Decision making under uncertainty with Bayesian filters

Alexey Sergeevich Narykov

A Thesis Submitted for the Degree of
Doctor of Philosophy in Electrical Engineering

HERIOT WATT UNIVERSITY
THE UNIVERSITY of EDINBURGH

July 2020

Abstract

This work is concerned with exploiting Bayesian filters for decision making under uncertainty. The kind of decision making that is formally suitable for problems requiring finding optimal (non-sensing) actions as well as optimal answers/statements. Specifically, the focus will be on filters for spatial point processes which model nature as a population of indistinguishable objects. Previous works have been limited to translating the problem of point estimation into loss functions compatible with object populations. Whereas the present work systematically constructs a number of novel loss functions that give rise to a class of statistical problems beyond point estimation, which have not been appropriately formalized yet. We obtain closed-form solutions to those problems (expressions computing optimal statements and corresponding minimized expected values of loss), and implement the solutions with a variety of approximate filters: the classical PHD filter, the Panjer PHD (PPHD) filter, and the Cardinalized PHD (CPHD) filter. We offer practical interpretations of the introduced problems, such as the estimation of risk value attached to an uncertain object population, and demonstrate selected implementations through numerical simulations. Overall, this work extends the variety of problems solvable using information from Bayesian filters, and reduces the amount of avoidable losses in such problems when compared to conventional approaches.

Contents

Acknowledgments

I would like to express my gratitude to Prof. Yvan Petillot, Prof. Mike E. Davies, and Dr. James Hopgood for making this all possible. I acknowledge the academic support of the University Defence Research Collaboration in Signal Processing, and the financial support of James Watt Scholarship at Heriot-Watt University.

I am indebted to Dr. Daniel E. Clark for inviting me to join his research team. I would like to express my appreciation to everyone in the team, including José Franco, Vibhav Bharti, Flávio Eler and Murat Üney. Most importantly, I wish to mention Emmanuel Delande and Isabel Schlangen whose excellent research is the foundation of this work. Thank you for introducing me to the subject and always being there to help.

I wish to thank the reviewers of this thesis, Prof. Simon Julier of University College London and Dr. Yoann Altmann of Heriot-Watt University, for taking interest in my research and helping to improve the present manuscript.

On a personal note, I wish to thank the people I met in Edinburgh that have become my close friends and a strong support network: José, Gala, [Eddy, Toms, Eric, Kima], and Nicolas. I wish to thank my family for their unconditional support, which has been a huge privilege. Thank you to Marianne for her kindness, humour, and helping me get unstuck in the most critical moments.

Chapter 1

Introduction

This chapters sets the context of the performed work, and introduces important concepts. Section 1.1 introduces the procedure of decision making, and describes that it is formally suitable to produce commands (that lead to actions) and statements. Section 1.2 clarifies the connection of sensing to decision making in Bayesian settings. Section 1.3 addresses the above procedures in the context of Bayesian filtering. Following the background information, Section 1.4 provides a statement of the problem and set out the objective for this work. Section 1.5 outlines the approach followed by this work. Section 1.6 highlights the contributions.

1.1 Terminal decision making (no sensing)

This section presents basic elements and definitions that provide settings for decision making that lead to the development of decision making under uncertainty or terminal analysis. We found it useful to present this material by tracing the origins of statistical decision making in game theory.

Decision making, which will be in the focus of this thesis, is best introduced along with its parent process that often gets overlooked: problem solving. Consider the following quote [101]:

> The work of managers, of scientists, of engineers, of lawyers — the work that steers the course of society and its economic and governmental organizations — is largely work of making decisions and solving problems. It is work of choosing issues that require attention, setting goals, finding suitable courses of action, and evaluating and choosing among alternative actions. The first three of these activities — fixing agendas, setting

goals, and designing actions — are usually called *problem solving*; the last, evaluating and choosing, is usually called *decision making*.

Problem solving establishes the context in which decision making will take place. As a consequence, decision making is commonly analysed in settings where the problem has been already framed, the goals are set and alternative courses of action are specified. In the scope of this thesis we will rely on well-defined formalisations which originate from game theory, as discussed in the next section.

1.1.1 Two-player zero-sum game against nature

The origins of statistical decision theory, which is concerned with a broad class of decisions under non-certainty discussed later in Subsection 1.1.4, are rooted in the game theory. Specifically, it was Wald who recognised its value for *systematic* interpretation of statistical procedures developed by Fisher (such as point estimation and hypothesis testing) as zero-sum games against nature. In this case 'nature' is interpreted as a fictitious player having no known goal [60, 114]. For this player the set of actions is replaced by the set of states, and it has no utility function in the sense of Von Neumann-Morgenstern [78]. Another player is often called 'a statistician' or 'a decision maker'.

In principle, the circumstances when decision making is encountered are then presented using two sets of variables. The first set of variables $\mathcal{A}$, those under control of the player, represent all possible acts. The second set $\mathcal{S}$ is outside the control of the player, and represents possible states of nature (equivalently, world or environment). These two sets are used to specify a loss function. This function is one of the central elements that was introduced by Wald, which he termed 'weight function', and it was his interpretation of the player's utility.[1] This function is the key element in formulation of a decision problem, and is a gateway to communicate the player's preferences.

Definition 1.1.1 (State space). *The set $\mathcal{S}$ of possible states of nature is called the state space.*

Definition 1.1.2 (Act space). *The set $\mathcal{A}$ of available options (i.e. actions, answers, conclusions, decisions, etc.) is called the act space. Following [83], these options are described as terminal acts to distinguish them from sensing acts and acts of*

[1] One of the consequences is that utility axioms of Von Neumann-Morgenstern are not strictly followed. A ubiquitous squared error model, that will be presented later in Section 1.5, violates one of the axioms as it is not bounded from above, as discussed by Durrant-Whyte in [27, p. 147].

	s_1	s_2
a_1	(a_1, s_1)	(a_1, s_2)
a_2	(a_2, s_1)	(a_2, s_2)

Table 1.1: Matrix of outcomes

	s_1	s_2
a_1	$L(a_1, s_1)$	$L(a_1, s_2)$
a_2	$L(a_2, s_1)$	$L(a_2, s_2)$

Table 1.2: Matrix of costs

experimentation, which are acts of different nature that are not considered at this point.

Definition 1.1.3 (Loss function)**.** *Loss function is a function $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$ that attaches a value of loss to every possible outcome*[2] *(a, s) of selecting an option $a \in \mathcal{A}$ when the state of nature is $s \in \mathcal{S}$.*

The actual outcomes of alternative courses of action depend on the joint behaviour of the decision maker and nature. In the simplest case, when the sets of actions and states of nature are represented by two points, i.e. $\mathcal{A} = \{a_1, a_2\}$ and $\mathcal{S} = \{s_1, s_2\}$, this function $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$ is a simple 2×2 matrix where to each outcome (see Table 1.1) a certain cost is prescribed (see Table 1.1). Having introduced the loss function, we can move on to define two realms of decision making.

Next we are going to describe how the loss function is employed to formulate decision making as an optimization problem.

1.1.2 Realms of decision making

A completely different situation is when the actual state of nature is not known. This leads to the conditions when the outcome is non-certain. We are now at the point of a large watershed in the theories of making decisions. Basically, there are two different viewpoints on how this situation is to be treated. One school of thought states that it is always possible to assign probabilities to the states of the world. This is refers to the category of decision making which we call decision making under uncertainty. Another school of thought state that it is never possible to assign probabilities to the states of the world.

Let us present a classification of decision-making conditions. We are inspired by the classification by Luce and Raiffa [66, p. 13], which, in turn, originates from Knight [54]. We suggest that the field of (normative) decision-making can be partitioned according to whether a decision among candidate options (candidate actions, or candidate answers) is made under conditions of:

[2]Note a more general formulation involves explicit specification of the space of outcomes, and an outcome function, which is absorbed here in the definition of loss function.

(a) *certainty*, if each option is known to lead invariably to a specific outcome;

(b) *non-certainty*, if either action has as its outcome a set of possible outcomes, but where the probabilities of these outcomes are completely unknown or are not even meaningful;

(c) *uncertainty*, if each option leads to one of a set of possible specific outcomes occurring with a known probability;

(d) *assumed certainty equivalence*, if each option leads to one of a set of possible specific outcomes occurring with a known probability, but the decision-maker assumes that the outcome is known to lead invariably to a specific outcome associated with the estimated state of nature, or with the summary of the state such as its mean.

Schematically, this classification is illustrated on Figure 1.1. Decision making under uncertainty and decision making under non-certainty are the most studied cases, and are often found under alternative titles collected in Table 1.3. In this connection, it is important to bring in the classification which can be traced to Knight [54] which is often used in the literature. Decision making under uncertainty, as addressed by this thesis corresponds to decision making under [Knightian] risk. This should be remembered, as many of the results found in decision-making literature are stated exactly in these terms.

The remainder of this section will discuss each decision-making realm with respect to using a function of loss, which was defined in Subsection 1.1.1.

According to [66, p. 13], decision making under uncertainty and decision making under non-certainty, as defined in the classification above, can be extended to include the opportunity of processing new observations. As far as decision making under uncertainty is concerned, consideration of this possibility will be deferred until Section 1.2, and therefore the solutions produced prior to that will be denoted with the subscript ϕ. As far as decision making under non-certainty is concerned, this possibility will not be considered in this thesis, but discussions in the relevant context can be found in [38, 73, 93], and otherwise general discussions on decision making in such settings can be found in [9, 12, 94].

1.1.3 Decision making under certainty

Perhaps a trivial case of decision making is decision making under certainty. Clearly, when the state of nature is known, it is possible to get a direct access to the value

DM under uncertainty	DM under non-certainty
Bayesian decision theory	Classical decision theory
DM under [Knightian] risk	DM under [Knightian] uncertainty
	DM under ambiguity

Table 1.3: Alternative titles for non-deterministic decision making (DM)

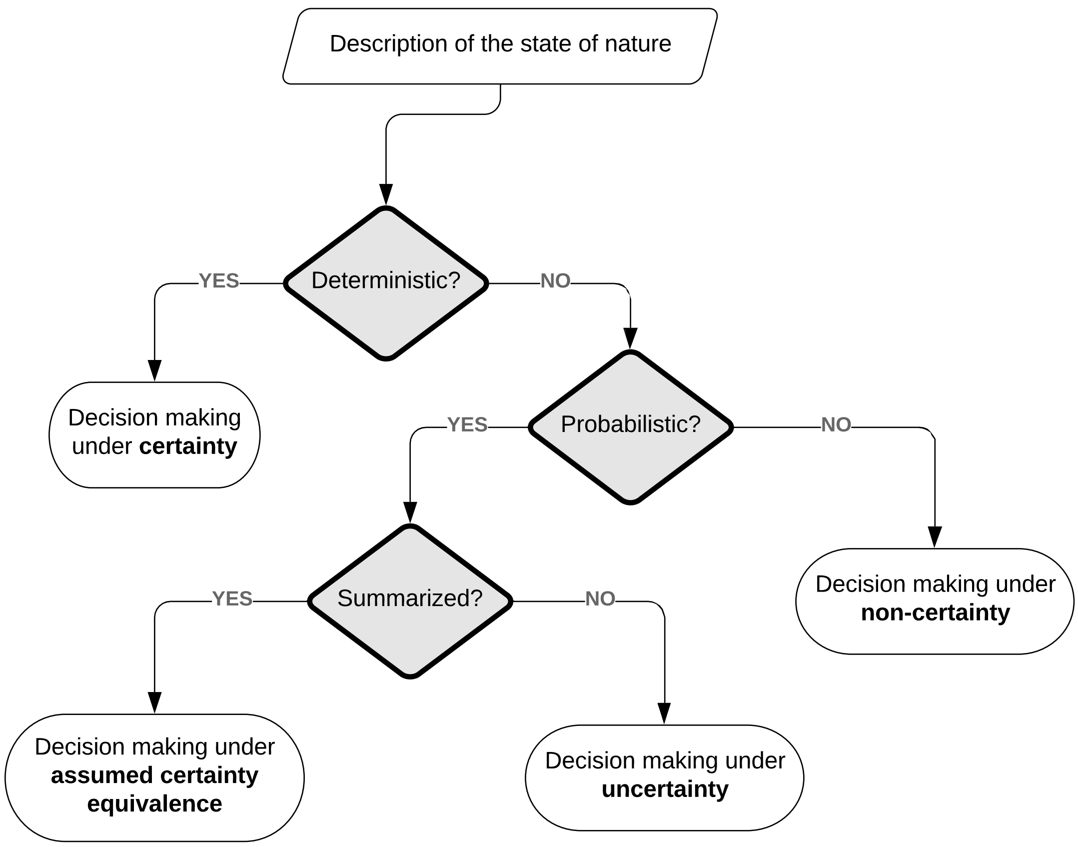


Figure 1.1: The realms of decision making depending on the conditions leading to a decision (e.g. an action or answer). The focus of this thesis is on probabilistic decision making. Decision making under certainty will be instrumental in evaluation of developed decision-making algorithms. Decision making under non-certainty will only be briefly mentioned.

of loss associated with an outcome for any considered action. However, in general finding an optimal solution may involve certain difficulties if the associated optimisation problem is complicated. Difficulties associated with making such decisions are *technical* [12, p. 14], as opposed to conceptual difficulties associated with non-deterministic decision making.

Proposition 1.1.4 (Minimum loss principle)**.** *When the state of nature $s \in \mathcal{S}$ is known, a decision-maker with loss $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$, should chose an action $a \in \mathcal{A}$ that minimizes their losses, i.e.*

$$a_\phi^{UC} = \arg\min_{a \in \mathcal{A}} L(a, s), \tag{1.1}$$

that corresponds to the optimised loss value

$$\rho_\phi^{UC} = L(a_\phi^{UC}, s). \tag{1.2}$$

Various problems commonly falling under operations research belong to this category, see [66, Sec. 2.2.] and [50, Ch. 2].

1.1.4 Decision making under non-certainty

Decision making under non-certainty (also referred to as *classical decision theory*, *decision-making under [Knightian] uncertainty*). It can be loosely described as decision theory without a prior distribution.

Proposition 1.1.5 (Minimax principle)**.** *When nature takes state in $\mathcal{S}$, a decision-maker with loss $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$, should choose an option $a \in \mathcal{A}$ that minimizes their minimax loss, i.e.*

$$a_\phi^{UN} = \arg\min_{a \in \mathcal{A}} \left[\max_{s \in \mathcal{S}} L(a, s) \right], \tag{1.3}$$

that corresponds to the minimax loss value

$$\rho_\phi^{UN} = L(a_\phi^{UN}, \arg\max_{s \in \mathcal{S}} L(a_\phi^{UN}, s)). \tag{1.4}$$

1.1.5 Decision making under uncertainty

Decision making under uncertainty (also referred to as *Bayesian decision theory*, or *decision-making under [Knightian] risk*) will be the central subject of this thesis. This is the situation when probabilities are known, see Figure 1.2a. We suppose that there is a not-yet encountered state of nature described probabilistically as a continuous random variable S, with p_{S} being a specification of current beliefs about the possible states (in the form of a probability density function if the state space is continuous, and the probability mass function if the state space is discrete). If the probability density over the various states of nature is available, the decision problem under non-certainty is converted to one under uncertainty.

Bayesian decision theory strives for good results on average. According to the minimum expected loss principle,[3] a rational decision maker is interested in selecting the option that minimises expected value of loss.

Proposition 1.1.6 (Minimum expected loss principle [12]). *When nature is described by a random variable* S*, a decision-maker with loss* $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$*, should choose an option* $a \in \mathcal{A}$ *that minimizes their expected loss, i.e.*

$$a_\phi^{UU} = \arg\min_{a \in \mathcal{A}} \mathbb{E}[L(a, \mathrm{S})] \tag{1.5a}$$

$$= \arg\min_{a \in \mathcal{A}} \int L(a, s) p_{\mathrm{S}}(s) \mathrm{d}s \tag{1.5b}$$

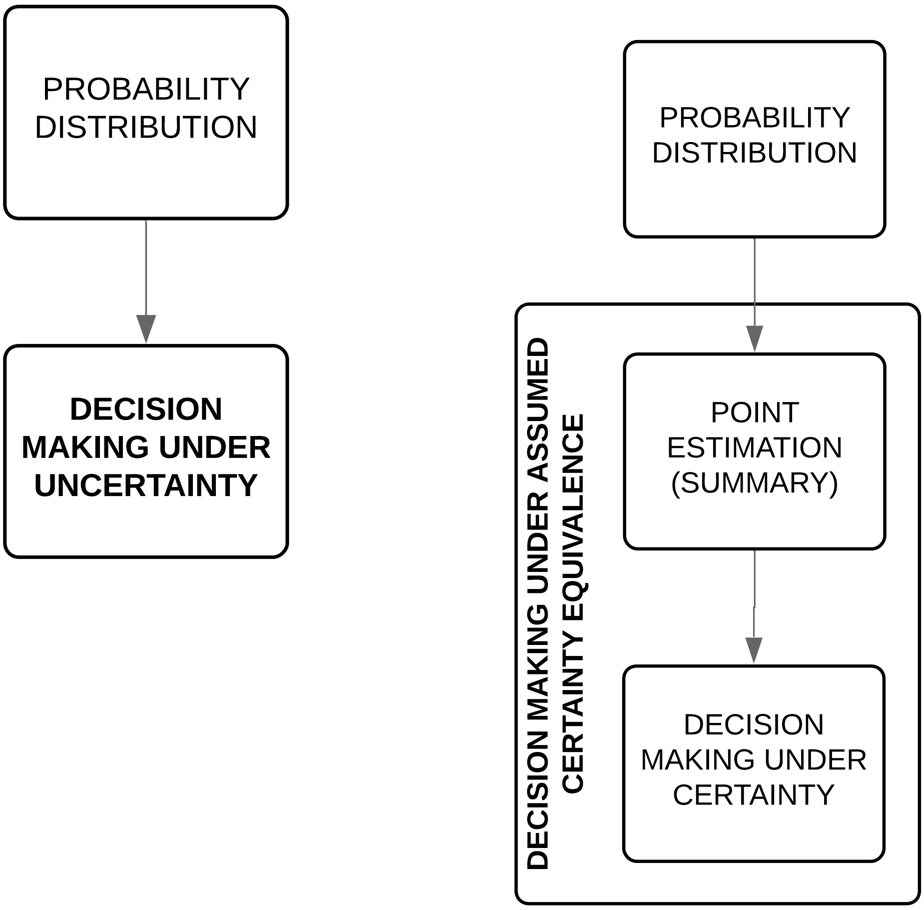
that corresponds to the optimised expected loss value

$$\rho_\phi^{UU} = \mathbb{E}[L(a_\phi^{UU}, \mathrm{S})]. \tag{1.6}$$

The pair $(a_\phi^{UU}, \rho_\phi^{UU})$ forms what we call an optimal solution to decision under uncertainty. Note that a stand alone optimal option a^{UU} would not form a *complete* solution if not accompanied by ρ^{UU}, which is an indicator of decision's quality.

Statement (answer) vs command (non-sensing action) It is important to recognize that the formalism of decision making under uncertainty is applicable for both guiding actions, and producing answers or statements [12]. When decision making concerns supporting a non-sensing action, the resulting decision is a command, e.g. a decision to take or leave an umbrella (depending on the weather forecast representing the nature) in the umbrella problem, or a decision concerning the level of stocks (depending on the demand representing the nature) in the inventory problem. When decision making concerns producing an answer (related to the uncertain state of nature described by p_{S}), the outcome is a statement, e.g. summarizing the current beliefs in a form of a point estimate.

[3] Alternatively, the maximum expected utility principle.



(a) Decision making under uncertainty.

(b) Decision making under assumed certainty equivalence.

Figure 1.2: Probabilistic decision making.

1.1.6 Decision making under assumed certainty equivalence

In this section we formalize a certain heuristic that is used for decision making when probabilities are available. It substitutes the procedure of decision making under uncertainty for the terminal decision (Figure 1.2a) by a cascade of two other decision procedures (Figure 1.2b): the problem of point estimation (decision making under uncertainty) and the actual terminal decision (decision making under certainty).

There may be various benefits to organizing decision making this way. For example, it results into optimization procedure that is simply not as involved as that resulting from following the minimum expected loss principle [28, p. 19] of Proposition 1.1.6. Otherwise, it is possible that the two problems will belong to two different decision makers (thus it is possible that the specific form of terminal loss is not disclosed to the decision maker that solves the problem of point estimation). It is also possible that decision are made at distinct moments of time (thus the knowledge of terminal loss is not required when the first estimation decision is made [12, p. 14]). From this perspective, decision making under uncertainty as described in Section 1.1.5 can be seen as excessively intrusive and restrictive.

Proposition 1.1.7 (Minimum loss principle under certainty equivalence). *When the state of nature is described by a random variable* S*, a decision maker with loss* $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$ *should obtain a summary* $\hat{s} \in \mathcal{S}$ *of the state and then choose an option* $a \in \mathcal{A}$ *that minimizes their losses with respect to the summary* $\hat{s}$*, i.e.*

$$a_\phi^{CE} = \arg\min_{a\in\mathcal{A}} L(a, \hat{s}), \tag{1.7}$$

that corresponds to the optimized loss value

$$\rho_\phi^{CE} = L(a_\phi^{CE}, \hat{s}). \tag{1.8}$$

Unfortunately, the solution will not be Bayes-optimal in general (some cases will be mentioned later in Section 1.3). So it is only an *assumption* that the solution is equivalent to the optimal solution produced following the minimum expected loss principle. This is known as the assumed (or forced, or heuristic) certainty equivalence design technique [72, p. 241], [83, Sec. 6.2.1]. Accordingly, certain favourable features of decision making under uncertainty, such as evaluation of decision quality in the form of Bayes expected loss could not be recovered in such settings. As mentioned in [92, 100] for many problems the eventual sub-optimality is often judged acceptable and could be in many cases tolerated.

1.2 Integrating sensing with decision making under uncertainty

The previous section has introduced the idea of *terminal* decision making, and it hasn't addressed the possibility of collecting new observations (data) before making a decision. In this section we are going to consider situations when it is possible to collect new data before making the terminal decision The focus will be on decisions made under uncertainty, i.e. when prior knowledge is available. In principle, Bayesian decision theory would distinguished two radically different procedures: terminal analysis (that is decision making under uncertainty for the posterior density conditioned on new data) and preposterior analysis (that concerns choosing on a way to collect new data).

Ultimately, these procedures can be attributed to an agent operating in an unknown environment (nature), see Figure 1.3. The agent can acquire information about its environment using a sensor. However, sensor measurements are noisy, and there are usually many things that cannot be sensed directly. As a result, the agent maintains some belief about the state of nature. The agent can make statements about the environment and also influence the environment through its effectors. These processes will now be described more formally.

1.2.1 Terminal analysis

Probabilistic decision making is associated with situations where probabilistic description of the state of nature is available. Let us first focus on using new observations in decision making under uncertainty in Subsection 1.1.5.

If additional information $z \in \mathcal{Z}$ (where $\mathcal{Z}$ is an observation space) is obtained which is probabilistically related to s by $p(z|s)$, then the best option a is that which minimizes the *posterior* expected loss.

Definition 1.2.1 (Bayes theorem)**.** *The posterior probability density describing a random variable* S *is given by*

$$p_{\mathrm{S}}(s|z) = \frac{p(z|s)p_0(s)}{\int p(z|s)p_0(s)\mathrm{d}s}, \tag{1.9}$$

where $p_0(\cdot)$ is the prior probability density, and $p(\cdot|\cdot)$ is a measurement likelihood.

Proposition 1.2.2 (Minimum expected posterior loss principle [9])**.** *When nature is described by a random variable* S *and its posterior probability density $p_{\mathrm{S}}(s|z)$, a*

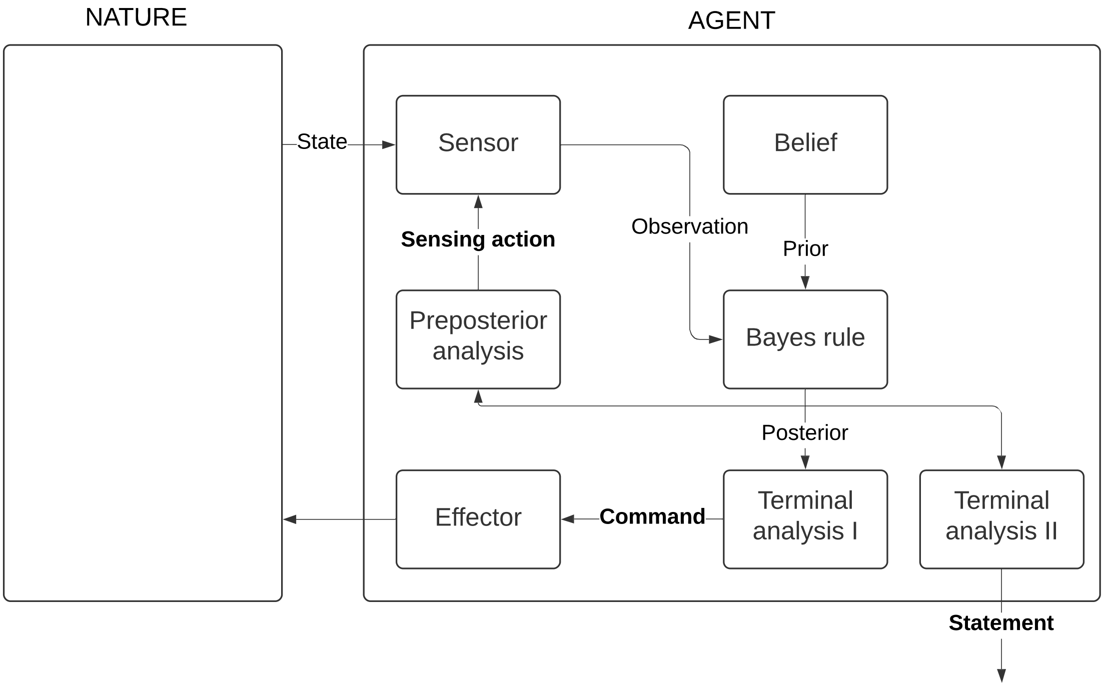


Figure 1.3: Integrating sensing with decision making under uncertainty. Note that terminal analysis, depending on the problem, can lead either to an action (via a command), or a statement; and the sensing action is selected in a different procedure called preposterior analysis.

decision-maker with loss $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$*, should choose an option* $a \in \mathcal{A}$ *that minimizes their posterior expected loss, i.e.*

$$a^{UU} = \arg\min_{a \in \mathcal{A}} \int L(a, s) p_{\mathrm{S}}(s|z) \mathrm{d}s, \tag{1.10}$$

that corresponds to the optimised posterior *expected loss value*

$$\rho^{UU} = \min_{a \in \mathcal{A}} \int L(a, s) p_{\mathrm{S}}(s|z) \mathrm{d}s \tag{1.11a}$$

$$= \int L(a^{UU}, s) p_{\mathrm{S}}(s|z) \mathrm{d}s. \tag{1.11b}$$

Remark 1.2.3. *In [9] the optimal action* a^{UU} *is termed Bayes action, and the optimised value of expected loss* ρ^{UU} *is Bayes expected loss.*

One way to refer to decision making under uncertainty is *terminal analysis*; the decision maker uses the posterior probability density (1.9) to find the best option from the set of terminal acts $\mathcal{A}$. It is important to distinguish this procedure from *preposterior analysis*, which is concerned with selecting the mode of sensing used to collect the observation z, and commonly studies as 'sensor management'.

1.2.2 Preposterior analysis

This analysis is nonterminal since it will be eventually followed by the terminal decision, like deciding on a command that triggers a non-sensing action or deciding on a statement.

Definition 1.2.4 (Preposterior analysis)**.** *The optimal sensing action is given by [64, Eq. 4.7], [18, Eq. 2]*

$$u^* = \arg\min_{u \in \mathcal{U}} \int_{\mathcal{Z}} \min_{a \in \mathcal{A}} \int_{\mathcal{S}} L(a, s) p_{\mathrm{S}}(s|z, u) p(z|u) \mathrm{d}s \mathrm{d}z, \tag{1.12}$$

where $p(z|u)$ *is a measurement likelihood conditioned on the sensor control input* u*.*

Note that (1.11a) explicitly enters the expression (1.12) used for finding the best possible sensing action, and this is the reason why the ability to compute the expected loss value (and not only optimal terminal action) is essential for sensor management.

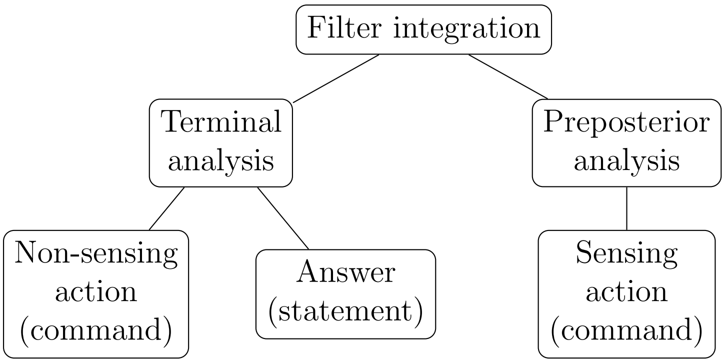


Figure 1.4: Informal mapping of various integration instances for a Bayesian filter.

1.3 Exploitation of Bayesian filters

Recursive Bayesian filters is a type of sensor data processing algorithms that are aimed to provide a probabilistic description of an uncertain dynamic system using partial sensor observations (Figure 1.5).[4] An important element of such filters is the model of system dynamics, which captures the system's evolution over time and has a capacity to model non-sensing actions that affect the system evolution. A variety of ways to employ Bayesian filters is sketched on Figure 1.4, and is discussed next.

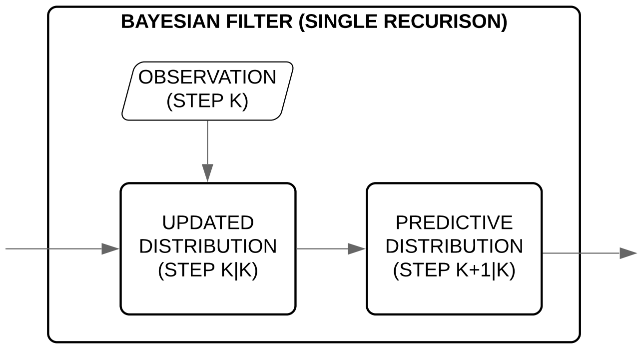


Figure 1.5: A single recursion of a Bayesian filter.

1.3.1 Related to terminal analysis

When considering Bayesian filters for terminal analysis, it should be made clear that decisions falling into this category are those which are dependent on the state of the

[4] Bayesian filters can be seen as special kind of signal processing algorithms, which are often employed in algorithmic solutions with the purpose of state estimation. As mentioned by Gustafsson in [37], these solutions are "conceptually different (although algorithmically similar)" to those of signal estimation and fault detection, however "the close links between these areas are clearly under-estimated in literature."

uncertain system but do not lead to the change in our knowledge about the said system.[5] Specifically, we distinguish two groups of problems in this category:

(1) those that (potentially) lead to an impact to the dynamic system;

(2) those that result into a statement (or a report) about the system.

The first group comprises problems of stochastic optimal control where the focus is on producing a non-sensing action, command or control law, that ensures that the dynamic system behaves in a desirable way, e.g. in guidance and navigation. In this case the decision's impact is explicitly taken into account within the model of the dynamic system, see e.g. an elementary exposition in [30]. Furthermore, there is a closely related body of problems, which approach the decision's impact differently and do not account for it in the model of the dynamic system. Instead, the sought decision directs an external effector to affect the system in the open loop (or 'fire and forget') manner, such as in the cases of firing a weapon [81], dropping a package [116, Sec. 2.3], or cueing an external sensor [55, 77].

The second group of problems is concerned with producing a statement related to the state of the uncertain dynamic system, and does not ultimately lead to an impact or to performing an action of any kind. In context of Bayesian filtering, the predominant problem of this kind is focused on producing a point estimate of the system state [44]. Filtering information is rarely used for decision making under uncertainty that leads to a statement (as opposed to a command) beyond point estimation, with some notable exceptions in [25, 51, 59, 119].

Overall, the textbooks on Bayesian filtering are commonly focused on a single type of terminal decision that concerns producing a point estimate from the filtering information. Consider the following statements:

- "distributions alone have no use in many practical applications; we need finite-dimensional summaries (point estimates)" [93, p. 20];
- "a lot of practical and operational applications require a point estimate" [14];
- "without a Bayes-optimal estimator of the multitarget state... the information in... [multitarget posterior density] is not available for practical use" [68].

Although the possibility of using filtering densities for terminal analysis is not completely ruled out, it appears that terminal decision making is most commonly

[5] This is in contrast to preposterior analysis considered next, where produced decisions are not only related to the state of the system, but also affect the knowledge about the system.

performed based on point estimates (in place of full distributions). Using point estimates is also continent for human operators and decision makers. This brings us to the following conclusion. Unless the point estimate is in the focus of terminal decision making, Bayesian filters are primarily exploited under assumed certainty equivalence, as discussed in Subsection 1.1.6. In other words, the produced decisions do not minimise the posterior expected loss (following Proposition 1.2.2), and generally are sub-optimal.

Remark 1.3.1 (When certainty equivalence holds). *There are certain conditions under which this isomorphism between the optimal decisions under uncertainty and the optimal decisions in an equivalent certainty context (using a point estimate) is valid [28, p. 19]. Specifically, certainty equivalence is Bayes-optimal for the linear-quadratic-Gaussian*[6] *(LQG) problems in optimal control [4], where the state of the world is summarized by its mean value. However, once the LQG conditions are relaxed certainty equivalence no longer applies. As far as state of the art filtering algorithms are concerned, the LQG conditions are not valid since extracting the mean of the distribution is not meaningful [25, 70].*

1.3.2 Related to preposterior analysis (sensor management)

The process of selecting sensing actions is commonly referred to as sensor management or, within the context of Bayesian statistics, preposterior analysis (Figure 1.4). Ultimately, such actions *improve our knowledge about the uncertain dynamic system.* In context of Bayesian filtering, the problem of sensor management is often introduced as a Partially Observable Markov Decision Process (POMDP) [39], thus stressing a possibly sequential nature of the management process. However, Aoki et al. [2] avoid describing the problem as POMDP because the term "does not distinguish the sensor management problem from the regular control problem" that is focused on actions affecting the state of the controlled dynamic system.

Overall, the literature on sensor management is vast and poorly systematized. Ultimately, the discussion on various approaches to sensor methods is focused on the formulation of an optimization objective. According to Kreucher et al. [57], Bayesian approaches to sensor management can be divided into *information-driven* and *task-driven* approaches.

[6] Where the world dynamics are linear, the terminal loss is squared error, and the process noise is additive Gaussian.

One of the most popular ways to formulate a management objective is by using measures of uncertainty developed in information theory, this is the reason why management approaches of this kind are referred to as information-driven [3, 87]. The idea behind this approach is to employ information measures in place of the expression of Bayes expected loss in (1.12). Despite their theoretical appeal, such algorithms are difficult to justify in practice as it is not clear whether the optimization objective supports the terminal decision.

Other Bayesian approaches are commonly referred to as task-driven. Although references like [57] attribute a considerable number of possible objectives to this category, the common example that uses a formalization compatible with (1.12) is focused on point estimation of the system state [2]. Such approaches are appealing in practice, since they offer a way to take the terminal decision into account when selecting a sensing action. Furthermore, they offer a possibility to formulate sensor management as the problem of minimizing the sensor resources spent, see e.g. [48, 119], complementing the direct problem of maximizing efficiency of a given sensor resource.

1.4 Problem formulation

Problem statement Bayesian filters are traditionally exploited in terminal analysis and preposterior analysis such as if terminal decision is concerned exclusively with the problem of point estimation of the system state. However, in practice the problems are diverse and may be distinct from that of point estimation. As a result, Bayesian filters are exploited in those distinct terminal decision procedures under assumed certainty equivalence. This leads to the situation where results of terminal decision making are suboptimal (underinformed and of unknown quality), and sensor management is misdirected (does not acknowledge the actual problem).

Objective To exploit Bayesian filters for terminal analysis in problems requiring finding optimal statements (as opposed to optimal commands), which cannot be reduced to the problem of point estimation.

Challenges Formulation of a problem such that it could be eventually approached in the context of Bayesian filters is complicated by a number of factors:

(i) *Complicated state space.* The set of all possible states of nature is one of the key elements in the formulation of a decision procedure. Bayesian decision theory is commonly concerned with Euclidean spaces. In contrast, the new generation of Bayesian filters is operating on state spaces required to accommodate realization of a point process [98], which are not metric spaces. For such spaces, even the simple problem of point estimation could not be easily resolved; the consequence is that most algorithms rely on heuristics to produce point estimates.

(ii) *Dependent act space.* Bayesian decision theory and Bayesian filtering rely on a number of overlapping elements, including the probability density function and the underlying state space. One element that is characteristic for a decision procedure is the act space. It is a known fact that decision theory literature is predominantly focused on problems characterised by loss functions that require that the space $\mathcal{A}$ coincides with the state space $\mathcal{S}$ [89]; in turn, $\mathcal{S}$ is either selected such as $\mathcal{S} \subseteq \mathbb{R}^d$ with $d \in \mathbb{N}$ being the number of coordinates, or $\mathcal{S} = \{0, 1\}$, i.e. both $\mathcal{A}$ and $\mathcal{S}$ consist of two points. This corresponds to the standard problems of point estimation[7] and of detection (the testing of hypothesis)[8]. Nevertheless, in general there is no requirement for two spaces to be dependent [113], and statistical decision theory should be able to address problems which in the words of Wald "have not yet been treated" [113]. And the consequence is that it may be difficult to express considerably new decision procedures, i.e. those where the act space is not dependent on the state space, within the standard framework.[9]

(iii) *Missing loss functions.* Although it is claimed that statistical decision theory "formally encompasses an enormous range of problems" [10], literature is predominantly concerned with "certain standard loss functions" [9] that also lead to the point estimation problem. As far as the current practice of decision making with Bayesian filters is concerned, it has decoupled the filtering algorithm from the problem being solved, and effectively avoided the need to consider, let alone design, loss functions. In addition to that, the context of Bayesian filtering contains a number of ad hoc solutions, which could have been presented within the general framework of Bayesian decision theory, but instead have been formulated as detection or point estimation procedures on

[7] This includes functions such as squared error, zero-one error, absolute error, LINEX, BLINEX, and others e.g. [79, 99, 103, 106, 115].

[8] Notably, the third standard problem of interval estimation cannot be adequately expressed in decision-theoretic terms [32, 63].

[9] Note that decisions described here are strictly focused on choosing terminal actions, and do not concern sensing actions, i.e. experiments.

alternative state spaces, e.g. [51, 59], or as statistical inference procedures, e.g. [25].

(iv) *Limited probabilistic information.* The settings of Bayesian decision theory commonly assume that the complete information about the probability density is available in some form, possibly via sufficient statistics. In contrast, filtering solutions, especially coming from the new generation of filters, often rely on propagation of statistical quantities, which may be insufficient for decision making.

1.5 Proposed approach

In this thesis we address the challenges presented in the previous section by proposing a new definition of a loss function as a function composition. We attempt to decouple the act space from a possibly complicated state space, and introduce the possibility of reconfiguring the loss function and formalising new problems leading to decision making under uncertainty.

Definition 1.5.1 (Loss function)**.** *Loss function is a function $L : \mathcal{A} \times \mathcal{S} \to \mathbb{R}_0^+$ that attaches a value of loss to every possible outcome of selecting an option $a \in \mathcal{A}$ when the state of nature is $s \in \mathcal{S}$, and is defined as a composition*

$$L(a, s) := l(a, q(s)), \tag{1.13}$$

where $q : \mathcal{S} \to \mathcal{A}$ is a query function *attaching an ideal option $\breve{a}$ to each possible state of nature s, and $l : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_0^+$ is a* query loss function *attaching losses to each outcome $(a, \breve{a})$ of accepting option a when the ideal choice is $\breve{a}$.*

The value of introducing of mapping $q : \mathcal{S} \to \mathcal{A}$ has been once recognized in the domain of robotics,[10] where it is simply referred to as 'a transformation' and was used to communicate the problem-specific information needs. Specifically, the following transformations have been highlighted [38]: parameter reductions, reductions to discrete spaces, continuous transformations, combinations of the above.

Overall, we found it useful to think of the mapping $q : \mathcal{S} \to \mathcal{A}$ as a problem-specific or a user-defined query, or information need. For example, it could define an externally specified feature, e.g. by taking into account contextual information that does not belong to the model of nature. Accordingly, the problems could be

[10] The work in [38] was discovered during the preparation of the thesis. However, it does not integrate the query function within the overall loss function L, as followed by this thesis.

formalised in a new way: using function q to communicate a 'query', and function l to describe losses associated with making errors in the 'answer' (e.g. overestimating or underestimating). It is also easy to see that the revised formulation is equivalent to the standard definition in Definition 1.1.3 if q is a linear function on $\mathcal{S}$.

To the best of our understanding, in the context of Bayesian statistics such formulations (where the same prior is used for distinct problems expressed via appropriate terminal loss functions) have been avoided because of the difficulties associated with formulating prior information. Specifically, a different prior might be used depending on the considered problem [11], because of the fear that the prior will dominate the data. Fortunately, in Bayesian filtering objective prior information is always available. As described in [35, p. 11], this may be the maximum detection range of a sensor, or the maximum achievable speed of an object.

Subsequent design choices The developments in this work are guided by the following overarching principles and set limitations:

(a) The state space is determined by the considered filtering algorithm, where it is defined as the minimal set of data that is sufficient to uniquely describe the dynamical behaviour of the system.

(b) It may not be possible to produce solutions for arbitrary act spaces, so additional limitations will have to be introduced (e.g. consider $\mathcal{A} = \mathbb{R}$).

(c) The same form of a query function may be appropriate for more than a single problem. Therefore, it is reasonable to introduce query functions formally, without justifying their physical meaning.

(d) Query loss functions essentially fulfil the same role as loss function used in standard procedures, e.g. point estimation (see Appendix A.1), and therefore the standard models of loss can be re-used.

1.6 Thesis outline and key contributions

Chapter 1 This chapter introduces the subject of this work. Specific contributions are:

- Section 1.1 offers the author's perspective on various realms of decision making, with the specific focus on the integration of the loss function, tracing the origins to game theory:

- Section 1.1.6 articulates the principle of decision making under assumed certainty equivalence, a simple heuristic that is commonly used in place of the minimum expected loss principle.

- Section 1.5 offers a revised definition of the loss function as a composition of a query function and a matching query loss function.

Chapter 2 This chapter gives an introduction to recursive Bayesian filtering for the classical case of a single surely present dynamic object, and demonstrates application of filtering information in decision making under uncertainty. The novel contributions are:

- Subsection 2.3.3 interprets ad hoc solutions to various problems, including the problems of threat assessment and binary classification, using the new formalization of the loss function.

Chapter 3 This chapter discusses recursive Bayesian filtering for the case of dynamic object populations, introduces the point process formalism, and demonstrates application of filtering information in decision making under uncertainty. The novel contributions are:

- Subsection 3.5.3 interprets the problem of regional cardinality estimation within the new decision-making formalism.

Chapter 4 This chapter uses the new decision-making formalism to develop optimal solutions in the context of spatial point processes. The novel contributions are:

- Section 4.1 presents an optimal solution to the class of problems expressed with an arbitrary real-valued query function and the squared error query loss.

- Section 4.2 introduces the sum query, and obtains the corresponding solution which can be expressed via the lower-order statistical moment of a point process.

- Section 4.3 introduces the product query, and obtains the corresponding solution which can be expressed via the probability generating functional (p.g.fl.) of a point process.

Chapter 5 This chapter implements the optimal solutions for a number of practical Bayesian filters, including the classical PHD filter, the Panjer PHD filter, and the Cardinalized PHD filter. The novel contributions are:

- Section 5.2 develops expressions of certain lower-order statistical moments and p.g.fl.s that correspond to the updated and predicted distributions in the considered filters. Note that expressions of the first-order moment is not a novel result.
- Section 5.3 obtains solutions corresponding to the sum query:
  - Section 5.3.1 produces expressions for the updated point process.
  - Section 5.3.2 produces expressions for the predictive point process.
- Section 5.4 obtains solutions corresponding to the product query:
  - Section 5.4.1 produces expressions for the updated point process.
  - Section 5.4.2 produces expressions for the predictive point process.

Chapter 6 This chapter is dedicated to demonstration of the developed solutions using simulated data, with the focus on the update step of the SMC-PHD filter. The discussion is focused on the problem of subjective decision-theoretic inference. The novel contributions are:

- Section 6.2 contains a parallel presentation of solutions developed for three distinct realms of decision making (and includes the pseudocode for implementation):
  - Section 6.2.1 is focused on decision making under certainty, which produces ideal (clairvoyant) answers for the known ground truth.
  - Section 6.2.2 is focused on conventional algorithm for decision making under assumed certainty equivalence, which produces suboptimal answers using heuristics for extracting the system state.
  - Section 6.2.3 is focused on developed algorithms for decision making under uncertainty, and implements solutions developed in Sections 5.3.1 and 5.4.1.
- Section 6.3 offers a number of practical query functions (including those based on the developments in Appendix A.3) which give rise to three meaningful problems of subjective decision-theoretic inference.

- Section 6.4 provides the simulation results that, for the developed algorithms, demonstrate the consistency of quality indicators and the capacity to outperform the conventional algorithm.

Chapter 7 This chapter provides a summary of the developments and offers possible directions of the future research.

Appendix A.1 This appendix presents certain query loss functions used in this work.

Appendix A.2 This appendix presents a concept of threat function, which models the probability that an object produces a negative impact on a threatened asset.

Appendix A.3 This appendix constructs a probabilistic model describing the value of damage incurred by a vulnerable asset as a result of a simultaneous detonation of multiple weapons. Specific contributions are:

- Section A.3.1 develops a general expression of the expected damage value, denoted as *risk*.
- Section A.3.2 develops a special case of the expression in the additive form, denoted as *sigma-risk*.
- Section A.3.3 develops a special case of the expression in the multiplicative form, denoted as *pi-risk*.

Publications An early technical result, which corresponds to developments in Sections 4.2.1 and 5.3.1, has been reported in the form of a conference publication:

- A. Narykov, E. Delande, D.E. Clark, P. Thomas, and Y. Petillot. Second-Order Statistics for Threat Assessment with the PHD Filter. In 2017 Sensor Signal Processing for Defence Conference (SSPD), pp. 1–5. IEEE, 2017.

Chapter 2

Recursive filtering for a single object

One of the key challenges for decision making under uncertainty is in obtaining the probability distribution of various states of nature. A body of algorithms that are successfully dealing with this issue, albeit in dynamic settings, are associated with recursive Bayesian estimation. Such algorithms, commonly called Bayesian filters, are constructed with the aim to sequentially describe the state evolution of an uncertain dynamic system using partial observations.

This chapter is based on a standard exposition of the Bayes filter, which originates from [40], and focuses on a system that is represented by a single *surely present* dynamic object.[1] It also addresses decision-making with the Bayes filter, and makes use of the revised formulation from Definition 1.5.1. It will be used to *analyse* the standard problem of point estimation, which is associated with producing a system state summary that removing the accumulated uncertainty. Furthermore, it will be used to *synthesize* a number of other problems previously not considered as alternatives to the point estimation.

The content of this chapter is as follows. In Section 2.1 we discuss the concept of optimality in the filtering context. In Section 2.2 we present the modelling details that lead to a single-object Bayes filter. In Section 2.3 we move on to address decision-making with the Bayes filter. Section 2.4 provides the summary.

[1] By stating that the object is 'surely present' we explicitly exclude a significantly more advanced Bernoulli filter [88] from the consideration. This filter is sometimes described as a 'single-object filter' [112], but conceptually belongs to the new generation of filters for object populations described next in Section 3.

2.1 Optimal filtering

Recursive Bayesian estimation, with a prominent example of the Kalman filter [40, 49], is arguably one of the most successful application of Bayesian statistics. The Bayesian filter is a recursive algorithm that generates a probabilistic description of an uncertain dynamic system by incorporating all information that can be provided to it. It processes all available observations, regardless of their accuracy, with the use of [71]:

- knowledge of the models describing the object dynamics and measurement systems in the absence of noise;
- the statistical description of the process and observation noises;
- all information that is available about initial object state.

The algorithm is called *recursive*, or *sequential*, because it is capable of constructing posterior distribution from arriving observations, rather than operating on a complete sequence.

One of the central concepts in recursive filtering is optimality. A filter is called *optimal*,[2] if it seeks a probability distribution that is "correctly calculated" [56, Sec. 3.1.1], it is "exact and complete" [85, p. 6]. An optimal filter can be exploited for producing point estimates that would be optimal in Bayes sense, i.e. a resulting estimate minimizes the expected value of some explicitly specified loss function.[3] However, in general, it is not necessary for a filtering algorithm to address the problem of point estimation. Additionally, one can consult the following statements:

- "this conditional distribution offers a complete solution to the filtering problem" [44, p. 145-146];
- "the purpose of Bayesian filtering is to compute the marginal posterior distribution or filtering distribution of the state" [93, p. 54];
- "the optimal solution to the nonlinear filtering problem requires that a complete description of the conditional[4] probability density is maintained" [47].

[2]Sometimes optimal filters are referred to as exact filters [24, 85].

[3]It may also happen that filter produces optimal point estimates in the process of its operation, as it is in the Kalman filter, which is a minimum mean square error (MMSE) estimator.

[4]Conditioned on all available observations.

2.2 Bayes filter for a surely present object

Let us consider the discrete-time state-space approach to the modelling of dynamical systems. At each time step $k \in \mathbb{N}$, the system is described by its state vector x_k that takes values in a state space $\mathcal{X} \subseteq \mathbb{R}^{n_x}$, where n_x are dimensions of the vector x_k. The system state itself is hidden, but its noisy measurement vectors z_k in measurement space $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ are available, where n_z are dimensions of the vector z_k.

The time evolution of the state vector is described by a stochastic model in the form

$$x_k = f_k(x_{k-1}, v_{k-1}), \tag{2.1}$$

where f_k is a possibly nonlinear function of the state x_{k-1}, and $v_{1:k}$ is an i.i.d. process noise sequence. Another way to describe the time evolution is using a Markov transition kernel $\pi_k(\cdot|\cdot)$.

The measurement vector is modelled using a measurement equation given by

$$z_k = h_k(x_k, w_k), \tag{2.2}$$

where h_k is a possibly nonlinear function of the state x_k, and $w_{1:k}$ is an i.i.d. measurement noise sequence. This model specifies at time k how any given state vector x_k and measurement noise w_k are taken into a measurement vector z_k. Another way to describe the measurement process is by a likelihood function $g_k(\cdot|\cdot)$, where $g_k(z_k|x_k)$ evaluates the adequacy of the state x_k when guessed against the measurement z_k.

Recursive filtering is concerned with sequentially describing the uncertain state of the dynamic system from the measurement history $z_{1:k} = (z_1, \cdots, z_k)$. Using the models introduced above, the required probability density function $p_k(x_k|z_{1:k})$ can be recursively propagated by the Bayes filter in two stages [40]:

$$p_{X_{k|k-1}}(x_k|z_{1:k-1}) = \int \pi_{k|k-1}(x_k|x) p_{X_{k-1}}(x|z_{1:k-1}) \mathrm{d}x, \tag{2.3}$$

$$p_{X_k}(x_k|z_{1:k}) = \frac{g_k(z_k|x_k) p_{X_{k|k-1}}(x_k|z_{1:k-1})}{\int g_k(z_k|x) p_{X_{k|k-1}}(x|z_{1:k-1}) \mathrm{d}x}, \tag{2.4}$$

for an initial distribution $p_{X_0}(x_0)$. These Bayesian discrete-time recursive equations constitute the foundation for optimal filtering for a single surely present dynamic object when observed by a sensor with unity detection probability.

2.3 Decision making under uncertainty

This chapter uses Definition 1.5.1 to explore decision-making using information from the Bayes filter. In Subsection 2.3.1, we outline a decision-making process leveraged by information from the Bayes filter. Subsection 2.3.2 analyses point estimation as a special case of a decision procedure. In Subsection 2.3.3, we synthesise a number of decision-making procedures that solve problems other than standard point estimation.

2.3.1 Revised decision procedure

Given the posterior density p_{X_k} (or equally predictive density $p_{X_{k|k-1}}$), it is possible to solve any problem for a formalized decision process in Definition 1.5.1. This requires that the act space $\mathcal{A}$ as well as the loss function $L : \mathcal{A} \times \mathcal{X} \to \mathbb{R}_0^+$, which is a composition of a query function $q : \mathcal{X} \to \mathcal{A}$ and a query loss function $l : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_0^+$, are *additionally* introduced, as they are not specified by the model in Section 2.2.

Proposition 2.3.1 (Posterior Bayes-optimal solution)**.** *For an uncertain system described by a random vector X_k on $\mathcal{X}$, a Bayes-optimal solution to the problem formalized by $L : \mathcal{A} \times \mathcal{X} \to \mathbb{R}_0^+$ is a tuple (a_k, ρ_k), where posterior Bayes action a_k is given by*

$$a_k = \arg\min_{a \in \mathcal{A}} \mathbb{E}[L(a, X_k)] \tag{2.5a}$$

$$= \arg\min_{a \in \mathcal{A}} \int l(a, q(x)) p_{X_k}(x|z_{1:k}) \mathrm{d}x, \tag{2.5b}$$

and posterior Bayes expected loss ρ_k is given by

$$\rho_k = \mathbb{E}[L(a_k, X_k)] \tag{2.6a}$$

$$= \int l(a_k, q(x)) p_{X_k}(x|z_{1:k}) \mathrm{d}x. \tag{2.6b}$$

Proposition 2.3.2 (Predictive Bayes-optimal solution)**.** *For an uncertain system described by a random vector $X_{k+1|k}$ on $\mathcal{X}$, a Bayes-optimal solution to the problem formalized by $L : \mathcal{A} \times \mathcal{X} \to \mathbb{R}_0^+$ is a tuple $(a_{k+1|k}, \rho_{k+1|k})$, where predictive Bayes action $a_{k+1|k}$ is given by*

$$a_{k+1|k} = \arg\min_{a \in \mathcal{A}} \mathbb{E}[L(a, X_{k+1|k})] \tag{2.7a}$$

$$= \arg\min_{a \in \mathcal{A}} \int l(a, q(x)) p_{X_{k+1|k}}(x|z_{1:k}) \mathrm{d}x, \tag{2.7b}$$

and predictive Bayes expected loss $\rho_{k+1|k}$ *is given by*

$$\rho_{k+1|k} = \mathbb{E}[L(a_{k+1|k}, X_{k+1|k})] \tag{2.8a}$$

$$= \int l(a_{k+1|k}, q(x)) p_{X_{k+1|k}}(x|z_{1:k}) \mathrm{d}x. \tag{2.8b}$$

2.3.2 Point estimation of the system state

Point estimation of the system state is often perceived as an inference procedure. However, it is recognized that it also has an interpretation as a decision-making procedure, see e.g. [12]. Accordingly, point estimation of the system state is a singular most popular decision-making procedure in the context of Bayesian filtering that is being solved optimally.

Bayes-optimal point estimation Bayes-optimal state estimation is a special kind of a problem, where the act space is selected to coincide with the state space, i.e. $\mathcal{A} = \mathcal{X}$. From the perspective of Definition 1.5.1, the problem of point estimation of a system state is characterised by the fact that the query function is specified as a simple identity function on the corresponding state space.

Definition 2.3.3 (Identity query)**.**
An identity query function $q_{\mathbb{I}} : \mathcal{X} \to \mathcal{X}$ *is given by*

$$q_{\mathbb{I}}(x) := \mathbb{I}_{\mathcal{X}}(x) \tag{2.9a}$$

$$= x. \tag{2.9b}$$

Proposition 2.3.4 (Posterior Bayes-optimal state estimation)**.** *For an uncertain system described by a random vector* X_k *on* $\mathcal{X}$*, a Bayes-optimal solution to the problem of point estimation is given by a tuple* $(\hat{x}_k, \rho_k)$*, where posterior Bayes estimate* $\hat{x}_k$ *is given by*

$$\hat{x}_k = \arg\min_{x_k \in \mathcal{X}} \mathbb{E}[l(x_k, q_{\mathbb{I}}(X_k))] \tag{2.10a}$$

$$= \arg\min_{x_k \in \mathcal{X}} \int l(x_k, x) p_{X_k}(x|z_{1:k}) \mathrm{d}x, \tag{2.10b}$$

and posterior Bayes expected loss ρ_k *is given by*

$$\rho_k = \mathbb{E}[l(\hat{x}_k, X_k)] \tag{2.11a}$$

$$= \int l(\hat{x}_k, x) p_{X_k}(x|z_{1:k}) \mathrm{d}x, \tag{2.11b}$$

for a query loss function of a kind $l : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_0^+$.

Note that we explicitly stated the value of posterior Bayes expected loss ρ_k as a part of the solution. In much of literature this quantity is dismissed, and the discussion is stressed on the estimate itself. This is unfortunate because expected loss is an indicator of the estimation quality. It may be used for other kind of decisions, such as of dismissing/accepting the estimate, or requesting additional measurements.

A popular algorithm is the minimum mean squared error estimation given in Example 2.3.5.

Example 2.3.5 (MMSE estimation)**.** *When the query loss function is modelled as squared error function* l_2 *in* (A.1)*, a Bayes-optimal solution of the estimation problem is a tuple* $(\hat{x}_k^{MMSE}, \rho_k^{MMSE})$ *is simply the conditional mean of the distribution given by*

$$
\begin{aligned}
\hat{x}_k^{MMSE} &= \arg\min_{x_k \in \mathcal{X}} \mathbb{E}[l_2(x_k, X_k)] && (2.12a)\\
&= \arg\min_{x_k \in \mathcal{X}} \int (x_k - x)^2 p_{X_k}(x|z_{1:k}) \mathrm{d}x, && (2.12b)\\
&= \int x p_{X_k}(x|z_{1:k}) \mathrm{d}x && (2.12c)\\
&= \mathbb{E}[X_k] && (2.12d)
\end{aligned}
$$

with the associated Bayes expected loss given by

$$
\begin{aligned}
\rho_k^{MMSE} &= \int (\hat{x}_k^{MMSE} - x)^2 p_k(x|z_{1:k}) \mathrm{d}x, && (2.13a)\\
&= \int x^2 p_{X_k}(x|z_{1:k}) \mathrm{d}x - \left[\int x p_{X_k}(x|z_{1:k}) \mathrm{d}x \right]^2 && (2.13b)\\
&= \mathbb{E}[X_k^2] - \mathbb{E}[X_k]^2 && (2.13c)\\
&= \mathrm{var}[X_k]. && (2.13d)
\end{aligned}
$$

The expression in (2.12) is also referred to as the Expected A Posteriori (EAP) estimator or the Minimum Variance (MV) estimator.

MAP estimation Another popular, and sometimes preferred [14, 15, 33, 91], technique is the maximum a posteriori probability (MAP) estimator

$$\hat{x}_k^{MAP} = \arg\sup_{x \in \mathcal{X}} p_{X_k}(x|z_{1:k}). \tag{2.14}$$

It is a Bayesian estimator that is not optimal in Bayes sense, i.e. it relies on the posterior distribution but it does not minimise expected value of any loss function. However, MAP estimator can be seen as an approximation to a Bayes estimator under conditions described in [7].

Remark 2.3.6. *It is perhaps worthwhile to highlight the close relation of MAP estimation to other point estimation techniques. Provided that that the posterior density is* $p_{X_k}(x|z_{1:k}) \propto g_k(z_k|x) p_{X_{k|k-1}}(x|z_{1:k-1})$*, it is easy to demonstrate its relation to the maximum likelihood estimator:*

$$\hat{x}_k^{MAP} = \arg\sup_{x\in\mathcal{X}} g_k(z_k|x) p_{X_{k|k-1}}(x|z_{1:k-1}), \tag{2.15}$$

$$\hat{x}_k^{ML} = \arg\sup_{x\in\mathcal{X}} g_k(z_k|x). \tag{2.16}$$

Note that maximum likelihood estimation is a non-Bayesian approach since it does not rely on a prior distribution $p_{X_k}(x|z_{1:k-1})$*. However, it can be interpreted as MAP estimation with a uniform prior.*

2.3.3 Other problems

In this subsection we offer three distinct problems that can be solved using information from the Bayes filter.

Definition 2.3.7 (Linear query)**.**
A linear query function $q_{\text{lin}} : \mathcal{X} \to \mathcal{X}$ *is given by*

$$q_{\text{lin}}(x) := Kx + C \tag{2.17}$$

for constant coefficients K *and* C *on* $\mathbb{R}$*.*

Example 2.3.8 (Predictive linear estimation)**.** *In the decision-theoretic settings, the problem can be defined by a composition of a linear query function* q_{lin} *in* (2.17) *and the squared error query loss function* l_2 *in* (A.1)*.*

For an object described with a random variable $X_{k+1|k}$ *with the predictive density* $p_{X_{k+1|k}}(x|z_{1:k})$*, the Bayesian solution to the problem of predictive linear estimation is a tuple* $(a_{\text{lin},k+1|k}, \rho_{\text{lin},k+1|k})$ *given by the Bayes action*

$$a_{\text{lin},k+1|k} = \arg\min_{a\in\mathbb{R}} \mathbb{E}[l_2(a, q_{\text{lin}}(X_{k+1|k}))] \tag{2.18a}$$

$$= \arg\min_{a\in\mathbb{R}} \int (a - q_{\text{lin}}(x))^2 p_{X_{k+1|k}}(x|z_{1:k}) \mathrm{d}x \tag{2.18b}$$

$$= K \int p_{X_{k+1|k}}(x|z_{1:k})\mathrm{d}x + C \tag{2.18c}$$

$$= K\mathbb{E}[X_{k+1|k}] + C \tag{2.18d}$$

and the associated Bayes expected loss

$$\rho_{\mathrm{lin},k+1|k} = \int (a_{\mathrm{lin},k+1|k} - q_{\mathrm{lin}}(x))^2 p_{X_{k+1|k}}(x|z_{1:k})\mathrm{d}x_k \tag{2.19a}$$

$$= K^2 \mathrm{var}[X_{k+1|k}]. \tag{2.19b}$$

When interest lies in queries defined on distinct target spaces, such as in the case of threat level assessment for an individual object presented in Example 2.3.10, the query function can be defined as follows.

Definition 2.3.9 (Object threat level query)**.**
A query function $q_\tau : \mathcal{X} \to \mathbb{R}$ *given by*

$$q_\tau(x) = \tau(x, x_A), \tag{2.20}$$

where function $\tau : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *returns a probability that detonation of a weapon in state* x *hits the asset in state* x_A *(such models are covered in Appendix A.2), evaluates to the threat level of object in state* x*.*

Example 2.3.10 (Threat estimation, inspired by [51])**.** *In the decision-theoretic settings, the problem of threat estimation can be defined by a composition of a query function* q_τ *returning an object's threat level* (2.20) *and the squared error query loss function* l_2 (A.1)*.*

For an object is described with a random variable X_k *with the posterior* $p_{X_k}(x|z_{1:k})$ *and the squared error query loss in* (A.1)*, the Bayesian solution to the problem of threat estimation is a tuple* $(a_{\tau,k}, \rho_{\tau,k})$ *given by the Bayes action*

$$a_{\tau,k} = \arg\min_{a\in\mathbb{R}} \mathbb{E}[l_2(a, q_\tau(X_k))] \tag{2.21a}$$

$$= \arg\min_{a\in\mathbb{R}} \int (a - q_\tau(x))^2 p_{X_k}(x|z_{1:k})\mathrm{d}x \tag{2.21b}$$

$$= \int \tau(x, x_A) p_{X_k}(x|z_{1:k})\mathrm{d}x \tag{2.21c}$$

$$= \mathbb{E}[\tau(X_k, x_A)] \tag{2.21d}$$

and the associated Bayes expected loss

$$\rho_{\tau,k} = \int (a_{\tau,k} - \tau(x, x_A))^2 p_{X_k}(x_k|z_{1:k})\mathrm{d}x_k \tag{2.22a}$$

$$= \mathrm{var}[\tau(X_k, x_A)]. \tag{2.22b}$$

Definition 2.3.11 (Binary classification)**.**
A query function $q_B : \mathcal{X} \to \{0, 1\}$ *given by*

$$q_B(x) = \mathbb{1}_B(x), \tag{2.23}$$

where $\mathbb{1}_B : \mathcal{X} \to \{0, 1\}$ *is an indicator function for an arbitrary region* $B \subset \mathcal{X}$*, evaluates whether an object in state* x *belongs to class/region* B*.*

Example 2.3.12 (Regional discrimination, inspired by [25, 59])**.** *In the decision-theoretic settings, the problem of regional discrimination can be specified by a composition of a query function* $q_B : \mathcal{X} \to \mathcal{A}$ *with* $\mathcal{A} = \{0, 1\}$ *from* (2.23)*, and a suitable query loss functions. Such loss function for* $\mathcal{A} = \{0, 1\}$ *is* $l_C : \mathcal{A} \times \mathcal{A} \to \{C_{00}, C_{01}, C_{10}, C_{11}\}$ *defined in* (A.2) *as*

$$l_C(a, \breve{a}) = \begin{cases} C_{00}, & a = 0, \breve{a} = 0, \\ C_{01}, & a = 0, \breve{a} = 1, \\ C_{10}, & a = 1, \breve{a} = 0, \\ C_{11}, & a = 1, \breve{a} = 1 \end{cases}$$

$$= C_{a0} + (C_{a1} - C_{a0})\mathbb{1}_1(\breve{a})$$

The overall loss associated with reporting an answer $a \in \{0, 1\}$ *when object is in the state* x *can be written as*

$$L_C(a, x) = l_C(a, q_B(x)) \tag{2.24a}$$

$$= C_{a0} + (C_{a1} - C_{a0})\mathbb{1}_1(q_B(x)) \tag{2.24b}$$

$$= C_{a0} + (C_{a1} - C_{a0})\mathbb{1}_B(x). \tag{2.24c}$$

For an object described with a random variable X_k *with the posterior* $p_{X_k}(x|z_{1:k})$*, the Bayesian solution to the problem of regional discrimination is a tuple* $(a_{B,k}, \rho_{B,k})$

given by the Bayes action

$$a_{B,k} = \arg\min_{a\in\{0,1\}} \mathbb{E}[L(a, X_k)] \tag{2.25a}$$
$$= \arg\min_{a\in\{0,1\}} \int \Big[C_{a0} + (C_{a1} - C_{a0})\mathbb{1}_B(x)\Big] p_{X_k}(x|z_{1:k})\mathrm{d}x \tag{2.25b}$$
$$= \arg\min_{a\in\{0,1\}} \Big[C_{a0} + (C_{a1} - C_{a0}) \int_B p_{X_k}(x|z_{1:k})\mathrm{d}x\Big], \tag{2.25c}$$

and the associated Bayes expected loss

$$\rho_{B,k} = \mathbb{E}[L_C(\bar{a}, X_k)] \tag{2.26}$$
$$= C_{\bar{a}0} + (C_{\bar{a}1} - C_{\bar{a}0}) \int_B p_{X_k}(x|z_{1:k})\mathrm{d}x, \tag{2.27}$$

where $\bar{a} := a_{B,k}$ *refers to the optimal action and different notation is used for convenience.*

Furthermore, if the costs are set to penalise the errors exclusively, $C_{01} = C_{10} = 1$ *and* $C_{00} = C_{11} = 0$*, we arrive at a minimum probability of error detector defined by the following expressions*

$$\mathbb{E}[L_C(0, X_k)] = \int_B p_{X_k}(x|z_{1:k})\mathrm{d}x, \tag{2.28}$$
$$\mathbb{E}[L_C(1, X_k)] = 1 - \int_B p_{X_k}(x|z_{1:k})\mathrm{d}x. \tag{2.29}$$

The first expression simply computes the probability of the object being in region B*, whereas the second is the probability that the object is not in the region. Accordingly, if the probability of object being inside the region is smaller, the answer* 0 *must be reported, and otherwise it is* 1.

2.4 Summary

In this chapter we introduced Bayesian filtering for a single surely present object. Uncertainty in the state of such dynamic system is modelled by a random vector. We considered the new formulation of the loss function, which is given in Definition 1.5.1, in the single-object filtering context. We were able to formulate a general class of statistical problems which require filtering information to be solved, and managed to produce a number of analytic solutions for instances such as threat estimation or binary discrimination. To the best of our knowledge, these problems have not been appropriately treated in decision theoretic framework yet.

Chapter 3

Recursive filtering for an object population

Since the development of the Kalman filter in 1960 [49], the possibility of constructing optimal algorithms has, arguably, been reserved for systems that could be faithfully modelled by a vector-valued random variable [24]. Although the theory of point processes, or point fields, emerged around the same time [75], its definite application for modelling dynamic systems has not been realised until recently [70]. This development has brought a paradigm shift of nature modelling from a single surely present dynamic object to a time-varying population of such objects.

The new paradigm posed a wealth of new challenges to the theory of recursive filtering. In particular, it brought into question the applicability of optimal point estimators, such as an MMSE algorithm. As a result, when the need comes to solve the problem of point estimation of the system state, literature often has little to offer but heuristics for extraction of a *multi-object state estimates.*

This chapter culminates with a presentation of a number of filters for spatial point processes. Eventually, the focus will be on approximate filters, including the classical Probability Hypothesis Density (PHD) filter [68] as well as its developments, such as the Cardinalized Probability Hypothesis Density (CPHD) filter [67] and Panjer Probability Hypothesis Density (PPHD) filter [95].[1]

The content of this chapter is as follows. In Section 3.1 we discuss the concept of optimality in the context of filters for object populations. Section 3.2 presents probabilistic methods for modelling uncertain object populations. In Section 3.3 we introduce the Bayes filter for an indistinguishable object population. In Section 3.5

[1] In this thesis the abbreviation PPHD is not to be interpreted as the Particle PHD filter, as found in [21, 22, 110], which is another name for the Sequential Monte-Carlo (SMC) implementation of the classical PHD filter, i.e. the SMC-PHD filter.

we discuss decision making with the Bayes filter. Section 3.4 presents approximate PHD filters and their current exploitation for decision making. Section 3.6 provides the summary.

3.1 Optimal filtering

The Bayes filter, as it is introduced in Chapter 2, is fundamentally based on the idea that the uncertain state of a dynamic system, as well as its partial observations, can be satisfactorily modelled by random vectors. This mathematical abstraction is well suited for describing a *single object* (phenomenon) of interest in the absence of clutter, i.e. possible unwanted objects. Some authors find it necessary to stress that this phenomenon is 'always-on' ('surely present' [74] or 'permanently active or present' [86]), as it is not experiencing birth of death. To this we wish to add that this phenomenon is 'cooperative' in the sense that it never fails to return a measurement.

This abstraction is well suited for many practical problems. Due to its natural fit, it has revolutionised the area of control systems (think of chemical plants or automotive engines) [4, 5] and facilitated navigation techniques culminating with the Apollo's lunar mission [97]. However, when it comes to applying it for area surveillance, which is focused on by time-varying number of 'non-cooperative' objects, its direct application is not possible.

For a long time, a common philosophy has been to build a multi-object filtering solutions in the 'bottom-up' manner, essentially as a combination of separate single-object filters. And it is only recently that 'top-down' filters became widespread, thanks to the new abstract framework, originating from the engineering interpretation of point process theory [75, 104] called Finite Set Statistics (FISST) [70], that replaced *vector-valued* random variables by what is called *finite-set-valued* (or, simply, *set-valued* [17]) random variables.

3.1.1 Bottom-up approaches

As mentioned before, traditional algorithms designed for object populations addressed recursive filtering as a combination of separate single-object filters, with their outputs integrated within an additional algorithm which returns a global output. Overall, these algorithms are designed such as to offer the ability to distinguish the objects and to naturally describe each of them. Such algorithms are reviewed in [82, 110].

Recursive filtering could include simple one-to-one association of measurements to filters (in case of the Nearest-Neighbour (NN) algorithm, see [102, 105]), which cannot cope well with data ambiguities. Alternatively, it is executed by maintaining all hypotheses of possible combinations (in case of the Multiple Hypothesis Tracking (MHT) filter, see [84]) or by selecting the most likely of those combinations (in case of the Joint Probabilistic Data Association (JPDA) filter, see [31]). These algorithms can become computationally involved when the number of objects is high or if there are many ambiguities in the data.

Despite these algorithms have been successfully implemented for real-life surveillance, their theoretical consistency is not as clear as that of an isolated single-object filter. Accordingly, the question is whether these algorithms are consistent with the Bayesian paradigm, and if so, whether they are optimal, i.e. they produce the exact posterior distribution of the object population. In particular, this concern has been raised in [112, p. 8-9], [70, p. 340-341]. As a response to this, there is an ongoing effort to recover the modelling assumptions under which the algorithms (or their variations) would be provably optimal, see e.g. [17, 117].

3.1.2 Top-down approaches

An alternative approach to recursive filtering views the population of objects as a single entity, or meta-object, and the set of measurements as a single meta-measurement, or observation. This constitutes the 'top-down' approach. Accordingly, the uncertain dynamic system (and its observations) is no longer described by a random vector but by a random finite set, or point process. This approach is introduced in terms of point process theory in [19], and in the context of FISST in [110, 111].

This new abstraction leads to the construction of the optimal Bayes filter analogous to the single-object filter, but will require methods from point process theory described in Section 3.2. This algorithm is very general and complex, so to be tractable, it requires successive approximations and simplifications. There are at least two strategies to approximate it: parametrized density approximation algorithms (for the case of Bernoulli and multi-Bernoulli filters, see [88]), and moment-approximation algorithms (see Section 3.4). Specifically, moment-approximation algorithms avoid the combinatorial problem that arises from data association, as they consider objects in a population as indistinguishable.

3.2 Probabilistic methods for representing object populations

3.2.1 Spatial point processes

In this work, the objects of interest have individual states x in some d-dimensional state space $\mathcal{X} \subset \mathbb{R}^d$, typically consisting of position, velocity and class variables. A point process Φ on $\mathcal{X}$ is a random variable on the process space $\mathfrak{X} = \bigcup_{n=0}^{\infty} \mathcal{X}^n$, i.e. the space of all finite sequences of points in $\mathcal{X}$, whose number of elements *and* element states are unknown and (possibly) time-varying. A realisation of Φ is a sequence[2] $\varphi = (x_{1:n}) \in \mathcal{X}^n$, representing a *population* of n objects with states $x_i \in \mathcal{X}$. In the context of multi-object filtering, this sequence depicts a specific multi-object configuration.

More formally, a point process Φ on $\mathcal{X}$ is a measurable mapping

$$\Phi : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathfrak{X}, \mathcal{B}(\mathfrak{X})) \tag{3.1}$$

from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, where Ω is a sample space; $\mathcal{F}$ is a σ-algebra on Ω; $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$; $\mathcal{B}(\mathfrak{X})$ is the Borel σ-algebra on $\mathfrak{X}$ [104].

Definition 3.2.1 (Spatial point process)**.** *A point process Φ on $\mathcal{X}$ is a random variable on the space $\mathfrak{X}$ of finite sequences in $\mathcal{X}$. A realisation of Φ is a sequence $\varphi = (x_1, \cdots, x_n) \in \mathcal{X}^n$, describing a group of n objects with states $x_i \in \mathcal{X}$ where both n and all x_i are random.*

As for usual real-valued random variables, a point process is described by its probability distribution P_Φ on $\mathfrak{X}$. The probability distribution is always defined as a symmetric function, so that the order of points in a realisation is irrelevant for statistical purposes and the permutations of φ—such as (x_1, x_2) and (x_2, x_1)—are equally probable. In addition, a point process is called *simple* if the probability distribution is such that realisations are sequences of points that are pairwise distinct almost surely, i.e. φ does not contain repetitions. For the rest of the thesis, all point processes are assumed to be simple. In that case, it admits a Radon-Nikodym derivative (w.r.t. the reference measure) denoted as p_Φ, which is a probability density of a point process Φ.

[2]Here $(x_{1:n})$ denotes the sequence $(x_1, \ldots, x_n)$.

3.2.2 Statistical moments and p.g.fl.

In this subsection we are going to present various quantities that are used to communicate information about a point process.

Statistical moments It rarely happens such that the complete knowledge about the point process Φ contained in a probability density p_Φ is available. What may be available is certain limited descriptions of Φ contained in its statistical moments. It is possible to obtain factorial and non-factorial moment densities for any order, but their construction is rather involved as compared to regular random variables. In this thesis, we are going to focus only on the lower-order moments that contain the most information. Specifically, we are going to focus on two quantities: the first-order non-factorial moment density, and the second-order factorial moment density.[3]

Definition 3.2.2 (First-order moment density). *For a point process Φ, the first-order statistical moment density, also called intensity function, or simply intensity, is defined as*

$$\mu_\Phi(x) := \int \left(\sum_{x_i \in \varphi} \delta_x(x_i) \right) p_\Phi(\varphi) \mathrm{d}\varphi \tag{3.2a}$$

$$= \sum_{n \geq 1} \int_{\mathcal{X}^n} \left(\sum_{1 \leq i \leq n} \delta_x(x_i) \right) p_\Phi^{(n)}(x_{1:n}) \mathrm{d}(x_{1:n}), \tag{3.2b}$$

where the following holds $\int \delta_y(x) f(x) \mathrm{d}x = f(y)$.

Definition 3.2.3 (Second-order factorial moment density). *For a point process Φ, the second-order factorial statistical moment density is defined as*

$$\nu_\Phi(x, \bar{x}) := \int \left(\sum_{x_i, x_j \in \varphi}^{i \neq j} \delta_x(x_i) \delta_{\bar{x}}(x_j) \right) p_\Phi(\varphi) \mathrm{d}\varphi \tag{3.3a}$$

$$= \sum_{n \geq 0} \int_{\mathcal{X}^n} \left(\sum_{1 \leq i,j \leq n}^{i \neq j} \delta_x(x_i) \delta_{\bar{x}}(x_j) \right) p_\Phi^{(n)}(x_{1:n}) \mathrm{d}(x_{1:n}), \tag{3.3b}$$

where the following holds $\int \delta_y(x) f(x) \mathrm{d}x = f(y)$.

Probability Generating Functional (p.g.fl.) One of the ways to study point processes is by expressing them using different presentation, such as through the

[3] Definitions 3.2.2 and 3.2.3 exhaust the variety of densities available for the moments of first two order. The reason is that for the first-order moments the densities coincide, whereas the density of the second non-factorial moment is not defined as it violates the assumption of process simplicity.

Probability Generating Functional (p.g.fl.). For example, it is possible to obtain expressions for various statistical moments for a point process by differentiating its p.g.fl. P.g.fl.s play a role similar to that of the Fourier transformation for signal processing and the probability generating functions for discrete random variables.

Definition 3.2.4 (Probability generating functional)**.** *For a point process Φ, the probability generating functional (p.g.fl.) is defined as an expectation*

$$\mathcal{G}_\Phi(h) := \mathbb{E}\Big[\prod_{x\in\Phi} h(x)\Big] \tag{3.4}$$

$$= \int \Big(\prod_{x\in\varphi} h(x)\Big) p_\Phi(\varphi)\mathrm{d}\varphi \tag{3.5}$$

$$= \sum_{n\geq 0}\int_{\mathcal{X}^n}\Big[\prod_{1\leq i\leq n} h(x_i)\Big] p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}), \tag{3.6}$$

where $h : \mathcal{X} \to [0,1]$ a test function.

3.2.3 Some elementary spatial point processes

Three different point processes are discussed below to illustrate the formulation of point processes on specific examples. All of them will be used later on to model different phenomena, resulting in different population filters. Examples of point processes that have been used in the context of multi-object filtering are given below.

Poisson point process is the most important of the considered point processes. Because of its convenient mathematical properties it has been often used to create simple models of natural and man-made phenomena. Two other considered point processes offer a possibility to build more realistic, higher fidelity models.

Definition 3.2.5 (Generalised factorial and binomial coefficient)**.**
Consider a real number $x \in \mathbb{R}$ and a non-negative integer $n \in \mathbb{N}$.

(a) The Pochhammer symbol *or rising factorial $x_{n\uparrow}$ is given by*

$$(x)_{n\uparrow} := \prod_{0\leq i\leq n-1} (x+i), \quad x_{0\uparrow} := 1. \tag{3.7}$$

(b) In the same manner, the falling factorial $x_{n\downarrow}$ is given by

$$(x)_{n\downarrow} := \prod_{0\leq i\leq n-1} (x-i), \quad x_{0\downarrow} := 1. \tag{3.8}$$

(c) Using (3.8), *the* generalised coefficient $\binom{x}{n}$ *is defined as*

$$\binom{x}{n} = \frac{x_{n\downarrow}}{n!} \tag{3.9}$$

Definition 3.2.6 (i.i.d. cluster process)**.**
An independent and identically distributed (i.i.d.) cluster process with cardinality distribution ρ on $\mathbb{N}$ and spatial distribution c on $\mathcal{X}$ describes a population of objects whose size is described by ρ, and whose states are i.i.d. according to c.

Its p.g.fl. is of the form

$$\mathcal{G}_{\text{i.i.d.}} = \sum_{n\geq 0} \rho(n) \left[\int_{\mathcal{X}} h(x)c(x)\mathrm{d}x \right]. \tag{3.10}$$

Lower-order statistics $(\mu_{\text{i.i.d.}}, \nu_{\text{i.i.d.}})$ *of this process are given by*

$$\mu_{\text{i.i.d.}}(x) = \left(\sum_{n\geq 0} n\rho(n) \right) c(x), \tag{3.11}$$

$$\nu_{\text{i.i.d.}}(x, \bar{x}) = \left(\sum_{n\geq 0} (n^2 - n)\rho(n) \right) c(x)c(\bar{x}). \tag{3.12}$$

Definition 3.2.7 (Panjer process [95])**.**
A Panjer process with parameters $\alpha \in \mathbb{R}_{>0}$ and $\beta \in \mathbb{R}_{>0}$ (or $\alpha \in \mathbb{Z}_{<0}$ and $\beta \in \mathbb{R}_{<0}$) and spatial distribution c is an i.i.d. cluster process with spatial distribution c, whose size is Panjer distributed with parameters α and β:

$$\rho_{\text{Panjer}}(n) = \binom{-\alpha}{n} \left(1 + \frac{1}{\beta}\right)^{-\alpha} \left(\frac{-1}{\beta + 1}\right)^{n}. \tag{3.13}$$

Its p.g.fl. is of the form

$$\mathcal{G}_{\text{Panjer}} = \left(1 + \frac{1}{\beta} \int_{\mathcal{X}} h(x)c(x)\mathrm{d}x \right)^{-\alpha}. \tag{3.14}$$

Lower-order statistics $(\mu_{\text{Panjer}}, \nu_{\text{Panjer}})$ *of this process are given by*

$$\mu_{\text{Panjer}}(x) = \alpha\beta^{-1}c(x), \tag{3.15}$$

$$\nu_{\text{Panjer}}(x, \bar{x}) = (\alpha)_2\beta^{-2}c(x)c(\bar{x}), \tag{3.16}$$

where $(\xi)_n$ *is a rising factorial symbol for any* $\xi \in \mathbb{R}$ *and* $n \in \mathbb{N}$*.*

Panjer point process can be completely described by its intensity function $\mu_\natural$ and the variance in object number $\mathrm{var}_{\mathrm{Panjer}}(\mathcal{X})$ [95]. At the same time, knowing the density ν_{Panjer} on top of that would introduce no additional information.

Definition 3.2.8 (Poisson process)**.**
A Poisson process with parameter $\lambda \geq 0$ *and spatial distribution* s *is an i.i.d. cluster process with spatial distribution* c*, whose size is Poisson distributed with rate* λ*:*

$$\rho_{\mathrm{Poisson}}(n) = e^{-\lambda}\frac{\lambda^n}{n!}. \tag{3.17}$$

Its p.g.fl. is of the form

$$\mathcal{G}_{\mathrm{Poisson}} = \exp\left(\int_{\mathcal{X}} [1 - h(x)]c(x)\mathrm{d}x\right). \tag{3.18}$$

Lower-order statistics $(\mu_{\mathrm{Poisson}}, \nu_{\mathrm{Poisson}})$ *of this process are given by*

$$\mu_{\mathrm{Poisson}}(x) = \lambda c(x), \tag{3.19}$$

$$\nu_{\mathrm{Poisson}}(x, \bar{x}) = \lambda^2 c(x)c(\bar{x}). \tag{3.20}$$

Note that Poisson process can be sufficiently described by its intensity function μ_{Poisson}. At the same time, the second-order moment density ν_{Poisson} carries no additional information.

3.3 Bayes filter for an object population

The Bayes filter for spatial point processes is a natural extension of the Bayes filter for a single surely present object in Subsection 2.2 to the case of object populations. The resulting Bayes recursion at time step k consists of the time prediction and data update steps given by:

$$p_{\Phi_{k|k-1}}(\varphi|Z_{1:k-1}) = \int T_{k|k-1}(\varphi|\bar{\varphi})p_{\Phi_{k-1}}(\bar{\varphi}|Z_{1:k-1})\mathrm{d}\bar{\varphi}, \tag{3.21}$$

$$p_{\Phi_k}(\varphi|Z_{1:k}) = \frac{L_k(Z_k|\varphi)p_{\Phi_{k|k-1}}(\varphi|Z_{1:k-1})}{\int L_k(Z_k|\bar{\varphi})p_{\Phi_{k|k-1}}(\bar{\varphi}|Z_{1:k-1})\mathrm{d}\bar{\varphi}}, \tag{3.22}$$

where $p_{\Phi_k}(\varphi|Z_{1:k-1})$ and $p_{\Phi_k}(\varphi|Z_{1:k})$ are probability densities, respectively, of the predicted and updated object process Φ at step k, which are conditioned on the set of available multi-object observations; Z_k is the set of measurements collected at time k; $T_{k|k-1}(\varphi|\bar{\varphi})$ is the multi-object transition kernel, describing the time evolution of the population of objects and encapsulates the underlying models of object birth, motion, spawning and death; and $L_k(Z_k|\varphi)$ is the population likelihood function, describing the sensor observation process and encapsulates the underlying models of object detection, object-originating measurements, and false alarms.

Equivalent expressions of the filter can be established through generating functionals. Through this perspective, the p.g.fl.s of the predicted and updated processes are [23, 25]:

$$\mathcal{G}_{\Phi_{k|k-1}}[h|Z_{1:k-1}] = \int\int\left[\prod_{x_i\in\varphi} h(x_i)\right] T_{k|k-1}(\varphi|\bar{\varphi}) p_{\Phi_{k-1}}(\bar{\varphi}|Z_{1:k-1})\mathrm{d}\varphi\mathrm{d}\bar{\varphi}, \tag{3.23}$$

$$\mathcal{G}_{\Phi_k}[h|Z_k] = \frac{\displaystyle\int\left[\prod_{x_i\in\varphi} h(x_i)\right] L_k(Z_k|\varphi) p_{\Phi_{k|k-1}}(\varphi)\mathrm{d}\varphi}{\displaystyle\int\left[\prod_{x_i\in\varphi} h(x_i)\right] L_k(Z_k|\varphi) p_{\Phi_{k|k-1}}(\varphi)\mathrm{d}\varphi\bigg|_{h=1}}. \tag{3.24}$$

3.4 Moment-approximation filters

The Bayes filter for an object population as presented in Section 3.3 is commonly regarded as intractable in practice because of the combinatorial nature of probability densities as well as complications associated with the curse of dimensionality [70]. In order to overcome this recognised complication, a number of principled approximations have been suggested that are based on the idea of propagating a limited number of statistical moments (such as those in Section 3.2.2). The Probability Hypothesis Density (PHD) [68], the Panjer PHD (PPHD) [95] and Cardinalized PHD (CPHD) [67] filters are such approximations that are based on first moments and some additional information about cardinality distributions. This section presents the filtering recursions largely following the overview in [96], with notations originally inspired by [25, 95, 109].

3.4.1 Useful notations

Let us define some specific terms which will be common to all filters described below, up to the specific prior and predicted intensities $\mu^{\bullet}_{k-1}$ and $\mu^{\bullet}_{k|k-1}$ with $\bullet \in \{\flat, \sharp, \natural\}$ indicating the relevant filter as defined later in text. The survival term is defined as

$$\mu_{s,k}(x) = p_{s,k}(\bar{x})\pi_{k|k-1}(x|\bar{x})\mu^{\bullet}_{k-1}(\bar{x}). \quad (3.25)$$

The missed detection term and association terms for any $z \in Z_k$ are defined, respectively, as

$$\mu^{\phi}_{k}(x) = (1 - p_{d,k}(x))\mu^{\bullet}_{k|k-1}(x), \quad (3.26)$$

$$\mu^{z}_{k}(x) = g_k(x|z)p_{d,k}(x)\mu^{\bullet}_{k|k-1}(x), \quad (3.27)$$

Furthermore, the following form of notation for intensity integrals is extensively used in the presentation:

$$\mu(\mathcal{X}) = \int_{\mathcal{X}} \mu(x)\mathrm{d}x. \quad (3.28)$$

3.4.2 The PHD filter

Among the filtering algorithms presented in this section, the PHD filter was introduced first and is the most popular filter. It was developed by Mahler in [68] under the assumption that the number of predicted objects, as well as the cardinality of false alarms, is Poisson distributed. We are going to refer to this filter using the superscript $\flat$.

Proposition 3.4.1 (PHD recursion [95])**.**

(a) The predicted first-order moment density is given by

$$\mu^{\flat}_{k|k-1}(x) = \mu_{b,k}(x) + \mu_{s,k}(x), \quad (3.29)$$

with survival intensity (3.25)*, where* $\bullet = \flat$*.*

(b) The updated first-order moment density with Poisson distributed prediction and false alarm models is obtained as

$$\mu^{\flat}_{k}(x) = \mu^{\phi}_{k}(x) + \sum_{z \in Z_k} \frac{\mu^{z}_{k}(x)}{\mu_{fa,k}(z) + \mu^{z}_{k}(\mathcal{X})} \quad (3.30)$$

with missed detection term (3.26) *and association term* (3.27)*, where* $\bullet = \flat$*, and* $\mu_{fa,k}$ *is the intensity of false alarms.*

In principle, the filtering recursion in the PHD filter cannot be computed exactly. However, tractable implementations include Gaussian Mixture (GM) [107] and Sequential Monte Carlo (SMC) [108] based algorithms. The GM implementation assumes that intensity is a Gaussian Mixture, and requires that every object and associated measurement follow a linear and Gaussian model. The SMC implementation approximates the intensity function by a set of weighted particles and does not require any assumptions regarding the dynamics of the objects.

3.4.3 The Panjer PHD filter

The Panjer PHD (also called the Second-order PHD filter) was introduced in [95] as a development of the PHD filter that includes additional second-order information, specifically by propagating not only the mean number of objects, but also variance in the mean number of objects. It is based on the assumption that the number of predicted objects, as well as the cardinality of false alarms, is Panjer distributed. The Panjer point process generalizes the Poisson point process, as demonstrated in [95], and, therefore, it is less restrictive. Furthermore, it avoids the computationally expensive propagation of the complete cardinality distribution, as it is done in the CPHD filter. The Panjer distribution is sufficiently characterized by two parameters, and they stand in direct correspondence with the distribution's mean and variance. As a consequence, it is possible to propagate both the mean and variance of the cardinality distribution in a filtering recursion. We are going to refer to this filter using the superscript $\natural$.

Let $\alpha_{k|k-1}$, $\beta_{k|k-1}$ and $\alpha_{\mathrm{fa},k}$, $\beta_{\mathrm{fa},k}$ be, respectively, the parameters of the predicted object and false alarm processes at time k. Define the terms

$$Y_u(Z) := \sum_{j=0}^{|Z|} \frac{(\alpha_{k|k-1})_{j+u}}{(\beta_{k|k-1})^{j+u}} \frac{(\alpha_{\mathrm{fa},k})_{|Z|-j}}{(\beta_{\mathrm{fa},k}+1)^{|Z|-j}} F_{\mathrm{d}}^{-j-u} e_j(Z) \tag{3.31}$$

for any $Z \subseteq Z_k$, where F_{d} is the scalar given by

$$F_{\mathrm{d}} := \int \left[1 + \frac{p_{\mathrm{d},k}(x)}{\beta_{k|k-1}}\right] \mu^{\natural}_{k|k-1}(x)\mathrm{d}x, \tag{3.32}$$

and e_j is the j-th elementary symmetric function

$$e_j(Z) := \sum_{\substack{Z' \subseteq Z \\ |Z'|=j}} \prod_{z \in Z'} \frac{\mu_k^z(\mathcal{X})}{c_{\mathrm{fa},k}(z)}, \tag{3.33}$$

where $c_{\mathrm{fa},k}$ denotes the spatial density of false alarms at time k, and

$$\mu_k^z(\mathcal{X}) = \int \mu_k^z(x)\mathrm{d}x, \tag{3.34}$$

where the association term (3.27) is defined for $\bullet = \natural$. Furthermore, define terms for $u = \{1, 2\}$ via

$$l_u^\natural(\phi) := \frac{Y_u(Z_k)}{Y_0(Z_k)} \quad \text{and} \quad l_u^\natural(z) := \frac{Y_u(Z_k \setminus \{z\})}{Y_0(Z_k)}. \tag{3.35}$$

Variance prediction is constructed based on the second-order factorial moment ν_k which generally cannot be obtained when the predicted information is limited to $\mu^\flat_{k|k-1}$ and $\mathrm{var}^\flat_{k|k-1}$ only. However, the assumption that $p_{s,k}(x) = p_{\mathrm{s},k}$ is uniform for all x over the state space $\mathcal{X}$ leads to the following recursion.

Proposition 3.4.2 (Panjer PHD recursion [95])**.**

(a) Under assumption that $p_{\mathrm{s},k}(x) = p_{\mathrm{s},k}$ *is constant for any* $x \in \mathcal{X}$ *at time* k*. In the manner of* (3.29) *and* (3.49)*, the predicted first-order moment density of the Panjer PHD filter is given by*

$$\mu^\natural_{k|k-1}(x) = \mu_{b,k}(x) + \mu_{\mathrm{s},k}(x), \tag{3.36}$$

with survival intensity (3.25)*, where* $\bullet = \natural$*. The predicted variance in the whole state space* $\mathcal{X}$ *is given by*

$$\mathrm{var}^\natural_{k|k-1}(\mathcal{X}) = \mathrm{var}_{\mathrm{b},k}(\mathcal{X}) + \mathrm{var}_{\mathrm{s},k}(\mathcal{X}), \tag{3.37}$$

where $\mathrm{var}_{\mathrm{b},k}$ *is the variance of the birth process and* $\mathrm{var}_{\mathrm{s},k}$ *is the variance of the predicted process describing the surviving objects which is given by*

$$\mathrm{var}_{\mathrm{s},k}(\mathcal{X}) = p^2_{\mathrm{s},k}\mathrm{var}_{k-1}(\mathcal{X}) + p_{\mathrm{s},k}[1 - p_{\mathrm{s},k}]\mu_{k-1}(\mathcal{X}). \tag{3.38}$$

(b) Obtain terms $\alpha_{k|k-1}$ *and* $\beta_{k|k-1}$ *using*

$$\alpha_{k|k-1} = \frac{\mu^\natural_{k|k-1}(\mathcal{X})^2}{\mathrm{var}^\natural_{k|k-1}(\mathcal{X}) - \mu^\natural_{k|k-1}(\mathcal{X})}, \tag{3.39}$$

$$\beta_{k|k-1} = \frac{\mu^\natural_{k|k-1}(\mathcal{X})}{\mathrm{var}^\natural_{k|k-1}(\mathcal{X}) - \mu^\natural_{k|k-1}(\mathcal{X})}. \tag{3.40}$$

Then, the updated first-order moment density with Panjer distributed prediction and false alarm models is obtained with

$$\mu_k^{\natural}(x) = \mu_k^{\phi}(x) l_1^{\natural}(\phi) + \sum_{z \in Z_k} \frac{\mu_k^z(x)}{c_{\mathrm{fa},k}(z)} l_1^{\natural}(z), \tag{3.41}$$

with missed detection term (3.26) *and association term* (3.27)*, where* $\bullet = \natural$*. The updated variance is given by*

$$\begin{aligned}\mathrm{var}_k^{\natural}(\mathcal{X}) =& \mu_k^{\natural}(\mathcal{X}) + \mu_k^{\phi}(\mathcal{X})[l_2^{\natural}(\phi) - l_1^{\natural}(\phi)^2] \\ &+ 2\mu_k^{\phi}(\mathcal{X}) \sum_{z \in Z_k} \frac{\mu_k^z(\mathcal{X})}{c_{\mathrm{fa},k}(z)} [l_2^{\natural}(z) - l_1^{\natural}(\phi) l_1^{\natural}(z)] \\ &+ \sum_{z, \bar{z} \in Z_k} \frac{\mu_k^z(\mathcal{X})}{c_{\mathrm{fa},k}(z)} \frac{\mu_k^{\bar{z}}(\mathcal{X})}{c_{\mathrm{fa},k}(\bar{z})} [l_2^{\natural,\neq}(z, \bar{z}) - l_1^{\natural}(z) l_1^{\natural}(\bar{z})],\end{aligned} \tag{3.42}$$

such that $l_2^{\natural,\neq}(z, \bar{z})$ *is obtained as* $l_2^{\natural}(z, \bar{z})$ *if* $z \neq \bar{z}$ *and zero otherwise.*

Closed-form Gaussian Mixture implementation of the Panjer PHD filter is proposed in [95].

3.4.4 The Cardinalized PHD filter

After the need for a filter that propagates higher-order information about the number of objects was expressed in [29], Mahler introduced the CPHD filter in [67]. In place of taking any particular assumption on the nature of the cardinality distribution, this algorithm estimates the distribution together with the intensity of the point process. We are going to refer to this filter using the superscript $\sharp$.

In the following ρ_k denotes the cardinality distribution of the object population, and ρ_{b} and ρ_{fa} denote, respectively, the birth and false alarm cardinality distributions. Similarly to the PPHD filter, the CPHD filter update has additional terms l_u which depend on the cardinality distribution. We denote the discrete and continuous inner product by

$$\langle f, g \rangle = \int f(x) g(x) dx \qquad \text{(continuous case)}, \tag{3.43}$$

$$\langle f, g \rangle = \sum_{n \geq 0} f(n) g(n) \qquad \text{(discrete case)}. \tag{3.44}$$

Using the notations of [109], we define the terms $\Upsilon^d[\mu, Z]$ with

$$\Upsilon^d[\mu, Z](n) = \sum_{j=0}^{\min(|Z|, n-u)} \frac{n!(|Z|-j)!}{(n-(j+d))!} \rho_{\text{fa}}(|Z|-j) \frac{\mu_k^\phi(\mathcal{X})^{n-(j+d)}}{\mu_{k|k-1}^\sharp(\mathcal{X})^n} e_j(Z), \tag{3.45}$$

where $e_j(Z)$ denotes the elementary symmetric functions defined by (cf. Equation (3.33))

$$e_j(Z) := \sum_{\substack{Z' \subseteq Z \\ |Z'|=j}} \prod_{z \in Z'} \frac{\mu_k^z(\mathcal{X}) \mathrm{d}x}{c_{\text{fa},k}(z)}, \tag{3.46}$$

where $c_{\text{fa},k}$ denotes the spatial density of false alarms at time k, and

$$\mu_k^z(\mathcal{X}) = \int \mu_k^z(x) \mathrm{d}x, \tag{3.47}$$

where the association term (3.27) is defined for $\bullet = \sharp$. This leads to the terms

$$l_1^\sharp(\phi) = \frac{\langle \Upsilon^1[\mu, Z], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle} \quad \text{and} \quad l_1^\sharp(z) = \frac{\langle \Upsilon^1[\mu, Z \setminus \{z\}], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle}. \tag{3.48}$$

As mentioned above, the CPHD filter propagates both the intensity function as well as the cardinality distribution of the object process.

Proposition 3.4.3 (CPHD recursion [67])**.**

(a) In the manner of (3.29)*, the predicted first-order moment density is given by*

$$\mu_{k|k-1}^\sharp(x) = \mu_{\text{b},k}(x) + \mu_{\text{s},k}(x), \tag{3.49}$$

with survival intensity (3.25)*, where* $\bullet = \sharp$*. The predicted object cardinality distribution is obtained with*

$$\rho_{k|k-1}(n) = \sum_{j=1}^{n} \rho_{\text{b}}(n-j) S[\mu_{k-1}^\sharp, \rho_{k-1}](j) \tag{3.50}$$

for any $n \in \mathbb{N}$ *with*

$$S[\mu, \rho](j) = \sum_{i=j}^{\infty} \binom{i}{j} \frac{\langle p_{\text{s},k}, \mu \rangle^j \langle (1-p_{\text{s},k}), \mu \rangle^{i-j}}{\langle 1, \mu \rangle^i} \rho(i). \tag{3.51}$$

(b) The updated first-order moment density with i.i.d. cluster distributed prediction

and false alarm models is obtained as

$$\mu_k^\sharp(x) = \mu_k^\phi(x) l_1^\sharp(\phi) + \sum_{z \in Z_k} \frac{\mu_k^z(x)}{c_{\mathrm{fa},k}(z)} l_1^\sharp(z), \tag{3.52}$$

with missed detection term (3.26) *and association term* (3.27)*, where* $\bullet = \sharp$*. The updated object cardinality distribution for any* $n \in \mathbb{N}$ *is given by*

$$\rho_k(n) = \frac{\Upsilon^0[\mu_{k|k-1}^\sharp, Z](n)\rho_{k|k-1}(n)}{\langle \Upsilon^0[\mu_{k|k-1}^\sharp, Z], \rho_{k|k-1}\rangle}. \tag{3.53}$$

Tractable implementations of the CPHD filter with SMC methods follow as straightforward extensions [108] as well as through GM closed form solutions in [109].

3.5 Decision making under uncertainty

In this section we are going to consider exploitation of information from Bayesian filters for an object population in decision making.

As far as decision making with object populations is concerned, literature has predominantly focused on the problem of point estimation of the system state. In the scope of this section we are going to consider the problem of producing a point estimate, commonly referred to as a *multi-object state estimate*.[4]

To the best of our knowledge, the filtering information has not been considered for making any other decisions consistent with the problem formulation offered by this thesis. However, some results concerning computation of regional statistics [25, 95] can be *interpreted* in this manner using the formulation for the loss function proposed by this thesis.

3.5.1 Revised decision procedure

Proposition 3.5.1 (Posterior Bayes-optimal decision)**.** *For an uncertain system described by a point process* Φ_k *on* $\mathcal{X}$*, a Bayes-optimal solution to a decision problem is given by a pair* (a_k, ρ_k)*, where Bayes action* a_k *is given by*

$$a_k = \arg\min_{a \in \mathcal{A}} \mathbb{E}[L(a, \Phi_k)] \tag{3.54}$$

[4] Engineering approaches that are commonly used to extract a system state in practical filter implementations will be discussed in Subsection 3.5.4.

$$= \arg\min_{a \in \mathcal{A}} \int l(a, q(\varphi)) p_{\Phi_k}(\varphi | Z_{1:k}) \mathrm{d}\varphi, \tag{3.55}$$

and Bayes expected loss ρ_k is given by

$$\rho_k = \mathbb{E}[L(a_k, \Phi_k)] \tag{3.56}$$

$$= \int l(a_k, q(x)) p_{\Phi_k}(\varphi | Z_{1:k}) \mathrm{d}\varphi. \tag{3.57}$$

3.5.2 Point estimation of the system state

Proposition 3.5.2 (Posterior Bayes-optimal state estimate)**.** *For an uncertain system described by a point process Φ_k on $\mathcal{X}$, a Bayes-optimal solution to the problem of point state estimation is given by a pair $(\hat{\varphi}_k, \rho_{e,k})$, where Bayes estimate $\hat{\varphi}_k$ is given by*

$$\hat{\varphi}_k = \arg\min_{\varphi_k \in \mathfrak{X}} \mathbb{E}[L(\varphi_k, \Phi_k)] \tag{3.58}$$

$$= \arg\min_{\varphi_k \in \mathfrak{X}} \int L(\varphi_k, \varphi) p_{\Phi_k}(\varphi | Z_{1:k}) \mathrm{d}\varphi, \tag{3.59}$$

and associated Bayes expected loss $\rho_{e,k}$ is given by

$$\rho_{e,k} = \mathbb{E}[L(\hat{\varphi}_k, \Phi_k)] \tag{3.60}$$

$$= \int L(\hat{\varphi}_k, \varphi) p_{\Phi_k}(\varphi | Z_{1:k}) \mathrm{d}\varphi, \tag{3.61}$$

for a loss function of a kind $L : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}_0^+$.

Query loss functions, like (A.1), which are commonly used for state estimation, are not applicable here. This is because such function are commonly formulated based on the miss-distance (or error metric), and it is the case that is not defined on the point process state space $\mathfrak{X}$, what is also discussed in [98].[5] Inapplicability of MMSE estimation based on (A.1) in the point process context is well recognised [70].

Optimal MMOSPA estimation Although the population state space itself is not equipped with a metric, there have been a considerable effort to define such metric. One of the outcomes is the Optimal Sub-Pattern Assignment (OSPA) metric [98], which has been a valuable tool for the cross-evaluation of filtering algorithms.[6]

[5] An alternative to defining the query loss function using error $a - \breve{a}$ may be found in [13, 43].
[6] Without taking object labelling into account.

This metric establish distance between two populations and relies on additional user-defined parameters c and p that specify, respectively, sensitivity to errors in location and cardinality.

Definition 3.5.3 (OSPA metric [98]). *For sensitivity parameters* $1 \leq p \leq \infty$ *and* $c > 0$*, the Optimal SubPattern Assignment (OSPA) distance between two populations* $\varphi, \varphi' \in \mathfrak{X}$ *with possibly distinct cardinality is defined as*

$$\bar{d}_p^{(c)}(\varphi, \varphi') := \frac{1}{|\varphi'|}\left[\min_{\pi \in \Pi_{|\varphi'|}} \sum_{i=1}^{|\varphi|} d^{(c)}(x_i, x'_{\pi(i)})^p + c^p(|\varphi'| - |\varphi|)\right] \tag{3.62}$$

for $|\varphi| \leq |\varphi'|$*, and* $\bar{d}_p^{(c)}(\varphi, \varphi') := \bar{d}_p^{(c)}(\varphi', \varphi)$ *for* $|\varphi| > |\varphi'|$*, where* $d^{(c)}(x, x') = \min(c, d(x, x'))$ *is a distance between two individuals in states* $x, x' \in \mathcal{X}$ *with cut-off* c*, and* Π_n *is the set of permutations on* $\{1, 2, \cdots, n\}$ *for any* $n \in \mathbb{N}$.

The possibility of using this metric as an overall loss function to produce state estimates was first proposed in the form of the Minimum Mean OSPA (MMOSPA) estimator in [36]. Following Proposition 3.5.2, this estimate is defined as [8, Eq. 6]

$$\hat{\varphi}_k = \arg\min_{\varphi_k \in \mathfrak{X}} \mathbb{E}[\bar{d}_p^{(c)}(\varphi_k, \Phi_k)^2] \tag{3.63}$$

$$= \arg\min_{\varphi_k \in \mathfrak{X}} \int \bar{d}_p^{(c)}(\varphi_k, \varphi)^2 p_{\Phi_k}(\varphi | Z_{1:k}) \mathrm{d}\varphi. \tag{3.64}$$

Unfortunately, computation of such estimate is very challenging in real-life scenarios, e.g. when the number of objects is not known or exceeds two. As a consequence few practical results have been demonstrated. Furthermore, it is not clear how to establish a relation between the problem being solved by the user and the sensitivity parameters in a principled manner.

Generalized MAP estimation Two algorithms denoted 'GMAP-I' and 'GMAP-II' were proposed in [34] as generalizations (or global 'analogs') of the regular Maximum A Posteriori (MAP) estimator (see Section 2.3.2). It was claimed that these estimators are Bayes-optimal for the loss function presented in [34, p. 192]. Various updates to these estimators were presented in [70], however, no explicit expressions to compute the expected value of loss, as a valuable indicator of quality, have been provided.

Overall, we find application of these algorithms problematic. First, we were not able to find any applications of these estimators in the context of moment approximation filters.[7] Second, since the MAP estimator itself is not a Bayes-optimal estimator [7], it brings into question Bayes-optimality of its generalizations to non-Euclidean spaces.

Heuristics for point estimation of the system state Numerical implementations of PHD filters can be equipped with state extraction algorithms that are not optimal in Bayes sense:

(a) *Gaussian Mixture (GM) implementations.* Approximate state estimation in the GM-PHD filter [107], (as well as in the GM-PPHD filter [95] and in the GM-CPHD filter [109]) concerns first computing the expected number of objects in the scene, and then using the posterior intensity to extract the corresponding number of mixture components with the highest weights as state estimates.

(b) *Sequential Monte Carlo (SMC) implementations.* Approximate state estimation in the SMC-PHD filter in [108] (which can be easily extended to the PPHD and the CPHD cases) concerns first computing the expected number of objects in the scene, and then using this number to partition the set of particles into a number of clusters representing the objects. The centres of these clusters are then used as state estimates.

3.5.3 Optimal posterior regional cardinality estimation

Strictly speaking, the material as presented next follows from the developments in Chapter 4, where a more general result is produced. However, the same results can be extracted from [25, 95], where they may not have been stated explicitly and produced within a slightly different context.

Definition 3.5.4 (Regional enumeration).
A query function $q_{\Sigma_B} : \mathfrak{X} \to \mathbb{R}$ given by

$$q_{\Sigma_B}(\varphi) := \sum_{x \in \varphi} \mathbb{1}_B(x), \tag{3.65}$$

where $\mathbb{1}_B : \mathcal{X} \to \{0, 1\}$ is an indicator function for an arbitrary region $B \subset \mathcal{X}$, evaluates to the number of objects in φ that belong to region B.

Theorem 3.5.5 (Posterior regional cardinality estimation [25, 95]).
For a process Φ_k, the query function (3.65) *and the squared error query loss* (A.1),

[7] Some available results [6, 52] concern multi-Bernoulli filters which are point process based filters of different nature than PHD filters.

the Bayesian solution to the problem of regional cardinality estimation on $B \subset \mathcal{X}$ is a tuple $(a_{\Sigma_B,k}, \rho_{\Sigma_B,k})$ obtained from process statistics $(\mu_{\Phi_k}, \nu_{\Phi_k})$ with Bayes action given by

$$a_{\Sigma_B,k} = \arg\min_{a\in\mathbb{R}} \mathbb{E}[l_2(a, q_{\Sigma_B}(\Phi_k))] \tag{3.66a}$$

$$= \arg\min_{a\in\mathbb{R}} \int (a - q_{\Sigma_B}(\varphi))^2 p_{\Phi_k}(\varphi|Z_k)\mathrm{d}\varphi \tag{3.66b}$$

$$= \int_B \mu_{\Phi_k}(x)\mathrm{d}x, \tag{3.66c}$$

and Bayes expected loss given by

$$\rho_{\Sigma_B,k} = \mathbb{E}[l_2(a_k, q_{\Sigma_B}(\Phi_k))] \tag{3.67a}$$

$$= \int (a_k - q_{\Sigma_B}(\varphi))^2 p_{\Phi_k}(\varphi|Z_k)\mathrm{d}\varphi \tag{3.67b}$$

$$= \int_B \mu_\varphi(x)\mathrm{d}x\left(1 - \int_B \mu_{\Phi_k}(x)\mathrm{d}x\right) + \int_{B\times B} \nu_{\Phi_k}(x,\bar{x})\mathrm{d}(x,\bar{x}). \tag{3.67c}$$

Corollary 3.5.6 (Regional cardinality estimation (PHD filter) [25])**.** *For the updated PHD filter in Proposition 3.4.1 and the squared error query loss* (A.1)*, the Bayesian solution $(a^\flat_{\Sigma_B,k}, \rho^\flat_{\Sigma_B,k})$ to the regional cardinality estimation problem on $B \subset \mathcal{X}$ is given by*

$$a^\flat_{\Sigma_B,k} = \mu^\phi_k(B) + \sum_{z\in Z_k} \frac{\mu^z_k(B)}{\mu_{\mathrm{fa},k}(z) + \mu^z_k(\mathcal{X})}, \tag{3.68}$$

$$\rho^\flat_{\Sigma_B,k} = \mu^\phi_k(B) + \sum_{z\in Z_k} \frac{\mu^z_k(B)}{\mu_{\mathrm{fa},k}(z) + \mu^z_k(\mathcal{X})}\left(1 - \sum_{z\in Z_k} \frac{\mu^z_k(B)}{\mu_{\mathrm{fa},k}(z) + \mu^z_k(\mathcal{X})}\right). \tag{3.69}$$

Corollary 3.5.7 (Regional cardinality estimation (PPHD filter) [95])**.** *For the updated PPHD filter in Proposition 3.4.2 and the squared error query loss* (A.1)*, the Bayesian solution $(a^\natural_{\Sigma_B,k}, \rho^\natural_{\Sigma_B,k})$ to the regional cardinality estimation problem on $B \subset \mathcal{X}$ is given by*

$$a^\natural_{\Sigma_B,k} = \mu^\phi_k(B) l^\natural_1(\phi) + \sum_{z\in Z_k} \frac{\mu^z_k(B)}{\mu_{\mathrm{fa},k}(z) + \mu^z_k(\mathcal{X})} l^\natural_1(z), \tag{3.70}$$

$$\begin{aligned}\rho^\natural_{\Sigma_B,k} =& \mu^\natural_k(B) + \mu^\phi_k(B)[l^\natural_2(\phi) - l^\natural_1(\phi)^2] + 2\mu^\phi_k(B) \sum_{z\in Z_k} \frac{\mu^z_k(B)}{c_{\mathrm{fa},k}(z)}[l^\natural_2(z) - l^\natural_1(\phi) l^\natural_1(z)] \\ &+ \sum_{z,\bar{z}\in Z_k} \frac{\mu^z_k(B)}{c_{\mathrm{fa},k}(z)} \frac{\mu^{\bar{z}}_k(B)}{c_{\mathrm{fa},k}(\bar{z})}[l^{\natural,\neq}_2(z,\bar{z}) - l^\natural_1(z) l^\natural_1(\bar{z})],\end{aligned} \tag{3.71}$$

where the $l^{\natural}$ terms for time k are presented in (3.35) *such that $l_2^{\natural,\neq}(z,\bar{z})$ is obtained as $l_2^{\natural}(z,\bar{z})$ if $z \neq \bar{z}$ and zero otherwise.*

Corollary 3.5.8 (Regional cardinality estimation (CPHD filter)[25])**.** *For the updated CPHD filter in Proposition 3.4.3 and the squared error query loss* (A.1)*, the Bayesian solution $(a^{\sharp}_{\Sigma_B,k}, \rho^{\sharp}_{\Sigma_B,k})$ to the regional cardinality estimation problem on $B \subset \mathcal{X}$ is given by*

$$a^{\sharp}_{\Sigma_B,k} = \mu_k^{\phi}(B) l_1^{\sharp}(\phi) + \sum_{z \in Z_k} \frac{\mu_k^z(B)}{\mu_{\mathrm{fa},k}(z) + \mu_k^z(\mathcal{X})} l_1^{\sharp}(z), \tag{3.72}$$

$$\begin{aligned}\rho^{\sharp}_{\Sigma_B,k} =& \mu_k^{\sharp}(B) + \mu_k^{\phi}(B)[l_2^{\sharp}(\phi) - l_1^{\sharp}(\phi)^2] + 2\mu_k^{\phi}(B) \sum_{z \in Z_k} \frac{\mu_k^z(B)}{c_{\mathrm{fa},k}(z)} [l_2^{\sharp}(z) - l_1^{\sharp}(\phi) l_1^{\sharp}(z)] \\ &+ \sum_{z,\bar{z} \in Z_k} \frac{\mu_k^z(B)}{c_{\mathrm{fa},k}(z)} \frac{\mu_k^{\bar{z}}(B)}{c_{\mathrm{fa},k}(\bar{z})} [l_2^{\sharp,\neq}(z,\bar{z}) - l_1^{\sharp}(z) l_1^{\sharp}(\bar{z})],\end{aligned} \tag{3.73}$$

where the first-order $l_1^{\sharp}$ terms for time k are from the original filter recursions in (3.48)*, and* additional *second-order $l_2^{\sharp}$ terms are [25]*

$$\begin{cases} l_2^{\sharp}(\phi) := \dfrac{\langle \Upsilon^2[\mu, Z], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle}, \\ l_2^{\sharp}(z) := \dfrac{\langle \Upsilon^2[\mu, Z \setminus \{z\}], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle}, \\ l_2^{\sharp}(z,\bar{z}) := \dfrac{\langle \Upsilon^2[\mu, Z \setminus \{z,\bar{z}\}], \rho_{k|k-1} \rangle}{\langle \Upsilon^0[\mu, Z], \rho_{k|k-1} \rangle}, \end{cases} \tag{3.74}$$

such that $l_2^{\sharp,\neq}(z,\bar{z})$ is obtained as $l_2^{\sharp}(z,\bar{z})$ if $z \neq \bar{z}$ and zero otherwise.

3.5.4 Ad hoc solutions to other problems

Particular solutions can be obtained using state estimates for a definition of a query function. This approach is closely related to decision making under assumed certainty equivalence, but it does not take a (query) loss function into account. As a consequence, these solutions referred to as ad hoc, i.e. solutions that are designed for specific problem.

Definition 3.5.9 (Centroid)**.**
A query function $q_{\mathrm{centroid}} : \mathfrak{X} \to \mathcal{X}$ given by [20, Eq. 32]

$$q_{\mathrm{centroid}}(\varphi) := \frac{1}{|\varphi|} \sum_{x \in \varphi} x, \tag{3.75}$$

where $|\cdot|$ *is a cardinality of the set.*

Definition 3.5.10 (Regional density)**.**
A query function $q_{\text{density}} : \mathfrak{X} \to \mathbb{R}$ *given by*

$$q_{\text{density}}(\varphi) := \frac{1}{B} \sum_{x \in \varphi} \mathbb{1}_B(x). \tag{3.76}$$

3.6 Summary

In this chapter we addressed the new generation of Bayesian filters that infer probabilistic description of an object population from partial data. In contrast to early recursive filters for object populations, the new algorithms are designed based on the explicitly stated modelling assumptions. Designing these filters have required new methods for describing uncertain populations, which are available from point process theory. In general, the Bayes filter for an object population is intractable, so we have presented a number of practical moment approximation filters, including the classical PHD filter, the Panjer PHD filter, and the Cardinalized PHD filter.

Decision making under uncertainty using such filters has been focused on the problem of point estimation. However, this problem has not received an optimal solution due to the nature of the point process state space. As a result, most implementations of such algorithms rely on various heuristics to produce the state estimate. Nevertheless, we used the new decision-theoretic framework to interpret the problem of optimal regional cardinality estimation.

Chapter 4

Decision making with spatial point processes

Spatial point processes are commonly associated with their application in the context of statistical inference. Among its goals is producing quantities characterising the process and its probability distribution. As presented in the previous chapter, the new generation of Bayesian filters is primarily focused on the first-order moment of the process [68] and additional cardinality statistics [25, 67, 95].

In this chapter, we are going to exploit spatial point processes for decision making under uncertainty. As far as first-order filtering is concerned, optimal state estimation using point processes has been deemed a challenging problem because of the difficulties associated with implementing the loss functions as described in Subsection 3.5.2. As a response to this, we are going to take a principled look at decision problems beyond the point estimation of the system state. We will show that problems that are *more specialized* can have optimal solutions in closed form. Furthermore, in certain situations they can also be expressed using a limited variety of point process statistics, which might eventually be extracted from the Bayesian filters.

The content of this chapter is as follows. In Section 4.1 we present a model of the decision-making process for a real-valued query function under squared error query loss, and develop its closed-form solution. In Section 4.2, we develop solutions for the sum query, which are expressed using lower-order statistics of a point process. In Section 4.3, we develop solutions for the product query, which are expressed using p.g.fl.s of a point process. Each solution is considered for a number of elementary point processes and for the superposition of point processes. Section 4.4 offers a summary.

4.1 Optimal solution for a real-valued query

In this section we approach decision making with a spatial point process as defined in Section 3.2.1. The design choices outlined in Section 1.5 lead to the construction of the following loss function.

Assumptions 4.1.1 (Squared error-in-answer loss). *The amount of loss associated with reporting an answer $a \in \mathbb{R}$ when the true state of nature is $\varphi \in \mathfrak{X}$ is given by*

$$L_{\text{sq}}(a, \varphi) := l_2(a, q_{\mathbb{R}}(\varphi)) \tag{4.1a}$$
$$= (a - q_{\mathbb{R}}(\varphi))^2 , \tag{4.1b}$$

where $q_{\mathbb{R}} : \mathfrak{X} \to \mathbb{R}$ is an arbitrary real-valued query function, and $l_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_0^+$ is the squared error query loss function in (A.1)*.*

For the loss function in (4.1), following the minimum expected loss principle in Proposition 1.2.2 results into the solution presented in Theorem 4.1.2.

Theorem 4.1.2 (Optimal solution for L_{sq})**.** *Optimal solution to the statistical problem characterised by the loss function L_{sq} in* (4.1) *is obtained as* $(a_{\text{sq},\Phi}, \rho_{\text{sq},\Phi})$ *for a point process Φ on $\mathcal{X}$ with*

$$a_{\text{sq},\Phi} = \mathbb{E}[q(\Phi)], \tag{4.2}$$
$$\rho_{\text{sq},\Phi} = \text{var}[q(\Phi)]. \tag{4.3}$$

Proof. Let us consider a function f, such that

$$f(a) = \mathbb{E}[L_{\text{sq}}(a, \Phi)] \tag{4.4a}$$
$$= \mathbb{E}[(a - q(\Phi))^2] \tag{4.4b}$$
$$= a^2 - 2a\mathbb{E}[q(\Phi)] + \mathbb{E}[q(\Phi)^2]. \tag{4.4c}$$

An extremum of the function f in point a is found as $f'(a) = 0$. The first derivative is found to be

$$f'(a) = 2a - 2\mathbb{E}[q(\Phi)], \tag{4.5}$$

which gives us the extremum for

$$a = \mathbb{E}[q(\Phi)]. \tag{4.6}$$

This point is a minimum if $f''(a) > 0$, and this is true since $f''(a) = 2$. Finally, the minimum that is reached by this function is

$$\begin{aligned} f(\mathbb{E}[q(\Phi)]) &= \mathbb{E}[L_{\text{sq}}(\mathbb{E}[q(\Phi)], \Phi)] && (4.7\text{a}) \\ &= \mathbb{E}[(\mathbb{E}[q(\Phi)] - q(\Phi))^2] && (4.7\text{b}) \\ &= \mathbb{E}[q(\Phi)]^2 - 2\mathbb{E}[q(\Phi)]\mathbb{E}[q(\Phi)] + \mathbb{E}[q(\Phi)^2] && (4.7\text{c}) \\ &= \mathbb{E}[q(\Phi)^2] - \mathbb{E}[q(\Phi)]^2 && (4.7\text{d}) \\ &= \text{var}[q(\Phi)]. && (4.7\text{e}) \end{aligned}$$

□

Corollary 4.1.3 (Explicit expressions)**.** *For a point process Φ on $\mathcal{X}$ described by its projection density $p_{\Phi}^{(n)}(x_{1:n})$ for $n \in \mathbb{N}$, the optimal solution $(a_{\text{sq},\Phi}, \rho_{\text{sq},\Phi})$ in Theorem 4.1.2 can be written as*

$$a_{\text{sq},\Phi} = \sum_{n \geq 0} \int q(x_{1:n}) p_{\Phi}^{(n)}(x_{1:n}) \mathrm{d}(x_{1:n}), \tag{4.8}$$

$$\rho_{\text{sq},\Phi} = \sum_{n \geq 0} \int q(x_{1:n})^2 p_{\Phi}^{(n)}(x_{1:n}) \mathrm{d}(x_{1:n}) - \left[\sum_{n \geq 0} \int q(x_{1:n}) p_{\Phi}^{(n)}(x_{1:n}) \mathrm{d}(x_{1:n}) \right]^2. \tag{4.9}$$

4.2 Optimal solutions for the sum query

In Section 4.1 we obtained the optimal solution to the problem characterised by the squared error-in-answer loss (4.1). This loss function is originally defined for an arbitrary real-valued query $q_{\mathbb{R}} : \mathfrak{X} \to \mathbb{R}$. In this subsection we additionally assume that the query function is of the additive form.

Definition 4.2.1 (Sum query)**.**
A real-valued query is called the sum query and defined as

$$q_{\Sigma}(\varphi) := K \sum_{x \in \varphi} m(x) + C \tag{4.10}$$

for a real-valued function $m : \mathcal{X} \to \mathbb{R}$, and constant coefficients K and C on $\mathbb{R}$.

Note that expression in (4.10) can be seen as a generalization of the query function for regional enumeration given in (3.65).

Remark 4.2.2 (Empty sum)**.** *An empty sum (nullary sum) is a summation where the number of elements is zero. The value of any empty sum of numbers, by convention, is the additive identity, i.e. zero.*

Accordingly, we can construct a specialized loss function.

Definition 4.2.3 (Squared error-in-sum loss). *The amount of loss associated with taking an action $a \in \mathbb{R}$ when the true state of environment is $\varphi \in \mathfrak{X}$ is given by*

$$L_\Sigma(a, \varphi) := \left(a - q_\Sigma(\varphi)\right)^2, \tag{4.11}$$

where the sum query $q_\Sigma : \mathfrak{X} \to \mathbb{R}$ is given in (4.10).

4.2.1 General solution

For a point process Φ described by its first-order moment density μ_Φ and second-order factorial moment density ν_Φ, we define additional notations:

$$F_\Phi[m] := \int m(x)\mu_\Phi(x)\mathrm{d}x, \tag{4.12}$$

$$Q_\Phi[m] := \int m(x)m(\bar{x})\nu_\Phi(x, \bar{x})\mathrm{d}(x, \bar{x}), \tag{4.13}$$

where $m : \mathcal{X} \to \mathbb{R}$ is a function.

Theorem 4.2.4 (Optimal solution for L_Σ [76]).
Optimal solution to the problem characterised by the loss function L_Σ in (4.11) *is obtained as $(a_{\Sigma,\Phi}, \rho_{\Sigma,\Phi})$ for the point process Φ from its statistics (μ_Φ, ν_Φ) with*

$$a_{\Sigma,\Phi} = K F_\Phi[m] + C, \tag{4.14}$$

$$\rho_{\Sigma,\Phi} = K^2 \left[F_\Phi[m^2] - F_\Phi[m]^2 + Q_\Phi[m] \right]. \tag{4.15}$$

This technical result (in a slightly more specialized form) was originally presented in the conference publication [76], and the proof is given next.

Proof. Let us obtain the solution $(a_{\Sigma,\Phi}, \rho_{\Sigma,\Phi})$ to the decision problem characterised by L_Σ and the object process Φ. First we focus on the optimal action $a_{\Sigma,\Phi}$. From (4.8) and (4.10) we can write

$$a_{\Sigma,\Phi} = \sum_{n\geq 0} \int \left(K \sum_{1\leq i\leq n} m(x_i) + C \right) p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}), \tag{4.16a}$$

$$= K \sum_{n\geq 0} \int \left(\sum_{1\leq i\leq n} m(x_i) \right) p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) + C, \tag{4.16b}$$

then using Campbell's theorem [104, p. 103], yields the optimal action

$$a_{\Sigma,\Phi} = K \int m(x)\mu_\Phi(x)\mathrm{d}x + C. \tag{4.16c}$$

Next we focus on the minimised expected loss $\rho_{\Sigma,\Phi}$ given in (4.7d) by

$$\rho_{\Sigma,\Phi} = \mathbb{E}[q_\Sigma(\Phi)^2] - \mathbb{E}[q_\Sigma(\Phi)]^2. \tag{4.17}$$

The expected value $\mathbb{E}[q_\Sigma(\Phi)^2]$ is obtained from (4.8) and (4.10) and written as

$$\begin{aligned}
\mathbb{E}[q_\Sigma(\Phi)^2] =& \sum_{n\geq 0} \int \left[K \sum_{1\leq i\leq n} m(x_i) + C \right]^2 p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) && (4.18\text{a})\\
=& K^2 \sum_{n\geq 0} \int \left[\sum_{1\leq i\leq n} m(x_i)^2 \right] p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n})\\
&+ K^2 \sum_{n\geq 0} \int \left[\sum_{1\leq i,j\leq n}^{\neq} m(x_i)m(x_j) \right] p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n})\\
&+ 2K \sum_{n\geq 0} \int \left[\sum_{1\leq i\leq n} m(x_i) \right] p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) + C^2 && (4.18\text{b})\\
=& K^2 \int m(x)^2 \mu_\Phi(x)\mathrm{d}x + K^2 \int m(x)m(\bar{x})\nu_\Phi(x,\bar{x})\mathrm{d}(x,\bar{x})\\
&+ 2K \int m(x)\mu_\Phi(x)\mathrm{d}x + C^2, && (4.18\text{c})
\end{aligned}$$

where ν_Φ is the second-order *factorial* moment density of the point process Φ. At the same time, following (4.2), the expression of $\mathbb{E}[q_\Sigma(\Phi)]$ is given by (4.16c).

Substituting (4.18c) and (4.16c) into (4.17) yields an expression of the expected loss value $\rho_{\Sigma,\Phi}$. □

Corollary 4.2.5 (Regional statistics [25])**.**
For $K = 1$, $C = 0$ and $m(\cdot) = \mathbb{1}_B(\cdot)$, where $\mathbb{1}_B$ is an indicator function $\mathbb{1}_B : \mathcal{X} \to [0, 1]$ for an arbitrary region $B \subset \mathcal{X}$ such that $\mathbb{1}_B(x)$ is equal to 1 if $x \in B$ and 0 otherwise, the optimal solution in Theorem 4.2.4 expresses statistics on the number of objects for a point process Φ on $\mathcal{X}$ described with its statistics (μ_Φ, ν_Φ) is given by

$$\mu(B) = F[\mathbb{1}_B], \tag{4.19}$$

$$\mathrm{var}(B) = F[\mathbb{1}_B](1 - F[\mathbb{1}_B]) + Q[\mathbb{1}_B]. \tag{4.20}$$

4.2.2 Elementary point processes

We consider three different elementary point processes to illustrate decision making for the squared error-in-sum loss L_Σ in (4.11). All of the point processes have been introduced in Subsection 3.2.3.

Corollary 4.2.6 (Solution for a Poisson point process and L_Σ)**.** *For a Poisson process with parameter $\lambda_\flat$ and spatial distribution $c_\flat$, optimal solution to the decision problem characterised by the loss function L_Σ in* (4.11) *is obtained as* $(a_{\Sigma,\flat}, \rho_{\Sigma,\flat})$ *with*

$$a_{\Sigma,\flat} = K\lambda_\flat \int m(x)c_\flat(x)\mathrm{d}x + C, \tag{4.21}$$

$$\rho_{\Sigma,\flat} = K^2\lambda_\flat \int m(x)^2 c_\flat(x)\mathrm{d}x. \tag{4.22}$$

Corollary 4.2.7 (Solution for a Panjer point process and L_Σ)**.** *For a Panjer process with parameters $\alpha_\natural$ and $\beta_\natural$ and spatial distribution $c_\natural$, optimal solution to the decision problem characterised by the loss function L_Σ in* (4.11) *is obtained as* $(a_{\Sigma,\natural}, \rho_{\Sigma,\natural})$ *with*

$$a_{\Sigma,\natural} = K\frac{\alpha_\natural}{\beta_\natural} \int m(x)c_\natural(x)\mathrm{d}x + C, \tag{4.23}$$

$$\rho_{\Sigma,\natural} = K^2\left[\frac{\alpha_\natural}{\beta_\natural} \int m(x)^2 c_\natural(x)\mathrm{d}x + \left[\frac{\alpha_\natural}{\beta_\natural^2} \int m(x)c_\natural(x)\mathrm{d}x\right]^2\right] \tag{4.24}$$

Corollary 4.2.8 (Solution for an i.i.d. cluster point process and L_Σ)**.** *For an i.i.d. cluster process with cardinality distribution $\rho_\sharp$ and spatial distribution $c_\sharp$, optimal solution to the decision problem characterised by the loss function L_Σ in* (4.11) *is obtained as* $(a_{\Sigma,\sharp}, \rho_{\Sigma,\sharp})$ *with*

$$a_{\Sigma,\sharp} = K\sum_{n\geq 0} n\rho_\sharp(n) \int m(x)c_\sharp(x)\mathrm{d}x + C, \tag{4.25}$$

$$\begin{aligned}\rho_{\Sigma,\sharp} &= K^2\sum_{n\geq 0} n\rho_\sharp(n) \int m(x)^2 c_\sharp(x)\mathrm{d}x \\ &\quad + K^2\left(\sum_{n\geq 0}(n^2-n)\rho_\sharp(n) + \left[\sum_{n\geq 0} n\rho_\sharp(n)\right]^2\right)\left(\int m(x)c_\sharp(x)\mathrm{d}x\right)^2\end{aligned} \tag{4.26}$$

4.2.3 Superimposed processes

Superposition occurs when one is not interested in the individual realizations of independent point processes, but only in the union of the realizations. If we denote by Ψ the union of point processes with known p.g.fl.s , then Ψ is also a point process.

Corollary 4.2.9 (Solution for superimposed processes and L_Σ).
For a superposition Ψ of N point processes $\cup_{1\le i\le N}\Phi_i$ described by their p.g.fl.s $\mathcal{G}_{\Phi_i}$, a Bayesian solution $(a_{\Sigma,\Psi}, \rho_{\Sigma,\Psi})$ for the loss function L_Σ in (4.11) is given by

$$a_{\Sigma,\Psi} = K \sum_{1\le i\le N} F_{\Phi_i}[m] + C, \tag{4.27}$$

$$\rho_{\Sigma,\Psi} = K^2 \sum_{1\le i\le N} \Big[F_{\Phi_i}[m^2] - F_{\Phi_i}[m]^2 + Q_{\Phi_i}[m]\Big]. \tag{4.28}$$

4.3 Optimal solutions for the product query

In Section 4.1 we obtained the optimal solution to the decision problem characterised by the squared error-in-action loss (4.1). This loss function is originally defined for an arbitrary real-valued query $q_\mathbb{R} : \mathfrak{X} \to \mathbb{R}$. In this subsection we additionally specify that the query function is of a multiplicative form.

Definition 4.3.1 (Product query).
A real-valued query is called the product query and defined as

$$q_\Pi(\varphi) := K \prod_{x\in\varphi} m(x) + C \tag{4.29}$$

for a function $m : \mathcal{X} \to [0,1]$, and constant coefficients K and C on $\mathbb{R}$.

Note the difference in the definition of m in Definitions 4.2.1 and 4.3.1.

Remark 4.3.2 (Empty product). *An empty product (nullary product) is the outcome of multiplying no factors. The value of any empty product, by convention, is equal to the multiplicative identity 1.*

Accordingly, we can construct a specialized loss function.

Definition 4.3.3 (Squared error-in-product loss). *The amount of loss associated with taking an action $a \in \mathbb{R}$ when the true state of environment is $\varphi \in \mathfrak{X}$ is given by*

$$L_\Pi(a, \varphi) := (a - q_\Pi(\varphi))^2\,, \tag{4.30}$$

where the product query $q_\Pi : \mathfrak{X} \to \mathbb{R}$ is given in (4.29).

4.3.1 General solution

Recall the definition of the probability generating functional (p.g.fl.) $\mathcal{G}_\Phi$ for a point process Φ given in (3.6).

Theorem 4.3.4 (Optimal solution for L_Π).
Optimal solution to the problem characterised by the loss function L_Π in (4.30) is obtained as $(a_{\Pi,\Phi}, \rho_{\Pi,\Phi})$ for the point process Φ from its p.g.fl. $\mathcal{G}_\Phi$ with

$$a_{\Pi,\Phi} = K\mathcal{G}_\Phi(m) + C, \tag{4.31}$$

$$\rho_{\Pi,\Phi} = K^2\big(\mathcal{G}_\Phi(m^2) - \mathcal{G}_\Phi(m)^2\big). \tag{4.32}$$

Proof. Let us obtain the solution $(a_{\Pi,\Phi}, \rho_{\Pi,\Phi})$ to the decision problem characterised by L_Π and the object process Φ. First we focus on the optimal action $a_{\Pi,\Phi}$. From (4.8) and (4.10), we can write

$$a_{\Pi,\Phi} = \sum_{n\geq 0}\int\left[K\prod_{1\leq i\leq n} m(x_i) + C\right]p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) \tag{4.33}$$

$$= K\sum_{n\geq 0}\int\left[\prod_{1\leq i\leq n} m(x_i)\right]p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) + C. \tag{4.34}$$

An expression for the optimal action $a_{\Pi,\Phi}$ is obtained from (4.34) using the definition of the p.g.fl. in (3.6) when considering that $m : \mathcal{X} \to [0, 1]$.

Next we focus on the minimised expected loss $\rho_{\Sigma,\Phi}$ given in (4.9) by

$$\rho_{\Pi,\Phi} = \sum_{n\geq 0}\int q_\Pi(x_{1:n})^2 p_\Phi(x_{1:n})\mathrm{d}(x_{1:n}) - \left[\sum_{n\geq 0}\int q_\Pi(x_{1:n})p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n})\right]^2 \tag{4.35}$$

$$\begin{aligned} &= \sum_{n\geq 0}\int\left[K\prod_{1\leq i\leq n} m(x_i) + C\right]^2 p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) \\ &\quad - \left[\sum_{n\geq 0}\int\left[K\prod_{1\leq i\leq n} m(x_i) + C\right]p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n})\right]^2 \end{aligned} \tag{4.36}$$

$$\begin{aligned} &= K^2\sum_{n\geq 0}\int\left[\prod_{1\leq i\leq n} m(x_i)^2\right]p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n}) \\ &\quad - K^2\left[\sum_{n\geq 0}\int\left[\prod_{1\leq i\leq n} m(x_i)\right]p_\Phi^{(n)}(x_{1:n})\mathrm{d}(x_{1:n})\right]^2 \end{aligned} \tag{4.37}$$

An expression for the minimized expected loss $\rho_{\Pi,\Phi}$ is obtained from (4.37) using the definition of p.g.fl. in (3.6).

□

4.3.2 Elementary point processes

We consider three different point processes to illustrate decision making for the squared error-in-product loss. All of the point processes have been introduced in Subsection 3.2.3.

Corollary 4.3.5 (Solution for a Poisson point process and L_Π)**.** *For a Poisson process with parameter $\lambda_\flat$ and spatial distribution $c_\flat$, optimal solution to the decision problem characterised by the loss function L_Π in* (4.30) *is obtained as $(a_{\Pi,\flat}, \rho_{\Pi,\flat})$ with*

$$a_{\Pi,\flat} = K\exp\left(\lambda_\flat \int [m(x)-1]c_\flat(x)\mathrm{d}x\right) + C, \tag{4.38}$$

$$\rho_{\Pi,\flat} = K^2\left[\exp\left(\lambda_\flat \int [m(x)^2-1]c_\flat(x)\mathrm{d}x\right) - \exp\left(2\lambda_\flat \int [m(x)-1]c_\flat(x)\mathrm{d}x\right)\right]. \tag{4.39}$$

Corollary 4.3.6 (Solution for a Panjer point process and L_Π)**.** *For a Panjer process with parameters $\alpha_\natural$ and $\beta_\natural$ and spatial distribution $c_\natural$, optimal solution to the decision problem characterised by the loss function L_Π in* (4.30) *is obtained as $(a_{\Pi,\natural}, \rho_{\Pi,\natural})$ with*

$$a_{\Pi,\natural} = K\left[1 + \frac{1}{\beta_\natural}\int [1-m(x)]c_\natural(x)\mathrm{d}x\right]^{-\alpha_\natural} + C, \tag{4.40}$$

$$\rho_{\Pi,\natural} = K^2\left[\left[1 + \frac{1}{\beta_\natural}\int [1-m(x)^2]c_\natural(x)\mathrm{d}x\right]^{-\alpha_\natural} - \left[1 + \frac{1}{\beta_\natural}\int [1-m(x)]c_\natural(x)\mathrm{d}x\right]^{-2\alpha_\natural}\right]. \tag{4.41}$$

Corollary 4.3.7 (Solution for an i.i.d. cluster point process and L_Π)**.** *For an i.i.d. cluster process with cardinality distribution $\rho_\sharp$ and spatial distribution $c_\sharp$, optimal solution to the decision problem characterised by the loss function L_Π in* (4.30) *is obtained as $(a_{\Pi,\sharp}, \rho_{\Pi,\sharp})$ with*

$$a_{\Pi,\sharp} = K\sum_{n\geq 0}\rho_\sharp(n)\left[\int m(x)c_\sharp(x)\mathrm{d}x\right]^n + C, \tag{4.42}$$

$$\rho_{\Pi,\sharp} = K^2 \left[\sum_{n \geq 0} \rho_\sharp(n) \left[\int m(x)^2 c_\sharp(x) \mathrm{d}x \right]^n - \left[\sum_{n \geq 0} \rho_\sharp(n) \left[\int m(x)^2 c_\sharp(x) \mathrm{d}x \right]^n \right]^2 \right]. \tag{4.43}$$

4.3.3 Superimposed processes

Corollary 4.3.8 (Solution for superimposed processes and L_Π)**.**
Optimal solution to the decision problem characterised by the loss function L_Π *in* (4.30) *is obtained as* $(a_{\Pi,\Psi}, \rho_{\Pi,\Psi})$ *for a superposition* Ψ *of* N *point processes* $\cup_{1\leq i\leq N}\Phi_i$ *described by their p.g.fl.'s* $\mathcal{G}_{\Phi_i}$ *with*

$$a_{\Pi,\Psi} = K \prod_{1\leq i\leq N} \mathcal{G}_{\Phi_i}(m) + C, \tag{4.44}$$

$$\rho_{\Pi,\Psi} = K^2 \left[\prod_{1\leq i\leq N} \mathcal{G}_{\Phi_i}(m^2) - \prod_{1\leq i\leq N} \mathcal{G}_{\Phi_i}(m)^2 \right]. \tag{4.45}$$

Proof. The solution is obtained from Theorem 4.3.4 using the property that p.g.fl. of a superposition is simply $\mathcal{G}_\Psi(h) = \prod_{1\leq i\leq N} \mathcal{G}_{\Phi_i}(h)$. □

4.4 Summary

In this chapter we aimed to employ point processes for decision making under uncertainty. In order to address problems beyond basic point estimation, we had to formulate novel loss functions. As a consequence, we were able to formulate a special class of problems of subjective decision-theoretic inference.

It was found that it is indeed possible to obtain the corresponding optimal solution.We focused on certain specific forms of the query function compatible with the point process state space: the sum query and the product query. We were able to derive optimal solutions that are expressed in closed form through quantities commonly used in characterising a point process: densities of first- and second-order factorial moments and p.g.fl.s.

To be useful in practice, these solutions need to be integrated with Bayesian filters. Specifically, the challenge is to extract the necessary statistical quantities from the practical first-order filters, which is the subject of the next chapter.

Chapter 5

Optimal solutions with moment approximation filters

In the previous chapter we formulated a class of problems involving decision making under uncertainty when the unknown state of the world is described by a point process. For a number of problems, we were able to produce optimal Bayesian solutions which are expressed through a limited number of quantities commonly used to describe a point process.

In this chapter we are going to implement these solutions using information from recursive Bayesian filtering algorithms. Such implementation would offer an opportunity to incorporate available evidence about the world state before a decision is made, or to postpone the moment at which the decision is made. For this, the quantities of interest will have to be obtained from the considered filters. However, this is associated with a number of difficulties if practical filters for spatial point processes are used due to the approximate nature of the algorithms. Nevertheless, we overcome this obstacle using the developments from the point process theory, and obtain solutions for a number of filters, namely, the classical Probability Hypothesis Density (PHD) filter, the Panjer PHD (PPHD) filter, and the Cardinalized PHD (CPHD) filter.

The content of this chapter is as follows. We begin by restating the assumptions underlying the moment approximation filters and introduce additional notations in Section 5.1. In Section 5.2 we obtain the required quantities. Finally, we implement the Bayesian solutions for the considered PHD filters: in Section 5.3 for the squared error-in-sum loss, and in Section 5.4 for the squared error-in-product loss. Section 5.5 summarises the findings.

5.1 Prerequisites

We shall first revisit assumptions involved in filter derivations and introduce useful notations.

5.1.1 Filtering assumptions

Operation of a Bayesian filter involves the data update step, when new data are integrated into the probabilistic description maintained by the filter on a dynamic system, and the time update (or prediction) step when this probabilistic description is propagated in time.

In the data update step, a point process described by $\Phi_{k|k-1}$, which is predicted from time step $k-1$, is updated to Φ_k (sometimes denoted $\Phi_{k|k}$) using the observation Z_k collected by some sensor at current time k. This step relies on the following assumptions.

Assumptions 5.1.1 (Data update step)**.**

(a) The predicted object process $\Phi_{k|k-1}$ is Poisson with rate $\lambda_k^\flat$ and spatial distribution $c_k^\flat$ in the PHD filter, Panjer with parameters $\alpha_{k|k-1}^\natural$ and $\beta_{k|k-1}^\natural$ and spatial distribution $c_k^\natural$ in the PPHD filter, or i.i.d. cluster with cardinality distribution $\rho_{k|k-1}^\sharp$ and spatial distribution $c_k^\sharp$ in the CPHD filter. The intensity of the predicted process is denoted by $\mu_{k|k-1}$.

(b) The measurements originating from object detections are generated independently from each other.

(c) An object with state $x \in \mathcal{X}$ is detected with probability $p_{\mathrm{d},k}(x)$; if so, it produces a measurement whose state is distributed according to a likelihood $g_k(\cdot|x)$.

(d) At time k, the process describing false alarms produced by the sensor is Poisson with rate $\lambda_{\mathrm{fa},k}^\flat$ and spatial distribution $c_{\mathrm{fa},k}^\flat$ in the PHD filter, Panjer with parameters $\alpha_{\mathrm{fa},k}^\natural$ and $\beta_{\mathrm{fa},k}^\natural$ and spatial distribution $c_{\mathrm{fa},k}^\natural$ in the PPHD filter, or i.i.d. cluster with cardinality distribution $\rho_{\mathrm{fa},k}^\sharp$ and spatial distribution $c_{\mathrm{fa},k}^\sharp$ in the CPHD filter.

The predicted object process $\Phi_{k+1|k}$ is obtained from the posterior process Φ_k using knowledge on the dynamical behaviour of the objects. The assumptions of the time update step are as follows.

Assumptions 5.1.2 (Time update step)**.**

(a) The objects evolve independently from each other.

(b) An object with state $x \in \mathcal{X}$ at time k survived to the current time $k+1$ with probability $p_{s,k+1}(x)$ or simply $p_s(x)$; if it did so, its state evolved according to a Markov transition kernel $\pi_{k+1|k}(\cdot|x)$ or $\pi(\cdot|x)$. In the PPHD filter, the probability of survival is uniform over the state space, i.e. $p_{\mathrm{s}}(x) := p_{\mathrm{s}}$ for any $x \in \mathcal{X}$.

(c) New objects entered the scene between time k and $k+1$ independently of the existing objects and described by a birth process Φ^{b}_{k+1} with statistics $(\mu^{\mathrm{b}}_{k+1}, \nu^{\mathrm{b}}_{k+1})$ and the p.g.fl. $\mathcal{G}^{\mathrm{b}}_{k+1}$, which is a Poisson process in the PHD filter, a Panjer process in the PPHD filter and an i.i.d. cluster process in the CPHD filter.[1]

5.1.2 Additional notations

Now we introduce a number of additional notations that will be useful in presentation of the extracted process information and implemented Bayesian solutions.

Recall missed detection term μ^{ϕ}_k and association terms μ^z_k for any $z \in Z_k$ which are expressed using predicted intensity $\mu_{k|k-1}$ as

$$\mu^{\phi}_k(x) = (1 - p_{\mathrm{d},k}(x))\mu_{k|k-1}(x), \tag{5.1}$$

$$\mu^{z}_k(x) = g_k(x|z)p_{\mathrm{d},k}(x)\mu_{k|k-1}(x). \tag{5.2}$$

We introduce following notations using the above quantities, including the predicted intensity, so that

$$F_k[m] := \int m(x)\mu_{k|k-1}(x)\mathrm{d}x, \tag{5.3}$$

$$F^{\phi}_k[m] := \int m(x)\mu^{\phi}_k(x)\mathrm{d}x, \tag{5.4}$$

$$F^{z}_k[m] := \int m(x)\mu^{z}_k(x)\mathrm{d}x, \tag{5.5}$$

for a function $m : \mathcal{X} \to \mathbb{R}$, as well as using birth process statistics $(\mu^{\mathrm{b}}_{k+1}, \nu^{\mathrm{b}}_{k+1})$ so that

$$F^{\mathrm{b}}_{k+1}[m] := \int m(x)\mu^{\mathrm{b}}_{k+1}(x)\mathrm{d}x, \tag{5.6}$$

$$Q^{\mathrm{b}}_{k+1}[m] := \int m(x)m(\bar{x})\nu^{\mathrm{b}}_{k+1}(x)\mathrm{d}(x,\bar{x}). \tag{5.7}$$

[1] Expressions of various statistical information on these elementary point processes was obtained in Chapter 3.

The survival process for an object with state x' at the previous time step can be described with a Bernoulli point process with parameter $p_s(x')$ and spatial distribution $\pi(\cdot|x')$ that is [95]

$$\mathcal{G}_{\mathrm{s}}(m|x') = 1 - p_{\mathrm{s}}(x') + p_{\mathrm{s}}(x') \int m(x)\pi(x|x')\mathrm{d}x. \tag{5.8}$$

Additionally to (5.8), we define $\mathcal{G}_s^1(m|\cdot)$ as

$$\mathcal{G}_s^1(m|x') := p_{\mathrm{s}}(x') \int m(x)\pi(x|x')\mathrm{d}x. \tag{5.9}$$

The above notations are applicable to every considered filter. Next, we shall introduce certain filter-specific terms.

In the course of development we noticed similarities in the expressions of implemented Bayesian solutions, and in order to promote this fact we developed the new ℓ-notations, similar to those first introduced in [25] and [95], though including the dependence on m. For this we first recall the rising factorial and the falling factorial in Definition 3.2.5.

Next, inspired by notations found in [25] and [95], we define the following intermediate Y-terms. In these definitions, the time subscripts on the Y terms are omitted for the sake of simplicity.

Definition 5.1.3 (Y-term for the PHD filter).
For the PHD filter at time k, a supporting term Y_u of order $u \in \{0,1,2\}$ is defined for a function $m : \mathcal{X} \to \mathbb{R}$ as

$$Y_u^{\flat}[m](Z) = e^{F_k^{\phi}[m]} \prod_{z \in Z} (F_k^z[m] + \lambda_{\mathrm{fa},k} c_{\mathrm{fa},k}^{\flat}(z)). \tag{5.10}$$

Definition 5.1.4 (Y-term for the PPHD filter).
For the PPHD filter at time k, a supporting term Y_u of order $u \in \{0,1,2\}$ is defined for a function $m : \mathcal{X} \to \mathbb{R}$ as

$$Y_u^{\natural}[m](Z) = \prod_{z \in Z} c_{\mathrm{fa},k}^{\natural}(z) \sum_{j=0}^{|Z|} \frac{(\alpha_{k|k-1}^{\natural})_{(j+u)\uparrow}}{(\alpha_{k|k-1}^{\natural} + F_k[1] - F_k^{\phi}[m])^{\alpha_{k|k-1}^{\natural}+j+u}} \frac{(\alpha_{\mathrm{fa},k}^{\natural})_{|Z|-j}}{(\beta_{\mathrm{fa},k}^{\natural} + 1)^{|Z|-j}} \cdot \sum_{\substack{I \subseteq Z \\ |I|=j}} \prod_{z \in I} \frac{F_k^z[m]}{c_{\mathrm{fa},k}^{\natural}(z)}. \tag{5.11}$$

Definition 5.1.5 (Y-term for the CPHD filter).

For the CPHD filter at time k, a supporting term Y_u of order $u \in \{0, 1, 2\}$ is defined for a function $m : \mathcal{X} \to \mathbb{R}$ as

$$Y_u^{\sharp}[m](Z) := \prod_{z\in Z} c_{\mathrm{fa},k}^{\sharp}(z) \sum_{n\geq 0} \rho_{k|k-1}^{\sharp}(n) \sum_{j=0}^{\min(|Z|,n-u)} \frac{n!(|Z|-(j+u))!}{(n-(j+u))!} \rho_{\mathrm{fa},k}^{\sharp}(|Z|-j)$$
$$\cdot \frac{F_k^{\phi}[m]^{n-(j+u)}}{F_k[1]^n} \sum_{\substack{I\subseteq Z\\ |I|=j}} \prod_{z\in I} \frac{F_k^z[m]}{c_{\mathrm{fa},k}^{\sharp}(z)}. \tag{5.12}$$

Finally, we use the above Y-terms to define the ℓ terms.

Definition 5.1.6 (ℓ-terms)**.**
For $u \in \mathbb{N}$ and any $z \in Z_k$, the ℓ-terms are given by

$$\ell_u^{\phi}[m] := \frac{Y_u^{\bullet}[m](Z_k)}{Y_0^{\bullet}[1](Z_k)}, \ell_u^{z}[m] := \frac{Y_u^{\bullet}[m](Z_k \setminus \{z\})}{Y_0^{\bullet}[1](Z_k)}, \tag{5.13}$$

where $\bullet \in \{\sharp, \natural, \flat\}$ indicates the filtering solution involved in producing of the Y-terms, respectively, the CPHD filter of (5.12)*, the PPHD of* (5.11)*, or the PHD of* (5.10)*.*

5.2 Point process information from PHD filters

We found that Bayesian solutions obtained in Chapter 4 explicitly rely on a number of quantities commonly used to describe a point process. Expressions of these quantities, such as densities of the lower-order statistical moments of an object process Φ_k and its p.g.fl., can naturally be obtained from the multi-object Bayes' filter using their definitions. However, in most real life scenarios its filtering recursion is not computationally tractable, so instead approximate solutions that propagate incomplete information are used. The classical Probability Hypothesis Density (PHD) [68] filter is perhaps the most popular approximation to the multi-object Bayes' filter, whereas the Panjer PHD (PPHD) and the Cardinalized PHD (CPHD) filters are its extensions constructed to propagate more information about the number of objects.

These filters approximate the predicted object process $\Phi_{k|k-1}$ by one of the elementary point processes presented in Chapter 3, either by a Poisson process (PHD filter), a Panjer process (PPHD filter), or an i.i.d. cluster process (CPHD filter). A Poisson process would be completely described by its intensity function μ_Φ, and this is the statistic propagated by the PHD filter [68]. A Panjer process is described by μ_Φ and variance in the object number $\mathrm{var}_\Phi(\mathcal{X})$, both propagated by the PPHD filter [95]. Analogously, a i.i.d. cluster process is described by μ_Φ and cardinality distribution ρ, both propagated by the CPHD filter [67].

In this section we are going to obtain the required quantities for both the updated and predicted processes. In particular, these are the first moment density, the second factorial moment density and the p.g.fl. of the point process.

5.2.1 Updated point process

The updated process Φ_k is not, in the general case, Poisson (respectively Panjer, i.i.d. cluster), even if the predicted process $\Phi_{k|k-1}$ is; that is, the updated probability distribution P_{Φ_k} is not completely described by the output of the PHD (respectively PPHD, CPHD) filter. Subsequently, it would not be possible, in general, to retrieve the second-order moment density ν_k from intensity μ_k (respectively intensity μ_k and variance $\mathrm{var}_k(\mathcal{X})$, intensity μ_k and cardinality distribution ρ_k) using the expression in Definition 3.2.8 (respectively Definition 3.2.7, Definition 3.2.6). Instead, one could obtain additional expressions for computing this density from intermediate quantities available from a filter's update step [25].

Proposition 5.2.1 (Intensity update [67, 68, 95])**.**
Under Assumptions 5.1.1, a probability hypothesis density function μ_k describing the updated object process Φ_k is given by

$$\mu_k(x) = \mu_k^\phi(x)\ell_1^\phi[1] + \sum_{z\in Z_k} \mu_k^z(x)\ell_1^z[1], \tag{5.14}$$

where the ℓ terms for time k are defined in (5.13)*.*

A general expression (5.14) of the updated intensity μ_k is developed for the PHD filter in [68], for the PPHD filter in [95], and for the CPHD filter in [67].

Proposition 5.2.2 (Second factorial moment density update)**.**
Under Assumptions 5.1.1, a second-order factorial moment density ν_k describing the updated object process Φ_k is given by

$$\begin{aligned}\nu_k(x,\bar{x}) =& \mu_k^\phi(x)\mu_k^\phi(\bar{x})\ell_2^\phi[1] + \mu_k^\phi(x)\sum_{z\in Z_k}\mu_k^z(\bar{x})\ell_2^z[1] \\ &+ \mu_k^\phi(\bar{x})\sum_{z\in Z_k}\mu_k^z(x)\ell_2^z[1] + \sum_{z,\bar{z}\in Z_k}\mu_k^z(x)\mu_k^{\bar{z}}(\bar{x})\ell_2^{\neq,z,\bar{z}}[1],\end{aligned} \tag{5.15}$$

where the ℓ-terms at time k are defined in (5.13)*, and $\ell_2^{\neq,z,\bar{z}}$ is obtained as $\ell_2^{\{z,\bar{z}\}}$ if $z \neq \bar{z}$ and zero otherwise.*

The proof of results related to extracting the second-order factorial moment (5.15) from the respective filters is given next.

Proof. Let us obtain the second-order factorial moment density describing the *updated* object process. Considering [104, Eq. 4.3.4], the second-order factorial moment can be obtained from the second-order *non-factorial* moment, should a suitable expression for the *updated* process be available. Specific expressions for the factorial moment density are dependent on the employed filter and given for the PHD filter in [25, Eq. 31], for the PPHD filter in [95, Eq. 79], and for the CPHD filter in [25, Eq. 29]. Substituting corresponding equations into [104, Eq. 4.3.4] leads to the desired result. □

Proposition 5.2.3 (PGFL update).
Under Assumptions 5.1.1, a p.g.fl. $\mathcal{G}_k$ of the updated object process Φ_k is given by

$$\mathcal{G}_k(h|Z_k) = \ell_0^{\phi}[h] \tag{5.16}$$

for a test function $h : \mathcal{X} \to [0,1]$, where the ℓ term for time k is defined in (5.13)*.*

Proof. We wish to obtain PGFLs describing the updated point process in the classical PHD, the PPHD and the CPHD filters, where for the general multi-object Bayes' filter the p.g.fl. is given in (3.24). for a test function $h : \mathcal{X} \to [0,1]$, where $p_{\Phi_{k|k-1}}$ is the probability density of the predicted process $\Phi_{k|k-1}$, and L_k is the multi-measurement/multi-object likelihood.

Let us begin with obtaining the p.g.fl. $\mathcal{G}_k^{\sharp}[h|Z_k]$ for the CPHD filter. This result can be produced by defining the terms L_k and $p_{\Phi_{k|k-1}}$, closely following the developments in [25]. As presented in Assumptions 5.1.1, in the CPHD filter the predicted process $\Phi_{k|k-1}$ is assumed to be an i.i.d. cluster described by intensity $\mu_{k|k-1}^{\sharp}$ and full cardinality $\rho_{k|k-1}^{\sharp}$, whereas the false alarm process is also an i.i.d. cluster process with spatial density $c_{\mathrm{fa},k}^{\sharp}$ and cardinality $\rho_{\mathrm{fa},k}^{\sharp}$. The cardinality $\rho_{k|k-1}^{\sharp}$ is linked to the term $F_k[1]$ through

$$F_k[1] = \sum_{n\geq 1} n\rho_{k|k-1}^{\sharp}(n). \tag{5.17}$$

The intensity $\mu_{k|k-1}^{\sharp}$ and the cardinality distribution $\rho_{k|k-1}^{\sharp}$ also completely determine the predicted process

$$\forall x_{1:n} \in \mathcal{X}^n, p_{\Phi_{k|k-1}}^{(n)}(x_{1:n}) = \rho_{k|k-1}^{\sharp}(n) \prod_{i=1}^{n} \frac{\mu_{k|k-1}^{\sharp}(x_i)}{F_k[1]}. \tag{5.18}$$

For the CPHD filter a multi-measurement/multi-object likelihood L_k in (3.24) is given by [25]

$$L_k^{\sharp}(Z_k|x_{1:n}) = \sum_{\pi \in \Pi_{|Z_k|,n}} \pi_{\phi}! \rho_{\text{fa},k}^{\sharp}(\pi_{\phi}) \prod_{\substack{z_i \in Z_k \\ (i,\phi)\in\pi}} c_{\text{fa},k}^{\sharp}(z_i) \prod_{(i,j)\in\pi} P(z_i|x_j) \prod_{(\phi,j)\in\pi} P(\phi|x_j), \tag{5.19}$$

where P are the single-measurement/single-target observation kernels, $\Pi_{|Z_k|,n}$ is the set of all the partitions of indexes $\{i_1, \cdots, i_{|Z_k|}, j_1, \cdots, ..., j_n\}$ solely composed of tuples of the form (i_a, j_b) (target x_{j_b} is detected and produces measurement z_{i_a}), (ϕ, j_b) (target x_{j_b} is not detected), or (i_a, ϕ) (measurement z_{i_a} is a false alarm), and $\pi_{\phi} = \sharp\{i|(i,\phi) \in \pi\}$ is the number of clutter measurements given by partition π.

Having described the necessary terms, we can simplify the expression of the numerator in (3.24) for the CPHD filter as

$$\sum_{n\geq 0} \int \left[\prod_{i=1}^{n} h(x_i)\right] L_k^{\sharp}(Z_k|x_{1:n}) p^{(n)}_{\Phi_{k|k-1}}(x_{1:n}) \mathrm{d}x_{1:n} \tag{5.20a}$$

$$= \sum_{n\geq 0} \rho_{k|k-1}^{\sharp}(n) \int \left[\prod_{i=1}^{n} h(x_i)\right] L_k^{\sharp}(Z_k|x_{1:n}) \prod_{i=1}^{n} \frac{\mu_{k|k-1}^{\sharp}(x_i)\mathrm{d}x_i}{F_k[1]} \tag{5.20b}$$

$$= \sum_{n\geq 0} \rho_{k|k-1}^{\sharp}(n) \int L_k^{\sharp}(Z_k|x_{1:n}) \prod_{i=1}^{n} \frac{h(x_i)\mu_{k|k-1}^{\sharp}(x_i)\mathrm{d}x_i}{F_k[1]} \tag{5.20c}$$

$$= \sum_{n\geq 0} \rho_{k|k-1}^{\sharp}(n) \sum_{\pi\in\Pi_{|Z_k|,n}} \pi_{\phi}! \rho_{\text{fa},k}^{\sharp}(\pi_{\phi}) \prod_{(i,\phi)\in\pi} c_{\text{fa},k}^{\sharp}(z_i) \prod_{(i,j)\in\pi} \frac{F_{k|k-1}^{z_i}[h]}{F_k[1]} \cdot \prod_{(\phi,j)\in\pi} \frac{F_k^{\phi}[h]}{F_k[1]} \tag{5.20d}$$

$$= \prod_{z\in Z_k} c_{\text{fa},k}^{\sharp}(z) \sum_{n\geq 0} \rho_{k|k-1}^{\sharp}(n) \sum_{j=0}^{\min(|Z_k|,n)} \frac{n!(|Z_k|-j)!}{(n-j)!} \rho_{\text{fa},k}^{\sharp}(|Z_k|-j) \frac{F_k^{\phi}[h]^{n-j}}{F_k[1]^n}$$
$$\cdot \sum_{\substack{Z\subseteq Z_k \\ |Z|=j}} \prod_{z\in Z} \frac{F_k^{z}[h]}{c_{\text{fa},k}^{\sharp}(z)}. \tag{5.20e}$$

After substituting (5.20e) into (3.24) and cancelling the multiplying constant $\prod_{z\in Z_k} c_{\text{fa},k}^{\sharp}(z)$, we can write

$$\mathcal{G}_{\Phi_k}^{\sharp}[h|Z_k] = \frac{Y_0^{\sharp}[h](Z_k)}{Y_0^{\sharp}[1](Z_k)}, \tag{5.21}$$

with $Y_0^\sharp$ defined in (5.12) for the CPHD filter.

Following from this, we need to obtain an expression of $\mathcal{G}^\natural_{\Phi_k}[h|Z_k]$ for the PPHD filter. Since a Panjer process is a specific case of an i.i.d. cluster process, we start from the CPHD result in (5.20e) with the additional assumptions presented in Assumptions 5.1.1 stating that $\Phi_{k|k-1}$ is a Panjer process described by intensity $\mu^\natural_{k|k-1}$ and parameters $\alpha^\natural_{k|k-1}$ and $\beta^\natural_{k|k-1}$, the cardinality pmf is obtained with

$$\rho^\natural_{k|k-1}(n) = \binom{-\alpha^\natural_{k|k-1}}{n}\left(1+\frac{1}{\beta^\natural_{k|k-1}}\right)^{-\alpha^\natural_{k|k-1}}\left(\frac{-1}{\beta^\natural_{k|k-1}+1}\right)^n, \tag{5.22}$$

and the false alarm process is also Panjer with spatial density $c^\natural_{\mathrm{fa},k}$ and parameters $\alpha^\natural_{\mathrm{fa},k}$ and $\beta^\natural_{\mathrm{fa},k}$.

The multiplying constant $\prod_{z\in Z_k} c^\natural_{\mathrm{fa},k}(z)\big(1+(\beta^\natural_{k|k-1})^{-1}\big)^{-\alpha^\natural_{k|k-1}}$ will be absorbed by the sign of proportion after we expand $\rho^\natural_{k|k-1}$ in

$$\begin{aligned}
&\prod_{z\in Z_k} c^\natural_{\mathrm{fa},k}(z)\sum_{n\gtrsim 0}\rho^\natural_{k|k-1}(n)\sum_{j=0}^{\min(|Z_k|,n)}\frac{n!(|Z_k|-j)!}{(n-j)!}\rho^\natural_{\mathrm{fa},k}(|Z_k|-j)\frac{F^\phi_k[h]^{n-j}}{F_k[1]^n}\\
&\quad\cdot\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F^z_k[h]}{c^\natural_{\mathrm{fa},k}(z)}\\
&\propto\sum_{n\gtrsim 0}\sum_{j=0}^{\min(|Z_k|,n)}\binom{-\alpha^\natural_{k|k-1}}{n}\left(\frac{-1}{1+\beta^\natural_{k|k-1}}\right)^n\frac{n!(|Z_k|-j)!}{(n-j)!}\frac{F^\phi_k[h]^{n-j}}{F_k[1]^n}\rho^\natural_{\mathrm{fa},k}(|Z_k|-j)\\
&\quad\cdot\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F^z_k[h]}{c^\natural_{\mathrm{fa},k}(z)},
\end{aligned}\tag{5.23a}$$

next we expand the binomial coefficient using the identity of a falling factorial so that $n!$ cancel out in

$$\begin{aligned}
&=\sum_{n\gtrsim 0}\sum_{j=0}^{\min(|Z_k|,n)}\left(\frac{-1}{1+\beta^\natural_{k|k-1}}\right)^n\frac{(-\alpha^\natural_{k|k-1})_{n\downarrow}}{(n-j)!}\frac{F^\phi_k[h]^{n-j}}{F_k[1]^n}(|Z_k|-j)!\rho^\natural_{\mathrm{fa},k}(|Z_k|-j)\\
&\quad\cdot\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F^z_k[h]}{c^\natural_{\mathrm{fa},k}(z)},
\end{aligned}\tag{5.23b}$$

now we are going to eliminate the dependency on n by using $\sum_{n\geq 0}\binom{m}{n}x^n = (1+x)^m$ and recognizing that $(-\alpha^\natural_{k|k-1})_{n\downarrow} = (-\alpha^\natural_{k|k-1} - j)_{(n-j)\downarrow}\big((-\alpha^\natural_{k|k-1} - j)_{-j\downarrow}\big)^{-1}$ which yields

$$= \sum_{j=0}^{|Z_k|}\left(1 + \frac{-F_k^\phi[h]}{F_k[1](1+\beta^\natural_{k|k-1})}\right)^{-\alpha^\natural_{k|k-1}-j} \frac{1}{(-\alpha^\natural_{k|k-1} - j)_{-j\downarrow}}\left(\frac{-1}{F_k[1](1+\beta^\natural_{k|k-1})}\right)^j$$
$$\cdot (|Z_k|-j)!\rho^\natural_{\mathrm{fa},k}(|Z_k|-j)\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_k^z[h]}{c^\natural_{\mathrm{fa},k}(z)} \tag{5.23c}$$
$$= \sum_{j=0}^{|Z_k|}\frac{(\alpha^\natural_{k|k-1} + F_k[1] - F_k^\phi[h])^{-\alpha^\natural_{k|k-1}-j}}{(F_k[1](1+\beta^\natural_{k|k-1}))^{-\alpha^\natural_{k|k-1}}}\frac{(-1)^j}{(-\alpha - j)_{-j\downarrow}}(|Z_k|-j)!\rho^\natural_{\mathrm{fa},k}(|Z_k|-j)$$
$$\cdot \sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_k^z[h]}{c^\natural_{\mathrm{fa},k}(z)}, \tag{5.23d}$$

next the proportion sign absorbs the constant $(F_k[1](1+\beta^\natural_{k|k-1}))^{\alpha^\natural_{k|k-1}}$, and we rely on identities $(x)_{-n\uparrow} = \frac{1}{(x-n)_{n\uparrow}}$ and $(x)_{n\uparrow} = (-1)^n(-x)_{n\downarrow}$ to write

$$\propto \sum_{j=0}^{|Z_k|}\frac{(\alpha^\natural_{k|k-1})_{j\uparrow}}{(\alpha^\natural_{k|k-1} + F_k[1] - F_k^\phi[h])^{\alpha^\natural_{k|k-1}+j}}(|Z_k|-j)!\rho^\natural_{\mathrm{fa},k}(|Z_k|-j)\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_k^z[h]}{c^\natural_{\mathrm{fa},k}(z)}, \tag{5.23e}$$

now we finally expand $\rho^\natural_{\mathrm{fa},k}$ to immediately absorb $\big(1 + (\beta^\natural_{\mathrm{fa},k})^{-1}\big)^{-\alpha^\natural_{\mathrm{fa},k}}$ in the proportion sign and expand the binomial coefficient what immediately cancels the term $(|Z_k|-j)!$ and using the identity $(x)_{-n\uparrow} = \frac{1}{(x-n)_{n\uparrow}}$ we write

$$\propto \sum_{j=0}^{|Z_k|}\frac{(\alpha^\natural_{k|k-1})_{j\uparrow}}{(\alpha^\natural_{k|k-1} + F_k[1] - F_k^\phi[h])^{\alpha^\natural_{k|k-1}+j}}\frac{(\alpha^\natural_{\mathrm{fa},k})_{|Z_k|-j}}{(\beta^\natural_{\mathrm{fa},k}+1)^{|Z_k|-j}}\sum_{\substack{Z\subseteq Z_k\\|Z|=j}}\prod_{z\in Z}\frac{F_k^z[h]}{c^\natural_{\mathrm{fa},k}(z)}. \tag{5.23f}$$

After substituting (5.23f) into (3.24), we cancel the multiplying constants absorbed in the development leading to (5.23f) which yields

$$\mathcal{G}^\natural_{\Phi_k}[h|Z_k] = \frac{Y_0^\natural[h](Z_k)}{Y_0^\natural[1](Z_k)}, \tag{5.24}$$

with $Y_0^\natural$ defined in (5.11) for the PPHD filter.

Following from this, we need to obtain an expression of $\mathcal{G}^\flat_{\Phi_k}[h|Z_k]$ for the PHD filter. Since a Poisson process is a limit case of a Panjer process, we could start from the PPHD result in (5.23f) and follow the development in [95]. However, we shall use the fact that a Poisson process is a specific case of an i.i.d. cluster process, and use the CPHD result in (5.20e) with the additional assumptions presented in Assumptions 5.1.1 stating that $\Phi_{k|k-1}$ is a Poisson process described by intensity $\mu^\flat_{k|k-1}$, so the cardinality pmf is obtained with

$$\rho^\flat_{k|k-1}(n) = e^{-\mu^\flat_{k|k-1}(\mathcal{X})} \frac{\mu^\flat_{k|k-1}(\mathcal{X})^n}{n!}, \tag{5.25}$$

and the false alarm process is also Poisson with spatial density $c^\flat_{\text{fa},k}$ and parameter $\lambda^\flat_{\text{fa},k}$.

$$\begin{aligned}
&\prod_{z \in Z_k} c^\flat_{\text{fa},k}(z) \sum_{n \gtrsim 0} \rho^\flat_{k|k-1}(n) \sum_{j=0}^{\min(|Z_k|,n)} \frac{n!(|Z_k|-j)!}{(n-j)!} \rho^\flat_{\text{fa},k}(|Z_k|-j) \frac{F^\phi_k[h]^{n-j}}{F_k[1]^n} \\
&\quad \cdot \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} \frac{F^z_k[h]}{c^\flat_{\text{fa},k}(z)} &(5.26a)\\
&= \prod_{z \in Z_k} c^\flat_{\text{fa},k}(z) \sum_{n \gtrsim 0} \sum_{j=0}^{\min(|Z_k|,n)} \frac{1}{(n-j)!} (\lambda^\flat_{\text{fa},k})^{|Z_k|-j} F^\phi_k[h]^{n-j} \cdot \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} \frac{F^z_k[h]}{c^\flat_{\text{fa},k}(z)} &(5.26b)\\
&= \prod_{z \in Z_k} c^\flat_{\text{fa},k}(z) \sum_{j=0}^{|Z_k|} \left(\sum_{n \gtrsim j} \frac{F^\phi_k[h]^{n-j}}{(n-j)!} \right) (\lambda^\flat_{\text{fa},k})^{|Z_k|-j} \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} \frac{F^z_k[h]}{c^\flat_{\text{fa},k}(z)} &(5.26c)\\
&= e^{F^\phi_k[h]} \prod_{z \in Z_k} c^\flat_{\text{fa},k}(z) \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} \frac{F^z_k[h]}{c^\flat_{\text{fa},k}(z)} \prod_{z \notin Z_k} \frac{\lambda^\flat_{\text{fa},k} c^\flat_{\text{fa},k}(z)}{c^\flat_{\text{fa},k}(z)} &(5.26d)\\
&= e^{F^\phi_k[h]} \sum_{\substack{Z \subseteq Z_k \\ |Z|=j}} \prod_{z \in Z} F^z_k[h] \prod_{z \notin Z} \lambda^\flat_{\text{fa},k} c^\flat_{\text{fa},k}(z) &(5.26e)\\
&= e^{F^\phi_k[h]} \prod_{z \in Z_k} (F^z_k[h] + \lambda^\flat_{\text{fa},k} c^\flat_{\text{fa},k}(z)) &(5.26f)
\end{aligned}$$

Substituting (5.26f) into (3.24) yields

$$\mathcal{G}^\flat_{\Phi_k}[h|Z_k] = \frac{Y^\flat_0[h](Z_k)}{Y^\flat_0[1](Z_k)}, \tag{5.27}$$

with $Y^\flat_0$ defined in (5.10) for the PHD filter.

Finally, the similarity of (5.21), (5.24) and (5.27) leads to the desired result when definitions in (5.13) are used for $\bullet \in \{\sharp, \natural, \flat\}$ corresponding, respectively, to the CPHD filter, the PPHD filter, or the PHD filter. □

5.2.2 Predictive point process

Next we obtain information describing the predicted process $\Phi_{k+1|k}$ in the considered PHD filters.

Proposition 5.2.4 (Intensity prediction [67, 68, 95]).
Under Assumptions 5.1.2, a probability hypothesis density function $\mu_{k+1|k}$ describing the updated object process $\Phi_{k+1|k}$ is given by

$$\mu_{k+1|k}(x) = \mu^{\mathrm{b}}_{k+1}(x) + \mu^{\mathrm{s}}_{k+1}(x) \tag{5.28}$$

where μ^{s}_{k+1} is the intensity of the process describing the surviving objects

$$\mu^{\mathrm{s}}_{k+1}(x) := \int p_{\mathrm{s}}(x')\pi(x|x')\mu_k(x')\mathrm{d}x', \tag{5.29}$$

and μ^{b}_{k+1} is the intensity of the newborn process Φ^{b}_{k+1}.

The general expression (5.28) of the predicted intensity $\mu_{k+1|k}$ is developed for the PHD filter in [68], and adopted in the PPHD filter [95] and the CPHD filter [67, 95].

Considering that information contained in ν_k of (5.15) is not maintained by the filters (i.e. it is discarded and not propagated to next time step), its expression for the predicted moment $\nu_{k+1|k}$ should be obtained additionally to usual $\mu_{k+1|k}$.

Proposition 5.2.5 (Second factorial moment density prediction).
Under Assumptions 5.1.1 and 5.1.2, a second-order factorial moment density $\nu_{k+1|k}$ describing the updated object process $\Phi_{k+1|k}$ is given by

$$\nu_{k+1|k}(x,\bar{x}) = \mu^{\mathrm{b}}_{k+1}(x)\mu^{\mathrm{s}}_{k+1}(\bar{x}) + \mu^{\mathrm{b}}_{k+1}(\bar{x})\mu^{\mathrm{s}}_{k+1}(x) + \nu^{\mathrm{b}}_{k+1}(x,\bar{x}) + \nu^{\mathrm{s}}_{k+1}(x,\bar{x}), \tag{5.30}$$

where ν^{s}_{k+1} is the second-order factorial moment of the process describing the surviving objects

$$\begin{aligned}\nu^{\mathrm{s}}_{k+1}(x,\bar{x}) := &\ell^{\phi}_2[1] \int p_{\mathrm{s}}(x')p_{\mathrm{s}}(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}')\mu^{\phi}_k(x')\mu^{\phi}_k(\bar{x}')\mathrm{d}(x',\bar{x}') \\ &+ 2\sum_{z\in Z_k} \ell^{z}_2[1] \int p_{\mathrm{s}}(x')p_{\mathrm{s}}(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}')\mu^{\phi}_k(x')\mu^{z}_k(\bar{x}')\mathrm{d}(x',\bar{x}')\end{aligned}$$

$$+ \sum_{z,\bar{z}\in Z_k} \ell_2^{\neq,z,\bar{z}}[1] \int p_s(x')p_s(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}')\mu_k^z(x')\mu_k^{\bar{z}}(\bar{x}')\mathrm{d}(x',\bar{x}'), \tag{5.31}$$

where the ℓ terms for time k are presented in Definition 5.1.6 such that $\ell_2^{\neq,z,\bar{z}}$ is obtained as $\ell_2^{\{z,\bar{z}\}}$ if $z \neq \bar{z}$ and zero otherwise, and ν_{k+1}^{b} is the second-order factorial moment describing the newborn objects.

The proof related to the densities of the second-order factorial moment (5.30) is given next.

Proof. Let us develop an expression of the second-order factorial moment density of the *predicted* object process. According to [104, Eq. 4.3.4], the second-order factorial moment can be obtained from the second-order *non-factorial* moment. The second-order non-factorial moment of the *predicted* process can be obtained from [95] by substituting [95, Eq. 58] and [95, Eq. 61] in [95, Eq. 53d]. Substituting this result into [104, Eq. 4.3.4] leads to (5.30), where the density of the second-order moment ν_{k+1}^{s} describing the persisting objects in (5.30) is found to be[2]

$$\nu_{k+1}^{\mathrm{s}}(x,\bar{x}) = \int \pi(x|x')\pi(\bar{x}|\bar{x}')\nu_k(x',\bar{x}')\mathrm{d}(x',\bar{x}'), \tag{5.32}$$

where ν_k is the density of the updated process. An explicit expression of density ν_k simultaneously for the PHD, the PPHD and the CPHD filters in (5.15) when substituted to (5.32) results into

$$\begin{aligned}
\nu_{k+1}^{\mathrm{s}}(x,\bar{x}) = &\int p_s(x')p_s(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}')\mu_k^{\phi}(x')\mu_k^{\phi}(\bar{x}')\ell_2^{\phi}[1]\mathrm{d}(x',\bar{x}') \\
&+ \int p_s(x')p_s(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}')\mu_k^{\phi}(x') \sum_{z\in Z_k} \mu_k^z(\bar{x}')\ell_2^z[1]\mathrm{d}(x',\bar{x}') \\
&+ \int p_s(x')p_s(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}')\mu_k^{\phi}(\bar{x}') \sum_{z\in Z_k} \mu_k^z(x')\ell_2^z[1]\mathrm{d}(x',\bar{x}') \\
&+ \int p_s(x')p_s(\bar{x}')\pi(x|x')\pi(\bar{x}|\bar{x}') \sum_{z,\bar{z}\in Z_k} \mu_k^z(x')\mu_k^{\bar{z}}(\bar{x}')\ell_2^{\neq,z,\bar{z}}[1]\mathrm{d}(x',\bar{x}').
\end{aligned} \tag{5.33}$$

After bringing the sums outside of the respective integrals in (5.33), the resulting expression can is used in (5.30) which yields the desired result. □

[2] This possibility was first outlined in [75].

Proposition 5.2.6 (P.g.fl. prediction [25, 67, 95]).
Under Assumptions 5.1.1 and 5.1.2, a p.g.fl. $\mathcal{G}_{k+1|k}$ *of the predicted object process* $\Phi_{k+1|k}$ *is given by*

$$\mathcal{G}_{k+1|k}(h|Z_k) = \mathcal{G}^{\mathrm{b}}_{k+1}(h)\ell^{\phi}_0[\mathcal{G}_{\mathrm{s}}(h)], \tag{5.34}$$

for a test function $h : \mathcal{X} \to [0, 1]$*, where the* ℓ*-term for time* k *is defined in* (5.13)*, and* $\mathcal{G}^{\mathrm{b}}_{k+1}$ *is the p.g.fl. of the newborn process* Φ^{b}_{k+1}*.*

Proof. As described in [95], the p.g.fl. of the predicted object process is expressed in the form

$$\mathcal{G}_{k+1|k}(h) = \mathcal{G}^{\mathrm{b}}_{k+1}(h)\mathcal{G}_k(\mathcal{G}_{\mathrm{s}}(h|\cdot)), \tag{5.35}$$

where the multiplicative structure is due to independence of the newborn objects and those surviving from the previous time step; and the composition appears because the survival process applies to each preexisting object described by the updated object process Φ_k from the previous time step. Substituting the p.g.fl. of the birth process as defined in Assumptions 5.1.2 and the p.g.fl of the updated process (5.16) to (5.35) yields the desired result.

□

5.3 Optimal solutions for the sum query

The rest of the chapter is concerned with constructing Bayesian solutions to a number of problems using point process information obtained from various moment-approximation filters. At arbitrary time $k > 0$ for a compatible loss function, a Bayesian solution is given by a tuple (a_k, ρ_k) of the Bayes action a_k (which is the optimal decision) and the Bayes expected loss ρ_k (which is the optimised value of quality associated with the optimal decision). We are going to develop expressions of (a_k, ρ_k) using process information inferred using PHD filters: for an updated process Φ_k and for a predicted process $\Phi_{k+1|k}$.

5.3.1 Updated point process

Now we obtain the Bayesian solution of a decision problem associated with loss function L_Σ for an updated (posterior) process Φ_k.

Theorem 5.3.1 (Posterior Bayesian solution for L_Σ).
For an updated process Φ_k obtained from a filter under Assumptions 5.1.1 and the loss function L_Σ in (4.11), the Bayesian solution $(a_{\Sigma,k}, \rho_{\Sigma,k})$ is given by

$$a_{\Sigma,k} = K(F_k^\phi[m]\ell_1^\phi[1] + \sum_{z\in Z_k} F_k^z[m]\ell_1^z[1]) + C, \tag{5.36}$$

$$\begin{aligned}\rho_{\Sigma,k} =& K^2\bigg(F_k^\phi[m^2]\ell_1^\phi[1] + \sum_{z\in Z_k} F_k^z[m^2]\ell_1^z[1] + F_k^\phi[m]^2\Big[\ell_2^\phi[1] - \ell_1^\phi[1]^2\Big]\\ &+ 2F_k^\phi[m]\sum_{z\in Z_k} F_k^z[m]\Big[\ell_2^z[1] - \ell_1^\phi[1]\ell_1^z[1]\Big]\\ &+ \sum_{z,\bar{z}\in Z_k} F_k^z[m]F_k^{\bar{z}}[m]\Big[\ell_2^{\neq,z,\bar{z}}[1] - \ell_1^z[1]\ell_1^{\bar{z}}[1]\Big]\bigg),\end{aligned} \tag{5.37}$$

where the ℓ terms for time k are presented in Definition 5.1.6 such that $\ell_2^{\neq,z,\bar{z}}$ is obtained as $\ell_2^{\{z,\bar{z}\}}$ if $z \neq \bar{z}$ and zero otherwise.

For the classic PHD filter, expression analogous to (5.36) was first presented in [69], and expression analogous to (5.37) was first presented in [76]. Expressions (5.36) and (5.37) are novel results when considered for the PPHD and the CPHD filters. The proof is given next.

Proof. Let us obtain expressions of the Bayesian solution $(a_{\Sigma,k}, \rho_{\Sigma,k})$ for the updated object process Φ_k and loss L_Σ. This is done by substituting the process statistics (μ_k, ν_k) exposed in Propositions 5.2.1 and 5.2.2 to the Bayesian solution given in Theorem 4.2.4.

We now obtain an expression for the Bayes action $a_{\Sigma,k}$. Substituting (5.14) to (4.14) gives

$$a_{\Sigma,k} = K\bigg(\int m(x)\mu_k^\phi(x)\ell_1^\phi[1]\mathrm{d}x + \int m(x)\sum_{z\in Z_k}\mu_k^z(x)\ell_1^z[1]\mathrm{d}x\bigg) + C, \tag{5.38}$$

this, after bringing the sum outside of the integral, yields the desired result in (5.36) when notations F^ϕ and F^z given respectively in (5.4) and (5.5) are used.

Next we focus on the Bayes expected loss $\rho_{\Sigma,k}$ for the updated process Φ_k and loss L_Σ. Substituting (5.14) and (5.15) to (4.15) yields

$$\begin{aligned}\rho_{\Sigma,k} =& K^2\bigg(\int m(x)^2\mu_k^\phi(x)\ell_1^\phi[1_\mathcal{X}]\mathrm{d}x + \int m(x)^2\sum_{z\in Z_k}\mu_k^z(x)\ell_1^z[1]\mathrm{d}x\\ &+ \int m(x)m(\bar{x})\mu_k^\phi(x)\mu_k^\phi(\bar{x})\ell_2^\phi[1]\mathrm{d}(x,\bar{x})\end{aligned}$$

$$+ 2\int m(x)m(\bar{x})\mu_k^{\phi}(x)\sum_{z\in Z_k}\mu_k^z(\bar{x})\ell_2^z[1]\mathrm{d}(x,\bar{x})$$
$$+ \int m(x)m(\bar{x})\sum_{z,\bar{z}\in Z_k}\mu_k^z(x)\mu_k^{\bar{z}}(\bar{x})\ell_2^{\neq,z,\bar{z}}[1]\mathrm{d}(x,\bar{x})$$
$$- \int m(x)m(\bar{x})\mu_k^{\phi}(x)\mu_k^{\phi}(\bar{x})\ell_1^{\phi}[1]^2\mathrm{d}(x,\bar{x})$$
$$- 2\int m(x)m(\bar{x})\mu_k^{\phi}(x)\ell_1^{\phi}[1]\sum_{z\in Z_k}\mu_k^z(\bar{x})\ell_1^z[1]\mathrm{d}(x,\bar{x})$$
$$- \int m(x)m(\bar{x})\sum_{z,\bar{z}\in Z_k}\mu_k^z(x)\ell_1^z[1]\mu_k^{\bar{z}}(\bar{x})\ell_1^{\bar{z}}[1]\mathrm{d}(x,\bar{x})\Bigg). \tag{5.39}$$

Substituting (5.4) and (5.5) into (5.39) yields the desired result in (5.37). □

5.3.2 Predictive point process

The Bayesian solution of a decision problem associated with loss function L_Σ for a predicted process $\Phi_{k+1|k}$ is presented next.

Corollary 5.3.2 (Predictive Bayesian solution for L_Σ)**.**
For a predicted process $\Phi_{k+1|k}$ obtained from a filter under Assumptions 5.1.1 and 5.1.2 and the loss function L_Σ in (4.11)*, the Bayesian solution* $(a_{\Sigma,k+1|k}, \rho_{\Sigma,k+1|k})$ *is given by*

$$a_{\Sigma,k+1|k} = a^{\mathrm{s}}_{\Sigma,k+1} + a^{\mathrm{b}}_{\Sigma,k+1} - C, \tag{5.40}$$
$$\rho_{\Sigma,k+1|k} = \rho^{\mathrm{s}}_{\Sigma,k+1} + \rho^{\mathrm{b}}_{\Sigma,k+1}, \tag{5.41}$$

where $(a^{\mathrm{s}}_{\Sigma,k+1}, \rho^{\mathrm{s}}_{\Sigma,k+1})$ *is a solution corresponding to the persisting objects that is obtained from the updated process Φ_k using $\mathcal{G}_s^1$ in* (5.9) *so that*

$$a^{\mathrm{s}}_{\Sigma,k+1} = K\Bigg(F_k^{\phi}[\mathcal{G}_s^1(m)]\ell_1^{\phi}[1] + \sum_{z\in Z_k}F_k^z[\mathcal{G}_s^1(m)]\ell_1^z[1]\Bigg) + C, \tag{5.42}$$
$$\rho^{\mathrm{s}}_{\Sigma,k+1} = K^2\Bigg(F_k^{\phi}[\mathcal{G}_s^1(m)^2]\ell_1^{\phi}[1] + \sum_{z\in Z_k}F_k^z[\mathcal{G}_s^1(m)^2]\ell_1^z[1] + F_k^{\phi}[\mathcal{G}_s^1(m)]^2\Big[\ell_2^{\phi}[1] - \ell_1^{\phi}[1]^2\Big]$$
$$+ 2F_k^{\phi}[\mathcal{G}_s^1(m)]\sum_{z\in Z_k}F_k^z[\mathcal{G}_s^1(m)]\Big[\ell_2^z[1] - \ell_1^{\phi}[1]\ell_1^z[1]\Big]$$
$$+ \sum_{z,\bar{z}\in Z_k}F_k^z[\mathcal{G}_s^1(m)]F^{\bar{z}}[\mathcal{G}_s^1(m)]\Big[\ell_2^{\neq,z,\bar{z}}[1] - \ell_1^z[1]\ell_1^{\bar{z}}[1]\Big]\Bigg), \tag{5.43}$$

where the ℓ terms for time k are presented in Definition 5.1.6 such that $\ell_2^{\neq,z,\bar{z}}$ is *obtained as* $\ell_2^{\{z,\bar{z}\}}$ *if* $z \neq \bar{z}$ *and zero otherwise, and* $(a^{\mathrm{b}}_{\Sigma,k+1}, \rho^{\mathrm{b}}_{\Sigma,k+1})$ *is a solution corresponding to the newborn objects*

$$a^{\mathrm{b}}_{\Sigma,k+1} = KF^{\mathrm{b}}_{k+1}[m] + C; \tag{5.44}$$
$$\rho^{\mathrm{b}}_{\Sigma,k+1} = K^2\big(F^{\mathrm{b}}_{k+1}[m^2] - F^{\mathrm{b}}_{k+1}[m]^2 + Q^{\mathrm{b}}_{k+1}[m]\big), \tag{5.45}$$

where for time $k+1$ *the* F*-term is defined in* (5.6)*, and the* Q*-term is defined in* (5.7)*.*

Proof. The proof is analogous to the one of Theorem 5.3.1 and obtained by substituting process statistics $(\mu_{k+1|k}, \nu_{k+1|k})$ presented in (5.28) and (5.30), respectively, to (4.14) to obtain the Bayes action and to (4.15) to obtain the Bayes expected loss. The terms corresponding to the birth process in (5.28) and (5.30) are then simply expressed using (5.6) and (5.7).

□

5.4 Optimal solutions for the product query

5.4.1 Updated point process

Now we obtain the Bayesian solution of a decision problem associated with loss function L_Π for an updated (posterior) process Φ_k.

Theorem 5.4.1 (Posterior Bayesian solution for L_Π)**.**
For an updated process Φ_k *obtained from a filter under Assumptions 5.1.1 and the loss function* L_Π *in* (4.30)*, the Bayesian solution* $(a_{\Pi,k}, \rho_{\Pi,k})$ *is given by*

$$a_{\Pi,k} = K\ell_0^\phi[m] + C, \tag{5.46}$$
$$\rho_{\Pi,k} = K^2\big(\ell_0^\phi[m^2] - \ell_0^\phi[m]^2\big), \tag{5.47}$$

where the ℓ *terms for time* k *are presented in Definition 5.1.6.*

Proof. Definition of a p.g.fl. in (5.16) when substituted to (4.31) yields an expression of the Bayes action, and when substituted to (4.32) yields an expression of the Bayes expected loss. □

5.4.2 Predictive point process

The Bayesian solution of a decision problem associated with loss function L_Π for a predicted process $\Phi_{k+1|k}$ is presented next.

Corollary 5.4.2 (Predictive Bayesian solution for L_Π).
For a predicted process $\Phi_{k+1|k}$ obtained from a filter under Assumptions 5.1.1 and 5.1.2 and the loss function L_Π in in (4.30)*, the Bayesian solution* $(a_{\Pi,k+1|k}, \rho_{\Pi,k+1|k})$ *is given by*

$$a_{\Pi,k+1|k} = K\mathcal{G}^{\mathrm{b}}_{k+1}(m)\ell_0^{\phi}[\mathcal{G}_{\mathrm{s}}(m)] + C, \tag{5.48}$$

$$\rho_{\Pi,k+1|k} = K^2\big(\mathcal{G}^{\mathrm{b}}_{k+1}(m^2)\ell_0^{\phi}[\mathcal{G}_{\mathrm{s}}(m)^2] - \mathcal{G}^{\mathrm{b}}_{k+1}(m)^2\ell_0^{\phi}[\mathcal{G}_{\mathrm{s}}(m)]^2\big), \tag{5.49}$$

where the ℓ terms for time k are presented in Definition 5.1.6.

Proof. The proof simply follows from Proposition 4.3.8 that presents the solution for superimposed point processes. □

5.5 Summary

In this chapter we aimed to implement certain optimal Bayesian solutions developed in Chapter 4 using information from filters for spatial point processes, namely, the classical Probability Hypothesis Density (PHD) filter, the Panjer PHD (PPHD) filter, and the Cardinalized PHD (CPHD) filter. However, we found that these filters are not sufficiently 'informative' in their standard implementation to construct the sought expressions directly. Nevertheless, it was shown that the missing quantities describing the object process in a considered filter can be extracted from the data update step using tools of the point process theory. We were able to extract quantities describing updated and predicted object processes; specifically, the densities of lower-order moments and p.g.fl.s. Ultimately, due to the recursive nature of the algorithms we produced two sets of solutions: one for the moment of time right after the available evidence has been incorporated, and another for the predicted moment of time, before any new evidence becomes available.

Chapter 6

Simulated decision making with the SMC-PHD filter

This chapter simulates probabilistic decision making with Probability Hypothesis Density (PHD) filters. We are going to implement the approach developed in Chapter 5 and compare its performance to that of conventional approach. It will be shown that the developed algorithms, on top of their theoretical soundness, equip optimal decisions with indicators of their quality and have a capacity to provide improvements in performance. The demonstration will be primarily focused on the the classical PHD filter, and results for the CPHD filter will be provided for comparison.

The content of this chapter as follows. In Section 6.1 we present the Sequential Monte Carlo (SMC) PHD filter as a mean to describe uncertainty in the environment.[1]. In Section 6.2 we present in parallel various decision making approaches: the one operating under certainty in Subsection 6.2.1; and two probabilistic algorithms: one operating under assumed certainty equivalence in Subsection 6.2.2 and one operating under uncertainty in Subsection 6.2.3. Section 6.3 assembles three practical query functions, including models of risk developed in Appendix A.3. Finally, in Section 6.4 we perform simulations to evaluate the performance of the algorithms. Section 6.5 provides summary.

[1]This means that as of now we are not going to address filter implementations using Gaussian Mixtures [107]. Furthermore, we leave out decision implementation of the PPHD and the CPHD filters despite being obtained in Chapter 5, and despite selected results obtained with the CPHD are presented in this chapter.

6.1 State uncertainty with the SMC-PHD filter

In practice the state of environment, or world, is often unknown. Instead, the decision-maker maintains a belief about uncertain aspects of environment inferred from collected observations. A possibility is that the belief is maintained and recursively updated by a Bayesian filter.

In this thesis chapter we are focused on the PHD filter, and its Sequential Monte Carlo (SMC) implementation presented below. SMC implementation is particularly suitable for handling non-linear phenomena, which comes at considerable computational costs. We offer a revised presentation from [108], which omits the step of 'state extraction' (it is interpreted as an integral part of the conventional decision making algorithm described in Subsection 6.2.2).

The filter operates in the circumstances where the number of objects is unknown and time-varying. Recursive equations of the PHD filter are given in Proposition 3.4.1 and restated here for convenience

$$\mu_{k|k-1}(x) = \mu_k^b(x) + \int p_{\mathrm{s}}(x')\pi(x|x')\mu_{k|k-1}(x')\mathrm{d}x', \tag{6.1}$$

$$\mu_k(x) = \mu_{k|k-1}^{\phi}(x) + \sum_{z\in Z_k} \frac{\mu_{k|k-1}^{z}(x)}{\int \mu_{k|k-1}^{z}(x)\mathrm{d}x + \kappa_{\mathrm{fa},k}(z)}. \tag{6.2}$$

The objects are described on the population state space $\mathfrak{X} = \bigcup_{n\geqslant 0} \mathcal{X}^n$ with individual states on $\mathcal{X}$. Here $\mathcal{X} \subset \mathbb{R}^{d_x}$ denotes the d_x-dimensional state space describing the state of an individual object. Specifically, the state of an object is represented by a vector

$$x = [\mathrm{x}, \dot{\mathrm{x}}, \mathrm{y}, \dot{\mathrm{y}}]^T, \tag{6.3}$$

where position of the object is denoted by $[\mathrm{x}, \mathrm{y}]^T$ and velocity is $[\dot{\mathrm{x}}, \dot{\mathrm{y}}]^T$. Movement of an individual object is modelled by a nearly constant velocity (CV) model

$$\pi(x|x') = \mathcal{N}(x; Fx' + \mu_w, \Sigma_w), \tag{6.4}$$

with additive zero-mean white Gaussian process noise described by its mean and covariance matrix

$$\mu_w = \mathbb{O}_{4\times 1},\ \Sigma_w = \mathbb{I}_2 \otimes \begin{bmatrix} \Delta t^3/3 & \Delta t^2/2 \\ \Delta t^2/2 & \Delta t \end{bmatrix} \sigma_w^2, \tag{6.5}$$

where σ_w is the standard deviation of velocity increments. The transition matrix F

is given by

$$F = \mathbb{I}_2 \otimes \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \tag{6.6}$$

where $\mathbb{I}_2$ is an identity matrix with dimensions 2×2, $\otimes$ is the Kronecker product, and Δt is the sampling interval.

The objects arrive spontaneously according to a Poisson point process with rate λ_{b} and intensity μ^{b} given by

$$\mu^{\mathrm{b}}(x) = \lambda_{\mathrm{b}} \mathcal{N}(x; x_{\mathrm{b}}, \Sigma_{\mathrm{b}}), \tag{6.7}$$

with Gaussian spatial density described by its mean x_{b} and covariance Σ_{b}:

$$x_{\mathrm{b}} = \begin{bmatrix} \mathrm{x_b} \\ \dot{\mathrm{x}}_\mathrm{b} \\ \mathrm{y_b} \\ \dot{\mathrm{y}}_\mathrm{b} \end{bmatrix}, \; \Sigma_{\mathrm{b}} = \begin{bmatrix} \sigma_{\mathrm{x_b x_b}} & 0 & 0 & 0 \\ 0 & \sigma_{\dot{\mathrm{x}}_\mathrm{b}\dot{\mathrm{x}}_\mathrm{b}} & 0 & 0 \\ 0 & 0 & \sigma_{\mathrm{y_b y_b}} & 0 \\ 0 & 0 & 0 & \sigma_{\dot{\mathrm{y}}_\mathrm{b}\dot{\mathrm{y}}_\mathrm{b}} \end{bmatrix}, \tag{6.8}$$

The probability of survival of an individual object is given by p_s.

Information about objects is collected by a single radar-like sensor in state

$$x_s = [\mathrm{x}_s, \dot{\mathrm{x}}_s, \mathrm{y}_s, \dot{\mathrm{y}}_s]^T \tag{6.9}$$

which measures range and bearing of individual objects. The measurement likelihood is given by

$$g(z|x) = \mathcal{N}(z; h(x) + \mu_v, \Sigma_v), \tag{6.10}$$

with measurement vector

$$h(x) = \begin{bmatrix} r(x, x_s) \\ b(x, x_s) \end{bmatrix}, \tag{6.11}$$

where range and bearing measurements for an object in state x are modelled as

$$r(x, x_s) = \sqrt{(\mathrm{x} - \mathrm{x}_s)^2 + (\mathrm{y} - \mathrm{y}_s)^2}, \tag{6.12}$$

$$b(x, x_s) = \operatorname{atan2}\left(\frac{\mathrm{x} - \mathrm{x}_s}{\mathrm{y} - \mathrm{y}_s}\right), \tag{6.13}$$

with additive zero-mean white Gaussian noise described by its mean μ_v the covariance matrix Σ_v:

$$\mu_\nu = \mathbb{O}_{2\times 1},\ \Sigma_v = \begin{bmatrix} \sigma_r^2 & \sigma_r\sigma_b \\ \sigma_b\sigma_r & \sigma_b^2 \end{bmatrix}, \tag{6.14}$$

where σ_r and σ_b are standard deviations in range and bearing, respectively. These measurements are immersed in false alarms described with intensity $\kappa_{\text{fa}}(z) = \lambda_{\text{fa}} c_{\text{fa}}(z)$.

SMC-PHD implementation of the time update step of Equation (6.1) is presented in Algorithm 6.1, and of the data update step in (6.2) is presented in Algorithm 6.2. Implementation is focused on the propagation of a limited number of particles describing the filtering intensity, where at time step k, $N_{b,k}$ is the number of birth particles, and N_{k-1} is the number of particles used to describe the intensity of the posterior process. In the simulation, the number of birth particles will be set to 1000 per expected target (and overall no less than 1000); the number of particles to describe the posterior intensity is set to 1000 per target.

Algorithm 6.1: Prediction step (time update)

Input:
Posterior intensity: $\{w_{k-1}^{(i)}, x_{k-1}^{(i)}\}_{i=1}^{N_{k-1}}$
Newborn intensity: $\{w_{b,k}^{(i)}, x_{b,k}^{(i)}\}_{i=1}^{N_{b,k}}$

Survival process
for $1 \leqslant i \leqslant N_{k-1}$ **do**
  $w_{k|k-1}^{(i)} \leftarrow p_s(x_{k-1}^{(i)}) w_{k-1}^{(i)}$ `// Update particle weights`
  $v_{k-1}^{(i)} \leftarrow \mathcal{N}(\mu_w, \Sigma_w)$
  $x_{k|k-1}^{(i)} \leftarrow F x_{k-1}^{(i)} + v_{k-1}^{(i)}$ `// Propagate particles`
end
Newborn process
for $1 \leqslant j \leqslant N_{b,k}$ **do**
  $x_{b,k}^{(j)} \leftarrow \mathcal{N}(x_{\text{b}}, \Sigma_{\text{b}})$
  $w_{b,k}^{(j)} = \lambda_{\text{b}} / N_{b,k}$
  $\{w_{k|k-1}^{(N_{k-1}+j)}, x_{k|k-1}^{(N_{k-1}+j)}\} \leftarrow \{w_{b,k}^{(j)}, x_{b,k}^{(j)}\}$ `// Append birth particle`
end
$N_{k|k-1} \leftarrow N_{k-1} + N_{b,k}$

Output:
Predicted intensity: $\{w_{k|k-1}^{(i)}, x_{k|k-1}^{(i)}\}_{i=1}^{N_{k|k-1}}$

Algorithm 6.2: Update step (data update) with extra outputs

Input:
Predicted intensity: $\{w^{(i)}_{k|k-1}, x^{(i)}_{k|k-1}\}_{i=1}^{N_{k|k-1}}$
Current measurements: Z_k

Missed detection and measurement components
for $1 \leqslant i \leqslant N_{k|k-1}$ **do**
    $w^{(i),\phi}_{k|k-1} \leftarrow (1 - p_{\mathrm{d},k}(x^{(i)}_{k|k-1}))w^{(i)}_{k|k-1}$ `// Missed detection components`
    for $z \in Z_k$ **do**
        $w^{(i),z}_{k|k-1} \leftarrow g(z|x^{(i)}_{k|k-1})p_{\mathrm{d},k}(x^{(i)}_{k|k-1})w^{(i)}_{k|k-1}$ `// Measurement components`
    end
end
Data update
for $1 \leqslant i \leqslant N_{k|k-1}$ **do**
    for $z \in Z_k$ **do**
        $\bar{w}^{(i),z}_{k|k-1} \leftarrow \frac{w^{(i),z}_{k|k-1}}{\sum_{1 \leq i' \leq N_{k|k-1}} w^{(i'),z}_{k|k-1} + \kappa_{\mathrm{fa},k}(z)}$ `// Normalize contributions`
    end
    $w^{(i)}_k \leftarrow w^{(i),\phi}_{k|k-1} + \sum_{z \in Z} \bar{w}^{(i),z}_{k|k-1}$ `// Update particle weights`
    $x^{(i)}_k \leftarrow x^{(i)}_{k|k-1}$
    $N_k \leftarrow N_{k|k-1}$
end

Output:
Updated intensity: $\{w^{(i)}_k, x^{(i)}_k\}_{i=1}^{N_k}$
Missed detection component: $\{w^{(i),\phi}_{k|k-1}, x^{(i)}_{k|k-1}\}_{i=1}^{N_{k|k-1}}$
Measurements components: $\{\{w^{(i),z}_{k|k-1}, x^{(i)}_{k|k-1}\}_{i=1}^{N_{k|k-1}}\}_{z \in Z_k}$

6.2 Algorithmic implementations

In this section we are going to expose specific decision-making algorithms that will be later used for simulations. Conceptually, these algorithms correspond to three different decision making strategies first outlined in Section 1.1:

- decision making under uncertainty (UC);
- decision making under assumed certainty equivalence (CE);
- decision making under certainty (UU).

The idea behind is that decisions produced under certainty are required to establish how good are the decisions produced by the algorithms operating in non-deterministic conditions. At the same time, decisions produced under assumed certainty equivalence (which is the state of the art strategy), are required in order to establish in what way decision making under uncertainty advances the state of the art.

In the following we will demonstrate how these strategies are to be implemented *for the update step of the SMC-PHD filter.*

6.2.1 Decision making under certainty (UC)

Decision making under certainty is characterised by the fact the state of environment is directly accessible by the decision-maker and that each action "is known to lead invariably to a specific outcome" [66, p. 13]. Such conditions are covered by the minimum loss principle stated in Proposition 1.1.4. Accordingly, this is simply an optimisation problem for the considered overall loss function.

Although eventually we are going to be focused on specific loss functions defined in Chapter 4, the loss considered in this subsection is the square error-in-answer loss (4.1), which permits any arbitrary real-valued query (beyond the sum (4.10) and the product (4.29) query functions). As a consequence, the solution is unnecessarily more general, but the presentation is more compact.

Theorem 6.2.1 (UC solution for an arbitrary real-valued query)**.** *When in environment $\varphi_k \in \mathfrak{X}$, a decision-maker with squared error-in-answer loss* (4.1) *following the ML principle in Proposition 1.1.4 reports* $(a^{UC}_{\mathbb{R},k}, \rho^{UC}_{\mathbb{R},k})$ *where answer is*

$$a^{UC}_{\mathbb{R},k} = q_{\mathbb{R}}(\varphi_k), \tag{6.15}$$

that corresponds to the optimised loss value

$$\begin{aligned} \rho^{UC}_{\mathbb{R},k} &= L_{\text{sq}}(a^{UC}_{\mathbb{R},k}, \varphi_k) && (6.16)\\ &= 0. && (6.17) \end{aligned}$$

Proof. Let us consider a function f coinciding with the loss (4.1) and given by

$$\begin{aligned} f(a) &= L_{\text{sq}}(a, \varphi_k) && (6.18a)\\ &= (a - q_{\mathbb{R}}(\varphi_k))^2. && (6.18b) \end{aligned}$$

An extremum of the function f in point a is found as $f'(a) = 0$. The first derivative is found to be

$$f'(a) = 2(a - q_{\mathbb{R}}(\varphi_k)), \tag{6.19}$$

which gives us the extremum for

$$a = q_{\mathbb{R}}(\varphi_k). \tag{6.20}$$

This point is a minimum if $f''(a) > 0$, and this is true since $f''(a) = 2$. □

The clairvoyant answers $(a^{UC}_{\mathbb{R},k}, 0)$ can be obtained at each time step using Algorithm 6.3 thanks to the ground truth state available in the simulation.

Algorithm 6.3: UC solution for an arbitrary real-valued query function

Input:
Ground truth state: φ_k
Compute the proposed answer
$a^{UC}_{\mathbb{R},k} \leftarrow q_{\mathbb{R}}(\varphi_k)$ `// Query function`
$\rho^{UC}_{\mathbb{R},k} \leftarrow 0$ `// Overall loss`
Output:
Proposed answer: $a^{UC}_{\mathbb{R},k}$ `// Clairvoyant answer`
Quality indicator: $\rho^{UC}_{\mathbb{R},k} = 0$ `// Actual loss`

Finally, it could be argued that there was no need to address the optimisation problem, and one could instead simply use a query function to generate ideal answers. However, this approach would not take any additional information that might be present in the query loss function into account. For example, the query loss function carries information about the minimum attainable loss, which may differ from the commonly set value of 0 for an arbitrary query loss. Furthermore, such presentation reinforces the fact that the same overall loss function is used across different decision making strategies.

6.2.2 Decision making under assumed certainty equivalence (CE)

Decision making under assumed certainty equivalence is a sub-optimal that aims to replicate the simplicity of decision making under certainty, while operating in the conditions of uncertainty. Effectively, it is a combination of two algorithm: an algorithm that produces a summary of the uncertain state of the world, and an algorithm for decision making under certainty as in Proposition (1.1.4). Note that this is a conventional strategy in the context of Bayesian filtering.

How does one produces a state summary with the SMC-PHD filter? As discussed in Chapter 3, there exists no tractable procedure to produce an optimal state estimate. Instead, heuristics, such as k-means, and expectation-maximization (EM) are employed to extract the state from the updated intensity. In this implementation we use the k-means algorithm, as it appears to provide more accurate estimates when used for state extraction in the PHD filter [22]. A variation of k-means clustering algorithm is presented in Algorithm 6.4, whereas the final implementation is after [108].

Theorem 6.2.2 (CE solution for an arbitrary real-valued query)**.** *In uncertain environment described in the PHD filter by the point process Φ_k, a decision-maker with the loss function L_{sq} (4.1) following Proposition 1.1.7, summarizes Φ_k by $\tilde{\varphi}_k$ (Algorithm 6.4) and reports the answer $(a^{CE}_{\mathbb{R},k}, \rho^{CE}_{\mathbb{R},k})$ as prescribed by UC solution* (6.15) *with*

$$a^{CE}_{\mathbb{R},k} = q_{\mathbb{R}}(\tilde{\varphi}_k), \tag{6.21}$$

that corresponds to the optimized loss value

$$\begin{aligned} \rho^{CE}_{\mathbb{R},k} &= L_{\text{sq}}(a^{CE}_{\mathbb{R},k}, \tilde{\varphi}_k) && \text{(6.22a)} \\ &= 0. && \text{(6.22b)} \end{aligned}$$

Note that the answers produced by this algorithm are associated with loss values 0, which is only true when the certainty equivalence holds.

A pseudocode for the solution in Theorem 6.2.2 is given in Algorithm 6.5.

Algorithm 6.4: k-means clustering (adapted from [22])

Input:
Approximate objects number: k
Particle states: $\{x_k^{(i)}\}_{i=1}^{N_k}$

Step 0. (Initialisation)
$j \leftarrow 1$
for $1 \leqslant n \leqslant \mathsf{k}$ **do**
  $q \leftarrow \mathcal{U}\{1, N_k\}$ // Draw a random number
  $m_{k,n}^{(j)} \leftarrow x_k^{(q)}$
end
$\Delta = \epsilon$
while $\Delta \geqslant \epsilon$ **do**
  $j \leftarrow j + 1$
  for $1 \leqslant n \leqslant \mathsf{k}$ **do**
    Step 1. (Partition)
    $P_{k,n}^{(j)} = \{x \in \{x_k^{(i)}\}_{i=1}^{N_k} : \left\| x - m_{k,n}^{(j-1)} \right\|^2 \leqslant \left\| x - m_{k,n'}^{(j-1)} \right\|^2 \ \forall n',\ 1 \leqslant n' \leqslant \mathsf{k}\}$
    Step 2. (Recalculate centres)
    $m_{k,n}^{(j)} = \frac{1}{\left|P_{k,n}^{(j)}\right|} \sum_{x \in P_{k,n}^{(j)}} x$
  end
  $\Delta \leftarrow \left| \sum_{i=1}^{N_k} \sum_{n=1}^{\mathsf{k}} \left\| x_k^{(i)} - m_{k,n}^{(j)} \right\| - \sum_{i=1}^{N_k} \sum_{n=1}^{\mathsf{k}} \left\| x_k^{(i)} - m_{k,n}^{(j-1)} \right\| \right|$
end
Step 3. (Calculate covariances of partitions)
for $1 \leqslant n \leqslant \mathsf{k}$ **do**
  $S_{k,n} \leftarrow \mathrm{cov}(P_{k,n}^{(j)})$
end
$\tilde{\varphi}_k \leftarrow \{m_{k,n}^{(j)}\}_{n=1}^{\mathsf{k}}$

Output:
Means and covariances: $\{(m_{k,n}^{(j)}, S_{k,n})\}_{n=1}^{\mathsf{k}}$
State summary: $\tilde{\varphi}_k$

6.2.3 Decision making under uncertainty (UU)

Decision making under uncertainty is a Bayes-optimal approach that faithfully implements the principle of minimum expected loss in Proposition 1.2.2. Its usual complexity is partially alleviated in solutions obtained in Chapters 4 and 5 due to the favourable properties of the employed squared error query loss, and just a limited set of real-valued query functions.

To simplify the presentation of solutions corresponding to the PHD filter, we restate the F-terms introduced in Subsection 5.1.2 defined for any $z \in Z_k$ as

$$F_k^z[m] = \int m(x)\mu_{k|k-1}^z(x)\mathrm{d}x, \tag{6.23}$$

$$F_k^\phi[m] = \int m(x)\mu_{k|k-1}^\phi(x)\mathrm{d}x. \tag{6.24}$$

Theorem 6.2.3 (UU solution for the sum query)**.** *In uncertain environment described by the PHD filter with point process Φ_k, a decision-maker with the loss function L_{sq} (4.1) and the sum query q_Σ (4.10) reports $(a_{\Sigma,k}^{UU}, \rho_{\Sigma,k}^{UU})$ answer*

$$a_{\Sigma,k}^{UU} = K\left[F_k^\phi[m_k] + \sum_{z\in Z_k}\frac{F_k^z[m_k]}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)}\right] + C, \tag{6.25}$$

that corresponds to the optimized expected loss value

$$\rho_{\Sigma,k}^{UU} = K^2\left[F_k^\phi[m_k^2] + \sum_{z\in Z_k}\left[\frac{F_k^z[m_k^2]}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)} - \left(\frac{F_k^z[m_k]}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)}\right)^2\right]\right]. \tag{6.26}$$

Proof. The result follows from Theorem 5.3.1 under Assumptions 5.1.1 for the PHD filter. □

Theorem 6.2.4 (UU solution for the product query)**.** *In uncertain environment described by the PHD filter with point process Φ_k, a decision-maker with the loss function L_{sq} (4.1) and the product query q_Π (4.29) reports $(a_{\Pi,k}^{UU}, \rho_{\Pi,k}^{UU})$ answer*

$$a_{\Pi,k}^{UU} = K\frac{e^{F_k^\phi[m_k]}}{e^{F_k^\phi[1]}}\prod_{z\in Z_k}\frac{F_k^z[m_k] + \kappa_{\mathrm{fa},k}(z)}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)} + C, \tag{6.27}$$

that corresponds to the optimized expected loss value

$$\rho_{\Pi,k}^{UU} = K^2\left[\frac{e^{F_k^\phi[m_k^2]}}{e^{F_k^\phi[1]}}\prod_{z\in Z_k}\frac{F_k^z[m_k^2] + \kappa_{\mathrm{fa},k}(z)}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)} - \left(\frac{e^{F_k^\phi[m_k]}}{e^{F_k^\phi[1]}}\prod_{z\in Z_k}\frac{F_k^z[m_k] + \kappa_{\mathrm{fa},k}(z)}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)}\right)^2\right]. \tag{6.28}$$

Proof. The result follows from Theorem 5.4.1 under Assumptions 5.1.1 for the PHD filter. □

Pseudocode implementing solutions in Theorems 6.2.3 and 6.2.4 are given respectively in Algorithms 6.6 and 6.7. Note that implementations do not rely on the updated intensity in the filter (as in k-means clustering of the assumed certainty equivalent approach), but on the extra terms extracted from the update step in Algorithm 6.2.

6.3 Practical query functions: total cardinality, sigma-risk, and pi-risk

Algorithmic solutions outlined in Section 6.2 offer a possibility of using the SMC-PHD filter in probabilistic decision making for specific real-valued query functions. When the query loss is modelled by a square error function, the optimal answer is given by the mean value of the query answer, and the value of associated expected loss is the variance around this mean. To be truly useful, these query functions should communicate specific *practical* queries that are found in the operation of real-life systems observing object populations.

The first query function is related to counting the total number of objects in the whole region. In principle, this is equivalent to setting $B = \mathcal{X}$ in Definition 3.5.4. It is a straightforward function, and presented here for the sake of completeness.

Definition 6.3.1 (Total cardinality)**.** *Total cardinality is a degenerate case of the sum query q_Σ in* (4.10) *with $C = 0$, $K = 1$, and $m(\cdot) = \mathbb{1}_\mathcal{X}(\cdot)$, and is written as*

$$q_{\Sigma_\mathcal{X}}(\varphi) := \sum_{x \in \varphi} \mathbb{1}_\mathcal{X}(x) \tag{6.29}$$

where $\mathbb{1}_\mathcal{X} : \mathcal{X} \to \{0, 1\}$ is an indicator function defined over a single object state space $\mathcal{X}$.

Other considered query functions emerge from the defence context and are focused on a model of operational risk. Specifically, in Appendix A.3 we develop custom models of risk as perceived by an asset in state x_A from a group of threatening objects. The object states φ are pulled into the model as detonation points for some weapons capable of producing impact on a distance. For each object, the probability that the object will damage the asset is modelled by a threat function $\tau : \mathcal{X} \times \mathcal{X} \to [0, 1]$. Specifically, we are going to focus on the Gaussian functions, as presented in Appendix A.2. Furthermore, assumed is the worst-case situation when weapons detonate simultaneously.

The risk model is constructed such that each weapon is ascribed with a damaging capacity of subtracting a certain fixed value $d > 0$ from the asset's total value

Algorithm 6.5: CE solution for an arbitrary real-valued query

Input:
Updated intensity: $\{w_k^{(i)}, x_k^{(i)}\}_{i=1}^{N_k}$
Obtain approximate object number
$\mathsf{k} = \left\lfloor \sum_{1 \leqslant i \leqslant N_k} w_k^{(i)} \right\rfloor$
Extract the state summary $\tilde{\varphi}_k$ **from** k **and** $\{x_k^{(i)}\}_{i=1}^{N_k}$ **with** k-means **clustering in Algorithm 6.4**
Compute the proposed answer
$a_{\mathbb{R},k}^{CE} \leftarrow q_{\mathbb{R}}(\tilde{\varphi}_k)$ `// Query function`
$\rho_{\mathbb{R},k}^{CE} \leftarrow 0$ `// Overall loss`
Output:
Proposed answer: $a_{\mathbb{R},k}^{CE}$ `// Sub-optimal`
Quality indicator: $\rho_{\mathbb{R},k}^{CE} = 0$ `// Assumed loss`

Algorithm 6.6: UU solution for the sum query

Input:
Missed detection components: $\{w_{k|k-1}^{(i),\phi}, x_{k|k-1}^{(i)}\}_{i=1}^{N_{k|k-1}}$
Measurement components: $\{\{w_{k|k-1}^{(i),z}, x_{k|k-1}^{(i)}\}_{i=1}^{N_{k|k-1}}\}_{z \in Z_k}$
F-terms for missed detections and measurements
$F_k^{\phi}[m_k] \leftarrow \sum_{1 \leqslant i \leqslant N_{k|k-1}} m_k(x_{k|k-1}^{(i)}) w_{k|k-1}^{(i),\phi}$
$F_k^{\phi}[m_k^2] \leftarrow \sum_{1 \leqslant i \leqslant N_{k|k-1}} m_k(x_{k|k-1}^{(i)})^2 w_{k|k-1}^{(i),\phi}$
for $z \in Z_k$ **do**
  $F_k^z[1] \leftarrow \sum_{1 \leqslant i \leqslant N_{k|k-1}} w_{k|k-1}^{(i),z}$
  $F_k^z[m_k] \leftarrow \sum_{1 \leqslant i \leqslant N_{k|k-1}} m_k(x_{k|k-1}^{(i)}) w_{k|k-1}^{(i),z}$
  $F_k^z[m_k^2] \leftarrow \sum_{1 \leqslant i \leqslant N_{k|k-1}} m_k(x_{k|k-1}^{(i)})^2 w_{k|k-1}^{(i),z}$
end
Compute optimal Bayesian solution
$a_{\Sigma,k}^{UU} \leftarrow K\left[F_k^{\phi}[m_k] + \sum_{z \in Z_k} \frac{F_k^z[m_k]}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)}\right] + C$
$\rho_{\Sigma,k}^{UU} \leftarrow K^2\left[F_k^{\phi}[m_k^2] + \sum_{z \in Z_k} \left[\frac{F_k^z[m_k^2]}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)} - \left(\frac{F_k^z[m_k]}{F_k^z[1] + \kappa_{\mathrm{fa},k}(z)}\right)^2\right]\right]$
Output:
Proposed answer: $a_{\Sigma,k}^{UU}$ `// Bayes-optimal`
Quality indicator: $\rho_{\Sigma,k}^{UU}$ `// Bayes expected loss`

Algorithm 6.7: UU solution for the product query

Input:

Missed detection components: $\{w^{(i),\phi}_{k|k-1}, x^{(i)}_{k|k-1}\}_{i=1}^{N_{k|k-1}}$

Measurement components: $\{\{w^{(i),z}_{k|k-1}, x^{(i)}_{k|k-1}\}_{i=1}^{N_{k|k-1}}\}_{z\in Z_k}$

F-terms for missed detections and measurements

$F^{\phi}_k[1] \leftarrow \sum_{1\leqslant i\leqslant N_{k|k-1}} w^{(i),\phi}_{k|k-1}$

$F^{\phi}_k[m_k] \leftarrow \sum_{1\leqslant i\leqslant N_{k|k-1}} m_k(x^{(i)}_{k|k-1}) w^{(i),\phi}_{k|k-1}$

$F^{\phi}_k[m_k^2] \leftarrow \sum_{1\leqslant i\leqslant N_{k|k-1}} m_k(x^{(i)}_{k|k-1})^2 w^{(i),\phi}_{k|k-1}$

for $z \in Z_k$ **do**

- $F^{z}_k[1] \leftarrow \sum_{1\leqslant i\leqslant N_{k|k-1}} w^{(i),z}_{k|k-1}$
- $F^{z}_k[m_k] \leftarrow \sum_{1\leqslant i\leqslant N_{k|k-1}} m_k(x^{(i)}_{k|k-1}) w^{(i),z}_{k|k-1}$
- $F^{z}_k[m_k^2] \leftarrow \sum_{1\leqslant i\leqslant N_{k|k-1}} m_k(x^{(i)}_{k|k-1})^2 w^{(i),z}_{k|k-1}$

end

Compute optimal Bayesian solution

$a^{UU}_{\Pi,k} \leftarrow K \frac{e^{F^{\phi}_k[m_k]}}{e^{F^{\phi}_k[1]}} \prod_{z\in Z_k} \frac{F^z_k[m_k]+\kappa_{\mathrm{fa},k}(z)}{F^z_k[1]+\kappa_{\mathrm{fa},k}(z)} + C$

$\rho^{UU}_{\Pi,k} \leftarrow K^2 \left[\frac{e^{F^{\phi}_k[m_k^2]}}{e^{F^{\phi}_k[1]}} \prod_{z\in Z_k} \frac{F^z_k[m_k^2]+\kappa_{\mathrm{fa},k}(z)}{F^z_k[1]+\kappa_{\mathrm{fa},k}(z)} - \left(\frac{e^{F^{\phi}_k[m_k]}}{e^{F^{\phi}_k[1]}} \prod_{z\in Z_k} \frac{F^z_k[m_k]+\kappa_{\mathrm{fa},k}(z)}{F^z_k[1]+\kappa_{\mathrm{fa},k}(z)} \right)^2 \right]$

Output:

Proposed answer: $a^{UU}_{\Pi,k}$ `// Bayes-optimal`

Expected loss: $\rho^{UU}_{\Pi,k}$ `// Bayes expected loss`

$V_A > 0$. The total expected damage induced by a group of weapons in state φ is the manifestation of risk, as described in Appendix A.3. Unfortunately, a general expression (A.11d) of risk for arbitrary values of V_A and d cannot be implemented using neither the sum nor the product queries. However, under certain relation between V_A and d, the general expression can be reduced to forms compatible with the query functions.

The first risk-focused query implements the risk model in (A.13) corresponding to condition $V_A \geq n \cdot d$ for any $n \in \mathbb{N}$, where it is assumed that the produced damage can never exhaust the diminishing asset value. The second risk-focused query implements the risk model in (A.15) corresponding to condition $V_A < d$, where it is assumed that a single successful hit is sufficient to completely eliminate the asset value.

Definition 6.3.2 (Sigma-risk). *Sigma-risk is a query function which is obtained from the sum query in* (4.10) *by setting* $C = 0$, $K = d$, *and* $m(\cdot) = \tau(\cdot, x_A)$, *and is written as*

$$q_{r_\Sigma}(\varphi) := d \sum_{x \in \varphi} \tau(x, x_A) \tag{6.30}$$

where d *is the damaging capacity of a weapon.*

Definition 6.3.3 (Pi-risk). *Pi-risk is a query function which is obtained from the product query in* (4.29) *by setting* $C = V_A$, $K = -V_A$, $m(\cdot) = 1 - \tau(\cdot, x_A)$, *and is written as*

$$q_{r_\Pi}(\varphi) := V_A \left[1 - \prod_{x \in \varphi} \left[1 - \tau(x, x_A) \right] \right] \tag{6.31}$$

where V_A *is the diminishing asset value.*

6.4 Simulated problems

In this section we analyse performance of the probabilistic algorithms using simulated data. We are interested to contrast performance of the proposed optimal approach (for decision making under uncertainty) to that of the conventional sub-optimal approach (for decision making under certainty equivalence). Primarily, the focus will be on the application of the PHD filter for data processing, but some results will be presented for the CPHD filter. Note, however, that implementation details for the SMC-CPHD filter are omitted from this thesis.

All analysis is based on the same underlying ground truth scenario presented in in Section 6.4.1, which involves a time-varying number of dynamic objects. This ground truth is initially used to determine clairvoyant answers for the problems introduced in Section 6.3. Then, in Section 6.4.3, it is used to simulate sensor data and perform Monte Carlo analysis.

6.4.1 Ground truth scenario

At the core of the simulation is the scenario, which is a recorded sequence of states occupied by an evolving object population over a specified number of time steps. The scenario lasts for $T = 50$ time steps with the sampling interval $\Delta t = 1\,\mathrm{s}$. In the ground truth, objects arrive at time step $[10, 20, 30, 40]$ with states sampled from a Gaussian distribution described with

$$x_{\mathrm{b}} = \begin{bmatrix} 500 \\ 0 \\ 500 \\ 0 \end{bmatrix}, \; \Sigma_{\mathrm{b}} = \begin{bmatrix} 5 \times 10^4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 5 \times 10^4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.32}$$

and velocity vectors are then corrected to be oriented towards the mean location, with absolute values of $10\mathrm{m\,s^{-1}}$. In the filter they are described as arriving according to a Poisson point process (6.7) with rate $\lambda_{\mathrm{b}} = 0.1$. Each object moves according to the linear dynamics (6.4) and (slight) additive Gaussian zero mean process noise $\sigma_w = 0.5$. Position ground truth over 50 time steps is displayed in Figure 6.1. Figure 6.2 plots the individual x and y components of each object against time.

The ground truth will enter the simulations in two distinct ways. Firstly, it will be used to generate ideal, or clairvoyant, answers by the algorithm operating under certainty that is given in Algorithm 6.3. Secondly, it will be used to generate noisy sensor observations, which are processed by the update step of the SMC-PHD filter given in Algorithm 6.2.

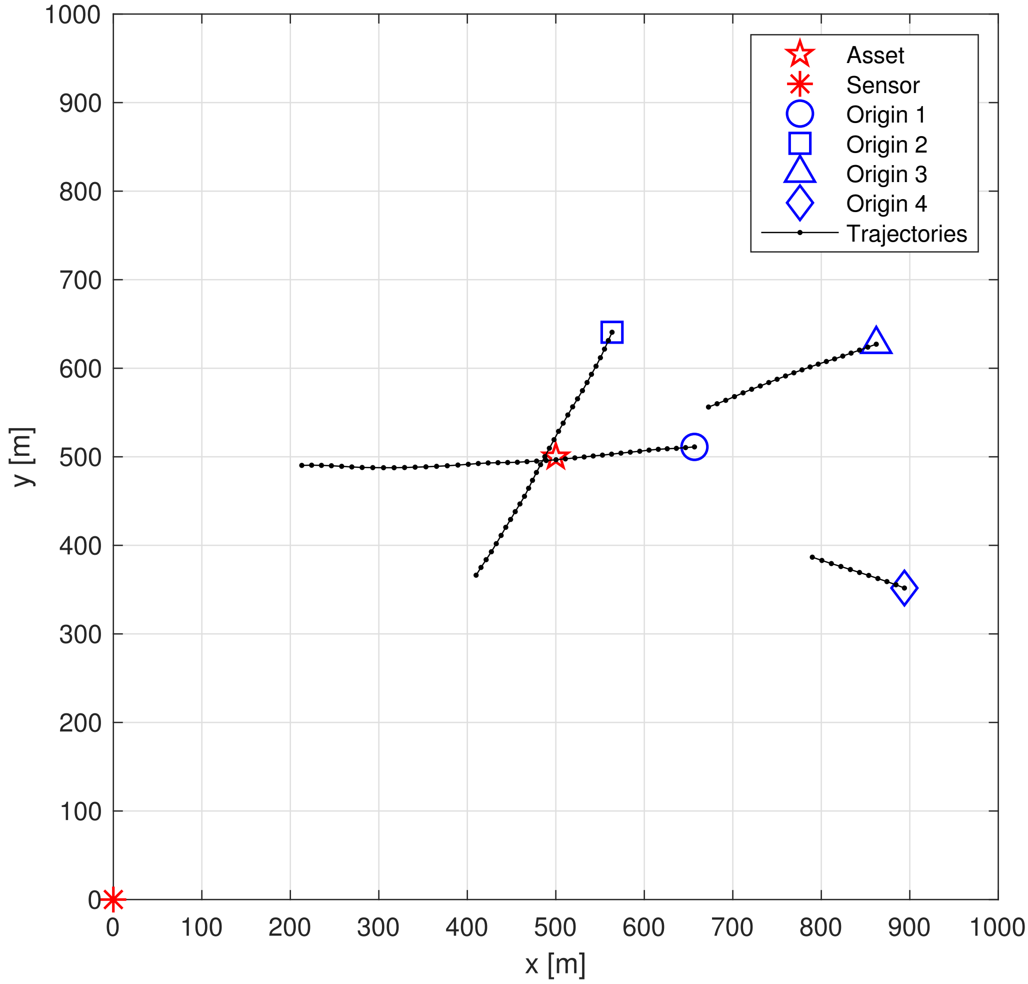


Figure 6.1: Ground truth: position plots of 5 object tracks superimposed over 50 time steps. The asset x_A is located in $[500, 500]^T$, and the sensor x_s is located at the origin.

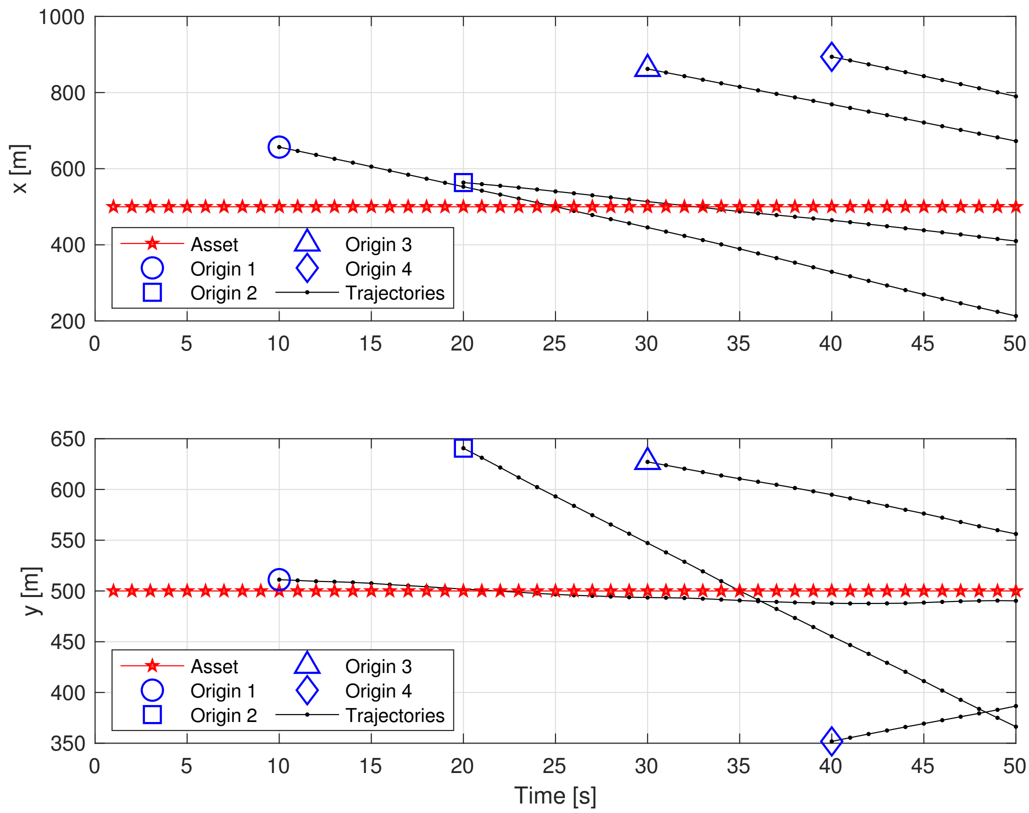


Figure 6.2: Ground truth: plots of position components, x and y, of the 4 true object tracks against time, showing the different start times.

6.4.2 Deterministic decision making (UC)

Let us now focus on the generation of the clairvoyant (ideal) decisions. These are obtained as results of decision making under certainty with Algorithm 6.3, when the ground truth is generated as described above and is directly accessible.

In the simulation we are using three distinct query functions introduced in Section 6.3. The context parameters used to specify the risk-related queries are presented in Table 6.1. Note that these query functions are constructed using the definition of a threat function, which in turn is a function of the object-to-asset geometry. Figure 6.3 demonstrates the intermediate quantities produced in response to the ground truth. Those include the values of threat that, for an individual object, grows after its appearance and drops once the object turns away from the asset. The values of threat will be further aggregated by the risk-related queries: additively in the sigma-risk model (6.30), and multiplicatively in the pi-risk model (6.31).

The produced clairvoyant decisions are presented in Figure 6.4. Note that although the ground truth is the same across the decision problems, the behaviour of the clairvoyant answers differs significantly. Let us next elaborate on the behaviour of each query over the course of the scenario.

Table 6.1: Parameters in the risk-based queries

Contextual parameter	Sigma-risk	Pi-risk
Sensitivity coefficient, b_{range} (m)	86.6	86.6
Sensitivity coefficient, b_{angle} (rad)	0.5	0.5
Asset position, (m)	$[500, 500]^T$	$[500, 500]^T$
Asset value, V_A	$> 100n$	1000
Hit value, d	100	> 1000

Behaviour of the total cardinality query $q_{\Sigma_{\mathcal{X}}}$ is presented in Figure 6.4(a). Naturally, it simply indicates the change in the number of objects at certain moments (specifically, the arrival of objects as they do not disappear). Since the query is evaluated over the whole state space $\mathcal{X}$, the cardinality value is not sensitive to the object locations, as opposed to regional cardinality [25].

The sigma-risk query q_{r_Σ} is presented in Figure 6.4(b). Despite this query is of the same nature as the total cardinality (i.e. a sum query), its behaviour is significantly different. It also displays slight variation between steps 10 and 20 when object cardinality doesn't change but the threat level does. And significantly drops when the threat level drops. This highlights that the risk value is sensitive to the evolving spatial configuration of the object population, and not simply to the number of objects.

Finally, the pi-risk query q_{r_Π} is presented in Figure 6.4(c). This query is also sensitive to the configuration of object population, but in addition we can observe that risk value saturates to the value of asset V_A on steps 20 to 24.

Recall that these are ideal decision for a decision-maker, which incur no avoidable losses at each time step. In practice, the ground truth will not be directly accessible by a decision maker, and so the answers will have to be produced using the filtering information. This will result in certain amount of (squared error) losses each time the ground truth is encountered, a situation that is studied next.

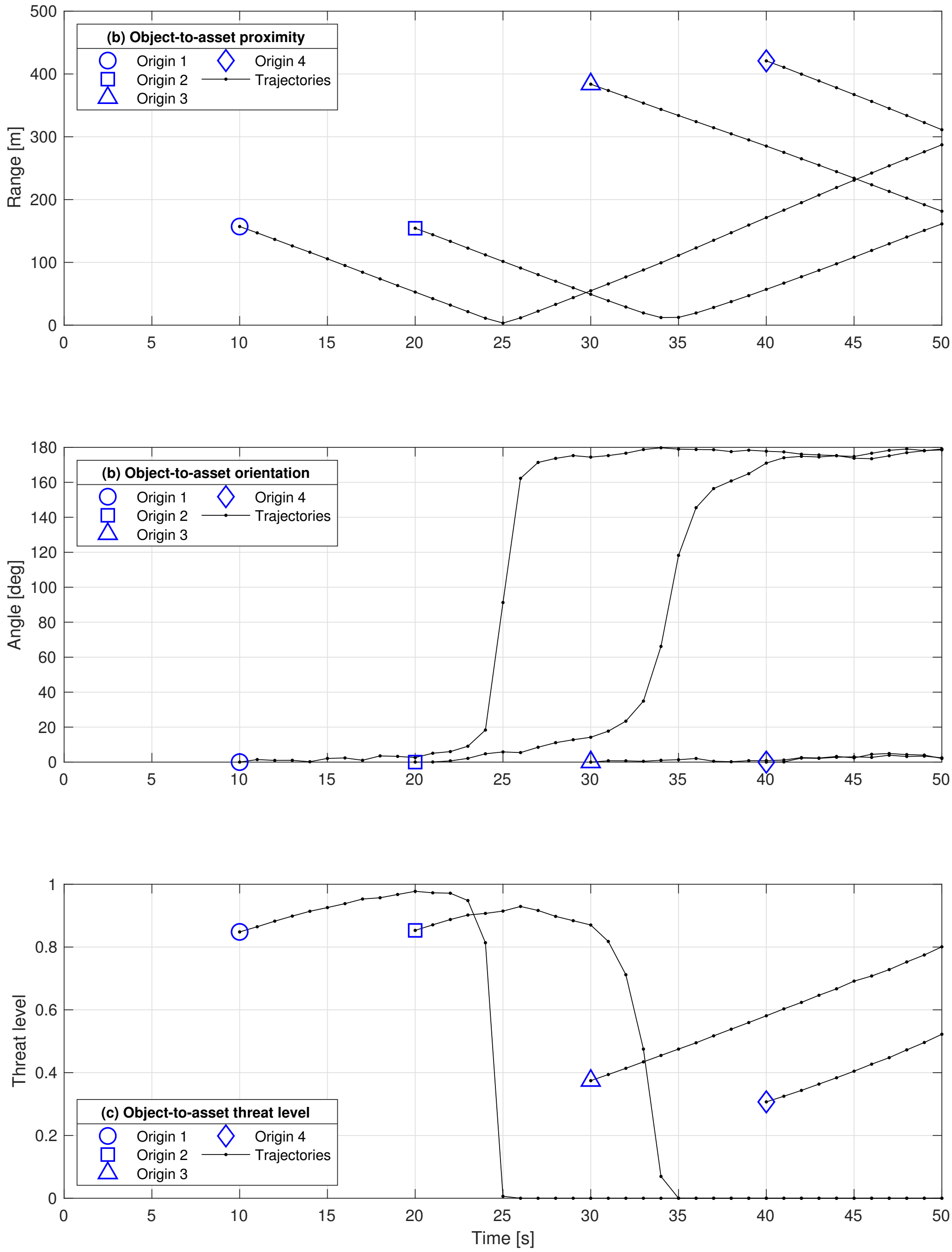


Figure 6.3: External object attributes (proximity, orientation, threat level) computed using the ground truth trajectories and the asset state: (a) proximity to the asset, which indicates an object's capability to damage the asset; (b) orientation to the asset, which indicates an object's intent to damage the asset; (c) threat level (a function of proximity and orientation), which models the probability that an object hits the asset at a current time step.

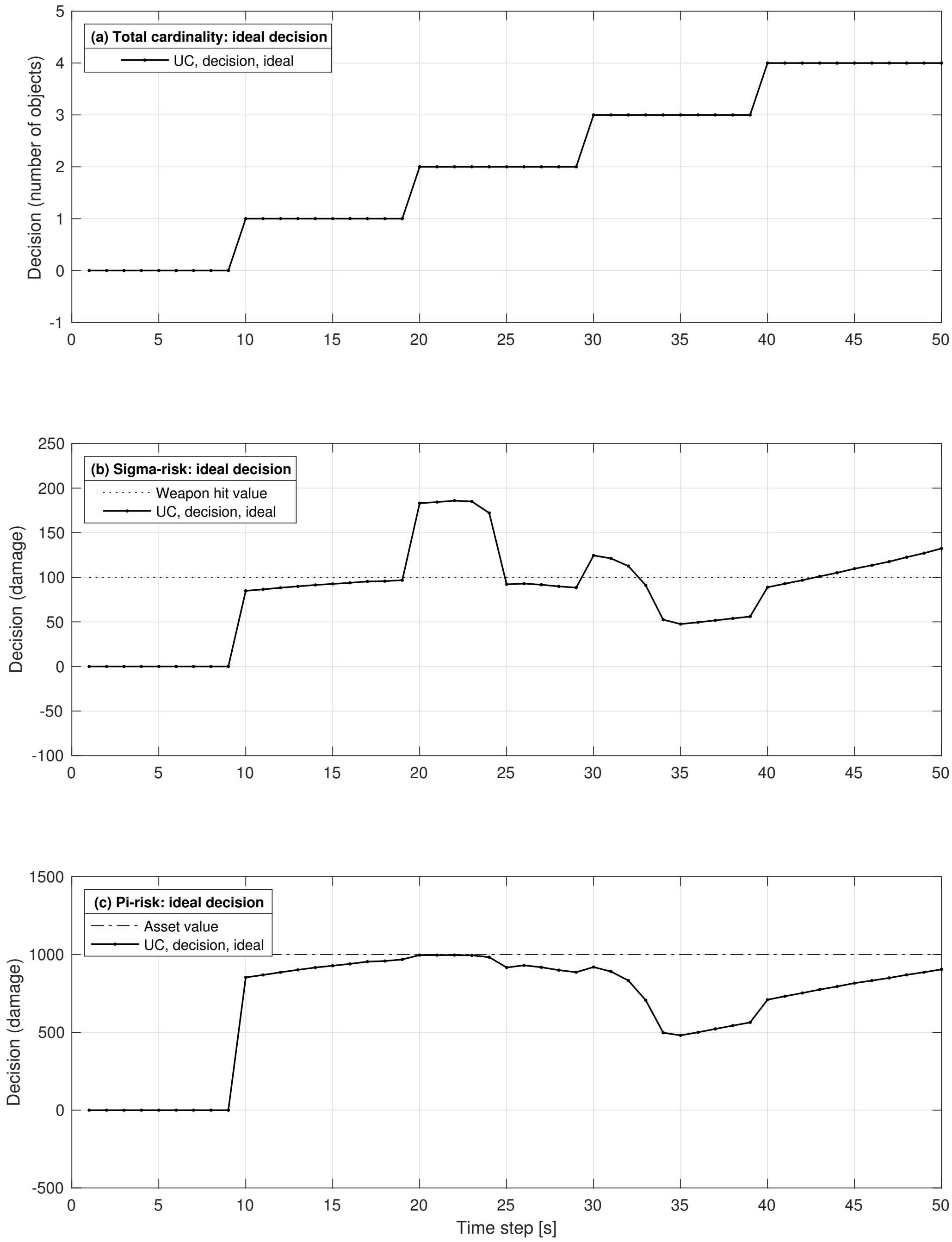


Figure 6.4: Clairvoyant (or ideal) decisions, i.e. decisions produced under certainty. Depending on the employed query function, different decisions are made in response to the same underlying ground truth scenario. Decisions concerning the estimation risk are presented next to the reference values: a weapon hit value d in (b), and the asset value V_A in (c).

6.4.3 Probabilistic decision making (UU and CE)

In the previous section we demonstrated *decision making under certainty* (UC), in circumstances when direct access to the ground truth is available. For the considered problems, the produced ideal (or clairvoyant) decisions are guaranteed to result into no avoidable losses.

In this section we consider the circumstances when the ground truth is not directly accessible, and so that decision making is approached probabilistically, using information extracted from sensed data. This inevitably results into some amount of loss associated with implementing the decision; the loss which otherwise could be avoided if the state of dynamic system was known with certainty.

We consider two approaches that permit decision making in probabilistic settings:

- *Decision making under assumed certainty equivalence* (CE), a conventional approach that focuses on removing uncertainty from the state of the dynamic system as essential step for decision making. Once the system's state is estimated (possibly, in Bayes-optimal sense), the subsequent procedure is equivalent to that of decision making under certainty. The output of the algorithm is an answer which is not guaranteed to be optimal in any sense.

- *Decision making under uncertainty* (UU), a proposed approach that focuses on future losses associated with implementing a decision, and weighs the alternative decisions to find that which minimizes the expected value of loss. This approach preserves uncertainty in the state of the dynamic system, and uses all available information to make a decision. The output of the algorithm is the optimal decision that is accompanied by an indicator of its quality, which describes the amount of loss expected from implementing the decision.

A preliminary comparison of the approaches reveals that the output of the proposed approach is *richer* than that of the conventional approach, in a sense that the produced decision is additionally equipped by the indicator of its quality. In general, it may be difficult to adequately visualize such output, but it is fairly straightforward in this study. Specifically, since the query loss function is modelled by the squared error function, the pair of quantities at the output (the decision and its quality) directly correspond, respectively, to the mean and variance of a random variable that is constructed using a query function to map from the point process state space to the real line. Subsequently, since the quality coincides with the variance, it can be easily presented next to the optimal decision (the mean) as a ± 1 standard deviation (or square root of the variance) from the mean value.

The results are averaged over M Monte Carlo runs, i.e. the scenario is executed M times and for each run a new sequence of measurement is produced. For example, for a query q on time step $k \in \{1, \cdots, T\}$ and run $i \in \{1, \cdots, M\}$, the decision produced by the conventional algorithm is given by $a_{q,k}^{CE,(i)}$, and decision produced by the proposed algorithm is given by $a_{q,k}^{UU,(i)}$ and equipped with an expected loss value $\rho_{q,k}^{UU,(i)}$. The mean decision values over M samples can be produced using

$$\mu_{a_{q,k}^{CE}} = \frac{1}{M} \sum_{i=1}^{M} a_{q,k}^{CE,(i)} \quad \text{or} \quad \mu_{a_{q,k}^{UU}} = \frac{1}{M} \sum_{i=1}^{M} a_{q,k}^{UU,(i)}, \tag{6.33}$$

depending on the used approach. Provided that the query loss is modelled by the squared error loss, the expected loss value $\rho_{q,k}^{UU,(i)}$ can be interpreted as the variance of a random variable, and thus can be used to compute its standard deviation:

$$\sigma_{q,k}^{UU,(i)} = \sqrt{\rho_{q,k}^{UU,(i)}}. \tag{6.34}$$

Accordingly, in the proposed approach for decision making under uncertainty, using M samples, it is possible to obtain the mean variance

$$\mu_{\rho_{q,k}^{UU}} = \frac{1}{M} \sum_{i=1}^{M} \rho_{q,k}^{UU,(i)}, \tag{6.35}$$

as well as the mean standard deviation

$$\mu_{\sigma_{q,k}^{UU}} = \frac{1}{M} \sum_{i=1}^{M} \sigma_{q,k}^{UU,(i)} \tag{6.36a}$$

$$= \frac{1}{M} \sum_{i=1}^{M} \sqrt{\rho_{q,k}^{UU,(i)}} \tag{6.36b}$$

Once again, the latter two sample means are only available when decision making under uncertainty is performed, and is not available from the conventional approach.

Figure 6.5 offers a simple presentation of the outputs for two approaches to decision making. Together with the ideal decisions (black lines with markers) that were first demonstrated in Figure 6.4, it presents decisions computed using the conventional approach (blue lines with markers), and decisions computed using the proposed approach (red lines with markers) along with confidence intervals (red lines without markers) defined by the value of standard deviation in (6.36). Initial comparison of the produced decisions to the ideal values reveals that both algorithms perform fairly well. The confidence interval provided by the proposed approach

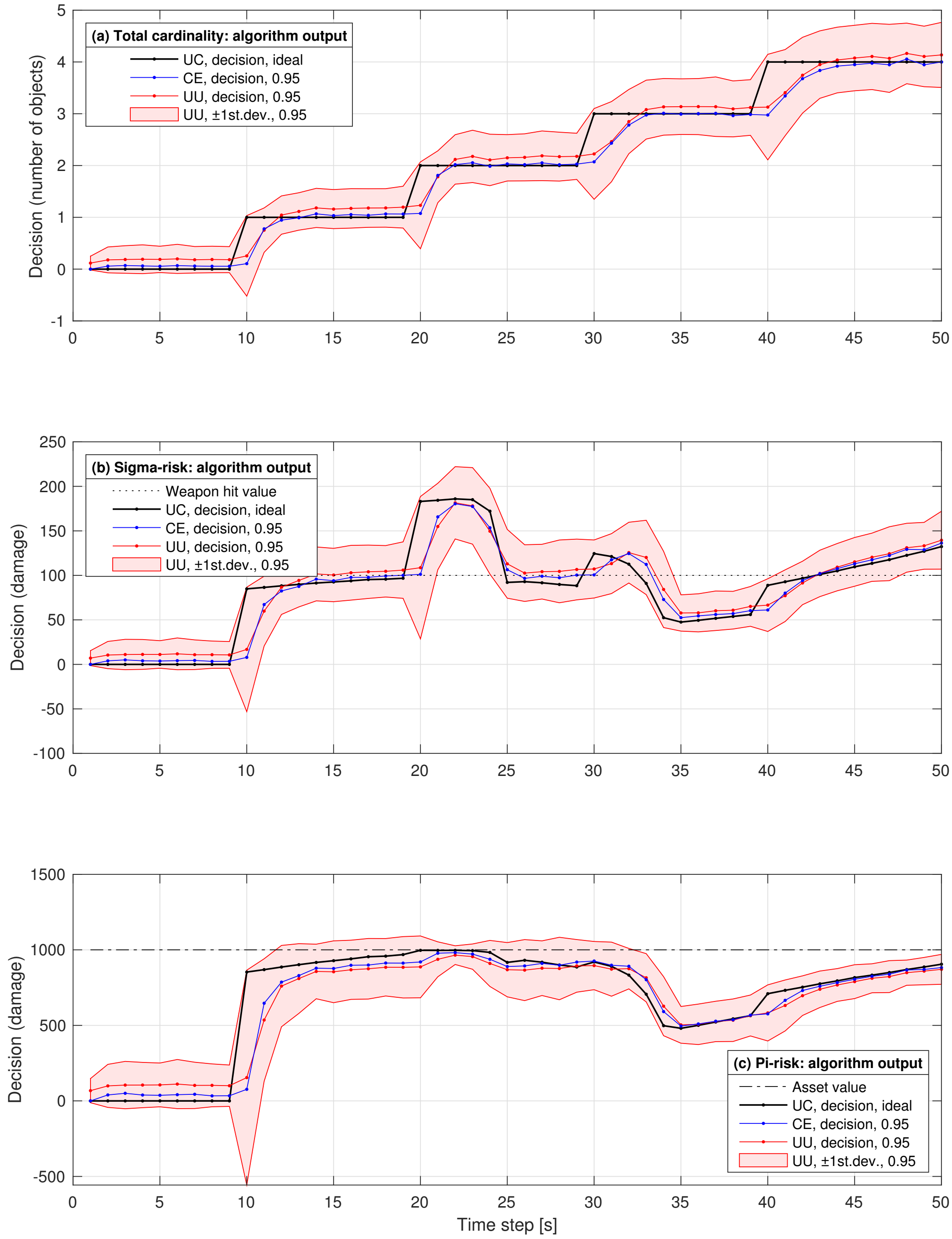


Figure 6.5: Operation of the probabilistic decision making algorithms based on the SMC-PHD filter and a sensor model with parameters given under 'Sensor 1' in Table 6.2. Decisions concerning the risk estimation are presented next to the reference values: a weapon hit value d in (b), and the asset value V_A in (c). The results are averaged over 500 Monte Carlo runs.

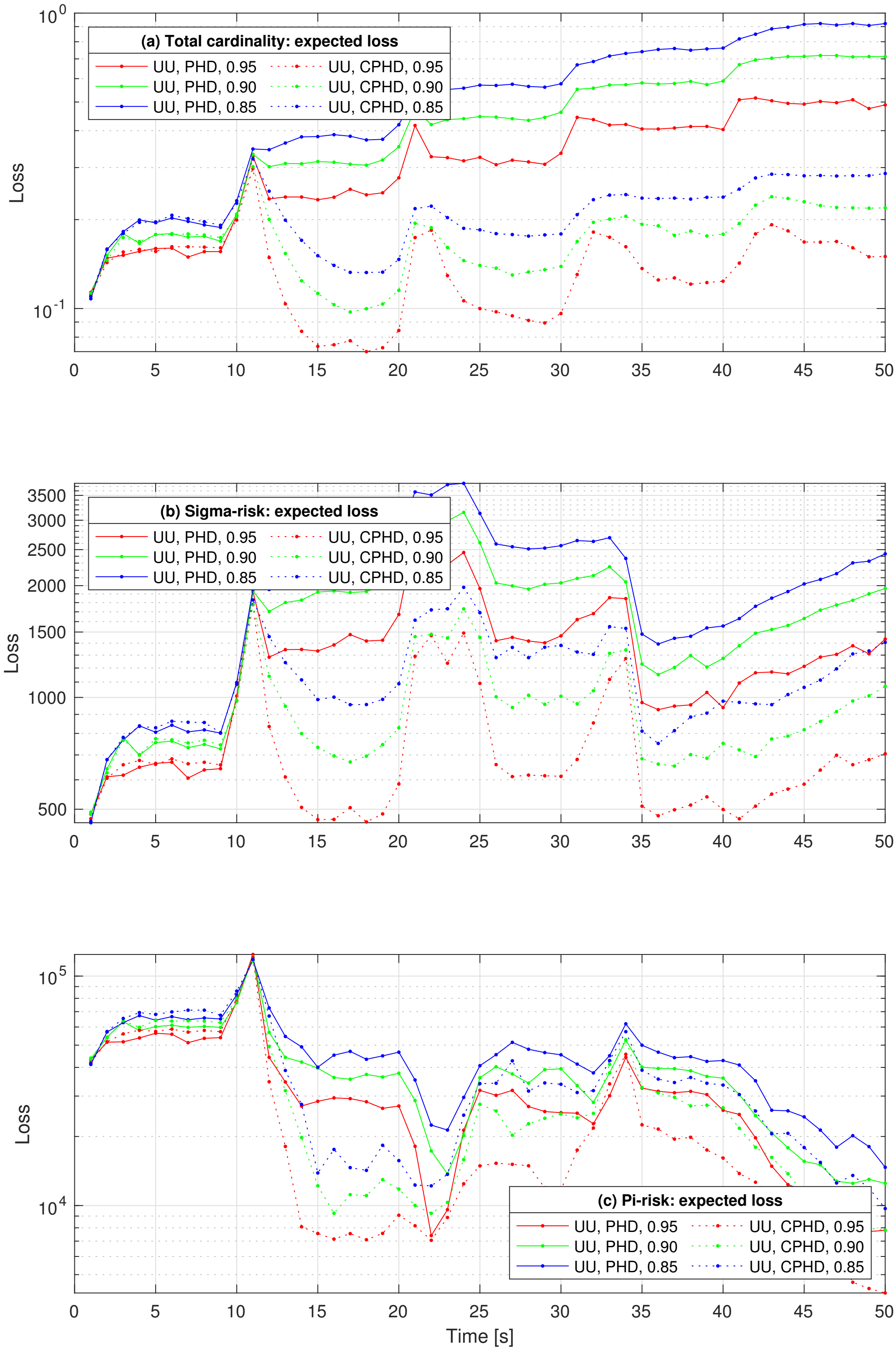


Figure 6.6: Expected values of squared error loss (i.e. the variance) associated with optimal decisions. The values are produced using the proposed approach for decision making under uncertainty using two filters (SMC-PHD and SMC-CPHD) and three sensors in Table 6.2. Note that the values of standard deviation used to plot the confidence intervals in Figure 6.5 for the sensor with $p_d = 0.95$ correspond to the solid blue line with markers. Results are averaged over 500 Monte Carlo runs.

appears be a valuable addition, as it tend to include the ideal decision. It is worth noting that the decision making results obtained using the proposed approach in Figure 6.5(a), which are related to the total cardinality query, are analogous to the cardinality estimation results presented in [25, Fig. 4] and [95, Fig. 4].

Table 6.2: Sensor parameters used in simulations

	Sensor 1	Sensor 2	Sensor 3
Probability of detection, p_d	0.95	0.90	0.85
Standard deviation in range, σ_r	5 m	5 m	5 m
Standard deviation in bearing, σ_b	1°	1°	1°

Figure 6.6 offers a study of the output in the proposed algorithm for decision making under uncertainty. Specifically, it presents the expected loss values for three sensor models (see Table 6.2): $p_d = 0.95$ (red lines), $p_d = 0.90$ (green lines), and $p_d = 0.85$ (blue lines). Note that results are obtained following (6.35). It appears that the value of expected loss decreases as the probability of detection grows, this can be explained by the fact that uncertainty decreases as detection quality improves. For the sake of comparison, additionally to the values produced with the SMC-PHD filter (solid lines), the figure presents the values obtained with the SMC-CPHD filter (dotted lines). The expected loss values obtained using the CPHD filter are overall smaller, this is due to the underlying modelling assumptions in the filter (i.i.d. cluster point process) that allow for a more refined representation of uncertainty (as compared to Poisson process). As far as the total cardinality query is concerned, the observed behaviour of the variance is consistent with that observed in [25].

Although both approaches to decision making exhibit a fairly good performance, we need to perform a quantitative analysis of the resulting loss next. This is to establish whether the proposed approach offers reduction in this resulting loss, and whether the values of expected loss can be used to predict the actual resulting loss.

6.4.4 Analysis of the resulting loss

In this section we focus on the resulting loss that emerge from implementing decisions made using the proposed decision making approach instead of implementing ideal decisions. For a query q at time step $k \in \{1, \cdots, T\}$ on the run $i \in \{1, \cdots, M\}$, the loss associated with implementing the decision $a_{q,k}^{UU,(i)}$ made under uncertainty is given by

$$L_q(a_{q,k}^{UU,(i)}, \varphi_k) = (a_{q,k}^{UU,(i)} - q(\varphi_k))^2, \tag{6.37}$$

where φ_k is the known ground truth, and the overall loss function L_q is a composition of the squared error query loss and a real-valued query q, and the loss associated with implementing the ideal decision $a_{q,k}^{UC} = q(\varphi_k)$ is

$$L_q(a_{q,k}^{UC}, \varphi_k) = 0. \tag{6.38}$$

Subsequently, the resulting loss from not implementing the ideal action is given by

$$\Delta L_{q,k}^{UU,(i)} = L_q(a_{q,k}^{UU,(i)}, \varphi_k) - L_q(a_{q,k}^{UC}, \varphi_k) \tag{6.39a}$$
$$= (a_{q,k}^{UU,(i)} - q(\varphi_k))^2. \tag{6.39b}$$

For M samples obtained in the Monte Carlo runs, we can obtain the sample mean of the resulting loss

$$\mu_{\Delta L_{q,k}^{UU}} = \frac{1}{M} \sum_{i=1}^{M} \Delta L_{q,k}^{UU,(i)} \tag{6.40a}$$
$$= \frac{1}{M} \sum_{i=1}^{M} (a_{q,k}^{UU,(i)} - q(\varphi_k))^2. \tag{6.40b}$$

In the following, we study the loss value (6.40) in three different contexts to establish the utility of the proposed decision making approach (UU).

Loss reduction due to switching from the CE to the UU approach First, we compare the proposed (UU) and conventional (CE) approaches to establish whether the proposed approach offers the reduction in resulting loss, for both the PHD and CPHD filters. Similarly to (6.40), on time step k the mean sample value of the resulting loss in the conventional algorithm is obtained with

$$\mu_{\Delta L_{q,k}^{CE}} = \frac{1}{M} \sum_{i=1}^{M} (a_{q,k}^{CE,(i)} - q(\varphi_k))^2. \tag{6.41}$$

Figure 6.7 presents the mean values of the resulting loss for both approaches using the SMC-PHD filter. In both approaches, the observed results are consistent with the fact that lower probability of detection results into higher uncertainty in the dynamic system, and thus leads to the higher values of resulting loss. An analogous result for the SMC-CPHD filter is not included here.

Overall, the proposed approach appears to outperform the conventional algorithm, as the values of resulting loss are smaller during the most of scenario. Notably, the conventional algorithm may offer superior performance at times when the ideal decisions fall into the extreme values, e.g. in the absence of objects or when the damage value in pi-risk gets saturated. This behaviour of the resulting loss in the conventional algorithm is due to the fact that it relies on the hard decision of extracting the system state to make a decision, and so is likely to point precisely at the extreme situation, e.g. the absence of objects.

As far as the total resulting loss over the whole length T of scenario is concerned, the proposed algorithm offers a reduction in loss that can be quantified by the ratio

$$\eta_q^{\text{PHD}} = \left[1 - \frac{\sum_{k=1}^{T} \mu^{\text{PHD}}_{\Delta L^{UU}_{q,k}}}{\sum_{k=1}^{T} \mu^{\text{PHD}}_{\Delta L^{CE}_{q,k}}} \right] \times 100\% \quad \text{or} \quad \eta_q^{\text{CPHD}} = \left[1 - \frac{\sum_{k=1}^{T} \mu^{\text{CPHD}}_{\Delta L^{UU}_{q,k}}}{\sum_{k=1}^{T} \mu^{\text{CPHD}}_{\Delta L^{CE}_{q,k}}} \right] \times 100\%, \tag{6.42}$$

depending on the used filter.

Figure 6.8(a) demonstrates the loss reduction due to application of the proposed approach for both the SMC-PHD and SMC-CPHD filters. For the PHD filter the proposed approach offers a 15-30 percent loss reduction, whereas for the CPHD filter the reduction is up to nearly 35 percent. The higher loss reduction is achieved for the risk-based queries (as opposed to the total cardinality), and for all queries this reduction gradually decreases with improving sensor's probability of detection.

Loss reduction due to switching from the PHD to the CPHD filter Next, we evaluate the proposed (UU) and conventional (CE) approaches to establish whether moving from the PHD filter to the CPHD filter offers the reduction in resulting loss.

Figure 6.9 presents the mean values of resulting loss for the proposed approach using the SMC-PHD and the SMC-CPHD filter. For both algorithms, the observed results are consistent with the fact that lower probability of detection results into higher uncertainty in the dynamic system, and thus leads to the higher values of resulting loss. An analogous result for the conventional approach is not included here.

Overall, implementing the proposed approach using the CPHD filter appears to outperform the PHD implementation, as the values of resulting loss are smaller during the most of scenario. Notably, the PHD implementation may offer superior performance at times when there is a change in the number of objects. This is due to a known conservativeness of cardinality estimation in the CPHD filter.

As far as the total resulting loss over the whole length T of scenario is concerned,

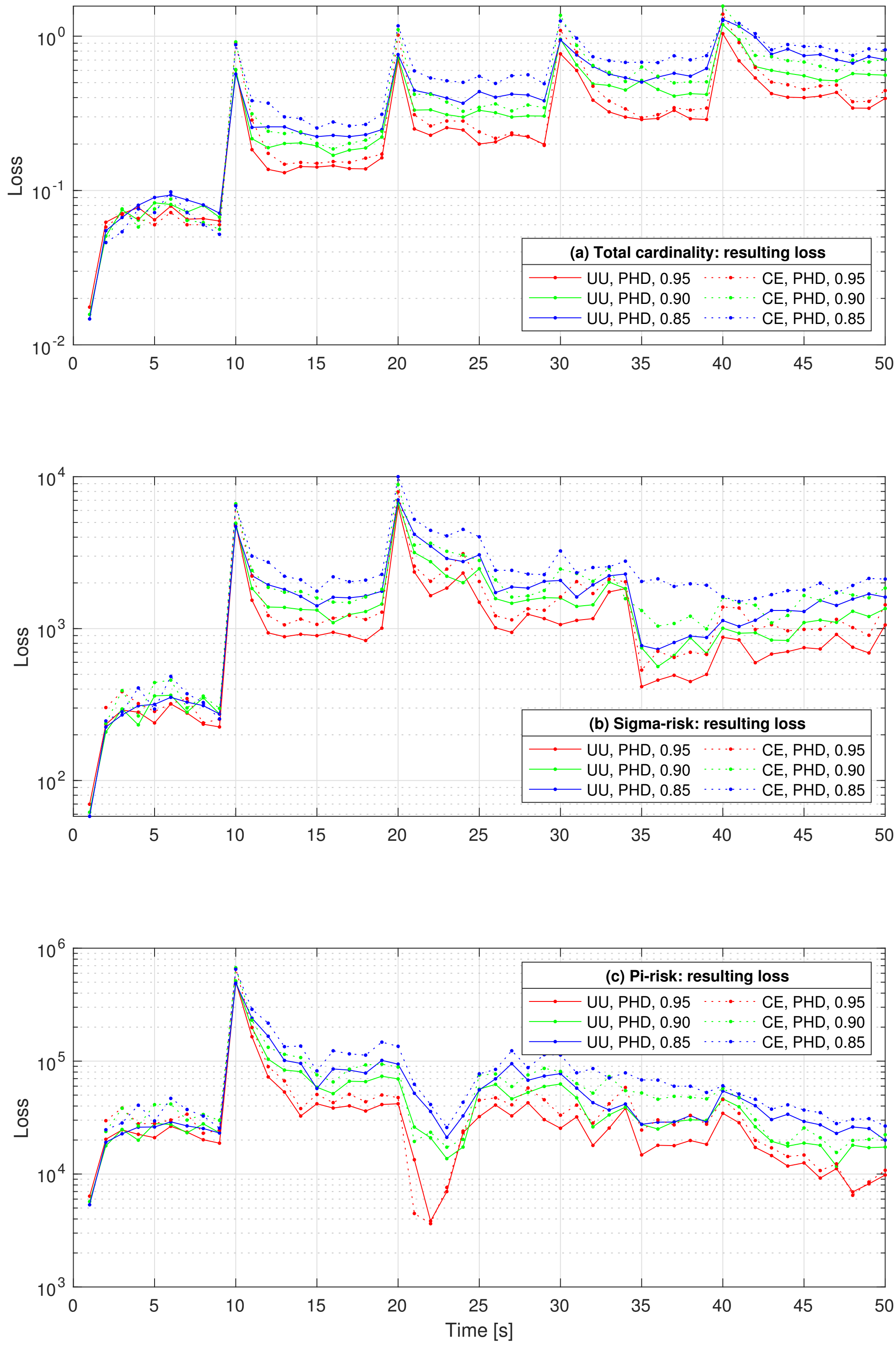


Figure 6.7: The values of resulting loss obtained using two approaches to decision making (CE and UU), for the SMC-PHD filter and three sensors in Table 6.2. The results are averaged over 500 Monte Carlo runs.

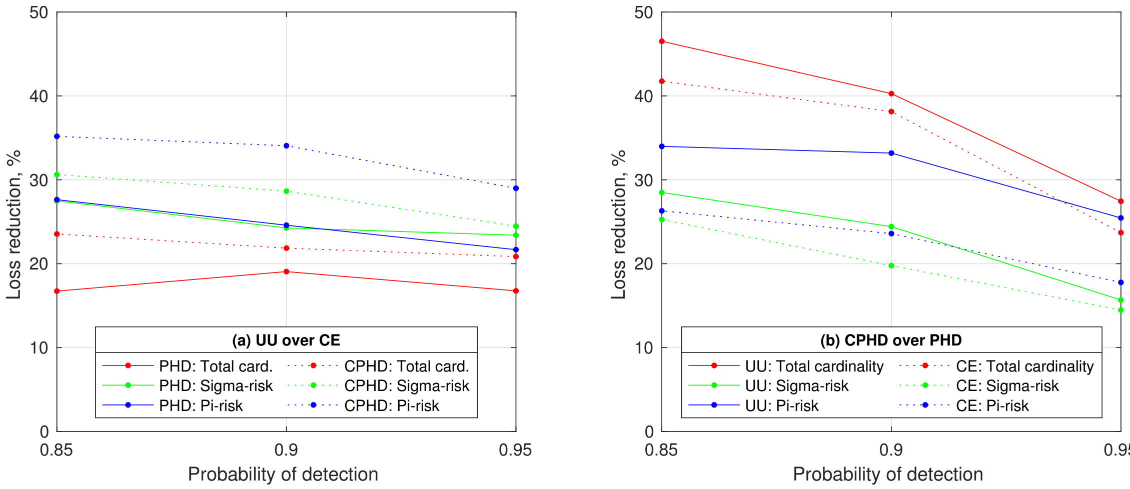


Figure 6.8: Analysis of the total resulting loss over the whole length T of scenario: (a) percentage reduction due to switching from the conventional (CE) to the proposed (UU) approach to decision making; (b) percentage reduction due to switching from the PHD to the CPHD filter.

the loss reduction due to moving from the PHD filter to the CPHD filter can be quantified by the ratio

$$\eta_q^{CE} = \left[1 - \frac{\sum_{k=1}^{T} \mu^{\text{CPHD}}_{\Delta L^{CE}_{q,k}}}{\sum_{k=1}^{T} \mu^{\text{PHD}}_{\Delta L^{CE}_{q,k}}}\right] \times 100\% \quad \text{or} \quad \eta_q^{UU} = \left[1 - \frac{\sum_{k=1}^{T} \mu^{\text{CPHD}}_{\Delta L^{UU}_{q,k}}}{\sum_{k=1}^{T} \mu^{\text{PHD}}_{\Delta L^{UU}_{q,k}}}\right] \times 100\%, \tag{6.43}$$

depending on the used decision making approach.

Figure 6.8(b) demonstrates the loss reduction due to the application of the SMC-CPHD filter instead of the SMC-PHD filter. For either of the decision making approaches, it appears beneficial to use the CPHD filter; however, the reduction is higher for the proposed approach.

Specifically, the proposed approach offers the loss reduction for up to 45 percent, and the reduction is no less than 15 percent in both approaches. The higher loss reduction is achieved for the total cardinality query (a result that is consistent with the fact that the CPHD filter is designed to improve the cardinality estimates), and for all queries this reduction slightly decreases with the growing sensor's probability of detection.

Loss prediction using the expected loss values in the proposed algorithm

Finally, we use the SMC-PHD filter to establish whether the expected loss values available from the proposed (UU) approach predict the resulting values of loss. An

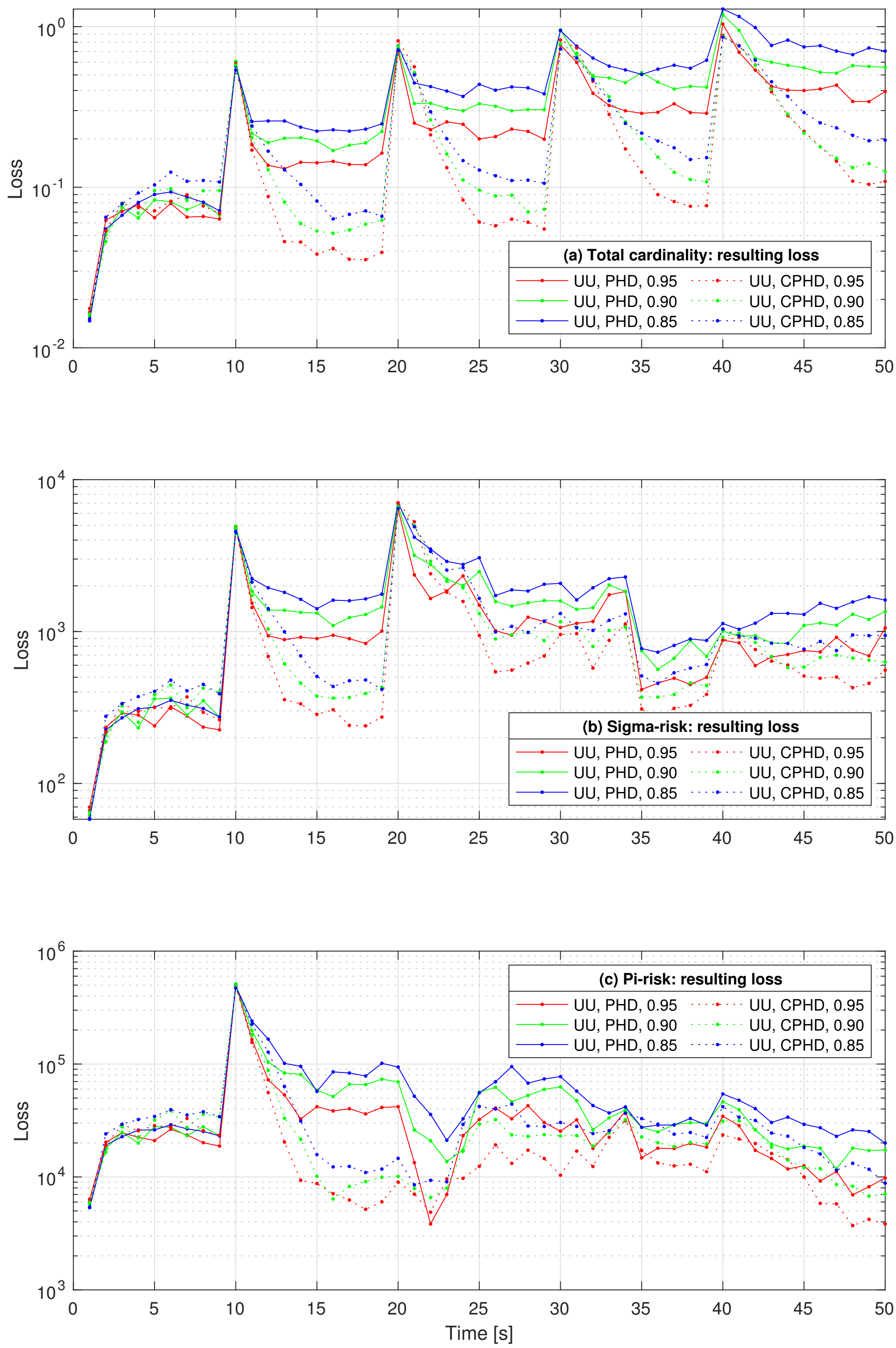


Figure 6.9: The values of resulting loss obtained using the proposed approach to decision making (UU), for the SMC-PHD and SMC-CPHD filters and three sensors in Table 6.2. Note that the values obtained for the PHD filter are equivalent to those obtained using the proposed approach in Figure 6.7. The results are averaged over 500 Monte Carlo runs.

analogous result for the SMC-CPHD filter is not included here.

For each time step k, we establish an interval of values over M samples (Monte Carlo runs) that are characteristic of the expected loss values. The interval is specified using the sample mean expected loss value $\mu_{\rho_{q,k}^{UU}}$ in (6.35), and the sample standard deviation of the expected loss given by

$$\sigma_{\rho_{q,k}^{UU}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\rho_{q,k}^{UU,(i)} - \mu_{\rho_{q,k}^{UU}}\right)^2}. \tag{6.44}$$

Figure 6.10 presents the above interval of the expected loss values (± 1 standard deviation spread around the mean) for the duration of scenario, as well as the resulting loss values (6.40). It appears that for the most of the scenario the values of resulting loss are within the interval, which indicates that the values of expected loss are useful for predicting the resulting loss values. Nevertheless, the interval fails to includes the values of resulting loss at the very beginning of the scenario, when the influence of the prior knowledge used for filter initialization is the strongest, and, possibly, on time steps when the number of objects changes.

6.5 Summary

We have implemented two approaches to decision making in probabilistic settings: the conventional approach (which explicitly relies on an extracted state summary) and the proposed algorithm (which makes decisions in the face of uncertainty in the system state). The implementation was primarily focused on the SMC-PHD filter and its update step. The same set of overall loss functions was used, which guarantees that the same decision problem has been addressed across the algorithms. Specifically, in the context of this chapter we were focused on three practical query functions: the total cardinality, the sigma-risk (Appendix A.3.2) and the pi-risk (Appendix A.3.3). The use of Monte Carlo simulations have revealed that decision making using the proposed approach can improve the result offered by the conventional approach for both the SMC-PHD and the SMC-CPHD filters. This is in addition to the unique capacity of the developed approach to equip produced decisions with indicators of their quality: the values of expected loss available when using the approach are indicatory of the resulting loss values.

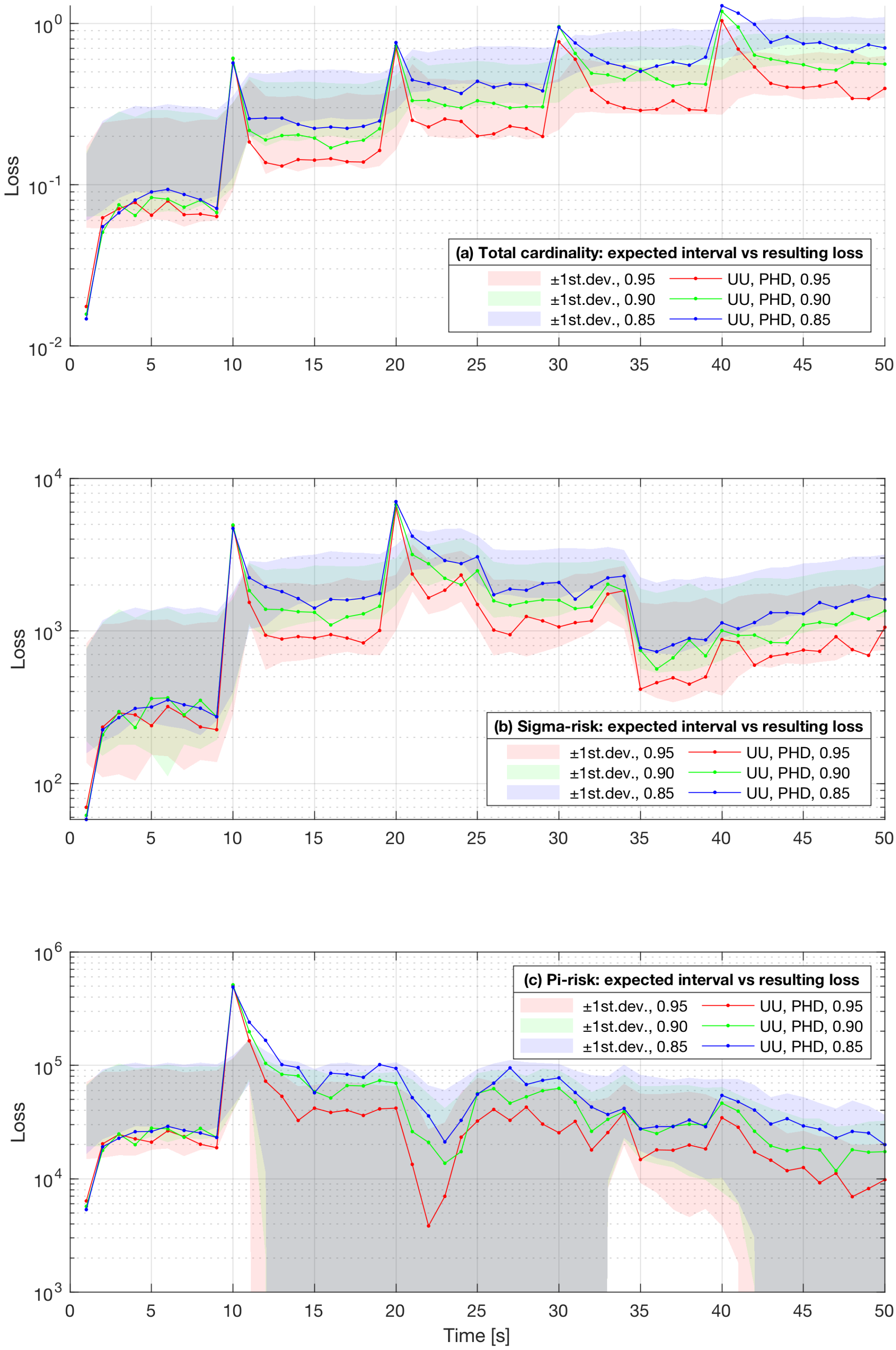


Figure 6.10: Resulting loss values presented over the intervals characterising the values of expected loss, using ±1 standard deviation around the mean values. Note that the resulting loss values are equivalent to those obtained using the SMC-PHD filter with proposed approach in Figures 6.7 and 6.9. The results are obtained over 500 Monte Carlo runs.

Chapter 7

Conclusions and future work

This work revolved around decision making with Bayesian filters. The convention is to use filtering information indirectly: to summarize the filtering distribution in the form of a point estimate before it could be used to produce a decision. Despite it is an efficient and seemingly versatile heuristic, the resulting decisions are of unknown quality and per definition underinformed. Therefore, the objective was set to develop a scheme that uses filtering distribution directly—in the spirit of Bayesian decision theory,—and produces a decision (based on all available information) which is also equipped by an uncertainty-sensitive indicator of decision quality. In the context of this thesis the two schemes were respectively denoted as *decision making under assumed certainty equivalence* and *decision making under uncertainty.*

A decision-making procedure is essentially constructed around the loss function which is defined over the product of act and state spaces. Provided that the state space is firmly determined by the filtering algorithm, the problem that gave rise to the decision circumstances would have to be expressed through selection of the act space and definition of a compatible loss function.

Specifically, this work was focused on the class of problems that could be interpreted as the problems of subjective statistical inference. While finding a way to accommodate this possibility, we found it reasonable to model the loss function as a composition of a query function and a query loss function. This generalizes the problem of point estimation, which is conventionally seen as a fairly objective procedure and correspond to the query that maps the state space on itself. In contrast, we were focused on queries that are mappings from the state space onto the subset of the real line. Throughout the thesis, the query loss was predominantly implemented by the ubiquitous squared error function, which is commonly used in the problems of inference.

As first attempt, the developed decision-theoretic formalism was used to address various statistical procedures previously developed in the context of Bayesian filtering. Along with the standard Bayesian point estimation, we interpreted the problems of threat assessment and object discrimination in the context of Kalman-like filters, and the problem of regional cardinality estimation in the context of PHD filters. An important finding was that the problem of point estimation does not have a tractable solution in the context of PHD filters (because of the complexity of the underlying state space); nevertheless, such solution is available for a rather subjective problem of regional cardinality estimation.

Based on these promising results, we moved on to synthesize more intricate inference procedures in the context of PHD filters. It might not possible to address every query function that maps on the subset of the real line, but already two considerably distinct functions would be sufficient to expose the need for and the value of the developed decision-theoretic formalism. Therefore, we limited ourselves to the following functions: the sum aggregation, which generalizes the query of regional enumeration, and the product aggregation, which is a completely novel proposal. Note that at this point the query functions were purely formal and corresponding to no underlying physical models.

Although Bayesian decision theory is commonly concerned with processing of new observations, in general, the decision procedure can be formulated without any reference to the observation. As a consequence, we were able to formulate the following strategy for developing solutions to the problem of subjective inference in the context of Bayesian filtering. First, we had to develop solutions for an abstract random element, which is a point process describing the state of an uncertain dynamic system. Second, we had to implement the solutions using information available from Bayesian filters. For the two considered query functions, we were able to produce optimal solutions that are expressed through quantities commonly used to characterise a point process: its lower-order statistical moments and p.g.fl. The possibility to obtain these solutions is largely due to the favourable properties of the squared error query loss.

Next, we used the formalism of point process theory to extract the necessary quantities from the update step of practical Bayesian filters: the classic PHD filter, the Panjer PHD filter, and the Cardinalized PHD filter. Some of the required quantities are not commonly maintained by the filters, so we had to propagate them additionally to produce quantities for the prediction step. Ultimately, we were able to produce two sets of solutions, corresponding both to the update step as well as to the prediction step.

Finally, we decided to demonstrate the developed results using simulated data. We abandoned the abstract nature of the query functions, and offered their physically meaningful interpretations. In addition to the function of total cardinality, we developed a model that evaluates to the value of risk (or expected damage to a vulnerable asset) attached to a population of objects. Specifically, we offered two approximations to this function which are well compatible with the sum and product aggregations. As expected, the simulation results have exposed that the developed approach has a novel ability of equipping the produced decisions with indicators of their quality. Variations in the level of decision quality have been consistent with the variations of sensor quality (used probability of detection). We also used those query functions along with the squared error query loss to compose overall loss functions, which are suitable to model decision making under assumed certainty equivalence for filters equipped with a state extraction algorithm. Overall, the developed approach to decision making has demonstrated its capacity to outperform the conventional approach.

Let us now highlight some shortcomings of this work, and suggest a number of ways for further developments. Admittedly, this work was focused on the very limited set of query functions that capture the essence of the problem that gives rise to the decision circumstances. Implementing other functions, such as those found in Section 3.5.4, is an exciting avenue for future research. Furthermore, the solutions have been produced for the squared error query loss, which is appropriate for estimation problems, but has to be replaced by other physically meaningful functions if practical problems are to be addressed (e.g. by zero-one error). Finally, this work has not addressed the possibility of improving decision quality via sensor management, which is a natural development stemming from this work.

Appendix A

A.1 Query loss functions

Let us briefly visit two standard query loss functions that are suitable for decisions making situations when $\mathcal{A} \subseteq \mathbb{R}$ and $\mathcal{A} = \{0, 1\}$. The *squared error* function in Definition A.1.1 is one of the most commonly used functions in decision-theoretic context. Originally proposed by Gauss in 1810, who has explicitly acknowledged its arbitrary nature and was defending it on grounds of simplicity [89]. The same reasoning has been used by Wiener, as reported in [5]. The second query loss function is a *cost matrix* in Definition A.1.2 (cf. Table 1.2), which correspond to numerous problems that lead to conclusions, such as detection (or hypothesis testing) [53], and in other practical problems that lead to actions such as in the classical umbrella problem.

Definition A.1.1 (Squared error query loss). *The amount of loss associated with reporting an answer $a \in \mathcal{A}$ when the correct (or ideal) answer is $\check{a} \in \mathcal{A}$, and $\mathcal{A} \subseteq \mathbb{R}$, is given by*

$$l_2(a, \check{a}) := (a - \check{a})^2. \tag{A.1}$$

Definition A.1.2 (Cost matrix query loss).
The amount of loss associated with accepting an answer $a \in \mathcal{A}$ when the correct (or ideal) answer is $\check{a} \in \mathcal{A}$, and $\mathcal{A} = \{0, 1\}$, is given by

$$l_C(a, \check{a}) = \begin{cases} C_{00}, & a = 0, \check{a} = 0, \\ C_{01}, & a = 0, \check{a} = 1, \\ C_{10}, & a = 1, \check{a} = 0, \\ C_{11}, & a = 1, \check{a} = 1 \end{cases} \tag{A.2a}$$

$$= C_{a0}\mathbb{1}_0(\check{a}) + C_{a1}\mathbb{1}_1(\check{a}) \tag{A.2b}$$

$$= C_{a0} + (C_{a1} - C_{a0})\mathbb{1}_1(\breve{a}) \tag{A.2c}$$

where $\mathbb{1}_0 : \{0,1\} \to \{0,1\}$ *and* $\mathbb{1}_1 : \{0,1\} \to \{0,1\}$ *are indicator functions evaluating, respectively, whether a value from* $\{0,1\}$ *is equal to* 0 *or to* 1*.*

A.2 Threat functions

Consider a surveillance area $\mathcal{X}$ that contains a vulnerable asset in state x_A. From the asset's perspective, the presence of any other object x in the area may lead to the asset's damage. The probability of an object-asset interaction that leads to non-negligible negative consequences is given by the function $\tau : \mathcal{X} \times \mathcal{X} \to [0,1]$, which we call a *threat function*.

A simple way to model the threat function is to define it as a product of subjective object attributes:

$$\tau(x, x_A) := \mathrm{c}(x, x_A) \cdot \mathrm{i}(x, x_A), \tag{A.3}$$

where $\mathrm{c} : \mathcal{X} \times \mathcal{X} \to [0,1]$ models the object's capability to damage the asset, and $\mathrm{i} : \mathcal{X} \times \mathcal{X} \to [0,1]$ models the object's intent to damage the asset. Accordingly, an object which have no capability or intent to damage the asset, or lacks either attribute, will exhibit low level of threat. In contrast, an object that can be attributed high values of capability and intent, will produce a considerable level of threat.

For the sake of demonstration, it is convenient to construct a threat function using basic relational parameters, which can be produced from the kinematic states of the considered entities. It is then possible to express capability based on the range between the object and the vulnerable asset (Section A.2.1), and express intent based on the angle between the object's heading and bearing to the asset (Section A.2.2).

A.2.1 Capability factor

A simple instance of a threatening object is an explosive device. It is commonly assumed that impact projected by the weapon is omnidirectional, and so the capability to produce damage depends exclusively on the distance from the impact point. Provided that the impact point coincides with the object's state x, it becomes possible to model the probability of damaging the asset in the known state x_A. Next we present a selection of functions (introduced as damage functions in

[65]), which can be used to model this probability. More formally, a capability function $\mathrm{c} : \mathcal{X} \times \mathcal{X} \to [0, 1]$ evaluates the probability that a single weapon detonation in point x damages a threatened asset in state x_A.

Definition A.2.1 (Cookie-cutter function)**.** *The cookie-cutter function is given by*

$$\mathrm{c}(x, x_A) = \begin{cases} 1, & r(x, x_A) \leq r_0, \\ 0, & r(x, x_A) > r_0, \end{cases} \tag{A.4}$$

where $r : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *evaluates range between the impact point* x *and the asset* x_A*, and* $r_0 > 0$ *is called the lethal range.*

Definition A.2.2 (Gaussian function)**.** *The Gaussian (or normal) function is given by*

$$\mathrm{c}(x, x_A) = \exp\left(-\frac{r(x, x_A)^2}{2b^2}\right) \tag{A.5}$$

where $r : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *evaluates range between the impact point* x *and the asset* x_A*, and* b *is a parameter.*

Definition A.2.3 (Exponential function)**.** *The exponential function is given by*

$$\mathrm{c}(x, x_A) = \exp\left(-\frac{r(x, x_A)}{b}\right) \tag{A.6}$$

where $r : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *evaluates range between the impact point* x *and the asset* x_A*, and* b *is a parameter.*

Definition A.2.4 (Lognormal function)**.** *The lognormal function is given by*

$$\mathrm{c}(x, x_A) = \frac{1}{2}\left(1 - \mathrm{erf}\left[\frac{\ln\left(\frac{r(x, x_A)}{\alpha}\right)}{\sqrt{2}\beta}\right]\right), \tag{A.7}$$

where $r : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *evaluates range between the impact point* x *and the asset* x_A*,* $\mathrm{erf}(\cdot)$ *is the error function*

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \tag{A.8}$$

and α *and* β *are parameters.*

A.2.2 Intent factor

The model of an object's intent comprises various aspects, including those specifying interaction with the object's point of interest [1, 62], as well as those quantifying the degree of object's hostility with respect to that point [45, 61]. Provided that the point of interest is represented by the vulnerable asset, a simple model of hostile intent can be constructed based on the angle that measures deviation between the object's heading and the bearing to the asset. This deviation is among most basic indicators revealing hostile intent [80]. Eventually, the object's intent function $\mathrm{i} : \mathcal{X} \times \mathcal{X} \to [0, 1]$ can be modelled using functions similar to those presented in Section A.2.1, or using [26, Eq. 16].

A.3 Risk functions

We have developed a model that computes the value of expected damage anticipated by a vulnerable asset when a group of weapons in a known state detonates in its proximity. It is constructed using basic tools of probability theory and a threat function in Appendix A.2 modelling the probability that a detonation will impact the asset.In addition, it models the value of the asset itself V_A and the fixed value d potentially removed by a weapon. In general, such expression may be rather difficult to deal with, but by limiting ourselves to specific relations among those values we are able to produce two compact expressions in additive and multiplicative forms. Overall, these expressions represent useful and physically meaningful models of certain real-life phenomena which can be modelled as a mapping from the population state space $\mathfrak{X}$ to the real line $\mathbb{R}$.

A.3.1 Risk model

Consider a set of impact points modelled by a sequence $\varphi = (x_1, \cdots, x_n) \in \mathfrak{X}$. A random variable D_i describing the amount of damage projected from the i-th impact point x_i to the asset in x_A is defined as

$$\mathrm{D}_i = d \cdot \mathrm{H}_i, \tag{A.9}$$

where d is a constant describing the damaging capacity of a weapon, and H_i is a Bernoulli random variable describing the boolean-valued outcome whether the hit was successful or not. The random variable H_i takes the value 1 with probability $\tau(x_i, x_A)$ and the value 0 with probability $1 - \tau(x_i, x_A)$, where $\tau : \mathcal{X} \times \mathcal{X} \to [0, 1]$ is a threat function from Appendix A.2.

For an asset of value $V_A > 0$, a random variable describing the total damage anticipated from a set of impact points $\varphi \in \mathfrak{X}$ is given by

$$\mathrm{D}_\varphi = \min\left(V_A, \sum_{1\le i\le n} \mathrm{D}_i\right), \tag{A.10}$$

where $\{\mathrm{D}_i\}$ are independent.

Describing damage probabilistically is common in literature on risk assessment, see e.g. [46, Fig. 1] and [58, Fig. 1a]. Although it is commonly studied using statistics that describe the tail of the distribution (quantiles), we shall focus on the mean value of damage. The expected value of damage r, which we call *risk*, is given by

$$\begin{aligned} r(\varphi) &= \mathbb{E}[\mathrm{D}_\varphi] && \text{(A.11a)}\\ &= \mathbb{E}\left[\min\left(V_A, \sum_{1\le i\le n} \mathrm{D}_i\right)\right] && \text{(A.11b)}\\ &= \mathbb{E}\left[\min\left(V_A, d\sum_{1\le i\le n} \mathrm{H}_i\right)\right] && \text{(A.11c)}\\ &= \sum_{0\le k\le|\varphi|} \min(V_A, k\cdot d)\cdot p_\mathrm{K}(k), && \text{(A.11d)} \end{aligned}$$

where p_K is a probability mass function of a Poisson binomial distribution describing the number of detonations that have hit the asset and given by

$$p_\mathrm{K}(k) = \sum_{S\in F_k}\prod_{i\in S}\tau(x_i, x_A)\prod_{j\in S^c}(1-\tau(x_j, x_A)), \tag{A.12}$$

where F_k is the set of all subsets of k integers that can be selected from the set $\{1, \cdots, n\}$, and S^c is the complement of S. This expression may difficult to compute in general, therefore certain assumptions will be made.

A.3.2 Sigma-risk of the robust asset

Assumptions A.3.1 (Robust asset)**.** *$V_A \ge n\cdot d$, $n \in \mathbb{N}$.*

Proposition A.3.2 (Sigma-risk)**.** *Under Assumptions A.3.1 of robust asset, the* *risk associated with a group of weapons in state φ is given by*

$$r_\Sigma(\varphi) = d \sum_{x \in \varphi} \tau(x, x_A), \tag{A.13}$$

and is called sigma-risk.

Proof. Consider expression of the expected damage in (A.11b). Provided that the sum $\sum_{1 \le i \le n} \mathrm{D}_i$ can never exceed the value of V_A (due to Assumption A.3.1), the expression reduces to

$$\begin{aligned} \mathbb{E}[\mathrm{D}_\varphi] &= \mathbb{E}\left[\sum_{1 \le i \le n} \mathrm{D}_i\right] && \text{(A.14a)} \\ &= \sum_{1 \le i \le n} \mathbb{E}[\mathrm{D}_i] && \text{(A.14b)} \\ &= \sum_{1 \le i \le n} (d \cdot \tau(x_i, x_A) + 0 \cdot (1 - \tau(x_i, x_A))) && \text{(A.14c)} \\ &= d \sum_{1 \le i \le n} \tau(x_i, x_A) && \text{(A.14d)} \end{aligned}$$

□

The model in (A.13) is a novel development, which is not paralleled in literature, and to the best of our understanding can serve as a physically meaningful alternative to the hypothetical concept of population threat which is defined to be additive across the objects [42, 69, 76, 118].

A.3.3 Pi-risk of the fragile asset

Assumptions A.3.3 (Fragile asset)**.** $V_A < d$.

Proposition A.3.4 (Pi-risk)**.** *Under Assumptions A.3.3 of fragile asset, the risk associated with a group of weapons in state φ is given by*

$$r_\Pi(\varphi) = V_A\left[1 - \prod_{x \in \varphi} \left[1 - \tau(x, x_A)\right]\right], \tag{A.15}$$

and is called pi-risk.

Proof. Consider expression of the expected damage in (A.11d) under Assumption

A.3.3:

$$\mathbb{E}[\mathrm{D}_\varphi] = \sum_{0\le k\le|\varphi|} \min(V_A, k\cdot d)\cdot p_{\mathrm{K}}(k) \tag{A.16a}$$

$$= \min(V_A, 0)\cdot p_{\mathrm{K}}(0) + \sum_{1\le k\le|\varphi|} \min(V_A, k\cdot d)\cdot p_{\mathrm{K}}(k) \tag{A.16b}$$

$$= V_A \sum_{1\le k\le|\varphi|} p_{\mathrm{K}}(k) \tag{A.16c}$$

$$= V_A(1 - p_{\mathrm{K}}(0)). \tag{A.16d}$$

From (A.12) we can determine the probability of 0 successful hits

$$p_{\mathrm{K}}(0) = \prod_{j\in\{1,\cdots,n\}} \big[1 - \tau(x_j, x_A)\big]. \tag{A.17}$$

Substituting (A.17) into (A.16d) yields the desired result. □

The model (A.15) of pi-risk is a fundamental result, as one can recognize its close relation to the expressions of 'risk' in [16, Eq. 1] and 'vulnerability' in [41, Eq. 18]. Furthermore, in case $V_A = 1$, expression in (A.15) reduces to the 'probability of kill', the probability that the asset is destroyed after weapon detonation, which is a quantity of importance in the algorithms for threat evaluation and impact assessment [90], and can be seen as an extension of a threat function of Appendix A.2 to the case of multiple threatening objects.

Bibliography

[1] Bashar I Ahmad, James K Murphy, Patrick M Langdon, and Simon J Godsill. Bayesian intent prediction in object tracking using bridging distributions. *IEEE transactions on cybernetics*, 48(1):215–227, 2016.

[2] Edson Hiroshi Aoki, Arunabha Bagchi, Pranab Mandal, and Yvo Boers. Particle filter approximations for general open loop and open loop feedback sensor management. *Memorandum/Department of Applied Mathematics*, (1952), 2011.

[3] Edson Hiroshi Aoki, Arunabha Bagchi, Pranab Mandal, and Yvo Boers. A theoretical analysis of bayes-optimal multi-target tracking and labelling. *Department of AppliedMathematics, University of Twente, Enschede*, 1953, 2011.

[4] Michael Athans. The role and use of the stochastic linear-quadratic-gaussian problem in control system design. *IEEE transactions on automatic control*, 16(6):529–552, 1971.

[5] Tamer Basar. *Twenty-five seminal papers in control*. Wiley-IEEE Press, 2000.

[6] Erkan Baser, Michael McDonald, Thia Kirubarajan, and Murat Efe. A joint multitarget estimator for the joint target detection and tracking filter. *IEEE Transactions on Signal Processing*, 63(15):3857–3871, 2015.

[7] Robert Bassett and Julio Deride. Maximum a posteriori estimators as a limit of bayes estimators. *Mathematical Programming*, 174(1-2):129–144, 2019.

[8] Marcus Baum, Peter Willett, and Uwe D Hanebeck. Polynomial-time algorithms for the exact mmospa estimate of a multi-object probability density represented by particles. *IEEE Transactions on Signal Processing*, 63(10):2476–2484, 2015.

[9] James O Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media, 1985.

[10] James O Berger. Statistical decision theory. In *Time Series and Statistics*, pages 277–284. Springer, 1990.

[11] J. M. Bernardo. Integrated objective bayesian estimation and hypothesis testing (with discussion). In *Bayesian Statistics, Proceedings of the Ninth Valencia Meeting*, volume 9, pages 1–68, 2011.

[12] José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.

[13] David Blackwell and Meyer A Girshick. *Theory of games and statistical decisions*. Wiley, 1954.

[14] Yvo Boers, Hans Driessen, and Arun Bagchi. Point estimation for jump markov systems: Various map estimators. In *Information Fusion, 2009. FUSION'09. 12th International Conference on*, pages 33–40. IEEE, 2009.

[15] Yvo Boers, Hans Driessen, Arunabha Bagchi, and Pranab Mandal. Particle filter based entropy. In *2010 13th International Conference on Information Fusion*, pages 1–8. IEEE, 2010.

[16] Fok Bolderheij and Piet Van Genderen. Mission driven sensor management. In *Information Fusion, 2004 7th International Conference on*. IEEE, 2004.

[17] Edmund Brekke and Mandar Chitre. Relationship between finite set statistics and the multiple hypothesis tracker. *IEEE Transactions on Aerospace and Electronic Systems*, 54(4):1902–1917, 2018.

[18] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[19] Daniel Clark, Emmanuel Delande, and Jérémie Houssineau. *Basic concepts for multi-object estimation (lecture notes)*. 2016.

[20] Daniel Clark and Simon Godsill. Group target tracking with the gaussian mixture probability hypothesis density filter. In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pages 149–154. IEEE, 2007.

[21] Daniel Clark, Ioseba Tena Ruiz, Yvan Petillot, and Judith Bell. Particle phd filter multiple target tracking in sonar image. *IEEE Transactions on Aerospace and Electronic Systems*, 43(1):409–416, 2007.

[22] Daniel E Clark and Judith Bell. Multi-target state estimation and track continuity for the particle phd filter. *IEEE Transactions on Aerospace and Electronic Systems*, 43(4):1441–1453, 2007.

[23] Daniel E Clark and Jeremie Houssineau. Faa di bruno's formula for gateaux differentials and interacting stochastic population processes. *arXiv preprint arXiv:1202.0264*, 2012.

[24] Fred Daum. Nonlinear filters: beyond the kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, 20(8):57–69, 2005.

[25] E. Delande, M. Üney, J. Houssineau, and D. E. Clark. Regional variance for multi-object filtering. *IEEE Transactions on Signal Processing*, 62(13):3415–3428, July 2014.

[26] Maarten Ditzel, Leon Kester, Sebastiaan van den Broek, and Martin van Rijn. Cross-layer utility-based system optimization. In *Proceedings of the 16th International Conference on Information Fusion*, pages 507–514. IEEE, 2013.

[27] Hugh Durrant-Whyte. Multi sensor data fusion. version 1.2. 2001.

[28] John Eatwell, Murray Milgate, and Peter Newman. *Utility and probability*. Springer, 1990.

[29] Ozgur Erdinc, Peter Willett, and Yaakov Bar-Shalom. Probability hypothesis density filter for multitarget multisensor tracking. In *2005 7th International Conference on Information Fusion*, volume 1, pages 8–pp. IEEE, 2005.

[30] Ramsey Faragher. Understanding the basis of the kalman filter via a simple and intuitive derivation [lecture notes]. *IEEE Signal Processing Magazine*, 29(5):128–132, 2012.

[31] Thomas Fortmann, Yaakov Bar-Shalom, and Molly Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE journal of Oceanic Engineering*, 8(3):173–184, 1983.

[32] Nicola Giocoli. From wald to savage: homo economicus becomes a bayesian statistician. *Journal of the History of the Behavioral Sciences*, 49(1):63–95, 2013.

[33] Simon Godsill, Arnaud Doucet, and Mike West. Maximum a posteriori sequence estimation using monte carlo particle filters. *Annals of the Institute of Statistical Mathematics*, 53(1):82–96, 2001.

[34] Irwin R Goodman, Ronald P Mahler, and Hung T Nguyen. *Mathematics of data fusion*, volume 37. Springer Science & Business Media, 2013.

[35] Neil Gordon. *Bayesian methods for tracking.* PhD thesis, Imperial College London (University of London), 1993.

[36] Marco Guerriero, Lennart Svensson, Daniel Svensson, and Peter Willett. Shooting two birds with two bullets: How to find minimum mean ospa estimates. In *2010 13th International Conference on Information Fusion*, pages 1–8. IEEE, 2010.

[37] Fredrik Gustafsson. *Adaptive filtering and change detection*, volume 1. Citeseer, 2000.

[38] Gregory D Hager. *Task-Directed Sensor Fusion and Planning: A Computational Approach*, volume 99. Springer Science & Business Media, 1990.

[39] Alfred O Hero and Douglas Cochran. Sensor management: Past, present, and future. *IEEE Sensors Journal*, 11(12):3064–3075, 2011.

[40] YC Ho and Robert Lee. A bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9(4):333–339, 1964.

[41] John R Hoffman, Eric Sorensen, Chad A Stelzig, Ronald PS Mahler, Adel I El-Fallah, and Mark G Alford. Scientific performance estimation of robustness and threat. In *Signal Processing, Sensor Fusion, and Target Recognition XI*, volume 4729, pages 248–259. International Society for Optics and Photonics, 2002.

[42] William J Horrey, Christopher D Wickens, Richard Strauss, Alex Kirlik, and Thomas R Stewart. Supporting situation assessment through attention guidance and diagnostic aiding: The benefits and costs of display enhancement on judgment skill. *Adaptive perspectives on human technology interaction: Methods and models for cognitive engineering and human-computer interaction*, pages 55–70, 2006.

[43] AFM Islam. *Loss functions, utility functions and Bayesian sample size determination.* PhD thesis, Queen Mary, University of London, 2011.

[44] Andrew H Jazwinski. *Stochastic processes and filtering theory.* Dover Publications Inc., 2007.

[45] Fredrik Johansson and Goran Falkman. Implementation and integration of a bayesian network for prediction of tactical intention into a ground target simulator. In *2006 9th international conference on information fusion*, pages 1–7. IEEE, 2006.

[46] Douglas E Johnston and Petar M Djurić. The science behind risk management. *IEEE Signal Processing Magazine*, 28(5):26–36, 2011.

[47] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–193. International Society for Optics and Photonics, 1997.

[48] Michael Kalandros. Covariance control for multisensor systems. *IEEE Transactions on Aerospace and Electronic Systems*, 38(4):1138–1157, 2002.

[49] Rudolph Emil Kalman et al. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[50] Sheen T Kassouf. *Normative decision making.* Prentice Hall, 1970.

[51] Fotios Katsilieris, Hans Driessen, and Alexander Yarovoy. Radar resource management for improved situational awareness. In *Radar Conference (Radar), 2014 International*, pages 1–6. IEEE, 2014.

[52] Fotios Katsilieris, Hans Driessen, and Alexander Yarovoy. A simple sensor management criterion for radar beam control based on the cb-member filter. In *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6. IEEE, 2015.

[53] Steven M Kay. *Fundamentals of statistical signal processing. Part II Detection theory.* Prentice-Hall, Inc., 1993.

[54] Frank H Knight. *Risk, uncertainty and profit.* Hart, Schaffner & Marx, New York, 1921.

[55] Wolfgang Koch. Adaptive parameter control for phased-array tracking. In *Signal and Data Processing of Small Targets 1999*, volume 3809, pages 444–455. International Society for Optics and Photonics, 1999.

[56] Wolfgang Koch. *Tracking and sensor data fusion.* Springer, 2016.

[57] Chris Kreucher, Alfred O Hero, and Keith Kastella. A comparison of task driven and information driven sensor management for target tracking. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 4004–4009. IEEE, 2005.

[58] Edouard Kujawski and Gregory A Miller. Quantitative risk-based analysis for military counterterrorism systems. *Systems Engineering*, 10(4):273–289, 2007.

[59] L. R. M. Lacle and J. N. Driessen. Velocity-based track discrimination algorithms. In *IEE Target Tracking: Algorithms and Applications (Ref. No. 2001/174)*, volume 1, pages 4/1–4/4 vol.1, Oct 2001.

[60] Steven M LaValle. *Planning algorithms*. Cambridge university press, 2006.

[61] KD Lee and J Lunas. Hybrid model for intent estimation. In *Sixth International Conference of Information Fusion, 2003. Proceedings of the*, volume 2, pages 1215–1222. IEEE, 2003.

[62] C. M. Leon, L. S. Mihaylova, and H. Driessen. Tracking of interacting targets. In *2017 20th International Conference on Information Fusion (Fusion)*, pages 1–8, July 2017.

[63] Dennis V Lindley. The 1988 wald memorial lectures: the present position in bayesian statistics. *Statistical Science*, pages 44–65, 1990.

[64] Dennis Victor Lindley. *Bayesian statistics, a review*, volume 2. SIAM, 1972.

[65] Thomas W Lucas. Damage functions and estimates of fratricide and collateral damage. *Naval Research Logistics (NRL)*, 50(4):306–321, 2003.

[66] R Duncan Luce and Howard Raiffa. Games and decisions: Introduction and critical survey. 1957.

[67] R. Mahler. PHD filters of higher order in target number. *IEEE Transactions on Aerospace and Electronic Systems*, 43(4):1523–1543, oct 2007.

[68] Ronald PS Mahler. Multitarget bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic systems*, 39(4):1152–1178, 2003.

[69] Ronald PS Mahler. Target preference in multitarget sensor management: A unified approach. In *Defense and Security*, pages 210–221. International Society for Optics and Photonics, 2004.

[70] Ronald PS Mahler. *Statistical multisource-multitarget information fusion.* Artech House, Inc., 2007.

[71] Peter S Maybeck. *Stochastic models, estimation, and control*, volume 1. Academic press, 1979.

[72] P.S. Maybeck. *Stochastic Models, Estimation, and Control*, volume 3. Academic press, 1982.

[73] Raymond McKendall. Statistical decision theory for sensor fusion. *Technical Reports (CIS)*, page 575, 1990.

[74] David Middleton and Raffaele Esposito. Simultaneous optimum detection and estimation of signals in noise. *IEEE transactions on information theory*, 14(3):434–444, 1968.

[75] JE Moyal. The general theory of stochastic population processes. *Acta mathematica*, 108(1):1–31, 1962.

[76] Alexey Narykov, Emmanuel Delande, Daniel Clark, Paul Thomas, and Yvan Petillot. Second-order statistics for threat assessment with the PHD filter. In *2017 Sensor Signal Processing for Defence Conference (SSPD).* IEEE, dec 2017.

[77] Alexey S Narykov, Oleg A Krasnov, and Alexander Yarovoy. Effectiveness-based radar resource management for target tracking. In *2014 International Radar Conference*, pages 1–5. IEEE, 2014.

[78] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior, 2nd ed.* Princeton University Press, 1947.

[79] Jan Gerhard Norstrom. The use of precautionary loss functions in risk analysis. *IEEE Transactions on reliability*, 45(3):400–403, 1996.

[80] MG Oxenham. Enhancing situation awareness for air defence via automated threat analysis. In *Information Fusion, 2003. Proceedings of the Sixth International Conference of*, volume 2, pages 1086–1093. IEEE, 2003.

[81] Dimitri Papageorgiou and Maxim Raykin. A risk-based approach to sensor resource management. In *Advances in Cooperative Control and Optimization*, pages 129–144. Springer, 2007.

[82] GW Pulford. Taxonomy of multiple target tracking methods. *IEE Proceedings-Radar, Sonar and Navigation*, 152(5):291–304, 2005.

[83] Howard Raiffa. Applied statistical decision theory. 1974.

[84] Donald Reid. An algorithm for tracking multiple targets. *IEEE transactions on Automatic Control*, 24(6):843–854, 1979.

[85] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House, 2003.

[86] Branko Ristic. *Particle filters for random set models*, volume 798. Springer.

[87] Branko Ristic, Ba-Ngu Vo, and Daniel Clark. A note on the reward function for phd filters with sensor control. *IEEE Transactions on Aerospace and Electronic Systems*, 47(2):1521–1529, 2011.

[88] Branko Ristic, Ba-Tuong Vo, Ba-Ngu Vo, and Alfonso Farina. A tutorial on bernoulli filters: theory, implementation and applications. *IEEE Transactions on Signal Processing*, 61(13):3406–3430, 2013.

[89] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation.* Springer Science & Business Media, 2007.

[90] Jean Roy, Stéphane Paradis, and Mohamad Allouche. Threat evaluation for impact assessment in situation analysis systems. In *Signal processing, sensor fusion, and target recognition XI*, volume 4729, pages 329–341. International Society for Optics and Photonics, 2002.

[91] Saikat Saha, Yvo Boers, Hans Driessen, Pranab Kumar Mandal, and Arunabha Bagchi. Particle based map state estimation: A comparison. In *2009 12th International Conference on Information Fusion*, pages 278–283. IEEE, 2009.

[92] DJ Salmond, NO Everett, and NJ Gordont. Target tracking and guidance using particles. In *Proceedings of the 2001 American Control Conference.(Cat. No. 01CH37148)*, volume 6, pages 4387–4392. IEEE, 2001.

[93] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[94] Mark J Schervish. *Theory of statistics.* Springer Science & Business Media, 2012.

[95] I. Schlangen, E. D. Delande, J. Houssineau, and D. E. Clark. A second-order phd filter with mean and variance in target number. *IEEE Transactions on Signal Processing*, 66(1):48–63, Jan 2018.

[96] Isabel Schlangen, Daniel E Clark, and Emmanuel D Delande. Single-cluster phd filter methods for joint multi-object filtering and parameter estimation. *arXiv preprint arXiv:1705.05312*, 2017.

[97] Stanley F Schmidt. The kalman filter-its recognition and development for aerospace applications. *Journal of Guidance and Control*, 4(1):4–7, 1981.

[98] Dominic Schuhmacher, Ba-Tuong Vo, and Ba-Ngu Vo. A consistent metric for performance evaluation of multi-object filters. *IEEE Transactions on Signal Processing*, 56(8):3447–3457, 2008.

[99] Seymour Sherman. Non-mean-square error criteria. *IRE Transactions on Information Theory*, 4(3):125–126, 1958.

[100] Josef Shinar and Turetsky Vladimir. What happens when certainty equivalence is not valid?: Is there an optimal estimator for terminal guidance? *Annual Reviews in Control*, 27(2):119–130, 2003.

[101] Herbert A Simon, George B Dantzig, Robin Hogarth, Charles R Plott, Howard Raiffa, Thomas C Schelling, Kenneth A Shepsle, Richard Thaler, Amos Tversky, and Sidney Winter. Decision making and problem solving. *Interfaces*, 17(5):11–31, 1987.

[102] R Singer and R Sea. New results in optimizing surveillance system tracking and data correlation performance in dense multitarget environments. *IEEE Transactions on Automatic Control*, 18(6):571–582, 1973.

[103] Fred A Spiring. The reflected normal loss function. *Canadian Journal of Statistics*, 21(3):321–330, 1993.

[104] Dietrich Stoyan, Wilfrid S Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 1995.

[105] Jeffrey K Uhlmann. Algorithms for multiple-target tracking. *American Scientist*, 80(2):128–141, 1992.

[106] Hal R Varian. A bayesian approach to real estate assessment. *Studies in Bayesian econometric and statistics in Honor of Leonard J. Savage*, pages 195–208, 1975.

[107] B-N Vo and W-K Ma. The gaussian mixture probability hypothesis density filter. *IEEE Transactions on signal processing*, 54(11):4091–4104, 2006.

[108] B-N Vo, Sumeetpal Singh, and Arnaud Doucet. Sequential monte carlo methods for multitarget filtering with random finite sets. *IEEE Transactions on Aerospace and electronic systems*, 41(4):1224–1245, 2005.

[109] B. T. Vo, B. N. Vo, and A. Cantoni. Analytic implementations of the cardinalized probability hypothesis density filter. *IEEE Transactions on Signal Processing*, 55(7):3553–3567, July 2007.

[110] Ba-ngu Vo, Mahendra Mallick, Yaakov Bar-shalom, Stefano Coraluppi, Richard Osborne III, Ronald Mahler, and Ba-tuong Vo. *Multitarget Tracking*, pages 1–15. American Cancer Society, 2015.

[111] Ba-Ngu Vo, Ba-Tuong VO, and Daniel Clark. *Bayesian Multiple Target Filtering Using Random Finite Sets*, chapter 3, pages 75–126. John Wiley and Sons, Ltd, 2014.

[112] Ba Tuong Vo. *Random finite sets in multi-object filtering.* PhD thesis, University of Western Australia, 2008.

[113] Abraham Wald. Contributions to the theory of statistical estimation and testing hypotheses. *The Annals of Mathematical Statistics*, 10(4):299–326, 1939.

[114] Abraham Wald. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, pages 265–280, 1945.

[115] Daling Wen and Martin S Levy. Blinex: A bounded asymmetric loss function with application to bayesian estimation. *Communications in Statistics-Theory and Methods*, 30(1):147–153, 2001.

[116] Jason L Williams. *Information theoretic sensor management.* PhD thesis, Massachusetts Institute of Technology, 2007.

[117] Jason L Williams. Marginal multi-bernoulli filters: Rfs derivation of mht, jipda, and association-based member. *IEEE Transactions on Aerospace and Electronic Systems*, 51(3):1664–1687, 2015.

[118] Mark Witkowski, Gareth White, Panos Louvieris, Gokce Gorbil, Erol Gelenbe, and Lorraine Dodd. High-level information fusion and mission planning

in highly anisotropic threat spaces. In *11th International Conference on Information Fusion, 2008*, pages 1–8. IEEE, 2008.

[119] JH Zwaga and H Driessen. Tracking performance constrained mfr parameter control: applying constraints on prediction accuracy. In *Information Fusion, 2005 8th International Conference on*, volume 1, pages 6–pp. IEEE, 2005.