1$, $\sigma^2 = 1$, $\sigma_h^2 = 1$, and $\mu = 2$. As can been seen, even for moderate intensities up to $I = 100$, the range of $\log(\eta(\boldsymbol{z}, \boldsymbol{s}))$ extends from around -500 to 500. Hence, $\eta(\boldsymbol{z}, \boldsymbol{s})$ has extremely narrow localized high peaks (approaching dirac delta functions) impeding the application of efficient numerical integration techniques.

Therefore, we propose a different estimator in Algorithm 3 where we select the highest peaks of the posterior PHD $D(\boldsymbol{z}|\boldsymbol{s})$ that lie above a certain threshold as estimates. The parameter N_s in Algorithm 3 should be chosen such that there is at least one $\bar{\boldsymbol{z}} \in \bar{Z}$ in the vicinity of each peak

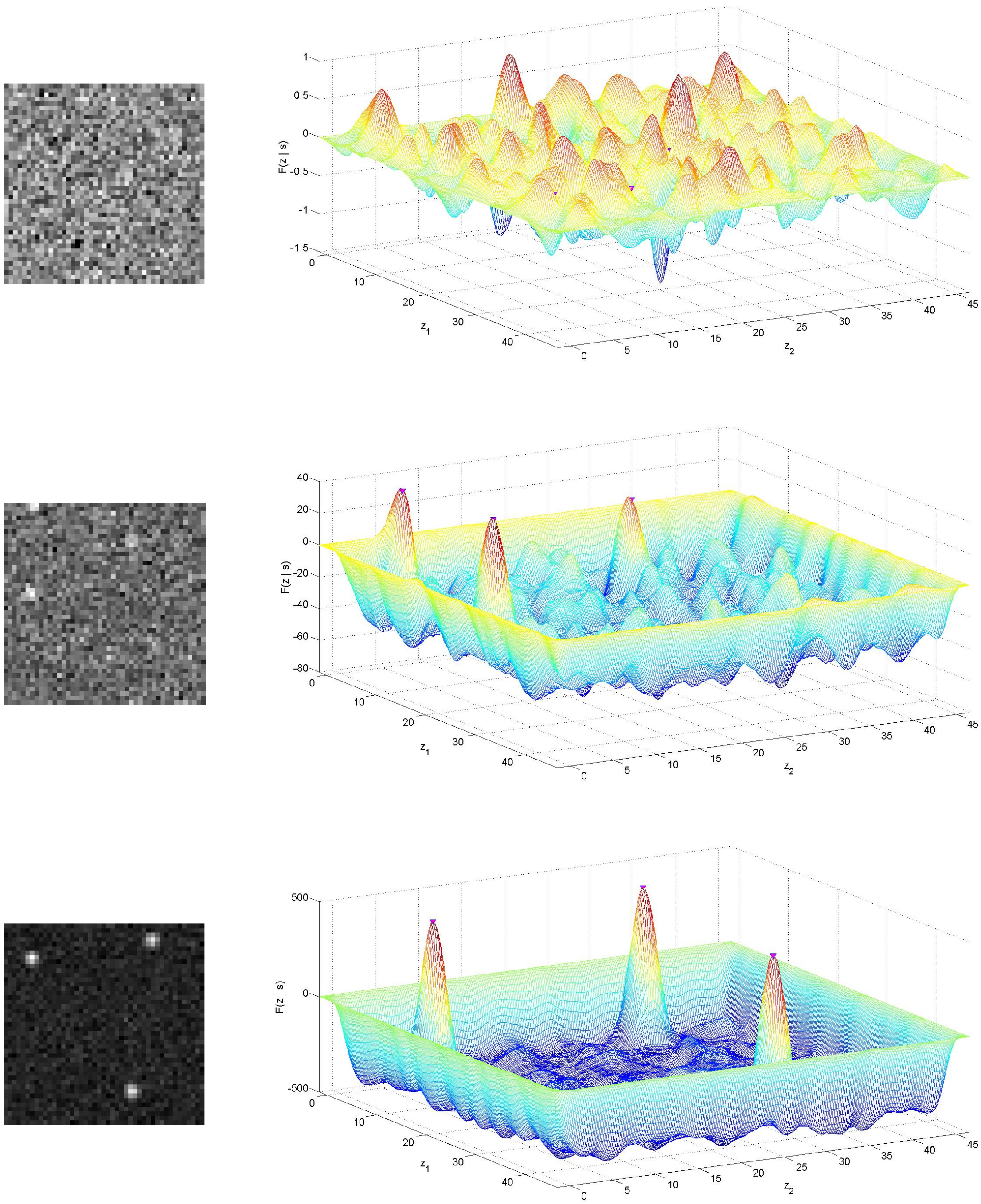

Fig. 5.3: *Different image realzations and their corresponding cost functions $F(\boldsymbol{z}) = \log(\eta(\boldsymbol{z}, \boldsymbol{s}))$. The source intensities I of the point spread function are from top to bottom $I = 1, 30, 100$. Actual object positions are indicated by purple markers.*

corresponding to an actual object to ensure that the subsequent gradient ascents reach these maxima (cf. Figure 5.3). In general, the threshold value F_t for discarding noise peaks depends on the noise variance σ^2 and the source intensity of the objects I. Ideally, it should be set such that all object maxima are kept and all noise maxima are discarded, which of course cannot be

Algorithm 3 Estimator for One Image

Uniformly sample $F(\boldsymbol{z}|\boldsymbol{s})$ at the pixel grid positions $\boldsymbol{z} = (a\Delta, b\Delta)$ with $a = 1, \ldots, m_1$, $b = 1, \ldots, m_2$ and collect the N_s highest points in $\bar{Z} = \{\bar{\boldsymbol{z}}_1, \ldots, \bar{\boldsymbol{z}}_{N_s}\}$.
for all $\bar{\boldsymbol{z}}_i \in \bar{Z}$ **do**
    Perform a gradient ascent algorithm with $\bar{\boldsymbol{z}}_i$ as initialization, a maximum of N_g iterations and step size $\Delta_g = 1/I^2$, resulting in a local maxima $\tilde{\boldsymbol{z}}_i$.
end for
Discard all local maxima $\tilde{\boldsymbol{z}}_i$ below a threshold level, $F(\tilde{\boldsymbol{z}}_i|\boldsymbol{s}) < F_t$.
Collect the $\tilde{N}_s$ remaining maxima in the set $\tilde{Z} = \{\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_{\tilde{N}_s}\}$.
$\hat{Z} = \emptyset$, $j = 1$.
while $\tilde{Z} \neq \emptyset$ **do**
    Select an arbitrary element $\tilde{\boldsymbol{z}} \in \tilde{Z}$.
    Find all $\tilde{\boldsymbol{z}}_i \in \tilde{Z}$ with $\|\tilde{\boldsymbol{z}}_i - \tilde{\boldsymbol{z}}\| < D$ and collect them in the set S_j.
    Add $\hat{\boldsymbol{z}}_j = \arg\max_{\tilde{\boldsymbol{z}}_i \in S_j} F(\tilde{\boldsymbol{z}}_i|\boldsymbol{s})$ as an estimate $\hat{Z} \leftarrow \hat{Z} \cup \{\hat{\boldsymbol{z}}_j\}$.
    Discard all points in S_j from $\tilde{Z}$.
    $j \leftarrow j + 1$.
end while
Output the estimate $\hat{Z} = \{\hat{\boldsymbol{z}}_1, \ldots, \hat{\boldsymbol{z}}_{\hat{N}}\}$.

achieved for all parameter pairs σ^2 and I (cf. Figure 5.3 (top) where the actual objects position are drowned in noise). Finally, the parameter D should be chosen such that all local maxima estimates of the gradient ascents corresponding to the same peak are clustered together.

5.3.2 Two Images Scenario

In this section, we extend our scenario with a single image from the previous section to the case where two partly overlapping images are observed. We use the same parameters as discussed in Section 5.3.1 with the main difference that we now have two instead of one image. The first image $\boldsymbol{s}_1$ occupies a region of the state space as indicated in Figure 5.2. The setup of the second image $\boldsymbol{s}_2$ is identical except that its upper left corner is located at the position $\boldsymbol{p} \in \mathbb{R}^2$. Given an object with state $\boldsymbol{z}$ we model the random pixel values $\mathsf{s}_i^{(j)}$ of image $i = 1, 2$ by (cf. (5.48))

$$\mathsf{s}_i^{(j)} = \begin{cases} \mathsf{n}_i^{(j)} + h_i(j, \boldsymbol{z}) & \text{if } j \in T_i(\boldsymbol{z}), \\ \mathsf{n}_i^{(j)} & \text{if } j \notin T_i(\boldsymbol{z}), \end{cases} \qquad j = 1, \ldots, m. \tag{5.60}$$

Here, $\mathsf{n}_i^{(j)}$ is zero mean white Gaussian noise with variance $\sigma_i^2 = 1$ and $h_i(j, \boldsymbol{z})$ is the point spread function for image i as in (5.49) The conditional PDF of pixel $\mathsf{s}_i^{(j)}$ given the state $\boldsymbol{z}$ can therefore be written as (cf. (5.50))

$$p(\mathsf{s}_i^{(j)}|\boldsymbol{z}) = \begin{cases} \varphi_i(s_i^{(j)}, \boldsymbol{z}) & \text{if } j \in T_i(\boldsymbol{z}), \\ \phi_i(s_i^{(j)}) & \text{if } j \notin T_i(\boldsymbol{z}), \end{cases} \tag{5.61}$$

where (cf. (5.51) and (5.52))

$$\varphi_i(s_i^{(j)}, \boldsymbol{z}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(s_i^{(j)} - h_i(j, \boldsymbol{z}))^2}{2\sigma_i^2}\right), \tag{5.62}$$

$$\phi_i(s_i^{(j)}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(s_i^{(j)})^2}{2\sigma_i^2}\right). \tag{5.63}$$

Again, we neglect the cases where two or more objects illuminate the same pixel in the same image and approximate the local likelihood functions to be separable (cf. (5.23) and (5.24))

$$g_1(\boldsymbol{s}_1|X) = \pi_1(\boldsymbol{s}_1) \prod_{\boldsymbol{x}\in X} \eta_1(\boldsymbol{x}, \boldsymbol{s}_1), \tag{5.64}$$

$$g_2(\boldsymbol{s}_2|Y) = \pi_2(\boldsymbol{s}_2) \prod_{\boldsymbol{y}\in Y} \eta_2(\boldsymbol{y}, \boldsymbol{s}_2), \tag{5.65}$$

where for $i = 1, 2$ (cf. (5.25) and (5.26))

$$\pi_i(\boldsymbol{s}_i) = \prod_{j=1}^{m} \phi_i(s_i^{(j)}) = \prod_{j=1}^{m} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{[s_i^{(j)}]^2}{2\sigma_i^2}\right), \tag{5.66}$$

$$\eta_i(\boldsymbol{z}, \boldsymbol{s}_i) = \prod_{j\in T_i(\boldsymbol{z})} \frac{\varphi_i(s_i^{(j)}, \boldsymbol{z})}{\phi_i(s_i^{(j)})} \tag{5.67}$$

$$= \prod_{j\in T_i(\boldsymbol{z})} \frac{\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(s_i^{(j)} - h_i(j,\boldsymbol{z}))^2}{2\sigma_i^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{[s_i^{(j)}]^2}{2\sigma_i^2}\right)} \tag{5.68}$$

$$= \exp\left(\sum_{j\in T_i(\boldsymbol{z})} \frac{2s_i^{(j)} h_i(j, \boldsymbol{z}) - h_i^2(j, \boldsymbol{z})}{2\sigma_i^2}\right). \tag{5.69}$$

Under this premise, our results from Section 5.2 hold approximately and the posterior PHD is given by (cf. (5.47))

$$D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2) = \mu\psi(\boldsymbol{x})\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2). \tag{5.70}$$

The RFS Z is chosen as a Poisson RFS with mean $\mu = 2$ and uniformly distributed elementary vector PDF $\psi(\boldsymbol{z}) \equiv \frac{1}{\lambda_2(R)}$. With this assumption, the peak localization of the posterior PHD $D(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ is equivalent to finding the local maxima of

$$F(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2) \triangleq \log(\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)) + \log(\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2)), \tag{5.71}$$

We use an estimation algorithm analogous to Algorithm 3, with the difference being that we now operate on $F(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ instead of $F(\boldsymbol{z}|\boldsymbol{s})$.

Algorithm 4 Estimator for Two Images

Uniformly sample $F(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ at the pixel grid positions $\boldsymbol{x} = (a\Delta, b\Delta)$ with $a = 1, \ldots, m_1$, $b = 1, \ldots, m_2$ and collect the N_s highest points in $\bar{X} = \{\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_{N_s}\}$.
for all $\bar{\boldsymbol{x}}_i \in \bar{X}$ **do**
    Perform a gradient ascent algorithm with $\bar{\boldsymbol{x}}_i$ as initialization, a maximum of N_g iterations and step size $\Delta_g = 1/I^2$, resulting in a local maxima $\tilde{\boldsymbol{x}}_i$.
end for
Discard all local maxima $\tilde{\boldsymbol{x}}_i$ below a threshold level, $F(\tilde{\boldsymbol{x}}_i|\boldsymbol{s}_1, \boldsymbol{s}_2) < F_t$.
Collect the $\tilde{N}_s$ remaining maxima in the set $\tilde{X} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{\tilde{N}_s}\}$.
$\hat{X} = \emptyset$, $j = 1$.
while $\tilde{X} \neq \emptyset$ **do**
    Select an arbitrary element $\tilde{\boldsymbol{x}} \in \tilde{X}$.
    Find all $\tilde{\boldsymbol{x}}_i \in \tilde{X}$ with $\|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}\| < D$ and collect them in the set S_j.
    Add $\hat{\boldsymbol{x}}_j = \arg\max_{\tilde{\boldsymbol{x}}_i \in S_j} F(\tilde{\boldsymbol{x}}_i|\boldsymbol{s}_1, \boldsymbol{s}_2)$ as an estimate $\hat{X} \leftarrow \hat{X} \cup \{\hat{\boldsymbol{x}}_j\}$.
    Discard all points in S_j from $\tilde{X}$.
    $j \leftarrow j + 1$.
end while
Output the estimate $\hat{X} = \{\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{\hat{N}}\}$.

The top of Figure 5.4 shows typical realizations of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ for $I = 30$, $\Delta = 1$, $\sigma_i^2 = 1$, $\sigma_h^2 = 1$, $\mu = 2$, and $\boldsymbol{p} = (14.05, 14.05)$, where the red corners indicate the shared region of both images. The corresponding cost functions $F(\boldsymbol{x}|\boldsymbol{s}_1)$ (utilizing just the single image $\boldsymbol{s}_1$ as in (5.59)) and $F(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ (utilizing both images) are shown in the middle and bottom part, respectively. As can be seen by comparing the two cost functions, the contribution of image $\boldsymbol{s}_2$ (via the right term in (5.71)) is an enhancement of the target peaks and a reduction of the background noise in the overlap region, thereby facilitating object detection. Furthermore, the maxima of the peaks in the overlap region are better aligned with the actual object positions (indicated by purple markers), thereby reducing the localization error (see next section). Note however that the two cost functions are identical outside the overlap region, which means that the detection and localization of objects outside this region does not improve compared with the single image case. Under the assumption of uniformly distributed objects we therefore expect a gain in the estimator performance, which should increase with the size of the overlap region.

5.3.3 Performance Evaluation

To assess the performance of our estimators we use the optimal subpattern assignment (OSPA) metric defined as follows [52].

Definition 5.8: *Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} \subseteq \mathbb{R}^N$ and $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\} \subseteq \mathbb{R}^N$ be two finite sets with $m \leq n$. Furthermore, let*

$$d^{(c)}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \min(c, \|\boldsymbol{x} - \boldsymbol{y}\|), \tag{5.72}$$

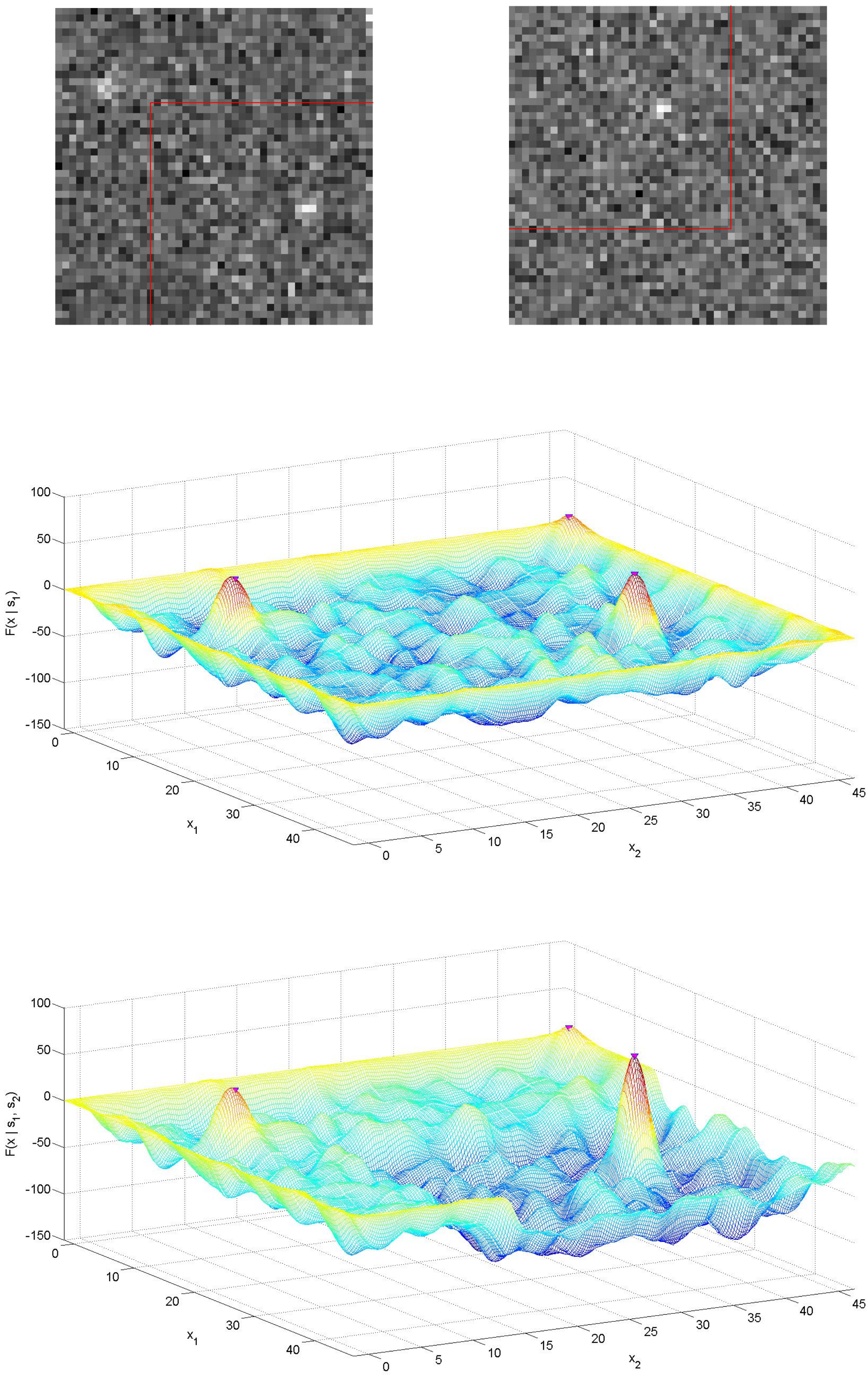

Fig. 5.4: *Typical image realizations of $\boldsymbol{s}_1$ (top left) and $\boldsymbol{s}_2$ (top right) for $I = 30$. The red corners indicate the shared region of both images. The corresponding cost functions $F(\boldsymbol{x}|\boldsymbol{s}_1)$ (utilizing just the single image $\boldsymbol{s}_1$) and $F(\boldsymbol{x}|\boldsymbol{s}_1, \boldsymbol{s}_2)$ (utilizing both images) are shown in the middle and bottom part, respectively. The purple markers indicate actual object positions.*

for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ with some $c > 0$. Here, $\|\boldsymbol{x}\|$ denotes the Euclidean norm on $\mathbb{R}^N$.

For $p \geq 1$, the p-th order OSPA metric with cutoff c is defined by

$$\bar{d}_p^{(c)}(X, Y) \triangleq \left(\frac{1}{n} \left(\min_{\sigma} \sum_{i=1}^{m} d^{(c)}(\boldsymbol{x}_i, \boldsymbol{y}_{\sigma_i})^p + c^p(n-m) \right) \right)^{\frac{1}{p}}, \tag{5.73}$$

where the minimization is performed over all $n!$ permutations σ over $1, \ldots, n$.

Note that for $p = 1$, the OSPA metric can be split into two additive terms

$$\bar{d}_1^{(c)}(X, Y) = d_{elem}^{(c)}(X, Y) + d_{card}^{(c)}(X, Y), \tag{5.74}$$

with

$$d_{elem}^{(c)}(X, Y) \triangleq \frac{1}{n} \min_{\sigma} \sum_{i=1}^{m} d^{(c)}(\boldsymbol{x}_i, \boldsymbol{y}_{\sigma_i}), \tag{5.75}$$

$$d_{card}^{(c)}(X, Y) \triangleq \frac{1}{n} c(n-m). \tag{5.76}$$

Here, $d_{elem}^{(c)}(X, Y)$ can be interpreted as a "per element distance" between X and Y. For each one of the $\binom{n}{m}$ m-ary subsets of Y, the $m!$ permutations are searched to yield a minimum of the sum in (5.75). The subset of Y yielding the lowest of these sums are the elements of Y nearest to the elements of X. The "cardinality distance" $d_{card}^{(c)}(X, Y)$ takes account of the different number of elements in X and Y.

The following parameters were used for our simulations:

- Data model parameters: $\Delta = 1$, $\sigma_i^2 = 1$, $\sigma_h^2 = 1$, and $\mu = 2$.
- Algorithm parameters: $N_s = 20$, $N_g = 100$, $D = 1.5\sqrt{2}$,

$$F_t = \begin{cases} a_1 \exp\left[-\frac{(I-a_2)^2}{a_3^2} \right] & \text{if } I < 30, \\ 20 & \text{if } I \geq 30, \end{cases} \tag{5.77}$$

  with $a_1 = 42.96$, $a_2 = 57.33$, and $a_3 = 31.21$. Here, the parameter F_t was determined empirically by noting suitable values for different realizations and performing curve fitting on the observations.

For our simulations we used an OSPA metric with $p = 1$ and cutoff $c = 30$. We calculated the Monte Carlo averaged OSPA distance $\bar{d}_1^{(30)}(X, \hat{X})$ between the actual objects X and our estimates $\hat{X}$ with 1000 trials for 50 values of the source intensity I from 1 to 100 and various overlap regions. The results are shown in Figure 5.5. We refer to $d_{elem}^{(30)}(X, \hat{X})$ and $d_{card}^{(30)}(X, \hat{X})$ as the *localization error* and the *cardinality error*, respectively. Their Monte Carlo averaged values are shown in Figure 5.5 as well.

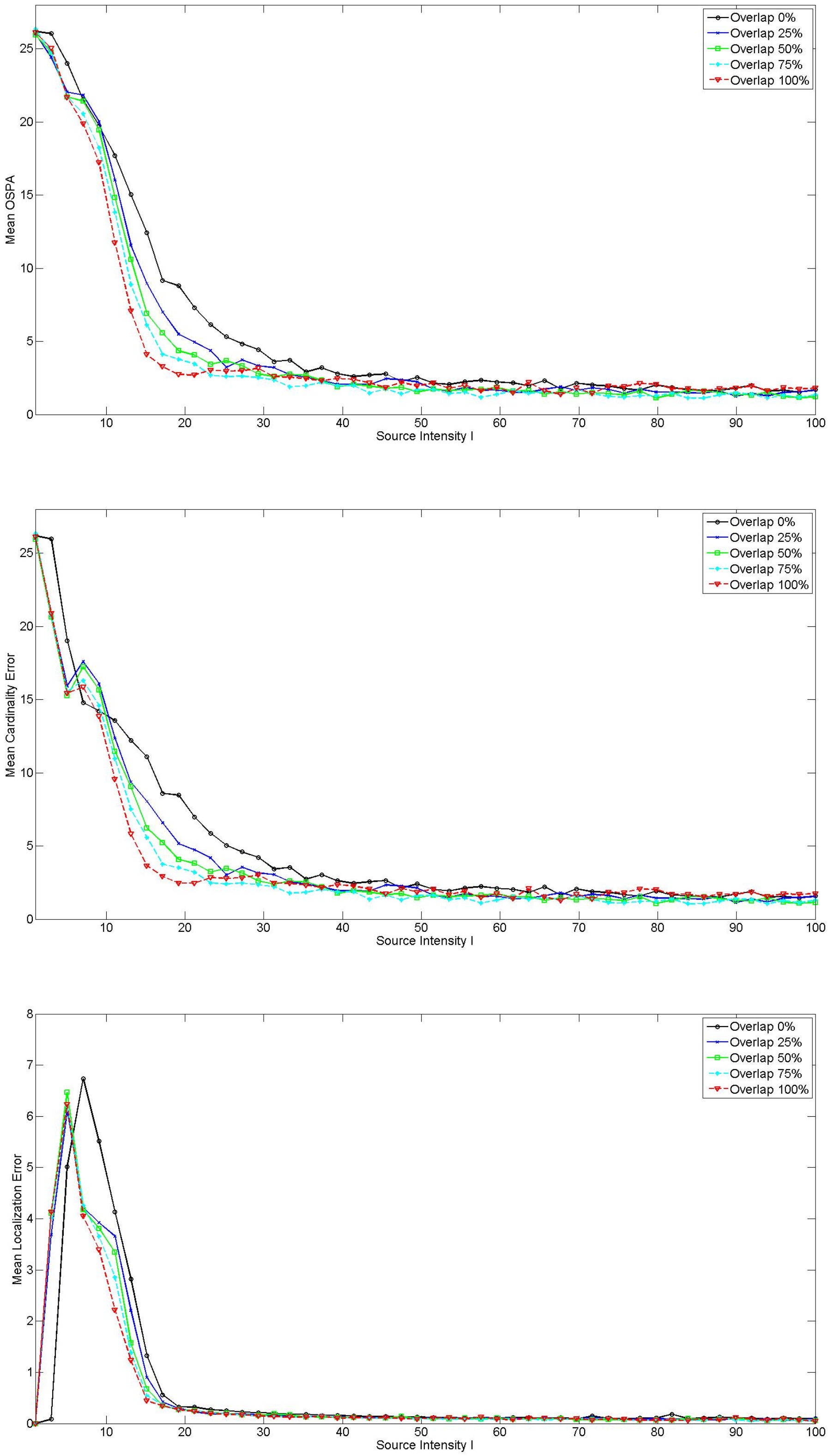

Fig. 5.5: *Monte Carlo averaged OSPA metric (top), cardinality error (middle), and localization error (bottom) for* $p = 1$ *and cutoff* $c = 30$.

As can be seen from the OSPA plot (top), using both images with Algorithm 4 generally outperforms Algorithm 3, which uses a single image (corresponding to 0% overlap). Also (as expected), bigger overlap regions achieve higher gains (with 100% overlap corresponding to two independent observations of the same scene). The biggest gains are achieved from around $I = 10$ to $I = 20$, that is for low “SNR” values, and decreases for higher values of I.

Comparing the cardinality error (middle) and the localization error (bottom), we see that the cardinality error is the main contribution to the overall OSPA distance. Interestingly, for I values from around 5 to 10, the algorithms incorporating both images are yielding a higher cardinality error than the single image estimator, which is probably caused by the hard detection threshold used. The error floor of around 2.5 for the cardinality error is due to the approximations of the local likelihood functions (5.64) and (5.65) and because of the clustering of objects which are near enough to each other such that their pixel sets in the image overlaps.

For intensity values bigger than 20, the localization error is mostly negligible compared to the cardinality error. The weaker localization performance of the algorithms utilizing two images from $I = 1$ to about $I = 5$, is caused by the inferior detection capability of the single image algorithm as shown in the cardinality error plot. From around $I = 5$ onward the object localization improves compared to the single image approach.

We want to point out, that there are certainly better implementations of our algorithms, since for an optimum performance the step sizes and discarding thresholds should be carefully matched to the source intensity values and noise variances.

Chapter 6

Conclusion

In this work, we provided a principled introduction to the theory of random finite sets with an emphasis on accessibility for readers completely new to the field. We then applied this theory to the problem of jointly estimating the number and states of an unknown and random number of objects based on image observations. We derived estimators based on the posterior probability hypothesis density for two different scenarios with a single image and with two partly overlapping images. Our simulations demonstrated a performance gain achieved by utilizing information from both images compared to the case where the estimation is based on only a single image.

The proposed estimation algorithms are rather simple and partly based on heuristics. Further improvements could be achieved by adopting more sophisticated gradient ascent algorithms, e.g., line search approaches to find better step sizes and speed up convergence. Moreover, finding methods to robustly and accurately approximate the posterior expected number of objects (which we circumvented by using a simple discarding scheme involving empirically determined thresholds) could yield another performance gain, especially with respect to cardinality errors in the low-SNR regime.

In our simulations, we investigated a two-dimensional scenario where most parameters were fixed. Further research is required to analyze the effects of different combinations of, e.g., image resolutions and sizes, prior distributions, noise variances, and object sizes. The study of scenarios involving more complex state spaces presents another venue for future research. Examples include the use of random source intensities for objects or augmenting the object states with a discrete attribute to distinguish between different object types with different appearances and orientations. Finally, it would be desirable to generalize our results to scenarios with more than two overlapping images. Depending on the image topologies considered, this could possibly lead to a form of information exchange between adjacent images that is similar to conventional message passing algorithms for estimating random vectors.

Appendix A

Mathematical Proofs

A.1 Proof of Lemma 3.3

The proof presented here is outlined in [46, p. 48]. First we prove that $\mathcal{X}_{\geq k} \in \mathcal{B}(\mathcal{X})$ by showing that $\mathcal{X}_{\geq k}$ is the union of specific base members of the hit-or-miss topology $\mathcal{T}$. Since $\mathcal{T}$ consists of all unions of base members (cf. Definitions 3.1 and 2.4), it then immediately follows that $\mathcal{X}_{\geq k} \in \mathcal{T}$. Furthermore, because $\mathcal{B}(\mathcal{X}) = \sigma(\mathcal{T})$ (see (3.4)), according to Definition 2.2 we have $\mathcal{T} \subseteq \mathcal{B}(\mathcal{X})$ and therefore $\mathcal{X}_{\geq k} \in \mathcal{B}(\mathcal{X})$.

Consider those base members $\mathcal{A} \in \mathcal{D}$ of the hit-or-miss topology $\mathcal{T}$ (cf. Definition 3.1) that have the following form:

- Let $n = k$, thus we have exactly k open sets $O_1, \dots, O_k$.
- Furthermore, let these open sets be disjoint: $O_i \cap O_j = \emptyset$ if $i \neq j$.
- Let $K = \emptyset$, which gives $\mathcal{X}^{\emptyset} = \{X \in \mathcal{X} : X \cap \emptyset = \emptyset\} = \mathcal{X}$.

According to (3.3), every base member $\mathcal{A}$ fulling these requirements can be written as

$$\mathcal{A} = \mathcal{X} \cap \mathcal{X}_{O_1} \cap \dots \cap \mathcal{X}_{O_k} = \mathcal{X}_{O_1} \cap \dots \cap \mathcal{X}_{O_k}. \tag{A.1}$$

Therefore, the collection of all these specific base members is

$$\mathcal{D}^{(k)} \triangleq \{\mathcal{X}_{O_1} \cap \dots \cap \mathcal{X}_{O_k} : O_i \in \mathcal{O} \text{ and } O_i \cap O_j = \emptyset \text{ if } i \neq j\}. \tag{A.2}$$

Denote the union of all these base members by

$$\mathcal{A}^{(k)} \triangleq \bigcup_{\mathcal{A} \in \mathcal{D}^{(k)}} \mathcal{A} \tag{A.3}$$

Of course, $\mathcal{A}^{(k)} \in \mathcal{T}$ since it is the union of base members of $\mathcal{T}$.

We will show that $\mathcal{A}^{(k)} = \mathcal{X}_{\geq k}$ by first showing $\mathcal{A}^{(k)} \subseteq \mathcal{X}_{\geq k}$ and then $\mathcal{X}_{\geq k} \subseteq \mathcal{A}^{(k)}$. Suppose $X \in \mathcal{A}^{(k)}$. By (A.3) and (A.2), there exist at least k disjoint $O_1, \dots, O_k$ such that $X \cap O_i \neq \emptyset$, for $i = 1, \ldots, k$. Since all O_i are disjoint, $|X| \geq k$, or equivalently $X \in \mathcal{X}_{\geq k}$. Hence, we have shown that $X \in \mathcal{A}^{(k)}$ implies $X \in \mathcal{X}_{\geq k}$. We thus conclude that $\mathcal{A}^{(k)} \subseteq \mathcal{X}_{\geq k}$.

Conversely, suppose $X \in \mathcal{X}_{\geq k}$. Then there are at least k different $\boldsymbol{x}_i \in \mathbb{R}^N$ such that $\boldsymbol{x}_i \in X$ for $i = 1, \ldots, k$. Since the underlying space is $\mathbb{R}^N$, there are k disjoint open sets $O_1, \ldots, O_k$ with $\boldsymbol{x}_i \in O_i$. Thus, $X \cap O_i \neq \emptyset$, which implies $X \in \mathcal{X}_{O_1} \cap \cdots \cap \mathcal{X}_{O_k}$. By (A.3) and (A.2), it follows $X \in \mathcal{A}^{(k)}$. Hence, we have shown that $X \in \mathcal{X}_{\geq k}$ implies $X \in \mathcal{A}^{(k)}$. We thus conclude that $\mathcal{X}_{\geq k} \subseteq \mathcal{A}^{(k)}$.

We will next show that $\mathcal{X}_{\leq k} \in \mathcal{B}(\mathcal{X})$. Observe that $\mathcal{X}_{\leq k} = (\mathcal{X}_{\geq k+1})^c$. Since $\mathcal{X}_{\geq k+1} \in \mathcal{B}(\mathcal{X})$ as shown above, by Definition 2.1, $(\mathcal{X}_{\geq k+1})^c \in \mathcal{B}(\mathcal{X})$, and therefore $\mathcal{X}_{\leq k} \in \mathcal{B}(\mathcal{X})$.

Finally, note that $\mathcal{X}_k = \mathcal{X}_{\leq k} \cap \mathcal{X}_{\geq k}$. Because $\mathcal{X}_{\leq k} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{X}_{\geq k} \in \mathcal{B}(\mathcal{X})$, and because $\mathcal{B}(\mathcal{X})$ is closed under countable intersections, we conclude $\mathcal{X}_k \in \mathcal{B}(\mathcal{X})$.

A.2 Proof of Lemma 3.5

Consider those base members $\mathcal{A} \in \mathcal{D}$ of the hit-or-miss topology $\mathcal{T}$ (cf. Definition 3.1) that have the following form:

- Let $n = 1$, thus we have only one open set $O_1 = O$.
- Let $K = \emptyset$, which gives $\mathcal{X}^{\emptyset} = \{X \in \mathcal{X} : X \cap \emptyset = \emptyset\} = \mathcal{X}$.

According to (3.3), every base member $\mathcal{A} \in \mathcal{D}$ fulfilling these requirements can be written as

$$\mathcal{A} = \mathcal{X} \cap \mathcal{X}_O = \mathcal{X}_O, \tag{A.4}$$

for some open set $O \in \mathcal{O}$. By Definition 2.4, $\mathcal{A} \in \mathcal{T}$. Furthermore, by Definitions 2.5 and 2.2, $\mathcal{A} \in \mathcal{B}(\mathcal{X})$. We take the complement

$$(\mathcal{A})^c = (\mathcal{X}_O)^c = \mathcal{X}^O. \tag{A.5}$$

Because of Definition 2.1, $(\mathcal{A})^c \in \mathcal{B}(\mathcal{X})$ and thus

$$\mathcal{X}^O \in \mathcal{B}(\mathcal{X}). \tag{A.6}$$

Given any closed set $A \subseteq \mathbb{R}^N$, we put $O = A^c$, which yields

$$\mathcal{X}^O = \mathcal{X}^{A^c} = \{X \in \mathcal{X} : X \cap A^c = \emptyset\} = \{X \in \mathcal{X} : X \subseteq A\} = \mathcal{A}(A). \tag{A.7}$$

Hence, it follows from (A.6) that $\mathcal{A}(A) \in \mathcal{B}(\mathcal{X})$ for all closed subsets A of $\mathbb{R}^N$.

A.3 Proof of Lemma 3.7

According to [42, Proposition 3.2.1], χ_k is measurable iff

$$\chi_k^{-1}(\mathcal{A}) \in \mathcal{B}(\mathbb{X}_k), \tag{A.8}$$

for every $\mathcal{A} \in \mathcal{T}_k$, where $\mathcal{T}_k$ is the hit-or-miss topology $\mathcal{T}$ on $\mathcal{X}$ from Definition 3.1 restricted to $\mathcal{X}_k \subseteq \mathcal{X}$. That is, $\mathcal{T}_k$ is generated by the base [53, Theorem 6.3]

$$\mathcal{D}_k = \left\{ (\mathcal{X}_k)^K \cap (\mathcal{X}_k)_{O_1} \cap \cdots \cap (\mathcal{X}_k)_{O_n} : K \in \mathcal{K}, O_i \in \mathcal{O}, n \geq 1 \right\}. \tag{A.9}$$

We will show that for every $\mathcal{A} \in \mathcal{T}_k$,

$$\chi_k^{-1}(\mathcal{A}) \in \mathcal{T}_{\mathbb{X}_k}, \tag{A.10}$$

where $\mathcal{T}_{\mathbb{X}_k}$ is the standard topology $\mathcal{T}_{\mathbb{R}^{kN}}$ on $\mathbb{R}^{kN}$ restricted to $\mathbb{X}_k$, given by [53, Definition 6.1]

$$\mathcal{T}_{\mathbb{X}_k} = \{ O \cap \mathbb{X}_k : O \in \mathcal{T}_{\mathbb{R}^{kN}} \}. \tag{A.11}$$

Since $\mathcal{B}(\mathbb{X}_k) = \sigma(\mathcal{T}_{\mathbb{X}_k})$ due to [42, Theorem 1.8.1], $\mathcal{T}_{\mathbb{X}_k} \subseteq \mathcal{B}(\mathbb{X}_k)$. Consequently, (A.10) implies (A.8).

Because $\mathcal{T}_k$ is generated by $\mathcal{D}_k$, every element $\mathcal{A} \in \mathcal{T}_k$ can be written as the union of base members

$$\mathcal{A} = \bigcup_{i \in I} \mathcal{E}_i, \tag{A.12}$$

where $\mathcal{E}_i \in \mathcal{D}_k$, with some index set I. Using (2.17) together with property (b) from Definition 2.3, it follows that $\chi_k^{-1}(\mathcal{A}) \in \mathcal{T}_{\mathbb{X}_k}$ if

$$\chi_k^{-1}(\mathcal{E}) \in \mathcal{T}_{\mathbb{X}_k}, \tag{A.13}$$

for every $\mathcal{E} \in \mathcal{D}_k$. Since every base member $\mathcal{E} \in \mathcal{D}_k$ itself can be written as an intersection (cf. (A.9))

$$\mathcal{E} = (\mathcal{X}_k)^K \cap (\mathcal{X}_k)_{O_1} \cap \cdots \cap (\mathcal{X}_k)_{O_n}, \tag{A.14}$$

using (2.18) together with property (c) from Definition 2.3, it follows that $\chi_k^{-1}(\mathcal{E}) \in \mathcal{T}_{\mathbb{X}_k}$ if

$$\chi_k^{-1}((\mathcal{X}_k)^K) \in \mathcal{T}_{\mathbb{X}_k}, \tag{A.15}$$

for any compact set $K \in \mathcal{K}$, and

$$\chi_k^{-1}((\mathcal{X}_k)_O) \in \mathcal{T}_{\mathbb{X}_k}, \tag{A.16}$$

for any open set $O \in \mathcal{O}$.

To show the former, note that

$$\chi_k^{-1}\big((\mathcal{X}_k)^K\big) = \chi_k^{-1}\big(\{X \in \mathcal{X}_k : X \cap K = \emptyset\}\big) \tag{A.17}$$

$$= \chi_k^{-1}\big(\{\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\} \in \mathcal{X}_k : \boldsymbol{x}_1 \notin K \wedge \cdots \wedge \boldsymbol{x}_k \notin K\}\big) \tag{A.18}$$

$$= \big\{(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in \mathbb{X}_k : \boldsymbol{x}_1 \notin K \wedge \cdots \wedge \boldsymbol{x}_k \notin K\big\} \tag{A.19}$$

$$= (\underbrace{K^c \times \cdots \times K^c}_{k \text{ times}}) \cap \mathbb{X}_k. \tag{A.20}$$

Since K^c is open, $K^c \in \mathcal{T}_{\mathbb{R}^N}$. According to [53, p. 53], $\mathcal{T}_{\mathbb{R}^{kN}}$ is the product topology $\mathcal{T}_{\mathbb{R}^N} \times \cdots \times \mathcal{T}_{\mathbb{R}^N}$. Therefore, $(K^c \times \cdots \times K^c) \in \mathcal{T}_{\mathbb{R}^{kN}}$. Because of (A.11)

$$(\underbrace{K^c \times \cdots \times K^c}_{k \text{ times}}) \cap \mathbb{X}_k \in \mathcal{T}_{\mathbb{X}_k}, \tag{A.21}$$

and therefore $\chi_k^{-1}\big((\mathcal{X}_k)^K\big) \in \mathcal{T}_{\mathbb{X}_k}$.

For the latter, we have

$$\chi_k^{-1}\big((\mathcal{X}_k)_O\big) = \chi_k^{-1}\big(\{X \in \mathcal{X}_k : X \cap O \neq \emptyset\}\big) \tag{A.22}$$

$$= \chi_k^{-1}\big(\{\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\} \in \mathcal{X}_k : \boldsymbol{x}_1 \in O \vee \cdots \vee \boldsymbol{x}_k \in O\}\big) \tag{A.23}$$

$$= \big\{(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in \mathbb{X}_k : \boldsymbol{x}_1 \in O \vee \cdots \vee \boldsymbol{x}_k \in O\big\} \tag{A.24}$$

$$= \big((O \times \mathbb{R}^{(k-1)N}) \cup (\mathbb{R}^N \times O \times \mathbb{R}^{(k-2)N}) \cup \cdots$$

$$\cdots \cup (\mathbb{R}^{(k-1)N} \times O)\big) \cap \mathbb{X}_k. \tag{A.25}$$

Since $O \in \mathcal{T}_{\mathbb{R}^N}$ and $\mathbb{R}^N \in \mathcal{T}_{\mathbb{R}^N}$, all the cartesian products above are contained in the product topology $\mathcal{T}_{\mathbb{R}^{kN}}$. Hence, their union is an element of $\mathcal{T}_{\mathbb{R}^{kN}}$. Because of (A.11) the intersection with $\mathbb{X}_k$ is included in $\mathcal{T}_{\mathbb{X}_k}$. Thus, $\chi_k^{-1}\big((\mathcal{X}_k)_O\big) \in \mathcal{T}_{\mathbb{X}_k}$. This concludes the proof that χ_k is measurable.

A.4 Proof of Lemma 3.9

We need to check the properties in Definition 2.7. Obviously, $\mu_{\mathcal{X}}$ is nonnegative, since all $\mu_{\mathcal{X}_k}$ in (3.43) are nonnegative. Moreover, $\mu_{\mathcal{X}_0}(\emptyset) = \mathbf{1}_\emptyset(\emptyset) = 0$ since $\emptyset \notin \emptyset$. Furthermore, for $k \in \mathbb{N}$, $\mu_{\mathcal{X}_k}(\emptyset) = \lambda_{kN}(\chi_k^{-1}(\emptyset)) = \lambda_{kN}(\emptyset) = 0$. Thus, $\mu_{\mathcal{X}}(\emptyset) = 0$. Finally, suppose $\{\mathcal{A}^{(i)}\} \subseteq \mathcal{B}(\mathcal{X})$, $i \in \mathbb{N}$, is a countable collection of pairwise disjoint events. Then, using (3.43),

$$\mu_{\mathcal{X}}\left(\bigcup_{i=1}^{\infty} \mathcal{A}^{(i)}\right) = \sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}\left(\left(\bigcup_{i=1}^{\infty} \mathcal{A}^{(i)}\right) \cap \mathcal{X}_k\right) = \sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}\left(\bigcup_{i=1}^{\infty} (\mathcal{A}^{(i)} \cap \mathcal{X}_k)\right) \tag{A.26}$$

$$= \sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}\left(\bigcup_{i=1}^{\infty} \mathcal{A}_k^{(i)}\right), \tag{A.27}$$

where $\mathcal{A}_k^{(i)} = \mathcal{A}^{(i)} \cap \mathcal{X}_k$. Because the $\mu_{\mathcal{X}_k}$ are measures and $\mathcal{A}_k^{(i)} \cap \mathcal{A}_k^{(j)} = \emptyset$ if $i \neq j$, using property (b) in Definition 2.7, we get for the last expression:

$$\sum_{k=0}^{\infty} \mu_{\mathcal{X}_k}\left(\bigcup_{i=1}^{\infty} \mathcal{A}_k^{(i)}\right) = \sum_{k=0}^{\infty}\sum_{i=1}^{\infty} \mu_{\mathcal{X}_k}(\mathcal{A}_k^{(i)}) = \sum_{i=1}^{\infty} \mu_{\mathcal{X}}(\mathcal{A}^{(i)}), \tag{A.28}$$

where the last equality follows since $\mu_{\mathcal{X}_k}(\mathcal{A}_k^{(i)}) \geq 0$ and therefore the summation order can be changed.

Hence, we have shown that

$$\mu_{\mathcal{X}}\left(\bigcup_{i=1}^{\infty} \mathcal{A}^{(i)}\right) = \sum_{i=1}^{\infty} \mu_{\mathcal{X}}(\mathcal{A}^{(i)}). \tag{A.29}$$

This concludes the proof.

A.5 Proof of Theorem 3.15

Suppose X and Y are statistically independent, i.e. (see Definition 3.12),

$$P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) = P_{\mathsf{X}}(\mathcal{A})P_{\mathsf{Y}}(\mathcal{E}), \tag{A.30}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{X})$ and $\mathcal{E} \in \mathcal{B}(\mathcal{Y})$. It follows that (cf. (3.59))

$$\begin{aligned} P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) &= \int_{\mathcal{A}} f_{\mathsf{X}}(X)\mathrm{d}\mu_{\mathcal{X}}(X) \int_{\mathcal{E}} f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{Y}}(Y) &\text{(A.31)}\\ &= \int_{\mathcal{A}}\int_{\mathcal{E}} f_{\mathsf{X}}(X)f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{X}}(X)\mathrm{d}\mu_{\mathcal{Y}}(Y), &\text{(A.32)}\end{aligned}$$

where Fubini's theorem (Theorem 2.19) has been used. Because the product measure is unique, by comparing with (3.135), we see that

$$f_{\mathsf{X},\mathsf{Y}}(X,Y) = f_{\mathsf{X}}(X)f_{\mathsf{Y}}(Y) \qquad \text{a.e.} \tag{A.33}$$

Conversely, assume that

$$f_{\mathsf{X},\mathsf{Y}}(X,Y) = f_{\mathsf{X}}(X)f_{\mathsf{Y}}(Y) \qquad \text{a.e.} \tag{A.34}$$

Inserting in (3.135), we obtain

$$\begin{aligned} P_{\mathsf{X},\mathsf{Y}}(\mathcal{A} \times \mathcal{E}) &= \int_{\mathcal{A}}\int_{\mathcal{E}} f_{\mathsf{X}}(X)f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{X}}(X)\mathrm{d}\mu_{\mathcal{Y}}(Y) &\text{(A.35)}\\ &= \int_{\mathcal{A}} f_{\mathsf{X}}(X)\mathrm{d}\mu_{\mathcal{X}}(X) \int_{\mathcal{E}} f_{\mathsf{Y}}(Y)\mathrm{d}\mu_{\mathcal{Y}}(Y) &\text{(A.36)}\\ &= P_{\mathsf{X}}(\mathcal{A})P_{\mathsf{Y}}(\mathcal{E}). &\text{(A.37)}\end{aligned}$$

This concludes the proof.

A.6 Proof of Theorem 4.6

To show that ν_{X} is a measure on $\mathcal{B}(\mathbb{R}^N)$ we need to check properties (a) and (b) in Definition 2.7.

(a) Since $|X \cap \emptyset| = |\emptyset| = 0$, we obtain upon inserting $S = \emptyset$ in (4.63)

$$\nu_{\mathsf{X}}(\emptyset) = \int_{\mathbb{R}^N} |X \cap \emptyset| f_{\mathsf{X}}(X)\delta X = \int_{\mathbb{R}^N} 0 f_{\mathsf{X}}(X)\delta X = 0. \tag{A.38}$$

(b) Let $S_i \in \mathcal{B}(\mathbb{R}^N)$, $i \in \mathbb{N}$, be a sequence of pairwise disjoint sets. Then

$$\nu_{\mathsf{X}}\left(\bigcup_{i=1}^{\infty} S_i\right) = \int_{\mathbb{R}^N} \left| X \cap \bigcup_{i=1}^{\infty} S_i \right| f_{\mathsf{X}}(X)\delta X = \int_{\mathbb{R}^N} \left| \bigcup_{i=1}^{\infty} (X \cap S_i) \right| f_{\mathsf{X}}(X)\delta X. \tag{A.39}$$

Since $S_i \cap S_j = \emptyset$ if $i \neq j$, it follows that $(X \cap S_i) \cap (X \cap S_j) = \emptyset$ if $i \neq j$. That is, the sequence of finite sets $X \cap S_i$, $i \in \mathbb{N}$, is pairwise disjoint. Because for disjoint finite sets A and B, we have $|A \cup B| = |A| + |B|$, it follows that $|\bigcup_{i=1}^{\infty}(X \cap S_i)| = \sum_{i=1}^{\infty} |X \cap S_i|$. Inserting this in (A.39) we obtain

$$\nu_{\mathsf{X}}\left(\bigcup_{i=1}^{\infty} S_i\right) = \int_{\mathbb{R}^N} \sum_{i=1}^{\infty} |X \cap S_i| \, f_{\mathsf{X}}(X)\delta X = \sum_{i=1}^{\infty} \int_{\mathbb{R}^N} |X \cap S_i| \, f_{\mathsf{X}}(X)\delta X = \sum_{i=1}^{\infty} \nu_{\mathsf{X}}(S_i), \tag{A.40}$$

where the linearity of the set integral (4.5) has been used. This concludes the proof that ν_{X} is a measure on $\mathcal{B}(\mathbb{R}^N)$.

To prove $\nu_{\mathsf{X}} \ll \lambda_N$ we need to show that $\lambda_N(S) = 0$ implies $\nu_{\mathsf{X}}(S) = 0$. Suppose $\lambda_N(S) = 0$ and note that

$$|X \cap S| = \begin{cases} 0 & \text{if } X = \emptyset, \\ \sum_{\boldsymbol{x} \in X} \mathbf{1}_S(\boldsymbol{x}) & \text{if } X \neq \emptyset, \end{cases} \tag{A.41}$$

where $\mathbf{1}_S$ is the indicator function of S. The EVFs corresponding to $|X \cap S|$ are therefore

$$|\chi_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \cap S| = \sum_{i=1}^{k} \mathbf{1}_S(\boldsymbol{x}_i), \qquad k \in \mathbb{N}_0 \tag{A.42}$$

where by convention $\sum_{i=1}^{0} \mathbf{1}_S(\boldsymbol{x}_i) = 0$. Using (4.3) we expand the set integral in (4.63) and insert the EVFs for $|X \cap S|$ to obtain

$$\nu_{\mathsf{X}}(S) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{\mathbb{R}^{kN}} \left(\sum_{i=1}^{k} \mathbf{1}_S(\boldsymbol{x}_i) \right) \tilde{f}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \tag{A.43}$$

$$= \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i=1}^{k} \underbrace{\int_{\mathbb{R}^{kN}} \mathbf{1}_S(\boldsymbol{x}_i) \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)}_{=I}. \tag{A.44}$$

Using Fubini's theorem (Theorem 2.19), the integral I can be written as

$$I = \int_{\mathbb{R}^N} \cdots \int_{\mathbb{R}^N} \mathbf{1}_S(\boldsymbol{x}_i) \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_N(\boldsymbol{x}_1) \ldots \lambda_N(\boldsymbol{x}_k) \tag{A.45}$$

$$= \int_{\mathbb{R}^N} \cdots \underbrace{\int_S}_{i\text{-th position}} \cdots \int_{\mathbb{R}^N} \tilde{f}_{\mathsf{X}}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \mathrm{d}\lambda_N(\boldsymbol{x}_1) \ldots \lambda_N(\boldsymbol{x}_k) \tag{A.46}$$

Since we assumed $\lambda_N(S) = 0$, we have $I = 0$ and therefore $\nu_{\mathsf{X}}(S) = 0$. This concludes the proof of $\nu_{\mathsf{X}} \ll \lambda_N$.

A.7 Proof of Proposition 5.1

Inserting the likelihood function (5.3) in (5.7), we obtain the posterior as

$$f(Z|\boldsymbol{s}) = C(\boldsymbol{s}) \left(\prod_{\boldsymbol{z} \in Z} \eta(\boldsymbol{z}, \boldsymbol{s}) \right) f_{\mathsf{Z}}(Z), \tag{A.47}$$

where $C(\boldsymbol{s})$ is a normalization constant given by

$$C(\boldsymbol{s}) = \left[\int_R \left(\prod_{\boldsymbol{z} \in Z} \eta(\boldsymbol{z}, \boldsymbol{s}) \right) f_{\mathsf{Z}}(Z) \delta Z \right]^{-1}. \tag{A.48}$$

Using (5.11), the posterior EVDs are therefore

$$\tilde{f}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k|\boldsymbol{s}) = C(\boldsymbol{s}) \left(\prod_{i=1}^k \eta(\boldsymbol{z}_i, \boldsymbol{s}) \right) e^{-\mu} \mu^k \prod_{i=1}^k \psi(\boldsymbol{z}_i) \tag{A.49}$$

$$= C(\boldsymbol{s}) e^{-\mu} \mu^k \prod_{i=1}^k \eta(\boldsymbol{z}_i, \boldsymbol{s}) \psi(\boldsymbol{z}_i). \tag{A.50}$$

Because the EVDs normalize to $k! p_{|Z|}(k|\boldsymbol{s})$ (cf. (3.69)), we have

$$p_{|Z|}(k|\boldsymbol{s}) = \frac{1}{k!} \int_{R^k} \tilde{f}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k|\boldsymbol{s}) \mathrm{d}\lambda_{kN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \tag{A.51}$$

$$= C(\boldsymbol{s}) \frac{e^{-\mu}\mu^k}{k!} \int_{R^k} \prod_{i=1}^k \eta(\boldsymbol{z}_i, \boldsymbol{s}) \psi(\boldsymbol{z}_i) \mathrm{d}\lambda_{kN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \tag{A.52}$$

$$= C(\boldsymbol{s}) p_{|\mathsf{Z}|}(k) \prod_{i=1}^k \int_R \eta(\boldsymbol{z}_i, \boldsymbol{s}) \psi(\boldsymbol{z}_i) \mathrm{d}\lambda_N(\boldsymbol{z}_i) \tag{A.53}$$

$$= C(\boldsymbol{s}) p_{|\mathsf{Z}|}(k) \alpha(\boldsymbol{s})^k. \tag{A.54}$$

Using $\sum_{k=0}^{\infty} p_{|Z|}(k|\boldsymbol{s}) = 1$, this allows us to calculate the normalization constant $C(\boldsymbol{s})$ as

$$[C(\boldsymbol{s})]^{-1} = \sum_{k=0}^{\infty} p_{|Z|}(k) \alpha(\boldsymbol{s})^k \tag{A.55}$$

$$= e^{-\mu} \underbrace{\sum_{k=0}^{\infty} \frac{[\mu\alpha(\boldsymbol{s})]^k}{k!}}_{=e^{\mu\alpha(\boldsymbol{s})}} \tag{A.56}$$

$$= e^{-\mu(1-\alpha(\boldsymbol{s}))}. \tag{A.57}$$

Multiplying (A.50) with $\frac{\alpha(\boldsymbol{s})^k}{\alpha(\boldsymbol{s})^k}$ and inserting $C(\boldsymbol{s})$ finally yields

$$\tilde{f}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k | \boldsymbol{s}) = \frac{\alpha(\boldsymbol{s})^k}{\alpha(\boldsymbol{s})^k} e^{\mu} e^{-\mu\alpha(\boldsymbol{s})} e^{-\mu} \mu^k \prod_{i=1}^{k} \eta(\boldsymbol{z}_i, \boldsymbol{s}) \psi(\boldsymbol{z}_i) \tag{A.58}$$

$$= e^{-\mu\alpha(\boldsymbol{s})} [\mu\alpha(\boldsymbol{s})]^k \prod_{i=1}^{k} \frac{\eta(\boldsymbol{z}_i, \boldsymbol{s}) \psi(\boldsymbol{z}_i)}{\alpha(\boldsymbol{s})} \tag{A.59}$$

$$= e^{-\nu(\boldsymbol{s})} \nu(\boldsymbol{s})^k \prod_{i=1}^{k} f(\boldsymbol{z}_i | \boldsymbol{s}). \tag{A.60}$$

A.8 Proof of Lemma 5.3

Expanding the set integral as in (4.3), we have

$$\int_{R_1 \cup R_2} f(Z) \delta Z = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{(R_1 \cup R_2)^k} \tilde{f}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \mathrm{d}\lambda_{kN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \tag{A.61}$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{k} \binom{k}{n} \int_{(R_2 \setminus R_1)^{k-n}} \int_{R_1^n} \tilde{f}^{(k)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) \mathrm{d}\lambda_{nN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$$

$$\mathrm{d}\lambda_{(k-n)N}(\boldsymbol{z}_{n+1}, \ldots, \boldsymbol{z}_k) \tag{A.62}$$

Using the bijective mapping $(k, n) \to (j = k - n, n)$ for $n \leq k$, this becomes

$$\int_{R_1 \cup R_2} f(Z) \delta Z = \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \frac{1}{n! j!} \int_{(R_2 \setminus R_1)^j} \int_{R_1^n} \tilde{f}^{(n+j)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n+j}) \mathrm{d}\lambda_{nN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$$

$$\mathrm{d}\lambda_{jN}(\boldsymbol{z}_{n+1}, \ldots, \boldsymbol{z}_{n+j}) \tag{A.63}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{R_1^n} \left[\sum_{j=0}^{\infty} \frac{1}{j!} \int_{(R_2 \setminus R_1)^j} \tilde{f}^{(n+j)}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n, \boldsymbol{z}_{n+1}, \ldots, \boldsymbol{z}_{n+j}) \right. \tag{A.64}$$

$$\left. \mathrm{d}\lambda_{jN}(\boldsymbol{z}_{n+1}, \ldots, \boldsymbol{z}_{n+j}) \right] \mathrm{d}\lambda_{nN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) \tag{A.65}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{R_1^n} \left[\int_{R_2 \setminus R_1} f(\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\} \cup Y) \delta Y \right] \mathrm{d}\lambda_{nN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) \tag{A.66}$$

$$= \int_{R_2 \setminus R_1} \left[\sum_{n=0}^{\infty} \frac{1}{n!} \int_{R_1^n} f(\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\} \cup Y) \mathrm{d}\lambda_{nN}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) \right] \delta Y \tag{A.67}$$

$$= \int_{R_2 \setminus R_1} \int_{R_1} f(X \cup Y) \delta X \delta Y. \tag{A.68}$$

A.9 Proof of Proposition 5.5

Consider the event

$$\mathcal{X}(A_1) \times \mathcal{Y}(A_2) = \{(X, Y) \in \mathcal{X}(R_1) \times \mathcal{Y}(R_2) : X \subseteq A_1 \wedge Y \subseteq A_2\}, \tag{A.69}$$

with $A_1 \subseteq R_1$ and $A_2 \subseteq R_2$. This event is equivalent to the event

$$\mathcal{A}' = \{Z \in \mathcal{Z}(R) : Z \cap R_1 \subseteq A_1 \wedge Z \cap R_2 \subseteq A_2\}. \tag{A.70}$$

Note that $Z = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\} \in \mathcal{A}'$ iff

$$(\boldsymbol{z}_i \in A_1 \vee \boldsymbol{z}_i \in R_{2\setminus 1}) \wedge (\boldsymbol{z}_i \in A_2 \vee \boldsymbol{z}_i \in R_{1\setminus 2}), \qquad i = 1, \ldots, n. \tag{A.71}$$

Expanding this expression yields the equivalent condition

$$\begin{aligned}(\boldsymbol{z}_i \in A_1 \wedge \boldsymbol{z}_i \in A_2) \vee (\boldsymbol{z}_i \in R_{2\setminus 1} \wedge \boldsymbol{z}_i \in A_2)&\\ \vee (\boldsymbol{z}_i \in A_1 \wedge \boldsymbol{z}_i \in R_{1\setminus 2}) \vee \underbrace{(\boldsymbol{z}_i \in R_{2\setminus 1} \wedge \boldsymbol{z}_i \in R_{1\setminus 2})}_{\equiv\text{false}}&,\end{aligned} \tag{A.72}$$

which is equivalent to

$$Z \subseteq (A_1 \cap A_2) \cup (A_2 \cap R_{2\setminus 1}) \cup (A_1 \cap R_{1\setminus 2}). \tag{A.73}$$

Thus,

$$\mathcal{A}' = \mathcal{Z}\big((A_1 \cap A_2) \cup (A_2 \cap R_{2\setminus 1}) \cup (A_1 \cap R_{1\setminus 2})\big). \tag{A.74}$$

Continuing, we have

$$\Pr(\mathsf{X} \subseteq A_1 \wedge \mathsf{Y} \subseteq A_2) = \int_{A_2} \int_{A_1} f(X, Y) \delta X \delta Y \tag{A.75}$$

$$= \int_{(A_1 \cap A_2) \cup (A_2 \cap R_{2\setminus 1}) \cup (A_1 \cap R_{1\setminus 2})} f_{\mathsf{Z}}(Z) \delta Z. \tag{A.76}$$

Using Lemma 5.3 the last expression becomes

$$\Pr(\mathsf{X} \subseteq A_1 \wedge \mathsf{Y} \subseteq A_2) = \int_{(A_2 \cap R_{2\setminus 1}) \cup (A_1 \cap R_{1\setminus 2})} \int_{A_1 \cap A_2} f_{\mathsf{Z}}(X_{12} \cup W) \delta X_{12} \delta W \tag{A.77}$$

$$= \int_{A_1 \cap R_{1\setminus 2}} \int_{A_2 \cap R_{2\setminus 1}} \left(\int_{A_1 \cap A_2} f_{\mathsf{Z}}(X_{12} \cup X_1 \cup Y_2) \delta X_{12} \right) \delta Y_2 \delta X_1. \tag{A.78}$$

Since

$$A_1 = (A_1 \cap R_{1\setminus 2}) \cup (A_1 \cap R_2), \tag{A.79}$$

$$A_2 = (A_2 \cap R_{2\setminus 1}) \cup (A_2 \cap R_1), \tag{A.80}$$

we can use Lemma 5.3 to write (A.75) as

$$\Pr(\mathsf{X} \subseteq A_1 \wedge \mathsf{Y} \subseteq A_2) = \int_{A_2 \cap R_1} \int_{A_2 \cap R_{2\setminus 1}} \int_{A_1} f(X, Y_2 \cup Y_{12}) \delta X \delta Y_2 \delta Y_{12} \tag{A.81}$$

$$\begin{aligned}&= \int_{A_2 \cap R_1} \int_{A_2 \cap R_{2\setminus 1}} \int_{A_1 \cap R_2} \int_{A_1 \cap R_{1\setminus 2}} f(X_1 \cup X_{12}, Y_2 \cup Y_{12})\\ &\qquad \delta X_1 \delta X_{12} \delta Y_2 \delta Y_{12}\end{aligned} \tag{A.82}$$

$$\begin{aligned}&= \int_{A_1 \cap R_{1\setminus 2}} \int_{A_2 \cap R_{2\setminus 1}} \Big(\int_{A_2 \cap R_1} \int_{A_1 \cap R_2} f(X_1 \cup X_{12}, Y_2 \cup Y_{12})\\ &\qquad \delta X_{12} \delta Y_{12} \Big) \delta Y_2 \delta X_1.\end{aligned} \tag{A.83}$$

Comparing with (A.78) we obtain

$$\int_{A_1 \cap A_2} f_{\mathsf{Z}}(X_{12} \cup X_1 \cup Y_2) \delta X_{12} = \underbrace{\int_{A_2 \cap R_1} \int_{A_1 \cap R_2} f(X_1 \cup X_{12}, Y_2 \cup Y_{12}) \delta X_{12} \delta Y_{12}}_{=K}. \tag{A.84}$$

We make the following ansatz for the joint pdf

$$f(X, Y) = f^*(X, Y) \delta_{X_{12}}(Y_{12}), \tag{A.85}$$

where $\delta_{X_{12}}(Y_{12})$ is the dirac delta function on the set space $\mathcal{Z}(R)$ as introduced in (5.39) with $X_{12} = X \cap R_{12}$ and $Y_{12} = Y \cap R_{12}$. Since $\mathsf{X} = \mathsf{Z} \cap R_1$ and $\mathsf{Y} = \mathsf{Z} \cap R_2$, the elements of X contained in the region R_{12} must be exactly the same as the elements of Y contained in R_{12}. The delta function in (A.85) makes sure that all pairs (X, Y) that do not fulfill this requirement are assigned zero probability. Inserting (A.85) in the right-hand side of (A.84), we obtain

$$K = \int_{A_1 \cap R_2} \underbrace{\int_{A_2 \cap R_1} f^*(X_1 \cup X_{12}, Y_2 \cup Y_{12}) \delta_{X_{12}}(Y_{12}) \delta Y_{12}}_{=L} \delta X_{12} \tag{A.86}$$

Using the sifting property of the Dirac delta (5.39), the inner integral evaluates to

$$L = \begin{cases} f^*(X_1 \cup X_{12}, Y_2 \cup X_{12}) & \text{if } X_{12} \subseteq A_2 \cap R_1, \\ 0 & \text{otherwise.} \end{cases} \tag{A.87}$$

Hence the integration area of the outer integral reduces to $\mathcal{Z}(A_1 \cap A_2)$, yielding

$$K = \int_{A_1 \cap A_2} f^*(X_1 \cup X_{12}, Y_2 \cup X_{12}) \delta X_{12}. \tag{A.88}$$

Consulting (A.84), we see that

$$f^*(X_1 \cup X_{12}, Y_2 \cup X_{12}) = f_{\mathsf{Z}}(X_{12} \cup X_1 \cup Y_2), \tag{A.89}$$

and consequently with $X = X_1 \cup X_{12}$ and $Y = Y_2 \cup Y_{12}$ we finally obtain

$$f(X, Y) = f_{\mathsf{Z}}(X \cup Y) \delta_{X_{12}}(Y_{12}). \tag{A.90}$$

A.10 Proof of Proposition 5.6

Using Lemma 5.3 and Proposition 5.5 we develop the posterior (5.33) as follows:

$$\begin{aligned}
f(X|\boldsymbol{s}_1,\boldsymbol{s}_2) &= C\int_{R_2} g_1(\boldsymbol{s}_1|X)g_2(\boldsymbol{s}_2|Y)f(X,Y)\delta Y && \text{(A.91)}\\
&= Cg_1(\boldsymbol{s}_1|X)\int_{R_{2\setminus 1}}\int_{R_{12}} g_2(\boldsymbol{s}_2|Y_2\cup Y_{12})f(X,Y_2\cup Y_{12})\delta Y_{12}\delta Y_2 && \text{(A.92)}\\
&= Cg_1(\boldsymbol{s}_1|X)\int_{R_{2\setminus 1}}\int_{R_{12}} g_2(\boldsymbol{s}_2|Y_2\cup Y_{12})f_{\mathsf{Z}}(X\cup Y_2\cup Y_{12})\delta_{X_{12}}(Y_{12})\delta Y_{12}\delta Y_2 && \text{(A.93)}\\
&= Cg_1(\boldsymbol{s}_1|X)\int_{R_{2\setminus 1}} g_2(\boldsymbol{s}_2|Y_2\cup X_{12})f_{\mathsf{Z}}(X\cup Y_2\cup X_{12})\delta Y_2 && \text{(A.94)}\\
&= Cg_1(\boldsymbol{s}_1|X)\int_{R_{2\setminus 1}} g_2(\boldsymbol{s}_2|Y_2\cup X_{12})f_{\mathsf{Z}}(X\cup Y_2)\delta Y_2, && \text{(A.95)}
\end{aligned}$$

where $X = X_1 \cup X_{12}$, $X_{12} = X\cap R_{12}$ and $X_1 = X\cap R_{1\setminus 2}$ (and similarly for Y). Inserting the local likelihood function g_2 from (5.24) we obtain

$$\begin{aligned}
f(X|\boldsymbol{s}_1,\boldsymbol{s}_2) &= Cg_1(\boldsymbol{s}_1|X)\int_{R_{2\setminus 1}} \pi_2(\boldsymbol{s}_2)\prod_{\boldsymbol{y}\in Y_2}\eta_2(\boldsymbol{y},\boldsymbol{s}_2)\prod_{\boldsymbol{x}\in X_{12}}\eta_2(\boldsymbol{x},\boldsymbol{s}_2)f_{\mathsf{Z}}(X\cup Y_2)\delta Y_2 && \text{(A.96)}\\
&= Cg_1(\boldsymbol{s}_1|X)\pi_2(\boldsymbol{s}_2)\prod_{\boldsymbol{x}\in X_{12}}\eta_2(\boldsymbol{x},\boldsymbol{s}_2)\underbrace{\int_{R_{2\setminus 1}}\prod_{\boldsymbol{y}\in Y_2}\eta_2(\boldsymbol{y},\boldsymbol{s}_2)f_{\mathsf{Z}}(X\cup Y_2)\delta Y_2}_{=K}. && \text{(A.97)}
\end{aligned}$$

Expanding the set integral as in (4.3) and inserting the prior (5.40), the integral term K becomes

$$\begin{aligned}
K &= \sum_{k=0}^{\infty}\frac{1}{k!}\int_{R_{2\setminus 1}^k}\left[\prod_{i=1}^k\eta_2(\boldsymbol{y}_i,\boldsymbol{s}_2)\right]\left[e^{-\mu}\mu^{k+|X|}\prod_{i=1}^k\psi(\boldsymbol{y}_i)\prod_{\boldsymbol{x}\in X}\psi(\boldsymbol{x})\right]\mathrm{d}\lambda_{kN}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_k) && \text{(A.98)}\\
&= e^{-\mu}\mu^{|X|}\prod_{\boldsymbol{x}\in X}\psi(\boldsymbol{x})\sum_{k=0}^{\infty}\frac{\mu^k}{k!}\int_{R_{2\setminus 1}^k}\prod_{i=1}^k\eta_2(\boldsymbol{y}_i,\boldsymbol{s}_2)\psi(\boldsymbol{y}_i)\mathrm{d}\lambda_{kN}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_k) && \text{(A.99)}\\
&= e^{-\mu}\mu^{|X|}\prod_{\boldsymbol{x}\in X}\psi(\boldsymbol{x})\sum_{k=0}^{\infty}\frac{\mu^k}{k!}\prod_{i=1}^k\underbrace{\int_{R_{2\setminus 1}}\eta_2(\boldsymbol{y},\boldsymbol{s}_2)\psi(\boldsymbol{y})\mathrm{d}\lambda_N(\boldsymbol{y})}_{=\alpha_2(\boldsymbol{s}_2)} && \text{(A.100)}\\
&= e^{-\mu}\mu^{|X|}\prod_{\boldsymbol{x}\in X}\psi(\boldsymbol{x})\underbrace{\sum_{k=0}^{\infty}\frac{[\mu\alpha_2(\boldsymbol{s}_2)]^k}{k!}}_{=e^{\mu\alpha_2(\boldsymbol{s}_2)}} && \text{(A.101)}\\
&= e^{\mu(\alpha_2(\boldsymbol{s}_2)-1)}\mu^{|X|}\prod_{\boldsymbol{x}\in X}\psi(\boldsymbol{x}). && \text{(A.102)}
\end{aligned}$$

Inserting this and the likelihood function $g_1(\boldsymbol{s}_1|X)$ from (5.23) in (A.97) we obtain

$$\begin{aligned}
f(X|\boldsymbol{s}_1,\boldsymbol{s}_2) &= C\pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2)\mu^{|X|}e^{\mu(\alpha_2(\boldsymbol{s}_2)-1)}\left[\prod_{\boldsymbol{x}\in X}\eta_1(\boldsymbol{x},\boldsymbol{s}_1)\psi(\boldsymbol{x})\right]\left[\prod_{\boldsymbol{x}\in X_{12}}\eta_2(\boldsymbol{x},\boldsymbol{s}_2)\right] && \text{(A.103)}\\
&= C\pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2)\mu^{|X|}e^{\mu(\alpha_2(\boldsymbol{s}_2)-1)}\prod_{\boldsymbol{x}\in X}\psi(\boldsymbol{x})\eta_1(\boldsymbol{x},\boldsymbol{s}_1)\gamma_2(\boldsymbol{x},\boldsymbol{s}_2), && \text{(A.104)}
\end{aligned}$$

with $\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2)$ as defined in (5.44).

Because the EVDs normalize to $k! p_{|\mathsf{X}|}(k|\boldsymbol{s}_1, \boldsymbol{s}_2)$ (cf. (3.69)), we have

$$p_{|\mathsf{X}|}(k|\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{1}{k!} \int_{R_1^k} \tilde{f}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k|\boldsymbol{s}_1, \boldsymbol{s}_2) \mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \tag{A.105}$$

$$= C\pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2) e^{\mu\alpha_2(\boldsymbol{s}_2)} \underbrace{\frac{e^{-\mu}\mu^k}{k!}}_{=p_{|\mathsf{Z}|}(k)} \int_{R_1^k} \prod_{i=1}^{k} \psi(\boldsymbol{x}_i)\eta_1(\boldsymbol{x}_i, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}_i, \boldsymbol{s}_2)$$

$$\mathrm{d}\lambda_{kN}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \tag{A.106}$$

$$= C\pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2) e^{\mu\alpha_2(\boldsymbol{s}_2)} p_{|\mathsf{Z}|}(k) \prod_{i=1}^{k} \underbrace{\int_{R_1} \psi(\boldsymbol{x})\eta_1(\boldsymbol{x}, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}, \boldsymbol{s}_2) \mathrm{d}\lambda_N(\boldsymbol{x})}_{=\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)} \tag{A.107}$$

$$= C\pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2) e^{\mu\alpha_2(\boldsymbol{s}_2)} p_{|\mathsf{Z}|}(k) \varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)^k \tag{A.108}$$

Using $\sum_{k=0}^{\infty} p_{|\mathsf{X}|}(k|\boldsymbol{s}_1, \boldsymbol{s}_2) = 1$, this allows us to calculate the normalization constant C as

$$C^{-1} = \pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2) e^{\mu\alpha_2(\boldsymbol{s}_2)} \sum_{k=0}^{\infty} p_{|\mathsf{Z}|}(k) \varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)^k \tag{A.109}$$

$$= \pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2) e^{\mu\alpha_2(\boldsymbol{s}_2)} e^{-\mu} \underbrace{\sum_{k=0}^{\infty} \frac{[\mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)]^k}{k!}}_{e^{\mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)}} \tag{A.110}$$

$$= \pi_1(\boldsymbol{s}_1)\pi_2(\boldsymbol{s}_2) e^{\mu(\alpha_2(\boldsymbol{s}_2) + \varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2) - 1)}. \tag{A.111}$$

Multiplying (A.104) with $\frac{\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)^k}{\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)^k}$ and inserting C finally yields

$$\tilde{f}^{(k)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k|\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)^k}{\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)^k} e^{-\mu\alpha_2(\boldsymbol{s}_2)} e^{-\mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)} e^{\mu} \mu^k e^{\mu\alpha_2(\boldsymbol{s}_2)} e^{-\mu}$$

$$\prod_{i=1}^{k} \psi(\boldsymbol{x}_i)\eta_1(\boldsymbol{x}_i, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}_i, \boldsymbol{s}_2) \tag{A.112}$$

$$= e^{-\mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)} [\mu\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)]^k \prod_{i=1}^{k} \frac{\psi(\boldsymbol{x}_i)\eta_1(\boldsymbol{x}_i, \boldsymbol{s}_1)\gamma_2(\boldsymbol{x}_i, \boldsymbol{s}_2)}{\varepsilon_1(\boldsymbol{s}_1, \boldsymbol{s}_2)} \tag{A.113}$$

$$= e^{-\nu}\nu^k \prod_{i=1}^{k} f(\boldsymbol{x}_i|\boldsymbol{s}_1, \boldsymbol{s}_2). \tag{A.114}$$

Bibliography

[1] S. Särkkä, *Bayesian Filtering and Smoothing.* Cambridge University Press, 2013.

[2] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking.* Yaakov Bar-Shalom, 1995.

[3] I. Goodman, R. Mahler, and H. Nguyen, *Mathematics of Data Fusion.* Kluwer Academic Publishers, 1997.

[4] G. Matheron, *Random Sets and Integral Geometry.* John Wiley & Sons, 1975.

[5] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes, Vol. I.* Springer, 2003.

[6] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Eletronic Systems*, vol. 39, pp. 1152–1178, October 2003.

[7] K. Panta, B.-N. Vo, S. Singh, and A. Doucet, "Probability hypothesis density filter versus multiple hypothesis tracking," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (I. Kadar, ed.), vol. 5429, pp. 284–295, August 2004.

[8] D. Svensson, J. Wintenby, and L. Svensson, "Performance evaluation of MHT and GM-CPHD in a ground target tracking scenario," in *Proceedings of the 12th International Conference on Information Fusion*, (Seattle, WA), pp. 300–307, July 2009.

[9] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, pp. 1224–1245, October 2005.

[10] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4091–4104, November 2006.

[11] D. E. Clark and J. Bell, "Convergence results for the particle PHD filter," *IEEE Transactions on Signal Processing*, vol. 54, pp. 2652–2661, July 2006.

[12] A. Johansen, S. Singh, A. Doucet, and B.-N. Vo, "Convergence of the SMC implementation of the PHD filter," *Methodology and Computing in Applied Probability*, vol. 8, no. 2, pp. 265–291, 2006.

[13] D. Clark and B.-N. Vo, "Convergence analysis of the Gaussian mixture PHD filter," *IEEE Transactions on Signal Processing*, vol. 55, pp. 1204–1212, April 2007.

[14] R. Mahler, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, pp. 1523–1543, October 2007.

[15] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, pp. 3553–3567, July 2007.

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

[17] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," *IEEE Transactions on Signal Processing*, vol. 57, pp. 409–423, February 2009.

[18] B. T. Vo, *Random Finite Sets in Multi-Object Filtering.* PhD thesis, The University of Western Australia, 2008.

[19] D. Clark and J. Bell, "Bayesian multiple target tracking in forward scan sonar images using the PHD filter," in *IEE Proceedings Radar, Sonar and Navigation*, vol. 152, pp. 327–334, October 2005.

[20] N. T. Pham, W. Huang, and S. H. Ong, "Tracking multiple objects using probability hypothesis density filter and color measurements," in *2007 IEEE International Conference on Multimedia and Expo*, (Beijing, China), pp. 1511–1514, July 2007.

[21] E. Maggio, M. Taj, and A. Cavallaro, "Efficient multitarget visual tracking using random finite sets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1016–1027, August 2008.

[22] B. N. Vo, B. T. Vo, N. T. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *IEEE Transactions on Signal Processing*, vol. 58, pp. 5129–5141, October 2010.

[23] R. Hoseinnezhad, B.-N. Vo, B.-T. Vo, and D. Suter, "Visual tracking of numerous targets via multi-Bernoulli filtering of image data," *Pattern Recognition*, vol. 45, pp. 3625–3635, October 2012.

[24] F. Zhang, H. Stahle, A. Gaschler, C. Buckl, and A. Knoll, "Single camera visual odometry based on random finite set statistics," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (Vilamoura, Portugal), pp. 559–566, October 2012.

[25] R. Hoseinnezhad, B.-N. Vo, and B.-T. Vo, "Visual tracking in background subtracted image sequences via multi-Bernoulli filtering," *IEEE Transactions on Signal Processing*, vol. 61, pp. 392–397, January 2013.

[26] D. Clark, A.-T. Cemgil, P. Peeling, and S. Godsill, "Multi-object tracking of sinusoidal components in audio with the Gaussian mixture probability hypothesis density filter," in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, NY), pp. 339–342, October 2007.

[27] D. Angelosante, *Applications of Random Set Theory to Some Problems in Wireless Communications.* PhD thesis, Università degli Studi di Cassino, Scuola di Dottorato in Ingegneria, 2008.

[28] J. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, "A random-finite-set approach to Bayesian SLAM," *IEEE Transactions on Robotics*, vol. 27, pp. 268–282, April 2011.

[29] H. Deusch, S. Reuter, and K. Dietmayer, "The labeled multi-Bernoulli SLAM filter," *IEEE Signal Processing Letters*, vol. 22, pp. 1561–1565, October 2015.

[30] M. Fröhle, "Multi-target tracking for UWB channels using PHD filters," Master's thesis, Graz University of Technology, 2011.

[31] M. Zhou, "Urban terrain multiple target tracking using the probability hypothesis," Master's thesis, Arizona State University, 2011.

[32] R. Karrer, "A finite random set method for distributed multi-target tracking in the presence of multi-path propagation," Master's thesis, Vienna University of Technology, 2013.

[33] S. Reuter and K. Dietmayer, "Pedestrian tracking using random finite sets," in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, (Chicago, IL), pp. 1–8, July 2011.

[34] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, pp. 3460–3475, July 2013.

[35] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the Bayes multi-target tracking filter," *IEEE Transactions on Signal Processing*, vol. 62, pp. 6554–6567, December 2014.

[36] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 62, pp. 3246–3260, June 2014.

[37] S. Reuter, *Multi-Object Tracking Using Random Finite Sets.* PhD thesis, Ulm University, 2014.

[38] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and A. Graziano, "Consensus CPHD filter for distributed multitarget tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 508–520, June 2013.

[39] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and B.-N. Vo, "Average Kullback-Leibler divergence for random finite sets," in *2015 18th International Conference on Information Fusion*, (Washington, DC), pp. 1359–1366, July 2015.

[40] B. Wang, W. Yi, S. Li, M. R. Morelande, L. Kong, and X. Yang, "Distributed multi-target tracking via generalized multi-Bernoulli random finite sets," in *2015 18th International Conference on Information Fusion*, (Washington, DC), pp. 253–261, July 2015.

[41] B. Wang, W. Yi, S. Li, L. Kong, and X. Yang, "Distributed fusion with multi-Bernoulli filter based on generalized covariance intersection," in *2015 IEEE Radar Conference (RadarCon)*, (Arlington, VA), pp. 958–962, May 2015.

[42] S. Resnick, *A Probability Path.* Birkhäuser, 1999.

[43] P. Billingsley, *Probability and Measure.* John Wiley & Sons, 1995.

[44] D. Williams, *Probability with Martingales.* Cambridge University Press, 1991.

[45] G. Roussas, *An Introduction to Measure-Theoretic Probability.* Academic Press, 2014.

[46] M. Vihola, *Random Sets for Multitarget Tracking and Data Fusion.* PhD thesis, Tampere University of Technology, 2004.

[47] W. Rudin, *Principles of Mathematical Analysis.* McGraw-Hill, 1976.

[48] R. Mahler, “Random set theory for target tracking and identification,” in *Handbook of Multisensor Data Fusion* (D. L. Hall and J. Llinas, eds.), ch. 14, CRC Press LLC, 2001.

[49] R. Mahler, ““Statistics 101” for multisensor, multitarget data fusion,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, pp. 53–64, January 2004.

[50] R. Mahler, ““Statistics 102” for multisource-multitarget detection and tracking,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, pp. 376–389, June 2013.

[51] M. Capinski and P. E. Kopp, *Measure, Integral and Probability.* Springer, 2013.

[52] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, “A consistent metric for performance evaluation of multiobject filters,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 3447–3457, August 2008.

[53] S. Willard, *General Topology.* Addison-Wesley, 1970.