Representation and estimation of stochastic populations

Jérémie Houssineau

This dissertation is submitted for
the degree of Doctor of Philosophy

Heriot-Watt University
School of Engineering and Physical Sciences

August 2015

Abstract

This work is concerned with the representation and the estimation of populations composed of an uncertain and varying number of individuals which can randomly evolve in time. The existing solutions that address this type of problems make the assumption that all or none of the individuals are distinguishable. In other words, the focus is either on specific individuals or on the population as a whole. These approaches have complimentary advantages and drawbacks and the main objective in this work is to introduce a suitable representation for partially-indistinguishable populations. In order to fulfil this objective, a sufficiently versatile way of quantifying different types of uncertainties has to be studied. It is demonstrated that this can be achieved within a measure-theoretic Bayesian paradigm. The proposed representation of stochastic populations is then used for the introduction of various filtering algorithms from the most general to the most specific. The modelling possibilities and the accuracy of one of these filters are then demonstrated in different situations.

Acknowledgements

Starting at the beginning, I would like to express my gratitude towards Dr Dann Laneuville and Mr Alain Carof who supported my decision to start a Ph.D. thesis after the two years I spent working with them at DCNS, and who made this project possible by securing the associated funding. I am also particularly thankful to Dr Daniel Clark for accepting to be my supervisor and for giving me so much freedom in my research activities.

I am very much indebted to Dr Branko Ristić, Prof. Bá-Ngư Võ, Prof. Bá-Tương Võ, Prof. Pierre Del Moral and Prof. Martin Adams who welcomed me in their respective institutions and who had a considerable influence on my work by guiding and challenging the most fundamental ideas of it. I am equally grateful to the official reviewers of this thesis, Prof. Arnaud Doucet and Dr Simon Malham, for their helpful questions and comments about both the form and the content of the present manuscript. I can hardly overstate how much I owe to Dr Špela Ivekovič for her guidance about proper scientific (and English) writing.

Also, I would like to thank my colleagues and friends from Heriot-Watt University, in particular Sharad Nagappa for welcoming me in Edinburgh with a “Do you have a waterproof jacket? Because umbrellas are useless here”, Yan Pailhas and Nicolas Valeyrie for the countless climbing sessions we had, as well as Murat Üney, Emmanuel “Manu” Delande, Jose Franco, Yves Wiaux and Isabel Schlangen for the cheerfulness (and the occasional cakes) they bring to the lab.

Pascale, thank you for managing to keep me alive by preventing me from working too much over evenings and weekends, and thank you for contributing to this manuscript by carefully proofreading it, twice.

Contents

Introduction

STOCHASTIC populations are of central interest in many areas of engineering and physical sciences. They are defined in this work as collections of multiple individuals or objects which state and behaviour through time can be both uncertain and intrinsically random. Applications of this concept can be found at very different scales, from stars all the way to elementary particles, from macroscopic to microscopic. However, multiple challenges have to be faced when representing these stochastic populations: *a*) suitable mathematical expressions of given collections of individuals have to be introduced, *b*) the integration of all the types of uncertainties generally encountered also needs to be addressed, and *c*) depending on the type of information that is made available, the general structure needs to be simplified to meet a level of description that is, in some sense, both necessary and sufficient. The work starts in Chapter 1 with the study of general representations of uncertainty from which a versatile information fusion operation is deduced. After introducing some specific mathematical concepts and results, the above-listed questions will motivate the introduction of general population representations in the first sections of Chapter 2, while the last sections will be dedicated to the analysis of the different operations between representations that will form the basis of the following chapters.

Unlike standard single-individual estimation solutions, such as the Kalman filter [Kalman, 1960], the estimation of stochastic populations is a multi-faceted problem. Indeed, when multiple individuals are involved, the nature of estimation algorithms needs to evolve in order to take into account uncertainties at different levels, including those in *a*) the number of individuals in the considered stochastic population at all times, *b*) the uncertainty about which piece of information corresponds to which individual, *c*) when and where individuals appear and disappear from the scene, and *d*) the membership of individuals to different populations, whenever objects of multiple types need to be estimated. Most of these aspects have been considered previously, first through *bottom-up* approaches which start from the single-object problem and attempt to generalise toward multi-object configurations [Blackman, 1986], and later through *top-down* approaches which prioritise the population modelling [Mahler, 2007]. A different approach is proposed in Chapter 3, where the general population representations of Chapter 2 are used with the objective of formulating a general multi-object filter allowing for many aspects to be integrated while relying on intuitive modelling assumptions.

Although the solutions of Chapter 3 can be considered optimal when the assumptions they rely on are verified, they can still be challenging to process in practice due to their inherent complexity. To address this potential issue, approximated solutions can be found in different ways. An example of this type of approximation is proposed by Mahler [2003] and consists in propagating the first moment of the full solution only. A different way of simplifying the problem is explored in Chapter 4, where the full solution is marginalised into individual hypotheses. The two approaches are then compared on a range of scenarios.

Motivation

The concepts considered in this work differ significantly from the ones found in existing approaches. In order to justify the introduction of these new concepts, non-negligible improvements at different levels have to be demonstrated. We first present some prior work where the existing framework has been used and we discuss some of the limitations. Solutions and improvements brought by the proposed approach are then briefly discussed as a motivation.

Prior work

One of the most popular approaches is to model the population of interest with a point process [Daley and Vere-Jones, 2003] where each point represents an individual of the population. Point processes do not allow for distinguishing individuals and it is often the mean of the point process that is estimated. It is then natural to try to compute the variance of the point process as in [Delande et al., 2013a,b] and [Delande et al., 2014b]. Figure 0.1 illustrates that the mean of a point process is a function of space in general, and that the variance depicts the uncertainty around this mean.

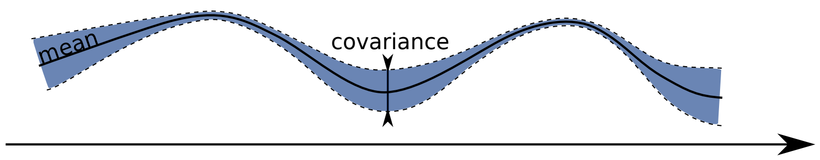


Figure 0.1: Illustration of the concept of variance of a point process

It is also natural to study the covariance associated to a stochastic population which depicts the relation between the number of individuals in two given areas of the space. However, the covariance of a point process can only account for the variation in the number of individuals through different cardinalities. In consequence, having access to a more refined representation of the uncertainty could lead to a more informative expression of the covariance which would account for variations between *configurations*.

It is usual to assume that the individuals in the population of interest are independent. Yet, there is a practical interest in taking into account interactions between individuals as well [Singh et al., 2009, Houssineau and Clark, 2014]. Figure 0.2 shows the variety of possible interactions between individuals of a population when proceeding to a transition between two state spaces X and Y.

However, considering interactions when modelling a population with a point process often leads to computationally demanding algorithms as these interactions will be applied to the whole population rather than *given* individuals. One of the motivations of this work is to find population representations that are sufficiently general to enable interactions to be modelled specifically. For instance, we might be interested in studying the possible interactions between manoeuvring targets only while assuming the others to be behaving independently.

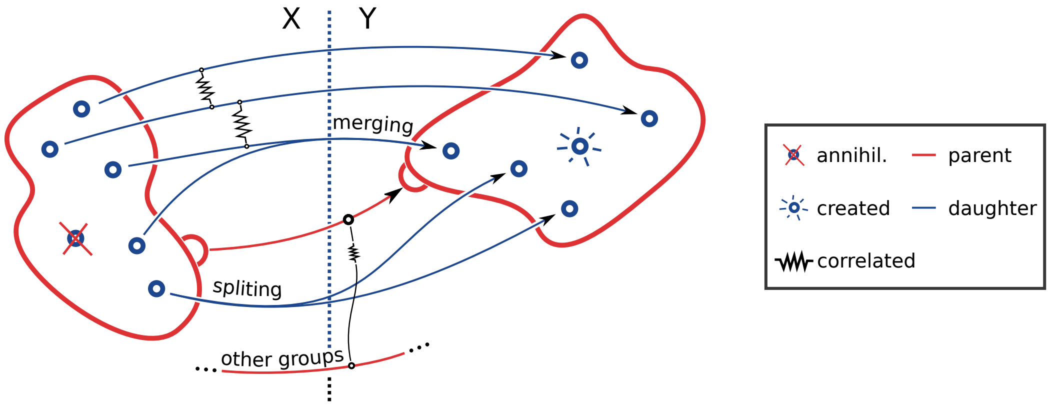
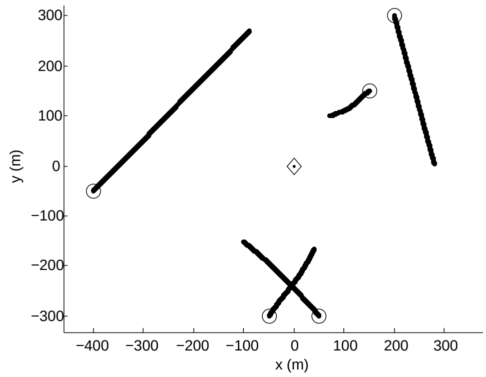


Figure 0.2: Interactions during the transition between the state spaces X and Y

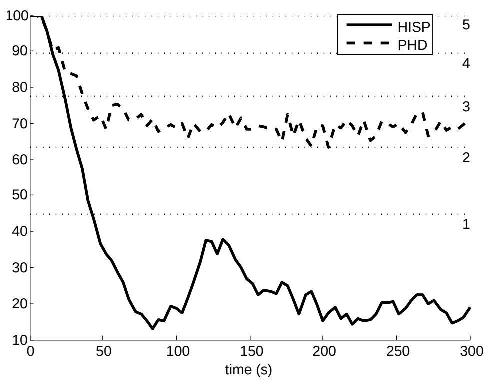
Overall, algorithms based on point processes such as the ones described in [Mahler, 2003, 2007] have been successful at addressing important challenges in multi-object estimation. One of the main the objectives in this work is to study how extensions of the existing concepts could address the remaining limitations.

Accuracy

The approximate solution that will be introduced in Chapter 4 under the name of **h**ypothesised filter for **i**ndependent **s**tochastic **p**opulations or HISP filter, displays a complexity that is linear in the number of hypotheses and in the number of observations. Among existing solutions in multi-object estimation, the most well-known algorithm with a complexity of the same order as the HISP filter is the **p**robability **h**ypothesis **d**ensity filter, or PHD filter [Mahler, 2003]. The results of a comparison between the accuracy of these two approaches are shown in Figure 0.3 using the OSPA distance [Schuhmacher et al., 2008], where OSPA stands for **o**ptimal **s**ub-**p**attern **a**ssignment. The OSPA distance can be understood as the distance between the estimated and the actual population state. In Figure 0.3b, the HISP filter shows a largely improved performance for a probability of detection of 0.5.



(a) Object's trajectories (-) and initial positions (○) as well as sensor position (◇)



(b) Averaged OSPA distance over 50 Monte Carlo runs versus time

Figure 0.3: Scenario (a) and compared performance of the HISP and PHD filters with 50% chance detection and 15 spurious observations per scan . In (b), the dotted line numbered n represents the OSPA distance when n objects are missed

Versatility

Other key advantages of the proposed approach are an increased versatility in population and sensor modelling. Two cases are considered here.

Finite-resolution sensor

The trajectory of an object as seen by a finite-resolution sensor is shown in Figure 0.4. In this case, the only information made available by the sensor is a set of *resolution cells* that are likely to contain an individual of the population. The position and velocity of the individuals in the scene have to be estimated out of this rather weak source of information. We will see in Section 4.3.3 how a sequential Monte Carlo implementation of the HISP filter can address this problem while maintaining the accuracy of the Kalman filter implementation when the non-Gaussianity is less severe.

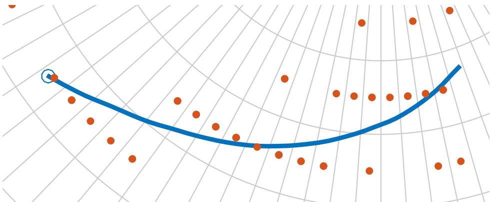
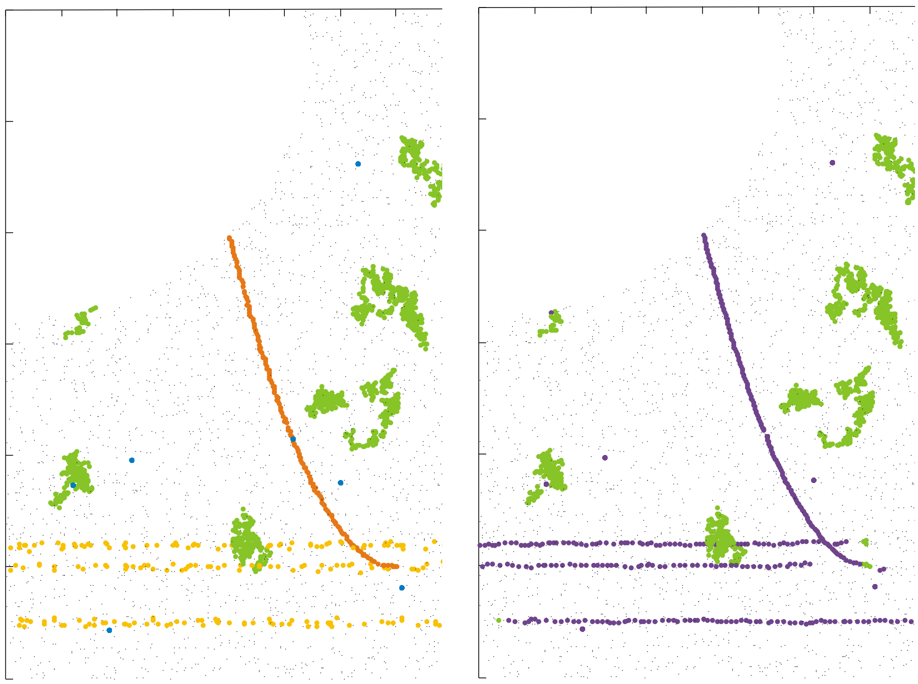


Figure 0.4: Trajectory (-) with initial position (∘) and accumulated observations (•) of an object with the resolution cells of the sensor in the background

Classification

An example of joint population estimation and classification is shown in Figure 0.5, where two sub-populations are distinguished based on their respective dynamical behaviour. This harbour-surveillance scenario is composed of boats going along a shipping lane, of fish moving according to a Brownian motion, and of an **a**utomatic **u**nderwater **v**ehicle (AUV) being sent toward the zone to be monitored. In this case, the HISP filter provides the estimated class of each individual in the scene in addition to their estimated state. This case will be studied in Section 4.3.2.



(a) Ground truth (b) Estimated

Figure 0.5: Accumulated view of the HISP filter's output (b) compared against ground truth (a). Each colour in (a) and (b) represents a different actual/estimated sub-population according to the colour code indicated in the table below

	Observations	Fish	Static targets	Boat	AUV
Ground truth	•	●	●	●	●
Estimated	•	●		●	

Colour codes for the sub-populations considered in Figure 0.5

Chapter 1

Representation of uncertainty

Uncertainty can be considered as being inherent to all types of knowledge about any given physical system and needs to be quantified in order to obtain meaningful characteristics for the considered system. In this chapter, we first present probability theory, since it is usually considered as the most efficient way of representing uncertainty. Then, some of the limitations of the usual concepts of probability theory are detailed. Finally, a novel way of representing different types of information is introduced and studied.

Notations

First of all, some of the notations that will be used throughout the work have to be defined:

a) The set of integers is denoted $\mathbb{Z}$, the sets of non-negative (positive) integers is denoted $\mathbb{N}$ ($\mathbb{N}^*$), and the set of real (non-negative) numbers is denoted $\mathbb{R}$ ($\mathbb{R}^+$)

b) We write $A \doteq B$ to indicate that A is defined as being equal to B

c) The set-builder notation is used with "s.t.", for **s**uch **t**hat, as a separator, e.g.,

$$\{x \text{ s.t. } x \in \mathbb{R},\ \ x = x^2\} = \{0, 1\}, \tag{1.1}$$

which can also be written more compactly as $\{x \in \mathbb{R} \text{ s.t. } x = x^2\} = \{0, 1\}$

d) The power set of a set A is denoted[1] $\wp(A)$

e) Set difference is denoted in two distinct ways:

$$\{x \in B \text{ s.t. } x \notin A\} = \begin{cases} B \setminus A & \text{if } A \nsubseteq B, \\ B - A & \text{if } A \subseteq B, \end{cases} \tag{1.2}$$

this is merely a notational convenience to underline whether all the elements of A are taken away from B or not, if A is a subset of a set $\mathbf{E}$ then the absolute complement of A in $\mathbf{E}$ is simply denoted A^c

f) Bijections are denoted by the arrow $\leftrightarrow$, the graph of a function f with domain $\mathbf{E}$ and codomain $\mathbf{F}$ is denoted $\mathrm{Gr}(f)$ and the image of f is denoted $\mathrm{Im}(f)$

[1] "$\wp$" is referred to as the "Weierstrass p".

g) For any set $\mathbf{E}$, the function on $\mathbf{E}$ which is everywhere equal to one is denoted $\mathbf{1}$ and the function everywhere equal to zero is denoted $\mathbf{0}$

h) For any $f, f' : \mathbf{E} \to \mathbf{F}$ and any $g, g' : \mathbf{E} \to \mathbb{R}$, we define the mappings $g \cdot g'$ and $f \ltimes f'$ on $\mathbf{E}$ as

$$g \cdot g' : x \mapsto g(x)g'(x) \in \mathbb{R} \tag{1.3a}$$
$$f \ltimes f' : x \mapsto (f(x), f'(x)) \in \mathbf{F} \times \mathbf{F}, \tag{1.3b}$$

and the mappings $g \rtimes g'$ and $f \times f'$ on $\mathbf{E} \times \mathbf{E}$ as

$$g \rtimes g' : (x, x') \mapsto g(x)g'(x') \in \mathbb{R}, \tag{1.4a}$$
$$f \times f' : (x, x') \mapsto (f(x), f'(x')) \in \mathbf{F} \times \mathbf{F} \tag{1.4b}$$

These notations will be extensively used in this chapter as well as in the subsequent chapters.

1.1 Probability theory

The fundamental concepts of measure and probability theory are first introduced, and many of the usual notations in these fields are defined at this occasion. Some of the limitations encountered when handling different types of uncertainty with probability theory are then explained and detailed.

1.1.1 Fundamental concepts

If $\mathbf{E}$ is a set and $\mathcal{E}$ is a σ-algebra of subsets of $\mathbf{E}$, then the space $(\mathbf{E}, \mathcal{E})$ is said to be a *measurable space* and forms a suitable basis for the introduction of most of the fundamental concepts of probability theory.

A countably-additive set function $m : \mathcal{E} \to [0, \infty]$ is said to be a *measure* on $(\mathbf{E}, \mathcal{E})$ and the space $(\mathbf{E}, \mathcal{E}, m)$ is referred to as a *measure space*. If it additionally holds that $m(\mathbf{E}) = 1$, then the measure m and the space $(\mathbf{E}, \mathcal{E}, m)$ are respectively called a *probability measure* and a *probability space*. In this case, any subset A in $\mathcal{E}$ is understood as an *event* and $m(A)$ is the probability for this event to happen. The set of measures and the set of probability measures on $\mathbf{E}$ are respectively denoted $\mathbf{M}(\mathbf{E})$ and $\mathbf{M}_1(\mathbf{E})$. We will also be interested in the subset of $\mathbf{M}(\mathbf{E})$ composed of integer-valued measures, or *counting measures*, which is denoted $\mathbf{N}(\mathbf{E})$. A simple example of measure on any given measurable space $(\mathbf{E}, \mathcal{E})$ is the *Dirac measure* at point $x \in \mathbf{E}$, denoted δ_x, which is characterised by

$$(\forall x \in \mathbf{E},\ \ \forall A \in \mathcal{E}) \qquad \delta_x(A) \doteq \mathbf{1}_A(x) \doteq \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise,} \end{cases} \tag{1.5}$$

where $\mathbf{1}_A$ is the indicator function of A.

A measure m on the measurable space $(\mathbf{E}, \mathcal{E})$ might take infinite values on $\mathcal{E}$, which is not always desirable. It is of interest to classify the measure m depending on the amount of “mass” that it gives to subsets of $\mathbf{E}$. The measure m is said to be *a)* *finite* if $m(\mathbf{E}) < \infty$, *b)* *σ-finite* if $\mathbf{E}$ can be seen as the countable union of sets in $\mathcal{E}$ of finite measure, and *c)* *locally-finite* if every point of $\mathbf{E}$ has a neighbourhood of finite measure. A measure m on $(\mathbf{E}, \mathcal{E})$ is said to be *complete* if every subset of a set in $\mathcal{E}$ of measure 0 is measurable.

The concept of *outer measure* is fundamental in measure theory and is defined as follows: an outer measure on $\mathbf{E}$ is a function $\mu : \wp(\mathbf{E}) \to [0, \infty]$ verifying

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

b) (Monotonicity) if $A \subseteq B \subseteq \mathbf{E}$ then $\mu(A) \leq \mu(B)$

c) (Countable sub-additivity) for every sequence $(A_n)_{n\in\mathbb{N}}$ of subsets of $\mathbf{E}$

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leq \sum_{n\in\mathbb{N}} \mu(A_n). \tag{1.6}$$

Outer measures allow for constructing both σ-algebras and measures on them via Carathéodory's method [Fremlin, 2000, Sect. 113], such as the Lebesgue measure on the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}$. If $\mathbf{F}$ is a set, μ and μ' are outer measures on $\mathbf{E}$ and $\mathbf{F}$ respectively and $f : \mathbf{E} \to \mathbf{F}$ is a function, then

$$C \subseteq \mathbf{F} \mapsto \mu(f^{-1}[C]) \qquad \text{and} \qquad A \subseteq \mathbf{E} \mapsto \mu'(f[A]) \tag{1.7}$$

are outer measures.

If $(\mathbf{F}, \mathcal{F})$ is another measurable space, then the *product σ-algebra* of $(\mathbf{E}, \mathcal{E})$ and $(\mathbf{F}, \mathcal{F})$, denoted $\mathcal{E} \otimes \mathcal{F}$, is the σ-algebra generated by the subsets of the form $A \times C$, with $A \in \mathcal{E}$ and $C \in \mathcal{F}$. If m and m' are measures on the spaces $\mathbf{E}$ and $\mathbf{F}$ respectively, then a *product measure*, denoted $m \rtimes m'$, can be defined on $(\mathbf{E} \times \mathbf{F}, \mathcal{E} \otimes \mathcal{F})$ by

$$(\forall A \in \mathcal{E}, \forall C \in \mathcal{F}) \qquad m \rtimes m'(A \times C) = m(A)m'(C). \tag{1.8}$$

If $(\mathbf{X}, \mathcal{T})$ is a topological space, then we can consider the σ-algebra generated by the open subsets in $\mathcal{T}$, called the *Borel σ-algebra* of $\mathbf{X}$ and denoted $\mathcal{B}(\mathbf{X}, \mathcal{T})$ or simply $\mathcal{B}(\mathbf{X})$ when there is no ambiguity on the considered topology. For any measure $m \in \mathbf{M}(\mathbf{X})$, the support of m is understood as being the set of all points $x \in \mathbf{X}$ for which every open neighbourhood U of x has positive measure, i.e., the support of m is

$$\mathrm{supp}(m) \doteq \{x \in \mathbf{X} \text{ s.t. } x \in U \in \mathcal{T} \Rightarrow m(U) > 0\}. \tag{1.9}$$

When $\mathbf{F}$ is the real line $\mathbb{R}$, we also consider the Lebesgue measure $\lambda \in \mathbf{M}(\mathbb{R})$ which can be defined on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ but also on the finer σ-algebra of Lebesgue measurable subsets.

A mapping $f : \mathbf{E} \to \mathbf{F}$ verifying $f^{-1}[A] \in \mathcal{E}$ for all $A \in \mathcal{F}$ is referred to as a *measurable mapping* and we denote $\mathbf{L}^0(\mathbf{E}, \mathbf{F})$ the set of all measurable mappings from $\mathbf{E}$ to $\mathbf{F}$. If $(\mathbf{F}, \mathcal{F}) = (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$, then the set $\mathbf{L}^0(\mathbf{E}, \mathbf{F})$ is simply denoted $\mathbf{L}^0(\mathbf{E})$. If m is a measure on $(\mathbf{E}, \mathcal{E})$ then a measure m' can be induced on $(\mathbf{F}, \mathcal{F})$ through

$$(\forall A \in \mathcal{F}) \qquad m'(A) = m(f^{-1}[A]). \tag{1.10}$$

The measure m' is referred to as the *pushforward measure of* m and is denoted f_*m. This concept is useful to express a change of variable in the context of measure theory [Bogachev, 2007, Vol. I, Sect. 3.6]. The σ-algebra generated by f on $\mathbf{E}$, denoted $\sigma(f)$, is defined as the set of inverse images of all the measurable subsets in $\mathcal{F}$, that is

$$\sigma(f) = \{f^{-1}[A] \text{ s.t. } A \in \mathcal{F}\}. \tag{1.11}$$

The σ-algebra $\sigma(f)$ is said to be a *sub-σ-algebra of $\mathcal{E}$* because it verifies $\sigma(f) \subseteq \mathcal{E}$. Other useful sub-$\sigma$-algebras include the ones generated by countable measurable partitions of $\mathbf{E}$. The following remark is purely technical and is indicated with a * for this reason.

Remark.* The direct image of a Borel subset by a continuous (hence measurable) function is called an *analytic set* and verifies some properties even though it is not a Borel set in general. As related by Dudley [2002, Note 13.2, p. 500], Lebesgue initially thought that the projection of a Borel subset of $\mathbb{R}^2$ onto the real line would be Borel again [Lebesgue, 1905]. Later on, Souslin [1917] found an error in Lebesgue's work and introduced the concept of analytic set. Borel and analytic sets form the basis of *descriptive set theory* [Kechris, 1995].

Another way of transforming measures, called the *Boltzmann-Gibbs transformation*, can be defined as follows: let $f : \mathbf{E} \to [0, \infty)$ be a bounded measurable function and let $\Psi_f : \mathbf{M}(\mathbf{E}) \to \mathbf{M}_1(\mathbf{E})$ be defined as[2]

$$\Psi_f(m)(\mathrm{d}x) \stackrel{\int}{=} \frac{1}{m(f)} f(x) m(\mathrm{d}x), \tag{1.12}$$

where it is assumed that

$$\mathrm{m}(f) \doteq \int f(x) m(\mathrm{d}x) > 0. \tag{1.13}$$

This transformation reduces to Bayes' theorem when f is a likelihood function.

If $(\Omega, \Sigma, \mathbb{P})$ is a probability space then a measurable mapping $X \in \mathbf{L}^0(\Omega, \mathbf{E})$ is said to be an $\mathbf{E}$-valued *random variable*[3]. The pushforward $X_*\mathbb{P}$ is the *law*, or *distribution*, of the random variable X. Two random variables $X \in \mathbf{L}^0(\Omega, \mathbf{E})$ and $X' \in \mathbf{L}^0(\Omega, \mathbf{F})$ can be joined to form a single random variable $X \ltimes X' \in \mathbf{L}^0(\Omega, \mathbf{E} \times \mathbf{F})$ and the law of $X \ltimes X'$ is referred to as the *joint law* of X and X'. The random variables X and X' are said to be *independent* if their joint law can be expressed as a function of their respective laws p and p' as $p \rtimes p' \in \mathbf{M}_1(\mathbf{E} \times \mathbf{F})$.

If $\mathbf{X}$ is a sufficiently nice[4] topological space, then a *point process* on $\mathbf{X}$ can be defined as a random counting measure, that is a random variable in the space $\mathbf{N}(\mathbf{X})$ of integer-valued measures. Another concept that permeates all of probability theory is the concept of stochastic kernel, commonly defined as follows.

Definition 1.1. A stochastic kernel is a map $q : \mathbf{E} \times \mathcal{F} \to \mathbb{R}$ such that

a) the mapping $x \mapsto q(x, B)$ is measurable for every $B \in \mathcal{F}$

b) the mapping $B \mapsto q(x, B)$ is a measure on $(\mathbf{F}, \mathcal{F})$ for every $x \in \mathbf{E}$.

If $q(x, \cdot)$ is in $\mathbf{M}_1(\mathbf{F})$ for every $x \in \mathbf{E}$, then q is said to be a *Markov kernel*.

A stochastic kernel on $\mathbf{E} \times \mathcal{F}$ is said to have $\mathbf{E}$ as a *source space* and $\mathbf{F}$ as a *target space*. The set of stochastic kernels from $(\mathbf{E}, \mathcal{E})$ to $(\mathbf{F}, \mathcal{F})$ is denoted $\mathbf{K}(\mathbf{E}, \mathbf{F})$ while the set of Markov kernels on the same source and target spaces is denoted $\mathbf{K}_1(\mathbf{E}, \mathbf{F})$ and is a subset of $\mathbf{K}(\mathbf{E}, \mathbf{F})$. For any kernel $q \in \mathbf{K}(\mathbf{E}, \mathbf{F})$ and any measurable function $f \in \mathbf{L}^0(\mathbf{F})$, the measurable function $q(f) \in \mathbf{L}^0(\mathbf{E})$ is defined as

$$q(f) : x \mapsto \int f(y) q(x, \mathrm{d}y). \tag{1.14}$$

With these concepts and notations, the question of the modelling capabilities offered when considering probability measures can be addressed.

[2]For any $m, m' \in \mathbf{M}(\mathbf{E})$, the equality $m(\mathrm{d}x) \stackrel{\int}{=} m'(\mathrm{d}x)$ refers to $m(\chi) = m'(\chi)$ for any $\chi \in \mathbf{L}^0(\mathbf{E})$.
[3]Some authors prefer to keep the name "random variable" for measurable mappings in $\mathbf{L}^0(\Omega, \mathbb{R})$.
[4]See Appendix A.2 for more details.

1.1.2 Uncertainty and probability measures

Consider the complete probability space $(\Omega, \Sigma, \mathbb{P})$ and assume that this space is sufficiently large to model all the randomness we will be interested in. Let $X \in \mathbf{L}^0(\Omega, \mathbb{R})$ be a real-valued random variable and denote $p \doteq X_*\mathbb{P} \in \mathbf{M}_1(\mathbb{R})$ the law of X. Since the probability measure $\mathbb{P}$ is not known in general, the definition of the law p does not allow for determining the values taken by p on $\mathcal{B}(\mathbb{R})$ which is a problem if p is not completely known a priori. In some cases, partial knowledge about X can be translated into a probability measure on a sub-σ-algebra of $\mathbb{R}$, as in the following example.

Example 1.1. If we only know that X has its image in the Borel subset A in $\mathcal{B}(\mathbb{R})$ with probability $\alpha \in [0,1]$, then this information can be encoded via the sub-σ-algebra $\mathcal{A} \doteq \{\emptyset, A, A^c, \mathbb{R}\}$ of $\mathbb{R}$ with the probability measure p on $(\mathbb{R}, \mathcal{A})$ characterised by $p(A) = \alpha$. Similarly, if nothing is known about X, then this can be encoded via the *trivial* sub-σ-algebra $\{\emptyset, \mathbb{R}\}$.

The concept of sub-σ-algebra allows for considering different levels of knowledge and for this reason, is used for defining conditional expectations [Loève, 1978, Chapt. 27]. However, we will see in the next example that their use can become challenging in some situations.

Example 1.2. Let p be a probability measure on $(\mathbf{E}, \mathcal{E})$ and let p' be another probability measure on $(\mathbf{E}, \mathcal{E}')$, with $\mathcal{E}' \subset \mathcal{E}$, then for any $a \in (0,1)$, the probability measure $q_a = (1-a)p + ap'$ can only be defined on the coarsest σ-algebra, that is $\mathcal{E}'$. When considering the extreme case where $\mathcal{E}'$ is the trivial σ-algebra $\{\emptyset, \mathbf{E}\}$, it results that nothing is known about q_a, however small is a.

One way to bypass this drawback is to single out a finite reference measure λ in $\mathbf{M}(\mathbf{E})$ and to define an extended version of p' denoted $\bar{p}'$ as a *uniform* probability measure as follows

$$(\forall A \in \mathcal{E}) \qquad \bar{p}'(A) = \frac{\lambda(A)}{\lambda(\mathbf{E})}. \tag{1.15}$$

In this way, the probability of a given event in $\mathbf{E}$ is equal to the probability of any other event of the same "size" **w**ith **r**espect **t**o (w.r.t.) the measure λ. In other words, no area of the space $\mathbf{E}$ is preferred over any other. Besides the facts that a reference measure is required and that the size of the space is limited, this way of modelling the information is not completely equivalent to the absence of information. There exist ways of modelling uncertainty on a probability measure itself, such as with Dirichlet processes [Ferguson, 1973] and to some extent with Wishart distributions [Wishart, 1928]; yet, these solutions do not directly help with the non-informative case since they require additional parameters to be set up.

Another important aspect of probability theory is the gap between probability measures on countable and uncountable sets, as explained in the following example.

Example 1.3. Let $\mathbf{X}$ be a countable set equipped with its discrete σ-algebra $\wp(\mathbf{X})$ and assume that some physical systems can be uniquely characterised by its state in $\mathbf{X}$. If p and p' are laws of independent random variables on $\mathbf{X}$ representing some uncertainty about the same physical system, then the information contained in p and p' can be *fused* into a conditional probability measure $\hat{p}(\cdot \,|\, \Delta) \in \mathbf{M}_1(\mathbf{X})$, defined as

$$\hat{p}(B \,|\, \Delta) \doteq \frac{p \rtimes p'(B \times B \cap \Delta)}{p \rtimes p'(\Delta)}, \tag{1.16}$$

where $\Delta \doteq \{(x,x) \text{ s.t. } x \in \mathbf{X}\}$ is the diagonal of $\mathbf{X} \times \mathbf{X}$ and where p and p' are assumed to be *compatible*, i.e., that $p \rtimes p'(\Delta) \neq 0$. Let $w, w' : \mathbf{X} \to [0,1]$ be the probability mass functions induced by p and p' and characterised by

$$p = \sum_{x \in \mathbf{X}} w(x)\delta_x, \qquad \text{and} \qquad p' = \sum_{x \in \mathbf{X}} w'(x)\delta_x, \tag{1.17}$$

then the fused probability measure $\hat{p}(\cdot \,|\, \Delta)$ can be more naturally defined via its probability mass function $\hat{w}$ on $\mathbf{X}$, which is found to be

$$\hat{w} : x \mapsto \frac{w(x)w'(x)}{\sum_{y \in \mathbf{X}} w(y)w'(y)}. \tag{1.18}$$

However, if $\mathbf{X}$ is uncountable and equipped with its Borel σ-algebra $\mathcal{B}(\mathbf{X})$ and if the probability measures p and p' are diffuse on $\mathbf{X}$, then they will not be compatible by construction. Indeed, even though the diagonal Δ can still be defined and is measurable under extremely weak assumptions[5] on $\mathbf{X}$, it holds that $p \rtimes p'(\Delta) = 0$. This is caused by the strong assumption that the probabilities $p(B)$ and $p'(B)$ are known for all measurable subsets in $\mathcal{B}(\mathbf{X})$ and, because p and p' are diffuse, tend to zero when B reduces to a singleton. The introduction of an appropriately coarse sub-σ-algebra on $\mathbf{X}$, such as the one generated by a given countable partition, would allow for recovering some of the results that hold for countable spaces. However, such an approach will not be natural or intuitive in most of the situations.

Overall, there is a need for the introduction of additional concepts that could account for these non-informative types of knowledge. The objective in the next section is to find a way to complement the notions available in probability theory with these weaker forms of representations.

1.2 Measure and probabilistic constraint

Henceforth, we consider a space $\mathbf{X}$ which is assumed to be a closed (or open) subset of $\mathbb{R}^d$ for some $d > 0$ and which is equipped with its Borel σ-algebra $\mathcal{B}(\mathbf{X})$. The general definition and the properties of measure constraints are given in the section below, followed by examples of the more specific concept of probabilistic constraint.

1.2.1 Measure constraint

Using the notion of outer measure defined in the previous section as well as the technical results about the set $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$ detailed in Appendix A.3, we introduce the concept of *measure constraint* as follows.

Definition 1.2. Let M be a measure on $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$, if it holds that the function μ_M defined on the power set $\wp(\mathbf{X})$ of $\mathbf{X}$ as

$$\mu_M : A \mapsto M(F(A, \cdot)) \doteq \int F(A, f) M(\mathrm{d}f) \tag{1.19}$$

[5] See Remark 1.1 of Section 1.4

is an outer measure for a given collection of measurable functions $\{F(A,\cdot)\}_{A\subseteq \mathbf{X}}$ on $\mathbf{L}^0(\mathbf{X},\mathbb{R})$, then M is said to be a *measure constraint* on $\mathbf{X}$ with *characteristic function* F. If m is a finite measure on $\mathbf{X}$ verifying

$$m(B) \leq M(F(B,\cdot)) \tag{1.20}$$

for any $B \in \mathcal{B}(\mathbf{X})$ and

$$M(F(\mathbf{X},\cdot)) = m(\mathbf{X}), \tag{1.21}$$

then M is said to be a measure constraint for m.

The motivation behind the introduction of measure constraints is to partially describe a measure by limiting the mass in some areas while possibly leaving it unconstrained elsewhere. Measure constraints that would bound a measure from below could also be defined using the associated concept of *inner measure*. In general, a measure on $\mathbf{X}$ could be constrained by a measure on the set $\mathbf{L}^0(\mathbf{Y},\mathbb{R})$ of measurable functions on a different set $\mathbf{Y}$, as long as the associated characteristic function F is defined accordingly.

Remark. In Definition 1.2, the condition that μ_M is an outer measure is used to reduce the set of measures on $\mathbf{L}^0(\mathbf{X},\mathbb{R})$ that would verify (1.20) to the ones that have natural properties. As explained by Fremlin [2000, Sect. 113B]: "The idea of the *outer* measure of a set A is that it should be some kind of upper bound for the possible measure of A". In fact, the use of outer measures as a way of dealing with uncertainty has first been proposed by Fagin et al. [1990]. In particular, the condition of monotonocity imposes that if a given mass is allowed in a set A then at least the same mass should be allowed in a larger set $B \supseteq A$. Similarly, the condition of sub-additivity allows for reaching the maximum mass $m(\mathbf{1})$ in several disjoint sets while still verifying $M(F(\mathbf{X},\cdot)) = m(\mathbf{X})$.

For a given characteristic function F, we can introduce the weighted *semi-norm*[6] $\|\cdot\|_F$ on $\mathbf{M}(\mathbf{L}^0(\mathbf{X},\mathbb{R}))$ as

$$\|\cdot\|_F : M \mapsto |M(F(\mathbf{1},\cdot))|. \tag{1.22}$$

In the considered setup, it holds that $\|M\|_F = M(F(\mathbf{X},\cdot))$ since $M(F(\mathbf{X},\cdot))$ is positive. This approach is justified in the following remark.

Remark.* The set of finite signed measures together with the addition $+$ and the multiplication $\rtimes$ form a vector space. The fact that $\|\cdot\|_F$ is a semi-norm on this vector space is easy to verify. It is not a norm since $\|M\|_F = 0$ does not imply that M is the null measure in general.

In the following list of properties, the statements regarding outer measure are drawn from Fremlin [2000, Sect. 113].

Property 1.1. *Using the notation of Definition 1.2:*

1.1.1 *The condition* $\|M\|_F = m(\mathbf{X})$ *implies that the inequality* (1.20) *turns into an equality when* $B = \mathbf{X}$.

[6] A function from a given vector space to the corresponding field that is absolutely homogeneous and that verifies the triangle inequality. A semi-norm p that also verifies $(p(x) = 0) \Rightarrow (x = 0)$ is a norm.

1.1.2 As an outer measure, μ_M induces a σ-algebra $\mathcal{X}_M$ of subsets of $\mathbf{X}$ composed of sets A verifying

$$(\forall C \subseteq \mathbf{X}) \qquad \mu_M(C) = \mu_M(C \cap A) + \mu_M(C \cap A^c), \tag{1.23}$$

*which are referred to as μ_M-*measurable sets*. The measure space $(\mathbf{X}, \mathcal{X}_M, \mu_M)$ is a complete measure space.*

1.1.3 Let $(\mathbf{X}, \mathcal{X}, m)$ be a measure space, then m induces an outer measure m^* on $\mathbf{X}$ defined as follows

$$(\forall C \subseteq \mathbf{X}) \qquad m^*(C) = \inf\{\mu(A) \ \textit{s.t.}\ A \in \mathcal{X}, \ \ A \supseteq C\}. \tag{1.24}$$

We will be particularly interested in the situation where m is a probability measure, in which case M satisfies $\|M\|_F = m(\mathbf{X}) = 1$ and is said to be a *probabilistic constraint*. The advantage with condition (1.21) is that if m is a finite measure that is not a probability measure, then m and M can be renormalised to be respectively a probability measure and a probabilistic constraint by dividing the inequality (1.20) by the total mass $m(\mathbf{X})$.

Remark. A direct consequence of (1.20) is that a lower bound for $m(B)$ is also available for any $B \in \mathcal{B}(\mathbf{X})$ and is found to be

$$m(B) = m(\mathbf{X}) - m(B^c) \geq m(\mathbf{X}) - M(F(B^c, \cdot)). \tag{1.25}$$

The information provided by this lower bound is limited since μ_M is sub-additive and might reach $m(\mathbf{X})$ on any set $B' \subset \mathbf{X}$. In this case, (1.25) only implies that $m(B') \geq 0$ which is not informative.

An especially useful case is found when M is supported by the set $\mathbf{L}^\infty(\mathbf{X}, \mathbb{R})$ of bounded measurable functions and when there exists a function $F' : \wp(\mathbf{X}) \to \mathbf{L}^\infty(\mathbf{X})$ such that[7]

$$F : (A, f) \mapsto \|F'(A) \cdot f\|, \tag{1.26}$$

where $\|\cdot\|$ is the uniform norm on $\mathbf{L}^\infty(\mathbf{X}, \mathbb{R})$. This expression of F does not depend on the sign of f so that assuming that a constraint with such a characteristic function is defined on $\mathbf{L}^\infty(\mathbf{X})$ rather than $\mathbf{L}^0(\mathbf{X}, \mathbb{R})$ is not restrictive. All outer measures do not take the form assumed in (1.19) with F as in (1.26), but this case offers suitably varied configurations by combining a linear part and a very sub-additive part, that is the measure by M and the uniform norm respectively.

The case where F' is such that $F' : A \mapsto \mathbf{1}_A$, will be understood as the default situation in the sense that $F(A, f) = \|\mathbf{1}_A \cdot f\| = \sup_A f$ will be considered when the characteristic function of a measure constraint is not specified. The subset of probabilistic constraint with such a supremum-based characteristic function is denoted $\mathbf{C}_1(\mathbf{X})$ and the weighted semi-norm $\|\cdot\|_F$ is simply denoted $\|\cdot\|$ and verifies

$$\|\cdot\| : M \mapsto \int \|f\| M(\mathrm{d}f). \tag{1.27}$$

In the following property, $\mathcal{L}^\infty(\mathbf{X})$ denotes the Borel σ-algebra induced on $\mathbf{L}^\infty(\mathbf{X})$ by the topology on $\mathbf{L}^0(\mathbf{X})$ studied in Appendix A.3.

[7] We assume that $\sup : \mathbf{L}^\infty(\mathbf{X}) \to \mathbb{R}$ is measurable.

Property 1.2. *Let $M \in \mathbf{C}_1(\mathbf{X})$ be a probabilistic constraint, then:*

1.2.1 There exists a probabilistic constraint $M^\dagger \in \mathbf{C}_1(\mathbf{X})$ that is equivalent to M and which can be determined via the following rescaling procedure:

$$(\forall F \in \mathcal{L}^\infty(\mathbf{X})) \qquad M^\dagger(F) = \int \mathbf{1}_F(f^\dagger)\|f\|M(\mathrm{d}f), \tag{1.28}$$

where the function $f^\dagger \in \mathbf{L}^\infty(\mathbf{X})$ is defined as

$$f^\dagger = \begin{cases} \dfrac{f}{\|f\|} & \text{if } \|f\| \neq 0 \\ \mathbf{1} & \text{otherwise.} \end{cases} \tag{1.29}$$

Remark. The definition of $f^\dagger$ when $\|f\| = 0$ is irrelevant because of the form of (1.28). Yet, considering $f^\dagger = \mathbf{1}$ when $\|f\| = 0$ implies that $f^\dagger$ is always in the subset

$$\mathbf{L}(\mathbf{X}) \doteq \{f \in \mathbf{L}^\infty(\mathbf{X}) \text{ s.t. } \|f\| = 1\} \tag{1.30}$$

of measurable functions with uniform norm equal to one. Probabilistic constraints in the set $\mathbf{C}_1^*(\mathbf{X}) \doteq \mathbf{M}_1(\mathbf{L}(\mathbf{X}))$ will be referred to as *canonical probabilistic constraints.*

Proof. By construction, it holds that the support of $M^\dagger$ is included in $\mathbf{L}(\mathbf{X})$, so that $\|M^\dagger\| = M^\dagger(\mathbf{L}^\infty(\mathbf{X})) = 1$. The measure $M^\dagger$ is then both a probability measure and a probabilistic constraint. We now have to show that M and $M^\dagger$ constrain the same probability measures: Let $p \in \mathbf{M}_1(\mathbf{X})$ be a probability measure constrained by M, then

$$(\forall B \in \mathcal{B}(\mathbf{X})) \qquad m(B) \leq \int \|\mathbf{1}_B \cdot f\|M(\mathrm{d}f) = \int \|\mathbf{1}_B \cdot f^\dagger\|M'(\mathrm{d}f), \tag{1.31}$$

where $M'(\mathrm{d}f) \stackrel{f}{=} \|f\|M(\mathrm{d}f)$ so that, by a change of variable,

$$m(B) \leq \int \|\mathbf{1}_B \cdot f^\dagger\|M'(\mathrm{d}f) = \int \|\mathbf{1}_B \cdot f'\|M^\dagger(\mathrm{d}f'), \tag{1.32}$$

which terminates the proof. □

1.2.2 The unary operation of Property 1.2.1 does not affect the norm since the equality $\|M^\dagger\| = \|M\|$ holds by construction for any $M \in \mathbf{C}_1(\mathbf{X})$. It is also idempotent *and it distributes over the product $\rtimes$, i.e.,*

$$(M^\dagger)^\dagger = M^\dagger \qquad \text{and} \qquad (M \rtimes M')^\dagger = M^\dagger \rtimes M'^\dagger \tag{1.33}$$

hold for any $M, M' \in \mathbf{C}_1(\mathbf{X})$. Moreover, any canonical probabilistic constraint $P \in \mathbf{C}_1^*(\mathbf{X})$ verifies $P^\dagger = P$.

1.2.3 A natural way to approximate M by a probability measure in $\mathbf{M}_1(\mathbf{X})$ is to consider the least informative probability measure p that is constrained by M. This is possible when there exists a bounded S of $\mathbf{X}$ such that M is supported by functions in $\mathbf{L}^\infty(\mathbf{X})$ which support is in S. In this case, p is characterised by

$$(\forall B \in \mathcal{B}(\mathbf{X})) \qquad p(B) \propto \int \mathbf{1}_B(x)f(x)\lambda(\mathrm{d}x)M(\mathrm{d}f), \tag{1.34}$$

where λ is a reference measure on $\mathbf{X}$, e.g., the Lebesgue measure. The condition on the support of M guarantees that the normalisation factor $\int \lambda(f) M(\mathrm{d}f)$ is finite. An alternative point of view on this operation will be given in Property 1.4.

1.2.4 If $M^\dagger$ is supported by the set $\mathbf{I}(\mathbf{X})$ of measurable indicator functions, then the conditions (1.20) *and* (1.21) *can be replaced by "m agrees with the measure induced by μ_M", i.e.,*

$$(\forall B \in \mathcal{B}(\mathbf{X})) \qquad (B \in \mathcal{X}_M) \Rightarrow (m(B) = \mu_M(B)), \tag{1.35}$$

where $\mathcal{X}_M$ is the σ-algebra of μ_M-measurable subsets.

1.2.5 If $X, X' \in \mathbf{L}^0(\Omega, \mathbf{X})$ are two independent random variables on $\mathbf{X}$ with respective laws p and p' and if M and M' are probabilistic constraints for p and p' respectively, then the joint law $p \rtimes p' \in \mathbf{M}_1$ verifies

$$p \rtimes p'(\hat{B}) \leq \int \|\mathbf{1}_{\hat{B}} \cdot (f \rtimes f')\| M(\mathrm{d}f) M'(\mathrm{d}f') \tag{1.36}$$

for any $\hat{B} \in \mathcal{B}(\mathbf{X}) \otimes \mathcal{B}(\mathbf{X})$. This is only an implication since a joint probability measure with a constraint of the same form might not correspond to independent random variables.

We deduce from Property 1.2.5 that another concept is needed to describe this special class of joint probabilistic constraints.

Definition 1.3. Let $X, X' \in \mathbf{L}^0(\Omega, \mathbf{X})$ be two random variables on $\mathbf{X}$ with joint law $\hat{p}$, then X and X' are said to be *independently constrained* by $\hat{M} \in \mathbf{C}_1(\mathbf{X} \times \mathbf{X})$ if there exist $M, M' \in \mathbf{C}_1(\mathbf{X})$ such that

$$\hat{M}(\hat{F}) = \int \mathbf{1}_{\hat{F}}(f \rtimes f') M(\mathrm{d}f) M'(\mathrm{d}f') \tag{1.37}$$

for any $\hat{F} \in \mathcal{L}^\infty(\mathbf{X} \times \mathbf{X})$.

Definition 1.3 implies that any function $\hat{f}$ in the support of $\hat{M}$ is such that there exist f and f' in $\mathbf{L}^\infty(\mathbf{X})$ for which $\hat{f} = f \rtimes f'$. Functions of this type can be said to be *separable*. The concept of independently constrained random variables introduced in Definition 1.3 differs in general from the standard concept of independence since *a*) independently constrained random variables might actually be correlated, and *b*) independent random variables that are not known to be independent might be represented by a probabilistic constraint that does not exclude correlation. However, the concepts coincide when the involved probabilistic constraints are equivalent to probability measures.

Remark. Using the notations of Definition 1.3 and the notion of convolution of measures that will be defined in Section 1.4.1, the condition (1.37) for $\hat{M}$ to independently constrain X and X' can be restated as a convolution on the semigroup $(\mathbf{L}^\infty, \rtimes)$ as

$$\hat{M} = M * M'. \tag{1.38}$$

In order to illustrate the use of measure constraints for modelling uncertainty, several examples of probabilistic constraints are given in the next section.

1.2.2 Examples of probabilistic constraints

Throughout this section, X will denote a random variable from the probability space $(\Omega, \Sigma, \mathbb{P})$ to $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ and $P \in \mathbf{C}_1^*(\mathbf{X})$ will be a canonical probabilistic constraint for the law $p \doteq X_*\mathbb{P} \in \mathbf{M}_1(\mathbf{X})$ of X.

Uninformative case

If $P = \delta_{\mathbf{1}}$ where $\mathbf{1}$ is the function everywhere equal to 1, then the only information provided by P on p is that

$$p(B) \leq \int \|\mathbf{1}_B \cdot f\| P(\mathrm{d}f) = 1 \tag{1.39}$$

for any $B \in \mathcal{B}(\mathbf{X})$, which is non-informative since p is already known to be a probability measure. In other words, nothing is known about the random variable X.

Remark. A partial order $\leq$ can be defined on $\mathbf{L}^0(\mathbf{X})$ as

$$f \leq g \Leftrightarrow \forall x \in \mathbf{X}(f(x) \leq g(x)), \tag{1.40}$$

so that $(\mathbf{L}^0(\mathbf{X}), \leq)$ is a *partially ordered set*, or *poset*. The maximal element in the subset $\mathbf{L}(\mathbf{X})$ of $\mathbf{L}^0(\mathbf{X})$ is found to be the function $\mathbf{1}$. This highlights a possible connection between information content and the partial order $\leq$ on $\mathbf{L}(\mathbf{X})$.

Remark. The σ-algebra of μ_P-measurable sets is the trivial σ-algebra $\{\emptyset, \mathbf{X}\}$.

Indicator function

If there exists $A \in \mathcal{B}(\mathbf{X})$ such that $P = \delta_{\mathbf{1}_A}$, then it holds that

$$p(B) \leq \|\mathbf{1}_{A\cap B}\| = \begin{cases} 1 & \text{if } A \cap B \neq \emptyset \\ 0 & \text{otherwise,} \end{cases} \tag{1.41}$$

for any $B \in \mathcal{B}(\mathbf{X})$. As a probability measure, p is always less or equal to 1, so that the only informative part in the previous inequality is:

$$(\forall B \in \mathcal{B}(\mathbf{X})) \qquad (A \cap B = \emptyset) \Rightarrow (p(B) = 0), \tag{1.42}$$

that is, all the probability mass of B lies within A. This means that the random variable X is only known to be in A almost surely. A uniform distribution over A, assuming it can be defined, would not model the same type of knowledge as the corresponding interpretation would be “realisations of X are equally likely everywhere inside A”, whereas the interpretation associated with the probabilistic constraint $\delta_{\mathbf{1}_A}$ is “whatever the law of X, it is only known that realisations will be inside A”.

Remark.* The σ-algebra of μ_P-measurable subsets is found to be the completion of the σ-algebra $\mathcal{X}_B \doteq \{\emptyset, B, B^c, \mathbf{X}\}$ w.r.t. any measure m on $(\mathbf{X}, \mathcal{X}_B)$ verifying $m(B) > 0$ as well as $m(B^c) = 0$. Further details about completion procedures of σ-algebras are given in Appendix A.5.

Upper bound

If there exists $f \in \mathbf{L}(\mathbf{X})$ such that $P = \delta_f$ then for any $B \in \mathcal{B}(\mathbf{X})$,

$$p(B) \leq \int \|\mathbf{1}_B \cdot f'\| P(\mathrm{d}f') = \int \|\mathbf{1}_B \cdot f'\| \delta_f(\mathrm{d}f') = \|\mathbf{1}_B \cdot f\|. \tag{1.43}$$

In this case, δ_f can be seen as the simplest non-trivial form of probabilistic constraint. The two previous examples are special cases with $f = \mathbf{1}$ and $f = \mathbf{1}_A$.

Remark. Probabilistic constraints of this form are equivalent to *possibility distributions* [Negoita et al., 1978, Dubois and Prade, 1988, 1998, 2000] and are also related to the notion of *membership function* of a *fuzzy set* [Zadeh, 1965]. However, the approach considered in this work does not rely on the notion of fuzzy set and the form considered here is only used as a simple example of probabilistic constraint.

The following alternative formulation is useful for understanding the mechanism behind this simple type of probabilistic constraint.

Proposition 1.1. *Let $P \in \mathbf{C}_1(\mathbf{X})$ be a probabilistic constraint for $p \in \mathbf{M}_1(\mathbf{X})$ of the form $P = \delta_f$ with $f \in \mathbf{L}(\mathbf{X})$, then it holds that*

$$(\forall \alpha \in \operatorname{Im} f) \qquad f_* p([0, \alpha]) \leq \alpha. \tag{1.44}$$

Proof. Let C_α denote the interval $[0, \alpha]$ for any $\alpha \in \mathbb{R}^+$. We first prove that

$$(\forall A \in \mathcal{E}, \forall \alpha \in [0,1]) \qquad (A \subseteq f^{-1}[C_\alpha]) \Rightarrow (p(A) \leq \alpha) \tag{1.45}$$

holds in the following way: for any $A \in \mathcal{E}$, we have

$$\begin{aligned} p(A) \leq \|\mathbf{1}_A \cdot f\| &\Leftrightarrow p(A) \leq \inf\{\alpha' \text{ s.t. } f^{-1}[C_{\alpha'}] \supseteq A\} && (1.46a) \\ &\Leftrightarrow \forall \alpha \in \{\alpha' \text{ s.t. } f^{-1}[C_{\alpha'}] \supseteq A\}(p(A) \leq \alpha) && (1.46b) \\ &\Leftrightarrow \forall \alpha((A \subseteq f^{-1}[C_\alpha]) \Rightarrow (p(A) \leq \alpha)). && (1.46c) \end{aligned}$$

Also, note that if α is not in $\operatorname{Im} f$ then it holds that $f^{-1}[C_\alpha] = f^{-1}[C_{\alpha_-}]$ with

$$\alpha_- = \max\{\alpha' \in \operatorname{Im} f \text{ s.t. } \alpha' \leq \alpha\}, \tag{1.47}$$

so that (1.45) can be equivalently stated with $\alpha \in \operatorname{Im} f$ rather than $\alpha \in [0,1]$. Then, since f is measurable, the monotonicity of measures implies that

$$\forall A \in \mathcal{E}\Big((A \subseteq f^{-1}[C_\alpha]) \Rightarrow (p(A) \leq \alpha)\Big) \Leftrightarrow \Big(p(f^{-1}[C_\alpha]) \leq \alpha\Big) \tag{1.48}$$

holds for any $\alpha \in \operatorname{Im} f$, which terminates the proof. □

The alternative formulation of Proposition 1.1 is illustrated in the following example.

Example 1.4. Assume that the only knowledge about X is that it is almost surely outside a given measurable subset $A \subset \mathbf{X}$. Considering the measurable function $f = \mathbf{1}_{A^c}$, which satisfies $\operatorname{Im} f = \{0, 1\}$, we obtain from Proposition 1.1 that the only constraints on the law $p \doteq X_* \mathbb{P}$ of X are

$$f_* p(\{0\}) = p(A) = 0 \qquad \text{and} \qquad f_* p(\{0,1\}) = p(\mathbf{X}) = 1, \tag{1.49}$$

that is, it is only known that there is no probability mass in the subset A.

Combination of upper bounds

If the probabilistic constraint P is of the form $P = \sum_{i=1}^{N} a_i \delta_{f_i}$, so that the support of P is in the subset $\mathbf{I}(\mathbf{X}) \subset \mathbf{L}^0(\mathbf{X})$ of measurable indicator functions, then

$$p(B) \leq \int \|\mathbf{1}_B \cdot f'\| P(\mathrm{d}f') = \sum_{i=1}^{N} a_i \|\mathbf{1}_B \cdot f_i\| = \sum_{i=1}^{N} a_i \sup_B f_i \tag{1.50}$$

for any $B \in \mathcal{B}(\mathbf{X})$. This combination of upper bounds is not equivalent to a single bound in general and allows for modelling more accurate information. For instance, if $P = 0.5\delta_{\mathbf{1}_B} + 0.5\delta_{\mathbf{1}_{B'}}$ with $B \cap B' = \emptyset$, then 50% of the probability mass of p is in B and 50% is in B'.

Constraint based on a partition

If there exists a measurable countable partition π of $\mathbf{X}$ and if the probability measure P is of the form

$$P = \sum_{B \in \pi} q(B) \delta_{\mathbf{1}_B}, \tag{1.51}$$

where q is a probability measure on the sub-σ-algebra generated by π, then

$$(\forall B \in \pi) \qquad p(B) = q(B), \tag{1.52}$$

i.e., the information available about p is the one embedded into q.

Proof. From the definition of probabilistic constraint, we deduce that the inequality $p(B) \leq q(B)$ holds for any $B \in \pi$. Assume there exists $B \in \pi$ such that $p(B) < q(B)$. This assumption implies that

$$p(\mathbf{X}) = \sum_{B \in \pi} p(B) < \sum_{B \in \pi} q(B) = q(\mathbf{X}) = 1, \tag{1.53}$$

which is a contradiction since p is a probability measure. □

This example shows that probabilistic constraints are versatile enough to model the same level of information as with a sub-σ-algebra generated by a partition.

Remark. Following Property 1.1.3, we define the outer measure q^* induced by q on $\mathbf{X}$ as follows

$$(\forall C \subseteq \mathbf{X}) \qquad q^*(C) = \inf\{q(B) \text{ s.t. } B \in \sigma(\pi), \;\; B \supseteq C\}. \tag{1.54}$$

Since π is a partition of $\mathbf{X}$, it is easy to check that the outer measure μ_P verifying

$$\mu_P(C) = \sum_{B \in \pi \,:\, B \cap C \neq \emptyset} q(B) \tag{1.55}$$

for all subsets C of $\mathbf{X}$ is equal to the outer measure q^*.

Probability measure

Given that the singletons of $\mathbf{X}$ are measurable[8], if the support of P is in the subset $\mathbf{I}_{\mathrm{s}}(\mathbf{X}) \subset \mathbf{I}(\mathbf{X})$ of indicator functions on singletons, then there exists a measure q in $\mathbf{M}_1(\mathbf{X})$ such that, for any $B \in \mathcal{B}(\mathbf{X})$, it holds that

$$p(B) \leq \int_{\mathbf{I}_{\mathrm{s}}(\mathbf{X})} \|\mathbf{1}_B \cdot f\| P(\mathrm{d}f) = \int \mathbf{1}_B(x) q(\mathrm{d}x) = q(B), \tag{1.56}$$

from which we conclude that $p = q$. The proof of the equality is very similar to the proof for constraints based on a partition.

Plausibility

If P has the form

$$P = \sum_{A \in \mathcal{A}} g(A) \delta_{\mathbf{1}_A} \tag{1.57}$$

for some set $\mathcal{A}$ of measurable subsets and some function $g : \mathcal{A} \to \mathbb{R}^+$, then

$$p(B) \leq \int_{\mathbf{I}(\mathbf{X})} \|\mathbf{1}_B \cdot f\| P(\mathrm{d}f) = \sum_{A \in \mathcal{A} \,:\, A \cap B \neq \emptyset} g(A) \tag{1.58}$$

for any $B \in \mathcal{B}(\mathbf{X})$, and the probabilistic constraint P reduces to a *plausibility* as defined in the context of Dempster–Shafer theory [Dempster, 1967, Shafer, 1976]. A non-technical overview of the concepts of this theory is given by Williamson [1989, Sect. 4.4]. The two previous examples can be seen as special cases of plausibility where g is a measure on $\mathcal{A} = \pi$ or $\mathcal{A} = \mathbf{I}_{\mathrm{s}}(\mathbf{X})$. Various extensions of the concept of plausibility have been studied [Yen, 1990, Friedman and Halpern, 2001] and we do not claim that the approach considered here is more general. It is rather the considered structure that will prove to be beneficial when handling uncertainty.

The multiple cases given in this section show that probabilistic constraints can model information with various degrees of precision. In this work, we will be mainly interested in the most informative and the most uninformative cases.

1.3 Operations on measure constraints

It is essential to be able to adapt a probabilistic constraint when the underlying probability measure is transformed, such as when considering a pushforward for the considered measure w.r.t. a given measurable function. In the following sections, $\mathbf{X}_1$ and $\mathbf{X}_2$ are assumed to be closed (or open) subsets of possibly different Euclidean spaces. The technical results detailed in Appendix A.4 will be useful in this section.

1.3.1 Pushforward

The operation of pushforwarding measures from one measurable space to another is central to measure theory and should therefore be translated in terms of measure constraints. In other words, if M is a constraint for the measure m with characteristic function F and ξ is a measurable function, then what is the constraint for $\xi_* m$?

[8]This is the case with most of the usual topological spaces equipped with their Borel σ-algebras.

The objective is to find sufficient conditions for the existence of a measure constraint M' with characteristic function F verifying

$$M'(F(B,\cdot)) = M(F(\xi^{-1}[B],\cdot)) \tag{1.59}$$

for all $B \in \mathcal{B}(\mathbf{X}_2)$. It first appears from the properties of outer measures described in Section 1.1.1 that $A \mapsto M(F(\xi^{-1}[A],\cdot))$ is an outer measure, so that M' is possibly a measure constraint for $\xi_* m$. Using the results of Appendix A.4, the existence of such a measure constraint is proved in the next proposition when F takes the form

$$F : (B, f) \mapsto \|(F' \circ \mathbf{1}_B) \cdot f\| \tag{1.60}$$

for some suitable function $F' : \mathbb{R}^+ \to \mathbb{R}^+$.

Proposition 1.2. *Let M be a measure constraint on $\mathbf{X}_1$ with characteristic function F of the form* (1.60) *and let ξ be a measurable mapping from $\mathbf{X}_1$ to $\mathbf{X}_2$, then the measure constraint M' on $\mathbf{X}_2$ defined as the pushforward of M by the mapping T_ξ from $\mathbf{L}^\infty(\mathbf{X}_1)$ to $\mathbf{L}^\infty(\mathbf{X}_2)$ characterised by*

$$(\forall f \in \mathbf{L}^\infty(\mathbf{X}_1)) \qquad T_\xi(f) : y \mapsto \sup_{\xi^{-1}[\{y\}]} f \tag{1.61}$$

verifies (1.59)*.*

Proof. The first step is to rewrite the uniform norm over $\mathbf{X}_1$ in a suitable way as

$$\|(F' \circ \mathbf{1}_{\xi^{-1}[B]}) \cdot f\| = \sup_{y \in \mathbf{X}_2} \Big(\sup_{x \in \xi^{-1}[\{y\}]} \big(F'(\mathbf{1}_{\xi^{-1}[B]}(x)) f(x) \big) \Big) \tag{1.62a}$$

$$= \sup_{y \in \mathbf{X}_2} \Big(F'(\mathbf{1}_B(y)) \sup_{\xi^{-1}[\{y\}]} f \Big) \tag{1.62b}$$

for any $f \in \mathbf{L}^\infty(\mathbf{X}_1)$ and any $B \in \mathcal{B}(\mathbf{X}_2)$, where the second line is explained by the fact that the function $F' \circ \mathbf{1}_{\xi^{-1}[B]}$ is constant over $\xi^{-1}[\{y\}]$, and is equal to $F'(\mathbf{1}_B(y))$ everywhere on this subset. It holds that $T_\xi(f) \in \mathbf{L}^\infty(\mathbf{X}_2)$ by Corollary A.1 since $\{\xi^{-1}[\{y\}] \times \{y\} : y \in B\}$ is a measurable subset of $\mathbf{X}_1 \times \mathbf{X}_2$ for any $B \in \mathcal{B}(\mathbf{X}_2)$. The fact that T_ξ is measurable follows from the assumption that sup is measurable on $\mathbf{L}^\infty(\mathbf{X})$. It then holds that

$$\int \|(F' \circ \mathbf{1}_B) \cdot f'\| M'(\mathrm{d}f') = \int \|(F' \circ \mathbf{1}_{\xi^{-1}[B]}) \cdot f\| M(\mathrm{d}f), \tag{1.63}$$

where M' is defined as the pushforward $(T_\xi)_* M$. □

In the following example, the operation of marginalisation is translated to probabilistic constraints.

Example 1.5. Let $\mathbf{X}$ be the Cartesian product of $\mathbf{X}_1$ and $\mathbf{X}_2$ and let p be a given probability measure on $\mathbf{X}$. The objective is to project a given probabilistic constraint $P \in \mathbf{C}_1(\mathbf{X})$ for p into a probabilistic constraint P' in $\mathbf{C}_1(\mathbf{X}_2)$ for the corresponding marginal $p' \in \mathbf{M}_1(\mathbf{X}_2)$ characterised by $p'(B) = p(\mathbf{X}_1 \times B)$. Marginalisation can be performed on P' by using Proposition 1.2 with the canonical projection map $\xi : \mathbf{X} \to \mathbf{X}_2$, from which we find that

$$(\forall B \in \mathcal{B}(\mathbf{X}_2)) \qquad p(B) \leq \int \|\mathbf{1}_B \cdot f'\| (T_\xi)_* P(\mathrm{d}f'), \tag{1.64}$$

where the mapping $T_\xi : \mathbf{L}^\infty(\mathbf{X}) \to \mathbf{L}^\infty(\mathbf{X}_2)$ is found to be

$$T_\xi(f) : y \mapsto \sup_{\xi^{-1}[\{y\}]}(f) = \|f(\cdot, y)\|. \tag{1.65}$$

The fact that $T_\xi(f)$ is measurable can be obtained directly from Corollary A.1 by considering C constant and equal to $\mathbf{X}_1$.

Another case, studied in the next example, is the pushforward of a probabilistic constraint defined as being equivalent to a probability measure on a sub-σ-algebra.

Example 1.6. Considering again the case detailed in Section 1.2.2 where π is a measurable countable partition of $\mathbf{X}_1$ and where the probabilistic constraint P is of the form

$$P = \sum_{B\in\pi} p'(B)\delta_{\mathbf{1}_B}, \tag{1.66}$$

where p' is a probability measure on the sub-σ-algebra generated by π, then for any measurable mapping $\xi : \mathbf{X}_1 \to \mathbf{X}_2$ such that $\sigma(\xi) \subseteq \sigma(\pi)$, it holds that

$$(\forall B \in \pi) \qquad T_\xi(\mathbf{1}_B) : y \mapsto \sup_{\xi^{-1}[\{y\}]}(\mathbf{1}_B) = \mathbf{1}_{\xi(B)}(y), \tag{1.67}$$

that is, since $\xi[B]$ is a singleton,

$$T_\xi(\mathbf{1}_B) = \mathbf{1}_{\xi(B)} \in \mathbf{I}_{\mathrm{s}}(\mathbf{X}_2), \tag{1.68}$$

so that $(T_\xi)_*P$ has its support in $\mathbf{I}_{\mathrm{s}}(\mathbf{X}_2)$ and is therefore equivalent to a probability measure on $\mathbf{X}_2$.

The concepts of pushforward and marginalisation for probabilistic constraints are important and will be useful in practice. We now consider the converse operation, referred to as pullback and which will also contribute to the formulation of the main results in this work.

1.3.2 Pullback

The objective in this section is to understand in which situations it is possible to reverse the operation of pushforwarding measures and measure constraints. We first define what is understood as a pullback.

Definition 1.4. Let m be a measure on the space $\mathbf{X}_2$ and let ξ be a measurable mapping from $\mathbf{X}_1$ to $\mathbf{X}_2$, then a *pullback* measure, denoted ξ^*m, is a measure in $\mathbf{M}_1(\mathbf{X}_1)$ satisfying

$$\xi_*(\xi^*m) = m. \tag{1.69}$$

Note that there are possibly many pullback measures for a given measure so that the concept of measure constraint will be useful in order to represent this multiplicity, as in the following proposition. As with the pushforward, we want to obtain the existence of a measure constraint M' with characteristic function F for the pullback of a measure $m \in \mathbf{M}_1(\mathbf{X}_2)$ such that

$$M'(F(\xi^{-1}[B], \cdot)) = M(F(B, \cdot)) \tag{1.70}$$

holds for any $B \in \mathcal{B}(\mathbf{X}_2)$, where M is a measure constraint for m. It appears from the properties of outer measures described in Section 1.1.1 that $A \mapsto M(F(\xi[A], \cdot))$ is an outer measure, so that M' is possibly a measure constraint for the pullback measures of m.

Proposition 1.3. *Let M be a measure constraint on $\mathbf{X}_2$ with characteristic function $F : (B, f) \mapsto \|\mathbf{1}_B \cdot f\|$ and let ξ be a measurable mapping from $\mathbf{X}_1$ to $\mathbf{X}_2$, then the measure constraint M' on $\mathbf{X}_1$ defined as the pushforward of M by the mapping $T'_\xi : \mathbf{L}^\infty(\mathbf{X}_2) \to \mathbf{L}^\infty(\mathbf{X}_1)$ defined as*

$$T'_\xi : f \mapsto f \circ \xi \tag{1.71}$$

verifies (1.70).

Proof. We rewrite the uniform norm of $\mathbf{1}_B \cdot f$ as

$$\|\mathbf{1}_B \cdot f\| = \|\mathbf{1}_{\xi^{-1}[B]} \cdot (f \circ \xi)\| \tag{1.72}$$

for any $f \in \mathbf{L}^\infty(\mathbf{X}_2)$ and any $B \in \mathcal{B}(\mathbf{X}_2)$. Since the composition of measurable mappings is measurable, we verify that the codomain of T'_ξ is $\mathbf{L}^\infty(\mathbf{X}_1)$. Assuming that the mapping T'_ξ is itself measurable, we write

$$\int \|\mathbf{1}_B \cdot f\| M(\mathrm{d}f) = \int \|\mathbf{1}_{\xi^{-1}[B]} \cdot f'\| (T'_\xi)_* M(\mathrm{d}f'), \tag{1.73}$$

which terminates the proof of the proposition. □

Remark. Other transformations than T'_ξ could lead to a measure on $\mathbf{L}^\infty(\mathbf{X}_1)$ that verifies (1.70). However, these measures would not be measure constraints for all the pullback measures of m.

Example 1.7. If m is a known probability measure, then the corresponding canonical probabilistic constraint M would have its support in the set $\mathbf{I}_s(\mathbf{X}_2)$ of indicator functions of singletons in $\mathbf{X}_2$. The induced probabilistic constraint M' on $\mathbf{X}_1$ would then be of the form

$$M' = \sum_{B \in \pi} m'(B) \delta_{\mathbf{1}_B}, \tag{1.74}$$

where π is the partition generated by ξ on $\mathbf{X}_1$ and m' is the probability measure induced by m on the sub-σ-algebra generated by ξ. This example closes a loop started in Example 1.6 where a probabilistic constraint on a sub-σ-algebra generated by a partition was shown to induce a probability measure on $\mathbf{X}_2$.

Now equipped with the concepts of pushforward and pullback for measure constraints and measures, we can address the question of the fusion of different sources of information.

1.4 Fusion of probabilistic constraints

The objective in this section is to fuse independent sources of information about the same physical system. After a section on a technical aspect of measures, we first address the question of how to fuse probabilistic constraints, starting with the most natural case of the fusion of probabilistic constraints expressed on a state space. Then, we will see how this operation can be extended to more general spaces.

1.4.1 Convolution of measures

An operation that will prove to be of importance in the next section is the operation of convolution. In order to define it, additional concepts are required: *a*) a set with a binary operation form a *semigroup* if the binary operation is associative, and *b*) a semigroup is said to be a *topological* semigroup if the binary operation of the semigroup is continuous.

Remark.* The first results in this section are from [Fremlin, 2000, Sect. 444], where the general case of the convolution of finite "quasi-Radon" measures on a topological group is considered. We do not seek this level of generality since all finite measures on a Polish space have sufficiently nice properties (they are *Radon* measures, as defined in Appendix A.1).

We still consider that the set $\mathbf{X}$ is a closed (or open) subset of an Euclidean space[9] equipped with its Borel σ-algebra.

Definition 1.5. Let m and m' be two finite measures on a topological semigroup $(\mathbf{X}, \cdot)$, then the convolution $m * m'$ of m and m' is defined as

$$(\forall A \in \mathcal{B}(\mathbf{X})) \qquad m * m'(A) = m \rtimes m'(\{(x,y) \text{ s.t. } x \cdot y \in A\}). \tag{1.75}$$

The requirement for $(\mathbf{X}, \cdot)$ to be a topological semigroup is justified by the fact that the subset $\{(x,y) \text{ s.t. } x \cdot y \in A\}$ needs to be a Borel set in $\mathbf{X} \times \mathbf{X}$ for (1.75) to be well defined. The two following results are fundamental to the fusion of probabilistic constraints that will be tackled in the next section. These results are corollaries of [Fremlin, 2000, Sect. 444A-D].

Corollary 1.1. *The set function $m * m'$ defined in* (1.75) *is a measure on* $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$.

Corollary 1.2. *If $(\mathbf{X}, \cdot)$ is a topological semigroup, then it holds that*

$$m * (m' * m'') = (m * m') * m'' \tag{1.76}$$

for all finite measures m, m' and m'' on $\mathbf{X}$. If $(\mathbf{X}, \cdot)$ is commutative, then it holds that $m * m' = m' * m$ for all finite measures m and m' on $\mathbf{X}$.

The binary operation on the set of probabilistic constraints that will be introduced in the next section will form a *monoid*, that is, a semigroup with an identity element.

1.4.2 State-space fusion

We assume that there exists a *representative* set $\boldsymbol{\mathcal{X}}$ in which the state of the system of interest can be uniquely characterised and we denote ξ the projection map between $\boldsymbol{\mathcal{X}}$ and $\mathbf{X}$. Assuming that $\boldsymbol{\mathcal{X}}$ is also a closed (or open) subset of an Euclidean space, we equip it with its Borel σ-algebra $\mathcal{B}(\boldsymbol{\mathcal{X}})$.

Remark.* In general, the set $\boldsymbol{\mathcal{X}}$ does not have to be an Euclidean space. Indeed, when considering the σ-algebra on $\boldsymbol{\mathcal{X}}$ induced by the *initial topology* with respect to the mapping ξ, the mapping ξ becomes bi-measurable, and $\boldsymbol{\mathcal{X}}$ can be equipped with the reference measure induced by the one on $\mathbf{X}$.

[9] The results in this section are directly applicable to any Polish space.

Let $X \in \mathbf{L}^0(\Omega, \mathbf{X})$ and $X' \in \mathbf{L}^0(\Omega', \mathbf{X})$ be two random variables in $\mathbf{X}$ based on two complete probability spaces $(\Omega, \Sigma, \mathbb{P})$ and $(\Omega', \Sigma', \mathbb{P}')$ and consider the predicate

$$\text{“}X \text{ and } X' \text{ represent the same physical system”.} \tag{\#}$$

The random variables X and X' are defined on different probability spaces as they might originate from different representations of randomness. As a consequence, the predicate (#) cannot be expressed as $X = X'$. A random variable $X \times X'$ can be defined on the probability space $(\Omega \times \Omega', \Sigma \otimes \Sigma', \mathbb{P} \rtimes \mathbb{P}')$ of joint outcomes/events and it is assumed that the law of $X \times X'$ is the probability measure $p \rtimes p'$ where p and p' are the respective laws of X and X'. This can be seen as an assumption of independence for random variables expressed on different probability spaces.

The predicate (#) can be expressed as the event $\Delta = \{(x, x) \text{ s.t. } x \in \boldsymbol{\mathcal{X}}\}$ in the product σ-algebra $\mathcal{B}(\boldsymbol{\mathcal{X}}) \otimes \mathcal{B}(\boldsymbol{\mathcal{X}})$ because of the representativity of the set $\boldsymbol{\mathcal{X}}$. Yet, several systems could have the same state in $\mathbf{X}$ and (#) cannot be expressed as an event in $\mathcal{B}(\mathbf{X}) \otimes \mathcal{B}(\mathbf{X})$. In order to bypass this issue, we show how probabilistic constraints can be used on $\boldsymbol{\mathcal{X}}$ to incorporate the information contained in (#) for probability measures and probabilistic constraints on $\mathbf{X}$.

*Remark** 1.1. The diagonal Δ of $\boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{X}}$ is measurable in a separable metric space [Bogachev, 2007, Lemma 6.4.2], and remains measurable if $\boldsymbol{\mathcal{X}}$ is generalised to a Hausdorff topological space with a countable base. An interesting result of Nedoma [1957], detailed in Schechter [1996, Chapt. 21], is that the diagonal Δ is never measurable when the cardinality of $\boldsymbol{\mathcal{X}}$ is strictly larger than the cardinality of the continuum.

The fusion of different sources of information can now be formalised as the fusion of probabilistic constraints as shown in the following theorem. A similar rule has been proposed by Mahler [2005, 2007] in the context of fuzzy Dempster-Shafer theory [Yen, 1990].

Theorem 1.1. *Let $X \in \mathbf{L}^0(\Omega, \boldsymbol{\mathcal{X}})$ and $X' \in \mathbf{L}^0(\Omega', \boldsymbol{\mathcal{X}})$ be two random variables in $\boldsymbol{\mathcal{X}}$ and let P and P' be canonical probabilistic constraints for the laws of X and X' respectively, then the posterior probabilistic constraint, denoted $P \star P'$ and obtained by assuming that (#) holds, is characterised by the following convolution of measures on $(\mathbf{L}^\infty(\boldsymbol{\mathcal{X}}), \cdot)$:*

$$P \star P' = \frac{(P * P')^\dagger}{\|P * P'\|}. \tag{1.77}$$

Proof. The main idea of the proof is to build the joint law $p \rtimes p'$ in $\boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{X}}$ from the prior laws p and p' of X and X' and then to study the diagonal $\Delta = \{(x, x) \text{ s.t. } x \in \boldsymbol{\mathcal{X}}\}$. For this purpose, we define $\Delta(B)$ as the intersection $\Delta \cap B \times B$ for any $B \in \mathcal{B}(\boldsymbol{\mathcal{X}})$ and consider the posterior law $\hat{p}(\cdot \,|\, \#)$ characterised by

$$(\forall B \in \mathcal{B}(\boldsymbol{\mathcal{X}})) \qquad \hat{p}(B \,|\, \#) \propto p \rtimes p'(\Delta(B)). \tag{1.78}$$

The law $p \rtimes p'$ verifies that

$$(\forall B \times B' \in \mathcal{B}(\boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{X}})) \qquad p \rtimes p'(B \times B') \leq \int \|\mathbf{1}_{B \times B'} \cdot (f \rtimes f')\| P(\mathrm{d}f) P'(\mathrm{d}f'), \tag{1.79}$$

so that

$$p \rtimes p'(\Delta(B)) \leq \int \|\mathbf{1}_{\Delta(B)} \cdot (f \rtimes f')\| P(\mathrm{d}f) P'(\mathrm{d}f') \tag{1.80a}$$

$$\leq \int \|\mathbf{1}_B \cdot (f \cdot f')\| P(\mathrm{d}f) P'(\mathrm{d}f') \tag{1.80b}$$

holds for any $B \in \mathcal{B}(\boldsymbol{\mathcal{X}})$. We conclude that the posterior probabilistic constraint $P \star P'$ on $\boldsymbol{\mathcal{X}}$ verifies

$$P \star P'(F) = \frac{1}{C} \int \mathbf{1}_F((f \cdot f')^\dagger) \|f \cdot f'\| P(\mathrm{d}f) P'(\mathrm{d}f') \propto (P * P')^\dagger(F) \tag{1.81}$$

for any $F \in \mathcal{L}^\infty(\boldsymbol{\mathcal{X}})$. The possibility of the event $(\#)$ can then be translated on $\mathbf{L}^\infty(\boldsymbol{\mathcal{X}}) \times \mathbf{L}^\infty(\boldsymbol{\mathcal{X}})$ as the likelihood $\ell(\# \,|\, f, f') = \|f \cdot f'\|$, and the constant C can be determined by

$$C = P \rtimes P'(\ell(\# \,|\, \cdot)) = \int \|f \cdot f'\| P(\mathrm{d}f) P'(\mathrm{d}f') = \|P * P'\|. \tag{1.82}$$

The convolution of P and P' is well defined since $(\mathbf{L}^\infty(\boldsymbol{\mathcal{X}}), \cdot)$ is a topological semigroup, as shown in Appendix A.3. □

The meaning of the term "prior" used in the proof of Theorem 1.1 differs from the usual one. For instance, if P and P' in $\mathbf{C}_1(\boldsymbol{\mathcal{X}})$ describe respectively the information made available by an estimation algorithm and the knowledge brought by an observation of the physical system, then both P and P' are considered as prior information and the fused constraint is considered as the posterior information. This is caused by the fact that P and P' are used in a completely symmetrical way, unlike usual Bayesian estimation algorithms that would attempt to compute the posterior probability given the observation.

Remark. For any probabilistic constraints $P, P' \in \mathbf{C}_1(\boldsymbol{\mathcal{X}})$ and any upper bounds $f, f' \in \mathbf{L}(\boldsymbol{\mathcal{X}})$, the real numbers $\|P \star P'\|$ and $\|f \cdot f'\|$ can be seen as the *compatibility* between P and P' or f and f'. Indeed, $\|P \star P'\|$ is the possibility of the event $(\#)$ and

$$\|P \star P'\| = \|f \cdot f'\| \tag{1.83}$$

when $P = \delta_f$ and $P' = \delta_{f'}$, so that the same interpretation can be given to $\|f \cdot f'\|$. In the next chapters, the fusion operation will also be used with *sub-probabilities*, that is measures with total mass less than one, which model the law of a random variable as well as the probability for this random variable to be considered. The compatibility, which is also scaled by the latter, will be useful for assessing the probability of different *configurations.*

Property 1.3. *Using the notations of Theorem 1.1:*

1.3.1 Defining the likelihood function $\ell(\# \,|\, \cdot)$ on $\mathbf{L}^\infty(\boldsymbol{\mathcal{X}}) \times \mathbf{L}^\infty(\boldsymbol{\mathcal{X}})$ as

$$\ell(\# \,|\, \cdot) : (f, f') \mapsto \|f \cdot f'\|, \tag{1.84}$$

the posterior probability measure $P \star P'$ can be expressed for any $F \in \mathcal{L}^\infty(\boldsymbol{\mathcal{X}})$ as

$$P \star P'(F) = \frac{P \rtimes P'(\ell(\# \,|\, \cdot)\Phi(\cdot, F))}{P \rtimes P'(\ell(\# \,|\, \cdot))}, \tag{1.85}$$

where Φ is a Markov kernel from $\mathbf{L}^\infty(\boldsymbol{\mathcal{X}}) \times \mathbf{L}^\infty(\boldsymbol{\mathcal{X}})$ to $\mathbf{L}^\infty(\boldsymbol{\mathcal{X}})$ defined as

$$\Phi : ((f, f'), F) \mapsto \delta_{(f \cdot f')^\dagger}(F). \tag{1.86}$$

This shows that the canonical version of $P \star P'$ is a proper Bayes' posterior probability measure on $\mathbf{L}^\infty(\boldsymbol{\mathcal{X}})$. This is a fundamental result since it shows that the fusion of information can be performed in a fully measure-theoretic Bayesian paradigm. Some of the implications of this result will be studied in the current and next sections.

1.3.2 *If one of the probabilistic constraints, say P', is equivalent to a probability measure $p' \in \mathbf{M}_1(\boldsymbol{\mathcal{X}})$, then we can write by extension*

$$P \star p' = \frac{(P * p')^\dagger}{\|P * p'\|}, \tag{1.87}$$

where the convolution of P with p' is defined as

$$(\forall F \in \mathcal{L}^\infty(\boldsymbol{\mathcal{X}})) \qquad P * p'(F) \doteq \int \mathbf{1}_F(f(x)\mathbf{1}_{\{x\}}) p'(\mathrm{d}x) P(\mathrm{d}f). \tag{1.88}$$

One immediate objection to Theorem 1.1 is that the assumption that the set $\boldsymbol{\mathcal{X}}$ is representative is highly restrictive. Very often there is no interest in handling sophisticated state spaces, especially when the received information does not allow for such an in depth estimation. In the next corollary, we show that the result of Theorem 1.1 can be extended to a simpler state space as long as it verifies the following assumption[10]:

A.1 There exists a representative set in which the physical system can be uniquely characterised.

Corollary 1.3. *Under Assumption **A.1**, the result of Theorem 1.1 holds for random variables in the state space $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$.*

Proof. Let ξ be a given projection map from $\boldsymbol{\mathcal{X}}$ to $\mathbf{X}$. The operations of pushforward and pullback can be respectively expressed via the following mappings on the space of measurable functions:

$$(\forall f \in \mathbf{L}^\infty(\boldsymbol{\mathcal{X}})) \qquad T_\xi(f) : x \mapsto \sup_{\xi^{-1}[\{x\}]} f, \tag{1.89}$$

and

$$(\forall f \in \mathbf{L}^\infty(\mathbf{X})) \qquad T'_\xi(f) = f \circ \xi. \tag{1.90}$$

We compute the result of the pusforward of fused pullback functions as follows:

$$(\forall f, f' \in \mathbf{L}^\infty(\mathbf{X})) \qquad T_\xi\Big(T'_\xi(f) \cdot T'_\xi(f')\Big) = T_\xi((f \cdot f') \circ \xi), \tag{1.91}$$

so that

$$T_\xi\Big(T'_\xi(f) \cdot T'_\xi(f')\Big) : x \mapsto \sup_{\xi^{-1}[\{x\}]} T_\xi((f \cdot f') \circ \xi) = f \cdot f'. \tag{1.92}$$

This result indicates that the pushforward and the pullback via ξ do not need to be considered when fusing probabilistic constraints on the non-representative state space $\mathbf{X}$ and that, as a result, the fusion of probabilistic constraints on $\mathbf{X}$ and on $\boldsymbol{\mathcal{X}}$ can be performed in the same way. □

The following example details two simple cases for which only one of the elements in the probabilistic constraint is maintained through fusion.

Example 1.8. Using the notations of Theorem 1.1:

[10] Assumptions with the prefix "A." will hold for the rest of the manuscript.

1.8.1 If P has its support in $\mathbf{I}_s(\mathbf{X})$ and is therefore equivalent to $p \in \mathbf{M}_1(\mathbf{X})$ and if P' is of the form $P' = \delta_{f'}$, then the probabilistic constraint $P \star P' \in \mathbf{C}_1(\mathbf{X})$ satisfies

$$P \star P'(F) \propto \int \mathbf{1}_F((\mathbf{1}_{\{x\}} \cdot f')^\dagger) \|\mathbf{1}_{\{x\}} \cdot f'\| p(\mathrm{d}x) \tag{1.93a}$$

$$\propto \int \mathbf{1}_F((f'(x)\mathbf{1}_{\{x\}})^\dagger) f'(x) p(\mathrm{d}x). \tag{1.93b}$$

Since $f'(x)\mathbf{1}_{\{x\}}$ is either the null function or is supported by a singleton, $P \star P'$ has its support in $\mathbf{I}_s(\mathbf{X})$ and is equivalent to the measure $\hat{p} \in \mathbf{M}_1(\mathbf{X})$ defined as

$$\hat{p}(\mathrm{d}x) \stackrel{\int}{=} \Psi_{f'}(p)(\mathrm{d}x) \doteq \frac{1}{p(f')} f'(x) p(\mathrm{d}x), \tag{1.94}$$

that is, the law of the fused random variable is known to be $\Psi_{f'}(p) \in \mathbf{M}_1(\mathbf{X})$. The same type of result holds if the roles of P and P' are interchanged. If f' is actually a likelihood of the form ℓ_z, then $\hat{p}$ can be expressed as

$$\hat{p}(\mathrm{d}x) \stackrel{\int}{=} \frac{\ell_z(x) p(\mathrm{d}x)}{\int \ell_z(x') p(\mathrm{d}x')}, \tag{1.95}$$

which is the usual Bayes' posterior of p given z, where z is interpreted as an observation and p is the prior probability measure. The fact that ℓ_z is an element of $\mathbf{L}(\mathbf{X})$ rather than a probability density does not affect the result because of the normalisation factor $\int \ell_z(x') p(\mathrm{d}x')$.

1.8.2 If P and P' have the form $P = \delta_f$ and $P' = \delta_{f'}$, then the probabilistic constraint $P \star P' \in \mathbf{C}_1(\mathbf{X})$ satisfies

$$P \star P'(F) = \frac{\mathbf{1}_F((f \cdot f')^\dagger) \|f \cdot f'\|}{\|f \cdot f'\|} = \delta_{(f \cdot f')^\dagger}(F) \tag{1.96}$$

for any $F \in \mathbf{L}^\infty(\mathbf{X})$, that is, the fused probabilistic constraint is $\delta_{(f \cdot f')^\dagger} \in \mathbf{C}_1^*(\mathbf{X})$ and $(f \cdot f')^\dagger \in \mathbf{L}(\mathbf{X})$ can be seen as a fused upper bound.

The binary operation $\star$ on $\mathbf{C}_1(\mathbf{X})$ introduced in Theorem 1.1 has some additional properties that are of practical interest when devising estimation algorithm.

Theorem 1.2. *The space* $(\mathbf{C}_1(\mathbf{X}), \star)$ *is a commutative monoid.*

Proof. Theorem 1.1 together with Corollary 1.3 show that $\star$ is a proper binary operation on $\mathbf{C}_1(\mathbf{X})$ in the sense that it is found to be a relation from $\mathbf{C}_1(\mathbf{X}) \times \mathbf{C}_1(\mathbf{X})$ to $\mathbf{C}_1(\mathbf{X})$. In order to prove the result of the theorem, we need to show that $\star$ is also associative, has an identity element in $\mathbf{C}_1(\mathbf{X})$ and is commutative: let P, P' and P'' be probabilistic constraints in $\mathbf{C}_1(\mathbf{X})$, then

a) $\star$ is associative: Since the convolution of measures $*$ is already known to be associative, we only need to show that $((P * P')^\dagger * P'')^\dagger = (P * (P' * P'')^\dagger)^\dagger$ holds. This can be done by verifying that

$$\|f \cdot f'\| \|\hat{f} \cdot f''\| = \|f \cdot f' \cdot f''\| \tag{1.97}$$

and

$$\frac{\hat{f} \cdot f''}{\|\hat{f} \cdot f''\|} = \frac{f \cdot f' \cdot f''}{\|f \cdot f' \cdot f''\|} \tag{1.98}$$

hold with $\hat{f} = \frac{f \cdot f'}{\|f \cdot f'\|}$ for any $f, f', f'' \in \mathbf{L}^\infty(\mathbf{X})$.

b) $\star$ has an identity element: consider that $P' = \delta_{\mathbf{1}}$, then for any $F \in \mathcal{L}^{\infty}(\mathbf{X})$

$$P \star P'(F) \propto \int \mathbf{1}_F((f \cdot \mathbf{1})^{\dagger}) \|f\| P(\mathrm{d}f) = P(F), \tag{1.99}$$

with the coefficient of proportionality

$$\int \|f \cdot \mathbf{1}\| P(\mathrm{d}f) = 1, \tag{1.100}$$

so that $P \star P' = P$ and P' is the identity element of $(\mathbf{C}_1(\mathbf{X}), \star)$.

c) $\star$ is commutative: this property of the binary operation $\star$ can be deduced directly from Theorem 1.1 where P and P' appear to be used in an exchangeable way in the characterisation of $P \star P'$.

This terminates the proof. □

The absence of inverse element for the operator $\star$ on $\mathbf{C}_1(\mathbf{X})$ is due to the fact that it is not always possible to "forget" what has been learnt.

Remark. Because of the properties of the fusion operator $\star$ on $\mathbf{C}_1(\mathbf{X})$, we can write $P \star P' \star P''$ without brackets when fusing the probabilistic constraints P, P' and P''.

We study the relation between the proposed fusion operator and Dempster's rule of combination in the next example.

Example 1.9. Let P and P' be two canonical probabilistic constraints on $\mathbf{X}$ of the form

$$P = \sum_{A \in \mathcal{A}} m(A) \delta_{\mathbf{1}_A} \qquad \text{and} \qquad P' = \sum_{A' \in \mathcal{A}'} m'(A') \delta_{\mathbf{1}_{A'}} \tag{1.101}$$

for some sets $\mathcal{A}$ and $\mathcal{A}'$ of measurable subsets and some functions $m : \mathcal{A} \to [0,1]$ and $m' : \mathcal{A}' \to [0,1]$. The posterior probabilistic constraint $P \star P'$ on $\mathbf{X}$ verifies

$$P \star P'(F) \propto \sum_{A \cap A' \neq \emptyset} \mathbf{1}_F(\mathbf{1}_{A \cap A'}) m(A) m'(A') \tag{1.102}$$

for any $F \in \mathcal{L}^{\infty}(\mathbf{X})$, which can be re-expressed as

$$P \star P' = \frac{1}{\|P * P'\|} \sum_{A \cap A' \neq \emptyset} m(A) m'(A') \delta_{\mathbf{1}_{A \cap A'}}, \tag{1.103}$$

with

$$\|P * P'\| = \sum_{A \cap A' \neq \emptyset} m(A) m'(A') = 1 - \sum_{A \cap A' = \emptyset} m(A) m(A'), \tag{1.104}$$

so that $P \star P'$ reduces to the result of Dempster's rule of combination between m and m' [Dempster, 1968, Shafer, 1986]. This rule of combination has been questioned many times, starting with Zadeh [1979], but we show here that it can be seen as a special case of Bayes' theorem for probabilistic constraints. A discussion about the connections between Dempster's rule of combination and Bayes' theorem applied to second-order probabilities can be found in Baron [1987].

Property 1.4. *The approximation procedure described in Property 1.2.3 can be justified using the fusion operation when functions in the support of the probabilistic constraint $P \in \mathbf{C}_1(\mathbf{X})$ have their support in a given bounded subset S of $\mathbf{X}$. In this case, the probability measure $p \in \mathbf{M}_1(\mathbf{X})$ which approximates P can be deduced from the fusion of P with the uniform probability measure p' on A since the probabilistic constraint characterised by*

$$(\forall F \in \mathcal{L}^{\infty}(\mathbf{X})) \qquad P \star p' \propto \int \mathbf{1}_F((f(x)\mathbf{1}_{\{x\}})^{\dagger}) f(x) \lambda(\mathrm{d}x) P(\mathrm{d}f) \tag{1.105}$$

is equivalent to the approximation p defined in Property 1.2.3.

1.4.3 General fusion

In Section 1.4.2, it has been assumed that the set $\mathbf{X}$ on which the probabilistic constraints are defined can be understood as a "state space", with the consequence that the system of interest should be represented at a single point of $\mathbf{X}$. The way elements of $\mathbf{L}^{\infty}(\mathbf{X})$ are fused is highly dependent on this assumption. Yet, a different viewpoint must be considered when introducing probabilistic constraints on more general sets such as $\mathbf{M}_1(\mathbf{X})$ and $\mathbf{C}_1(\mathbf{X})$.

Remark.* In order to have a suitable topology defined on $\mathbf{L}^0(\mathbf{C}_1(\mathbf{X}))$, a reference measure must be introduced. As there is no natural reference measure on $\mathbf{C}_1(\mathbf{X})$, a case-by-case reference measure must be defined, e.g., as an integer-valued measure that singles out a countable number of elements of $\mathbf{C}_1(\mathbf{X})$, or as a parametric family of elements of $\mathbf{C}_1(\mathbf{X})$ which parameter lies in an Euclidean space.

Let $\mathbf{Y}$ be a topological space equipped with its Borel σ-algebra, let $X \in \mathbf{L}^0(\Omega, \mathbf{Y})$ and $X' \in \mathbf{L}^0(\Omega', \mathbf{Y})$ be two random variables in $\mathbf{Y}$ with respective laws $\boldsymbol{p}$ and $\boldsymbol{p}'$ and consider the predicate

$$\text{“}X \text{ and } X' \text{ represent the same physical system”.} \tag{\#}$$

We assume there is a given way of combining elements of $\mathbf{Y}$ which is characterised by a likelihood function $\ell(\# \,|\, \cdot) : \mathbf{Y} \times \mathbf{Y} \to [0, 1]$ and a surjective mapping $\theta : S \to \mathbf{Y}$, with $S \subseteq \mathbf{Y} \times \mathbf{Y}$, which defines a stochastic kernel $\Phi \in \mathbf{K}(\mathbf{Y} \times \mathbf{Y}, \mathbf{Y})$ as

$$\Phi((y, y'), \cdot) \doteq \begin{cases} \delta_{\theta(y,y')} & \text{if } (y, y') \in S \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{1.106}$$

We also assume that $\ell(\# \,|\, \cdot)$ verifies $\ell(\# \,|\, y, y') = 0$ for any $(y, y') \in S^{\mathrm{c}}$. Note that the kernel Φ becomes a Markov kernel when restricted to S. The posterior law $\hat{\boldsymbol{p}}$ is characterised by

$$(\forall B \in \mathcal{B}(\mathbf{Y})) \qquad \hat{\boldsymbol{p}}(B) = \frac{\boldsymbol{p} \rtimes \boldsymbol{p}'(\ell(\# \,|\, \cdot)\Phi(\cdot, B))}{\boldsymbol{p} \rtimes \boldsymbol{p}'(\ell(\# \,|\, \cdot))}. \tag{1.107}$$

Henceforth, we will write ℓ rather than $\ell(\# \,|\, \cdot)$ for the sake of compactness.

Example 1.10. If $\mathbf{Y}$ is equal to $\mathbf{L}^{\infty}(\mathbf{X})$, with $\boldsymbol{\mathcal{X}}$ the state space defined in Section 1.4.2, then the laws $\boldsymbol{p}$ and $\boldsymbol{p}'$ can be seen as probabilistic constraints on $\boldsymbol{\mathcal{X}}$. In this case, the natural choice for ℓ and θ is $\ell(f, f') = \|f \cdot f'\|$ and $\theta(f, f') = (f \cdot f')^{\dagger}$ for any f and f' in $\mathbf{L}^{\infty}(\boldsymbol{\mathcal{X}})$. The posterior law $\hat{\boldsymbol{p}}$ then takes the form

$$(\forall F \in \mathcal{L}^{\infty}(\mathbf{Y})) \qquad \hat{\boldsymbol{p}}(F) \propto \int \mathbf{1}_F((f \cdot f')^{\dagger}) \|f \cdot f'\| \boldsymbol{p}(\mathrm{d}f)\boldsymbol{p}'(\mathrm{d}f'), \tag{1.108}$$

which is exactly the expression of $\boldsymbol{p} \star \boldsymbol{p}'$ when $\boldsymbol{p}$ and $\boldsymbol{p}'$ are seen as probabilistic constraints on $\mathbf{X}$.

We assume that there is a reference measure on $\mathbf{Y}$ which enables probability measures to be defined on $\mathbf{L}^\infty(\mathbf{Y})$. The objective is to extend the fusion of probabilistic constraints to the case where $\mathbf{Y}$ might not be a state space.

Theorem 1.3. *Let*[11] $\mathfrak{C} \in \mathbf{L}^0(\Omega, \mathbf{Y})$ *and* $\mathfrak{C}' \in \mathbf{L}^0(\Omega', \mathbf{Y})$ *be two random variables in* $\mathbf{Y}$ *with respective probabilistic constraints* $\boldsymbol{P}$ *and* $\boldsymbol{P}'$*. When defined, the posterior probabilistic constraint* $\boldsymbol{P} \star \boldsymbol{P}'$*, obtained by assuming that* $\mathfrak{C}$ *and* $\mathfrak{C}'$ *represent the same physical system, is characterised by the following convolution of measures on* $(\mathbf{L}^\infty(\mathbf{Y}), \odot)$*:*

$$\boldsymbol{P} \star \boldsymbol{P}' = \frac{(\boldsymbol{P} * \boldsymbol{P}')^\dagger}{\|\boldsymbol{P} * \boldsymbol{P}'\|}, \tag{1.109}$$

where the binary operation $\odot$ *on* $\mathbf{L}^\infty(\mathbf{Y})$ *is characterised by*

$$f \odot f' : \hat{y} \mapsto \sup_{(y,y') \in \theta^{-1}[\{\hat{y}\}]} \ell(y, y') f(y) f'(y'). \tag{1.110}$$

Proof. Let $\boldsymbol{p}$ and $\boldsymbol{p}'$ denote the respective laws of $\mathfrak{C}$ and $\mathfrak{C}'$. Equation (1.107) shows that the measure $\boldsymbol{m}$ to be constrained is characterised by

$$(\forall B \in \mathcal{B}(\mathbf{Y})) \qquad \boldsymbol{m}(B) = \boldsymbol{p} \rtimes \boldsymbol{p}'(\ell\, \Phi(\cdot, B)), \tag{1.111}$$

and we find that

$$\boldsymbol{m}(B) \leq \int \sup_{(y,y') \,:\, \theta(y,y') \in B} \Big(\ell(y, y') f(y) f'(y')\Big) \boldsymbol{P}(\mathrm{d}f) \boldsymbol{P}'(\mathrm{d}f') \tag{1.112}$$

for all $B \in \mathcal{B}(\mathbf{Y})$. In order to express the right hand side of the previous inequality as a probabilistic constraint, the argument of the supremum has to be normalised, and we find that

$$\boldsymbol{P} \star \boldsymbol{P}'(F) = \frac{1}{C} \int \mathbf{1}_F((f \odot f')^\dagger) \|f \odot f'\| \boldsymbol{P}(\mathrm{d}f) \boldsymbol{P}'(\mathrm{d}f') \propto (\boldsymbol{P} \star \boldsymbol{P}')^\dagger \tag{1.113}$$

for all $F \in \mathcal{L}^\infty(\mathbf{Y})$, where the operation $\odot$ is defined as in the statement of the theorem. The constant is determined by the possibility for $\mathfrak{C}$ and $\mathfrak{C}'$ to represent the same physical system which is expressed as

$$C = \int \|f \odot f'\| \boldsymbol{P}(\mathrm{d}f) \boldsymbol{P}'(\mathrm{d}f') = \|\boldsymbol{P} \star \boldsymbol{P}'\|, \tag{1.114}$$

thus proving the result of the theorem. □

Remark. For the posterior probabilistic constraint of Theorem 1.3 to be well defined, the mappings θ and ℓ have to have sufficient properties for $(\mathbf{L}^\infty(\mathbf{Y}), \odot)$ to be a semi-group. Informally, the mapping θ can be loosely seen as a binary operation which has to be associative for $\odot$ to be associative. More rigorously, we can define an extension $\bar{\mathbf{Y}} \doteq \mathbf{Y} \cup \{\varphi\}$ of the set $\mathbf{Y}$ by an isolated point φ and extend θ and ℓ as follows:

$$(\forall (y, y') \notin S) \;\; \theta(y, y') = \varphi \qquad \text{and} \qquad (\forall (y, y') \notin \mathbf{Y}) \;\; \ell(y, y') = 0. \tag{1.115}$$

With these notations, the binary operation $\odot$ is associative when θ is associative and when

$$\ell(y, y')\ell(\theta(y, y'), y'') = \ell(y, \theta(y', y''))\ell(y', y'') \tag{1.116}$$

holds for any $y, y', y'' \in \mathbf{Y}$.

[11] $\mathfrak{C}$ is the letter C with the Fraktur font

The fusion of probabilistic constraints on $\mathbf{Y}$ can be seen as a Bayesian operation, as demonstrated in the following property.

Property 1.5. *Using the notations of Theorem 1.3:*

1.5.1 If we define a likelihood function as $\boldsymbol{\ell}(f, f') \doteq \|f \odot f'\|$ *and a Markov kernel* $\boldsymbol{\Phi}((f, f'), \cdot) \doteq \delta_{(f \odot f')^\dagger}$ *for any measurable function* $f, f' \in \mathbf{L}^\infty(\mathbf{Y})$*, then the probability measure* $\boldsymbol{P} \star \boldsymbol{P}'$ *can be expressed as*

$$\boldsymbol{P} \star \boldsymbol{P}'(F) = \frac{\boldsymbol{P} \rtimes \boldsymbol{P}'(\boldsymbol{\ell}\,\boldsymbol{\Phi}(\cdot, F))}{\boldsymbol{P} \rtimes \boldsymbol{P}'(\boldsymbol{\ell})} \tag{1.117}$$

for any $F \in \mathcal{L}^\infty(\mathbf{Y})$*, thus showing the Bayesian nature of the binary operation* $\star$ *on* $\mathbf{C}_1(\mathbf{Y})$*.*

1.5.2 When the mapping θ *is bijective, the function* $f \odot f'$ *is also characterised by the relation*

$$(\forall (y, y') \in S) \qquad f \odot f'(\theta(y, y')) = \ell(y, y') f(y) f'(y'). \tag{1.118}$$

Indeed, the supremum in (1.110) *is taken over* $\theta^{-1}[\{\hat{y}\}]$ *which is a singleton when* θ *is bijective.*

In the following example, we make the connection between state-space fusion and the more general fusion operation introduced in Theorem 1.3.

Example 1.11. Using the notations of Theorem 1.3:

1.11.1 If $\mathbf{Y}$ is equal to the state space $\mathbf{X}$ defined in Section 1.4.2, then the natural way to set ℓ and θ is to take $\ell(x, x') = \mathbf{1}_{\{x\}}(x')$ for all $x, x' \in \mathbf{X}$ and to define θ on the diagonal of $\mathbf{X} \times \mathbf{X}$ only, by $\theta(x, x) = x$ for all $x \in \mathbf{X}$. Since θ is bijective in this case, we can use the result of Property 1.5.2 to find that

$$f \odot f' : x \mapsto \ell(x, x) f(x) f'(x) = f(x) f'(x). \tag{1.119}$$

In other words, it holds that $f \odot f' = f \cdot f'$ and the result of Section 1.4.2 about fusion for state spaces is recovered. This confirms that the fusion operation introduced in this section is not different from the one of Section 1.4.2, it is instead a more general formulation of the same operation.

1.11.2 If $\mathbf{Y} = \mathbf{C}_1(\mathbf{X})$ where $\mathbf{X}$ is the state space defined in Section 1.4.2, then the natural way to set ℓ and θ is $\ell(P, P') = \|P * P'\|$ and $\theta(P, P') = P \star P'$ for any probabilistic constraints P and P' in $\mathbf{C}_1(\mathbf{X})$. In this case, we find that

$$f \odot f' : \hat{P} \mapsto \sup_{(P, P') \,:\, P \star P' = \hat{P}} \|P * P'\| f(P) f'(P'). \tag{1.120}$$

The set $\mathbf{C}_1(\mathbf{X})$ is one of the useful examples of sets on which the fusion operation takes a general form since two probabilistic constraints do not have to be equal to represent the same individual. Also, the mapping θ is surjective but not bijective in general for probabilistic constraints, as opposed to the case of Example 1.11.1.

1.5 Markov constraint

The concept of Markov kernel is an extremely useful one in the context of estimation theory since it is likely that the physical system of interest will be subject to uncertain dynamics and will only be partially and randomly observed. In this section, we consider an additional set $\mathbf{Y}$ which is assumed to be a closed or open subset of $\mathbb{R}^{d'}$ for a given $d' \in \mathbb{N}^*$. A Markov kernel of interest is the *identity* Markov kernel $I \in \mathbf{K}_1(\mathbf{X}, \mathbf{Y})$, defined as

$$(\forall x \in \mathbf{X}) \qquad I(x, \cdot) \doteq \delta_x. \tag{1.121}$$

In particular, if any two measurable spaces can be related to each other via the identity kernel then they can be considered as *statistically equivalent* for the representation of the corresponding physical system. This notion of equivalence allows for handling the fusion of probabilistic constraints on different spaces by expanding each considered space to a sufficiently general product space.

Example 1.12. Let P be a probabilistic constraint on $\mathbf{X}$ and let P' be a probabilistic constraint on $\mathbf{X} \times \mathbf{Y}$, then P can be extended to $\mathbf{X} \times \mathbf{Y}$ by considering the pullback probabilistic constraint

$$\bar{P} \doteq (T'_\xi)_* P, \tag{1.122}$$

where ξ is the canonical projection from $\mathbf{X} \times \mathbf{Y}$ to $\mathbf{X}$, so that the fusion of P and P' can be expressed as $\bar{P} \star P'$, although we just write $P \star P'$ when there is no ambiguity.

This way of fusing probabilistic constraints that are expressed on different spaces will be useful when dealing with stochastic kernels and the probabilistic constraints defined for them as follows.

Definition 1.6. Let q and Q be stochastic kernels in $\mathbf{K}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{K}(\mathbf{X}, \mathbf{L}^0(\mathbf{Y}))$ respectively, if there exists a function F such that $Q(x, \cdot)$ is a measure constraint with characteristic function F for all $x \in \mathbf{X}$, then Q is said to be a *kernel constraint* for the kernel q with characteristic function F.

We are also interested in the case where q is a Markov kernel in which case Q is said to be a *Markov constraint*. The set of kernel constraints and of Markov constraints from $\mathbf{X}$ to $\mathbf{Y}$ are respectively denoted $\mathbf{C}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{C}_1(\mathbf{X}, \mathbf{Y})$. Rather than considering the two concepts of probabilistic and Markov constraint separately, we show in the next proposition that these two notions can be made consistent when F is the function $(B, f) \mapsto \|\mathbf{1}_B \cdot f\|$.

Proposition 1.4. *Let Q be a Markov constraint in $\mathbf{C}_1(\mathbf{X}, \mathbf{Y})$, then an equivalent probabilistic constraint $\bar{Q}$ can be defined on $\mathbf{X} \times \mathbf{Y}$ as follows: for every $x \in \mathbf{X}$, let S_x denote the section of $\mathbf{L}^\infty(\mathbf{X} \times \mathbf{Y})$ made of the functions of the form $\mathbf{1}_{\{x\}} \rtimes f$, for any $x \in \mathbf{X}$ and any $f \in \mathbf{L}^\infty(\mathbf{Y})$, let $S \subseteq \mathbf{L}^\infty(\mathbf{X} \times \mathbf{Y})$ be defined as*

$$S \doteq \bigcup_{x \in \mathbf{X}} S_x, \tag{1.123}$$

and let $T_x : \mathbf{L}^\infty(\mathbf{Y}) \to S_x$ be defined as $T_x : f \mapsto \mathbf{1}_{\{x\}} \rtimes f$, then $\bar{Q}$ is characterised by $\bar{Q}|_{S^c} = \mathbf{0}$ and

$$(\forall x \in \mathbf{X}) \qquad \bar{Q}|_{S_x} = (T_x)_* Q(x, \cdot). \tag{1.124}$$

Proof. The mapping T_x is an isomorphism between $\mathbf{L}^\infty(\mathbf{Y})$ and S_x as it is invertible and bi-measurable for any $x \in \mathbf{X}$. As a result, the measure $Q(x, \cdot)$ is equivalent to the restriction $\bar{Q}|_{S_x}$ of the measure $\bar{Q}$ to the section S_x for any $x \in \mathbf{X}$. From this point of view, the full Markov kernel Q can be seen as equivalent to the restriction $\bar{Q}|_S$, and since it holds that $\bar{Q}|_{S^c} = \mathbf{0}$, we obtain the equivalence between Q and $\bar{Q}$. □

Remark.* The result of Proposition 1.4 can be seen as a weaker form of isomorphism mod 0 [Itô, 1984], where the "isomorphism" is found to be between a probability space and a collection of probability spaces.

Following the case of Example 1.12, if P is a probabilistic constraint on $\mathbf{X}$ and if Q is a Markov constraint in $\mathbf{C}_1(\mathbf{X}, \mathbf{Y})$, then the fusion of P and Q can be expressed as $\bar{P} \star \bar{Q}$, but will be denoted $P \star Q$ when there is no ambiguity. In this case, it holds that $\|P * Q\| = 1$ since Q, as a Markov kernel, does not hold any information on $\mathbf{X}$.

Summary

As a conclusion, the concept of measure constraint has been shown to behave well with respect to the usual operations of measure theory and demonstrates a high level of versatility, which will be useful for representing any type of knowledge. Also, this flexibility enables the study of fused probabilistic constraints, which is of practical interest when it comes to estimation.

Chapter 2

Representation of stochastic populations

THE objective in this chapter is to find a way of representing individuals and populations with different characteristics on given state spaces. After some preliminaries in Section 2.1, a natural representation of stochastic populations is introduced in Section 2.2 and the fusion of two instances of this type of representation is studied in Section 2.3. Finally, alternative formulations are defined in Section 2.4 as a basis for Chapter 3. Except for the background material of Section 2.1, the concepts studied in this chapter as well as the methods and algorithms introduced in the subsequent chapters are novel.

Throughout this chapter, $(\Omega, \Sigma, \mathbb{P})$ will denote a complete probability space.

2.1 Background

The first section is indicated with a * in order to highlight that it is purely technical and is not required for the understanding of the following sections. The second section, however, contains a short review of the important properties of permutations and equivalence relations which will play a fundamental role in the subsequent developments.

2.1.1 *Set theory

The objective in this section is to highlight a few considerations that are specific to set theory. We consider the most common axiomatic system known as ZFC, for **Z**ermelo-**F**raenkel set theory with the axiom of **c**hoice. One of the nine axioms of ZFC, which is actually also found in most of the other axiomatic systems, is the axiom of extensionality, often formulated as

$$\forall A \forall B (\forall x (x \in A \Leftrightarrow x \in B) \Rightarrow A = B). \tag{2.1}$$

In other words, sets composed of the same elements are equal. Although being natural, this axiom also implies that sets are unable to directly represent multiplicity, i.e., it holds that $\forall x (\{x, x\} = \{x\})$. This will prove to be of a considerable importance when defining populations since handling a simple collection of individuals would be less convenient than a set of individuals.

We will also need to answer the question of whether or not a given axiomatic system is appropriate for representing collections of individuals as sets. For instance, quasi-set theory has been introduced to deal with the collections of indistinguishable objects arising in quantum physics, and is based on two types of *ur-elements*, i.e., objects that are not sets but may be elements of a set, one for quantum particles and one for macroscopic objects. Quasi-sets have been named and studied by da Costa [1980] and Krause [1992], based on questions raised by Manin [1976].

2.1.2 Permutation and equivalence relation

The concepts of permutation and equivalence relation will be useful to describe populations with various types of relations and symmetries. Permutations are generally studied in the context of group theory and equivalence relations are studied under the framework of lattice theory [Birkhoff, 1984].

A set G together with a binary operation is said to be a *group* if the binary operation is associative, if there exists an identity element, and if each element of G has an inverse element. Groups are special instances of monoids and semigroups. We denote $\mathbf{S}(A, B)$ the set of bijections from A to B and $\mathrm{Sym}(A) \doteq (\mathbf{S}(A, A), \circ)$ is referred to as the *symmetric group* on the set A. A *permutation group* is defined as a subgroup of a symmetric group [Dixon and Mortimer, 1996]. The mapping from $\mathrm{Sym}(A) \times A \to A$ characterised by $(\nu, x) \mapsto \nu(x)$ is said to be an *action* of $\mathrm{Sym}(A)$ on A since it holds that

a) $\mathbf{1}(x) = x$ for all $x \in A$,

b) $\sigma(\nu(x)) = (\sigma \circ \nu)(x)$ for all $x \in A$ and all $\nu, \sigma \in \mathrm{Sym}(A)$.

In the following definition and theorem, we consider the general case of a group G acting on a set A and we denote $g.x$ the action of $g \in G$ on $x \in A$.

Definition 2.1. Let G be a group acting on a set A and consider any $x \in A$, then the orbit of x under G is defined as

$$G.x \doteq \{g.x \text{ s.t. } g \in G\}. \tag{2.2}$$

Theorem 2.1 ([Dixon and Mortimer, 1996, Theorem. 1.4A])**.** *Let G be a group acting on a set A and consider any $x, y \in A$, then the two orbits $G.x$ and $G.y$ are either equal (as sets) or disjoint, so that the set of all orbits is a partition of A.*

For a given $\sigma \in \mathrm{Sym}(A)$, we also define the set $G \doteq \{\sigma^n \text{ s.t. } n \in \mathbb{Z}\}$, where σ^n is the n-fold composition of σ with itself, which forms a group when equipped with the binary operation $\circ$. The orbit of a given $x \in A$ under G verifies

$$G.x = \{\sigma^n(x) \text{ s.t. } n \in \mathbb{Z}\}, \tag{2.3}$$

and the partition of A defined by the set of orbits of G is referred to as the *orbits of* σ. We now consider the concept of equivalence relation. The properties of equivalence relations will prove to be closely related to the one stated in Theorem 2.1 for groups.

Definition 2.2. A binary relation $\sim$ on a set A is an equivalence relation if and only if, for any $x, y, z \in A$, *a*) (Reflexivity) $x \sim x$

b) (Symmetry) if $x \sim y$ then $y \sim x$

c) (Transitivity) if $x \sim y$ and $y \sim z$ then $x \sim z$.

Let $\sim$ be an equivalence relation on the set A and consider any $x \in A$, then the set $\{y \in A \text{ s.t. } x \sim y\}$ is called the *equivalence class* of x and is denoted $[x]$.

Theorem 2.2 ([Dummit and Foote, 2004, Proposition 2])**.** *Let $\sim$ be an equivalence relation on a set A and consider any $x, y \in A$, then the two equivalence classes $[x]$ and $[y]$ are either equal (as sets) or disjoint, so that the set of all equivalence classes is a partition of A. Also, every partition of A induces a unique equivalence relation on A.*

To indicate that a given set A is equipped with a structure $\sim$, we will write $(A, \sim)$. The set of equivalence classes on a set is called the *quotient set* and is denoted $A/\sim$. It appears from Theorems 2.1 and 2.2 that equivalence relations and groups share some important properties. This fact will be useful when studying relations between individuals of a given population.

Henceforth, equivalence relations will be denoted either by symbols like $\sim$ or $\approx$ or by a Greek letter, ρ, η or τ in particular. As a consequence of Theorem 2.2, for any equivalence relation ρ on a set A, we can define $\Pi(\rho)$ as the partition induced by ρ on A. Let $\mathbf{\Pi}(A)$ denote the set of equivalence relations on A and define a partial order on $\mathbf{\Pi}(A)$ as follows:

$$(\rho \leq \tau) \Leftrightarrow \forall x, y \in A(x\rho y \Rightarrow x\tau y), \tag{2.4}$$

i.e., the partition $\Pi(\rho)$ is a refinement of $\Pi(\tau)$. The least element of $\mathbf{\Pi}(A)$, denoted O, is the equality relation and the greatest element, denoted I, is the degenerate partition having A as its only equivalence class.

Finally, for any set A and any partition π of A, let $\mathrm{Sym}(A, \pi)$ be the subset of $\mathrm{Sym}(A)$ composed of any permutation σ verifying that the orbits of σ is a refinement of π.

Proposition 2.1. *For any set A and any partition π of A, the subset* $\mathrm{Sym}(A, \pi)$ *is a subgroup of* $\mathrm{Sym}(A)$.

For a given equivalence relation ρ on a set A, we write without ambiguity $\mathrm{Sym}(A, \rho)$ rather than $\mathrm{Sym}(A, \Pi(\rho))$. These concepts will be essential for describing the relation between the individuals of a population in the following sections.

2.2 Stochastic population

The objective in this section is to find a natural way of representing a stochastic population. In this context, "natural" means that the representation should handle general stochastic populations without introducing redundancies. The underlying motivation is that a natural representation should not only be useful in theory when expressing different results and properties, but also in practice when devising approximation algorithms for the induced probability laws.

Remark. An example class of such approximation algorithms is composed of the **M**arkov **c**hain **M**onte **C**arlo (MCMC) methods [Hastings, 1970, Andrieu et al., 2003]. Indeed, in this case, redundancies are undesirable since there is no interest in exploring parts of the state space that would essentially represent the same configuration. See [Oh et al., 2004, Jiang et al., 2014, Vu et al., 2014] for applications in the case of multi-object estimation.

2.2.1 Describing a population

As in Section 1.4, we consider a representative set $\boldsymbol{\mathcal{X}}_{\mathrm{a}}$ in which individuals can be uniquely characterised. Because of this characterisation, a population, which can be intuitively understood as a collection of individuals, is formally defined as a subset of $\boldsymbol{\mathcal{X}}_{\mathrm{a}}$. The set of all possible populations is then defined as $\boldsymbol{\mathcal{X}} \doteq \wp(\boldsymbol{\mathcal{X}}_{\mathrm{a}})$, where $\wp$ denotes the power set.

An important aspect is that in practice, a more realistic set $\mathbf{X}$ needs to be considered for the representation of individuals. This set is seen as being a projection of the set $\boldsymbol{\mathcal{X}}_{\mathrm{a}}$ and we define $\xi : \boldsymbol{\mathcal{X}}_{\mathrm{a}} \to \mathbf{X}$ as the associated projection map. Such a simplification is required for most of the applications as the estimation of all the characteristics of an individual is not usually considered possible. For instance, the observation might not account for the shape, mass or composition of a given solid, so that only its centre of mass/volume can be inferred. One of the consequences of this simplified representation is that individuals might have the same state in $\mathbf{X}$. In the context of point process theory [Daley and Vere-Jones, 2003], processes that never have two individuals at the same point are called *simple*. Borrowing this term, we can impose that representations should not require *simplicity* in $\mathbf{X}$ in general.

We assume that $\mathbf{X}$ can be written as the union of an Euclidean space $\mathbf{X}^{\bullet}$ and an *isolated point* ψ which can be viewed as an *empty state*. For a given population $\mathcal{X} \in \boldsymbol{\mathcal{X}}$, the subset $\xi^{-1}[\mathbf{X}^{\bullet}]$ of individuals with image in $\mathbf{X}^{\bullet}$ is denoted $\mathcal{X}^{\bullet}$ and the subset $\xi^{-1}[\{\psi\}]$ of individuals at ψ is denoted $\mathcal{X}^{\psi}$. In consequence, any population $\mathcal{X} \in \boldsymbol{\mathcal{X}}$ can be expressed as

$$\mathcal{X} = \mathcal{X}^{\bullet} \uplus \mathcal{X}^{\psi}, \tag{2.5}$$

where $\uplus$ refers to a disjoint union, i.e., it holds that $\mathcal{X}^{\bullet} \cap \mathcal{X}^{\psi} = \emptyset$.

Individuals in $\mathcal{X}^{\psi}$ cannot be distinguished whatever information is made available about them since the set $\mathbf{X}$ cannot hold any of this information. Even for individuals in $\mathcal{X}^{\bullet}$, the aptitude to obtain specific information, or *observability*, might not be sufficient to tell some of them apart. Individuals that are in this situation are said to be *strongly indistinguishable*, i.e., they cannot be distinguished in their current states even with the best possible observations. Strongly indistinguishable individuals can be related through an equivalence relation $\tau \in \mathbf{\Pi}(\mathcal{X})$ defined as follows: two individuals $x, x' \in \mathcal{X}$ are strongly indistinguishable if and only if $x \tau x'$ holds[1]. The set

$$\boldsymbol{\mathcal{Y}} \doteq \{(\mathcal{X}, \tau) \text{ s.t. } \mathcal{X} \in \boldsymbol{\mathcal{X}},\ \ \tau \in \mathbf{\Pi}(\mathcal{X})\} \tag{2.6}$$

is introduced in order to represent partially-indistinguishable populations. We consider the following naming convention:

$$\text{not(strongly indistinguishable)} \Leftrightarrow \text{(weakly distinguishable)}. \tag{2.7}$$

[1] The fact that this defines an equivalence relation is easy to verify.

Even when some individuals are weakly distinguishable, it could happen that the available information is not sufficient to tell them apart. We then say that these individuals are *weakly indistinguishable*. This concept clearly depends on the observation of the system and might evolve if additional information is made available. To sum up, strong indistinguishability is a state-dependent concept while weak indistinguishability is a probabilistic concept.

The description of the uncertainty on a given population $\mathcal{X} \in \boldsymbol{\mathcal{X}}$ can be performed by associating every individual in $\mathcal{X}$ with a random variable on $\mathbf{X}$. This solution, however, does not describe the relation between the different distributions related to different individuals, in particular with strongly indistinguishable ones. A global representation of uncertainty is thus sought. One of the most usual ways of describing multiple spatial entities as a whole is given by the theory of point processes. However, this theory is built on the following principle:

> "We talk of the probability of finding a given number k of points in a set A: we do not give names to the individual points and ask for the probability of finding k specified individuals within the set A. Nevertheless, this latter approach is quite possible (indeed, natural) in contexts where the points refer to individual particles, animals, plants and so on." [Daley and Vere-Jones, 2003, p. 124]

Yet, we wish to model the partially-indistinguishable nature of the individuals in $\mathcal{X}$ without assuming that they are all indistinguishable, i.e., without assuming that $\tau = I$. The study of populations composed of indistinguishable individuals is already challenging due to the difficulty in finding a consistent way of describing multiple individuals within a single stochastic object. Examples of questions arising from this issue are: Should the individuals be ordered even though there is no natural way of defining the order? Should the individuals be assumed to be represented at different points of the state space in order to enable a set representation? Should the population be assumed finite in order to proceed to the analysis? There are different ways of answering these questions and each way has to be proved equivalent in some sense to the others [Moyal, 1962, Macchi, 1975]. The representation of partially indistinguishable populations raises many additional and equivalently difficult questions. Alternative representations of stochastic populations have to be found in order to tackle this issue.

Example of a representative set

In order to illustrate the use of the representative set $\boldsymbol{\mathcal{X}}_{\mathrm{a}}$ introduced in the previous section, a practical example is detailed here based on physical considerations. For this purpose, we consider elementary matter-related particles that we call "atoms", following the ancient Greek "$\alpha\tau o\mu o\varsigma$" meaning "indivisible", even though these particles are named "fermions" in contemporary particle physics. This choice is explained by the fact that, except for particle physics related applications, the actual definition of elementary particles is unimportant, only the fact that everything can be described as a set of "atoms" is useful.

We consider a set $\mathbf{X}_{\mathrm{a}}$, called an *atomic state space*, on which every atom of an isolated system can be uniquely characterised through all its intrinsic and extrinsic properties such as the position, momentum, mass, spin, etc. The set $\mathbf{X}_{\mathrm{a}}$ is assumed to be a closed (or open) subset of an Euclidean space. As a consequence of the

Pauli exclusion principle, every atom shall occupy a unique state in $\mathbf{X}_{\mathrm{a}}$ and hence, collections of atoms can be represented by a set. Any collection of atoms could be considered to be finite since it is known that there would be a number of atoms in the observable universe of the order of 10^{80} if all atoms were hydrogen atoms[2]. However, the objective is not to determine whether the universe is finite or not, and we will only assume that collections of atoms are countable[3].

Generally, individuals can be naturally seen as collections of atoms. In other words, an individual is thought as being a piece of matter. As every atom occupies a unique state in the atomic state space $\mathbf{X}_{\mathrm{a}}$, a collection of atoms can be treated as a set, making all the operations of set theory available. An individual can then be seen as a subset of $\mathbf{X}_{\mathrm{a}}$ and the individual representative set is defined as $\boldsymbol{\mathcal{X}}_{\mathrm{a}} \doteq \wp(\mathbf{X}_{\mathrm{a}})$. It is natural to require different individuals to be disjoint subsets of $\mathbf{X}_{\mathrm{a}}$ and we assume that an atom can belong to one individual only. Individuals following this assumption can overlap in the physical space, as long as they are composed of different atoms, e.g., two mixed liquids can be seen as two individuals of a population. We conclude by noticing that the population representative set $\boldsymbol{\mathcal{X}}$ can be defined in this example as

$$\boldsymbol{\mathcal{X}} \doteq \wp(\boldsymbol{\mathcal{X}}_{\mathrm{a}}) = \wp(\wp(\mathbf{X}_{\mathrm{a}})), \tag{2.8}$$

that is, a population is a set of sets of atoms.

2.2.2 Representing a population

Based on the set $\boldsymbol{\mathcal{X}}$ of of all possible populations and on the set $\mathbf{X}$ on which all individuals are represented, we describe a versatile way of introducing randomness in the states of the individuals in $\mathbf{X}$ which conveys the concept of strong indistinguishability.

For a given population

Let $\mathcal{Y} = (\mathcal{X}, \tau) \in \boldsymbol{\mathcal{Y}}$ be a partially-distinguishable population of interest, i.e., a set $\mathcal{X}$ of individuals characterised in $\boldsymbol{\mathcal{X}}_{\mathrm{a}}$ that is equipped with an equivalence relation τ connecting strongly indistinguishable individuals. The objective is to include the relation between the individuals of $\mathcal{X}$ in the probabilistic modelling of the population. We first introduce the set

$$\mathbf{F}_{\mathcal{Y}}(\mathbf{X}) \doteq \Big\{ f : \mathcal{X} \to \mathbf{X} \text{ s.t. } |\mathcal{X}^{\bullet}| < \infty, \;\; (\forall x, x' \in \mathcal{X}) \;\; f(x) = f'(x) \Rightarrow x \tau x' \Big\} \tag{2.9}$$

that is composed of mappings $f : \mathcal{X} \to \mathbf{X}$ that have finitely many individuals in $\mathcal{X}^{\bullet} = f^{-1}[\mathbf{X}^{\bullet}]$ and such that all individuals that have the same image through f are strongly indistinguishable. The condition $|\mathcal{X}^{\bullet}| < \infty$ facilitates the definition of various types of operations on individuals but can be relaxed without inducing major changes in the following results. The set $\mathcal{X}$ is used as a way of indexing the states in $\mathbf{X}$ and the actual knowledge of the full individual characteristics $x \in \mathcal{X}$ is not used. Otherwise, the state of an individual $x \in \mathcal{X}$ could be directly obtained from the projection $\xi(x) \in \mathbf{X}$. At the end of this section, we will derive a formulation that ensures that $\boldsymbol{\mathcal{X}}$ cannot be used to hold information on the state of individuals.

[2]Because the mass of the observable universe is estimated to be 1.45×10^{53}kg [Davies, 2008], and the mass of a hydrogen atom is 1.67×10^{-27}kg

[3]This is a convenient assumption as it implies that we do not have to worry about running out of atoms.

A suitable σ-algebra of subsets of $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$, denoted $\mathcal{F}^*_{\mathcal{Y}}(\mathbf{X})$ is introduced in the following remark.

Remark.* There is a natural topology on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ that is generated by open sets of the same form as

$$A = \{f \text{ s.t. } (\forall x \in \mathcal{X}) \ \ f(x) \in A_x\}, \tag{2.10}$$

where A_x is an open set in $\mathbf{X}$ that differs from $\{\psi\}$ for finitely many $x \in \mathcal{X}$ only. Note that $\{\psi\}$ is indeed open as an isolated point. This topology is denoted $\mathcal{T}^*_{\mathcal{Y}}$ and the corresponding Borel σ-algebra is denoted $\mathcal{F}^*_{\mathcal{Y}}(\mathbf{X})$. Representations of the population $\mathcal{X}$ can thus be given as random variables in the measurable space of mappings $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}^*_{\mathcal{Y}}(\mathbf{X}))$.

A random variable[4] $\mathfrak{F}$ from (Ω, Σ) to $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}^*_{\mathcal{Y}}(\mathbf{X}))$ represents all the individuals in $\mathcal{X}$ on $\mathbf{X}$. We first have to verify that every single individual in $\mathcal{X}$ can be characterised separately. This would thus be equivalent to a collection of possibly correlated random variables, since indistinguishability has not been taken into account yet.

Proposition 2.2. *Any random variable from (Ω, Σ) to $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}^*_{\mathcal{Y}}(\mathbf{X}))$ characterises the individuals in $\mathcal{X}$.*

Proof. We have to prove that the random variable $\mathfrak{F}$ generates a random variable X_x on $\mathbf{X}$ for any individual $x \in \mathcal{X}$. Let $C_x \in \mathcal{F}^*_{\mathcal{Y}}(\mathbf{X})$ be the measurable subset defined as

$$C_x = \{f \in \mathbf{F}_{\mathcal{Y}}(\mathbf{X}) \text{ s.t. } f(x) \in B_x\} \tag{2.11}$$

for a given B_x in $\mathcal{B}(\mathbf{X})$. The mapping $X_x \doteq \mathfrak{F}(\cdot)(x)$ from Ω to $\mathbf{X}$ is measurable since it holds that

$$X_x^{-1}[B_x] = (\mathfrak{F}(\cdot)(x))^{-1}[C_x] \in \Sigma \tag{2.12}$$

for any $B_x \in \mathcal{B}(\mathbf{X})$, which terminates the proof. □

The equivalence relation τ on $\mathcal{X}$ connecting strongly indistinguishable individuals has not been taken into account yet. When two individuals in $\mathcal{X}$ are strongly indistinguishable, we expect that individual characterisations would not be available, even when considering a specific outcome $\omega \in \Omega$. Random variables on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ that do not respect this constraint would be mistakenly distinguishing individuals that are strongly indistinguishable. The set $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ is then not fully satisfying as is does not ensure that indistinguishable individuals are well represented.

A natural way of circumventing this incomplete representation of the structured population $\mathcal{Y}$ is to make the set $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ coarser by "gluing" together functions that distinguish indistinguishable individuals.

Example 2.1. Suppose that $\mathcal{Y} = (\{x, x'\}, I)$, i.e., $\mathcal{Y}$ is made of two indistinguishable individuals so that $\mathcal{X}/\tau = \{\{x, x'\}\}$. Additionally suppose that $\mathbf{X} = \mathbf{X}^\bullet = \{\mathbf{x}, \mathbf{x}'\}$, i.e., there are only two possible states for the individuals x and x', and assume that $\mathbf{X}$ is also representative so that x and x' must have different states in $\mathbf{X}$. There are only $2! = 2$ different distinguishable outcomes f, g in $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ defined by their respective graph as $\{(x, \mathbf{x}), (x', \mathbf{x}')\}$ and $\{(x, \mathbf{x}'), (x', \mathbf{x})\}$. To ensure that the individuals x, x' are indistinguishable, one can glue together these two symmetrical outcomes and define a new set of functions as $\{\{f, g\}\}$ (note the additional curly brackets). There is now only one outcome $\{f, g\}$ that does not allow for distinguishing the individuals x and x' as required.

[4]$\mathfrak{F}$ is the letter F with the Fraktur font

Following Example 2.1 and recalling the notation $\mathrm{Sym}(\mathcal{X}, \tau)$ introduced in Section 2.1.2 for permutations agreeing with the equivalence relation τ, we introduce a binary relation on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ as follows.

Definition 2.3. A binary relation ρ on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ is said to be induced by the equivalence relation τ if it holds that

$$(\forall f, f' \in \mathbf{F}_{\mathcal{Y}}(\mathbf{X})) \qquad f \rho f' \Leftrightarrow \exists \sigma \in \mathrm{Sym}(\mathcal{X}, \tau)(f = f' \circ \sigma). \tag{2.13}$$

Intuitively, elements of $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ are related through a binary relation whenever they only differ by a permutation of indistinguishable individuals. This binary relation can be proved to have additional properties.

Property 2.1. *The equivalence relation τ induces a unique binary relation on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$, and this binary relation is an equivalence relation.*

The proof of Property 2.1 relies mostly on the group nature of $\mathrm{Sym}(\mathcal{X}, \tau)$, as a subgroup of $\mathrm{Sym}(\mathcal{X})$ (Proposition 2.1). Consequently, only the specific group properties of $\mathrm{Sym}(\mathcal{X}, \tau)$ will be invoked when proving that the induced binary relation is an equivalence relation.

Proof. (*Uniqueness*) Let ρ and ρ' be two binary relations induced by τ. We want to prove that $f \rho f' \Leftrightarrow f \rho' f'$ holds for any $f, f' \in \mathbf{F}_{\mathcal{Y}}(\mathbf{X})$. Let σ, σ' be the two permutations in $\mathrm{Sym}(\mathcal{X}, \tau)$ satisfying (2.13) for ρ and ρ' respectively. There exists σ'' in $\mathrm{Sym}(\mathcal{X}, \tau)$ such that $\sigma \circ \sigma'' = \sigma'$, proving the uniqueness.
(*Reflexivity*) The identity is in $\mathrm{Sym}(\mathcal{X}, \tau)$.
(*Symmetry*) Existence of an inverse element in $\mathrm{Sym}(\mathcal{X}, \tau)$.
(*Transitivity*) Closure of $\mathrm{Sym}(\mathcal{X}, \tau)$. $\square$

Let ρ denote the unique equivalence relation on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ induced by τ and let ξ_ρ be the quotient map from $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ to $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})/\rho$ induced by ρ. In the next remark, we introduce a σ-algebra of subsets of $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$, denoted $\mathcal{F}_{\mathcal{Y}}(\mathbf{X})$, which does not allow for distinguishing strongly indistinguishable individuals.

Remark.* Let $\mathcal{T}_{\mathcal{Y}}$ denote the initial topology on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ induced by the quotient map ξ_ρ. We can verify that $\mathcal{T}_{\mathcal{Y}} \subseteq \mathcal{T}^*_{\mathcal{Y}}$ holds, meaning that there are fewer open subsets in $\mathcal{T}_{\mathcal{Y}}$ when compared to $\mathcal{T}^*_{\mathcal{Y}}$. The Borel σ-algebra induced by $\mathcal{T}_{\mathcal{Y}}$ is denoted $\mathcal{F}_{\mathcal{Y}}(\mathbf{X})$. A reference measure on $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}_{\mathcal{Y}}(\mathbf{X}))$ can be easily deduced from the reference measure on $\mathbf{X}$, e.g., the Lebesgue measure on $\mathbf{X}$. Using these reference measures, we can show that the set $\mathbf{L}^\infty(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$ is a Polish space, so that probabilistic constraints can be defined on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$.

Remark. Random variables on $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}_{\mathcal{Y}}(\mathbf{X}))$ characterise subsets of indistinguishable individuals rather than individuals themselves. Indeed, the approach used in the proof of Proposition 2.2 for random variables in $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}^*_{\mathcal{Y}}(\mathbf{X}))$ cannot be applied anymore since events in $\mathcal{F}_{\mathcal{Y}}(\mathbf{X})$ cannot be specific about indistinguishable individuals.

Now equipped with suitable spaces for considering the representation of partially-indistinguishable populations, we study the properties of probability measures and probabilistic constraints on $(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}), \mathcal{F}_{\mathcal{Y}}(\mathbf{X}))$. Since populations have an intrinsic multivariate nature, it is natural to introduce a notion of independence for probabilistic constraints on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ as in the following definition.

Definition 2.4. The individuals in $\mathcal{X}$ are said to be *independently represented* by the probabilistic constraint P on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ if there exists a family $\{P_x\}_{x\in\mathcal{X}}$ of probabilistic constraints on $\mathbf{X}$ such that

$$P(H) = \int \mathbf{1}_H\Big(f \mapsto \prod_{x\in\mathcal{X}} h_x(f(x))\Big) \prod_{x\in\mathcal{X}} P_x(\mathrm{d}h_x) \tag{2.14}$$

for any $H \in \mathcal{L}^\infty(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$.

To understand the meaning of Definition 2.4, it is useful to note that a function h in $\mathbf{L}^\infty(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$ for which there exists a family $\{h_x\}_{x\in\mathcal{X}}$ of functions in $\mathbf{L}^\infty(\mathbf{X})$ verifying

$$h : f \mapsto \prod_{x\in\mathcal{X}} h_x(f(x)), \tag{2.15}$$

could be interpreted as a *separable* function. This creates a connection with the concept of independently constrained random variables introduced in Definition 1.3. Equation (2.14) can be seen as a $|\mathcal{X}|$-fold convolution based on the family of probability measures $\{P_x\}_{x\in\mathcal{X}}$ and we assume that the mapping $\{h_x\}_{x\in\mathcal{X}} \mapsto h$ induced by (2.15) is continuous for P to be well defined. The concept of independently represented individual is a generalisation of the notion of independently constrained random variables introduced in Definition 1.3.

The notion of weak indistinguishability that was introduced in Section 2.2.1 has not been translated into practical terms yet. As opposed to strongly indistinguishable individuals that are bound through the events in $\mathcal{F}_{\mathcal{Y}}(\mathbf{X})$, it just happens that there is no specific knowledge about weakly indistinguishable individuals. As a result, weak indistinguishability is a fully probabilistic concept. In order to formally define it, we introduce a mapping T_σ from $\mathbf{L}^\infty(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$ into itself for any given $\sigma \in \mathrm{Sym}(\mathcal{X})$ characterised by

$$(\forall h \in \mathbf{L}^\infty(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))) \qquad T_\sigma(h) : f \mapsto h(f \circ \sigma). \tag{2.16}$$

Mappings of this form describe a change of upper bound when some individuals are swapped. It is therefore suitable for expressing properties of symmetry for probabilistic constraints as in the following definition.

Definition 2.5. Let $\mathfrak{F}$ be a random variable on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ with probabilistic constraint P. The relation of weak indistinguishability induced by P on $\mathcal{X}$ is defined as

$$\eta = \sup\Big\{\eta' \in \mathbf{\Pi}(\mathcal{X}) \text{ s.t. } (\forall \sigma \in \mathrm{Sym}(\mathcal{X}, \eta'))\ \ P = (T_\sigma)_* P\Big\}. \tag{2.17}$$

The relation of weak indistinguishability is an equivalence relation by definition. Some of its properties are listed below.

Property 2.2. *Using the notations of Definition 2.5, we find that:*

2.2.1 Since $\mathbf{\Pi}(\mathcal{X})$ is only a partially ordered set, the greatest element of a given subset might not exist, but it is necessarily unique if it exists. We can show that the relation η of weak indistinguishability exists by verifying that any element $\eta' \neq \eta$ in the considered subset can only identify less symmetries than η. In other words, there exist at least two subsets in $\Pi(\eta')$ which union is a subset of $\Pi(\eta)$ so that $\Pi(\eta') \leq \Pi(\eta)$ holds for any η' in the subset of $\mathbf{\Pi}(\mathcal{X})$ of interest.

2.2.2 *It holds that $\eta \geq \tau$.*

Proof. The elements of $\mathbf{L}^\infty(\mathbf{F}_\mathcal{Y}(\mathbf{X}))$ are bounded measurable functions from $(\mathbf{F}_\mathcal{Y}(\mathbf{X}), \mathcal{F}_\mathcal{Y}(\mathbf{X}))$ to $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$, and as such, they are not able to distinguish individuals related by τ. Thus, for any given $\sigma \in \mathrm{Sym}(\mathcal{X}, \tau)$, it holds that $f \circ \sigma = f$ for any $f \in \mathbf{F}_\mathcal{Y}(\mathbf{X})$ so that T_σ is the identity on $\mathbf{L}^\infty(\mathbf{F}_\mathcal{Y}(\mathbf{X}))$. As a result, the equivalence relation τ is always in the set of which η is the greatest element. □

2.2.3 *If the probabilistic constraint P is equivalent to a probability measure p on $\mathbf{X}$, then for any $\sigma \in \mathrm{Sym}(\mathcal{X}, \eta)$ and any $F \in \mathcal{F}_\mathcal{Y}(\mathbf{X})$, it holds that*

$$p(F) = p(\{f \circ \sigma \ \ s.t. \ \ f \in F\}), \tag{2.18}$$

which means that the probability measure p is invariant to permutations of states of weakly indistinguishable individuals.

2.2.4 *If the individuals in $\mathcal{X}$ are independently represented by P and if η is the relation of weak indistinguishability induced by P, then for any pair (x, x') of individuals in $\mathcal{X}$, it holds that*

$$(x \eta x') \Leftrightarrow (P_x = P_{x'}). \tag{2.19}$$

The representation of strongly indistinguishable individuals by random variables on $(\mathbf{F}_\mathcal{Y}(\mathbf{X}), \mathcal{F}_\mathcal{Y}(\mathbf{X}))$ can be considered as satisfactory. Yet, the true population $\mathcal{Y}$ was supposed to be known so far, even though it is only used as an indexing set, but this cannot be assumed in general. It is thus necessary to find a way of dealing with unknown populations.

Stochastic representation

In order to formally define the uncertainties for an unknown population, an appropriate source of randomness must be introduced, and we consider the complete probability space $(\Omega, \Sigma, \mathbb{P})$ for that purpose. The most natural way to extend the results of the previous section to unknown populations is to consider the union of the sets $\mathbf{F}_\mathcal{Y}(\mathbf{X})$ and to simplify it using an equivalence relation as previously. However, this would make all individuals weakly indistinguishable as there would be no way of assessing events based on specific individuals without a means of indexing the distinguished ones. Since weak indistinguishability is a probabilistic concept, it is meaningful to work directly on the set $\mathbf{C}_1(\mathbf{F}_\mathcal{Y}(\mathbf{X}))$, or more generally on

$$\mathbf{C}_\mathbf{F}(\mathbf{X}) \doteq \bigcup_{\mathcal{Y} \in \boldsymbol{\mathcal{Y}}} \mathbf{C}_1(\mathbf{F}_\mathcal{Y}(\mathbf{X})). \tag{2.20}$$

It is then possible to simplify the set $\mathbf{C}_\mathbf{F}(\mathbf{X})$ while preserving the relations of indistinguishability between individuals. For any $\mathcal{Y}$ and $\mathcal{Y}'$ in $\boldsymbol{\mathcal{Y}}$ and any bijection ν in $\mathbf{S}(\mathcal{Y}, \mathcal{Y}')$, we introduce the mapping $T_\nu : \mathbf{L}^\infty(\mathbf{F}_{\mathcal{Y}'}(\mathbf{X})) \to \mathbf{L}^\infty(\mathbf{F}_\mathcal{Y}(\mathbf{X}))$ characterised by

$$(\forall h \in \mathbf{L}^\infty(\mathbf{F}_{\mathcal{Y}'}(\mathbf{X}))) \qquad T_\nu(h) : f \mapsto h(f \circ \nu). \tag{2.21}$$

This mapping describes how an upper bound associated to the population $\mathcal{Y}$ can be transformed into an upper bound for the population $\mathcal{Y}'$, when a mapping between $\mathcal{Y}$ and $\mathcal{Y}'$ is given. The mapping defined in (2.16) can be seen as a special case when $\mathcal{Y} = \mathcal{Y}'$.

Definition 2.6. Let $\mathcal{Y}, \mathcal{Y}' \in \boldsymbol{\mathcal{Y}}$ be two populations equipped with a relation of strong indistinguishability defined via $\mathcal{Y} \doteq (\mathcal{X}, \tau)$ and $\mathcal{Y}' \doteq (\mathcal{X}', \tau')$. The binary relations $\sim$ on $\boldsymbol{\mathcal{X}}$ and $\approx$ on $\boldsymbol{\mathcal{Y}}$ are defined as follows

$$\mathcal{X} \sim \mathcal{X}' \Leftrightarrow \exists \nu : \mathcal{X} \leftrightarrow \mathcal{X}', \tag{2.22a}$$

$$\mathcal{Y} \approx \mathcal{Y}' \Leftrightarrow \exists \nu : (\mathcal{X}, \tau) \overset{\sim}{\longleftrightarrow} (\mathcal{X}', \tau'), \tag{2.22b}$$

where $\leftrightarrow$ indicates a bijection and where $\overset{\sim}{\longleftrightarrow}$ indicates a relation-preserving bijection. Also, for any $P \in \mathbf{C}_1(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$ and any $P' \in \mathbf{C}_1(\mathbf{F}_{\mathcal{Y}'}(\mathbf{X}))$, let the binary relation $\boldsymbol{\rho}$ on $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ be defined as

$$P \boldsymbol{\rho} P' \Leftrightarrow \exists \nu : (\mathcal{X}, \tau) \overset{\sim}{\longleftrightarrow} (\mathcal{X}', \tau')\Big(P = (T_\nu)_* P'\Big). \tag{2.23}$$

Some of the properties of the binary relations introduced in Definition 2.6 are listed below.

Property 2.3. *Using the notation of Definition 2.6, we find that:*

2.3.1 The binary relations $\sim$, $\approx$ and $\boldsymbol{\rho}$ on the respective sets $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}$ and $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ are equivalence relations.

2.3.2 The relation $\sim$ on $\boldsymbol{\mathcal{X}}$ can be equivalently defined as

$$\mathcal{X} \sim \mathcal{X}' \Leftrightarrow |\mathcal{X}| = |\mathcal{X}'|, \tag{2.24}$$

but this formulation makes less explicit the parallel with the definition of the equivalence relations $\approx$ and $\boldsymbol{\rho}$.

Remark.* Cantor [1915] actually extended the concept of cardinality to infinite sets by stating that two sets have the same cardinality if and only if there exists a bijection between them.

Since each probabilistic constraint P in $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ is defined on a single population in $\boldsymbol{\mathcal{Y}}$, the latter can be recovered and will be denoted $\mathcal{Y}_P$ or $(\mathcal{X}_P, \tau_P)$. If individuals are independently represented by a given probabilistic constraint $P \in \mathbf{C}_1(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$ then the equivalence class $[P]$ of probabilistic constraints related to P via $\boldsymbol{\rho}$ is found to be

$$[P] = \Big\{P' \in \mathbf{C}_{\mathbf{F}}(\mathbf{X}) \text{ s.t. } \exists \nu : \mathcal{Y} \overset{\sim}{\longleftrightarrow} \mathcal{Y}_{P'}\Big(\forall x \in \mathcal{Y}(P_x = P'_{\nu(x)})\Big)\Big\}. \tag{2.25}$$

This result highlights the structure of the equivalence relation $\boldsymbol{\rho}$ and of the mapping T_ν in (2.23). In particular, if the family $\{P_x\}_{x \in \mathcal{X}}$ of individual constraints can only have given values in $\mathcal{P} \doteq \{P_i \text{ s.t. } i \in \mathbb{I}\}$, then the equivalence relation $\boldsymbol{\rho}$ between probabilistic constraints in $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ can be translated into an equivalence relation $\sim$ between mappings in the set

$$\boldsymbol{\theta}_{\mathbb{I}} \doteq \bigcup_{(\mathcal{X},\tau) \in \boldsymbol{\mathcal{Y}}} \{\theta : (\mathcal{X}, \tau) \to \mathbb{I} \text{ s.t. } (\forall x, x' \in \mathcal{X}) \;\; x \tau x' \Rightarrow \theta(x) = \theta(x')\}, \tag{2.26}$$

defined by

$$\theta \sim \theta' \Leftrightarrow \exists \nu : \mathrm{dom}(\theta) \overset{\sim}{\longleftrightarrow} \mathrm{dom}(\theta')(\theta = \theta' \circ \nu). \tag{2.27}$$

A given equivalence class in $\mathbf{C}_{\mathbf{F}}(\mathbf{X})/\boldsymbol{\rho}$ allows for describing the randomness of a population of a given size and structure without knowing the actual population state in

$\boldsymbol{\mathcal{X}}$ as required. Such an equivalence class is referred to as a *population representation* or simply as a *representation*, and when individuals are independently represented, the induced probabilistic constraints on $\mathbf{X}$ are called *individual representations*.

However, the size and structure of a population are generally unknown and possibly random, and there might be second-order uncertainties on the probabilistic constraints in $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ themselves. A case where these second-order uncertainties arise is found when the fusion of probabilistic constraints require some type of *data association*, i.e., when there is uncertainty on which individual in one representation corresponds to which individual in another representation. Such a case will be studied in the following sections and chapters as it underlies the estimation of stochastic populations. In the following remark, a suitable σ-algebra of subsets of $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ is introduced and denoted $\mathcal{C}_{\mathbf{F}}(\mathbf{X})$.

Remark.* Even when a topology on $\boldsymbol{\mathcal{Y}}$ is available, the corresponding topology on $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ would not be suitable for our purpose since it would allow for distinguishing representations based on a given population $\mathcal{X} \in \boldsymbol{\mathcal{X}}$. Instead, we consider the initial topology induced by the quotient map of $\boldsymbol{\rho}$ and we denote $\mathcal{C}_{\mathbf{F}}(\mathbf{X})$ the corresponding Borel σ-algebra. There is no natural reference measure on $\mathcal{C}_{\mathbf{F}}(\mathbf{X})$, but we assume that such a measure is given case by case via a countable subset or a parametric family of probabilistic constraints. We additionally assume that probabilistic constraints can be defined on $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$. Similarly, the σ-algebras on $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\theta}_{\mathbb{I}}$ are assumed to be induced by the natural topology on $\boldsymbol{\mathcal{Y}}/\approx$ and $\boldsymbol{\theta}_{\mathbb{I}}/\sim$ respectively.

A measurable mapping $\mathfrak{C}$ from (Ω, Σ) to $(\mathbf{C}_{\mathbf{F}}(\mathbf{X}), \mathcal{C}_{\mathbf{F}}(\mathbf{X}))$ describes all the uncertainties about the system of interest and is referred to as a *stochastic representation*. The interpretation of $\mathfrak{C}$ can be made easier by separating its law $\boldsymbol{p}_{\mathfrak{C}} = \mathfrak{C}_*\mathbb{P}$ into two parts:

$$(\forall C \in \mathcal{C}_{\mathbf{F}}(\mathbf{X})) \qquad \boldsymbol{p}_{\mathfrak{C}}(C) = p_{\mathfrak{Y}}(\boldsymbol{p}_{\mathfrak{C}\,|\,\mathfrak{Y}}(C\,|\,\cdot)), \tag{2.28}$$

where $p_{\mathfrak{Y}}$ is the law of the random population[5] $\mathfrak{Y}$ induced by $\mathfrak{C}$ on $\boldsymbol{\mathcal{Y}}$, and $\boldsymbol{p}_{\mathfrak{C}\,|\,\mathfrak{Y}}$ is a version of the conditional law of $\mathfrak{C}$ given $\mathfrak{Y}$. This separation of the randomness is straightforward but helps to interpret the behaviour of $\mathfrak{C}$: first a size and a structure $\mathcal{Y} \in \boldsymbol{\mathcal{Y}}$ are randomly selected for the population, then a probabilistic constraint on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ is drawn.

To make use of stochastic representations, we have to understand what kind of operations can be performed on them, and the objective in the next section is to study the operation of fusion for stochastic representations.

2.3 Fusion

As a basis for the fusion of stochastic representations, we first study the case where the population of interest is given in $\boldsymbol{\mathcal{X}}$. Understanding the properties of the fusion operation in this idealistic case will be useful to tackle the general one.

2.3.1 Fusion for a given population

Let $\mathcal{Y} \in \boldsymbol{\mathcal{Y}}$ and let $\mathfrak{F}$ and $\mathfrak{F}'$ be two random variables on $\mathbf{F}_{\mathcal{Y}}(\mathbf{X})$ with respective probabilistic constraints P and P'. Also, denote η and η' the relations of weak indistinguishability induced by P and P' respectively. Using Theorem 1.1, a state-space

[5] The letter $\mathfrak{Y}$ is the letter Y with the Fraktur font

fusion can be performed between P and P', and the fused probabilistic constraint $P \star P'$ verifies

$$P \star P'(H) \propto (P * P')^{\dagger} = \int \mathbf{1}_H((h \cdot h')^{\dagger}) \| h \cdot h' \| P(\mathrm{d}h) P'(\mathrm{d}h') \tag{2.29}$$

for any $H \in \mathcal{L}^{\infty}(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$. Note that it is not possible to fuse probabilistic constraints that are based on different structured populations in $\boldsymbol{\mathcal{Y}}$. Indeed, not only should the two constraints be based on the same set of individuals, but these individuals should also have the same relation of strong indistinguishability.

Remark. In the case of individuals being made of atoms, it would be tempting to fuse populations that are not equal but which are composed of the same atoms overall. However, this is not the role of the fusion operation and an additional Markov kernel would need to be included to describe the transition between these two populations.

The relations η and η', however, do not have to be equal since they only characterise a property of the probabilistic constraints P and P'. A natural question is: what is the relation of weak indistinguishability induced by $P \star P'$? This question is answered in the following proposition.

Proposition 2.3. *The relation of weak indistinguishability induced by $P \star P'$ is larger than the greatest lower bound of η and η' in $\mathbf{\Pi}(\mathcal{X})$ which is denoted $\eta \wedge \eta'$.*

Proof. Let σ be a permutation in $\mathrm{Sym}(\mathcal{X})$. We want to determine if, for any function $\hat{h}$ in the support of $P \star P'$, it holds that

$$(\forall f \in \mathbf{F}_{\mathcal{Y}}(\mathbf{X})) \qquad \hat{h}(f) = \hat{h}(f \circ \sigma). \tag{2.30}$$

Let $h, h' \in \mathbf{L}^{\infty}(\mathbf{F}_{\mathcal{Y}}(\mathbf{X}))$ be such that $(h \cdot h')^{\dagger} = \hat{h}$, then the previous equation can be expressed as

$$h \cdot h'(f) = h \cdot h'(f \circ \sigma) \tag{2.31a}$$

$$= h(f \circ \sigma) h'(f \circ \sigma) \tag{2.31b}$$

for any $f \in \mathbf{F}_{\mathcal{Y}}(\mathbf{X})$. This equation holds when σ is in $\mathrm{Sym}(\mathcal{X}, \eta) \cap \mathrm{Sym}(\mathcal{X}, \eta')$. In other words, σ must be in $\mathrm{Sym}(\mathcal{X}, \eta \wedge \eta')$, which terminates the proof. ☐

To understand the meaning of $\eta \wedge \eta'$, it is useful to note that $\Pi(\eta \wedge \eta')$ is the intersection between the partitions $\Pi(\eta)$ and $\Pi(\eta')$. In other words, two individuals are related via $\eta \wedge \eta'$ if they were related via both η and η'. Overall, the result of Proposition 2.3 indicates that weak indistinguishability tends to be reduced when probabilistic constraints are fused together, as opposed to strong indistinguishability which only evolves when the population evolves.

Example 2.2. Let $\{B_i\}_{i \in \mathbb{I}}$ and $\{B'_i\}_{i \in \mathbb{I}'}$ be two families of Borel subsets of $\mathbf{X}$, let θ and θ' be mappings from a weakly distinguishable population $(\mathcal{X}, O)$ to the index sets $\mathbb{I}$ and $\mathbb{I}'$ respectively, and let $P = \delta_h$ and $P' = \delta_{h'}$ with

$$h : f \mapsto \prod_{x \in \mathcal{X}} \mathbf{1}\lfloor B_{\theta(x)} \rfloor (f(x)) \qquad \text{and} \qquad h' : f \mapsto \prod_{x \in \mathcal{X}} \mathbf{1}\lfloor B'_{\theta'(x)} \rfloor (f(x)), \tag{2.32}$$

where $\mathbf{1}\lfloor B \rfloor$ is a notation for $\mathbf{1}_B$. The form of h and h' shows that individuals are independently represented by the probabilistic constraints P and P'. The relations of weak indistinguishability η and η', induced by P and P' respectively, verify

$$\Pi(\eta) = \{\theta^{-1}[\{i\}] \text{ s.t. } i \in \mathbb{I}\} \qquad \text{and} \qquad \Pi(\eta') = \{\theta'^{-1}[\{i\}] \text{ s.t. } i \in \mathbb{I}'\}. \tag{2.33}$$

For the sake of simplicity, we assume that $\|h \cdot h'\| = 1$ holds, which means that there is an intersection between $B_{\theta(x)}$ and $B'_{\theta'(x)}$ for all $x \in \mathcal{X}$. The fused probabilistic constraint is found to be $P \star P' = \delta_{(h \cdot h')^\dagger}$, where

$$h \cdot h' : f \mapsto \prod_{x \in \mathcal{X}} \mathbf{1}\lfloor \hat{B}_{\theta \ltimes \theta'(x)} \rfloor (f(x)), \tag{2.34}$$

with $\hat{B}_{(i,i')} = B_i \cap B'_{i'}$ for any $(i, i') \in \mathrm{Im}(\theta \ltimes \theta') \subseteq \mathbb{I} \times \mathbb{I}'$. We can verify that this probabilistic constraint induces the relation $\eta \wedge \eta'$ through the following equality

$$\Pi(\eta \wedge \eta') = \{(\theta \ltimes \theta')^{-1}[\{(i, i')\}] \text{ s.t. } (i, i') \in \mathrm{Im}(\theta \ltimes \theta')\}. \tag{2.35}$$

In practice, the exact composition of the population is not known, and stochastic representations have to be used in order to represent and fuse the information made available by different sources.

2.3.2 Fusion of stochastic representations

Let $\mathfrak{C}$ and $\mathfrak{C}'$ be two stochastic representations on $\mathbf{C_F}(\mathbf{X})$ with probabilistic constraints $\boldsymbol{P}$ and $\boldsymbol{P}'$. Because of the probabilistic nature of the space on which $\mathfrak{C}$ and $\mathfrak{C}'$ are defined, second-order fusion must be used between $\boldsymbol{P}$ and $\boldsymbol{P}'$. From Theorem 1.3, we deduce that

$$\boldsymbol{P} \star \boldsymbol{P}'(H) \propto \int \mathbf{1}_H((h \odot h')^\dagger) \|h \odot h'\| \boldsymbol{P}(\mathrm{d}h) \boldsymbol{P}'(\mathrm{d}h') \tag{2.36}$$

for any $H \in \mathcal{L}^\infty(\mathbf{C_F}(\mathbf{X}))$, where the binary operations $\star$ and $\odot$ are defined as in (1.109) and (1.110). The properties of the fusion operator $\star$ for probabilistic constraints on $\mathbf{C_F}(\mathbf{X})$ can be illustrated for simple cases as in the following examples.

Example 2.3. Assuming that all individuals are independently represented, i.e., that the considered upper bounds are supported by probabilistic constraints in $\mathbf{C_F}(\mathbf{X})$ which independently represent individuals, consider the following cases:

2.3.1 The probabilistic constraints $\boldsymbol{P}$ and $\boldsymbol{P}'$ are such that $\boldsymbol{P} = \delta_h$ and $\boldsymbol{P}' = \delta_{h'}$, with h and h' in $\mathbf{L}(\mathbf{C_F}(\mathbf{X}))$ such that

$$h : P \mapsto \prod_{x \in \mathcal{X}_P} \mathbf{1}_{\delta_f}(P_x) \qquad \text{and} \qquad h' : P \mapsto \prod_{x \in \mathcal{X}_P} \mathbf{1}_{\delta_{f'}}(P_x), \tag{2.37}$$

that is, there is no information on the size or the structure of the population and the only individual probabilistic constraints available are δ_f and $\delta_{f'}$. The fusion of h and h' is characterised by the measurable function $h \odot h'$ in $\mathbf{L}^\infty(\mathbf{C_F}(\mathbf{X}))$ which is defined as

$$h \odot h' : P \mapsto \prod_{x \in \mathcal{X}_P} \mathbf{1}\lfloor \delta_{(f \cdot f')^\dagger} \rfloor (P_x), \tag{2.38}$$

i.e., the only individual probabilistic constraint available is $\delta_{(f \cdot f')^\dagger}$, and it holds that

$$\|h \odot h'\| = \|f \cdot f'\|. \tag{2.39}$$

The obtained fused probabilistic constraint is $\boldsymbol{P} \star \boldsymbol{P}' = \delta_{(h \odot h')^\dagger}$. It appears that there is only one possible fused upper bound, regardless of the number of individuals in the underlying population, and that this upper bound allows for one representation only.

2.3.2 Let $\{f_i\}_{i\in\mathbb{I}}$ and $\{f'_i\}_{i\in\mathbb{I}'}$ be two indexed families of $n < \infty$ different functions in $\mathbf{L}^\infty(\mathbf{X})$, let θ and θ' be any two bijections in $\boldsymbol{\theta}_{\mathbb{I}}$ and $\boldsymbol{\theta}_{\mathbb{I}'}$ and let $\mathcal{Y} \in \boldsymbol{\mathcal{Y}}$ be any population of n distinguishable individuals. We assume that the probabilistic constraints $\boldsymbol{P}$ and $\boldsymbol{P}'$ are such that $\boldsymbol{P} = \delta_h$ and $\boldsymbol{P}' = \delta_{h'}$, with h in $\mathbf{L}(\mathbf{C_F}(\mathbf{X}))$ of the form

$$h : P \mapsto \mathbf{1}_{[\mathcal{Y}]}(\mathcal{Y}_P) \prod_{x\in\mathcal{X}_P} h\lfloor\theta_P(x)\rfloor(P_x), \tag{2.40}$$

where $h_i = \mathbf{1}\lfloor\delta_{f_i}\rfloor$ and θ_P is the element of $[\theta]$ verifying $\mathrm{dom}(\theta) = \mathcal{Y}_P$. The function h' is of a similar form but based on $\{f'_i\}_{i\in\mathbb{I}'}$ and θ'. In this case, the only information encoded in h (resp. h') is that the size of the population is equal to n and that the individual laws have one of the f_i's (resp. f'_i) as an upper bound. Assuming that $f_i \cdot f'_{i'} \neq f_j \cdot f'_{j'}$ whenever $i \neq j$ or $i' \neq j'$, the fusion of h and h' is characterised by the measurable function $h \odot h'$ in $\mathbf{L}^\infty(\mathbf{C_F}(\mathbf{X}))$ defined as

$$h \odot h' : P \mapsto \mathbf{1}_{[\mathcal{Y}]}(\mathcal{Y}_P) \sup_{\nu\in\mathbf{S}(\mathbb{I},\mathbb{I}')} C_\nu \prod_{x\in\mathcal{X}_P} \hat{h}\lfloor\theta_P^{(\nu)}(x)\rfloor(P_x), \tag{2.41}$$

where $\hat{h}_{(i,j)} = \mathbf{1}\lfloor\delta\lfloor(f_i \cdot f'_j)^\dagger\rfloor\rfloor$ for all $(i,j) \in \mathbb{I}\times\mathbb{I}'$, where $\theta^{(\nu)}$ is any bijection in $\boldsymbol{\theta}_{\mathrm{Gr}(\nu)}$, and where

$$C_\nu = \prod_{(i,j)\in\mathrm{Gr}(\nu)} \|f_i \cdot f'_j\|. \tag{2.42}$$

This means that only the size of the population is known and the individual probabilistic constraints available correspond to the different bijections between elements of $\{f_i\}_{i\in\mathbb{I}}$ and of $\{f'_i\}_{i\in\mathbb{I}'}$. The compatibility of h and h' is found to be

$$\|h \odot h'\| = \sup_{\nu\in\mathbf{S}(\mathbb{I},\mathbb{I}')} C_\nu. \tag{2.43}$$

The obtained probabilistic constraint is $\boldsymbol{P} \star \boldsymbol{P}' = \delta_{(h\odot h')^\dagger}$ since there is only one possible fused function $h \odot h'$ as previously. The number of different representations in this probabilistic constraint is $|\mathbf{S}(\mathbb{I},\mathbb{I}')| = n!$.

The two cases detailed in Example 2.3 show that the number of possible *a posteriori* representations can range from 1 to $n!$, where n is the number of individuals in each population. This aspect will be made more specific in Proposition 2.4 under additional assumptions.

2.4 Alternative formulations

The objective is now to show that the problem can be formulated on more standard sets than $\mathbf{C_F}(\mathbf{X})$. Two alternative formulations are studied, the first relies on integer-valued measures whereas the second is based on the definition of suitable product spaces. These two types of formulations already exist for point processes as described by Moyal [1962] and Itô [2013].

2.4.1 Assumptions

In order to simplify the statement of these two formulations, the following assumption will be considered:

A.2 Individuals are independently represented.

As mentioned before, this assumption does not imply that individuals are independent in general, but rather that the possible correlations between the individuals are not quantified. The subset of $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$ composed of probabilistic constraints for which all individuals are independently represented is denoted $\mathbf{C}^*_{\mathbf{F}}(\mathbf{X})$. We also formulate an assumption that will be of interest when devising practical estimation algorithms:

A.3 Stochastic populations take values in a finite subset of $\mathbf{C}^*_{\mathbf{F}}(\mathbf{X})/\boldsymbol{\rho}$.

A related assumption can be formulated for any two stochastic representations verifying Assumption **A.3** with the subsets $\mathcal{P}$ and $\mathcal{P}'$:

A.4 For any $P_1, P_2 \in \mathcal{P}$ and any $P'_1, P'_2 \in \mathcal{P}'$ such that $P_1 \neq P_2$ or $P'_1 \neq P'_2$, it holds that

$$P_1 \star P'_1 \neq P_2 \star P'_2. \tag{2.44}$$

Under these assumptions, the number of posterior representations can be determined from the relation of weak indistinguishability in the prior representations as in the following proposition. In order to state this result, we assume that elements of $\boldsymbol{\mathcal{Y}}/\approx$ can be understood as abstract sets, i.e., as sets with a given size and structure, but with unknown elements.

Proposition 2.4. *Let $(\mathcal{X}, \eta)$ and $(\mathcal{X}', \eta')$ be abstract sets in $\boldsymbol{\mathcal{Y}}/\approx$ and let the equivalence relation $\sim$ be defined, for any $\nu, \nu' \in \mathbf{S}(\mathcal{X}, \mathcal{X}')$, as*

$$\nu \sim \nu' \Leftrightarrow \exists(\sigma, \sigma') \in \mathrm{Sym}(\mathcal{X}, \eta) \times \mathrm{Sym}(\mathcal{X}', \eta')\Big(\nu = \sigma' \circ \nu' \circ \sigma\Big), \tag{2.45}$$

then the possible associations between $\mathcal{X}$ and $\mathcal{X}'$ are characterised by the equivalence classes in $\mathcal{S} \doteq \mathbf{S}(\mathcal{X}, \mathcal{X}')/\sim$.

Proposition 2.4 is consistent with Examples 2.3.1 and 2.3.2 since we can verify that $\sim$ is equal to I when all the individuals are indistinguishable, so that $|\mathcal{S}| = 1$, and $\sim$ is equal to O when they are all distinguishable, so that $|\mathcal{S}| = n!$ with n the number of individuals in $\mathcal{X}$ and $\mathcal{X}'$. In order to describe the possible associations, one could take an arbitrary bijection in each of the equivalence classes in $\mathcal{S}$. However, it is more natural to build a binary relation Θ_S on $\mathcal{X} \times \mathcal{X}'$ for every $S \in \mathcal{S}$, which relates individuals that have been identified in a given association:

$$(\forall (x, x') \in \mathcal{X} \times \mathcal{X}') \qquad x\Theta_S x' \Leftrightarrow \exists \nu \in S[(x, x') \in \mathrm{Gr}(\nu)]. \tag{2.46}$$

To go further with the analysis of the structure of a given association, we can verify that the sets

$$\pi = \Big\{\Theta_S^{-1}[\{x'\}] \text{ s.t. } x' \in \mathcal{X}'\Big\} \qquad \text{and} \qquad \pi' = \Big\{\Theta_S[\{x\}] \text{ s.t. } x \in \mathcal{X}\Big\} \tag{2.47}$$

form partitions of $\mathcal{X}$ and $\mathcal{X}'$ respectively, and that these partitions verify $\pi \leq \Pi(\eta)$ and $\pi' \leq \Pi(\eta')$. As expected from the result of Proposition 2.3, an association reduces the indistinguishability by identifying different indistinguishable individuals in $\mathcal{X}$ with distinguishable individuals in $\mathcal{X}'$ and conversely. To reduce the level of abstraction, a mapping $a_S : \pi \to \pi'$ can be defined through

$$(\forall (X, X') \in \pi \times \pi') \qquad (X, X') \in \mathrm{Gr}(a_S) \Leftrightarrow \Theta_S[X] = X'. \tag{2.48}$$

Overall, an *association* a_S can be defined for every $S \in \mathcal{S}$ and identifies subsets of indistinguishable sub-populations between $\mathcal{X}$ and $\mathcal{X}'$.

2.4.2 As an integer-valued measure

One of the most direct alternative formulations uses the concept of integer-valued measures or counting measures.

Remark.* From the results of Appendix A.2, it appears that $\mathbf{C}_1(\mathbf{X})$ is a Polish space so that $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$ is also Polish. The natural σ-algebra of subsets of $\mathbf{C}_1(\mathbf{X})$ is then the Borel σ-algebra.

A connection between the specific notion of population representation and the more common concept of counting measure is established in the following proposition.

Proposition 2.5. *The mapping* $\zeta : \mathbf{C}^*_{\mathbf{F}}(\mathbf{X}) \to \mathbf{N}(\mathbf{C}_1(\mathbf{X}))$*, defined as*

$$\zeta : P \mapsto \sum_{x \in \mathcal{X}_P} \delta_{P_x}, \tag{2.49}$$

is measurable.

Proposition 2.5 shows that laws of stochastic representations on $\mathbf{C}^*_{\mathbf{F}}(\mathbf{X})$ can be pushforwarded onto the set of counting measures on $\mathbf{C}_1(\mathbf{X})$. The transformation ζ introduced in this proposition does not preserve the representation of strong indistinguishability and is not bi-measurable as a consequence. This can be seen as beneficial in practice since the observability of strong indistinguishability is often out of reach. The only individuals that are known to be strongly indistinguishable in this case are the ones that are almost surely at the same point of the state space, i.e., the ones which law is known to be of the form $\delta_{\mathbf{x}}$ for some $\mathbf{x} \in \mathbf{X}$.

Remark. Let $\mathfrak{C}$ be a stochastic representation on $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$ with values in the finite set $\mathcal{P} \doteq \{P_i \text{ s.t. } i \in \mathbb{I}\}$, then any realisation μ of $\mathfrak{C}$ can be expressed as

$$\mu = \sum_{i \in \mathbb{I}} \boldsymbol{n}_i \delta_{P_i} \tag{2.50}$$

for some $\boldsymbol{n}$ in the set $\mathbb{N}^{\mathbb{I}}$ of families of non-negative integers indexed by $\mathbb{I}$. The measure μ can be denoted $\mu_{\boldsymbol{n}}$ to underline the multiplicity of each atom in $\mathcal{P}$. The probability measure P induced by $\mathfrak{C}$ on $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$ can then be expressed as

$$P(B) = \int \mathbf{1}_B(\mu_{\boldsymbol{n}}) \boldsymbol{c}(\mathrm{d}\boldsymbol{n}) \tag{2.51}$$

for any Borel subset B of $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$, where $\boldsymbol{c}$ is a probability measure on $\mathbb{N}^{\mathbb{I}}$. In consequence, the only probabilistic constraint that needs to be introduced is the one for the probability measure $\boldsymbol{c}$. This is due to the fact that there is no weak types of uncertainty on the individual probabilistic constraints themselves since they are known to be in $\mathcal{P}$. Let $\boldsymbol{p} \in \mathbf{C}_1(\mathbb{N}^{\mathbb{I}})$ be this probabilistic constraint, then the probability measure P can be bounded above, for any Borel subset B of $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$, by

$$P(B) \leq \int \sup_{\boldsymbol{n} \in \mathbb{N}^{\mathbb{I}}} (\mathbf{1}_B(\mu_{\boldsymbol{n}}) h(\boldsymbol{n})) \boldsymbol{p}(\mathrm{d}h). \tag{2.52}$$

The measure $\boldsymbol{p}$ can be seen as a probabilistic constraint for P with a characteristic function of the form (1.26), that is

$$F : (A, h) \mapsto \|F'(A) \cdot h\| \qquad \text{with} \qquad F' : A \mapsto \mathbf{1}_{\{\boldsymbol{n} \,:\, \mu_{\boldsymbol{n}} \in A\}}. \tag{2.53}$$

The fusion operation described in (2.36) can now be expressed for stochastic representations on $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$ as in the following corollary.

Corollary 2.1. *Let $\mathfrak{C}$ and $\mathfrak{C}'$ be two stochastic representations in $\mathbf{N}(\mathbf{C}_1(\mathbf{X}))$ with values in the respective sets $\mathcal{P} \doteq \{P_i \text{ s.t. } i \in \mathbb{I}\}$ and $\mathcal{P}' \doteq \{P'_i \text{ s.t. } i \in \mathbb{I}'\}$, and let $\boldsymbol{p} \in \mathbf{C}_1(\mathbb{N}^{\mathbb{I}})$ and $\boldsymbol{p}' \in \mathbf{C}_1(\mathbb{N}^{\mathbb{I}'})$ be probabilistic constraints for the multiplicity of $\mathfrak{C}$ and $\mathfrak{C}'$ respectively. Then the probabilistic constraint $\hat{\boldsymbol{p}}$ in $\mathbf{C}_1(\mathbb{N}^{\mathbb{I}\times\mathbb{I}'})$ is characterised by*

$$(\forall H \in \mathcal{L}^{\infty}(\mathbb{N}^{\mathbb{I}\times\mathbb{I}'})) \qquad \hat{\boldsymbol{p}}(H) \propto \int \mathbf{1}_H(\Lambda(h,h')^{\dagger})\, \|\Lambda(h,h')\| \boldsymbol{p}(\mathrm{d}h)\boldsymbol{p}'(\mathrm{d}h'), \tag{2.54}$$

where the mapping $\Lambda : \mathbf{L}^{\infty}(\mathbb{N}^{\mathbb{I}}) \times \mathbf{L}^{\infty}(\mathbb{N}^{\mathbb{I}'}) \to \mathbf{L}^{\infty}(\mathbb{N}^{\mathbb{I}\times\mathbb{I}'})$ is defined as

$$\Lambda : (h,h') \mapsto ((h \rtimes h') \circ \theta) \cdot w, \tag{2.55}$$

with θ denoting the mapping from $\mathbb{N}^{\mathbb{I}\times\mathbb{I}'}$ to $\mathbb{N}^{\mathbb{I}} \times \mathbb{N}^{\mathbb{I}'}$, which recovers the prior multiplicities corresponding to a given posterior multiplicity, defined as

$$\theta : \boldsymbol{m} \mapsto \Big(\Big\{ \sum\nolimits_{j\in\mathbb{I}'} \boldsymbol{m}_{i,j} \Big\}_{i\in\mathbb{I}}, \Big\{ \sum\nolimits_{i\in\mathbb{I}} \boldsymbol{m}_{i,j} \Big\}_{j\in\mathbb{I}'} \Big), \tag{2.56}$$

and with the weighting function $w \in \mathbf{L}^{\infty}(\mathbb{N}^{\mathbb{I}\times\mathbb{I}'})$ defined as

$$w : \boldsymbol{m} \mapsto \prod_{(i,j)\in\mathbb{I}\times\mathbb{I}'} \|P_i * P'_j\|^{m_{i,j}}. \tag{2.57}$$

The fusion introduced in Corollary 2.1 cannot be seen as a special case of the operations defined in Chapter 1 since a given pair $(\boldsymbol{n},\boldsymbol{n}') \in \mathbb{N}^{\mathbb{I}} \times \mathbb{N}^{\mathbb{I}'}$ of prior multiplicities induces possibly many posterior multiplicities, as described by $\theta^{-1}[\{(\boldsymbol{n},\boldsymbol{n}')\}]$. However, the probabilistic constraint $\hat{\boldsymbol{p}}$ still retains a Bayesian flavour since it is the direct consequence of the fusion of the underlying probabilistic constraints on $\mathbf{C}_{\mathbf{F}}(\mathbf{X})$. A similar result can be obtained directly when considering more than two prior constraints whenever the mappings θ and w are accordingly defined.

This way of representing stochastic populations will be extremely useful when performing filtering in the next chapter since finite collections of individual representations are often available in practice, so that Assumption **A.3** is verified.

2.4.3 As a joint probability measure

Another formulation can be established on product spaces when the involved probabilistic constraints are all equivalent to probability measures. For a given index set $\mathbb{I}$, we introduce the set $\mathbf{X}^{\times}$ as

$$\mathbf{X}^{\times} \doteq \{\psi_{\infty}\} \cup \bigcup_{k\geq 1} \mathbf{X}^k, \tag{2.58}$$

where the point state denoted ψ_{∞} represents the case where infinitely many individuals are at point ψ. Also, for given $I \subseteq \mathbb{I}$ and $\boldsymbol{n} \in (\mathbb{N}^*)^I$, we introduce the set $\mathbf{X}_I^{(\boldsymbol{n})}$ as

$$\mathbf{X}_I^{(\boldsymbol{n})} \doteq \{X \in (\mathbf{X}^{\times})^I \text{ s.t. } (\forall i \in I)\ X_i \in \mathbf{X}^{\boldsymbol{n}_i}\}. \tag{2.59}$$

In particular, if $I = \emptyset$, then $\boldsymbol{n}$ is the empty function in $\mathbb{N}^*$ and $\mathbf{X}_{\emptyset}^{(\boldsymbol{n})}$ can be seen as a point state representing the case where the population is empty. Finally, the sets $\mathbb{N}_{\mathbb{I}}$ and $\mathbb{X}_{\mathbb{I}}^{\times}$ are defined as

$$\mathbb{N}_{\mathbb{I}} \doteq \{(I,\boldsymbol{n}) \text{ s.t. } I \subseteq \mathbb{I},\ \boldsymbol{n} \in (\mathbb{N}^*)^I\} \qquad \text{and} \qquad \mathbf{M}_{\mathbb{I}}(\mathbf{X}) \doteq \bigcup_{(I,\boldsymbol{n})\in\mathbb{N}_{\mathbb{I}}} \mathbf{M}_1(\mathbf{X}_I^{(\boldsymbol{n})}). \tag{2.60}$$

Proposition 2.6. *Assuming that the stochastic representations of interest take values in the subset* $\mathcal{P} \doteq \{p_i \text{ s.t. } i \in \mathbb{I}\}$ *of* $\mathbf{M}_1(\mathbf{X})$*, let* ζ' *be the mapping between* $\mathbf{N}(\mathbf{M}_1(\mathbf{X}))$ *and* $\mathbf{M}_{\mathbb{I}}(\mathbf{X})$ *defined as*

$$\zeta' : \mu_{\boldsymbol{n}} \mapsto P(\,\cdot \mid I_{\boldsymbol{n}}, \breve{\boldsymbol{n}}), \tag{2.61}$$

where $I_{\boldsymbol{n}} \doteq \mathrm{supp}(\boldsymbol{n})$*, where* $\breve{\boldsymbol{n}}$ *is the restrictions of* $\boldsymbol{n}$ *to its support, and where for any* $(I, \boldsymbol{n}) \in \mathbb{N}_{\mathbb{I}}$*, the probability measure* $P(\cdot \mid I, \boldsymbol{n})$ *is defined as*

$$(\forall B \in \mathcal{B}(\mathbf{X}_I^{(\boldsymbol{n})})) \qquad P(B \mid I, \boldsymbol{n}) \doteq \prod_{i \in I} p_i^{\times \boldsymbol{n}_i}(B_i), \tag{2.62}$$

then ζ' *is a bijection.*

It is not necessary to establish a fusion operation for this alternative representation since the fusion can be performed on counting measures in $\mathbf{N}(\mathbf{M}_1(\mathbf{X}))$ and the result can then be expressed as in Proposition 2.6, whenever the individual probabilistic constraints involved are all equivalent to probability measures.

Summary

Starting from general considerations about the concepts of individual and population and about the partially indistinguishable knowledge that may be available about them, we went across increasingly general notions in an attempt to faithfully describe the multi-faceted nature of the corresponding uncertainties. After a suitable level of generality was reached, three ways of expressing the uncertainty about these complex systems have been introduced. The first one proved to be useful for finding the result of general operations, whereas the second and third ones conveniently allowed for the expression of practical results under some reasonable assumptions. In particular, the formulation based on counting measures will be extremely useful in the next chapter, when tackling the derivation of a general filtering algorithm.

Chapter 3

Estimation of stochastic populations

The existing solutions in the field of multi-object estimation can be divided into two classes of multi-object filters. One class consists of "classical" filters, such as the **m**ultiple **h**ypothesis **t**racking or MHT [Blackman, 1986], that are based on practical generalisations of single-object filters. The strength of these classical filters is their ability to distinguish the objects of interest and to naturally characterise each of them. The other class of filters comprises approaches based on point processes, such as the **p**robability **h**ypothesis **d**ensity (PHD) filter introduced by Mahler [2003]. These filters successfully propagate global statistics about the population of interest and integrate clutter and appearance of objects in a principled way. However, they do not naturally propagate specific information about objects because of the point process assumption of indistinguishability. One of the attempts to overcome this limitation can be found in Vo and Vo [2013], where marked point processes are used.

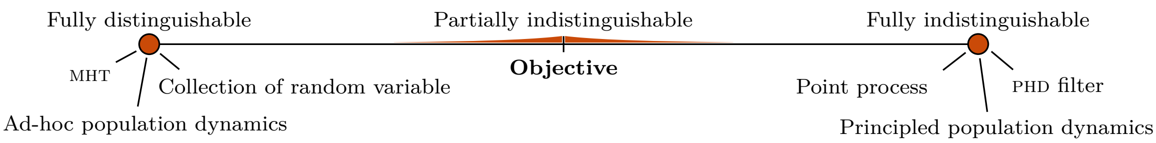


Figure 3.1: Estimation of partially-indistinguishable populations compared with existing approaches.

The possibility of conducting estimation using stochastic representations will be demonstrated in this chapter with the objective of finding a compromise between the existing methods as illustrated in Figure 3.1. For this purpose, Sections 3.1 and 3.2 will consider the target population as well as a given individual of this population. These will help to describe the basic facts about this population before introducing, in Section 3.2.2, the actual stochastic representation that will be used for filtering. The novelty in this chapter lies both in the versatility of the proposed algorithms and in the extent to which weak forms of uncertainty can be integrated.

Overall, this chapter can be treated as an example application of the concepts introduced in Chapter 2, although the set of assumptions that is considered is general enough to encompass many practical filtering problems. Simple generalisations that would generate a non-negligible notational burden will be given as remarks, but will not be integrated in the final filtering algorithm.

The successive stochastic representations used to model the evolution of the target population in time will be driven by some data that will be received from a sensing device, or sensor. Numerous types of sensors can be used to enhance our representation of the target population. Two general non-excluding types of sensors can be distinguished, the *counting* sensors and the *localising* sensors. Even though the emphasis will be put on the latter kind, both types can be used to enrich our knowledge of the target population in the present framework.

Without loss of generality, the time is indexed by the set $\mathbb{T} \doteq \mathbb{N}$, and for any $t \in \mathbb{T}$, we denote $\mathbb{T}_t \doteq \{t' \in \mathbb{T} \text{ s.t. } t' \leq t\}$ the set of time steps that are less or equal to t.

3.1 Individual modelling

We consider the spaces $\mathbf{X}_t$ and $\mathbf{Z}_t$ as state and observation spaces at time $t \in \mathbb{T}$. We assume that the state space $\mathbf{X}_t$ can be expressed as $\mathbf{X}_t = \{\psi\} \cup \mathbf{X}_t^\bullet$ where $\mathbf{X}_t^\bullet$ is an Euclidean space and ψ is an isolated point; similarly the observation space $\mathbf{Z}_t$ is expressed as $\mathbf{Z}_t = \{\phi\} \cup \mathbf{Z}_t^\bullet$. The isolated points ψ and ϕ are interpreted as an empty state and an empty observation respectively. The target population at time t is denoted $\mathcal{X}_t^\sharp$ and represents the set of individuals of interest. At time $t \in \mathbb{T}$, a given individual $x \in \mathcal{X}_t^\sharp$ is imaged in the state spaces $\mathbf{X}_t$ and $\mathbf{Z}_t$ through the projections $\xi_t : \boldsymbol{\mathcal{X}}_t^{\mathrm{a}} \to \mathbf{X}_t$ and $\xi'_t : \boldsymbol{\mathcal{X}}_t^{\mathrm{a}} \to \mathbf{Z}_t$ (recall that $\boldsymbol{\mathcal{X}}_{\mathrm{a}}$ was defined as a representative individual state space in Section 2.2.1). We assume the following:

A.5 the projection ξ_t verifies $(x \notin \mathcal{X}_t^\sharp) \Rightarrow (\xi_t(x) = \psi)$,

that is, ξ_t gives images in $\mathbf{X}_t^\bullet$ to individuals in the target population only. The same assumption is not made about ξ'_t since individuals that are not in the population of interest might still generate observations, in a way that strongly depends on the sensor(s). State histories in $\mathbb{X}_t$ and $\mathbb{Z}_t$ can then be formed for the individual x, where

$$\mathbb{X}_t = \bigtimes_{t' \in \mathbb{T}_t} \mathbf{X}_{t'} \qquad \text{and} \qquad \mathbb{Z}_t = \bigtimes_{t' \in \mathbb{T}_t} \mathbf{Z}_{t'}. \tag{3.1}$$

The empty states $\boldsymbol{\psi}_t$ and $\boldsymbol{\phi}_t$ in $\mathbb{X}_t$ and $\mathbb{Z}_t$ are the histories containing only the empty states ψ and ϕ respectively. The sets $\mathbb{X}_t^\bullet$ and $\mathbb{Z}_t^\bullet$ are respectively defined as $\mathbb{X}_t - \{\boldsymbol{\psi}_t\}$ and $\mathbb{Z}_t - \{\boldsymbol{\phi}_t\}$. The following assumptions about any individual x in $\mathcal{X}_t^\sharp$, $t \in \mathbb{T}$, are introduced in order to simplify the form of state histories in the different spaces involved.

A.6 Individuals exist in $\boldsymbol{\mathcal{X}}_t^{\mathrm{a}}$ for all times $t \in \mathbb{T}$.

A.7 Individuals can only continuously exist in $\mathbf{X}_t$ for a given time interval in $\mathbb{T}$: there exists a time interval $T = [t_+, t_-] \subseteq \mathbb{T}$, referred to as *time of presence*, which verifies

$$(\forall t \in \mathbb{T}) \qquad (t \in \mathbb{T} - T) \Leftrightarrow (\xi_t(x) = \psi). \tag{3.2}$$

A.8 Individuals have an image in $\mathbf{Z}_t^\bullet$ only when they have an image in $\mathbf{X}_t^\bullet$:

$$(\forall t \in \mathbb{T}) \qquad (t \in \mathbb{T} - T) \Rightarrow (\xi'_t(x) = \phi). \tag{3.3}$$

Assumptions **A.6** to **A.8** create a hierarchy between $\boldsymbol{\mathcal{X}}^{\mathrm{a}}_t$, $\mathbf{X}_t$ and $\mathbf{Z}_t$, as sets of decreasing sophistication. The consequence of Assumptions **A.7** and **A.8** is that x can be represented by a state and observation history $\boldsymbol{x}_t \in \mathbb{X}^{\bullet}_t$ and $\boldsymbol{z}_t \in \mathbb{Z}_t$ of the form

$$\boldsymbol{x}_t = (\psi, \dots, \psi, \mathbf{x}_{t_+}, \dots, \mathbf{x}_{t_-}, \psi, \dots, \psi), \tag{3.4a}$$

$$\boldsymbol{z}_t = (\phi, \dots, \phi, \mathbf{z}_{t_+}, \dots, \mathbf{z}_{t_-}, \phi, \dots, \phi), \tag{3.4b}$$

where $\mathbf{x}_t \in \mathbf{X}^{\bullet}_t$ and $\mathbf{z}_t \in \mathbf{Z}_t$ for any $t \in T$. Note that the times t_+ and t_- cannot be recovered from a given observation history $\boldsymbol{z}_t \in \mathbb{Z}_t$ as any/all $z_{t'}$ may be equal to ϕ.

3.1.1 Time transition

At this point, the image of the individual x in $\mathbf{X}_{t-1}$ and $\mathbf{X}_t$ is computed separately through ξ_{t-1} and ξ_t. However, as neither the state of x in $\boldsymbol{\mathcal{X}}^{\mathrm{a}}_{t'}$ nor the mappings $\xi_{t'}$ with $t' \in [0, t]$ are known in practice, there is a need for a transition from $\mathbb{X}_{t-1}$ to $\mathbb{X}_t$. The latter does not require the individual x to be numerically described in $\boldsymbol{\mathcal{X}}^{\mathrm{a}}_{t-1}$ when its state in $\mathbb{X}_{t-1}$ is given. Formally, a Markov kernel $q_{t|t-1} \in \mathbf{K}_1(\mathbb{X}_{t-1}, \mathbb{X}_t)$ is assumed to be given. The stochasticity of $q_{t|t-1}$ comes from the uncertainty on/randomness of the actual transition performed by x. As a consequence of Assumption **A.7**, the following interpretations can be conducted:

- A transition from $\boldsymbol{\psi}_{t-1}$ to $\mathbb{X}^{\bullet}_t$ is interpreted as a spontaneous appearance, and $t_+ = t$ for the newborn individual x.
- A transition from a state in $\boldsymbol{x}_{t-1} \in \mathbb{X}^{\bullet}_{t-1}$ to the state[1] $\boldsymbol{x}_{t-1} \times \psi$ in $\mathbb{X}_t$ is a spontaneous disappearance, and $t_- = t - 1$ for the individual x.

If the description of the state of the individual in $\mathbf{X}_t$ is sufficient and if the transition only depends on the previous state space $\mathbf{X}_{t-1}$, then the stochastic process associated with x is said to have the *Markov property* and the considered kernels can be taken in $\mathbf{K}_1(\mathbf{X}_{t-1}, \mathbf{X}_t)$.

We also assume that there might be some prior knowledge about the state $\xi_t(x)$ of the individual x in $\mathbf{X}_t$, i.e., the individual might be assumed to be in a specific area of the space with a given probability (such an area is usually given by environmental constraints like roads in traffic control or blood vessels in biomedical imaging). This is best modelled by a probabilistic constraint $P_t \in \mathbf{C}_1(\mathbf{X}_t)$.

3.1.2 Observation of an individual

As the spaces $\mathbf{Z}^{\bullet}_{t'}$ are assumed to be simpler, e.g., of lower dimension, than the respective state spaces $\mathbf{X}^{\bullet}_{t'}$, the transition from $\mathbb{Z}_{t-1}$ to $\mathbb{Z}_t$ is less accurately described than the transition from $\mathbb{X}_{t-1}$ to $\mathbb{X}_t$. In consequence, it is more precise to achieve a transition from $\mathbb{X}_t$ to $\mathbb{Z}_t$ directly. The knowledge of a kernel from $\mathbb{X}_t$ to $\mathbf{Z}_t$ is then sufficient. We will additionally assume that this transition can be described by a Markov kernel $\ell_t \in \mathbf{K}_1(\mathbf{X}_t, \mathbf{Z}_t)$. According to Assumption **A.8**, ℓ_t verifies $\ell_t(\psi, \{\phi\}) = 1$ for individuals of the target population.

Information about the individual x is gained through its observation via a sensor. At time $t \in \mathbb{T}$, the sensor operates in the observation space $\mathbf{Z}^{\bullet}_t$ and attempts to localise

[1] If $\boldsymbol{x}$ and $\mathbf{x}$ are respectively a sequence and a point in a given set, then $\boldsymbol{x} \times \mathbf{x}$ denotes the sequence $\boldsymbol{x}$ augmented with the point $\mathbf{x}$.

the individual in this space. When successful, it provides us with a subset A_t of $\mathbf{Z}_t^\bullet$ in which the individual has been detected. The subset A_t is called an *observation*. When unsuccessful, the sensor does not provide anything, the observation is the empty observation ϕ and we say that a *missed detection* occurred for x. A missed detection can occur either if the sensor fails to detect x while it has an image in $\mathbf{Z}_t^\bullet$ or if the image of x onto $\mathbf{Z}_t$ is ϕ, such as when x is not in the field of view of the sensor. The subset A_t of $\mathbf{Z}_t^\bullet$ is assumed to be a Borel set. The fusion of the information about x from the sensor together with its description on $\mathbf{X}_t$ is called an *observation update*.

3.2 Population modelling

3.2.1 Representation

Recall that a population $\mathcal{X}_t$ is a set of individuals that are given a state in $\mathbf{X}_t$ through the function ξ_t at time $t \in \mathbb{T}$. As in Section 2.2.1, we consider $\mathcal{X}_t^\bullet$ and $\mathcal{X}_t^\psi$ as the subsets of individuals with images via ξ_t in $\mathbf{X}_t^\bullet$ and at ψ respectively, so that the population $\mathcal{X}_t$ can be expressed as

$$\mathcal{X}_t = \mathcal{X}_t^\bullet \uplus \mathcal{X}_t^\psi. \tag{3.5}$$

This is one of the most fundamental partition of the population $\mathcal{X}_t$; yet, other partitions will prove to be useful for specifying the models required to perform time and observation updates.

Before time $t = 0$

Let $\mathbf{X}_-$ be the state space at any time $t < 0$, right before the beginning of the time interval $\mathbb{T}$ of interest. For the sake of simplicity, and without loss of generality, we assume that:

A.9 It holds that $\mathbf{X}_- = \{\psi\}$.

The time steps $t < 0$ are not represented in the individuals' state histories and observation histories as nothing happens at these times. The only possible probability measure on $\mathbf{X}_-$ is δ_ψ. Assumption **A.9** also implies that every individual in $\mathcal{X}_-$ has image ψ, i.e., $\mathcal{X}_-^\bullet = \emptyset$, which avoids intricate notations at time t by reducing the set of individuals to those which were born between time 0 and t.

From time $t-1$ to time t

First of all, every individual in $\mathcal{X}_{t-1}^\bullet$ is propagated to time t according to the model introduced in Section 3.1. Also, at time t, some of the individuals in the target population that had image $\boldsymbol{\psi}_{t-1}$ appear in $\mathbf{X}_t^\bullet$ for the first time, which is interpreted as *birth*. We assume that:

A.10 there is no specific prior knowledge on the individuals in $\mathcal{X}_t^\bullet$.

Assumption **A.10** is made for the sake of simplicity, as other choices would lead to more complex expressions for the time update. An example for which it would be necessary to relax this constraint is given by the case where a set of roads is available and the objective is to estimate which individual goes along which road. Assumption **A.10** does not however prevent from modelling environmental constraints as noted in Section 3.1.

Observation at time t

At time $t \in \mathbb{T}$, a family $\{A_t^z\}_{z \in Z_t}$ of observations in $\mathbf{Z}_t^\bullet$ indexed by the set Z_t is made available by the sensor. Each observation might be generated by several individuals in $\mathcal{X}_t$, but we assume that an individual cannot be related to several observations.

Remark. Considering more general forms of correspondence between populations allows for modelling "extended" individuals that could generate several distinct observations [Salmond and Gordon, 1999, Koch, 2008].

In order to model the existence of *spurious observations*, also referred to as *false positives*, we introduce another stochastic population $\mathcal{X}_t^\flat$ where the symbol "$\flat$" relates to an *outer* population that is not of direct interest but which might interfere with the observation of the target population[2]. Individuals in $\mathcal{X}_t^\flat$ have the image ψ in $\mathbf{X}_t$ as we assume that the projection ξ_t gives an image in $\mathbf{X}_t^\bullet$ to individuals in the target population only. The population $\mathcal{X}_t$ that has to be considered is then of the form

$$\mathcal{X}_t = \mathcal{X}_t^\sharp \uplus \mathcal{X}_t^\flat, \tag{3.6}$$

where $\mathcal{X}_t^\sharp$ denotes the target population, as mentioned before. It holds that $\mathcal{X}_t^\bullet \subseteq \mathcal{X}_t^\sharp$ whereas $\mathcal{X}_t^\psi \nsubseteq \mathcal{X}_t^\sharp$ whenever $\mathcal{X}_t^\flat$ is non-empty. The individuals in the outer population are strongly indistinguishable since they all have state ψ in $\mathbf{X}_t$.

Remark. The outer population should not be made of physical individuals as these might generate consistent observations in time and would be better represented by another stochastic population with, e.g., a dynamical model that differs from the one of $\mathcal{X}_t^\sharp$, see Section 3.2.4 for more details.

Note that Assumption **A.8** does not hold for individuals in $\mathcal{X}_t^\flat$ since these individuals may have an image in $\mathbf{Z}_t^\bullet$ without being in $\mathbf{X}_t^\bullet$. Every individual in $\mathcal{X}_t$ generates an image on the observation space $\mathbf{Z}_t$ through the mapping ξ_t'. Then, an actual observation is generated for each of them, either in $\mathbf{Z}_t^\bullet$ or at ϕ, depending on the characteristics of the sensor and of the capabilities of the detector. Notice that the set $\mathbf{Z}_t$ is used as an observation space and as a sensor space at the same time, with the difference that the image of an individual in the sensor space depends on the capabilities of the sensor.

Remark. Considering $\mathcal{X}_t^\flat$ as a population could allow for estimating the parameters that describe the false positives. However, these often exhibit highly non-Markovian behaviours and their dynamics is generally unknown. We will see in the following sections and chapters that we can however estimate the probability for an observation to be a false positive a posteriori, and this information can be useful for evaluating the characteristics of these spurious observations in a post-processing stage.

Description of the population

In order to describe the different classes of individuals in the population $\mathcal{X}_t$, we introduce the symbols "$\cdot t$" and "$:t$" that relate to these classes at time $t \in \mathbb{T}$, after the time filtering and after the observation filtering respectively. The most direct way to

[2]As a mnemonics, the symbol "$\flat$" is called *bémol* in French, which is used to refer, in addition to the music-related concept, to a flaw or a drawback, and hence well qualifies the false positives.

distinguish individuals is to look at their observation histories. For this reason, we consider the space $\bar{\mathbb{O}}_t$, defined as

$$\bar{\mathbb{O}}_t \doteq \bigtimes_{t' \in \mathbb{T}_t} \bar{Z}_t, \tag{3.7}$$

with $\bar{Z}_t \doteq Z_t \cup \{\phi\}$, so that $\boldsymbol{o}_t \in \bar{\mathbb{O}}_t$ takes the form

$$\boldsymbol{o}_t = (\phi, \dots, \phi, z_{t_+}, \dots, z_{t_-}, \phi, \dots, \phi). \tag{3.8}$$

The observation history $\boldsymbol{o}_t$ can also be referred to as the *observation path* of the individual x [Pace and Del Moral, 2013]. The set $\mathbb{O}_t = \bar{\mathbb{O}}_t - \{\boldsymbol{\phi}_t\}$ of non-empty observation paths is defined for consistency.

Under Assumptions **A.9** and **A.10**, the only sources of specific information at time $t \in \mathbb{T}$ lie in the description of the newborn individuals at different times and in the observations made before time t. The representations of individuals in $\mathcal{X}_t^\sharp$ after the time update can then be indexed by the set $\mathbb{I}^\sharp_{\cdot t}$, defined as follows:

$$\mathbb{I}^\sharp_{\cdot t} \doteq \Big\{(\sharp, T, \boldsymbol{o}) \text{ s.t. } (T, \boldsymbol{o}) \in \mathcal{F}_I(\mathbb{T}_t) \times \bar{\mathbb{O}}_{t-1}, \;\; (\forall t' \in \mathbb{T}_{t-1}) \;\; \boldsymbol{o}_{t'} \neq \phi \Rightarrow t' \in T\Big\}, \tag{3.9}$$

where $\mathcal{F}_I(\mathbb{T}_t)$ is the set of closed intervals of $\mathbb{T}_t$. The definition of $\mathbb{I}^\sharp_{\cdot t}$ can be interpreted as follows: time-updated representations are distinguished by their interval of existence up to time t and by their observation path, the latter being constrained to be empty at times where the individual does not exist. The set $\mathbb{I}^\sharp_{:t}$ indexing individual representations at time t after the observation update is defined similarly:

$$\mathbb{I}^\sharp_{:t} \doteq \Big\{(\sharp, T, \boldsymbol{o}) \text{ s.t. } (T, \boldsymbol{o}) \in \mathcal{F}_I(\mathbb{T}_t) \times \bar{\mathbb{O}}_t, \;\; (\forall t' \in \mathbb{T}_t) \;\; \boldsymbol{o}_{t'} \neq \phi \Rightarrow t' \in T\Big\}. \tag{3.10}$$

The only difference with $\mathbb{I}^\sharp_{\cdot t}$ is that observation paths are now defined up to time t, rather than up to time $t-1$.

The representations of the individuals in $\mathcal{X}_t^\sharp$, e.g., after the observation update, can then be indexed as follows: *a*) a detected individual born at time t_+ that went away at time t_-, is indexed by $(\sharp, [t_+, t_-], \boldsymbol{o}) \in \mathbb{I}^\sharp_{:t}$, *b*) an individual born at time t_+ that still exists at time t but that remains undetected has index $(\sharp, [t_+, t], \boldsymbol{\phi}_t)$, and *c*) an individual that is not born yet has index $(\sharp, \emptyset, \boldsymbol{\phi}_t)$. On the other hand, individuals in $\mathcal{X}_t^\flat$ never exist in $\mathbf{X}_t^\bullet$ and do not give consistent observations, which corresponds to the indices $\boldsymbol{i}^\flat_{\cdot t} = (\flat, \emptyset, \boldsymbol{\phi}_{t-1})$ and $\boldsymbol{i}^\flat_{:t} = (\flat, \emptyset, \boldsymbol{\phi}_t)$. The previously defined index sets is then extended in order to index the whole population $\mathcal{X}_t$ as $\mathbb{I}_{\boldsymbol{t}} \doteq \mathbb{I}^\sharp_{\boldsymbol{t}} \cup \{\boldsymbol{i}^\flat_{\boldsymbol{t}}\}$, where $\boldsymbol{t}$ is either "$\cdot t$" or "$:t$".

Remark. The index $\boldsymbol{i}^\flat_t$ chosen for the outer population $\mathcal{X}_t^\flat$ suggests that nothing is learned about this sub-population through time and observation filtering. We will see in the next sections that this is due to the fact that the number and respective behaviour of individuals in $\mathcal{X}_t^\flat$ is fully unknown. The only thing that could be recorded is the subset of $Z_{t'}$ that is assigned to $\mathcal{X}_{t'}^\flat$ at all times $t' \leq t$, but that information can be recovered from the subset of observations that is assigned to $\mathcal{X}_{t'}^\sharp$.

Remark. If we relax Assumption **A.10** and try to learn about which predefined path is taken by which individual, then the "identifier" of the path, say i, should be added to the individual indices which would be of the form $(\sharp, T, \boldsymbol{o}, i)$.

Some additional notations are required to express specific statements about subsets of the index sets $\mathbb{I}_{\cdot t}$ and $\mathbb{I}_{:t}$ as well as multiplicities related to them. In particular we consider the subsets $\mathbb{I}^{\bullet}_{\boldsymbol{t}}$ and $\mathbb{I}^{\psi}_{\boldsymbol{t}}$, where $\boldsymbol{t}$ is "$\cdot t$" or "$:t$", defined as

$$\mathbb{I}^{\bullet}_{\boldsymbol{t}} = \{(s, T, \boldsymbol{o}) \in \mathbb{I}_{\boldsymbol{t}} \text{ s.t. } T \ni t\} \qquad \text{and} \qquad \mathbb{I}^{\psi}_{\boldsymbol{t}} = \mathbb{I}_{\boldsymbol{t}} - \mathbb{I}^{\bullet}_{\boldsymbol{t}}. \tag{3.11}$$

The definition of these subsets is consistent with the partitioning $\mathcal{X}_t = \mathcal{X}^{\bullet}_t \uplus \mathcal{X}^{\psi}_t$ of the considered population since, e.g., individuals in $\mathcal{X}^{\bullet}_t$ will be naturally associated with indices in the subset $\mathbb{I}^{\bullet}_{\boldsymbol{t}}$ of $\mathbb{I}_{\boldsymbol{t}}$. As far as the multiplicities related to descriptions indexed by elements of $\mathbb{I}_{\boldsymbol{t}}$ are concerned, we consider the sets

$$\mathbb{M}_{\boldsymbol{t}} = \{\boldsymbol{n} \in \bar{\mathbb{N}}^{\mathbb{I}_{\boldsymbol{t}}} \text{ s.t. } (\forall \boldsymbol{i} \in \mathbb{I}^{\bullet}_{\boldsymbol{t}}) \;\; \boldsymbol{n}_{\boldsymbol{i}} < \infty\}, \tag{3.12a}$$

$$\mathbb{M}'_t = \{\boldsymbol{n} \in \bar{\mathbb{N}}^{\bar{Z}_t} \text{ s.t. } (\forall z \in Z_t) \;\; 0 < \boldsymbol{n}_z < \infty\}, \tag{3.12b}$$

where $\bar{\mathbb{N}} \doteq \mathbb{N} \cup \{\infty\}$. Note that an infinite multiplicity is authorised for individuals that are almost surely at points ψ and ϕ and that each actual observation is assumed to represent at least one individual in the population.

The objective is now to define the form of the stochastic representations involved in filtering, considering all the modelling choices that have been introduced so far. In particular, the index sets $\mathbb{I}_{\cdot t}$ and $\mathbb{I}_{:t}$ will be shown to be suited to the expression of the time-updated and observation-updated stochastic representations.

3.2.2 Stochastic representation

The last step before stating the filtering equations of our model is to express the law of the different stochastic representations that will be involved in the filtering algorithm. From Assumption **A.7**, individuals in the target population can appear and disappear only once from the state space across all time steps so that we use the terms *birth* and *death* to describe these transitions.

Before time $t = 0$

At time $t < 0$, the state space $\mathbf{X}_-$ verifies $\mathbf{X}_- = \{\psi\}$ and $\mathbf{M}_1(\mathbf{X}_-) = \{\delta_\psi\}$. As a result, counting measures on $\mathbf{C}_1(\mathbf{X}_-) = \{\delta_{f_-}\}$ with $f_- \doteq \mathbf{1}_\psi$ are of the form $\mu^{(n)}_- = n\delta\lfloor\delta_{f_-}\rfloor$, and the law of $\mathfrak{C}_-$ on $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_-))$ is reduced to

$$\mathbb{P}_- = \int c_-(\mathrm{d}n)\delta\lfloor\mu^{(n)}_-\rfloor, \tag{3.13a}$$

$$= \int c_-(\mathrm{d}n)\delta\Big\lfloor n\delta\lfloor\delta_{f_-}\rfloor\Big\rfloor, \tag{3.13b}$$

where $c_- \in \mathbf{M}_1(\bar{\mathbb{N}})$ is the law of the cardinality associated with $\mathfrak{C}_-$, with $\bar{\mathbb{N}} \doteq \mathbb{N} \cup \{\infty\}$. The extreme simplicity of the stochastic representation $\mathfrak{C}_-$ is particularly visible in (3.13b), where only Dirac measures appear, demonstrating the lack of diversity in the representation. Yet, $\mathfrak{C}_-$ allows for modelling that the only uncertainty before time $t = 0$ is on the cardinality of the target population. This can prove to be important when a prior knowledge about this cardinality is available, forcing all the subsequent births to comply with the given cardinality, as in Example 3.1 below.

Example 3.1. If the cardinality of the target population is known in advance to be $N \in \mathbb{N}$, or at least bounded by N, then a representation that seems probable at a given time t might become extremely unlikely later on if the associated number of existing and disappeared individuals becomes higher than N. This behaviour can be useful in challenging scenarios or when the computational resources are limited, in which case the maximum a posteriori for the stochastic representation $\mathfrak{C}_t$ will be the best configuration with no more than N individuals, including the disappeared ones.

If there is no known limitation on the cardinality of the target population, then one can allow for infinitely many individuals to appear in the future by setting *a*) $c_- = \delta_\infty$, or *b*) by introducing a probabilistic constraint $p_- \in \mathbf{C}_1(\mathbf{X}_-)$ for c_-, such that $p_- = \delta_{\mathbf{1}}$. In this case, $\mathfrak{C}_-$ becomes very uninformative, but still serves as a meaningful starting point. The difference between *a*) and *b*) is that the former asserts that the population is *known* to be infinitely large, whereas the latter models that nothing is known about the size of the population, and that this size is possibly infinite. In practice, the main difference is that *b*) allows for estimating the size of the population while *a*) does not.

Prior knowledge at time t

Under Assumption **A.10**, there is no specific prior knowledge about the individuals in the target population at time t, so that they are all indistinguishable in the corresponding stochastic representation $\mathfrak{C}_t$, and are all associated with one probabilistic constraint P_t in $\mathbf{C}_1(\mathbf{X}_t)$. As it is assumed that the kind of knowledge available does not allow for localising the individuals, it is meaningful to consider that P_t is of the form δ_{f_t} for some f_t in $\mathbf{L}(\mathbf{X}_t)$. We additionally assume that $f_t(\psi) = 1$, since there is no natural objection against individuals that are not present at time t. In consequence, the counting measures in $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_t))$ related to $\mathfrak{C}_t$ are of the form

$$\mu_t^{(n)} = n\delta\lfloor\delta_{f_t}\rfloor, \tag{3.14}$$

with $n \in \bar{\mathbb{N}}$, and the law $\mathbb{P}_t$ of $\mathfrak{C}_t$ is of the form

$$\mathbb{P}_t = \int c_t(\mathrm{d}n)\delta\lfloor\mu_t^{(n)}\rfloor, \tag{3.15}$$

where $c_t \in \mathbf{M}_1(\bar{\mathbb{N}})$ is the law of the cardinality associated with $\mathfrak{C}_t$. Assuming that the cardinality of the population is not statistically known in advance in general, we consider the probabilistic constraint $p_t \in \mathbf{C}_1(\bar{\mathbb{N}})$ for c_t to be of the form $p_t = \delta_{h_t}$, with, for instance, $h_t = \mathbf{1}$ in the completely uninformative case. Again, the description of $\mathfrak{C}_t$ is very simple, suggesting the fact that this representation is not highly informative. In particular, if both f_t and h_t are set to be everywhere equal to 1, then no information is contained in $\mathfrak{C}_t$. In this case, the time update does not need to be performed, and the observation prediction can be directly applied to the time-predicted stochastic representation.

Observation at time t

In this section, the objective is to incorporate the received observations $\{A_t^z\}_{z\in Z_t}$ into a single stochastic representation $\mathfrak{C}'_t$. First, each observation can be directly transformed into elements of $\mathbf{L}(\mathbf{Z}_t)$ by identifying a measurable subset with its indicator function. As a result, we obtain a collection of measurable functions $\{f_t^z\}_{z\in Z_t}$ with $f_t^z = \mathbf{1}_{A_t^z}$ for every $z \in Z_t$. It is also required to integrate another individual representation, which is not given by the sensor, namely $\mathbf{1}_{\{\phi\}} \in \mathbf{L}(\mathbf{Z}_t)$, and which will be associated with undetected individuals. From these considerations, it appears that all the counting measures on $\mathbf{C}_1(\mathbf{Z}_t)$ induced by $\mathfrak{C}'_t$ will have their support in the set

$$\{f_t^z \text{ s.t. } z \in Z_t\} \cup \{\mathbf{1}_{\{\phi\}}\}. \tag{3.16}$$

The element $\mathbf{1}_{\{\phi\}}$ is denoted f_t^ϕ, and A_t^ϕ is defined as the singleton $\{\phi\}$ for consistency.

Remark. The individual observations are put together as an indexed family since it could happen that two of them would be equal, possibly giving an indication about the number of individuals that are represented by them. This aspect can be specifically integrated in the law $\boldsymbol{c}'_t$ of the multiplicities in $\mathfrak{C}'_t$ or in the corresponding probabilistic constraint introduced below.

The counting measures in $\mathbf{N}(\mathbf{C}_1(\mathbf{Z}_t))$ are of the form

$$\mu_t^{\prime(\boldsymbol{n})} = \sum_{z \in \bar{Z}_t} \boldsymbol{n}_z \delta\lfloor \delta_{f_t^z} \rfloor \tag{3.17}$$

for a given $\boldsymbol{n}$ in the set $\mathbb{M}'_t$ defined in (3.12b). To specify $\mathfrak{C}'_t$, the law $\boldsymbol{c}'_t$ of the multiplicity on $\mathbb{M}'_t$ has to be given and two cases can be distinguished:

a) If the sensor provides no knowledge about the number of individuals represented by each observation, then one possible way is to introduce a probabilistic constraint $\boldsymbol{p}'_t \in \mathbf{C}_1(\mathbb{M}'_t)$ for $\boldsymbol{c}'_t$ such that $\boldsymbol{p}'_t = \delta_{\mathbf{1}}$.

b) Otherwise, if the sensor does provide a law c_z describing the number of individuals represented by the observation A_t^z, for every $z \in Z_t$, then a meaningful choice for $\boldsymbol{c}'_t$ is characterised by

$$\boldsymbol{c}'_t = \sum_{\boldsymbol{n} \in \mathbb{M}'_t} \left[\prod_{z \in \bar{Z}_t} w_z(\boldsymbol{n}_z) \right] \delta_{\boldsymbol{n}}, \tag{3.18}$$

where w_z denotes the probability mass function corresponding to c_z for any observation $z \in \bar{Z}_t$. The law c_ϕ characterising the number of undetected individuals might not be known since there might be no information on this cardinality. In this case, c_ϕ can be constrained by $p_\phi = \delta_{\mathbf{1}} \in \mathbf{C}_1(\bar{\mathbb{N}})$.

Overall, the law of the stochastic representation $\mathfrak{C}'_t$ takes the form

$$\mathbb{P}'_t = \int \boldsymbol{c}'_t(\mathrm{d}\boldsymbol{n}) \delta\left\lfloor \mu_t^{\prime(\boldsymbol{n})} \right\rfloor. \tag{3.19}$$

This stochastic representation contains the knowledge related to the observations of the sensor, and can then be joined and fused with, e.g., the observation-predicted stochastic representation that will be associated with the population $\mathcal{X}_t$.

Observation-updated knowledge at time $t-1$

In this section, the objective is to find a sufficiently general way of writing stochastic representations, considering the assumptions made in the previous sections. The fact that the selected form is general enough will be verified when proceeding to the time and observation updates at time t. If, after these two steps, the stochastic representation has a similar form, then this expression will be considered as *closed form*.

We assume that an individual probabilistic constraint $P_{\boldsymbol{i}} \in \mathbf{C}_1(\mathbf{X}_{t-1})$ is available for every index $\boldsymbol{i}$ in the set $\mathbb{I}_{:t-1}$ introduced in Section 3.2, so that the counting measures on $\mathbf{C}_1(\mathbf{X}_{t-1})$ induced by the stochastic representation $\mathfrak{C}_{:t-1}$ will be of the form

$$\mu^{(\boldsymbol{n})}_{:t-1} = \sum_{\boldsymbol{i} \in \mathbb{I}_{:t-1}} \boldsymbol{n}_{\boldsymbol{i}} \delta_{P_{\boldsymbol{i}}} \tag{3.20}$$

for a given $\boldsymbol{n} \in \mathbb{M}_{:t-1}$, where $\boldsymbol{n}_{\boldsymbol{i}}$ is the multiplicity at $P_{\boldsymbol{i}}$ for all $\boldsymbol{i} \in \mathbb{I}_{:t-1}$. The form of the law $\mathbb{P}_{:t-1}$ describing $\mathfrak{C}_{:t-1}$ is then given by

$$\mathbb{P}_{:t-1} = \int \boldsymbol{c}_{:t-1}(\mathrm{d}\boldsymbol{n}) \delta\lfloor \mu^{(\boldsymbol{n})}_{:t-1} \rfloor, \tag{3.21}$$

where $\boldsymbol{c}_{:t-1}$ is a probability measure on $\mathbb{M}_{:t-1}$ describing the multiplicity of each individual probabilistic constraint in $\mathfrak{C}_{:t-1}$. When the law $\boldsymbol{c}_{:t-1}$ is not fully known, it can be described by a probabilistic constraint $\boldsymbol{p}_{:t-1} \in \mathbf{C}_1(\mathbb{M}_{:t-1})$. A simpler expression of the law $\mathbb{P}_{:t-1}$ is given in the following remark in the case where all the individuals are distinguishable.

Remark 3.1. If all the individuals are distinguishable in the stochastic representation $\mathfrak{C}_{:t-1}$, then the probability mass function $\boldsymbol{w}_{:t-1}$ induced by $\boldsymbol{c}_{:t-1}$ must satisfy

$$(\forall \boldsymbol{n} \in \mathbb{M}_{:t-1}) \qquad \exists \boldsymbol{i} \in \mathbb{I}_{:t-1}(\boldsymbol{n}_{\boldsymbol{i}} > 1) \Rightarrow \boldsymbol{w}_{:t-1}(\boldsymbol{n}) = 0. \tag{3.22}$$

It is then more convenient to introduce another probability mass function $w_{:t-1}$ on $\wp(\mathbb{I}_{:t-1})$ such that

$$w_{:t-1}(I) = \begin{cases} \boldsymbol{w}_{:t-1}(\boldsymbol{n}) & \text{if } \mathrm{supp}(\boldsymbol{w}_{:t-1}) = I \\ 0 & \text{otherwise} \end{cases} \tag{3.23}$$

for any $I \subseteq \mathbb{I}_{:t-1}$. A simple expression of $\mathbb{P}_{:t-1}$ can then be given as

$$\mathbb{P}_{:t-1} = \sum_{I \subseteq \mathbb{I}_{:t-1}} w_{:t-1}(I) \delta\Big\lfloor \mu^{(I)}_{:t-1} \Big\rfloor, \tag{3.24}$$

where $\mu^{(I)}_{:t-1}$ is the simple counting measure with support $\{P_{\boldsymbol{i}} \text{ s.t. } \boldsymbol{i} \in I\}$. A subset I of $\mathbb{I}_{:t-1}$ can then be thought as being a *population hypothesis*, as it characterises one of the realisations of $\mathfrak{C}_{:t-1}$. This example illustrates one of the possible simplifications of a representation, obtained by making all the individuals distinguishable. This will be of interest at the end of the current chapter as well as in Chapter 4, where the objective is to derive practical estimation algorithms for stochastic populations.

It is useful to introduce some other symbols for representing some specific subpopulations of $\mathcal{X}_t$: the individuals with an index $(\sharp, T, \boldsymbol{o}) \in \mathbb{I}^{\sharp}_{:t-1}$ that

($\circ$) have been previously in the state space, i.e., such that $T \neq \emptyset$ and $T \not\ni t-1$,

($\boldsymbol{\psi}$) have never been in the state space, i.e., such that $T = \emptyset$.

The indices $\boldsymbol{i}^{\psi}_{\cdot t} \doteq (\sharp, \emptyset, \boldsymbol{\phi}_{t-1})$ and $\boldsymbol{i}^{\psi}_{:t} \doteq (\sharp, \emptyset, \boldsymbol{\phi}_t)$ are the only elements of $\mathbb{I}^{\psi}_{\cdot t}$ and $\mathbb{I}^{\psi}_{:t}$. Before proceeding to the time and observation filtering steps, it is required to define the stochastic population kernels that allow for relating representations on $\mathbf{X}_{t-1}$ and $\mathbf{X}_t$ as well as on $\mathbf{X}_t$ and $\mathbf{Z}_t$.

3.2.3 Stochastic representation for kernels

Up to this point, for the sake of simplicity, it has been assumed that individuals can be propagated in time or projected in the observation space using given kernels. However, in general, there might be uncertainty on how to propagate the individuals within a given population.

For the time prediction

We consider the following kernels:

(α) A Markov kernel q_α in $\mathbf{K}_1(\{\psi\}, \mathbf{X}^\bullet_t)$ that models appearance of new individuals. The multiplicity of such a kernel is driven by the law $c_\alpha \in \mathbf{M}_1(\mathbb{N})$.

(β) A stochastic kernel q_β in $\mathbf{K}(\mathbf{X}^\bullet_{t-1}, \mathbf{X}^\bullet_t)$ modelling the dynamics of the individuals in the target population $\mathcal{X}^\sharp_t$.

(γ) A stochastic kernel q_γ in $\mathbf{K}(\mathbf{X}^\bullet_{t-1}, \{\psi\})$ that models the disappearance of individuals.

(ι) The identity kernel $q_\iota(\psi, \cdot) = \delta_\psi$ in $\mathbf{K}_1(\{\psi\}, \{\psi\})$, called the *empty kernel*.

The set $\{\alpha, \beta, \gamma, \iota\}$ that indexes all the possible kernels is denoted $\mathbb{I}_{t|t-1}$. The kernels q_β and q_γ are assumed to verify the following property:

$$(\forall \mathbf{x} \in \mathbf{X}^\bullet_{t-1}) \qquad q_\gamma(\mathbf{x}, \{\psi\}) + \int q_\beta(\mathbf{x}, \mathrm{d}\mathbf{y}) = 1, \tag{3.25}$$

since an individual at point $\mathbf{x} \in \mathbf{X}^\bullet_{t-1}$ can be either propagated to $\mathbf{X}^\bullet_t$ or can disappear and be moved to ψ.

Remark 3.2. Note the following remarks on the above-mentioned kernels:

3.2.1 All the considered kernels, here in the time prediction and later in the observation prediction, can be made time dependent. The time index will not be indicated in general for the sake of conciseness.

3.2.2 The kernels q_β, q_γ and q_ι have complimentary source and target spaces and can be expressed as a single Markov kernel $M_{t|t-1}$ from $\mathbf{X}_{t-1}$ to $\mathbf{X}_t$ characterised, for all $\mathbf{x} \in \mathbf{X}^\bullet_{t-1}$ by

$$(\forall B \in \mathcal{B}(\mathbf{X}^\bullet_t)) \qquad M_{t|t-1}(\mathbf{x}, B) = q_\beta(\mathbf{x}, B), \tag{3.26}$$

and for all $\mathbf{x} \in \mathbf{X}_{t-1}$ by

$$M_{t|t-1}(\mathbf{x}, \{\psi\}) = \mathbf{1}_{\mathbf{X}^\bullet_{t-1}}(\mathbf{x}) q_\gamma(\mathbf{x}, \{\psi\}) + \mathbf{1}_{\{\psi\}}(\mathbf{x}) q_\iota(\psi, \{\psi\}). \tag{3.27}$$

Equation (3.25) ensures that $M_{t|t-1}(\mathbf{x}, \cdot) \in \mathbf{M}_1(\mathbf{X}_t)$ for any $\mathbf{x} \in \mathbf{X}^\bullet_{t-1}$. Another way to express the kernel $M_{t|t-1}$ is to rescale the kernels defining it by introducing a measurable function f_β and a Markov kernel M_β, defined for any $\mathbf{x} \in \mathbf{X}^\bullet_{t-1}$ as

$$f_\beta(\mathbf{x}) \doteq q_\beta(\mathbf{x}, \mathbf{X}^\bullet_t), \tag{3.28a}$$

$$f_\beta(\mathbf{x}) M_\beta(\mathbf{x}, \cdot) \doteq q_\beta(\mathbf{x}, \cdot). \tag{3.28b}$$

The function $f_\beta : \mathbf{X}^\bullet_{t-1} \to [0,1]$ models the probability that an individual with state $\mathbf{x} \in \mathbf{X}^\bullet_{t-1}$ persists to time t, and the Markov kernel $M_\beta(\mathbf{x}, \cdot)$ is uniquely defined whenever $f_\beta(\mathbf{x}) \neq 0$. In consequence, note that $q_\gamma(\mathbf{x}, \{\psi\}) = 1 - f_\beta(\mathbf{x})$ and that the Markov kernel $M_{t|t-1}$ can be equivalently characterised for any $\mathbf{x} \in \mathbf{X}^\bullet_{t-1}$ by

$$(\forall B \in \mathcal{B}(\mathbf{X}^\bullet_t)) \qquad M_{t|t-1}(\mathbf{x}, B) = f_\beta(\mathbf{x}) M_\beta(\mathbf{x}, B), \tag{3.29}$$

and for any $\mathbf{x} \in \mathbf{X}_{t-1}$ by

$$M_{t|t-1}(\mathbf{x}, \{\psi\}) = \mathbf{1}_{\mathbf{X}^\bullet_{t-1}}(\mathbf{x})(1 - f_\beta(\mathbf{x})) + \mathbf{1}_{\{\psi\}}(\mathbf{x})\delta_\psi. \tag{3.30}$$

The kernel $M_{t|t-1}$ can be used for all the individuals that appeared already, but does not allow for distinguishing persisting and disappeared individuals.

3.2.3 It is meaningful to model the birth of individuals as a kernel since what is really modelled is the number of individuals that appeared *between* time $t-1$ and t. For instance, if a birth rate is available, then c_α will depend on the duration of the considered time step, which is natural for a kernel-related quantity. Conversely, kernels should always model behaviours that depend either on both the source and the target spaces or on only one of the two spaces, but with a dependency on the duration of the time step. Birth can indeed be considered as a kernel, but, for instance, prior knowledge at time t should not be integrated in a kernel since this type of information only depends on the current time step.

3.2.4 Two possible generalisations of the above kernel models would be *a*) to split q_β into several kernels corresponding to several possible motion models, hence allowing for motion-based classification or for *jump Markov models* to be applied [Elliott et al., 1994, Doucet et al., 2001], and/or *b*) to allow for one individual to be associated with several kernels, hence modelling that one individual can possibly be propagated into two different independent individuals such as in cell division in biology. This type of dynamics is usually referred to as *spawning* in the target tracking literature.

For each $i \in \mathbb{I}_{t|t-1}$, we introduce a kernel constraint K_i for the stochastic kernel q_i which enables additional uncertainty on the transition performed by the individuals in $\mathcal{X}_{t-1}$. This is particularly useful for the kernel q_α that models the appearance of new individuals since, very often, the knowledge about such an event is extremely weak. In terms of multiplicity, the kernel q_ι is allowed to be used infinitely many times, so that multiplicities will be in the set

$$\mathbb{M}_{t|t-1} \doteq \{\boldsymbol{k} \in \bar{\mathbb{N}}^{\mathbb{I}_{t|t-1}} \text{ s.t. } (\forall i \in \mathbb{I}_{t|t-1} - \{\iota\}) \;\; \boldsymbol{k}_i < \infty\}. \tag{3.31}$$

Let $\mathfrak{H}_{t|t-1}$ be a stochastic representation in $\mathbf{N}(\mathbf{C}(\mathbf{X}_{t-1}, \mathbf{X}_t))$ describing the time-evolution of the considered population between times $t-1$ and t. Since only a finite number of kernels are possible, the counting measures on $\mathbf{C}(\mathbf{X}_{t-1}, \mathbf{X}_t)$ induced by $\mathfrak{H}_{t|t-1}$ will have their support in a finite set and can thus be expressed in the following form

$$\mu^{(\boldsymbol{k})}_{t|t-1} = \sum_{i \in \mathbb{I}_{t|t-1}} k_i \delta \lfloor K_i \rfloor, \tag{3.32}$$

where the indexed family $\boldsymbol{k}$ is in $\mathbb{M}_{t|t-1}$. The law $\mathbb{P}_{t|t-1}$ of the stochastic representation $\mathfrak{H}_{t|t-1}$ on $\mathbf{N}(\mathbf{C}(\mathbf{X}_{t-1}, \mathbf{X}_t))$ is then characterised by

$$\mathbb{P}_{t|t-1} = \int \boldsymbol{c}_{t|t-1}(\mathrm{d}\boldsymbol{k}) \delta \lfloor \mu^{(\boldsymbol{k})}_{t|t-1} \rfloor, \tag{3.33}$$

where $\boldsymbol{c}_{t|t-1}$ is the law on $\mathbb{M}_{t|t-1}$ describing the multiplicities of the atoms $\{K_i\}_{i \in \mathbb{I}_{t|t-1}}$. Assuming that only a probabilistic constraint $p_\alpha \in \mathbf{C}_1(\mathbb{N})$ is available for c_α and that the other cardinality distributions are unknown a priori, the law $\boldsymbol{c}_{t|t-1}$ is found to verify

$$(\forall B \in \mathcal{B}(\mathbb{M}_{t|t-1})) \qquad \boldsymbol{c}_{t|t-1}(B) \leq \int \sup_{\boldsymbol{k} \in B} (h(\boldsymbol{k}_\alpha)) p_\alpha(\mathrm{d}h). \tag{3.34}$$

The same approach can be used for observation-related stochastic representations.

For the observation prediction

Several types of kernels with complimentary source and target spaces could be defined for the observation prediction as in the previous section with time prediction. However, whereas time update does not allow for distinguishing existing and disappeared individuals, the observation update does contain enough information to make the distinction between the individuals that have been detected and the ones that have been undetected. In this situation, if there is no prior knowledge on the number of detected/undetected individuals, then an averaged kernel can be used, as in Remark 3.2.2. In consequence, two kernels are considered for the modelling of the observation of the target population at time t:

(d) A Markov kernel ℓ_{d} in $\mathbf{K}_1(\mathbf{X}^\bullet_t, \mathbf{Z}_t)$ describing the possible **d**etection of the individuals in the population $\mathcal{X}_t$,

(e) A Markov kernel $\ell_{\mathrm{e}}(\psi, \cdot) = \delta_\phi$ in $\mathbf{K}_1(\{\psi\}, \{\phi\})$ relating **e**mpty state and **e**mpty observation.

In order to model the creation of spurious observations from the individuals in the outer population, a kernel from the empty state ψ in $\mathbf{X}_t$ to the observation space $\mathbf{Z}_t$ has to be introduced. We propose two models for the creation of spurious observations, which correspond to two different statistical descriptions of the underlying mechanism.

M.1 A single Markov kernel $\ell_{\mathrm{f}} \in \mathbf{K}_1(\{\psi\}, \mathbf{Z}^\bullet_t)$ is introduced for spurious observations, also called **f**alse positives, and the uncertainty on the number of these spurious observations is translated into a law $c_{\mathrm{f}} \in \mathbf{M}_1(\mathbb{N})$ for the multiplicity of ℓ_{f}.

M.2 The sensor is understood as a *finite-resolution* sensor which can only generates observations in a countable measurable partition π'_t of $\mathbf{Z}^\bullet_t$. With each *observation cell* in π'_t is associated a unique index and the set of all these indices is denoted Z'_t. The partition π'_t induces a first sub-σ-algebra $\mathcal{A}_t$ on $\mathbf{Z}^\bullet_t$ and a second sub-σ-algebra $\bar{\mathcal{A}}_t$ on $\mathbf{Z}_t$. In this case, for each observation cell indexed by $z \in Z'_t$, a Markov kernel ℓ_z in $\mathbf{K}_1(\{\psi\}, \bar{\mathcal{A}}_t)$ is introduced such that $\ell_z(\psi, A_z) = 1$. Each kernel in $\{\ell_z \text{ s.t. } z \in Z'_t\}$ is then assumed to be used exactly once.

We consider that the set $\mathbb{I}_{t|t} \doteq \{\mathrm{d}, \mathrm{e}, \mathrm{f}\}$ indexes the kernels in $\mathfrak{H}_{t|t}$, where "f" refers to all the elements in Z'_t when the model **M.2** is considered. The set of multiplicities for the kernels indexed by $\mathbb{I}_{t|t}$ is denoted $\mathbb{M}_{t|t}$ and defined as

$$\mathbb{M}_{t|t} \doteq \{\boldsymbol{k} \in \bar{\mathbb{N}}^{\mathbb{I}_{t|t}} \text{ s.t. } (\forall i \in \{\mathrm{d}, \mathrm{f}\}) \;\; \boldsymbol{k}_i < \infty\}, \tag{3.35}$$

that is, the empty kernel ℓ_{e} is allowed to be used infinitely many times. These observation-related kernels might not be fully known and, for every $i \in \mathbb{I}_{t|t}$, we introduce a Markov constraint $K_i \in \mathbf{C}_1(\mathbf{X}_t, \mathbf{Z}_t)$ for the Markov kernel ℓ_i in order to enable a greater versatility in the modelling.

Let $\mathfrak{H}_{t|t}$ be a stochastic representation in $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_t, \mathbf{Z}_t))$ which describes the transition of the population $\mathcal{X}_t$ from the state space to the observation space at time t. The counting measures induced by $\mathfrak{H}_{t|t}$ on $\mathbf{C}_1(\mathbf{X}_t, \mathbf{Z}_t)$ take the form

$$\mu_{t|t}^{(\boldsymbol{k})} = \sum_{i \in \mathbb{I}_{t|t}} k_i \delta \lfloor K_i \rfloor, \tag{3.36}$$

where $\boldsymbol{k} = \{k_i\}_{i \in \mathbb{I}_{t|t}}$ is an indexed family in $\mathbb{M}_{t|t}$. The law $\mathbb{P}_{t|t}$ of $\mathfrak{H}_{t|t}$ can be formulated as

$$\mathbb{P}_{t|t} = \int \boldsymbol{c}_{t|t}(\mathrm{d}\boldsymbol{k}) \delta \lfloor \mu_{t|t}^{(\boldsymbol{k})} \rfloor, \tag{3.37}$$

where $\boldsymbol{c}_{t|t}$ is a law induced by $\mathfrak{H}_{t|t}$ on $\mathbb{M}_{t|t}$. Assuming that there is no prior information on the multiplicity of the kernels K_{d} and K_{e} and that only a probabilistic constraint $\boldsymbol{p}_{\mathrm{f}}$ on the multiplicity of K_{f} might be known, we find that the law $\boldsymbol{c}_{t|t}$ verifies

$$(\forall B \in \mathcal{B}(\mathbb{M}_{t|t})) \qquad \boldsymbol{c}_{t|t}(B) \leq \int \sup_{\boldsymbol{n} \in B} (h(\boldsymbol{n}_{\mathrm{f}})) \boldsymbol{p}_{\mathrm{f}}(\mathrm{d}h). \tag{3.38}$$

The probabilistic constraint $\boldsymbol{p}_{\mathrm{f}}$ on the set of families of natural integers indexed by $\mathbb{I}^{\mathrm{f}}_{t|t}$ models the number of false positives, with

M.1 $\mathbb{I}^{\mathrm{f}}_{t|t} = \{\mathrm{f}\}$ and $\boldsymbol{p}_{\mathrm{f}} \in \mathbf{C}_1(\mathbb{N})$

M.2 $\mathbb{I}^{\mathrm{f}}_{t|t} = Z'_t$, and $\boldsymbol{p}_{\mathrm{f}} \in \mathbf{C}_1(\mathbb{N}^{Z'_t})$ is such that each kernel is used exactly once.

Henceforth, the results will be stated under the model **M.2** since the false-positive-related quantity that is commonly given with a sensor is the probability of false positive per resolution cell, and global statistics are only deduced from this more fundamental sensor characteristic.

3.2.4 Classification

The term *classification* is to be understood in this context as the estimation of the probability for a given individual or for a set of indistinguishable individuals to belong to a pre-defined sub-population in the target population $\mathcal{X}^{\sharp}_t$. Examples of such pre-defined sub-populations are based on the introduction of: *a*) different dynamical models, e.g., we might be interested in estimating which individuals move according to a Brownian motion and which move in a more predictable way, such that with a constant velocity or acceleration, *b*) different observation models, so that extended individuals which tend to generate more variable observations can be singled out and tracked with a suitable uncertainty without affecting smaller individuals, and *c*) distinguishable prior knowledge at given times, as already suggested in Section 3.2.1.

We focus on the case of motion-based classification which is very often of interest and we assume that there are two sub-populations in $\mathcal{X}^{\sharp}_t$, each being described by a specific motion model.

Different models

The target population is assumed to be composed of two types of individuals being represented at time $t \in \mathbb{T}$ on one of the state spaces $\mathbf{X}_t^1$ and $\mathbf{X}_t^2$ and evolving in time according to one of the stochastic kernels

$$q_\beta^1 \in \mathbf{K}\Big(\mathbf{X}_{t-1}^1, \mathbf{X}_t^1\Big) \qquad \text{and} \qquad q_\beta^2 \in \mathbf{K}\Big(\mathbf{X}_{t-1}^2, \mathbf{X}_t^2\Big). \tag{3.39}$$

The birth of individuals in each of these sub-populations is also characterised separately via one of the two stochastic kernels $q_\alpha^{(i)} \in \mathbf{K}(\{\psi\}, \mathbf{X}_t^{(i)})$, $i = 1, 2$. The only difference with the case presented in the previous section is that the index set $\mathbb{I}_{t|t-1}$ is now defined as

$$\mathbb{I}_{t|t-1} \doteq \{\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma, \iota\}. \tag{3.40}$$

In terms of algorithmic complexity, such classification has a reasonable cost since the only additional multiplicity to consider is at birth, where the two kernels have to be tested for each appearing individual. Indeed, the kernel q_β^i can only be used for individual hypotheses that have been created with the birth kernel q_α^i, so that no additional algorithmic cost is induced by the propagation of the individual representations.

Single model with different parameters

Even if both considered sub-populations are represented on the same state space $\mathbf{X}_t$ and propagated through the same motion model q_β, some classification can still be performed. Indeed, the difference between the two sub-populations might lie in the values taken by part of their state. For instance, we might be interested in classifying the individuals which display a velocity above/below a given threshold. This type of classification can be performed via a numerical integration.

3.3 Filtering

It is now time to gather the different stochastic representations introduced in Section 3.2 into two different filtering steps: the time filtering and the observation filtering. Each of these two steps will be further divided into a prediction and an update step, resulting in the following scheme:

$$\underbrace{\text{time prediction} \to \text{time update}}_{\text{time filtering}} \longrightarrow \underbrace{\text{observation prediction} \to \text{observation update.}}_{\text{observation filtering}}$$

The filtering algorithm presented in this section encompasses many aspects of the filtering problem that are not usually considered, with the assumption that the dynamics and observation can be described by one-to-one individual kernels such as the ones introduced in Section 3.2.3. This algorithm will thus be referred to as the **bi**jective filter for **i**ndependent **s**tochastic **p**opulations, or simply as the BISP filter.

Remark. Among the existing multi-object estimation problems, two cases that would require considering one-to-many or many-to-one kernels can be underlined: *a*) the case of the estimation of extended objects [Salmond and Gordon, 1999, Koch, 2008] would require each individual in the time-updated representation to be possibly associated with several observations, and *b*) the so-called *superpositional-sensor* case, described by Thouin et al. [2011], where all the individuals are detected via a single observation which sums up all the individual signals and which requires the consideration of many-to-one observation kernels.

3.3.1 Time filtering

The objective in this section is to fuse the observation-updated stochastic representation $\mathfrak{C}_{:t-1}$ together with the stochastic representation $\mathfrak{C}_t$ of the prior knowledge at time t, both described in Section 3.2.2. This will first require $\mathfrak{C}_{:t-1}$ to be fused with the stochastic representation $\mathfrak{H}_{t|t-1}$ containing kernel-related information, before fusing the obtained representation with $\mathfrak{C}_t$. The stochastic representation resulting from this fusion will be said to be *time-updated*, and denoted $\mathfrak{C}_{\cdot t}$.

The first step is to fuse $\mathfrak{C}_{:t-1}$ with appropriate kernels from $\mathfrak{H}_{t|t-1}$ to allow for predicting the underlying individual descriptions from time $t-1$ to time t. Rather than considering the full set of joint indices $\mathbb{I}_{:t-1} \times \mathbb{I}_{t|t-1}$, we remove the pairs $(\flat, \alpha)$ and $(\circ, \alpha)$ in order to take into account Assumptions **A.5** and **A.7** respectively. The induced set of indices is defined as

$$\mathbb{I}_{t,t-1} \doteq \mathbb{I}_{:t-1} \times \mathbb{I}_{t|t-1} - (\mathbb{I}^{\circ}_{:t-1} \times \{\alpha\} \cup \{(\boldsymbol{i}^{\flat}_{:t-1}, \alpha)\}). \tag{3.41}$$

The set $\mathbb{M}_{t,t-1}$ is accordingly introduced as

$$\mathbb{M}_{t,t-1} \doteq \{\boldsymbol{m} \in \bar{\mathbb{N}}^{\mathbb{I}_{t,t-1}} \text{ s.t. } (\forall (\boldsymbol{i}, j) \in \mathbb{I}_{t,t-1} - \mathbb{I}^{\psi}_{t,t-1})\ \boldsymbol{m}_{\boldsymbol{i},j} < \infty\}, \tag{3.42}$$

where the symbol "ψ" on $\mathbb{I}_{t,t-1}$ refers to the individuals that are almost surely on ψ in both of the involved state spaces. We first state the result of the fusion of the stochastic representations $\mathfrak{C}_{:t-1}$, $\mathfrak{H}_{t|t-1}$ and $\mathfrak{C}_t$ on the set of counting measures on $\mathbf{C}_1(\mathbf{X}_{t-1} \times \mathbf{X}_t)$. The second step will then consist in reformulating the obtained stochastic representation with marginalised probabilistic constraints indexed by an appropriate set.

Lemma 3.1. *The law $\mathbb{P}_{t,t-1}$ of the time-updated stochastic representation $\mathfrak{C}_{t,t-1}$ on the space $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_{t-1} \times \mathbf{X}_t))$ is found to be*

$$\mathbb{P}_{t,t-1} = \int \boldsymbol{c}_{t,t-1}(\mathrm{d}\boldsymbol{m}) \delta\lfloor \mu^{(\boldsymbol{m})}_{t,t-1} \rfloor, \tag{3.43}$$

where the counting measure $\mu^{(\boldsymbol{m})}_{t,t-1}$ on $\mathbf{C}_1(\mathbf{X}_{t-1} \times \mathbf{X}_t)$ is of the form

$$\mu^{(\boldsymbol{m})}_{t,t-1} = \sum_{(\boldsymbol{i},j) \in \mathbb{I}_{t,t-1}} \boldsymbol{m}_{\boldsymbol{i},j} \delta\lfloor P_{\boldsymbol{i}} \star K_j \star P_t \rfloor. \tag{3.44}$$

and where the law $\boldsymbol{c}_{t,t-1}$ on $\mathbb{M}_{t,t-1}$ is bounded above by the probabilistic constraint $\boldsymbol{p}_{t,t-1}$, which verifies

$$\boldsymbol{p}_{t,t-1}(H) \propto \int \mathbf{1}_H(\Lambda_{t,t-1}(h,h')^{\dagger})\, \|\Lambda_{t,t-1}(h,h')\| \boldsymbol{p}_{:t-1}(\mathrm{d}h) \boldsymbol{p}_{t|t-1}(\mathrm{d}h') \tag{3.45}$$

for any H in $\mathcal{L}^{\infty}(\mathbb{M}_{t,t-1})$, where $\Lambda_{t,t-1}$ is a map from $\mathbf{L}^{\infty}(\mathbb{M}_{:t-1}) \times \mathbf{L}^{\infty}(\mathbb{M}_{t|t-1})$ to $\mathbf{L}^{\infty}(\mathbb{M}_{t,t-1})$ defined as

$$\Lambda_{t,t-1} : (h, h') \mapsto ((h \rtimes h') \circ \theta_{t,t-1}) \cdot w_{t,t-1}, \tag{3.46}$$

with $\theta_{t,t-1}$ the map recovering the prior multiplicities in $\mathbb{M}_{:t-1}$ and $\mathbb{M}_{t|t-1}$ from a given multiplicity in $\mathbb{M}_{t,t-1}$ and with the function $w_{t,t-1} : \mathbb{M}_{t,t-1} \to [0,1]$ defined as

$$w_{t,t-1} : \boldsymbol{m} \mapsto \prod_{(\boldsymbol{i},j)\in\mathbb{I}_{t,t-1}} \|P_{\boldsymbol{i}} \star K_j \star P_t\|^{\boldsymbol{m}_{\boldsymbol{i},j}}. \tag{3.47}$$

The next step is to ensure that the index set $\mathbb{I}_{\cdot t}$ is suitable for expressing the time updated stochastic representation $\mathfrak{C}_{\cdot t}$. We denote ν the bijection from $\mathbb{I}_{t,t-1}$ to $\mathbb{I}_{\cdot t}$ defined as

$$\nu : ((s,T,\boldsymbol{o}),k) \mapsto \begin{cases} (s, T \cup \{t\}, \boldsymbol{o}) & \text{if } k \in \{\alpha,\beta\} \\ (s,T,\boldsymbol{o}) & \text{otherwise.} \end{cases} \tag{3.48}$$

The mapping ν induces another mapping $\boldsymbol{\nu} : \boldsymbol{m} \mapsto \boldsymbol{m} \circ \nu^{-1}$ from $\mathbb{M}_{t,t-1}$ to $\mathbb{M}_{\cdot t}$, and the associated pushforwarding map for probabilistic constraint is denoted $T_{\boldsymbol{\nu}}$. Also, in order to complete the time filtering step, the probabilistic constraints on the product space $\mathbf{X}_{t-1} \times \mathbf{X}_t$ have to be marginalised. For this purpose, we introduce $\xi_{t,t-1}$ as the canonical projection map from $\mathbf{X}_{t-1} \times \mathbf{X}_t$ to $\mathbf{X}_t$ and we denote $T_{\xi_{t,t-1}}$ the corresponding pushforwarding map for probabilistic constraints.

Note that no specific index set has been defined for time-predicted representations since the prior-knowledge stochastic representation $\mathfrak{C}_t$ does not contain any distinguishable information and this set would be equal to the time-updated index set $\mathbb{I}_{\cdot t}$ as a result. Using the mapping ν, the multiplicities of the previous representations and of the kernels can be recovered from any indexed family in $\mathbb{M}_{t,t-1}$, thus allowing for the expression of the time-updated stochastic representation $\mathfrak{C}_{\cdot t}$ with the index set $\mathbb{I}_{\cdot t}$, as in the following theorem.

Theorem 3.1. *The law of the time-updated stochastic representation $\mathfrak{C}_{\cdot t}$ on the set $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_t))$ is found to be*

$$\mathbb{P}_{\cdot t} = \int \boldsymbol{c}_{\cdot t}(\mathrm{d}\boldsymbol{n})\delta\lfloor \mu_{\cdot t}^{(\boldsymbol{n})} \rfloor, \tag{3.49}$$

where the law $\boldsymbol{c}_{\cdot t} \in \mathbf{M}_1(\mathbb{M}_{\cdot t})$ is constrained by $\boldsymbol{p}_{\cdot t} = (T_{\boldsymbol{\nu}})_* \boldsymbol{p}_{t,t-1}$, and where for a given $\boldsymbol{n} \in \mathbb{M}_{\cdot t}$, the counting measure $\mu_{\cdot t}^{(\boldsymbol{n})}$ on $\mathbf{C}_1(\mathbf{X}_t)$ is such that

$$\mu_{\cdot t}^{(\boldsymbol{n})} = \sum_{\boldsymbol{i}\in\mathbb{I}_{\cdot t}} \boldsymbol{n}_{\boldsymbol{i}} \delta\lfloor P_{\boldsymbol{i}} \rfloor, \tag{3.50}$$

with $P_{\boldsymbol{i}} = (T_{\xi_{t,t-1}})_*(P_{\boldsymbol{j}} \star K_{j'} \star P_t)$, where $(\boldsymbol{j}, j') \doteq \nu^{-1}(\boldsymbol{i})$.

A straightforward but important consequence of Theorem 3.1 is concerned with the level of generality of the form proposed in Section 3.2.2 for the observation-updated stochastic representation and is expressed in the following corollary.

Corollary 3.1. *The form given in* (3.21) *is closed form under time filtering.*

We can now proceed to the observation prediction and update, in order to integrate the observations received from the sensor and verify that the result of Corollary 3.1 propagates to these additional steps.

3.3.2 Observation filtering

Even though the observation prediction relies on the same mechanisms as its time-related counterpart, the observation update will prove to be more sophisticated, as it can be anticipated when comparing the complexity of the stochastic representations $\mathfrak{C}_t$ and $\mathfrak{C}'_t$. Considering the product between the index sets $\mathbb{I}_{\cdot t}$ and $\mathbb{I}_{t|t}$, with the exception of the associations $(\circ, \mathrm{f})$ and $(\boldsymbol{\psi}, \mathrm{f})$ which are not considered under Assumption **A.8**, it appears that there is no uncertainty on which kernel to apply to which individual because of the source spaces of the kernels in $\mathbb{I}_{t|t}$. Also, the individuals propagated to $\mathbf{Z}_t$ via the kernel K_e cannot be identified with actual observations so that, overall, the observation filtering step is based on the index set $\mathbb{I}_{t,t}$ defined by

$$\mathbb{I}^\sharp_{t,t} \doteq \Big\{(\boldsymbol{i}, \mathrm{d}, z) \text{ s.t. } \boldsymbol{i} \in \mathbb{I}^\bullet_{\cdot t},\ \ z \in \bar{Z}_t\Big\} \cup \Big\{(\boldsymbol{i}, e, \phi) \text{ s.t. } \boldsymbol{i} \in \mathbb{I}^\circ_{\cdot t}\Big\} \cup \Big\{(\boldsymbol{i}^\psi_{\cdot t}, e, \phi)\Big\}, \qquad (3.51)$$

and by

$$\mathbb{I}^\flat_{t,t} \doteq \Big\{(\boldsymbol{i}^\flat_{\cdot t}, z, z) \text{ s.t. } z \in Z_t\Big\} \cup \Big\{(\boldsymbol{i}^\flat_{\cdot t}, z, \phi) \text{ s.t. } z \in Z'_t - Z_t\Big\}. \qquad (3.52)$$

The set $\mathbb{M}_{t,t}$ is accordingly introduced as

$$\mathbb{M}_{t,t} \doteq \{\boldsymbol{m} \in \bar{\mathbb{N}}^{\mathbb{I}_{t,t}} \text{ s.t. } (\forall (\boldsymbol{i}, j, z) \in \mathbb{I}_{t,t} - \mathbb{I}^\psi_{t,t})\ \ \boldsymbol{m}_{\boldsymbol{i},j,z} < \infty\}, \qquad (3.53)$$

where the symbol "ψ" on $\mathbb{I}_{t,t}$ refers to the individuals that are almost surely on ψ and on ϕ in the involved state and observation spaces. We can now proceed to the observation filtering step as in the following lemma.

Lemma 3.2. *The law $\mathbb{P}_{t,t}$ of the observation-predicted stochastic representation $\mathfrak{C}_{t,t}$ on $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_t \times \mathbf{Z}_t))$ is found to be*

$$\mathbb{P}_{t,t} = \int \boldsymbol{c}_{t,t}(\mathrm{d}\boldsymbol{m}) \delta\lfloor \mu^{(\boldsymbol{m})}_{t,t} \rfloor, \qquad (3.54)$$

where the counting measure $\mu^{(\boldsymbol{m})}_{t,t}$ on $\mathbf{C}_1(\mathbf{X}_t \times \mathbf{Z}_t)$ is of the form

$$\mu^{(\boldsymbol{m})}_{t,t} = \sum_{(\boldsymbol{i},j,z) \in \mathbb{I}_{t,t}} \boldsymbol{m}_{\boldsymbol{i},j,z} \delta\lfloor P_{\boldsymbol{i}} \star K_j \star \delta_{f^z_t} \rfloor, \qquad (3.55)$$

and where the law $\boldsymbol{c}_{t,t} \in \mathbf{M}_1(\mathbb{M}_{t,t})$ of the multiplicities in $\mathfrak{C}_{t,t}$ is bounded above by the probabilistic constraint $\boldsymbol{p}_{t,t}$, which verifies

$$\boldsymbol{p}_{t,t}(H) \propto \int \mathbf{1}_H(\Lambda_{t,t}(h, h', h'')^\dagger)\, \|\Lambda_{t,t}(h, h', h'')\| \boldsymbol{p}_{\cdot t}(\mathrm{d}h) \boldsymbol{p}_{t|t}(\mathrm{d}h') \boldsymbol{p}'_t(\mathrm{d}h'') \qquad (3.56)$$

for any H in $\mathcal{L}^\infty(\mathbb{M}_{t,t})$, where $\Lambda_{t,t}$ is a map from $\mathbf{L}^\infty(\mathbb{M}_{\cdot t}) \times \mathbf{L}^\infty(\mathbb{M}_{t|t}) \times \mathbf{L}^\infty(\mathbb{M}'_t)$ to $\mathbf{L}^\infty(\mathbb{M}_{t,t})$ defined as

$$\Lambda_{t,t} : (h, h', h'') \mapsto ((h \rtimes h' \rtimes h'') \circ \theta_{t,t}) \cdot w_{t,t}, \qquad (3.57)$$

with $\theta_{t,t}$ the map recovering the prior multiplicities in $\mathbb{M}_{\cdot t}$, $\mathbb{M}_{t|t}$ and $\mathbb{M}'_t$ from a given multiplicity in $\mathbb{M}_{t,t}$ and where the function $w_{t,t} : \mathbb{M}_{t,t} \to [0, 1]$ is defined as

$$w_{t,t} : \boldsymbol{m} \mapsto \prod_{(\boldsymbol{i},j,z) \in \mathbb{I}_{t,t}} \|P_{\boldsymbol{i}} \star K_j \star \delta_{f^z_t}\|^{\boldsymbol{m}_{\boldsymbol{i},j,z}}. \qquad (3.58)$$

Remark. Some detection algorithms provide a probability for an observation to be associated with the same individual as one of the previous observations, using image feature recognition for instance, and it is tempting to incorporate this information into the compatibility expressed via $w_{t,t}$ as an a priori information on the association. However, this integration can only be performed when the independence of the observations can be guaranteed, e.g., if the estimation algorithm does not already use this kind of information and if the feature recognition algorithm does not require more than the last collection of observations.

As with the time-filtering step, we introduce $\xi_{t,t}$ as the canonical projection map from the product space $\mathbf{X}_t \times \mathbf{Z}_t$ to $\mathbf{X}_t$ which enables the observation filtering step to be completed by integrating the information contained in the observations into the corresponding individual representations. The associated pushforwarding map for probabilistic constraints in the set $\mathbf{C}_1(\mathbf{X}_t \times \mathbf{Z}_t)$ is denoted $T_{\xi_{t,t}}$.

We are now in a position to express the result of the fusion of the stochastic representations $\mathfrak{C}_{\cdot t}$ and $\mathfrak{C}'_t$. Unlike the time filtering step, it is the update that is central to the observation filtering part. This can be related to the fact that time filtering is usually just composed of a prediction while observation filtering focuses on the update part. However, the filtering algorithm presented here is only an example that encompasses standard filtering, and other assumptions might lead to an algorithm that focuses mainly on time update and on observation prediction.

We denote ς the mapping from $\mathbb{I}_{t,t}$ to $\mathbb{I}_{:t}$ defined as

$$\varsigma : ((s, T, \boldsymbol{o}), k, z) \mapsto \begin{cases} (\sharp, T, \boldsymbol{o} \times z) & \text{if } s = \sharp \\ \boldsymbol{i}^{\flat}_{:t} & \text{if } s = \flat, \end{cases} \tag{3.59}$$

and we introduce the map $\boldsymbol{\varsigma} : \boldsymbol{m} \mapsto \boldsymbol{m} \circ \varsigma^{-1}$ from $\mathbb{M}_{t,t}$ to $\mathbb{M}_{:t}$ as well as the associated pushforwarding map $T_{\boldsymbol{\varsigma}}$. Notice that the restriction of ς to $\mathbb{I}^{\sharp}_{t,t}$ is bijective since only one kernel can be used for each sub-population in $\mathcal{X}^{\sharp}_t$. The observation-update step can now be formulated as follows.

Theorem 3.2. *The law $\mathbb{P}_{:t}$ on $\mathbf{N}(\mathbf{C}_1(\mathbf{X}_t))$ of the observation-updated stochastic population $\mathfrak{C}_{:t}$ is found to be*

$$\mathbb{P}_{:t} = \int \boldsymbol{c}_{:t}(\mathrm{d}\boldsymbol{n}) \delta\big\lfloor \mu^{(\boldsymbol{n})}_{:t} \big\rfloor, \tag{3.60}$$

where the law $\boldsymbol{c}_{:t} \in \mathbf{M}_1(\mathbb{M}_{:t})$ of the multiplicity in $\mathfrak{C}_{:t}$ is constrained by $\boldsymbol{p}_{:t} = (T_{\boldsymbol{\varsigma}})_* \boldsymbol{p}_{t,t}$ and where, for a given $\boldsymbol{n} \in \mathbb{M}_{:t}$, the counting measure $\mu^{(\boldsymbol{n})}_{:t}$ is of the form

$$\mu^{(\boldsymbol{n})}_{:t} = \sum_{\boldsymbol{i} \in \mathbb{I}_{:t}} \boldsymbol{n}_{\boldsymbol{i}} \delta \lfloor P_{\boldsymbol{i}} \rfloor, \tag{3.61}$$

with $P_{\boldsymbol{i}} = (T_{\xi_{t,t}})_*(P_{\boldsymbol{j}} \star K_{j'} \star \delta_{f^z_t})$ where $(\boldsymbol{j}, j', z) \doteq \varsigma^{-1}(\boldsymbol{i})$ for any $\boldsymbol{i} \in \mathbb{I}^{\sharp}_{:t}$ and with $P_{\boldsymbol{i}} = P_{\boldsymbol{j}}$ where $\boldsymbol{j} \doteq \boldsymbol{i}^{\flat}_{\cdot t}$ when $\boldsymbol{i} = \boldsymbol{i}^{\flat}_{:t}$.

The downside of the simplicity of Theorem 3.2 is that most of the associations have to be considered, resulting in a high-complexity algorithm. Indeed, except for the associations that have been discarded in (3.51) when forming the index set $\mathbb{I}_{t,t}$, all possibilities have to be treated on a case-by-case basis. A direct consequence of Theorem 3.2 and of Corollary 3.1 about the level of generality of the form proposed in Section 3.2.2 is expressed in the following corollary.

Corollary 3.2. *The form given in* (3.21) *is closed form under time and observation filtering.*

Remark. A usual concern in multi-object estimation is the integration of multiple synchronised sensors. When using stochastic representations within an "optimal" filter such as the BISP filter presented here, there are actually two equivalent ways of addressing this concern:

- By performing the observation filtering several times, integrating the sensor information in any order. The order of the sensors does not matter as long as there is no approximation, which indicates that the observation-filtering operation is commutative.
- By fusing the different observation stochastic representations together before applying the observation update on the time-updated stochastic representation. This can be seen as a property of associativity for the observation filtering.

Through the two previous sections, it has been proved that under relatively weak assumptions, it is possible to proceed to the estimation of any independent stochastic population including: *a*) the integration of indistinguishable prior knowledge on a given time step, *b*) the modelling of unbounded information such as for appearing individuals, *c*) the potential to associate several individuals with the same observation, hence allowing for *unresolved* individuals to be considered in a simple way, *d*) the possibility of estimating the probability for a given observation to be a false positive a posteriori, and *e*) the time of appearance and disappearance of individuals, enabling the estimation of lifetimes. Moreover some additional features could be added with minor rework, such as: *f*) multiple motion models such as jump Markov models, also enabling motion-based classification, *g*) independent spawning models, and *h*) non–Markovian dynamical or observation models, see [Cox, 1955, Feller, 1959] for general discussions and [Fanaswala and Krishnamurthy, 2013] for an application in the case of multi-object estimation.

3.4 The DISP filter

In spite of the multiple assumptions made throughout Sections 3.1 to 3.3, the filtering algorithm for stochastic representations introduced in the previous section has a remarkably high complexity. It is however possible to reduce this complexity without making approximations by introducing new assumptions on the detection of newborn individuals as well as on the law $\boldsymbol{c}'_t$, which controls the number of individuals represented by the received observations. The obtained filtering algorithm is called the DISP filter, which stands for filter for **d**istinguishable and **i**ndependent **s**tochastic **p**opulations.

Remark 3.3. The use of probability measures, especially on uncountable sets, can be seen as a limit case of the actual uncertainty since this degree of accuracy might never be reached in practice. The consequence of such a consideration is that representations based on probability measures should not be fused together, so that at least one representation in every considered fusion should be sufficiently uninformative. This aspect will also be of importance in Section 3.5 where performance assessment is studied.

3.4.1 Time filtering

Starting from the modelling choices and results of Sections 3.2.3 and 3.3.1, we consider some modifications and additional assumptions in order to express the time update in a simpler form.

First, we assume to know the Markov kernel q_α, which models the birth of an individual in the state space $\mathbf{X}_t$. The number of individuals possibly present at the beginning of the scenario is left unbounded and we obtain as a consequence that the number of individuals with symbol $\boldsymbol{\psi}$ is infinite, i.e., the probability measure $P_{\boldsymbol{\psi}}$ on the set

$$\mathbf{X}_t^{\times} \doteq \{\psi_\infty\} \cup \bigcup_{k \geq 1} \mathbf{X}_t^k, \tag{3.62}$$

which describes the individuals with index $\boldsymbol{\psi}$ is defined as $P_{\boldsymbol{\psi}} = \delta_{\psi_\infty}$. Under these conditions, it is no longer possible to limit the overall number of newborn individuals a priori. Since the law of the individuals with indices $\boldsymbol{\psi}$ and $\flat$ will not change in time, these individuals are not taken into account in the equations of the DISP filter. In addition, only the averaged kernel $M_{t|t-1}$ introduced in Remark 3.2.2 is considered for the propagation of existing individuals, so that a direct estimation of the lifetime of the individuals is not possible anymore. This kind of information can however be partially recovered as part of a post-processing algorithm. One of the consequences is that it is no longer necessary to distinguish representations of individuals that do not exist anymore, and the use of the symbol $\circ$ is not required.

Remark. Although there is no distinct death kernel, some representations might have all their mass at point ψ, and the symbol $\circ$ could be used to refer to these representations. However, this kind of considerations can be viewed as approximations since a non-zero threshold has to be set in order to distinguish representations that have a sufficient amount of probability mass in the state space and others.

Similarly, one can choose to average the representations for newborn and undetected individuals at time $t-1$, together with the newborn individuals at time t, in order to obtain a single representation for individuals that have never been detected. In this way, the only thing that is known about the time of birth of an individual is that it is less or equal to the time of first detection. Additionally, the symbol $\bullet$ is further subdivided using the observation paths, by distinguishing the individuals that: (m) have already been observed, with "m" for **m**easured, and (u) have never been observed, with "u" for **u**nobserved.

According to the above assumptions and notations, the observation-updated index set $\mathbb{I}_{:t-1}$ is simplified as follows:

a) through averaging, the indices of undetected individuals become sets of indices of the form

$$\{(\sharp, [t', t-1], \boldsymbol{\phi}_{t-1}) \text{ s.t. } 0 \leq t' \leq t-1\}, \tag{3.63}$$

and the only information left is that the associated representation corresponds to individuals that do exist at time t but have never been detected. The index that is used for these individuals is then expressed as $\boldsymbol{i}^{\mathrm{u}}_{:t-1} \doteq (\sharp, [\cdot, t-1], \boldsymbol{\phi}_{t-1})$, where $[\cdot, t-1]$ expresses the fact that the time of birth is unknown.

b) for the same reasons, a detected individual with observation path $\boldsymbol{o} \in \mathbb{O}_{t-1}$ is indexed by a set of the form

$$\{(\sharp, [t', t-1], \boldsymbol{o}) \text{ s.t. } 0 \leq t' \leq t-1\}, \tag{3.64}$$

in which there is no information about the existence before the first detection, so that this individual will now have $(\sharp, [\cdot, t], \boldsymbol{o})$ as index.

The considered DISP observation-updated index set is $\tilde{\mathbb{I}}_{:t-1} = \tilde{\mathbb{I}}^{\mathrm{m}}_{:t-1} \cup \{\boldsymbol{i}^{\mathrm{u}}_{:t-1}\}$, with

$$\tilde{\mathbb{I}}^{\mathrm{m}}_{:t-1} = \{(\sharp, [\cdot, t-1], \boldsymbol{o}) \text{ s.t. } \boldsymbol{o} \in \mathbb{O}_{t-1}\}. \tag{3.65}$$

The index set $\tilde{\mathbb{I}}_{\cdot t}$ is obtained in the same way. We assume that a given population hypothesis $I \subseteq \mathbb{I}_{:t-1}$ cannot have the same non-empty observation in any two of its observation paths, and we denote $\tilde{\wp}(\tilde{\mathbb{I}}_{:t-1})$ the set of subsets of indices with pairwise everywhere-disjoint observation paths. The reason for this assumption will be explained in the next section. Applying the idea of Remark 3.1 and using the fact that the detected individuals are distinguishable, we introduce the sets

$$\mathbf{X}_I^{(n)} \doteq \{X \in (\mathbf{X}_{t-1}^{\times})^{I \cup \{\boldsymbol{i}_{:t-1}^{\mathrm{u}}\}} \text{ s.t. } (\forall \boldsymbol{i} \in I) \;\; X_{\boldsymbol{i}} \in \mathbf{X}_{t-1}, \;\; X_{\boldsymbol{i}_{:t-1}^{\mathrm{u}}} = \mathbf{X}_{t-1}^{n}\} \tag{3.66a}$$

$$\mathbf{X}_I^{(0)} \doteq \mathbf{X}_{t-1}^{I} \tag{3.66b}$$

for any $I \in \tilde{\wp}(\tilde{\mathbb{I}}_{:t-1}^{\mathrm{m}})$ and any $n \in \mathbb{N}^*$. It then appears that the laws of interest can be expressed on the set

$$\tilde{\mathbf{X}}_{:t-1}^{\times} \doteq \bigcup_{I \in \tilde{\wp}(\tilde{\mathbb{I}}_{:t-1}^{\mathrm{m}})} \Big[\bigcup_{n \in \mathbb{N}} \mathbf{X}_I^{(n)} \Big]. \tag{3.67}$$

The law of the observation-updated stochastic representation $\mathfrak{C}_{:t-1}$ is assumed to be characterised by the conditionals

$$(\forall B \in \mathcal{B}(\mathbf{X}_I^{(n)})) \qquad P_{:t-1}(B \,|\, I, n) = P_{:t-1}^{\mathrm{u}}(B_{\mathrm{u}} \,|\, I, n) \prod_{\boldsymbol{i} \in I} p_{:t-1}^{(\boldsymbol{i})}(B_{\boldsymbol{i}}) \tag{3.68}$$

for any $n \in \mathbb{N}$ and any subset I in the set $\tilde{\wp}(\tilde{\mathbb{I}}_{:t-1}^{\mathrm{m}})$, and by the corresponding probability mass function $w_{:t-1}$ on $\tilde{\wp}(\tilde{\mathbb{I}}_{:t-1}^{\mathrm{m}}) \times \mathbb{N}$ indicating the probability to be in a given subset $\mathbf{X}_I^{(n)}$. The time update can then be expressed as in the following corollary, where $\tilde{\mathbf{X}}_{\cdot t}^{\times}$ is the equivalent of $\tilde{\mathbf{X}}_{:t-1}^{\times}$ for the index set $\tilde{\mathbb{I}}_{\cdot t}$.

Corollary 3.3. *The law $P_{\cdot t}$ of the time-updated stochastic representation $\mathfrak{C}_{\cdot t}$ on $\tilde{\mathbf{X}}_{\cdot t}^{\times}$ is characterised by the conditionals*

$$(\forall B \in \mathcal{B}(\mathbf{X}_I^{(n)})) \qquad P_{\cdot t}(B \,|\, I, n) = P_{\cdot t}^{\mathrm{u}}(B_{\mathrm{u}} \,|\, I, n) \prod_{\boldsymbol{i} \in I} p_{\cdot t}^{(\boldsymbol{i})}(B_{\boldsymbol{i}}) \tag{3.69}$$

for any $n \in \mathbb{N}$ and any subset I in the set $\tilde{\wp}(\tilde{\mathbb{I}}_{\cdot t}^{\mathrm{m}})$, and by the corresponding probability mass function $w_{\cdot t}$ on $\tilde{\wp}(\tilde{\mathbb{I}}_{\cdot t}^{\mathrm{m}}) \times \mathbb{N}$, where the probability measures $p_{\cdot t}^{(\boldsymbol{i})}$ have been propagated with the Markov kernel $M_{t|t-1}$ and updated with the prior knowledge f_t, and where

$$w_{\cdot t}(I, n) = \sum_{k+k'=n} w_{:t-1}(I, k) c_{\alpha}(k') \tag{3.70a}$$

$$P_{\cdot t}^{\mathrm{u}}(B_{\mathrm{u}} \,|\, I, n) \propto \sum_{k+k'=n} w_{:t-1}(I, k) c_{\alpha}(k') \tilde{P}_{\cdot t}^{\mathrm{u}}(B_{\mathrm{u}} \,|\, k, k') \tag{3.70b}$$

for any $I \in \tilde{\wp}(\tilde{\mathbb{I}}_{\cdot t}^{\mathrm{m}})$ and any $n \in \mathbb{N}$, where $\tilde{P}_{\cdot t}^{\mathrm{u}}(\cdot \,|\, k, k') \in \mathbf{M}_1(\mathbf{X}_t^{k+k'})$ is the symmetrical probability measure based on k' instances of $q_{\alpha}(\psi, \cdot)$ and on k instances of the time-updated law $p_{\cdot t}^{\mathrm{u}} \in \mathbf{M}_1(\mathbf{X}_t)$ of the unobserved individuals.

Except for the propagation of the observation-updated individual laws at time $t-1$, the only operation that is performed in the time filtering is the mixing of the previously undetected individuals with the newly appeared ones.

3.4.2 Observation filtering

As with the time update, we make some additional assumptions on the observation update in order to simplify the expressions obtained in Sections 3.2.3 and 3.3.2.

The main simplification is obtained by assuming that the observations received from the sensor can only represent one individual. This assumption not only simplifies the expression of the observation-related stochastic representation $\mathfrak{C}'_t$, but also implies that all the detected individuals are distinguishable, hence the name of the filter. Also, to maintain a fully probabilistic algorithm, it is assumed that there is a uniform prior on the data association, enabling the direct characterisation of the posterior probability mass function $w_{:t}$ as described in Remark 1.2.3. Finally, since only probability measures are considered for representations, and since there is no distinction between individuals in $\mathbf{X}^{\bullet}_t$ and at ψ, the extension of the observation kernel ℓ_{d} to $\mathbf{X}_t \times \mathcal{B}(\mathbf{Z}_t)$ is obtained by joining the kernels ℓ_{d} and ℓ_{e}.

We also consider the following notations:

a) for any $\boldsymbol{k} \doteq (s, T, \boldsymbol{o}) \in \tilde{\mathbb{I}}_{\cdot t}$ and any $z \in \bar{Z}_t$, let $\boldsymbol{k} \cdot z \in \tilde{\mathbb{I}}_{:t}$ represent the index $(s, T, \boldsymbol{o} \times z)$

b) the symbol "d" is associated to the individuals that have been newly **d**etected at time t, and the subset $\tilde{\mathbb{I}}^{\mathrm{d}}_{:t}$ of $\tilde{\mathbb{I}}^{\mathrm{m}}_{:t}$ is found to be

$$\tilde{\mathbb{I}}^{\mathrm{d}}_{:t} = \{\boldsymbol{i} \in \tilde{\mathbb{I}}^{\mathrm{m}}_{:t} \text{ s.t. } (\exists z \in Z_t) \;\; \boldsymbol{i} = \boldsymbol{i}^{\mathrm{u}}_{\cdot t} \cdot z\}. \tag{3.71}$$

The results presented in Section 3.3.2 can now be formulated on the space $\tilde{\mathbf{X}}^{\times}_{:t}$ as follows. We first proceed to the observation filtering of the predicted laws of individuals in the target population as

$$(\forall \boldsymbol{i} \in \tilde{\mathbb{I}}^{\sharp}_t, \forall z \in \bar{Z}_t) \qquad p^{(\boldsymbol{i}\cdot z)}_{:t} = \Psi_{\ell_{\mathrm{d}}(\cdot, A^z_t)}(p^{(\boldsymbol{i})}_{\cdot t}). \tag{3.72}$$

The individuals in the outer population do not change during the observation filtering but induce a weighting function $w^{\flat}_t : \tilde{\wp}(\tilde{\mathbb{I}}^{\mathrm{m}}_{:t}) \to [0,1]$ defined as

$$w^{\flat}_t : I \mapsto \Bigg[\prod_{z \in Z_t - Z_I} \ell_z(\psi, A^z_t)\Bigg]\Bigg[\prod_{z \in Z'_t - (Z_t - Z_I)} \ell_z(\psi, \{\phi\})\Bigg], \tag{3.73}$$

where $Z_I \subseteq Z_t$ is the subset of observations assigned to $\mathcal{X}^{\sharp}_t$ in the subset I. With these preliminary results, we are in position to state the observation filtering for the DISP filter.

Corollary 3.4. *The law $P_{:t}$ of the observation-updated stochastic representation $\mathfrak{C}_{:t}$ on $\tilde{\mathbf{X}}^{\times}_{:t}$ is characterised by the conditionals*

$$(\forall B \in \mathcal{B}(\mathbf{X}^{(n)}_I)) \qquad P_{:t}(B \,|\, I, n) = P^{\mathrm{u}}_{:t}(B_{\mathrm{u}} \,|\, I, n) \prod_{\boldsymbol{i} \in I} p^{(\boldsymbol{i})}_{:t}(B_{\boldsymbol{i}}) \tag{3.74}$$

for any $n \in \mathbb{N}$ and any $I \in \tilde{\wp}(\tilde{\mathbb{I}}^{\mathrm{m}}_{:t})$, and by the corresponding probability mass function $w_{:t}$ on $\tilde{\wp}(\tilde{\mathbb{I}}^{\mathrm{m}}_{:t}) \times \mathbb{N}$ expressed as

$$w_{:t}(I, n) \propto w^{\flat}_t(I) w_{\cdot t}(J, n + |I_{\mathrm{d}}|) \prod_{\boldsymbol{i}\cdot z \in I} p^{(\boldsymbol{i})}_{\cdot t}(\ell_t(\cdot, A^z_t)), \tag{3.75}$$

where J is the subset in $\tilde{\wp}(\tilde{\mathbb{I}}^{\mathrm{m}}_{\cdot t})$ from which I originates, i.e.,

$$J = \{\boldsymbol{i} \in \tilde{\mathbb{I}}^{\mathrm{m}}_{\cdot t} \textit{ s.t. } (\exists z \in \bar{Z}_t) \;\; \boldsymbol{i} \cdot z \in I\}. \tag{3.76}$$

The approach considered here for time and observation filtering is a consequence of the more general filtering equations derived in Section 3.3. In consequence, the mechanisms behind the two considered filtering steps do not have to be given explicitly.

3.4.3 Implementation

Even with the simplifications introduced in the previous sections, the complexity of the DISP filter remains high and approximations need to be applied to scale down the underlying combinatorial explosion. The most straightforward of these approximations is to apply pruning techniques by removing all the terms in $P_{\cdot t}$ and $P_{:t}$ that are not sufficiently likely to represent the target population. More details about the implementation of the DISP filter can be found in Delande et al. [2015]. Note that in this paper, the chosen birth model is slightly different.

3.4.4 Applications

The DISP filter has already been used for different applications including sensor control [Delande et al., 2014a] and space situational awareness [Frueh et al., 2015]. The main benefit obtained in sensor control when using the DISP filter is that the cost function can be designed to make the sensor focus on specific individuals in the population or, e.g., on the whole subpopulation of appearing individuals. As far as space situational awareness is concerned, the limitations in sensor coverage on space debris require estimation techniques that are both robust and responsive to the detection of appearing objects. The DISP filter, which does not rely on strong approximations, can be expected to perform well under these conditions.

3.5 Performance assessment

In this short section, we address the problem of performance assessment of estimation algorithms based on stochastic populations. For decades, assessment tools for multi-object filters were based either on single-object performance evaluation like the **r**oot **m**ean **s**quare **e**rror, or RMSE, for the estimation of a given object state, or on simple multi-object characteristics like the estimated number of objects. More recently, the OSPA distance [Schuhmacher et al., 2008] was introduced, where OSPA stands for **o**ptimal **s**ub-**p**attern **a**ssignment, and aimed at providing a global characterisation of the performance of a multi-object filter. The OSPA distance can be interpreted as the distance between the multi-object estimate and the ground truth. It requires two parameters to be set: one relates to the norm that is used in the single-object distance and the other one, called the *cut-off*, gives the arbitrary distance that is induced by inaccuracies in the estimated number of objects. Different aspects of the OSPA distance could be improved: *a*) it is difficult to interpret this distance on its own and comparison with other filters is always required, *b*) the choice of the cut-off is not always intuitive and greatly affects the obtained distance, and *c*) it does not take into account the observability of an object, e.g., a filter will be penalised for missing an object that has never been detected.

We suggest that the mechanism that has been used for time and observation filtering can be used again for performance assessment. Indeed, unlike standard filtering techniques where the state is updated given the observations, the two sources of information are treated in a completely symmetrical way in this work. Therefore, it becomes possible to fuse the outputs of two stochastic-population based filters, and the induced compatibility can be interpreted as a probability for the two filters to represent the same population. For this approach to be used as a performance assessment tool, one of the two filters has to be considered as a reference filter. Such a filter could provide an optimal estimate or even an *over-optimal* estimate. Obtaining an optimal estimate at each time step would be computationally demanding, but over-optimal estimates can be computed more easily. The most crude example of an over-optimal estimate is the ground truth itself, as in the OSPA distance, while more realistic over-optimal estimates could be found by assuming that the data association is known at every time step. In this case, the filter under assessment would not be penalised for missing an individual that is not discoverable in a given situation.

Remark. As already mentioned in Remark 3.3, representations based on probability measures will often be over optimistic in terms of quantification of the uncertainty. For instance, the compatibility between two diffused probability measures is always null. If both the filter to be assessed and the reference filter make use of probability measures to represent uncertainty, then the resulting compatibility will not be meaningful. One way to bypass this issue is to apply a Markov constraint on one or both of the considered measures in order to reduce the effect of these overly accurate representations.

Remark. It is possible to assess the performance of a filter that is not expressed with the same tools as the ones considered here by modelling the parts of the uncertainty that might be missing. For instance, the MHT provides a collection of posterior laws for the distinguished individuals but does not give a representation of the undetected ones. Yet, using the partially-indistinguishable representations introduced in this work, the output of the MHT can be completed with a fully non-informative representation for undetected individuals, hence modelling the fact that nothing is known about this aspect of the problem when using this specific estimation algorithm.

Summary

Two multi-object filters have been introduced in this chapter. The first one, referred to as the BISP filter, performs prediction and update under the main assumption that there is a one-to-one correspondence between the individuals in the stochastic representations involved. This filter allows for integrating a large range of stochastic population models, some of them being directly embedded in the filter, while others have only been mentioned. Without additional assumptions or approximations, the BISP filter would however be extremely challenging to compute except for small scale estimation problems.

The second filter, called the DISP filter, relies on stronger assumptions, the main one being that individuals are distinguished upon their first observation. The expression of the time and observation updates is also made more explicit, and their implementation becomes more straightforward as a result.

Chapter 4

The HISP filter

USING the results of the previous chapter, an additional multi-object filter, called the **h**ypothesised filter for **i**ndependent **s**tochastic **p**opulation or HISP filter, is derived and its efficiency is demonstrated on simulated data. The objective is to have recourse to some approximations in order to obtain a low-complexity filter. A first set of approximations, presented in Section 4.1, enables the expression of the filter, whereas a second set of approximations, explained in Section 4.2, allows for obtaining a complexity that is linear in the number of observations and in the number of hypotheses. We are now dealing with approximations rather than assumptions and the performance of the HISP filter needs to be verified in practice, and so is done in Section 4.3.

4.1 Derivation

Starting from the filtering equations derived for the DISP filter, the additional stages required to obtain a *local* filter that only propagates individual hypotheses are explained and expressions for the time and observation filtering steps are detailed.

4.1.1 Principle

The nature of the approximations allowing for deriving the HISP filter is twofold: *a*) to ignore, up to a certain degree, the coexistence or the non-coexistence of different individual representations, and *b*) to assess given individual associations while verifying, to a limited extent, the meaningfulness of others.

For the approximations of the nature of *a*), the state space needs to be further augmented at any time $t \in \mathbb{T}$ as $\bar{\mathbf{X}}_t \doteq \{\varphi\} \cup \mathbf{X}_t$. The additional state φ is associated with the fact that population hypotheses will turn into *individual hypotheses*, and therefore the credibility of representations needs to be assessed on an individual basis.

Approximations belonging to the category *b*) will be explained in detail later on, in Section 4.2, and are not required to obtain closed-form filtering equations.

4.1.2 Initialisation

Given the assumptions on times prior to $t = 0$ that are detailed in Section 3.2.2, the only kernel that applies from time $t_- < 0$ to time $t = 0$ is the birth kernel q_α and the empty kernel q_ι. For the HISP filter, we redefine a birth-related Markov kernel

$\hat{q}_\alpha$ in $\mathbf{K}_1(\{\psi\}, \bar{\mathbf{X}}_t)$ and we assume that the multiplicity of this kernel is known. The uncertainty on the number of appearing individuals is then induced by the amount of probability mass attributed by $\hat{q}_\alpha$ to φ and we can assume that

$$(\forall t \in \mathbb{T}, \forall B \in \mathcal{B}(\mathbf{X}_t)) \qquad q_\alpha(\psi, B) = \frac{\hat{q}_\alpha(\psi, B)}{\hat{q}_\alpha(\psi, \mathbf{X}_t)}, \tag{4.1}$$

that is, the law $q_\alpha(\psi, \cdot)$ is equal to the conditional probability measure induced by $\hat{q}_\alpha$ given that the represented individual exists. We assume that the considered number of appearing individuals $n_0^\alpha \in \mathbb{N}$ is larger than the number of observations at all times, so that all observations might correspond to newborn individuals. One possibility is to consider as many possibly appearing individuals as there are resolution cells in the sensor.

The law of $\mathfrak{C}_{\cdot 0}$ is then found to be $\hat{P}_{\cdot 0} = \hat{P}^{\mathrm{u}}_{\cdot 0}$ where $\hat{P}^{\mathrm{u}}_{\cdot 0}$ is a symmetrical probability measure based on n_0^α instances of the law $\hat{p}^{\mathrm{u}}_{\cdot 0} \in \mathbf{M}_1(\bar{\mathbf{X}}_0)$ defined as $\hat{p}^{\mathrm{u}}_{\cdot 0} \doteq \hat{q}_\alpha(\psi, \cdot)$, that is,

$$\hat{P}^{\mathrm{u}}_{\cdot 0} = (\hat{p}^{\mathrm{u}}_{\cdot 0})^{\times n_0^\alpha}. \tag{4.2}$$

4.1.3 Time filtering

The mechanisms behind the time filtering of the HISP filter are the same as the ones of the DISP filter, but the form of the law of $\mathfrak{C}_{:t-1}$ differs significantly and the involved Markov kernels have to be extended to $\bar{\mathbf{X}}_{t-1}$ and $\bar{\mathbf{X}}_t$. We assume there is a given number $n^{\mathrm{u}}_{:t-1}$ of undetected individuals at time $t-1$ after the observation update. The HISP index set is defined as

$$\hat{\mathbb{I}}_{:t-1} \doteq \tilde{\mathbb{I}}_{:t-1} \cup \{\boldsymbol{i}^{\flat}_{:t-1}\} = \tilde{\mathbb{I}}^{\mathrm{m}}_{:t-1} \cup \{\boldsymbol{i}^{\mathrm{u}}_{:t-1}, \boldsymbol{i}^{\flat}_{:t-1}\}, \tag{4.3}$$

and the set $\bar{\mathbf{X}}^{\times}_{:t-1}$ which characterises the individuals in $\hat{\mathbb{I}}^{\sharp}_{:t-1}$ is defined as

$$\bar{\mathbf{X}}^{\times}_{:t-1} \doteq \Big\{X \in (\bar{\mathbf{X}}^{\times}_{t-1})^{\hat{\mathbb{I}}^{\sharp}_{:t-1}} \text{ s.t. } (\forall \boldsymbol{i} \in \hat{\mathbb{I}}^{\mathrm{m}}_{:t-1}) \;\; X_{\boldsymbol{i}} \in \bar{\mathbf{X}}_{t-1}, \;\; X_{\boldsymbol{i}^{\mathrm{u}}_{:t-1}} \in \bar{\mathbf{X}}_{t-1}^{n^{\mathrm{u}}_{:t-1}}\Big\}. \tag{4.4}$$

The approximated law $\hat{P}_{:t-1}$ of the observation-filtered stochastic representation $\mathfrak{C}_{:t-1}$ is expressed on the set $\bar{\mathbf{X}}^{\times}_{:t-1}$ and is assumed to have the following form:

$$(\forall B \in \mathcal{B}(\bar{\mathbf{X}}^{\times}_{:t-1})) \qquad \hat{P}_{:t-1}(B) = \hat{P}^{\mathrm{u}}_{:t-1}(B_{\mathrm{u}}) \prod_{\boldsymbol{i} \in \hat{\mathbb{I}}^{\mathrm{m}}_{:t-1}} \hat{p}^{(\boldsymbol{i})}_{:t-1}(B_{\boldsymbol{i}}), \tag{4.5}$$

where $\hat{P}^{\mathrm{u}}_{:t-1}$ is a symmetrical probability measure based on the individual law $\hat{p}^{\mathrm{u}}_{:t-1}$. The form of $\hat{P}_{:t-1}$ shows that all the uncertainty is integrated at the individual level by balancing the individual probability mass between φ and $\mathbf{X}_{t-1}$. This is one of the strengths of the HISP filter since only a collection of posterior individual laws need to be considered for characterising this type of population law. The Markov kernel $M_{t|t-1}$ is extended to $\bar{\mathbf{X}}_{t-1} \times \mathcal{B}(\bar{\mathbf{X}}_t)$ by assuming that $M_{t|t-1}(\mathbf{x}, \{\varphi\}) = 0$ for any $\mathbf{x} \in \bar{\mathbf{X}}_{t-1}$ and that $M_{t|t-1}(\varphi, \{\varphi\}) = 1$. The time-filtering step can now be expressed as in the following corollary.

Corollary 4.1. *The approximated law $\hat{P}_{\cdot t}$ of the time-filtered stochastic representation $\mathfrak{C}_{\cdot t}$ is characterised by*

$$(\forall B \in \mathcal{B}(\bar{\mathbf{X}}^{\times}_{\cdot t})) \qquad \hat{P}_{\cdot t}(B) = \hat{P}^{\mathrm{u}}_{\cdot t}(B_{\mathrm{u}}) \prod_{\boldsymbol{i} \in \hat{\mathbb{I}}^{\mathrm{m}}_{\cdot t}} \hat{p}^{(\boldsymbol{i})}_{\cdot t}(B_{\boldsymbol{i}}), \tag{4.6}$$

where the law $\hat{P}^{\mathrm{u}}_{:t}$ results from the mixing of $\hat{P}^{\mathrm{u}}_{:t-1}$ and the n^{α}_{t} appearing individuals, with law $\hat{q}_{\alpha}$ describing the unobserved individuals.

As a result of the mixing of the previously unobserved individuals and the newly appeared ones, the union of these sub-populations becomes indistinguishable and the associated multiplicity is $n^{\mathrm{u}}_{:t} = n^{\mathrm{u}}_{:t-1} + n^{\alpha}_{t}$. It appears that the time-filtered law $\hat{P}_{:t}$ takes the same form as the observation-filtered law $\hat{P}_{:t-1}$ expressed in (4.5).

4.1.4 Observation filtering

Unlike the time filtering of the HISP filter detailed in the previous section, the expression of the observation-filtering step relies on additional mechanisms, when compared to the equivalent operation for the DISP filter. The index set $\hat{\mathbb{I}}_{t,t}$ is defined by $\hat{\mathbb{I}}^{\flat}_{t,t} \doteq \mathbb{I}^{\flat}_{t,t}$ and by

$$\hat{\mathbb{I}}^{\sharp}_{t,t} \doteq \{(\boldsymbol{i}, \mathrm{d}, z) \text{ s.t. } \boldsymbol{i} \in \hat{\mathbb{I}}^{\sharp}_{\cdot t},\ \ z \in \bar{Z}_t\}. \tag{4.7}$$

As in Section 3.3.2, we introduce a mapping ς from $\hat{\mathbb{I}}_{t,t}$ to $\hat{\mathbb{I}}_{:t}$ as

$$\varsigma : ((s, T, \boldsymbol{o}), k, z) \mapsto \begin{cases} (\sharp, T, \boldsymbol{o} \times z) & \text{if } s = \sharp \\ \boldsymbol{i}^{\flat}_{:t} & \text{if } s = \flat. \end{cases} \tag{4.8}$$

Also the Markov kernel ℓ_{d} and the Markov kernels ℓ_z, for all $z \in Z'_t$, are extended to $\bar{\mathbf{X}}_t \times \bar{\mathbf{Z}}_t$ by setting $\ell_s(\varphi, \{\phi\}) = 1$ for all $s \in \mathbb{I}_{t|t}$. The observation-filtering step for the HISP filter can now be expressed as follows.

Theorem 4.1. *The approximated law $\hat{P}_{:t}$ of the observation-updated stochastic representation $\mathfrak{C}_{:t}$ is characterised by*

$$(\forall B \in \mathcal{B}(\bar{\mathbf{X}}^{\times}_{:t})) \qquad \hat{P}_{:t}(B) = \hat{P}^{\mathrm{u}}_{:t}(B_{\mathrm{u}}) \prod_{\boldsymbol{i} \in \hat{\mathbb{I}}^{\mathrm{m}}_{:t}} \hat{p}^{(\boldsymbol{i})}_{:t}(B_{\boldsymbol{i}}), \tag{4.9}$$

where the marginalised posterior individual law $\hat{p}^{(\boldsymbol{i})}_{:t} \in \mathbf{M}_1(\bar{\mathbf{X}}_t)$ of the individual with index $\boldsymbol{i} \in \hat{\mathbb{I}}^{\sharp}_{:t}$ can be expressed by denoting $(\boldsymbol{k}, s, z) \doteq \varsigma^{-1}(\boldsymbol{i})$, as

$$\hat{p}^{(\boldsymbol{i})}_{:t} = \frac{w_{\mathrm{ex}}(\boldsymbol{k}, z) w^{(\boldsymbol{k},z)}_{:t}}{\sum_{z' \in \bar{Z}_t} w_{\mathrm{ex}}(\boldsymbol{k}, z') w^{(\boldsymbol{k},z')}_{:t}} \Psi_{\ell_{\mathrm{d}}(\cdot, A^z_t)}(\hat{p}^{(\boldsymbol{k})}_{\cdot t}), \tag{4.10a}$$

or, when $z \neq \phi$, as

$$\hat{p}^{(\boldsymbol{i})}_{:t} = \frac{w_{\mathrm{ex}}(\boldsymbol{k}, z) w^{(\boldsymbol{k},z)}_{:t}}{\sum_{\boldsymbol{k}' \in \hat{\mathbb{I}}_{\cdot t}} w_{\mathrm{ex}}(\boldsymbol{k}', z) w^{(\boldsymbol{k}',z)}_{:t}} \Psi_{\ell_{\mathrm{d}}(\cdot, A^z_t)}(\hat{p}^{(\boldsymbol{k})}_{\cdot t}), \tag{4.10b}$$

where, for any $(\boldsymbol{k}', s, z') \in \hat{\mathbb{I}}_{t,t}$, the real number $w^{(\boldsymbol{k}',z')}_{:t}$ is the total mass attributed to the association between $\boldsymbol{k}'$ and z' defined as

$$w^{(\boldsymbol{k}',z')}_{:t} \doteq \hat{p}^{(\boldsymbol{k}')}_{\cdot t}(\ell_s(\cdot, A^{z'}_t)) = \begin{cases} \hat{p}^{(\boldsymbol{k}')}_{\cdot t}(\ell_{\mathrm{d}}(\cdot, A^{z'}_t)) & \textit{if } \boldsymbol{k}' \in \hat{\mathbb{I}}^{\sharp}_{\cdot t} \\ \ell_{z'}(\psi, A^{z'}_t) & \textit{otherwise,} \end{cases} \tag{4.11}$$

and where the weighting function $w_{\mathrm{ex}} : \hat{\mathbb{I}}_{t,t} \to [0, 1]$ is such that: $w_{\mathrm{ex}}(\boldsymbol{k}', z')$ is the probability for the individuals with index in

$$\hat{\mathbb{I}}^{(\boldsymbol{k}')}_{\cdot t} \doteq (\hat{\mathbb{I}}^{\mathrm{m}}_{\cdot t} \setminus \{\boldsymbol{k}'\}) \cup \{\boldsymbol{i}^{\mathrm{u}}_{\cdot t}, \boldsymbol{i}^{\flat}_{\cdot t}\}, \tag{4.12}$$

and for the $n^{\mathrm{u}}_{\cdot t} - \mathbf{1}_{\boldsymbol{i}^{\mathrm{u}}_{\cdot t}}(\boldsymbol{k}')$ undetected individuals to be associated with the observations in $Z_t \setminus \{z'\}$.

Less formally, the weighting function w_{ex} at point $(\boldsymbol{k}', z')$ can be understood as the assessment of the compatibility between the time-predicted law and the collection of observations at the current time **ex**cluding the/an individual with index $\boldsymbol{k}'$ and the observation z'. Note that we consider $Z_t \setminus \{z\}$ rather than $\bar{Z}_t - \{z\}$ in the definition of w_{ex} since the empty observation ϕ might be associated with an arbitrary number of individuals, i.e., it is not because one individual is not detected that other individuals have to be detected.

Proof. Let $\mathrm{Adm}_t(\boldsymbol{k}, z)$ be the subset of $\tilde{\wp}(\hat{\mathbb{I}}_{:t})$ in which one individual with index $\boldsymbol{k} \in \hat{\mathbb{I}}_{\cdot t}$ is associated with $z \in \bar{Z}_t$. Also, let $P_{:t}$ be the DISP observation-updated law corresponding to the HISP time-updated law $\hat{P}_{\cdot t}$. The posterior marginal probability measure $\hat{p}^{(\boldsymbol{i})}_{:t} \in \mathbf{M}_1(\mathbf{X}_t)$ can be expressed as

$$(\forall C_{\boldsymbol{i}} \in \mathcal{B}(\bar{\mathbf{X}}_t)) \qquad \hat{p}^{(\boldsymbol{i})}_{:t}(C_{\boldsymbol{i}}) = \frac{1}{W_{:t}(\boldsymbol{k}, z)} \sum_{I \in \mathrm{Adm}_t(\boldsymbol{k},z)} w_{:t}(I, n^{\mathrm{u}}_{:t}) P_{:t}(B^{(\boldsymbol{i})}_I \,|\, I, n^{\mathrm{u}}_{:t}), \tag{4.13}$$

where the measurable subset $B \doteq B^{(\boldsymbol{i})}_I$ is such that $B_{\boldsymbol{i}} = C_{\boldsymbol{i}}$ and $B_{\boldsymbol{j}} = \bar{\mathbf{X}}_t$ for any $\boldsymbol{j} \in I_{\mathrm{m}} \setminus \{\boldsymbol{i}\}$, and such that $B_{\mathrm{u}} = C_{\mathrm{u}} \times (\bar{\mathbf{X}}_t)^{\times n^{\mathrm{u}}_t - 1}$ with $C_{\mathrm{u}} = \bar{\mathbf{X}}_t$ if $\boldsymbol{i} \neq \boldsymbol{i}^{\mathrm{u}}_{:t}$, i.e., only the part of $P_{:t}$ related to $\boldsymbol{i}$ is not marginalised, and where

$$W_{:t} : (\boldsymbol{k}, z) \mapsto \sum_{I \in \mathrm{Adm}_t(\boldsymbol{k},z)} w_{:t}(I, n^{\mathrm{u}}_{:t}). \tag{4.14}$$

The proof can then be divided into two parts: *a*) prove that there exists a function $w_{\mathrm{ex}} : \hat{\mathbb{I}}_{t,t} \to [0, 1]$ such that the sum in (4.13) can be factorised as in the numerators of (4.10), and *b*) prove that $P_{:t}$ can be equivalently expressed as the denominator of either (4.10a) or (4.10b). For any $\boldsymbol{i} \in \hat{\mathbb{I}}_{:t}$ with $(\boldsymbol{k}, s, z) \doteq \varsigma^{-1}(\boldsymbol{i})$, we introduce the measure $r^{(\boldsymbol{i})}_{:t} \in \mathbf{M}(\bar{\mathbf{X}}_t)$ characterised by $r^{(\boldsymbol{i})}_{:t}(\chi) = p^{(\boldsymbol{k})}_{\cdot t}(\chi \ell_s(\cdot, A^z_t))$ for any $\chi \in \mathbf{L}^0(\bar{\mathbf{X}}_t)$.

a) Expend the sum in (4.13) as follows

$$\sum_{I \in \mathrm{Adm}_t(\boldsymbol{k},z)} w_{:t}(I, n^{\mathrm{u}}_{:t}) P_{:t}(B \,|\, I, n^{\mathrm{u}}_{:t}) = r^{(\boldsymbol{i})}_{:t}(C_{\boldsymbol{i}}) \sum_{J \in \mathrm{Adm}'_t(\boldsymbol{k},z)} w_{:t}(J, n^{\mathrm{u}}_{\cdot t} - \mathbf{1}_{\boldsymbol{i}^{\mathrm{u}}_{\cdot t}}(\boldsymbol{k})), \tag{4.15}$$

where $\mathrm{Adm}'_t(\boldsymbol{k}, z)$ is the set of associations between $\hat{\mathbb{I}}^{(\boldsymbol{k})}_{\cdot t}$ and $Z_t \setminus \{z\}$. The function $w_{\mathrm{ex}} : \hat{\mathbb{I}}_{t,t} \to [0, 1]$ is then defined as

$$w_{\mathrm{ex}}(\boldsymbol{k}, z) = \sum_{J \in \mathrm{Adm}'_t(\boldsymbol{k},z)} w_{:t}(J, n^{\mathrm{u}}_{\cdot t} - \mathbf{1}_{\boldsymbol{i}^{\mathrm{u}}_{\cdot t}}(\boldsymbol{k})), \tag{4.16}$$

so that the probability measure $\hat{p}^{(\boldsymbol{i})}_{:t}$ is found to be

$$\hat{p}^{(\boldsymbol{i})}_{:t} = \frac{1}{W_{:t}(\boldsymbol{k}, z)} w_{\mathrm{ex}}(\boldsymbol{k}, z) r^{(\boldsymbol{i})}_{:t}. \tag{4.17}$$

b) It is sufficient to verify that for any $z \in \bar{Z}_t$ and any $\boldsymbol{k} \in \hat{\mathbb{I}}^{\sharp}_{\cdot t}$, it holds that

$$\bigcup_{\boldsymbol{k}' \in \hat{\mathbb{I}}_{\cdot t}} \mathrm{Adm}_t(\boldsymbol{k}', z) = \bigcup_{z' \in \bar{Z}_t} \mathrm{Adm}_t(\boldsymbol{k}, z') = \tilde{\wp}(\hat{\mathbb{I}}_{:t}) \tag{4.18}$$

since the constraint "$\boldsymbol{k}$ is associated with z" is loosen by summing over the indices in $\hat{\mathbb{I}}_{\cdot t}$ or over all observations in Z_t. We finally find that $W_{:t}$ is a constant function verifying

$$W_{:t} : (\boldsymbol{k}, z) \mapsto \sum_{I \in \tilde{\wp}(\hat{\mathbb{I}}_{:t})} w_{:t}(I, n^{\mathrm{u}}_t) = \sum_{\boldsymbol{k}' \in \hat{\mathbb{I}}_{\cdot t}} w_{\mathrm{ex}}(\boldsymbol{k}', z) w^{(\boldsymbol{k}', z)}_{:t} = \sum_{z' \in \bar{Z}_t} w_{\mathrm{ex}}(\boldsymbol{k}, z') w^{(\boldsymbol{k}, z')}_{:t}. \tag{4.19}$$

These two parts directly prove (4.10). The expression (4.9) of the modified law $\hat{P}_{:t}$ can then be deduced from the assumption of pairwise independence between all the individual laws. □

An important feature of the HISP filter can already be highlighted: an a posteriori probability of missed-detection can be computed through (4.10a) when $z = \phi$. Also, an a posteriori probability for an observation to be spurious could be obtained before marginalising when considering $\boldsymbol{k} = \boldsymbol{i}^{\flat}_{:t}$. The form of (4.10) reveals the fact that the collection of marginalised individual representations $\{\hat{p}^{(\boldsymbol{i})}_{:t}\}_{\boldsymbol{i}\in\hat{\mathbb{I}}^{\sharp}_{:t}}$ can be seen as single-object filters in interaction, where the "weight" of each single object filter is the first term on the right hand side of (4.10).

4.2 The HISP filter in practice

The HISP observation-filtering step is based on the expression of the posterior law $\hat{p}^{(\boldsymbol{i})}_t$ given the corresponding prior law, for any $\boldsymbol{i} \in \hat{\mathbb{I}}^{\sharp}_{:t}$. However, an explicit expression of the weighting function w_{ex} has not been given yet. Also, for practical reasons such as devising approximations, it is convenient to define a factorised form of w_{ex}, which can be deduced from the factorisation of the DISP observation-updated law $P_{:t}$ given in the following lemma. In this section, for the sake of compactness, we will use the symbols "u" and "♭" in place of the indices $\boldsymbol{i}^{\mathrm{u}}_{:t}$ and $\boldsymbol{i}^{\flat}_{:t}$ when there is no possible ambiguity.

Lemma 4.1. *The total mass $W_{:t} \doteq P_{:t}(\mathbf{1}) = 1$ in the probability measure $P_{:t}$ can be expressed in a factorised form as*

$$W_{:t} = C^{\phi}_t \bigg[\prod_{z\in Z_t} C^{\mathrm{u},\flat}_t(z) \bigg] \bigg[\sum_{I\in\tilde{\wp}(\hat{\mathbb{I}}^{\mathrm{m}}_{:t})} \bigg[\prod_{(\boldsymbol{k},s,z)\in\varsigma^{-1}(I_{\mathrm{d}})} \frac{w^{(\boldsymbol{k},z)}_{:t}}{w^{(\boldsymbol{k},\phi)}_{:t} C^{\mathrm{u},\flat}_t(z)} \bigg] \bigg], \tag{4.20}$$

where the constant C^{ϕ}_t is the joint probability for all the individuals in the full population $\mathcal{X}_t$ to be undetected, defined as

$$C^{\phi}_t = \Big[w^{(\mathrm{u},\phi)}_{:t} \Big]^{n^{\mathrm{u}}_{:t}} \bigg[\prod_{\boldsymbol{k}\in\hat{\mathbb{I}}^{\mathrm{m}}_{:t}} w^{(\boldsymbol{k},\phi)}_{:t} \bigg] \bigg[\prod_{z\in Z'_t} \ell_z(\psi, \{\phi\}) \bigg], \tag{4.21}$$

and where the function $C^{\mathrm{u},\flat}_t : Z_t \to [0,1]$ is defined as

$$C^{\mathrm{u},\flat}_t : z \mapsto \frac{w^{(\mathrm{u},z)}_{:t}}{w^{(\mathrm{u},\phi)}_{:t}} + v^{(\flat,z)}_{:t}, \tag{4.22}$$

with $v^{(\flat,z)}_{:t} = \ell_z(\psi, A^z_t)/\ell_z(\psi, \{\phi\})$.

Proof. The first step in proving the result is to rewrite the probability mass $w_{:t}(I, n^{\mathrm{u}}_{\hat{t}})$ in a suitable way, for any $I \in \tilde{\wp}(\hat{\mathbb{I}}_{:t})$. For this purpose let $J^{\mathrm{m}}_{\mathrm{d}}$ be the subset of $\hat{\mathbb{I}}_{:t}$ corresponding to individuals in $I^{\mathrm{m}}_{\mathrm{d}}$ that have already been measured and that have been detected at time t, let $Z^{\mathrm{m,d}}_I$ be the subset of Z_t containing the corresponding observations, let $\sigma(I) : J^{\mathrm{m}}_{\mathrm{d}} \leftrightarrow Z^{\mathrm{m,d}}_I$ be the bijection describing this identification, and let Z^{u}_I and $Z^{\flat}_I$ be the subsets of Z_t containing the observations associated with the undetected individuals and to the outer population respectively, then it holds that

$$Z_t = Z^{\mathrm{m,d}}_I \uplus Z^{\mathrm{u}}_I \uplus Z^{\flat}_I, \tag{4.23}$$

and $w_{:t}(I, n^{\mathrm{u}}_{:t})$ can be expressed as

$$w_{:t}(I, n^{\mathrm{u}}_{:t}) = C^{\phi}_t \Bigg[\prod_{(\boldsymbol{i},z)\in \mathrm{Gr}(\sigma(I))} \frac{w^{(\boldsymbol{i},z)}_{:t}}{w^{(\boldsymbol{i},\phi)}_{:t}} \Bigg] \Bigg[\prod_{z\in Z^{\mathrm{u}}_I} \frac{w^{(\mathrm{u},z)}_{:t}}{w^{(\mathrm{u},\phi)}_{:t}} \Bigg] \Bigg[\prod_{z\in Z^{\flat}_I} v^{(\flat,z)}_{:t} \Bigg]. \tag{4.24}$$

We can proceed to the second step of the proof by rewriting $W_{:t}$ as follows

$$W_{:t} = C^{\phi}_t \sum_{I\in\tilde{\wp}(\hat{\mathbb{I}}^{\mathrm{m}}_{:t})} \Bigg[\prod_{(\boldsymbol{i},z)\in \mathrm{Gr}(\sigma(I))} \frac{w^{(\boldsymbol{i},z)}_{:t}}{w^{(\boldsymbol{i},\phi)}_{:t}} \Bigg] \Bigg[\sum_{\substack{Z_{\mathrm{u}},Z_{\flat}\subseteq Z_t :\\ Z_{\mathrm{u}}\uplus Z_{\flat} = Z_t - Z^{\mathrm{d}}_I}} \Bigg[\prod_{z\in Z_{\mathrm{u}}} \frac{w^{(\mathrm{u},z)}_{:t}}{w^{(\mathrm{u},\phi)}_{:t}} \Bigg] \Bigg[\prod_{z\in Z_{\flat}} v^{(\flat,z)}_{:t} \Bigg] \Bigg], \tag{4.25}$$

where Z^{d}_I is defined as $Z^{\mathrm{m,d}}_I$ except that I only contains indices associated with previously measured individuals, so that the superscript "m" is now superfluous. We conclude by noticing that the sum over Z_{u} and $Z_{\flat}$ has a binomial form and can thus be factorised, so that

$$W_{:t} = C^{\phi}_t \sum_{I\in\tilde{\wp}(\hat{\mathbb{I}}^{\mathrm{m}}_{:t})} \Bigg[\prod_{(\boldsymbol{i},z)\in \mathrm{Gr}(\sigma(I))} \frac{w^{(\boldsymbol{i},z)}_{:t}}{w^{(\boldsymbol{i},\phi)}_{:t}} \Bigg] \Bigg[\prod_{z\in Z_t - Z^{\mathrm{d}}_I} C^{\mathrm{u},\flat}_t(z) \Bigg], \tag{4.26}$$

from which the desired result follows directly. □

Making use of the term $\mathrm{Adm}'_t(\boldsymbol{k}, z)$ defined in the proof of Theorem 4.1 as the set of associations between $\hat{\mathbb{I}}^{(\boldsymbol{k})}_{t}$ and $Z_t \setminus \{z\}$, the expressions of $W_{:t}$ and of w_{ex} can be given as a function of $w_{:t}$ in a closely related way:

$$W_{:t} = \sum_{I\in\tilde{\wp}(\hat{\mathbb{I}}^{\mathrm{m}}_{:t})} w_{:t}(I, n^{\mathrm{u}}_{:t}) \qquad \text{and} \qquad w_{\mathrm{ex}}(\boldsymbol{k}, z) = \sum_{J\in \mathrm{Adm}'_t(\boldsymbol{k},z)} w_{:t}(J, n^{\mathrm{u}}_{\cdot t} - \mathbf{1}_{\boldsymbol{i}^{\mathrm{u}}_{:t}}(\boldsymbol{k})). \tag{4.27}$$

A factorisation of the weighting function w_{ex} can then be found from Lemma 4.1 and contributes to the integration of the following hypotheses: let I and Z be subsets of $\hat{\mathbb{I}}^{\mathrm{m}}_{\cdot t}$ and Z_t respectively

H.1 for any $\boldsymbol{k} \in I$ and any $z, z' \in Z$, it holds that $w^{(\boldsymbol{k},z)}_{:t} w^{(\boldsymbol{k},z')}_{:t} \approx 0$,

H.2 for any $\boldsymbol{k}, \boldsymbol{k}' \in I$ and any $z \in Z$, it holds that $w^{(\boldsymbol{k},z)}_{:t} w^{(\boldsymbol{k}',z)}_{:t} \approx 0$.

Considering Hypothesis **H.1** for a given I and a given Z is equivalent to assuming that two observations in Z are unlikely to be associated with the same individual representation in I. Hypothesis **H.2** is the counterpart of **H.1**, for which two representations are unlikely to be associated with the same observation. These two assumptions allow for further factorising the expressions of the probability mass $W_{:t}$ and of the weighting function w_{ex}.

Proposition 4.1. *Assuming that* ***H.1*** *holds for the whole sets* $\hat{\mathbb{I}}^{\mathrm{m}}_{:t}$ *and* Z_t*, the probability mass* $W_{:t}$ *can be factorised as*

$$W_{:t} = C_t \prod_{\boldsymbol{k}\in\hat{\mathbb{I}}^{\mathrm{m}}_{\cdot t}} \Bigg[w^{(\boldsymbol{k},\phi)}_{:t} + \sum_{z\in Z_t} \frac{w^{(\boldsymbol{k},z)}_{:t}}{C^{\mathrm{u},\flat}_t(z)} \Bigg], \tag{4.28}$$

where

$$C_t = \Big[w^{(\mathrm{u},\phi)}_{:t} \Big]^{n^{\mathrm{u}}_{\cdot t}} \Bigg[\prod_{z\in Z'_t} \ell_z(\psi, \{\phi\}) \Bigg] \Bigg[\prod_{z\in Z_t} C^{\mathrm{u},\flat}_t(z) \Bigg]. \tag{4.29}$$

Also, assuming that **H.2** *holds for the sets* $\hat{\mathbb{I}}^{\mathrm{m}}_{:t}$ *and* Z_t*, the probability mass* $W_{:t}$ *can be factorised as*

$$W_{:t} = \left[\prod_{\boldsymbol{k} \in \hat{\mathbb{I}}^{\mathrm{m}}_{:t}} w_{:t}^{(\boldsymbol{k},\phi)} \right] \left[\prod_{z \in Z_t} \left[C_t^{\mathrm{u},\flat}(z) + \sum_{\boldsymbol{k} \in \hat{\mathbb{I}}^{\mathrm{m}}_{:t}} \frac{w_{:t}^{(\boldsymbol{k},z)}}{w_{:t}^{(\boldsymbol{k},\phi)}} \right] \right]. \tag{4.30}$$

Henceforth, we will focus on the Hypothesis **H.1** as results for Hypothesis **H.2** are very similar. As a consequence of Proposition 4.1, the weighting function w_{ex} can be re-expressed as follows.

Corollary 4.2. *For any* $\boldsymbol{k} \in \hat{\mathbb{I}}_{\cdot t}$ *and any* $z \in \bar{Z}_t$*, assuming that* ***H.1*** *holds for the subsets* $\hat{\mathbb{I}}^{(\boldsymbol{k})}_{\cdot t}$ *and* $Z_t \setminus \{z\}$*, the weighting function* w_{ex} *can be factorised as follows*

$$w_{\mathrm{ex}}(\boldsymbol{k}, z) = C'_t(\boldsymbol{k}, z) \prod_{\boldsymbol{k}' \in \hat{\mathbb{I}}^{\mathrm{m}}_{\cdot t} \cap \hat{\mathbb{I}}^{(\boldsymbol{k})}_{\cdot t}} \left[w_{:t}^{(\boldsymbol{k}',\phi)} + \sum_{z' \in Z_t \setminus \{z\}} \frac{w_{:t}^{(\boldsymbol{k}',z')}}{C_t^{\mathrm{u},\flat}(z')} \right], \tag{4.31}$$

where

$$C'_t(\boldsymbol{k}, z) = \left[w_{:t}^{(\mathrm{u},\phi)} \right]^{n^{\mathrm{u}}_{\cdot t} - \mathbf{1}_{\mathrm{u}}(\boldsymbol{k})} \left[\prod_{z \in Z'_t \setminus Z_\flat} \ell_z(\psi, \{\phi\}) \right] \left[\prod_{z \in Z_t \setminus \{z\}} C_t^{\mathrm{u},\flat}(z) \right], \tag{4.32}$$

with $Z_\flat$ *equal to* $\{z\}$ *when* $\boldsymbol{k} = \boldsymbol{i}^\flat_{\cdot t}$ *and* $\emptyset$ *otherwise.*

Corollary 4.2 is a direct consequence of Proposition 4.1 and does not require a separate proof. The interest in stating Proposition 4.1 first, even though only Corollary 4.2 will be useful in practice, lies in the relative simplicity brought by considering the whole population rather than a subset of it. An important property of the HISP filter that appears in Corollary 4.2 is that the whole weighting function w_{ex} can be computed with a complexity that is linear in the number of hypotheses and linear in the number of observations. This is because the values taken by w_{ex} are extremely close to each other.

4.3 Results

The HISP filter relies on assumptions and approximations that limit both the accuracy and the modelling possibilities when compared to the more general filters introduced in Section 3. However, it is shown in this section that the HISP filter still displays reasonable performance and versatility on a range of scenarios with estimation problems of different natures. In Section 4.3.1, the HISP filter is compared with the PHD filter for increasingly difficult multi-object estimation problems. The choice of the PHD filter for the comparison is motivated by the fact that both filters display the same algorithmic complexity. Then, in Section 4.3.2, the HISP filter is shown to handle a motion-based classification in a complex harbour surveillance problem. Finally, in Section 4.3.3, an implementation of the filter for non-linear observation models is shown to maintain or improve the performance of the corresponding linearised version.

4.3.1 Standard target tracking

The performance of the HISP filter is assessed and compared with the PHD filter on one scenario, for different probabilities of detection and different statistics for the spurious observations. We consider a sensor placed at the centre of the 2-D Cartesian plane that delivers range and bearing observations every 4 s during 300s. The size of the resolution cells of this sensor is $1° \times 15\,\mathrm{m}$. Considering small fixed random error and bias error, the standard deviation of the observations is $\sigma_r = 6.2\,\mathrm{m}$ and $\sigma_\theta = 4.5\,\mathrm{mrad}$ for a **s**ignal-to-**n**oise **r**atio (SNR) of 3 dB and $\sigma_r = 4.87\,\mathrm{m}$ and $\sigma_\theta = 3.5\,\mathrm{mrad}$ for a SNR of 5 dB. The range r is in $[50\,\mathrm{m}, 500\,\mathrm{m}]$ and the bearing θ is in $(-\pi, \pi]$.

The scenario comprises 5 objects with initial states $\mathbf{x}_1, \ldots, \mathbf{x}_5$, which are expressed in $[x, y, v_x, v_y]$ coordinates, with v_x and v_y the velocities in $\mathrm{m.s}^{-1}$ along the x- and y-axis, as

$$\begin{aligned} &\mathbf{x}_1 = [-400, -50, 1, 1.1], && \mathbf{x}_2 = [-50, -300, 0.4, 0.6], \\ &\mathbf{x}_3 = [50, -300, -0.4, 0.6], && \mathbf{x}_4 = [150, 150, -0.2, 0.2], \\ &\mathbf{x}_5 = [200, 300, 0.25, -1]. \end{aligned}$$

The motion of the objects is driven by a linear model in which the noise has variance $0.05\,\mathrm{m}^2.\mathrm{s}^{-4}$. We assume that the objects never spontaneously disappear. The scenario is depicted in Figure 4.1a. Note that objects 3 and 4 are crossing around $t = 120\,\mathrm{s}$.

We consider a **K**alman **f**ilter implementation of the HISP filter based on Hypothesis **H.1**, referred to as the KF-HISP filter. In this implementation, the detected and undetected hypotheses are updated through (4.10b) and (4.10a) respectively. As far as the PHD filter is concerned, we consider its Gaussian mixture implementation [Vo and Ma, 2006]. The non-linearity of the observation model is dealt with by an extended Kalman filter. To reduce the computational cost, pruning (with parameter $\tau = 10^{-5}$) and merging (with parameter $d_\mathrm{m} = 4$) are carried out on the collection of individual posterior laws as in standard mixture reduction techniques [Salmond, 1990]. For the HISP filter, merging two hypotheses means that these hypotheses are assumed to represent the same object, so that the sum of the probabilities of the merged hypotheses is limited to 1. An hypothesis is considered as "confirmed" if it has a probability of existence above $\tau_\mathrm{c} = 0.99$ or if it was previously confirmed and has a probability of existence above $\tau_\mathrm{uc} = 0.9$.

In the considered scenarios, the probability of birth is assumed to be constant across the state space $\mathbf{X}_t^\bullet$ and we denote $w_\alpha = q_\alpha(\psi, \mathbf{X}_t^\bullet)$ for any time t. The average number of spurious observations per time step is denoted $n_\flat$. The probability of detection is assumed to be constant across the state space and we denote $p_\mathrm{d} = \ell_\mathrm{d}(\mathbf{x}, \mathbf{Z}_t)$ for any $\mathbf{x} \in \mathbf{X}_t^\bullet$ and any time t. From the given characteristics of the sensor and for a given value of p_d, we deduce the probability for a single observation cell to produce a spurious observation and we denote it $w_\flat$. The approximate value of $n_\flat$ can then be deduced directly from the number of observations cells. We proceed to the performance assessment on 3 increasingly difficult scenarios.

The PHD and HISP filters have a similar computational time for the 3 considered scenarios even though the weighting function w_ex of the HISP filter has no counterpart in the PHD filter. This is the consequence of the higher discrimination capabilities of the HISP filter which relies less on the merging procedure to handle the set of hypotheses. In particular, the fact that an a posteriori probability of missed detection is computed in the HISP filter implies that the corresponding terms are often pruned while they are most often merged with another term in the PHD.

Case 1: High probability of detection (5dB)

We set $w_\alpha = 10^{-6}$ and $p_\mathrm{d} = 0.995$ so that $w_\flat = 7.67 \times 10^{-3}$ and $n_\flat \approx 83$. The OSPA distance, defined by Schuhmacher et al. [2008], is averaged over 50 Monte Carlo runs and is depicted in Figure 4.1b. Even though the estimation problem is not challenging with these parameters, there is a noticeable difference of performance between the two filters. This is mainly caused by the additional weighting function w_ex of the HISP filter which allows for a better discrimination between likely and unlikely hypotheses and which reduces the effects of association uncertainty on the overall performance. Note that the performance of the PHD filter approaches the one of the HISP filter and is even slightly higher when objects 3 and 4 cross. This can be due to the higher object state uncertainty in the PHD filter which facilitates merging.

Case 2: Low probability of detection (3dB)

We set $w_\alpha = 10^{-6}$ and $p_\mathrm{d} = 0.5$ so that $w_\flat = 1.34 \times 10^{-3}$ and $n_\flat \approx 15$. The average OSPA distance is shown in Figure 4.1c. The OSPA distance for the HISP filter is below the one of the PHD filter at all time. Due to the uncertainty on the association, the performance of the HISP filter decreases when objects 3 and 4 cross. The performance of the HISP filter in this case is mainly explained by the fact that it computes an a posteriori probability of detection, so that the prior probability, $p_\mathrm{d} = 0.5$ here, has a lower impact on the final result when compared to the PHD filter.

Case 3: High probability of false alarm (3dB)

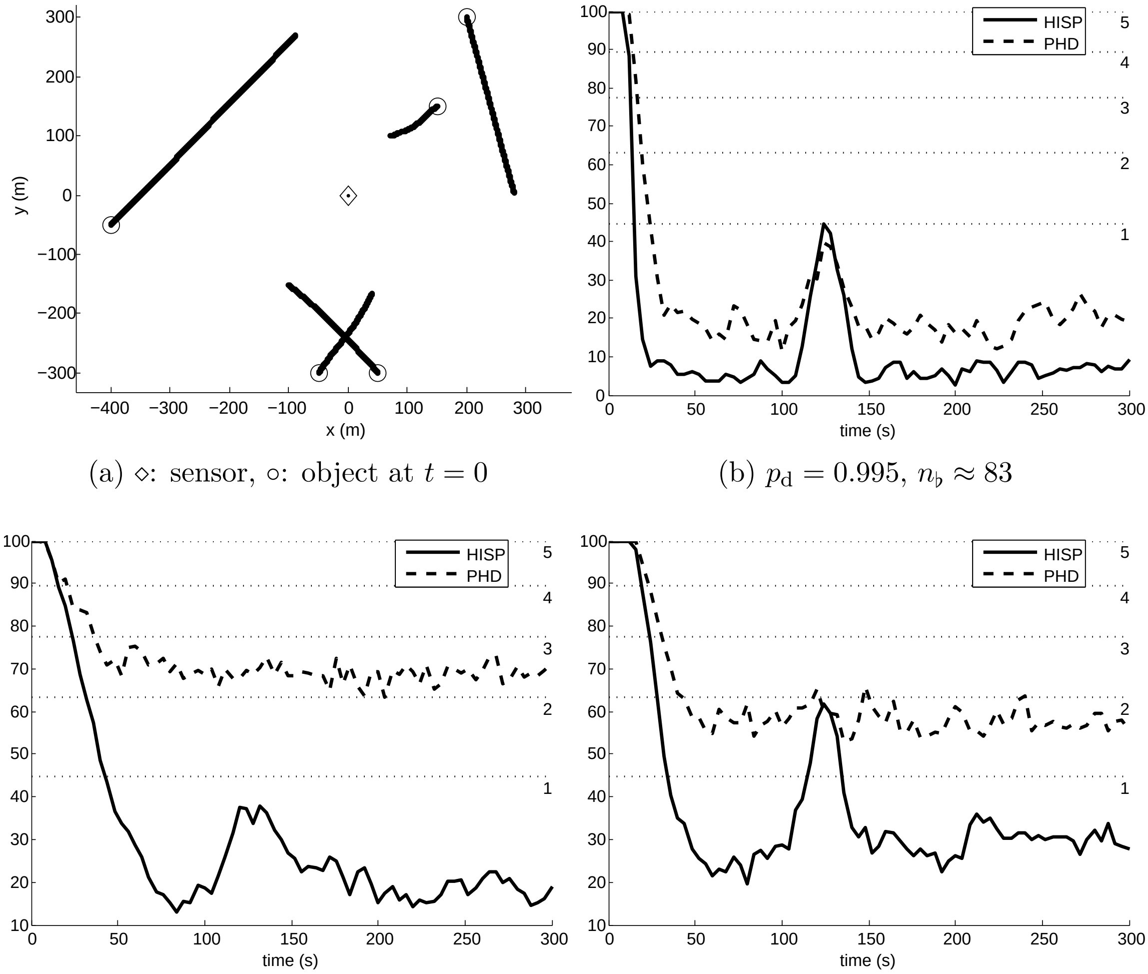
In this case, we set $w_\alpha = 5 \times 10^{-7}$ and $p_\mathrm{d} = 0.8$ so that $w_\flat = 1.54 \times 10^{-2}$ and $n_\flat \approx 167$. The average OSPA distance is shown in Figure 4.1d. The PHD filter, which is known to be robust to high numbers of spurious observations, behaves better than in Case 2 whereas the performance of the HISP filter is decreased in comparison. This behaviour highlights a fundamental difference in the assumptions and approximations that lead to the two filters: the PHD filter considers that all the individuals are indistinguishable while the HISP filter assumes that they are sufficiently separated.

4.3.2 Classification

Following the modelling introduced in Section 3.2.4, the KF-HISP filter is implemented[1] with two motion models in order to classify the target population into two sub-populations made of individuals evolving according to: *a*) a Brownian motion, and *b*) a constant velocity model. The objective is to show that classification can be performed naturally and efficiently within the multi-object filter when exploiting the concept of distinguishability.

Figure 4.2a displays the considered scenario: in a harbour environment, a restricted area is located close to a traffic area. The goal is to protect the restricted area from underwater threats. Figure 4.2b displays the geometry of the synthetic environment: 300×200 m area to survey, 15 m average depth with coarse sand or sandy mud sediment. The MIMO system is composed of 11 transmitters (Tx) located on the top and 11 receivers (Rx) located on the right, all the transducers are located at 7.5 m depth. The central frequency for the MIMO system is 30 kHz and the resolution cell 50 cm.

[1] The results from this section have been published in [Pailhas et al., 2014].



(a) $\diamond$: sensor, $\circ$: object at $t = 0$

(b) $p_{\mathrm{d}} = 0.995$, $n_{\flat} \approx 83$

(c) $p_{\mathrm{d}} = 0.5$, $n_{\flat} \approx 15$

(d) $p_{\mathrm{d}} = 0.8$, $n_{\flat} \approx 167$

Figure 4.1: OSPA distance versus time in Case 1 (b), Case 2 (c), Case 3 (d) on the scenario (a) over 50 Monte Carlo runs. HISP filter: solid line. PHD filter: dashed line. The dotted line numbered n represents the OSPA for a cardinality error of n without localisation error

Figure 4.2c displays the input to the multi-object tracker. Note that the detections have been colour-coded for display purposes only.

The probability of detection resulting from the above-described observation process is of 0.8 or more. The different types of motions have been grouped into two classes which are parametrised as follows *a*) the Brownian motion is assumed to have a noise of variance $3\,\mathrm{m}^2.\mathrm{s}^{-2}$ and *b*) the linear motion is initialised with a normal distribution on the velocity with a standard deviation of $3.5\,\mathrm{m}.\mathrm{s}^{-1}$ in each direction and considered additive noise has a variance of $0.01\mathrm{m}^2.\mathrm{s}^{-4}$.

The output of a KF-HISP filter is pictured in Figure 4.3 with two different types of seabeds: Figure 4.3b for coarse sand and Figure 4.3c for muddy sand. These figures show that the HISP filter managed to separate the fish from the other targets. This is made possible by estimating two different multi-target populations with two different dynamical models. More specifically, the coarse-sand scenario shown in Figure 4.3b has more false alarms than the muddy-sand scenario shown in Figure 4.3c. As a result, the estimation is made more difficult, e.g., the estimated positions of the fish are not as consistent as the one given for the muddy-sand scenario, the latter being closer to the ground truth.

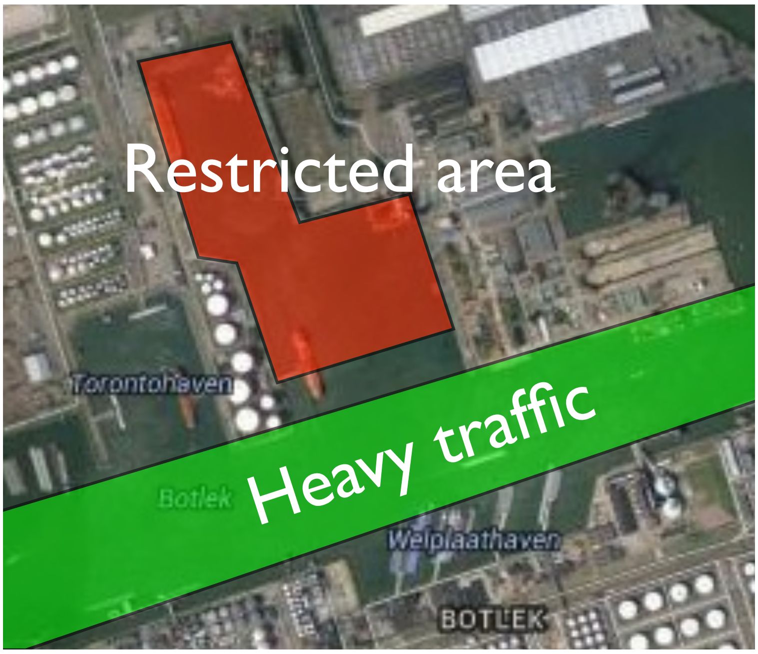
4.3.3 Finite resolution sensor

We consider[2] a finite-resolution range-bearing sensor with range between 20 and 500 meters in three different configurations of resolution cells, with a cell size of $5\,\mathrm{m} \times 1°$, $10\,\mathrm{m} \times 2°$, and $20\,\mathrm{m} \times 4°$. The sensor is located at the centre of the coordinate system. Observations are acquired synchronously every 0.1s and are generated as follows: each object is assumed to have an extension modelled by a Gaussian of standard deviation 2 m in each direction, is detected with probability 0.8, and the corresponding resolution cell is selected randomly according to the amount of probability mass that is induced in each cell by the Gaussian distribution mentioned above. An average of 5 spurious observations is triggered at each time step.

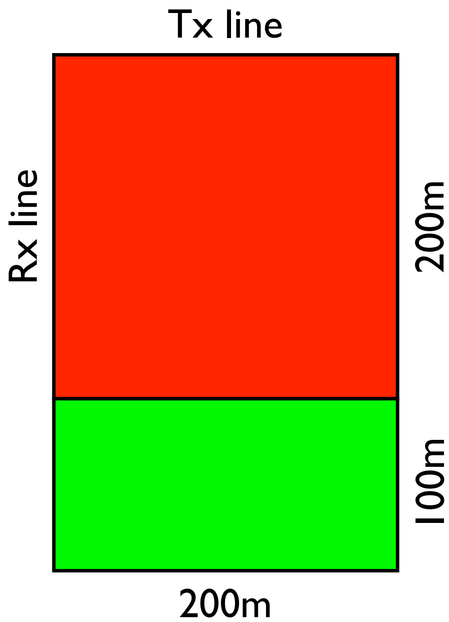
There are 10 objects in the field of view of the sensor evolving in the 2-dimensional Cartesian plane, with trajectories as shown in Figure 4.4. The motion model of these objects is described by a known and constant turn with a rate of $0.2\,\mathrm{rad}.\mathrm{s}^{-1}$ and with an additive Gaussian noise driven by a non-zero acceleration with 0 mean and standard deviation $10\,\mathrm{m}.\mathrm{s}^{-2}$. Figure 4.5 shows one of the objects with the three different resolution-cell sizes in the background.

In order to accurately represent the observation model, a **s**equential **M**onte **C**arlo (SMC) implementation is considered as in [Houssineau et al., 2015]. In this case, the approximation of the filter relies on a mixed particle association model, i.e., both the hypotheses and the law associated with each hypothesis are approximated using empirical measures [Del Moral, 2013, Pace and Del Moral, 2013]. The objective is to compare this SMC implementation of the HISP with a KF implementation. Since the considered observation model cannot be used directly in a Kalman filter, we represent resolution cells by a Gaussian centred on the cell and with a standard deviation equal to a quarter of the size of the cell in each direction. For the SMC implementation, 2500 particles are used for each hypothesis and 20 particles are used for modelling the extension. The same HISP parameters are used for both implementations, with a probability of disappearance of $1 - 10^{-4}$ and a confirmation threshold of 0.9.

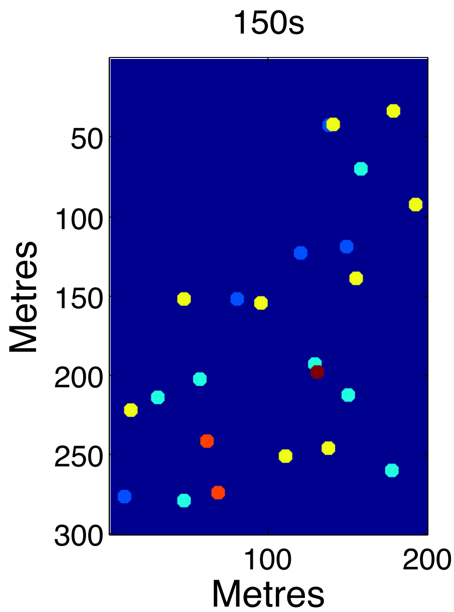
[2]The results from this section have been published in [Houssineau et al., 2015]



(a) Harbour scenario



(b) Geometry of the MIMO simulation

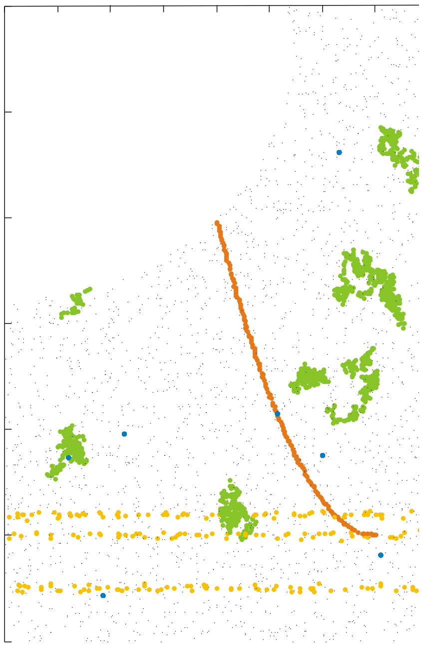


(c) Colour-coded observations (table below)

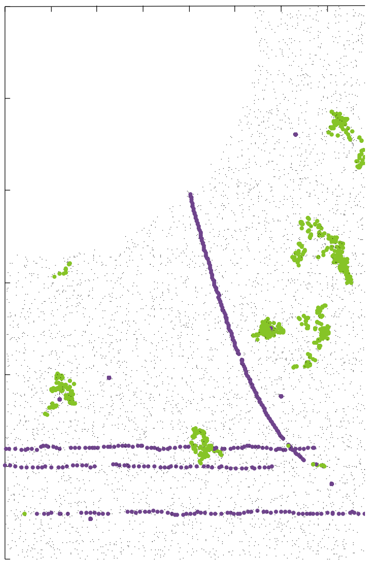
Figure 4.2: Example of real and synthetic scenarios of harbour surveillance. The sensor is a MIMO sonar with 11 transmitters (Tx) and 11 receivers (Rx)

	Fish	Static targets	Spurious observations	Boat	AUV
Observations	● (yellow)	● (cyan)	● (blue)	● (red)	● (dark red)

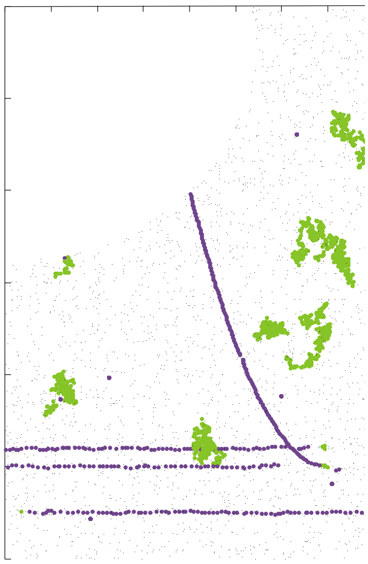
Colour code in Figure 4.2



(a) Ground truth



(b) Coarse sand



(c) Muddy sand

Figure 4.3: Accumulated view of the HISP filter's output (4.3b & 4.3c) compared with ground truth (4.3a). The color code is given in the table below

	Observations	Fish	Static targets	Boat	AUV
Ground truth	•	●	●	●	●
Estimated	•	●	●		

Colour code in Figure 4.3

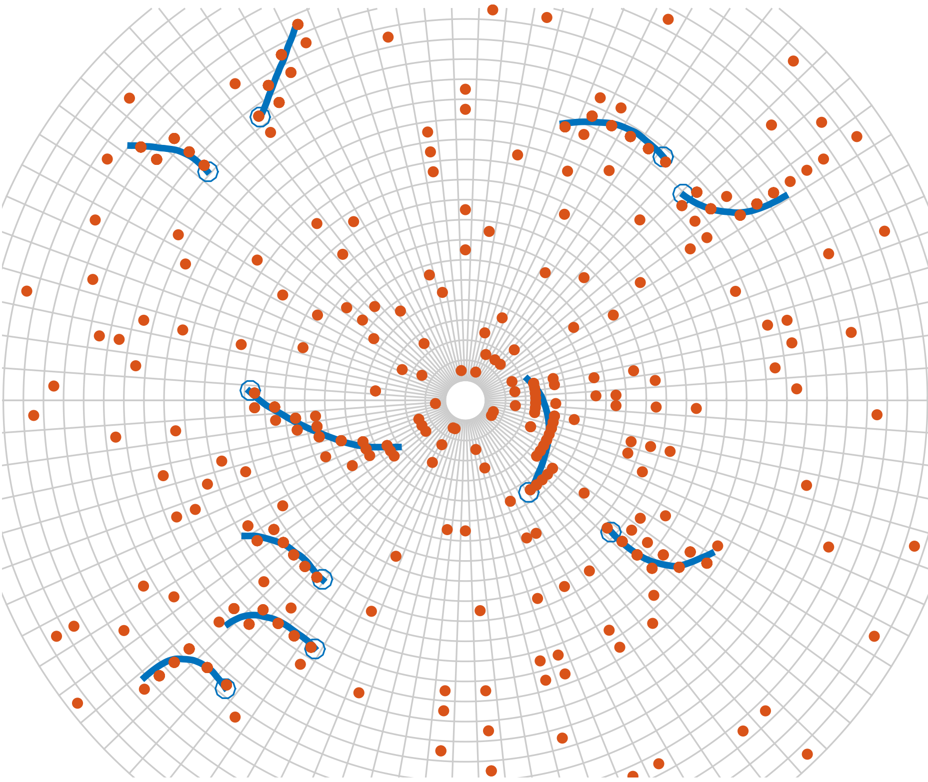
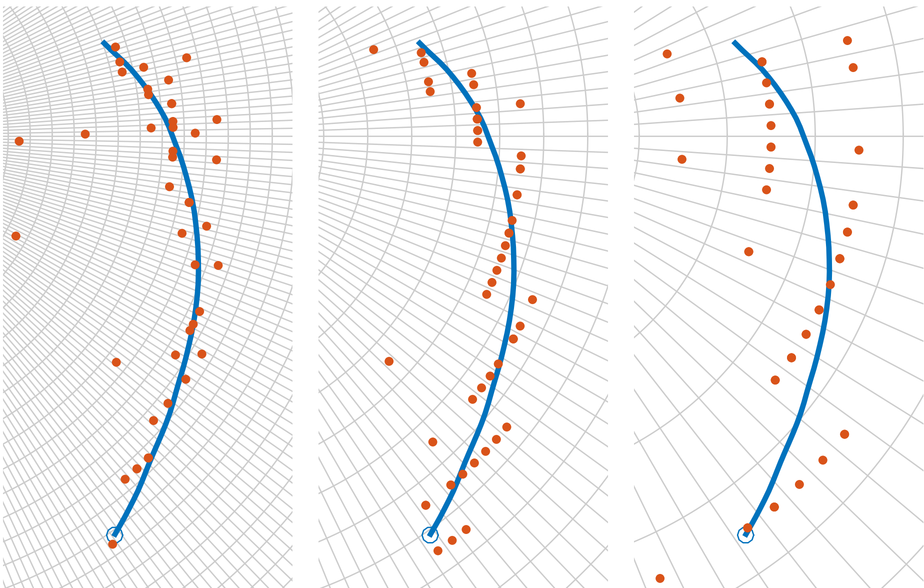


Figure 4.4: Trajectories (blue lines) with initial position (blue circle) and accumulated observations (orange bullets) for the 20 m × 4° observation cells



(a) 5m × 1° (b) 10m × 2° (c) 20m × 4°

Figure 4.5: Trajectories (blue lines) with initial position (blue circle) and accumulated observations (orange bullets, including false positives) of an object with increasing observation-cell sizes

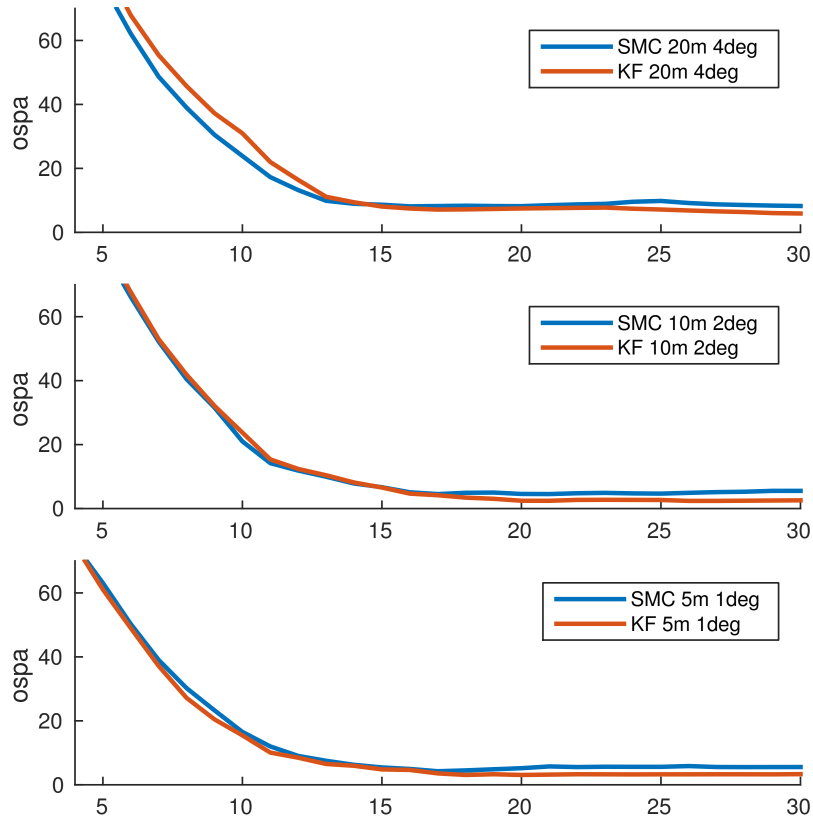


Figure 4.6: OSPA distance over 25 time steps and 100 Monte Carlo runs

Figure 4.6 compares the two implementations with different cell sizes in terms of OSPA distance, with a 2-norm and a cutoff of 100. The performance of the two implementations is similar, confirming that the SMC implementation does not bring the efficiency down while allowing for more diverse types of models to be used. The SMC version also has faster initialisation for larger resolution cells, but tends to be less accurate in the longer term.

Summary

A new multi-object filter for independent stochastic populations, called the HISP filter, has been derived and detailed. When studying this filter, it appeared that there is more than one way of using the observation-filtering equations and that there are different possible approximations as well as diverse applicable modelling alternatives. In this sense, the HISP filter can be seen as a general and computationally-affordable way of approaching the problem of multi-object estimation. The HISP filter allows for characterising each hypothesis separately thus giving a local picture of the underlying multi-object problem while controlling the level of approximation. Its efficiency has been compared with the performance of the PHD filter, since the two filters have equivalent complexities. The results show that the HISP filter outperforms the PHD filter in several cases with varying probabilities of detection and statistics for the spurious-observations.

Conclusion

While seeking a natural way to represent stochastic populations which could enable the formulation of versatile estimation algorithms, a variety of novel concepts and approaches have been introduced. Fulfilling this objective required a sufficiently general representation of the uncertainty, which in turn required both theoretical and practical considerations to be taken into account. A fundamental result arising from this general representation is the characterisation of the fusion of various levels of uncertainty, which has been written in a simple way. Equipped with these novel notions and results and with a way of representing partially-indistinguishable populations, we characterised the fusion of uncertainties about stochastic populations on spaces of decreasing sophistication. These different types of parametrisations allow for representations to be suited to a given purpose. Two estimation algorithms, namely the BISP and DISP filters, have been derived as an example of the characterisation and of the fusion of stochastic populations. The formulations of these algorithms made use of the range of possibilities that has been introduced and can serve as a basis for other problem-oriented estimation techniques. Finally, a more practical filter called the HISP filter has been proposed as an alternative to the computationally-demanding BISP and DISP filters. The HISP filter relies on an assumption of sparsity, in the sense that individuals are supposed to be sufficiently separated by the sensor. Its performance and versatility have been assessed on a range of scenarios.

Appendices

Appendix A

*Advanced measure theory

A presentation of the required non-trivial mathematical background is given in this section. Even though most of the concepts and results presented here can be found in various references, they are usually scattered across different areas. Here, the objective is to justify and motivate their use in a concise and consistent way. In particular, there is no detailed proof of known results, yet, the main steps of a proof will be given if they are likely to be insightful.

A.1 Radon measure and Polish space

This section is concerned with the concepts of Radon measure on Hausdorff spaces and the subsequent definition of a suitably fine σ-algebra on the set of probability measures. This commonly involves not only measure-theoretic arguments but also topological considerations.

The following concepts will be required: a topological space $(\mathbf{X}, \mathcal{T})$ is said to be *a*) *completely metrizable* if there exists at least one metric d on $\mathbf{X}$ such that $(\mathbf{X}, d)$ is a complete metric space and *b*) *separable* if it contains a countable dense subset.

In this section, we follow Schwartz [1973] for the definitions and for most of the results. Let $(\mathbf{X}, \mathcal{T})$ be a Hausdorff space [1] endowed with its Borel σ-algebra $\mathcal{B}(\mathbf{X})$.

Definition A.1. A Radon measure on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ is a measure μ which is locally finite and inner regular, i.e., for which it holds that

$$(\forall B \in \mathcal{B}(\mathbf{X})) \qquad \mu(B) = \sup\{\mu(K) \text{ s.t. } K \subseteq B, \;\; K \text{ compact}\}. \tag{A.1}$$

The Lebesgue measure on the Borel sets of any Euclidean space and the Dirac measure on any topological space are examples of Radon measures. Hausdorff topological spaces might prove to be too general for the purpose in this section. The objective is then to find sufficiently well-behaved spaces that allow for the standard operations of measure and probability theory to be applied without loosing too much generality. The spaces that can usually be found in the literature are the following, ordered by level of generality:

$$\text{Radon} \;\Leftarrow\; \text{Suslin} \;\Leftarrow\; \text{Lusin} \;\Leftarrow\; \text{Polish}. \tag{A.2}$$

The concept of Radon space is closely related to the one of Radon measure and is defined as follows.

[1] A Hausdorff space is a topological space in which different points have disjoint neighbourhoods

Definition A.2. A Hausdorff space $\mathbf{X}$ is called a Radon space if every finite Borel measure on $\mathbf{X}$ is a Radon measure.

As a direct consequence of Definition A.2, every probability measure on a Radon space is a Radon measure. Radon spaces have some nice properties: the class of Radon spaces is closed for countable topological sums, countable unions and intersections and for complementation. However, countable products of Radon spaces require every compact set in the involved spaces to be metrizable [Schwartz, 1973, Chapt. 2, §8, Thm. 8].

The definitions of Suslin and Lusin spaces are based on the one of Polish space and are not directly of interest here. However, we can note that closed and open subsets of a Suslin space are Suslin, and that countable products of Lusin spaces are Lusin. Also, every locally-finite Borel measure is a Radon measure in a Suslin space. Due to (A.2), these properties are also valid for Polish spaces, which are defined as follows.

Definition A.3. A Polish space is a separable completely-metrizable space.

See Fremlin [2000] for a relatively exhaustive list of Polish space properties. Note that, by assuming the space $\mathbf{X}$ to be locally compact, the following result holds:

$$\text{Polish} \;\Leftrightarrow\; \text{Lusin} \;\Leftrightarrow\; \text{Suslin} \;\Leftrightarrow\; \text{separable and metrizable.} \tag{A.3}$$

Some authors, including Daley and Vere-Jones [2003, 2008], prefer to consider one of the metrics on Polish spaces, so that they become **c**omplete **s**eparable **m**etric **s**paces, or CSMSs. Some other authors prefer to ignore the difference between the two concepts and use the term "Polish space" to refer to a CSMS.

A.2 Topology on the set of probability measure

We still consider the Hausdorff space $(\mathbf{X}, \mathcal{T})$. The objective in this section is to equip the set of Radon measures on $\mathbf{X}$ with an appropriate topology and/or σ-algebra.

Following Prokhorov [1956], we will see that the set $\mathbf{M}_1(\mathbf{X})$ of probability measure on $\mathbf{X}$ can be equipped with a topology that makes it Polish if $\mathbf{X}$ is itself a Polish space. To see that, we need a notion of convergence and a metric on $\mathbf{M}_1(\mathbf{X})$. References on this topic include Billingsley [1968] and Van Gaans.

A.2.1 Convergence and metric

Let d be a metric on $\mathbf{X}$ that is compatible with $\mathcal{T}$, and let $(P_n)_n$ be a sequence of probability measures in $\mathbf{M}_1(\mathbf{X})$. We say that $(P_n)_n$ *converges weakly* to the probability measure $P \in \mathbf{M}_1(\mathbf{X})$, written $P_n \Rightarrow P$, whenever

$$\int f(x) P_n(\mathrm{d}x) \to \int f(x) P(\mathrm{d}x) \tag{A.4}$$

for every bounded continuous real-valued function f on $\mathbf{X}$. Prokhorov introduced a distance on $\mathbf{M}_1(\mathbf{X})$, considering it as analogous to the Lévy distance, and defined it for any $P, P' \in \mathbf{M}_1(\mathbf{X})$ as

$$d_{\mathrm{P}}(P, P') \doteq \inf \Big\{ \epsilon > 0 \text{ s.t. } P(B) \leq P'(B_\epsilon) + \epsilon, \;\; P'(B) \leq P(B_\epsilon) + \epsilon, \;\; \forall B \in \mathcal{B}(\mathbf{X}) \Big\}, \tag{A.5}$$

where A_ϵ is the ϵ-neighbourhood of $A \subseteq \mathbf{X}$, defined as

$$A_\epsilon = \{x \in \mathbf{X} \text{ s.t. } (\exists y \in A) \;\; d(x,y) \leq \epsilon\}. \tag{A.6}$$

The metric d_P is called the Lévy-Prokhorov metric, or the Prokhorov metric for short. We can now consider the metric space $(\mathbf{M}_1(\mathbf{X}), d_\mathrm{P})$ as well as the topological space $(\mathbf{M}_1(\mathbf{X}), \mathcal{T}_\mathrm{P})$, where $\mathcal{T}_\mathrm{P}$ is the topology induced by the metric d_P. Assuming that $(\mathbf{X}, d)$ is separable, we can prove that convergence in the metric d_P is equivalent to weak convergence.

A.2.2 Properties

Assuming again that the space $(\mathbf{X}, d)$ is separable, it holds that it contains a countable dense subset D and we can prove that the set of atomic probability measures with atoms in D and rational masses is dense in $(\mathbf{M}_1(\mathbf{X}), d_\mathrm{P})$, so that we obtain the following result.

Theorem A.1. *If* $(\mathbf{X}, d)$ *is separable, then* $(\mathbf{M}_1(\mathbf{X}), d_\mathrm{P})$ *is separable.*

In other words, forgetting about the metrics, if $(\mathbf{X}, \mathcal{T})$ is a separable metrizable space, then so is $(\mathbf{M}_1(\mathbf{X}), \mathcal{T}_\mathrm{P})$. To show the completeness of $\mathbf{M}_1(\mathbf{X})$, additional concepts are required. Let $\mathcal{P}$ be a family of probability measures in $\mathbf{M}_1(\mathbf{X})$, we say that $\mathcal{P}$ is *a*) tight if for every $\epsilon > 0$ and all $P \in \mathcal{P}$ there exists a compact set K such that $P(K) > 1 - \epsilon$, and *b*) *relatively compact* if every sequence of elements of $\mathcal{P}$ contains a *weakly convergent subsequence*, i.e., for every sequence $(P_n)_n$ in $\mathcal{P}$, there exists a subsequence $(P_{n'})_{n'}$ and a probability measure P in $\mathbf{M}_1(\mathbf{X})$ such that $P_{n'} \Rightarrow P$.

Prokhorov theorem then tells us that if $\mathcal{P}$ is tight, then it is also relatively compact and, if $(\mathbf{X}, d)$ is separable and complete and if $\mathcal{P}$ is relatively compact, then it is tight. Proving that Cauchy sequences are tight in $(\mathbf{M}_1(\mathbf{X}), d_\mathrm{P})$ implies that they are also relatively compact by Prokhorov theorem, which in turns, implies their convergence. The following theorem holds as a result.

Theorem A.2. *If* $(\mathbf{X}, d)$ *is a complete separable metric space, then* $(\mathbf{M}_1(\mathbf{X}), d_\mathrm{P})$ *is complete.*

Combining Theorems A.1 and A.2, we obtain that $(\mathbf{M}_1(\mathbf{X}), \mathcal{T}_\mathrm{P})$ is a Polish space if $(\mathbf{X}, \mathcal{T})$ is Polish. In particular, this result enables the introduction of the set $\mathbf{M}_1(\mathbf{M}_1(\mathbf{X}))$ which can be itself equipped with an appropriate topology, and the corresponding Borel σ-algebra can be considered.

A.3 Topology on the set of measurable functions

A common approach for defining a topology on sets of functions is to check that the set of all continuous functions with compact support from the d-dimensional smooth manifold to the real line has sufficiently nice properties for this purpose [Bourbaki, 1969, Chapt. 9]. However, the interest here is in the more general concept of measurable functions, and a different approach has to be considered.

Let $(\mathbf{E}, \mathcal{E}, m)$ be a measure space, and recall that $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ is the set of real-valued measurable functions on $\mathbf{E}$. A pseudometric[2] ρ_A can be defined on $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ for every

[2] A pseudometric is a real-valued function which behaves like a metric except that the distance between two distinct points can be null

measurable subset $A \in \mathcal{E}$ of finite measure by

$$(\forall f, f' \in \mathbf{L}^0(\mathbf{E}, \mathbb{R})) \qquad \rho_A(f, f') \doteq \int \min\{|f(x) - f'(x)|, \mathbf{1}_A(x)\} m(\mathrm{d}x). \tag{A.7}$$

The family

$$\boldsymbol{\rho} = \{\rho_A \text{ s.t. } A \in \mathcal{E}, \ \ m(A) < \infty\} \tag{A.8}$$

induces a topology on $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ in the following way [Fremlin, 2000, Sect. 2A3F]: an open set G is a set such that for every $f \in G$, there are pseudometrics $\rho_{A_1}, \dots, \rho_{A_n} \in \boldsymbol{\rho}$ and $\delta > 0$ such that

$$\{f' \in \mathbf{L}^0(\mathbf{E}, \mathbb{R}) \text{ s.t. } \max_{i \le n} \rho_{A_i}(f, f') < \delta\} \subseteq G. \tag{A.9}$$

The topology induced by the family $\boldsymbol{\rho}$ is referred to as *topology of convergence in measure.*

Remark.* Topological spaces where the concept of "closeness" is defined by a family of pseudometrics are referred to as *uniform spaces* [Bourbaki, 1974, Chapt. II].

The following results, which can be deduced from [Fremlin, 2000, Sect. 245], will be useful for studying probability measures on the set of measurable functions.

Proposition A.1. *The topology of convergence in measure on* $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ *is completely metrizable if* m *is* σ*-finite.*

Proof. Denote $\mathcal{T}$ the topology of convergence in measure on $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$. From [Fremlin, 2000, Sect. 245E], we deduce that *a)* m is σ-finite if and only if $\mathcal{T}$ is metrizable, and *b)* if m is σ-finite then $\mathcal{T}$ is Hausdorff and $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ is complete under $\mathcal{T}$. □

Proposition A.2. *If* $(\mathbf{E}, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *for some* $d > 0$ *and if* m *is the Lebesgue measure on* $\mathbb{R}^d$*, then the set* $\mathbf{L}^0(\mathbb{R}^d, \mathbb{R})$ *is separable for the topology of convergence in measure.*

Proof. Let S be the set of simple functions of the form

$$\sum_{i \in \mathbb{I}} \alpha_i \mathbf{1}_{[a_i, b_i)}, \tag{A.10}$$

where $\mathbb{I}$ is a countable index set, $\{\alpha_i\}_{i \in \mathbb{I}}$ is a family of elements of the set $\mathbb{Q}$ of rational numbers, and $\{[a_i, b_i)\}_{i \in \mathbb{I}}$ is a family of intervals with $a_i, b_i \in \mathbb{Q}^d$. Then S is a countable subset of $\mathbf{L}^0(\mathbb{R}^d, \mathbb{R})$ which closure is equal to the set $\mathbf{L}^0(\mathbb{R}^d, \mathbb{R})$ itself. The existence of a dense countable subset shows that $\mathbf{L}^0(\mathbb{R}^d, \mathbb{R})$ is separable for the topology of convergence in measure. □

The results of Propositions A.1 and A.2 show that $\mathbf{L}^0(\mathbb{R}^d, \mathbb{R})$ with the Lebesgue measure can be seen as a Polish space when equipped with the topology of convergence in measure. For any measure space $(\mathbf{E}, \mathcal{E}, m)$, the topology of convergence in measure on $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ has some useful properties, detailed in the following proposition.

Property A.1 (Fremlin [2000, Sect. 245D])**.** *The binary operations* $+$ *and* $\cdot$ *on* $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$ *as well as the mapping*

$$\mathbb{R} \times \mathbf{L}^0(\mathbf{E}, \mathbb{R}) \ni (\alpha, f) \mapsto \alpha f \in \mathbf{L}^0(\mathbf{E}, \mathbb{R}) \tag{A.11a}$$

are continuous in the topology of convergence in measure.

The fact that the addition and the multiplication by a scalar are continuous operations makes the topology of convergence in measure a *linear space topology* on $\mathbf{L}^0(\mathbf{E}, \mathbb{R})$.

A.4 Measurability

Operations on probabilistic constraints require some more advanced topological considerations. In particular, we will need the concept of *universally measurable set*, which is related to subsets of a Polish space that are measurable w.r.t. every complete probability measure that measures all Borel subsets of the considered space.

The measurability of mappings defined with a supremum is of importance in this work and we consider the next theorem from Castaing and Valadier [1977] as well as the following corollary from Crauel [2003].

Theorem A.3 (Projection theorem [Castaing and Valadier, 1977, Theorem III.23]). *If* $\mathbf{X}$ *is a Polish space and if* $(\mathbf{E}, \mathcal{E})$ *is a measurable space, then the projection of* $A \subset \mathbf{X} \times \mathbf{E}$ *to* $\mathbf{E}$*, given by*

$$\pi_{\mathbf{E}}(A) = \Big\{ y \in \mathbf{E} \ \textit{s.t.} \ \exists x \in \mathbf{X}[(x, y) \in A] \Big\}, \tag{A.12}$$

is universally measurable for any $A \in \mathcal{B}(\mathbf{X}) \times \mathcal{E}$.

Corollary A.1 ([Crauel, 2003, Corollary 2.13]). *Let* $\mathbf{X}$ *be a Polish space, let* $(\mathbf{E}, \mathcal{E})$ *be a measurable space, let the function* $f : \mathbf{X} \times \mathbf{E} \to \mathbb{R}$ *be measurable and let* $y \mapsto C(y)$ *be a set-valued mapping on* $\mathbf{E}$ *such that* $\mathrm{Gr}(C)$ *is measurable, then the mapping*

$$y \mapsto \sup_{x \in C(y)} f(x, y) \tag{A.13}$$

is measurable w.r.t. the universally completed σ*-algebra of* $\mathcal{E}$.

A.5 Complete measure

The concept of complete measure, or more precisely of complete measure space, has been introduced in Section 1.1. The objective is now to study how any measure can be made complete and what the properties of complete measure spaces are.

Let $(\mathbf{E}, \mathcal{E}, \mu)$ be a measure space. We first perform a completion procedure on the σ-algebra $\mathcal{E}$: let $\hat{\mathcal{E}} \subseteq \wp(\mathbf{E})$ be defined as

$$\hat{\mathcal{E}} \doteq \Big\{ A \subseteq \mathbf{E} \ \text{s.t.} \ \exists A', A'' \in \mathcal{E}[A' \subseteq A \subseteq A'' \ \wedge \ \mu(A'' \setminus A') = 0] \Big\}, \tag{A.14}$$

then $\hat{\mathcal{E}}$ is a σ-algebra that contains $\mathcal{E}$. We then define an extention of μ to the σ-algebra $\hat{\mathcal{E}}$: let $\hat{\mu}$ be a set function on $\hat{\mathcal{E}}$ defined as

$$(\forall A \in \hat{\mathcal{E}}) \qquad \hat{\mu}(A) \doteq \inf\{\mu(A') \ \text{s.t.} \ A' \supseteq A, \ \ A' \in \mathcal{E}\}, \tag{A.15}$$

then $\hat{\mu}$ is a complete measure. The space $(\mathbf{E}, \hat{\mathcal{E}}, \hat{\mu})$ is called *the completion* of $(\mathbf{E}, \mathcal{E}, \mu)$ since it is uniquely defined and verifies $\hat{\mu}(A) = \mu(A)$ for all $A \in \mathcal{E}$.

We now study under which conditions the pushforward of a complete measure is also complete. To that purpose, it is useful to consider the σ-algebra induced by a function on its codomain as follows: let $(\mathbf{E}, \mathcal{E})$ be a measurable space, let $\mathbf{F}$ be a set and let $f : \mathbf{E} \to \mathbf{F}$ be a mapping, then the set $\mathcal{F} \subseteq \wp(\mathbf{F})$, defined as

$$\mathcal{F} \doteq \{A \subseteq \mathbf{F} \ \text{s.t.} \ f^{-1}[A] \in \mathcal{E}\} \tag{A.16}$$

is a σ-algebra of subsets of $\mathbf{F}$, and is referred to as the σ-algebra induced by f on $\mathbf{F}$. The mapping f is then measurable by construction.

Property A.2 ([Fremlin, 2000, Sect. 212B])**.** *Let $(\mathbf{E}, \mathcal{E}, \mu)$ be a complete measure space, let $\mathbf{F}$ be a set and let $f : \mathbf{E} \to \mathbf{F}$ be a mapping. Then, denoting $\mathcal{F}$ the σ-algebra induced by f on $\mathbf{F}$, it holds that $(\mathbf{F}, \mathcal{F}, f_*\mu)$ is a complete measure space.*

The procedure of completion of a measure does not change the values that the considered measure gives to subsets that were already measurable in the original σ-algebra. As a result, the completion of a measure does not affect probabilistic constraints that might have been defined for it. In consequence, we will consider the completion of a measure space rather than the measure space itself whenever required.

Index

Bibliography

C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

J. Baron. Second-order probabilities and belief functions. *Theory and Decision*, 23 (1):25–36, 1987.

P. Billingsley. *Convergence of probability measures*, volume 493. John Wiley & Sons, 1968.

G. Birkhoff. *Lattice Theory.* American Mathematical Society, 3rd edition, 1984.

S. S. Blackman. *Multiple-target tracking with radar applications*, volume 1. Artech House, Norwood, 1986.

V. I. Bogachev. *Measure Theory.* Springer, 2007.

N. Bourbaki. *Eléments de mathématiques, intégration.* Hermann, Paris, 1969.

N. Bourbaki. *Elements of Mathematics. General Topology, Part 1.* Hermann, Addison-Wesley, Paris, Reading, MA, 1974.

G. Cantor. *Contributions to the Founding of the Theory of Transfinite Numbers.* Courier Dover Publications, 1915.

C. Castaing and M. Valadier. *Convex analysis and measurable multifunctions.* Springer, 1977.

D. R. Cox. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 51, pages 433–441. Cambridge Univ Press, 1955.

H. Crauel. *Random probability measures on Polish spaces*, volume 11. CRC Press, 2003.

N. C. A. da Costa. *Ensaio sobre os fundamentos da lógica.* Editora Hucitec, São Paulo, 1980.

D. Daley and D. Vere-Jones. *An Introduction to The Theory of Point Processes.* Springer, 2nd edition, 2003.

D. Daley and D. Vere-Jones. *An Introduction to The Theory of Point Processes, Vol. II.* Springer, 2nd edition, 2008.

P. Davies. *The Goldilocks enigma: why is the universe just right for life?* Mariner Books, 2008.

P. Del Moral. *Mean field simulation for Monte Carlo integration.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 2013.

E. Delande, J. Houssineau, and D. Clark. PHD filtering with localised target number variance. In *SPIE Defense, Security, and Sensing*, 2013a.

E. Delande, J. Houssineau, and D. E. Clark. Localised variance in target number for the Cardinalized Hypothesis Density Filter,". In *Information Fusion, Proceedings of the 16th International Conference on*, 2013b.

E. Delande, J. Houssineau, and D. E. Clark. Performance metric in closed-loop sensor management for stochastic populations. In *IEEE Sensor Signal Processing for Defence (SSPD)*, 2014a.

E. Delande, M. Uney, J. Houssineau, and D. E. Clark. Regional variance for multi-object filtering. *Signal Processing, IEEE Transactions on*, 62(13):3415–3428, 2014b.

E. Delande, J. Houssineau, and D. E. Clark. A filter for distinguishable and independent populations. *ArXiv:1501.04671*, 2015.

A. P. Dempster. Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika*, 54(3-4):515–528, 1967.

A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 205–247, 1968.

J. Dixon and B. Mortimer. *Permutation Groups.* Graduate texts in mathematics. Springer, 1996.

A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *Signal Processing, IEEE Transactions on*, 49(3): 613–624, 2001.

D. Dubois and H. Prade. *Possibility theory.* Wiley Online Library, 1988.

D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In *Quantified representation of uncertainty and imprecision*, pages 169–226. Springer, 1998.

D. Dubois and H. Prade. *Fundamentals of fuzzy sets*, volume 7. Springer Science & Business Media, 2000.

R. M. Dudley. *Real analysis and probability*, volume 74. Cambridge University Press, 2002.

D. S. Dummit and R. M. Foote. *Abstract Algebra.* John Wiley and Sons, Inc., 2004.

R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models.* Springer, 1994.

R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and computation*, 87(1):78–128, 1990.

M. Fanaswala and V. Krishnamurthy. Detection of anomalous trajectory patterns in target tracking via stochastic context-free grammars and reciprocal process models. *Selected Topics in Signal Processing, IEEE Journal of*, 7(1):76–90, 2013.

W. Feller. Non-Markovian processes with the semigroup property. *The Annals of Mathematical Statistics*, 30(4):1252–1253, 1959.

T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

D. H. Fremlin. *Measure theory*. 4^{th} edition, 2000.

N. Friedman and J. Y. Halpern. Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685, 2001.

C. Frueh, E. Delande, J. Houssineau, and D. E. Clark. Multi-object filtering for space situational awareness. In *25th AAS/AIAA Space Flight Mechanics Meeting*, 2015.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

J. Houssineau and D. Clark. Bayesian estimation of multi-object systems with independently identically distributed correlations. In *Workshop on Statistical Signal Processing (SSP)*, 2014.

J. Houssineau, D. E. Clark, and P. Del Moral. A sequential Monte Carlo approximation of the HISP filter. In *European Signal Processing Conference (EUSIPCO)*, 2015.

K. Itô. *Introduction to probability theory*. Cambridge University Press, 1984.

K. Itô. *Stochastic processes: lectures given at Aarhus University*. Springer Science & Business Media, 2013.

L. Jiang, S. S. Singh, and S. Yıldırım. Bayesian tracking and parameter learning for non-linear multiple target tracking models. *arXiv preprint arXiv:1410.2046*, 2014.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

A. S. Kechris. *Classical descriptive set theory*, volume 156. Springer-Verlag New York, 1995.

J. W. Koch. Bayesian approach to extended object and cluster tracking using random matrices. *Aerospace and Electronic Systems, IEEE Transactions on*, 44(3):1042–1059, 2008.

D. Krause. On a quasi-set theory. *Notre Dame Journal of Formal Logic*, 33(3):402–411, 1992.

H. Lebesgue. Sur les fonctions représentables analytiquement. 1905.

M. Loève. Probability theory, vol. ii. *Graduate texts in mathematics*, 46:0–387, 1978.

O. Macchi. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, pages 83–122, 1975.

R. P. S. Mahler. Multitarget Bayes filtering via first-order multitarget moments. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1152–1178, 2003.

R. P. S. Mahler. Can the bayesian and dempster-shafer approaches be reconciled? yes. In *Information Fusion, 2005 8th International Conference on*, volume 2. IEEE, 2005.

R. P. S. Mahler. PHD filters of higher order in target number. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1523–1543, 2007.

Y. Manin. *Mathematical Developments Arising from Hilbert Problems*, volume 28 of *Proceedings of symposia in pure mathematics*. American Mathematical Society, 1976.

J. Moyal. The general theory of stochastic population processes. *Acta Mathematica*, pages 1–31, 1962.

J. Nedoma. Note on generalized random variables. In *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Liblice, 1956), Publishing House of the Czechoslovak Academy of Sciences, Prague*, pages 139–141, 1957.

C. V. Negoita, L. A. Zadeh, and H. J. Zimmermann. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.

S. Oh, S. Russell, and S. Sastry. Markov chain Monte Carlo data association for general multiple-target tracking problems. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 1, pages 735–742. IEEE, 2004.

M. Pace and P. Del Moral. Mean-field PHD filters based on generalized Feynman-Kac flow. *Journal of selected topics in signal processing, special issue on multi-target tracking*, 2013.

Y. Pailhas, J. Houssineau, E. Delande, Y. Petillot, and D. E. Clark. Tracking underwater objects using large MIMO sonar systems. In *International Conference on Underwater Acoustic*, 2014.

Y. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214, 1956.

D. J. Salmond. Mixture reduction algorithms for target tracking in clutter. In *OE/LASE'90, 14-19 Jan., Los Angeles, CA*, pages 434–445. International Society for Optics and Photonics, 1990.

D. J. Salmond and N. J. Gordon. Group and extended object tracking. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pages 284–296. International Society for Optics and Photonics, 1999.

E. Schechter. *Handbook of Analysis and its Foundations*. Academic Press, 1996.

D. Schuhmacher, B.-T. Vo, and B.-N. Vo. A consistent metric for performance evaluation of multi-object filters. *Signal Processing, IEEE Transactions on*, 56(8):3447–3457, 2008.

L. Schwartz. *Radon measures on arbitrary topological spaces and cylindrical measures*, volume 239. Oxford University Press London, 1973.

G. Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

G. Shafer. The combination of evidence. *International Journal of Intelligent Systems*, 1(3):155–179, 1986.

S. S. Singh, B.-N. Vo, A. Baddeley, and S. Zuyev. Filters for spatial point processes. *SIAM Journal on Control and Optimization*, 48:2275–2295, 2009.

M. Y. Souslin. Sur une définition des ensembles mesurables b sans nombres transfinis. 1917.

F. Thouin, S. Nannuru, and M. Coates. Multi-target tracking for measurement models with additive contributions. In *Information Fusion, Proceedings of the 14th International Conference on*, pages 1–8, 2011.

O. Van Gaans. Probability measures on metric spaces. Technical report, Delft University of Technology.

B.-N. Vo and W.-K. Ma. The Gaussian mixture probability hypothesis density filter. *Signal Processing, IEEE Transactions on*, 54(11):4091–4104, 2006.

B.-T. Vo and B.-N. Vo. Labeled random finite sets and multi-object conjugate priors. *Signal Processing, IEEE Transactions on*, 61(13):3460–3475, 2013.

T. Vu, B.-N. Vo, and R. Evans. A particle marginal Metropolis-Hastings multi-target tracker. *Signal Processing, IEEE Transactions on*, 62(15):3953–3964, Aug 2014.

R. C. Williamson. *Probabilistic arithmetic.* PhD thesis, University of Queensland, 1989.

J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52, 1928.

J. Yen. Generalizing the Dempster-Schafer theory to fuzzy sets. *Systems, Man and Cybernetics, IEEE Transactions on*, 20(3):559–570, 1990.

L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

L. A. Zadeh. On the validity of Dempster's rule of combination. 1979.