A NEW PARADIGM FOR ROBUST COMBINATORIAL OPTIMIZATION: USING PERSISTENCE AS A THEORY OF EVIDENCE

by

Tod Morrison

B.A., Metropolitan State College of Denver, 1995

M.S., University of Colorado Denver, 2007

A thesis submitted to the

University of Colorado Denver

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Applied Mathematics

2010

This thesis for the Doctor of Philosophy

degree by

Tod Morrison

has been approved

by

Harvey J. Greenberg

Stephen Billups

Alexander Engau

William Hart

Weldon Lodwick

Burt Simon

Date

Morrison, Tod (Ph.D., Applied Mathematics)

A New Paradigm for Robust Combinatorial Optimization:
Using Persistence as a Theory of Evidence

Thesis directed by Professor Harvey J. Greenberg

ABSTRACT

Robust optimization is a name given to a collection of mathematical programming techniques for proactively addressing uncertainties in mathematical optimization problems. This dissertation adds to this collection a new paradigm for combinatorial optimization by which the persistence of decisions is treated as evidence of robustness. We begin by introducing necessary terminology and developing the conceptual foundations for our notion of persistence. We then review the necessary elements of Dempster-Shafer Theory before presenting the main thesis: by treating persistence as evidence, we can select solutions based on their evidential robustness. Finally, we demonstrate the potential of this new approach with a discussion of some applications.

Central to this paradigm is the notion that by looking at how individual decisions (selections and rejections) persist across a given set of "acceptable" solutions, we can find solutions that are more robust to uncertainties in the underlying model. This set of solutions may come from any of a number of sources, such as expert opinion or historical decisions, or may be generated by algorithmic means. We develop one such algorithmic approach that produces a rank ordered list of optimal and near-optimal solutions by a method of successive exclusions. Once we have a collection of solutions to examine, we can model the "how" of decision persistence through a variety means: by looking at individual decision independently or by looking at sets of decisions and how they arise as complements or substitutes; and by considering selection and rejections alone or in combination.

Using the Dempster-Shafer Theory of Evidence, we construct a formal foundation in which we treat persistence as evidence of robustness, which we in turn use to build belief functions. Then, using these belief functions, we investigate conditioning on prior knowledge, and methods for combining persistence evidence from multiple sources (or multiple objectives) using various rules of combination, including an extension to Shafer's discount and combine method. Further, we investigate relations to Hamming distance and information theory. Finally, we apply our approach to two applications: the sensor placement problem, and portfolio tracking.

This abstract accurately represents the content of the candidate's thesis. I recommend its publication.

Signed ______________________________

Harvey J. Greenberg

ACKNOWLEDGMENT

First and foremost, I want to thank Dr. Harvey Greenberg who, more than simply being my thesis advisor, has been my mentor and, most importantly, my friend. Without Harvey's guidance and constant support throughout my graduate career, this dissertation could not have been completed. Additionally, I want to thank my committee: Dr. Steve Billups, Dr. Alex Engau, Dr. Bill Hart, Dr. Weldon Lodwick, and Dr. Burt Simon for their valuable input — most notably Steve and Bill for their thorough and careful review.

Finally, I would like to thank my parents for their continual support and encouragement; and Brian Peisley, Ron Nelson, and Andra Ferrara for their willingness to play sounding board and readiness to nag when necessary.

Tod Morrison
June 2010

CONTENTS

FIGURES

TABLES

Table

1. Introduction

Nearly all decision problems are plagued by some form of uncertainty. In mathematical models, this uncertainty may appear as erroneous, incomplete, or noisy data. This presents a number of interesting challenges to mathematical programming that have been addressed by a variety of techniques, both *reactive* and *proactive*. Reactive techniques, such as sensitivity analysis, are concerned with understanding how a particular solution changes under small perturbations in the data. Proactive techniques, instead, attempt to anticipate uncertainties and provide a solution that is less sensitive to changes in the data.

The mathematical notion of robustness has evolved since its origins in statistical decision theory [29]. As mathematical programming developed in the 1950's, three paradigms emerged: Mean-risk [23], Recourse [9], and Chance-constrained [8]. Building on the recourse model, Rosenhead [1968] proposed taking suboptimal decisions at an early stage in order to gain *flexibility* at subsequent stages. That was a non-mathematical proposal, and not all situations have opportunities for recourse. One example, which we explore in depth, is assigning sensors to placements in a water distribution system to detect the injection of a contaminant. A sensor, once affixed, cannot be moved — hence, there is no recourse. Another notion of robustness originates with robust statistics — that is, *volatility*. This was proposed by Mulvey, Vanderbei, and Zenios [24], using an objective function that rewarded expected returns and penalized variance. During the past decade, robust optimization has transformed from a special case designed to capture the discouragement of catastrophic outcomes of low probability towards a unifying framework of the classical paradigms (see Kouvelis and Yu [20] and Ben-Tal and Nemirovski [2, 3]). (Also see Greenberg [13, 14] for more details and an extensive bibliography, plus Greenberg and Morrison [15] for a broad range of robustness notions.)

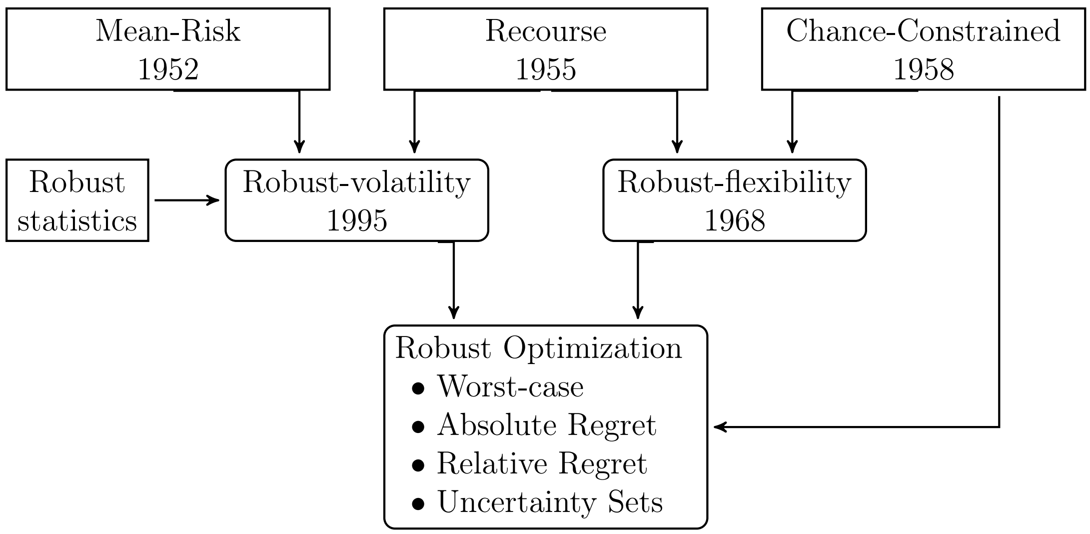


Figure 1.1: Chronology of stochastic programming models.

Common to all these approaches, a robust decision is a policy chosen in the presence of uncertainty, in which a trade-off is made between maximizing reward and minimizing risks inherent in its implementation. Robust optimization is a term given to a broad collection of mathematical programming methods that use some combination of algorithmic and modeling techniques to find robust solutions to uncertain optimization problems. This active approach to managing uncertainty places robust optimization into the category of proactive techniques for optimization under uncertainty.

This thesis adds to the modeling paradigms in three principal ways:

- Probability need not be the underlying measure of uncertainty.
- Worst-case outcomes could be ruled out separately, and remaining solutions are robust if they impart some *confidence* in their implementation. The greater the confidence, the more robust the solution.
- A decision-maker's confidence grows with greater *persistence* of solutions.

Our starting point is a given set of acceptable solutions, $X^K = \{x^1, \ldots, x^K\}$, from which we want to infer a robust solution. In particular, X^K can be an objective-ranked list, though we consider other ways to choose X^K. Intuitively, if an asset is selected in not only an optimal solution, but also in all of the 10-best solutions, we think this is a robust selection.

Brown et al. [5] recognized the impact of persistence on confidence in a computed solution. As one solves a series of scenarios for a model, with modest changes in the data, one expects to see similar solutions. That is a form of robustness that we call *confidence* in the solutions computed for implementation. They proceed to elaborate on how to take persistence into account for any mathematical program. The developments in this thesis are consistent with their goals, but we use a different approach than the elastic constraints given by Brown et al. [6].

Glover [12] introduced the use of persistence, which he called *consistency*, in variable assignment in the context of using the information to construct good solutions. Adams et al. [1] extended this by proving some properties of persistence, again for the purpose of developing a better solution algorithm. Slaney and Walsh [28] studied how persistence affects the computational complexity of combinatorial optimization problems in what they call the “backbone”.

The move from using persistence for solution computation to its impact on models with uncertainty came with Walsh [30] and was extended by Manandhar et al. [22]. Their observations about values persisting under different scenarios is related to our work, but they did not connect it to robust optimization. As far as we can tell, the first connection of persistence with robust optimization was by Bertsimas et al. [4]. They are concerned with finding persistence of a variable’s value and using that frequency as a probability measure for robustness. Our work differs from theirs in that we focus on how persistence can be used as *evidence*, and we do not limit it to single variables. We develop a new theory of robustness based on persistence, and we show how it applies to a variety of problems.

This approach has some boundaries, so it may not be the best choice for some decision-making situations. Our basic model is combinatorial: select a subset, whose reward (or cost) is uncertain. We assume there is no recourse possible. Once a subset is selected, it is implemented, and the decision-makers have no further control over outcomes (except to cross their fingers). Our focus is on notions of confidence and how persistence offers evidence that builds confidence. Besides building intuition for this approach, we propose a mathematical foundation, which we illustrate with familiar applications — knapsack problem and portfolio selection. We also go deeper into the sensor placement problem, giving numerical results to demonstrate how this approach could be used.

The rest of this thesis is organized as follows. Chapter 2 builds intuition for laying the formal foundation of persistence, and Chapter 3 presents what we need from the Dempster-Shafer Theory of Evidence. Then, Chapter 4 presents the main thesis: *persistence is evidence, and we can select solutions based on their evidential robustness.* Chapter 5 goes deeper into some applications to demonstrate the efficacy of this approach, and Chapter 6 presents conclusions and avenues for further research. See the *Mathematical Programming Glossary* [16] for basic terms; new terms are defined as needed.

2. Conceptual Foundations of Persistence

Most of our decisions are affected by a host of uncertainties. Situations of incomplete or erroneous data, unclear and possibly conflicting objectives and goals, and the potential for unforeseen complications in implementation warrant *robust* solutions. Robust solutions are designed to mitigate uncertainty, regardless of source. In this chapter, we develop the foundations for an approach to mathematical programming that define robustness by considering the extent of agreement among a collection of "acceptable" solutions (for instance, optimal and near-optimal).

2.1 Persistence of Selections

We begin our development with a discussion of some basic concepts and terminology. In order to ground and give context to this discussion, let us start with a simple example. Suppose, like a character from an Indiana Jones adventure, we have discovered the location of a legendary treasure room. Of course, like any good treasure room, this one is largely inaccessible and protected by all kinds of traps and pitfalls, so we will take only what we can fit in one small rucksack. Given that we will need to quickly decide what to take once we enter the room, we want to have a definite plan; we need to solve a knapsack (or perhaps, rucksack) problem. Unfortunately, due to the vagaries of historical records, we can only speculate on what we will find (some things may be lighter or heavier, or more or less valuable than expected), so our solution should allow for some extra flexibility.

Suppose after researching the likely contents of the treasure room our best guess at their weights and values gives the following knapsack problem:

$$\begin{aligned} &\max\ cx \\ &\text{s.t.}\quad ax \leq b, \\ &\qquad\quad x \in \{0,1\}^n, \end{aligned} \tag{2.1}$$

where the values, weights, and bag limit are

$$\begin{aligned} c &= [\ 44\ \ 49\ \ 39\ \ 43\ \ 11\ \ 10\ \ 10\ \ 41\ \ 5\ \ 25\ ] \\ a &= [\ 47\ \ 52\ \ 42\ \ 39\ \ 16\ \ \ 5\ \ 13\ \ 43\ \ 1\ \ 30\ ] \\ b &= 107. \end{aligned}$$

Our first step is to generate a collection of "acceptable" solutions. One approach is to use a method of *successive exclusions* to generate a collection of optimal and near-optimal solutions. We do this by first obtaining an initial optimal solution, x^1, with objective value z^1. For each x^k, we define the selection set: $S^k = \{i \in \mathcal{I} : x_i^k = 1\}$, where $\mathcal{I} = \{1, \ldots, n\}$ is the set of decisions corresponding to the decision variables, x_i. Then, starting with $k = 1$, we obtain S^{k+1} by solving our problem with the added constraints:

$$\sum_{i \in S^j} x_i \leq \left|S^j\right| - 1 \text{ for } j = 1, \ldots, k,$$

where $|S^j|$ is the magnitude of the selection set S^j. We stop when:

- (2.1) becomes infeasible;
- a specified maximum number of solutions is generated; or,
- the objective value falls below some threshold, such as $z^K \leq \alpha z^1$ for some $\alpha \in (0\,,1)$.

Stepping through in this way, we find alternative optima, if any, and continue generating solutions, allowing the objective value to decrease ensuring that no solutions are regenerated. Other stopping rules are possible, but, whatever rule is chosen, the result is $X^K = \{x^1, \ldots, x^K\}$, for some $K \geq 1$.

The rate at which the objective value changes as we generate new solutions is one measure of interest; a flat function implies that there are at least K alternative optima. We can see the result of applying this process to our treasure-room knapsack problem in Table 2.1, where the last row shows the total number of times each object is selected. Intuitively, object 6 seems highly preferred, while object 10 is never selected.

We define the *consensus set* is defined as:

$$C^k = S^1 \cap S^2 \cap \cdots \cap S^k. \tag{2.2}$$

We are interested in how the consensus set changes as alternative solutions are generated. Table 2.2 shows the consensus set for the best 10 knapsack solutions. Since the consensus set is defined in terms of set intersections, the set cannot increase as solutions are added. In fact, for our example, $|C^k| = 0$ for $k > 7$. Moreover, this process is highly order-dependent; in our example, since $z^4 = z^5 = z^6$, the fact that $|C^4| = 3$ is simply an effect of the order in which the

Table 2.1: Treasure-room knapsack solution sets.

k	Objective Value	Knapsack Assignment x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	112	1	0	0	1	0	1	1	0	1	0
2	110	0	0	0	1	1	1	0	1	1	0
3	109	0	0	0	1	0	1	1	1	1	0
4	108	0	0	1	1	1	1	0	0	1	0
5	108	1	1	0	0	0	1	0	0	1	0
6	108	1	0	0	1	1	1	0	0	0	0
7	107	0	0	1	1	0	1	1	0	1	0
8	107	0	1	0	1	0	0	1	0	1	0
9	107	1	0	0	1	0	1	1	0	0	0
10	107	0	1	0	1	0	1	0	0	1	0
11	106	0	0	1	0	1	1	0	1	1	0
12	105	0	0	1	0	0	1	1	1	1	0
13	105	0	1	0	0	0	1	0	1	1	0
14	105	0	0	0	1	1	1	0	1	0	0
	Total	4	4	4	10	5	13	6	6	11	0

solutions are generated. On the other hand, the relative slowness with which the optimal value decreases suggests that some form of robustness may be gained by considering near-optimal solutions — for instance, there may be near-optimal solutions that are less sensitive to variations in model parameters.

We can address the issue of order-dependence by considering the notion of *breakpoints*. Formally, breakpoints are values $k = b_0, b_1, \ldots$ in the sequence for which the objective value decreases (for maximization). For a maximization problem like the treasure-room knapsack problem, we begin with $b_0 = 1$ and recursively define $b_i \stackrel{\text{def}}{=} \min\{k : z^k < z^i\}$ while adopting the convention that

Table 2.2: Consensus sets for treasure-room knapsack problem.

k	1	2	3	4	5	6	7	8	9	10
z	112	110	109	108	108	108	107	107	107	107
x_1	✓									
x_4	✓	✓	✓	✓		*				
x_6	✓	✓	✓	✓	✓	✓	✓		*	*
x_7	✓									
x_9	✓	✓	✓	✓	✓					
$\|C\|$	5	3	3	3	2	1	1	0	0	0

(alternative consensus set elements are marked with asterisks)

$b_i \stackrel{\text{def}}{=} K+1$ if $z^k \geq z^i$ for all k. The recursion stops with m breakpoint regions:

$$\{1,\ldots,b_1-1\},\ \{b_1,\ldots,b_2-1\},\ \ldots,\ \{b_{m-1},\ldots,K\},$$

and $m+1$ breakpoints:

$$\begin{aligned}
b_0 &= 1 \\
b_1 &= \min\{k\ :\ z^k < z^1\} \\
&\vdots \\
b_i &= \min\{k\ :\ z^k < z^{b_{i-1}}\} \\
&\vdots \\
b_{m-1} &= \min\{k\ :\ z^k = z^K\} \\
b_m &= K+1.
\end{aligned}$$

This is illustrated graphically in Figure 2.1.

We define persistence as the continued presence of a selection s in the

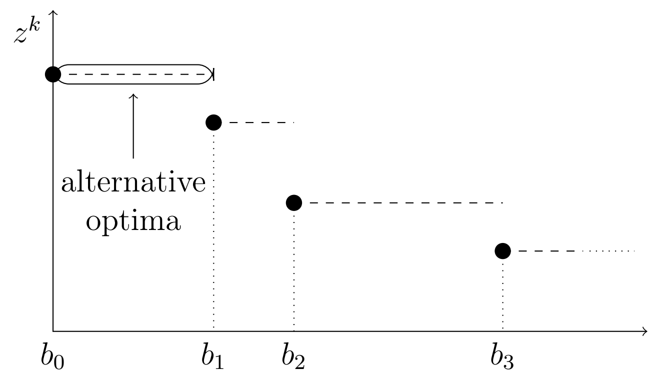


Figure 2.1: Breakpoints for the objective value.

collection of selection sets. One form, if there are alternative optima, is the percentage of solutions in the optimality region that contain s, that is $\left|\{k \in 1,\ldots,b_1-1 \,:\, x_s = 1\}\right|/(b_1-1)$. A stronger form of persistence is a measure of how long a selection remains in the consensus set. Because this can depend on the order in which the optimal-set is generated, we modify our definition of consensus set to only consider breakpoint regions in their entirety.

Formally, we define the *multisets*:

$$\mathcal{S}^i = \biguplus_{b_{i-1} \le k < b_i} S^k \text{ for } i = 1,\ldots,m. \tag{2.3}$$

Let $\mathcal{N}(x,\mathcal{X})$ denote the number of occurrences of element x in multiset $\mathcal{X}$. In particular, $\mathcal{N}(s,\mathcal{S}^1)$ is the number of optimal solution sets that contain the selection s. More generally,

$$\mathcal{N}(s,\mathcal{S}^i) = \left|\{k : s \in S^k \text{ for } b_{i-1} \le k < b_i\}\right| \tag{2.4}$$

gives the number of solution sets in the i-th breakpoint region that contain selection s.

For example, looking at the knapsack solution sets we see that the first decision, x_1, is selected twice in the fourth multiset (the first with multiple solutions), $\mathcal{S}^4$. Hence,

$$\mathcal{N}(1, \mathcal{S}^4) = \left|\{k : 1 \in S^k \text{ for } b_3 \leq k < b_4\}\right| = 2.$$

Now define the *consensus multisets* by their intersection:

$$\mathcal{C}^k = \mathcal{S}^1 \cap \cdots \cap \mathcal{S}^k \text{ for } k = 1, \ldots, K. \tag{2.5}$$

The number of occurrences satisfies $\mathcal{N}(s, \mathcal{C}^i) = \min\{\mathcal{N}(s, \mathcal{C}^{i-1}), \mathcal{N}(s, \mathcal{S}^i)\}$. If the solution set for each breakpoint region is unique, we have $b_i = i + 1$ and $\mathcal{S}^i = S^i$, in which case the consensus multiset reduces to the consensus set.

The *local persistence* of a selection s is its frequency within a breakpoint region:

$$\rho_i(s) = \frac{\mathcal{N}(s, \mathcal{S}^i)}{b_i - b_{i-1}}. \tag{2.6}$$

The numerator is the number of optimal or near-optimal solution sets that contain s, while the denominator is the number of solution sets in the breakpoint region, so $\rho_i(s) \in [\,0\,, 1\,]$. The greater the value of $\rho_i(s)$ the more persistent selection s is in the i-th breakpoint region.

Although there may be some value to considering the local persistence of a selection, particularly within the first breakpoint region when there are many alternate optima, this can be overly restrictive when the optimal value decreases relatively slowly. For instance, in our example problem we see a total drop in objective value of less than 7% between the first and fourteens solutions. In such cases, the extent of the data uncertainties can be more concerning than a slight loss in objective value, and the possibility of discovering highly persistent selections that might not be identified when considering only the first breakpoint warrants considering a broader notion of persistence.

The *global persistence* of selection s is its frequency for the horizon:

$$\gamma(s) = \frac{\mathcal{N}(s, \uplus_{1 \leq k \leq K} S^k)}{K}. \tag{2.7}$$

In this case, the numerator is the number of optimal and near-optimal solution sets that contain s across all breakpoints, or equivalently, across all K generated solutions.

To illustrate this, Table 2.3 extends the list of generated solutions from Table 2.1 by including the global persistence values for each of the items. Considering the global persistence values of the items, we see that x^3 satisfies the knapsack constraint and maximizes the total persistence, $\sum_s \gamma(s)$, with a solution value of 109, only slightly less than the optimal solution.

Looking at the global persistence value of an individual selection in isolation fails to capture interaction effects, so we must extend the notion of persistence to consider subsets of S. In the context of robust decision-making, we may feel confident in choosing s if $\gamma(s)$ is relatively large, but we want to have confidence in the entire set of selections. Two selections could be *substitutes*, where the

Table 2.3: Treasure-room knapsack solution sets with global persistence values.

k	Objective Value	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\sum_s \gamma(s)$
1	112	1	0	0	1	0	1	1	0	1	0	$^{44}/_{14}$
2	110	0	0	0	1	1	1	0	1	1	0	$^{45}/_{14}$
3	109	0	0	0	1	0	1	1	1	1	0	$^{46}/_{14}$
4	108	0	0	1	1	1	1	0	0	1	0	$^{43}/_{14}$
5	108	1	1	0	0	0	1	0	0	1	0	$^{32}/_{14}$
6	108	1	0	0	1	1	1	0	0	0	0	$^{32}/_{14}$
7	107	0	0	1	1	0	1	1	0	1	0	$^{44}/_{14}$
8	107	0	1	0	1	0	0	1	0	1	0	$^{31}/_{14}$
9	107	1	0	0	1	0	1	1	0	0	0	$^{33}/_{14}$
10	107	0	1	0	1	0	1	0	0	1	0	$^{38}/_{14}$
11	106	0	0	1	0	1	1	0	1	1	0	$^{39}/_{14}$
12	105	0	0	1	0	0	1	1	1	1	0	$^{40}/_{14}$
13	105	0	1	0	0	0	1	0	1	1	0	$^{34}/_{14}$
14	105	0	0	0	1	1	1	0	1	0	0	$^{34}/_{14}$
	$\gamma(v)$	$\frac{4}{14}$	$\frac{4}{14}$	$\frac{4}{14}$	$\frac{10}{14}$	$\frac{5}{14}$	$\frac{13}{14}$	$\frac{6}{14}$	$\frac{6}{14}$	$\frac{11}{14}$	0	

inclusion of one excludes the other. Or, they could be *complements*, where the inclusion of one includes the other — for instance, in our example solution list the second and third items both have global persistence values of $^4/_{11}$ but are never selected together, which we should consider when selecting based on persistence.

We address this issue by defining the *global selection persistence factor* of a set as:

$$\Gamma(S) = \frac{\left|\{k : S \subseteq S^k \text{ for } 1 \leq k \leq K\}\right|}{K} \text{ for } S \neq \emptyset \quad \text{and} \quad \Gamma(\emptyset) = 0.$$

In particular, $\Gamma(\{s\}) = \gamma(s)$. More generally, we have the following.

Theorem 2.1. $\Gamma(S) \leq \min_{s \in S} \gamma(s)$.

Proof: We have,

$$\left|\{k : S \subseteq S^k \text{ for } 1 \leq k \leq K\}\right| \leq \left|\{k : s \in S^k \text{ for } 1 \leq k \leq K\}\right|$$

for each $s \in S$. Equivalently, $\Gamma(S) \leq \gamma(s)$ for each $s \in S$. ■

Theorem 2.1 says that the global selection persistence factor of a subset cannot exceed the global persistence of any of its members. For example, considering the selection set $S = \{4, 5, 6\}$, we see that S is a subset of S^2, S^4, S^6, and S^{14}, giving it a global selection persistence factor, $\Gamma(S) = \frac{4}{14}$, while the global persistence values of the constituent selections are $\gamma(4) = \frac{10}{14}$, $\gamma(5) = \frac{5}{14}$, and $\gamma(6) = \frac{13}{14}$.

Table 2.4: Treasure-room solution sets with total global persistence factors.

k	Objective Value	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\sum_{S \subseteq S^k} \Gamma(S)$
1	112	1	0	0	1	0	1	1	0	1	0	$^{134}/_{14}$
2	110	0	0	0	1	1	1	0	1	1	0	$^{142}/_{14}$
3	109	0	0	0	1	0	1	1	1	1	0	$^{146}/_{14}$
4	108	0	0	1	1	1	1	0	0	1	0	$^{130}/_{14}$
5	108	1	1	0	0	0	1	0	0	1	0	$^{64}/_{14}$
6	108	1	0	0	1	1	1	0	0	0	0	$^{68}/_{14}$
7	107	0	0	1	1	0	1	1	0	1	0	$^{138}/_{14}$
8	107	0	1	0	1	0	0	1	0	1	0	$^{64}/_{14}$
9	107	1	0	0	1	0	1	1	0	0	0	$^{74}/_{14}$
10	107	0	1	0	1	0	1	0	0	1	0	$^{86}/_{14}$
11	106	0	0	1	0	1	1	0	1	1	0	$^{118}/_{14}$
12	105	0	0	1	0	0	1	1	1	1	0	$^{120}/_{14}$
13	105	0	1	0	0	0	1	0	1	1	0	$^{74}/_{14}$
14	105	0	0	0	1	1	1	0	1	0	0	$^{78}/_{14}$

In Table 2.4, we illustrate an application of global selection persistence factors, extending Table 2.1 to include with each solution its *total* global selection persistence factor, which we compute as the sum of Γ over all subsets of the corresponding selection set. Because of its ability to capture interactions among selections, considering the total Γ of solutions can, in some cases, allow us to distinguish among solutions with similar global persistence values. For instance, since the first and seventh solutions, x^1 and x^7, have the same total global persistence (Table 2.3), they could be seen as equally good when we consider decisions independently. However, the difference in their total global selection persistence factor (Table 2.4) suggests that x^7 is a better solution when interactions among selections are considered.

In computing the global selection persistence factor, a solution that selects many elements contributes the same weight to each as one that selects only a few. As a consequence, Γ will tend to give more weight to subsets of those solutions whose selection sets are relatively large. In the knapsack problem this would tend to favor low item-weight items. For instance, x_2, which has the highest item-weight at 52, is selected only by those solutions that select just four items and hence feasible selection sets, S, that include x_2 will have low values of $\Gamma(S)$ despite its high item-value. By contrast, while x_7 has a much lower item-value relative to its item-weight sets that include it will have higher Γ values. To addresses this situation, we propose an *adjusted global selection persistence factor* that normalizes the weight contributed by each solution.

One approach to normalizing the contribution of a solution is to define an adjusted global selection persistence factor, so that each solution in our list contributes an amount relative to the cardinality its solution set. Thus, we define the adjusted global selection persistence factor as:

$$\widehat{\Gamma}(S) = \frac{\displaystyle\sum_{\{1 \le k \le K \,:\, S \subseteq S^k\}} (2^{\left|S^k\right|} - 1)^{-1}}{K} \quad \text{for } S \neq \emptyset \quad \text{and} \quad \widehat{\Gamma}(\emptyset) = 0.$$

Note that each solution in our list of acceptable solutions contributes an equal weight, which is distributed over the subsets of that solution.

We contrast this normalized form with the unnormalized global selection persistence factor in Table 2.5. The effect of the adjustment can be seen when comparing the second and third solutions — using Γ the third solution scores higher, while with $\widehat{\Gamma}$ the second scores higher. We can explain this, in part, by the increased contribution of the fourteenth solution, which selects only four items and selects a proper subset of the items selected by the second solution.

2.2 Persistence of Rejections

As we have seen, defining persistence in terms of selections alone causes the global persistence to be greater for those solutions whose selection sets are relatively large. In the knapsack problem this favors low-weight items and does not account for the information provided by considering when items are rejected. For our example, item 1 is selected only four times; another perspective is that item 1 is *rejected* 10 times. The rejections offer information that is complementary to selections. This observation offers an alternative to using $\widehat{\Gamma}$ for dealing with the issues raised when the number of selections made varies between solutions. Instead, we address the problem with using selections alone by accounting

Table 2.5: Treasure-room solution sets with total adjusted global persistence factors.

k	Objective Value	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\sum_{S\subseteq S^k}\Gamma(S)$	$\sum_{S\subseteq S^k}\widehat{\Gamma}(S)$
		Knapsack Assignment											
1	112	1	0	0	1	0	1	1	0	1	0	$^{134}/_{14}$	0.42919
2	110	0	0	0	1	1	1	0	1	1	0	$^{142}/_{14}$	0.43779
3	109	0	0	0	1	0	1	1	1	1	0	$^{146}/_{14}$	0.43717
4	108	0	0	1	1	1	1	0	0	1	0	$^{130}/_{14}$	0.38065
5	108	1	1	0	0	0	1	0	0	1	0	$^{64}/_{14}$	0.24332
6	108	1	0	0	1	1	1	0	0	0	0	$^{68}/_{14}$	0.24762
7	107	0	0	1	1	0	1	1	0	1	0	$^{138}/_{14}$	0.39908
8	107	0	1	0	1	0	0	1	0	1	0	$^{64}/_{14}$	0.22857
9	107	1	0	0	1	0	1	1	0	0	0	$^{74}/_{14}$	0.25653
10	107	0	1	0	1	0	1	0	0	1	0	$^{86}/_{14}$	0.30876
11	106	0	0	1	0	1	1	0	1	1	0	$^{118}/_{14}$	0.33333
12	105	0	0	1	0	0	1	1	1	1	0	$^{120}/_{14}$	0.33303
13	105	0	1	0	0	0	1	0	1	1	0	$^{74}/_{14}$	0.26144
14	105	0	0	0	1	1	1	0	1	0	0	$^{78}/_{14}$	0.26083

for both rejections and selections in determining the overall persistence.

We could account for rejections by simply re-modeling with complementary decision variables, $x' = 1 - x$, and considering the persistence of selections for the complementary variables. However, despite the dependence, accounting for both selections and rejections in one framework does change the effect of final decision-making and the confidence we have in the result. An analogy, albeit imperfect, is grading by subtracting the number of wrong answers from the number of right ones, rather than just counting the percent of right ones. Another view is the statistical concept of hypothesis testing. Although there may not be enough selections to make a confident decision to select an asset, there could be enough rejections to make a confident decision to reject it; there could also be insufficient persistence of either decision, thus giving little or no confidence to either selection or rejection. More to the point, the consequences of a selection are different from that of a rejection. For example, rejecting an asset may cost a little due to its being better (in terms of lost opportunity) than what we finally choose, but selecting an asset may incur a significant downside.

Let $R^k = \neg S^k$, the rejection set, be the complement of S^k and define the global persistence of rejection:

$$\widetilde{\gamma}(r) = \frac{\mathcal{N}(r, \uplus_{1 \le k \le K} R^k)}{K}, \tag{2.8}$$

and the *global rejection persistence factor*:

$$\widetilde{\Gamma}(R) = \frac{\left|\{k \,:\, R \subseteq R^k \text{ for } 1 \le k \le K\}\right|}{K} \text{ for } R \ne \emptyset \quad \text{and} \quad \widetilde{\Gamma}(\emptyset) = 0. \tag{2.9}$$

Note that, $\widetilde{\Gamma}(\{r\}) = \widetilde{\gamma}(r)$ and, by analogy, Theorem 2.1 extends as:

$$\widetilde{\Gamma}(R) \le \min_{r \in R} \widetilde{\gamma}(r).$$

We can combine the information provided by considering the persistence of both selections and rejections in several ways. For now, consider their simple sum, the *combined global persistence factor*:

$$\Psi(S, R) = \Gamma(S) + \widetilde{\Gamma}(R) \quad \text{for} \quad S \cap R = \emptyset. \tag{2.10}$$

In particular, $\Psi(S, \emptyset) = \Gamma(S)$ and $\Psi(\emptyset, R) = \widetilde{\Gamma}(R)$.

Table 2.6: Total persistence values for example class of knapsack problems.

k	γ^k	$\widetilde{\gamma}^k$	Γ^k	$\widetilde{\Gamma}^k$	Ψ^k
1	$\frac{1}{n+1}$	$\frac{2(n-1)}{n+1}$	$\frac{1}{n+1}$	$\frac{2^{n-1}+n-2}{n+1}$	$\frac{(3)2^{n-1}+2n-4}{n+1}$
2	$\frac{(n-1)^2}{n+1}$	$\frac{n}{n+1}$	$\frac{(n+1)\,2^{n-2}-n}{n+1}$	$\frac{n}{n+1}$	$\frac{(2n+1)\,2^{n-1}-2n}{n+1}$
$3,\ldots,n+1$	$\frac{(n-2)(n-1)}{n+1}$	$\frac{n+2}{n+1}$	$\frac{(n+2)\,2^{n-3}-n}{n+1}$	$\frac{n+3}{n+1}$	$\frac{(3n+7)\,2^{n-2}-4n}{n+1}$

To illustrate these extensions, consider the class of knapsack problems for $n > 2$ given by:

$$\begin{aligned} &\max\ nx_1 + \sum_{j=2}^{n} x_j \\ &\text{s.t.}\quad (n-1)\,x_1 + \sum_{j=2}^{n} x_j \le n-1, \\ &\qquad x \in \{0,1\}^n. \end{aligned} \tag{2.11}$$

The optimal solution is $x^1 = \vec{\mathbf{e}}_1$ with $S^1 = \{1\}$, $R^1 = \{2,\ldots,n\}$, and objective value n. The second-best solution is $x^2 = \vec{\mathbf{1}} - \vec{\mathbf{e}}_1$ with an objective value of $n-1$. The next $n-1$ solutions, x^k for $k \in \{3,\ldots,n+1\}$, each have value $n-2$ and are, respectively, $x^k = \vec{\mathbf{1}} - \vec{\mathbf{e}}_1 - \vec{\mathbf{e}}_{k-1}$. Our objective-ranked list has the values shown in Table 2.6, where

$$\gamma^k = \sum_{s \in S^k} \gamma(s)$$

$$\widetilde{\gamma}^k = \sum_{r \in R^k} \widetilde{\gamma}(r)$$

$$\Gamma^k = \sum_{S \subseteq S^k} \Gamma(S)$$

$$
\begin{aligned}
&= \sum_{S\subseteq S^k:|S|=1} \Gamma(S) + \sum_{S\subseteq S^k:|S|\geq 2} \Gamma(S) \\
&= \begin{cases}
\gamma^k + 0 & = \dfrac{1}{n+1} & \text{if } k=1; \\
\gamma^k + \dfrac{1}{n+1}\displaystyle\sum_{i=2}^{n-1} \tbinom{n-1}{i}(n-i) & = \dfrac{(n+1)\,2^{n-2}-n}{n+1} & \text{if } k=2; \\
\gamma^k + \dfrac{1}{n+1}\displaystyle\sum_{i=2}^{n-2} \tbinom{n-2}{i}(n-i) & = \dfrac{(n+2)\,2^{n-3}-n}{n+1} & \text{if } 3\leq k\leq n+1.
\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\widetilde{\Gamma}^k &= \sum_{R\subseteq R^k} \widetilde{\Gamma}(R) \\
&= \sum_{R\subseteq R^k:|R|=1} \widetilde{\Gamma}(R) + \sum_{R\subseteq R^k:|R|\geq 2} \widetilde{\Gamma}(R) \\
&= \begin{cases}
\widetilde{\gamma}^k + \dfrac{2^{n-1}-(n-1)-1}{n+1} & = \dfrac{2^{n-1}+n-2}{n+1} & \text{if } k=1; \\
\widetilde{\gamma}^k + 0 & = \dfrac{n}{n+1} & \text{if } k=2; \\
\widetilde{\gamma}^k + \dfrac{1}{n+1} & = \dfrac{n+3}{n+1} & \text{if } 3\leq k\leq n+1.
\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\Psi^k &= \sum_{S\subseteq S^k} \sum_{R\subseteq R^k} \Psi(S,R) \\
&= 2^{\left|R^k\right|} \sum_{S\subseteq S^k} \Gamma(S) + 2^{\left|S^k\right|} \sum_{R\subseteq R^k} \widetilde{\Gamma}(R) \\
&= 2^{\left|R^k\right|}\Gamma^k + 2^{\left|S^k\right|}\widetilde{\Gamma}^k.
\end{aligned}
$$

In Table 2.7, we compare the rankings under various models of persistence of solutions to our example class of knapsack problems for different values of $n > 2$. For $n = 4$ and $n = 5$, the optimal solution is most persistent with respect only to the ranking of $\{\widetilde{\gamma}^k\}$ and $\{\widetilde{\Gamma}^k\}$. In all three cases, the second-best solution is the most persistent with respect to the rankings of $\{\gamma^k\}$, $\{\Gamma^k\}$, $\{\gamma^k + \widetilde{\gamma}^k\}$,

Table 2.7: Comparative rankings under various models of persistence.

n=3

k	x^k	γ^k	$\widetilde{\gamma}^k$	Γ^k	$\widetilde{\Gamma}^k$	Ψ^k
1	1 0 0	$\frac{1}{4}$	$\frac{4}{4}$	$\frac{1}{4}$	$\frac{4}{4}$	$\frac{14}{4}$
2	0 1 1	$\frac{4}{4}$	$\frac{3}{4}$	$\frac{5}{4}$	$\frac{3}{4}$	$\frac{22}{4}$
3	0 0 1	$\frac{2}{4}$	$\frac{5}{4}$	$\frac{2}{4}$	$\frac{5}{4}$	$\frac{20}{4}$
4	0 1 0	$\frac{2}{4}$	$\frac{5}{4}$	$\frac{2}{4}$	$\frac{5}{4}$	$\frac{20}{4}$

n=4

k	x^k	γ^k	$\widetilde{\gamma}^k$	Γ^k	$\widetilde{\Gamma}^k$	Ψ^k
1	1 0 0 0	$\frac{1}{5}$	$\frac{6}{5}$	$\frac{1}{5}$	$\frac{6}{5}$	$\frac{28}{5}$
2	0 1 1 1	$\frac{9}{5}$	$\frac{4}{5}$	$\frac{16}{5}$	$\frac{4}{5}$	$\frac{64}{5}$
3	0 0 1 1	$\frac{6}{5}$	$\frac{6}{5}$	$\frac{8}{5}$	$\frac{6}{5}$	$\frac{60}{5}$
4	0 1 0 1	$\frac{6}{5}$	$\frac{6}{5}$	$\frac{8}{5}$	$\frac{6}{5}$	$\frac{60}{5}$
5	0 1 1 0	$\frac{6}{5}$	$\frac{6}{5}$	$\frac{8}{5}$	$\frac{6}{5}$	$\frac{60}{5}$

n=5

k	x^k	γ^k	$\widetilde{\gamma}^k$	Γ^k	$\widetilde{\Gamma}^k$	Ψ^k
1	1 0 0 0 0	$\frac{1}{6}$	$\frac{8}{6}$	$\frac{1}{6}$	$\frac{8}{6}$	$\frac{54}{6}$
2	0 1 1 1 1	$\frac{16}{6}$	$\frac{5}{6}$	$\frac{43}{6}$	$\frac{5}{6}$	$\frac{166}{6}$
3	0 0 1 1 1	$\frac{12}{6}$	$\frac{7}{6}$	$\frac{23}{6}$	$\frac{7}{6}$	$\frac{156}{6}$
4	0 1 0 1 1	$\frac{12}{6}$	$\frac{7}{6}$	$\frac{23}{6}$	$\frac{7}{6}$	$\frac{156}{6}$
5	0 1 1 0 1	$\frac{12}{6}$	$\frac{7}{6}$	$\frac{23}{6}$	$\frac{7}{6}$	$\frac{156}{6}$
6	0 1 1 1 0	$\frac{12}{6}$	$\frac{7}{6}$	$\frac{23}{6}$	$\frac{7}{6}$	$\frac{156}{6}$

$\{\Gamma^k + \widetilde{\Gamma}^k\}$, and $\{\Psi^k\}$. The remaining solutions ($k \geq 3$) are the second-most persistent with respect to all rankings.

We further investigate the persistence rankings with respect to $\{\Psi^k\}$ by using the formulas from page 20. First, we note that for $n > 2$, the optimal solution is always the least persistent (that is, $\Psi^1 < \Psi^k$, for $k \neq 1$). We then investigate the relative rankings of Ψ^2 and Ψ^{3^+} (that is, Ψ^k for $k \geq 3$) by looking at their difference,

$$\Psi^2 - \Psi^{3^+} = \frac{(2n+1)\,2^{n-1} - 2n}{n+1} - \frac{(3n+7)\,2^{n-2} - 4n}{n+1} = \frac{(n-5)\,2^{n-2} + 2n}{n+1},$$

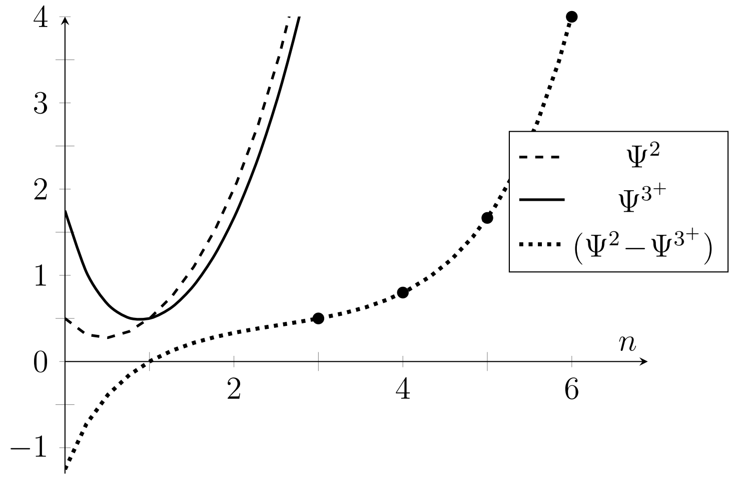


Figure 2.2: Plotting Ψ^k as a function of n.

where we see that $\Psi^2 > \Psi^{3^+}$ for $n > 2$ (illustrated graphically in Figure 2.2).

2.3 Normalization

So far in this chapter, we have developed several functions that seek to quantify the persistence of a given decision or solution. However, these functions may produce values on vastly different scales and are not readily comparable. Consider, for instance, a situation where the underlying problem is such that acceptable solutions make only a relatively few selections, rejecting most options. In this case, the total global rejection persistence factor, $\sum_{R \subseteq \neg S} \widetilde{\Gamma}(R)$, will be an order of magnitude larger than the total global selection persistence factor, $\sum_{S' \subseteq S} \Gamma(S')$. We can address these incompatibilities by normalizing the functions, rescaling their range with an appropriate divisor to produce values in the unit interval, $[\,0\,,1\,]$.

Even though any sufficiently large divisor can be used to produce values in the desired interval, the idea is to find an upper bound on the possible values that is meaningful, so some consideration of the underlying model should be given. For instance, recall that for a given decision (that is, a selection, s, or rejection, r), the global persistence values for selections and rejections are defined as

$$\gamma(s) = \frac{\mathcal{N}(s, \uplus_{1 \leq k \leq K} S^k)}{K} \qquad \text{and} \qquad \widetilde{\gamma}(r) = \frac{\mathcal{N}(r, \uplus_{1 \leq k \leq K} R^k)}{K},$$

respectively. We have $0 \leq \gamma(s) \leq 1$ and $0 \leq \widetilde{\gamma}(s) \leq 1$, so an easy upper bound for the total global persistence of selections, $\sum_{s\,:\,x_s=1} \gamma(s)$, or rejections, $\sum_{r\,:\,x_r=0} \widetilde{\gamma}(r)$, is simply the number of decision variables, n. Although this bound is non-arbitrary, it wholly disregards the information contained in the solutions set and reflects only a very coarse understanding of the underlying problem.

We can find a tighter bound by noting that the total global persistence of selections would be maximized by selecting everything if this were feasible, while the total global persistence of rejections would be maximized by rejecting all possible choices. Hence, we have

$$\max_x \sum_{s\,:\,x_s=1} \gamma(s) \leq \sum_{s=1}^{n} \gamma(s) = \frac{1}{K} \sum_{k=1}^{K} \left|S^k\right| \tag{2.12}$$

for selections, and

$$\max_x \sum_{s\,:\,x_s=1} \widetilde{\gamma}(s) \leq \sum_{s=1}^{n} \widetilde{\gamma}(r) = \frac{1}{K} \sum_{k=1}^{K} \left|R^k\right| \tag{2.13}$$

for rejections. Using the bounds from (2.12) and (2.13) as divisors, we define the normalized global persistence functions for selections and rejections as

$$\underline{\gamma}(s) = \frac{\mathcal{N}(s, \uplus_{1 \leq k \leq K} S^k)}{\sum_{k=1}^{K} |S^k|} \qquad \text{and} \qquad \underline{\widetilde{\gamma}}(r) = \frac{\mathcal{N}(r, \uplus_{1 \leq k \leq K} R^k)}{\sum_{k=1}^{K} |R^k|} \tag{2.14}$$

respectively, which give us the desired property that the total normalized global persistence values, $\sum_{s\,:\,x_s=0} \underline{\widetilde{\gamma}}(s)$ and $\sum_{r\,:\,x_r=0} \underline{\widetilde{\gamma}}(r)$, are appropriately adjusted to lie in the unit interval.

Considering the global persistence factor functions,

$$\Gamma(S) = \frac{\big|\{k \;:\; S \subseteq S^k \text{ for } 1 \leq k \leq K\}\big|}{K} \text{ for } S \neq \emptyset \quad \text{and} \quad \Gamma(\emptyset) = 0,$$

and,

$$\widetilde{\Gamma}(R) = \frac{\big|\{k \;:\; R \subseteq R^k \text{ for } 1 \leq k \leq K\}\big|}{K} \text{ for } R \neq \emptyset \quad \text{and} \quad \widetilde{\Gamma}(\emptyset) = 0,$$

we have $0 \leq \Gamma(S) \leq 1$ and $0 \leq \widetilde{\Gamma}(R) \leq 1$ for arbitrary decision sets $S, R \subseteq \mathcal{I}$. As noted though, the total global selection persistence factor, $\sum_{S' \subseteq S} \Gamma(S')$, and total global rejection persistence factor, $\sum_{R \subseteq \neg S} \widetilde{\Gamma}(R)$, may produce very large values.

As with the total global persistence functions, we can find a gross upper bound by simply examining the general form of the total Γ (or total $\widetilde{\Gamma}$) function without considering the information contained in the solutions set. Note that, since $0 \leq \Gamma(S) \leq 1$ for all $S \subseteq \mathcal{I}$ and $\Gamma(\emptyset) = 0$, the total Γ for the maximal selection set, $S = \mathcal{I}$, which sums over all $2^{|\mathcal{I}|} = 2^n$ subsets of $\mathcal{I}$, is always less than $(2^n - 1)$. By similar reasoning, we get the same bound for the total $\widetilde{\Gamma}$ using the maximal rejection set, $R = \mathcal{I}$.

Although this bound provides a sufficient divisor for rescaling both total Γ and total $\widetilde{\Gamma}$ onto the unit interval, it is a single divisor for both, so it does not fully address the very incompatibilities that we are attempting to adjust. We can improve on this bound by incorporating information from the solution set, in which case, given an arbitrary selection set $S \subseteq \mathcal{I}$ and rejection set $R \subseteq \mathcal{I}$, we have

$$\sum_{S' \subseteq S} \Gamma(S') \leq \frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|S^k\right|} - 1\right) \tag{2.15}$$

for global selection persistence factor, and

$$\sum_{R' \subseteq R} \widetilde{\Gamma}(R') \leq \frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|\neg S^k\right|} - 1\right) \tag{2.16}$$

for global rejection persistence factor.

These bounds can be significantly tighter than the gross bound, $(2^n - 1)$. Consider the knapsack problems from Section 2.2. For rejections we have,

$$\begin{aligned}
\sum_{R \subseteq \neg S} \widetilde{\Gamma}(R) &= \frac{1}{n+1} \sum_{k=1}^{n+1} \left(2^{\left|R^k\right|} - 1\right) \\
&= \frac{1}{n+1} \left[\left(2^{n-1} - 1\right) + \left(2^1 - 1\right) + \sum_{k=2}^{n} \left(2^2 - 1\right)\right] \\
&= \frac{1}{n+1} \left[2^{n-1} + 3(n-1)\right] < 2^n - 1.
\end{aligned}$$

Thus, for large n the gross bound will be on the order of $(2n + 2)$ times the bound given in (2.16).

Using the bounds in (2.15) and (2.16), define the normalized global persistence factor functions:

$$\underline{\Gamma}(S) = \frac{\left|\{k \,:\, S \subseteq S^k \text{ for } 1 \le k \le K\}\right|}{\sum_{k=1}^{K} \left(2^{\left|S^k\right|} - 1\right)} \text{ for } S \neq \emptyset \quad \text{and} \quad \underline{\Gamma}(\emptyset) = 0, \qquad (2.17)$$

and,

$$\widetilde{\underline{\Gamma}}(R) = \frac{\left|\{k \,:\, R \subseteq R^k \text{ for } 1 \le k \le K\}\right|}{\sum_{k=1}^{K} \left(2^{\left|\neg S^k\right|} - 1\right)} \text{ for } R \neq \emptyset \quad \text{and} \quad \widetilde{\underline{\Gamma}}(\emptyset) = 0. \qquad (2.18)$$

We can compare these normalized forms to the inherently normalized adjusted global selection persistence factor,

$$\widehat{\Gamma}(S) = \frac{\displaystyle\sum_{\{1 \le k \le K \,:\, S \subseteq S^k\}} (2^{\left|S^k\right|} - 1)^{-1}}{K} \text{ for } S \neq \emptyset \quad \text{and} \quad \widehat{\Gamma}(\emptyset) = 0,$$

which is already normalized, with

$$\sum_{S' \subseteq S} \widehat{\Gamma}(S') \le 1 \quad \text{for } S \subseteq \mathcal{I}$$

holding with equality when $S = \mathcal{I}$.

Another normalization is to consider the maximum possible value, taking into account that there is a limit, possibly due to a budget. For example, consider the knapsack problem from Section 2.2. Once $x_1 = 1$, all other variables must be zero, whereas $x_1 = 0$ allows all other variables to be 1. We can thus partition

(2.15) as:

$$\sum_{S' \subseteq S} \Gamma(S') \leq \begin{cases} \frac{1}{K} & \text{if } 1 \in S \\ 2^{n-1} - 1 & \text{otherwise.} \end{cases}$$

(Similarly for (2.16).)

Using problem-dependent bounds to normalize could raise new issues, especially as a problem is modified. Hence, we point this out only as an extreme, and note that *the normalization chosen can affect the ranking.* For this reason, we maintain a problem-independent normalization for the remainder of this study.

In deriving an appropriate divisor for the combined global persistence factor,

$$\Psi(S, R) = \Gamma(S) + \widetilde{\Gamma}(R) \text{ for } S \cap R = \emptyset,$$

we can use the bound information we have already found. First, we recall (from Section 2.2) that the total combined global persistence factor for a solution $V \subseteq \mathcal{I}$ can be written as

$$\begin{aligned} \sum_{\substack{S \subseteq V \\ R \subseteq \neg V}} \Psi(S, R) &= \sum_{\substack{S \subseteq V \\ R \subseteq \neg V}} \left(\Gamma(S) + \widetilde{\Gamma}(R) \right) \\ &= 2^{|\neg V|} \sum_{S \subseteq V} \Gamma(S) + 2^{|V|} \sum_{R \subseteq \neg V} \widetilde{\Gamma}(R). \end{aligned}$$

Then, using the bounds from (2.15) and (2.16), we have, for any $V \subseteq \mathcal{I}$,

$$\sum_{\substack{S \subseteq V \\ R \subseteq \neg V}} \Psi(S, R) \leq 2^{|\neg V|} \frac{\sum_{k=1}^{K} \left(2^{\left|S^k\right|} - 1 \right)}{K} + 2^{|V|} \frac{\sum_{k=1}^{K} \left(2^{\left|R^k\right|} - 1 \right)}{K}. \tag{2.19}$$

Finally, we determine an upper bound on the total Ψ for all $V \subseteq \mathcal{I}$ by maximizing the right-hand side of (2.19) with respect to V.

Theorem 2.2. *For* $A, B \geq 0$,

$$\max_{0 \leq v \leq N} \left\{ A(2^v) + B(2^{N-v}) \right\} = 2^N \max\{A, B\} + \min\{A, B\}.$$

Proof: The maximand is convex since its second derivative is

$$(\ln 2)^2 (A2^v + B2^{N-v}) \geq 0.$$

Thus, the maximum occurs at an end point:

$$\begin{aligned}
\max_{0 \leq v \leq N} \left\{ A(2^v) + B(2^{N-v}) \right\} &= \max_{v \in \{0, N\}} \left\{ A(2^v) + B(2^{N-v}) \right\} \\
&= \max\{A2^N + B, A + B2^N\} \\
&= 2^N \max\{A, B\} + \min\{A, B\}. \qquad \blacksquare
\end{aligned}$$

Applying Theorem 2.2 to (2.19), our upper bound on the total Ψ is

$$\begin{aligned}
\max_{V \subseteq \mathcal{I}} \sum_{\substack{S \subseteq V \\ R \subseteq \neg V}} \Psi(S, R) &= \max_{V \subseteq \mathcal{I}} \left[2^{|\neg V|} \sum_{S \subseteq V} \Gamma(S) + 2^{|V|} \sum_{R \subseteq \neg V} \widetilde{\Gamma}(R) \right] \\
&\leq \left(2^{|\mathcal{I}|}\right) \max \left[\frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|S^k\right|} - 1 \right),\ \frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|R^k\right|} - 1 \right) \right]
\end{aligned}$$

Moreover, letting

$$
\begin{aligned}
d = \left(2^{|\mathcal{I}|}\right) \max & \left[\frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|S^k\right|} - 1\right), \ \frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|R^k\right|} - 1\right) \right] \\
& + \min \left[\frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|S^k\right|} - 1\right), \ \frac{1}{K} \sum_{k=1}^{K} \left(2^{\left|R^k\right|} - 1\right) \right],
\end{aligned}
$$

we define the normalized combined global persistence factor,

$$
\underline{\Psi}(S, R) = \frac{\Gamma(S) + \widetilde{\Gamma}(R)}{d} \quad \text{for} \quad S \cap R = \emptyset. \tag{2.20}
$$

In summary, our exploration of normalization concludes that using problem-independent bounds for scaling offers simplicity and consistency, but tighter bounds are possible for specific problems.

2.4 A Note on the Domain of the Maximum-Persistence Model

In the treasure-room knapsack problem, maximizing the total global persistence subject to the original knapsack constraint gives a solution with a value very near the optimal of the original and well within the range of generated solutions. In general, though, there is no guarantee that there exists an x^* with

$$
x^* \in \underset{x}{\operatorname{argmax}} \sum_{s \,:\, x_s = 1} \gamma(s) \ : \ ax \leq b, x \in \{0,1\}^n
$$

and $cx^* \geq z^K$. We illustrate this by modifying our previous knapsack example slightly. Setting the weights equal to the values, so that

$$
a = c = [\ 44 \ \ 49 \ \ 39 \ \ 43 \ \ 11 \ \ 10 \ \ 10 \ \ 41 \ \ 5 \ \ 25\ ],
$$

the method of successive exclusions generates the results shown in Table 2.8.[1] In this case, we maximize global persistence by setting $x = [\,0\;0\;0\;0\;1\;1\;1\;1\;1\;1\,]$, which gives a total persistence value of $^{57}/_{19}$, but whose value in the original problem is only 102, well outside the range of the generated solutions.

Table 2.8: Modified treasure-room knapsack solutions ($a = c$).

k	Objective Value	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\sum_s \gamma(s)$
1	107	0	1	0	1	0	1	0	0	1	0	$^{34}/_{19}$
2	107	1	0	0	1	0	1	1	0	0	0	$^{34}/_{19}$
3	107	0	0	1	1	0	0	0	0	0	1	$^{17}/_{19}$
4	107	0	1	0	1	0	0	1	0	1	0	$^{34}/_{19}$
5	107	0	0	1	1	0	1	1	0	1	0	$^{45}/_{19}$
6	106	0	0	1	0	1	0	1	1	1	0	$^{48}/_{19}$
7	106	0	1	0	0	1	0	0	1	1	0	$^{37}/_{19}$
8	106	0	0	1	0	1	1	0	1	1	0	$^{48}/_{19}$
9	106	1	0	0	0	1	0	1	1	0	0	$^{37}/_{19}$
10	106	1	0	0	0	1	1	0	1	0	0	$^{37}/_{19}$
11	105	0	1	0	0	0	1	0	1	1	0	$^{39}/_{19}$
12	105	0	1	0	0	1	1	1	0	0	1	$^{41}/_{19}$
13	105	1	0	0	0	1	1	1	0	1	1	$^{50}/_{19}$
14	105	0	1	0	0	0	0	1	1	1	0	$^{39}/_{19}$
15	105	0	0	0	1	1	0	1	1	0	0	$^{39}/_{19}$
16	105	0	0	0	1	1	1	0	1	0	0	$^{39}/_{19}$
17	105	0	0	1	0	0	1	1	1	1	0	$^{50}/_{19}$
18	105	0	0	1	0	0	0	0	1	0	1	$^{22}/_{19}$
19	105	1	0	0	0	0	1	1	1	0	0	$^{39}/_{19}$
	$\gamma(v)$	$\frac{5}{19}$	$\frac{6}{19}$	$\frac{6}{19}$	$\frac{7}{19}$	$\frac{9}{19}$	$\frac{11}{19}$	$\frac{11}{19}$	$\frac{12}{19}$	$\frac{10}{19}$	$\frac{4}{19}$	

This same result extends to other methods of generating the collection of acceptable solutions. In general, we impose no restriction on the feasible domain

[1] For this example we set $K = 19$ in order to capture a complete breakpoint.

of the maximum-persistence problem that would restrict the optimal solution to be in the acceptable solution list beyond the constraints of the underlying model. Mathematically, this means that the collection of acceptable solutions is a subset of the considered decision space, that is $X^K \subseteq X$.

2.5 Summary

By considering the nature and extent of agreement among a collection of acceptable solutions, using the persistence of decisions, for both selections and rejections, to make these decisions suggests a greater confidence in the resulting solution. We develop an approach to uncertainty in combinatorial optimization that defines robustness in terms of maximizing persistence. First, however, we must determine an appropriate method by which a suitable collection of solutions will be generated, which will vary by application. Then, we must clarify what we mean by persistence, which can simply consider individual decisions independently, or can be extended to consider interactions among subsets of the decision space.

One approach to generate the required collection of solutions, though not the only approach, is by the method of successive exclusions, which produces a rank-ordered list of solutions by starting with an initial optimal solution and generates alternative solutions by excluding those solutions already found and re-solving our model. A consequence of using this approach is that the solutions are generated in order of decreasing optimality (with respect to the base model). Hence, by considering the persistence of a decision between successive breakpoints, where the optimal value changes, we can measure the local persistence of a decision across only those solutions that have the same objective value. Alternately, when the objective value changes very slowly between breakpoints, or when our collection of acceptable solutions is not rank-ordered, we can consider the global persistence of decisions.

If we wish to extend our notion of persistence to include not only individual decisions but also the interactions among sets of decision, we can measure persistence by the global selection persistence factor. Additionally, we can either consider selections or rejections alone or we can consider both selections and rejections in a combined measure. Then, because, in general, there is no reason for all the solutions in our list to select the same number of elements, we also investigate adjusting our measures of persistence to account for the size of the selection sets in which a given decision occurs. The result of this investigation is a normalized measure of persistence from which follows a deeper *and* more general discussion of comparability and normalization across all our various persistence measures.

3. Dempster-Shafer Theory of Evidence

Dempster-Shafer Theory (DST) is a mathematical theory of evidence that allows one to quantify the degree to which some source of evidence supports a particular proposition. It is an alternative to traditional probability theory in that this support can be allocated to sets or intervals without requiring assumptions about the degree of support given to their constituent elements. In the case where the evidence is sufficient to completely assign support to singleton sets, the model reduces to the traditional probability model. DST was initially developed by Shafer in his 1976 book, *A Mathematical Theory of Evidence* [27], as an expansion to Dempster's earlier work on the mathematical structure for upper and lower probabilities [10]. In that discussion, Shafer proposed calling set functions with the structure of Dempster's lower probabilities *belief functions*, and developed the implications of Dempster's rule for combining belief functions based on different bodies of evidence.

The theory is at once a theory of evidence, using belief functions to indicate the degree of support provided by a body of evidence, and a theory of probable reasoning, focusing on the combination of degrees of support stemming from distinct bodies of evidence. Our discussion begins with an examination of belief functions, their relatives, *plausibility functions*, and the mathematical representation of evidence. Following that we examine the combination of belief functions, beginning with a discussion of Dempster's rule of combination and its implications, then with a brief review of some alternative rules of combination including an investigation of a novel modification of Shafer's *discount & combine* method. Finally, we investigate some criticisms and alternate interpretations of DST that have arisen as the theory has evolved.

3.1 Belief Functions

A belief function assigns a value, representing the weight of evidence (or degree of support), to each of the subsets of a universe of possible events. Unlike traditional probability theory, DST relaxes the additivity axiom, so that beliefs can be formed from evidence that does not add up in some simple way. Formally, a belief function is defined as follows:

Definition 3.1. Given a nonempty universal set (or *frame of discernment*), Ω, let $\mathcal{P}(\Omega)$ denote its power set. A *belief function* is a function, $\mathrm{Bel} : \mathcal{P}(\Omega) \to [0, 1]$, such that

1. $\mathrm{Bel}(\emptyset) = 0$,
2. $\mathrm{Bel}(\Omega) = 1$, and
3. $\mathrm{Bel}(A \cup B) \geq \mathrm{Bel}(A) + \mathrm{Bel}(B) - \mathrm{Bel}(A \cap B) \quad \text{for } A, B \in \mathcal{P}(\Omega)$.

The first two conditions are the same axioms as in probability theory. The third condition is the property of *superadditivity*, which relaxes the additivity axiom of probability by allowing the belief allocated to a set to be greater than the sum of the beliefs allocated to its constituents.

Related to belief functions is another type of function for representing the plausibility of evidence in DST. Similar to belief functions, plausibility functions relax the additivity axiom of probability, but instead have the property of *subadditivity*, which means that the plausibility allocated to a set can be less than the sum of the plausibilities of its constituents. We formalize this in the following definition:

Definition 3.2. A *plausibility function* is a function $\mathrm{Plaus} : \mathcal{P}(\Omega) \to [0,1]$ such that

1. $\mathrm{Plaus}(\emptyset) = 0$,
2. $\mathrm{Plaus}(\Omega) = 1$, and
3. $\mathrm{Plaus}(A \cap B) \leq \mathrm{Plaus}(A) + \mathrm{Plaus}(B) - \mathrm{Plaus}(A \cup B) \quad \text{for } A, B \in \mathcal{P}(\Omega)$.

Given a belief function, Bel_m, its companion plausibility function is defined by the formula

$$\mathrm{Plaus}_m(A) = 1 - \mathrm{Bel}_m(\neg A). \tag{3.1}$$

We verify that this function is a plausibility function as defined by Definition 3.2, by confirming that

$$\begin{aligned}
\mathrm{Plaus}_m(\emptyset) &= 1 - \mathrm{Bel}_m(\Omega) = 0, \\
\mathrm{Plaus}_m(\Omega) &= 1 - \mathrm{Bel}_m(\emptyset) = 1, \text{ and} \\
\mathrm{Plaus}_m(A \cap B) &= 1 - \mathrm{Bel}_m(\neg A \cup \neg B) \\
&\leq 1 - [\mathrm{Bel}_m(\neg A) + \mathrm{Bel}_m(\neg B) - \mathrm{Bel}_m(\neg A \cap \neg B)] \\
&= 1 - \big[\big(1 - \mathrm{Plaus}_m(A)\big) + \big(1 - \mathrm{Plaus}_m(B)\big) - \big(1 - \mathrm{Plaus}_m(A \cup B)\big)\big] \\
&= \mathrm{Plaus}_m(A) + \mathrm{Plaus}_m(B) - \mathrm{Plaus}_m(A \cup B).
\end{aligned}$$

Like probability functions, belief functions encode the sum of all support for any subset of Ω. Unlike probability functions, however, the weakening of the additivity axiom allows some degree of support to be allocated to the whole of compound set without making any additional claims about the support given to any of its subsets. This behavior becomes explicit when we look at another set function, related to the belief function, called the *basic probability assignment* (bpa).

Definition 3.3. A bpa is a set-function $\mathrm{m}(x) : \mathcal{P}(\Omega) \rightarrow [\,0\,,1\,]$ with the following properties:

1. $\mathrm{m}(\emptyset) = 0,$
2. $\sum_{A\in\mathcal{P}(\Omega)} \mathrm{m}(A) = 1.$

The bpa, $\mathrm{m}(A)$, expresses the degree of support given to the set A but to no particular subset of A. The belief in A is then the sum of all support given to subsets of A. Thus, given a bpa, m, we can write the belief function based on m as

$$\mathrm{Bel}_{\mathrm{m}}(A) = \sum_{B\subseteq A} \mathrm{m}(B), \tag{3.2}$$

and we can write the associated plausibility function as

$$\mathrm{Plaus}_{\mathrm{m}}(A) = \sum_{B\cap A\neq\emptyset} \mathrm{m}(B). \tag{3.3}$$

Moreover, we can obtain the bpa from a given belief function via the formula:

$$\mathrm{m}(A) = \sum_{B\subseteq A} (-1)^{|A|-|B|}\,\mathrm{Bel}(B), \tag{3.4}$$

where $|A| - |B|$ gives the difference in cardinality of the two sets, and from a given plausibility function via the formula:

$$\mathrm{m}(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|}[1 - \mathrm{Plaus}(\neg B)]. \tag{3.5}$$

Using (3.2) and (3.3) and noting that $B \subseteq A \implies B \cap A \neq \emptyset$ for $A, B \neq \emptyset$, we can readily show another relation between belief functions and plausibility functions. We have, given a bpa, m,

$$\mathrm{Bel}_{\mathrm{m}}(A) = \sum_{B \subseteq A} \mathrm{m}(B) \leq \sum_{B \cap A \neq \emptyset} \mathrm{m}(B) = \mathrm{Plaus}_{\mathrm{m}}(A) \tag{3.6}$$

for all $A \subseteq \Omega$. In words, belief functions are bound above by their companion plausibility functions.

Sets to which a given bpa assigns support, those sets A for which $\mathrm{m}(A) > 0$, are called a *focal elements* [19]. We can express *total ignorance* in terms of the bpa by setting $\mathrm{m}(\Omega) = 1$ and $\mathrm{m}(A) = 0$ for all $A \neq \Omega$. This reflects that there is no evidence to support any subset of the universal set and the only focal element is the universal set, Ω. Using (3.2), we see that the corresponding belief function is $\mathrm{Bel}(\Omega) = 1$ and $\mathrm{Bel}(A) = 0$ for all $A \neq \Omega$, exactly the same as the bpa. Total ignorance is an extreme example of a class of belief functions, called *non-dogmatic* belief functions, whose corresponding bpa is such that $\mathrm{m}(\Omega) > 0$.

Some other special classes of belief functions are:

- *Dogmatic belief functions* that assign no support to the universal set (that is, $\mathrm{m}(\Omega) = 0$); and

- *Additive belief functions* that assign support *only* to singletons, so that

$$\mathrm{m}(A) = 0 \quad \forall A \subseteq \Omega \, : \, |A| > 1.$$

Additive belief functions are also referred to as *Bayesian* belief functions because of their relation to traditional probability functions, which are given in the following, known propositions [19].

Theorem 3.4. *A belief function* Bel *on a finite universe* Ω *is additive if, and only if, its corresponding bpa* m *is such that* $\mathrm{m}(\{x\}) = \mathrm{Bel}(\{x\})$ *and* $\mathrm{m}(A) = 0$ *for all* $A \subseteq \Omega$ *that are not singletons.*

Corollary 3.5. *For additive belief functions,*

$$\mathrm{Bel}(A \cup B) = \mathrm{Bel}(A) + \mathrm{Bel}(B) - \mathrm{Bel}(A \cap B) \quad \textit{for } A, B \subseteq \Omega.$$

Corollary 3.6. *For additive belief functions, we have* $\mathrm{Bel}(A) = \mathrm{Plaus}(A)$ *for all* $A \subseteq \Omega$.

3.1.1 Other Properties of Belief Functions

Let $\mathrm{m}_1, \ldots, \mathrm{m}_n$ be bpas with their associated belief functions, $\mathrm{Bel}_1, \ldots, \mathrm{Bel}_n$, and let

$$\sum_{i=1}^{n} \alpha_i = 1, \ \alpha_i \geq 0.$$

Then, the weighted sum of belief functions is a belief function:

$$\sum_{i=1}^{n} \alpha_i \operatorname{Bel}_i(\emptyset) = \sum_{i=1}^{n} \alpha_i \times 0 = 0 \tag{3.7a}$$

$$\sum_{i=1}^{n} \alpha_i \operatorname{Bel}_i(\Omega) = \sum_{i=1}^{n} \alpha_i \times 1 = 1 \tag{3.7b}$$

$$\begin{aligned} \sum_{i=1}^{n} \alpha_i \operatorname{Bel}_i(A \cup B) &\geq \sum_{i=1}^{n} \left[\alpha_i \operatorname{Bel}_i(A) + \alpha_i \operatorname{Bel}_i(B) - \alpha_i \operatorname{Bel}_i(A \cap B)\right] \qquad (3.7c) \\ &= \sum_{i=1}^{n} \alpha_i \operatorname{Bel}_i(A) + \sum_{i=1}^{n} \alpha_i \operatorname{Bel}_i(B) - \sum_{i=1}^{n} \alpha_i \operatorname{Bel}_i(A \cap B). \end{aligned}$$

3.2 Dempster's Rule of Combination

In general, distinct bodies of evidence over a common frame of discernment can be combined in a number of ways. Dempster's rule, originally presented by Dempster [10] and, subsequently, further developed by Shafer [27], was the first of these techniques specifically applied to combining belief functions. Given two independent bodies of evidence expressed by bpas, m_1 and m_2, *Dempster's rule of combination* is given by,

$$\mathrm{m}_{1,2}(A) = \frac{\sum_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}{1 - \underbrace{\sum_{X \cap Y = \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}_{= c_{1,2}}},$$

for $c_{1,2} \neq 1$. (Note, summations over X, Y are abbreviated to just the condition for notational convenience, for instance we write $X \cap Y = A$ where we want to sum over all X and Y such that $X \cap Y = A$.) Alternatively, because

$$\sum_{\substack{X \in \mathcal{P}(\Omega) \\ Y \in \mathcal{P}(\Omega)}} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) = \left(\sum_{X \in \mathcal{P}(\Omega)} \mathrm{m}_1(X) \right) \left(\sum_{Y \in \mathcal{P}(\Omega)} \mathrm{m}_2(Y) \right) = 1$$

implies

$$1 - c_{1,2} = 1 - \sum_{X \cap Y = \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) = \sum_{X \cap Y \neq \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y),$$

we can write Dempster's rule as

$$\mathrm{m}_{1,2}(A) = \frac{\sum\limits_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}{1 - \sum\limits_{X \cap Y = \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)} = \frac{\sum\limits_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}{\sum\limits_{X \cap Y \neq \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)},$$

with the requirement that $\sum\limits_{X \cap Y \neq \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) \neq 0$.

Using this expression of Dempster's rule we have,

$$\mathrm{m}_{(1,2),3}(A) = \frac{\sum\limits_{X \cap Y = A} \frac{\sum_{U \cap V = X} \mathrm{m}_1(U)\,\mathrm{m}_2(V)}{1 - c_{1,2}}\,\mathrm{m}_3(Y)}{\sum\limits_{X \cap Y \neq \emptyset} \frac{\sum_{U \cap V = X} \mathrm{m}_1(U)\,\mathrm{m}_2(V)}{1 - c_{1,2}}\,\mathrm{m}_3(Y)}$$

$$= \frac{\sum\limits_{U \cap V \cap Y = A} \mathrm{m}_1(U)\,\mathrm{m}_2(V)\,\mathrm{m}_3(Y)}{\sum\limits_{U \cap V \cap Y \neq \emptyset} \mathrm{m}_1(U)\,\mathrm{m}_2(V)\,\mathrm{m}_3(Y)}$$

$$= \frac{\sum\limits_{X \cap Y = A} \frac{\sum_{U \cap V = X} \mathrm{m}_1(U)\,\mathrm{m}_2(V)}{1 - c_{2,3}}\,\mathrm{m}_3(Y)}{\sum\limits_{X \cap Y \neq \emptyset} \frac{\sum_{U \cap V = X} \mathrm{m}_1(U)\,\mathrm{m}_2(V)}{1 - c_{2,3}}\,\mathrm{m}_3(Y)} = \mathrm{m}_{1,(2,3)}(A),$$

which shows that Dempster's rule is associative. Moreover, since

$$\mathrm{m}_{1,2}(A) = \frac{\sum\limits_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}{1 - \sum\limits_{X \cap Y = \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)} = \frac{\sum\limits_{X \cap Y = A} \mathrm{m}_2(X)\,\mathrm{m}_1(Y)}{1 - \sum\limits_{X \cap Y = \emptyset} \mathrm{m}_2(X)\,\mathrm{m}_1(Y)} = \mathrm{m}_{2,1}(A),$$

Dempster's rule of combination is also commutative. Thus, Dempster's rule is independent of the order in which evidence is combined.

In fact, given J distinct bpas, $\mathrm{m}_i : \mathcal{P}(\Omega) \to [\,0\,,1\,]$ for $i \in \{1, \ldots, J\}$, defined on the same frame of discernment, Ω, we have

$$\mathrm{m}_{1,\ldots,J}(A) = \frac{\sum\limits_{\bigcap_{i=1}^{J} X_i = A} \left(\prod\limits_{j=1}^{J} \mathrm{m}_j(X_i) \right)}{\sum\limits_{\bigcap_{i=1}^{J} X_i \neq \emptyset} \left(\prod\limits_{j=1}^{J} \mathrm{m}_j(X_i) \right)} = \frac{\sum\limits_{\bigcap_{i=1}^{J} X_i = A} \left(\prod\limits_{j=1}^{J} \mathrm{m}_j(X_i) \right)}{1 - \sum\limits_{\bigcap_{i=1}^{J} X_i = \emptyset} \left(\prod\limits_{j=1}^{J} \mathrm{m}_j(X_i) \right)}.$$

3.2.1 Alternate Rules of Combination

Dempster's rule can lead to counterintuitive results with highly conflicting evidence. The value of the term c in the denominator can be seen as a measure of the conflict. When $c = 0$, Dempster's rule is well justified, but for $c \neq 0$ the normalization factor, $1 - c$, has the effect of attributing any probability mass associated with conflict to the null set, completely ignoring any conflict.

As an example, consider a situation where two technicians are asked to diagnose a malfunctioning computer. The first believes that the problem is likely due to bad RAM though he considers the possibility of a failing CPU and gives the following values: $\mathrm{Bel}_1(\text{bad RAM}) = \mathrm{m}_1(\text{bad RAM}) = 0.98$, and $\mathrm{Bel}_1(\text{CPU failure}) = \mathrm{m}_1(\text{CPU failure}) = 0.02$. The second suspects a failing hard drive, but also recognizes a small possibility of a failing CPU, assigning the values: $\mathrm{Bel}_2(\text{Drive failure}) = \mathrm{m}_2(\text{Drive failure}) = 0.97$, and $\mathrm{Bel}_2(\text{CPU failure}) = \mathrm{m}_2(\text{CPU failure}) = 0.03$. Using these values we see that,

$$c = (0.98)(0.97) + (0.98)(0.03) + (0.02)(0.97) = 0.9994,$$

and thus,

$$\mathrm{m}_{1,2}(\text{CPU failure}) = \frac{(0.02)(0.03)}{1 - 0.9994} = 1,$$

which implies that $\mathrm{Bel}_{1,2}(\text{CPU failure}) = 1$. Dempster's rule of combination thus assigns complete support to the belief in a failing CPU, a diagnoses neither technician believes is very likely. This happens because each of the technicians rules out the preferred alternative of the other.

In order to address this anomaly, a number of alternative rules of combination have been developed (Sentz & Ferson [26] provides a good survey of these). We describe two here: Yager's rule [34] and Dubois and Prade's rule [11]. In the next section, we explore and develop a specialized extension of Shafer's discount & combine method [27].

Yager's Rule:

$$\mathrm{m}_{1,2}(A) = \begin{cases} \sum\limits_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) & \text{for } A \neq \emptyset \text{ and } A \neq \Omega \\ \mathrm{m}_1(\Omega)\,\mathrm{m}_2(\Omega) + \sum\limits_{X \cap Y = \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) & \text{for } A = \Omega \\ 0 & \text{for } A = \emptyset \end{cases}$$

Applying Yager's Rule to the technician example, we get the combined bpa:

$$\mathrm{m}_{1,2}(\text{bad RAM}) = 0, \qquad \mathrm{m}_{1,2}(\text{Drive failure}) = 0,$$
$$\mathrm{m}_{1,2}(\text{CPU failure}) = 0.0006, \qquad \mathrm{m}_{1,2}(\Omega) = 0.9994.$$

This gives a belief function that represents the technicians' disagreement with nearly total ignorance (see Table 3.1).

Dubois and Prade's Rule:

$$\mathrm{m}_{1,2}(A) = \begin{cases} \sum\limits_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) + \sum\limits_{\substack{X \cup Y = A \\ X \cap Y = \emptyset}} \mathrm{m}_1(X)\,\mathrm{m}_2(Y) & \text{for } A \neq \emptyset \\ 0 & \text{for } A = \emptyset \end{cases}$$

Applying Dubois and Prade's Rule to the technician example, we get the combined bpa:

$$\mathrm{m}_{1,2}(\text{bad RAM}) = 0, \qquad \mathrm{m}_{1,2}(\text{bad RAM or CPU failure}) = 0.0294,$$
$$\mathrm{m}_{1,2}(\text{CPU failure}) = 0.0006, \qquad \mathrm{m}_{1,2}(\text{bad RAM or Drive failure}) = 0.9506,$$
$$\mathrm{m}_{1,2}(\text{Drive failure}) = 0, \qquad \mathrm{m}_{1,2}(\text{CPU failure or Drive failure}) = 0.0194.$$

Thus, instead of putting the most support into Ω (as Yager's Rule does), this disjuncts the technicians' top choices.

We compare the resulting belief and plausibility values in Table 3.1. Note that Yager's rule results in nearly total ignorance, assigning belief values near zero while assigning very high plausibility values to everything. By contrast, though Dubois and Prade's rule also assigns very little belief to singletons, it does assign substantial belief to the set "bad RAM or Drive failure" and very low plausibility to "CPU failure."

Table 3.1: Combined belief and plausibility values for the technician example.

	Yager		Dubois & Prade	
	Bel	Plaus	Bel	Plaus
bad RAM	0.0000	0.9994	0.0000	0.9800
CPU failure	0.0006	1.0000	0.0006	0.0494
Drive failure	0.0000	0.9994	0.0000	0.9700
bad RAM or CPU failure	0.0006	1.0000	0.0300	1.0000
bad RAM or Drive failure	0.0000	0.9994	0.9506	0.9994
CPU or Drive failure	0.0006	1.0000	0.0200	1.0000
bad RAM or CPU or Drive failure	1.0000	1.0000	1.0000	1.0000

3.2.2 Discount & Combine – Combining Highly Conflicting Beliefs

As discussed in Section 3.2.1, Dempster's rule can be problematic when combining highly conflicting evidence. Mathematically, the denominator becomes zero, but more importantly, the conflict means we cannot believe the combination. This problem is particularly evident when combining highly conflicting, *dogmatic* belief functions. For such situations, Shafer proposed a method in chapter 11 of his book [27, p. 251–255] that discounts the belief functions by first applying a discount factor $0 < \alpha_i < 1$ to each belief function to reflect our confidence in that source of evidence, $\mathrm{Bel}_i(A)$, for all proper subsets A of Ω, so that

$$\mathrm{Bel}_i^{\alpha}(A) = \alpha_i \, \mathrm{Bel}_i(A) \text{ for } A \neq \Omega \quad \text{and} \quad \mathrm{Bel}_i^{\alpha}(\Omega) = 1,$$

and then combines them using Dempster's Rule. This discount & combine method is more appropriate for combining highly conflicting belief functions because it avoids the problem of discarding conflicting information seen in other rules of combination. In fact, for belief functions that are both highly conflicting *and* dogmatic, discounting the belief functions before combining them is essential in order to avoid discarding evidence, because the other methods assign zero belief to anything not supported by both evidence pools.

For combining highly-conflicting, dogmatic belief functions we propose a new method based on Shafer's discount & combine method, where we introduce an additional "re-normalizing" step to convert the resulting combined belief function back into a dogmatic belief function. We further show that for totally conflicting bpas (where $c = 1$) the combined belief function is equivalent to a weighted sum of the original belief functions.

Given totally conflicting, dogmatic bpas m_1 and m_2, define the non-dogmatic bpas:

$$\begin{aligned} \hat{\mathrm{m}}_1(A) &= \alpha\, \mathrm{m}_1(A) \text{ for } A \neq \Omega, \\ \hat{\mathrm{m}}_1(\Omega) &= 1 - \alpha, \end{aligned} \qquad \text{and} \qquad \begin{aligned} \hat{\mathrm{m}}_2(A) &= \beta\, \mathrm{m}_1(A) \text{ for } A \neq \Omega, \\ \hat{\mathrm{m}}_2(\Omega) &= 1 - \beta, \end{aligned}$$

where $0 < \alpha < 1$ and $0 < \beta < 1$. We can think of α and β as weights of trust on the evidence for m_1 and m_2, respectively; setting $\alpha = 0$ implies that we have no trust in m_1 (that is, $\hat{\mathrm{m}}_1$ reflects total ignorance), while setting $\alpha = 1$ implies total faith in m_1 with $\hat{\mathrm{m}}_1 = \mathrm{m}_1$.

We can then combine $\hat{\mathrm{m}}_1$ and $\hat{\mathrm{m}}_2$ using Dempster's rule:

$$\hat{\mathrm{m}}_{1,2}(A) = \frac{\sum\limits_{X \cap Y = A} \hat{\mathrm{m}}_1(X)\hat{\mathrm{m}}_2(Y)}{1 - \sum\limits_{X \cap Y = \emptyset} \hat{\mathrm{m}}_1(X)\hat{\mathrm{m}}_2(Y)} = \frac{\sum\limits_{X \cap Y = A} \hat{\mathrm{m}}_1(X)\hat{\mathrm{m}}_2(Y)}{\sum\limits_{X \cap Y \neq \emptyset} \hat{\mathrm{m}}_1(X)\hat{\mathrm{m}}_2(Y)},$$

for $A \neq \emptyset$ and $\hat{\mathrm{m}}_{1,2}(\emptyset) = 0$.

Since m_1 and m_2 are totally conflicting, we have, for $A \neq \Omega$,

$$\begin{aligned}
\hat{\mathrm{m}}_{1,2}(A) &= \frac{\alpha\,\mathrm{m}_1(A)(1-\beta) + (1-\alpha)\beta\,\mathrm{m}_2(A)}{\sum\limits_{A \neq \Omega} \alpha\,\mathrm{m}_1(A)(1-\beta) + \sum\limits_{A \neq \Omega} (1-\alpha)\beta\,\mathrm{m}_2(A) + (1-\alpha)(1-\beta)} \\
&= \frac{\alpha(1-\beta)\,\mathrm{m}_1(A) + (1-\alpha)\beta\,\mathrm{m}_2(A)}{\alpha(1-\beta) + (1-\alpha)\beta + (1-\alpha)(1-\beta)} \\
&= \frac{\alpha(1-\beta)\,\mathrm{m}_1(A) + (1-\alpha)\beta\,\mathrm{m}_2(A)}{1-\alpha\beta} \qquad (3.8)
\end{aligned}$$

and, for $A = \Omega$,

$$\hat{\mathrm{m}}_{1,2}(\Omega) = \frac{(1-\alpha)(1-\beta)}{1-\alpha\beta}. \qquad (3.9)$$

The reasons for restricting α and β to be strictly greater than zero and less than one become apparent when we examine the behavior of $\hat{\mathrm{m}}_{1,2}$ at these extremes. If, for example, we let $\alpha = 0$, we have $\hat{\mathrm{m}}_{1,2}(\Omega) = 1 - \beta$ and $\hat{\mathrm{m}}_{1,2}(A) = \beta\,\mathrm{m}_2(A)$, which is simply the non-dogmatic bpa $\hat{\mathrm{m}}_2$. At the other extreme, where $\alpha = 1$, we have $\hat{\mathrm{m}}_{1,2}(\Omega) = 0$ and $\hat{\mathrm{m}}_{1,2}(A) = \frac{(1-\beta)\,\mathrm{m}_1(A)}{1-\beta}$, which is undefined when $\beta = 1$ and equal to m_1 otherwise.

Finally, we produce the dogmatic combination, $\mathrm{m}_{1,2}$, by setting $\mathrm{m}_{1,2}(\Omega) = 0$

and normalizing:

$$\mathrm{m}_{1,2}(A) = \frac{\hat{\mathrm{m}}_{1,2}(A)}{1 - \hat{\mathrm{m}}_{1,2}(\Omega)} = \frac{\dfrac{\alpha(1-\beta)\,\mathrm{m}_1(A) + (1-\alpha)\beta\,\mathrm{m}_2(A)}{1-\alpha\beta}}{1 - \dfrac{(1-\alpha)(1-\beta)}{1-\alpha\beta}}$$

$$= \frac{\alpha(1-\beta)\,\mathrm{m}_1(A) + (1-\alpha)\beta\,\mathrm{m}_2(A)}{1-\alpha\beta-(1-\alpha)(1-\beta)}$$

$$= \frac{\alpha(1-\beta)\,\mathrm{m}_1(A) + (1-\alpha)\beta\,\mathrm{m}_2(A)}{\alpha - 2\alpha\beta + \beta}$$

$$= \frac{\alpha(1-\beta)}{\alpha - 2\alpha\beta + \beta}\,\mathrm{m}_1(A) + \frac{(1-\alpha)\beta}{\alpha - 2\alpha\beta + \beta}\,\mathrm{m}_2(A). \quad (3.10)$$

Note that, since

$$\frac{\alpha(1-\beta)}{\alpha - 2\alpha\beta + \beta} + \frac{(1-\alpha)\beta}{\alpha - 2\alpha\beta + \beta} = 1,$$

$\mathrm{m}_{1,2}$ is a weighted sum of m_1 and m_2. Moreover, in the case where $\alpha = \beta$ we have

$$\frac{\alpha(1-\beta)}{\alpha - 2\alpha\beta + \beta} = \frac{(1-\alpha)\beta}{\alpha - 2\alpha\beta + \beta} = \frac{1}{2}.$$

For the case of combining n "totally conflicting" belief functions, whose focal elements are completely disjoint, we have

$$\hat{\mathrm{m}}_{1\ldots n}(A) = \frac{\sum_{i=1}^{n}\left(\alpha_i \prod_{j\neq i}(1-\alpha_i)\right)\mathrm{m}_i(A)}{\sum_{i=1}^{n}\left(\alpha_i \prod_{j\neq i}(1-\alpha_i)\right) + \prod_{i=1}^{n}(1-\alpha_i)}$$

and, for $A = \Omega$,

$$\hat{\mathrm{m}}_{1\ldots n}(\Omega) = \frac{\prod_{i=1}^{n}(1-\alpha_i)}{\sum_{i=1}^{n}\left(\alpha_i \prod_{j\neq i}(1-\alpha_i)\right) + \prod_{i=1}^{n}(1-\alpha_i)}$$

where $0 \leq \alpha_i \leq 1$ is the discounting weight for m_i. Setting $\mathrm{m}_{1\ldots n}(\Omega) = 0$ and normalizing:

$$\mathrm{m}_{1\ldots n}(A) = \frac{\hat{\mathrm{m}}_{1\ldots n}(A)}{1 - \hat{\mathrm{m}}_{1\ldots n}(\Omega)} = \frac{\sum_{i=1}^{n}\left(\alpha_i \prod_{j\neq i}(1-\alpha_i)\right)\mathrm{m}_i(A)}{\sum_{i=1}^{n}\left(\alpha_i \prod_{j\neq i}(1-\alpha_i)\right)}.$$

Looking again at the case where all the discounting weights are equal, $\alpha_i = \alpha$ for all i, we have

$$\begin{aligned}\mathrm{m}_{1\ldots n}(A) = \frac{\sum_{i=1}^{n}\alpha(1-\alpha)^{n-1}\,\mathrm{m}_i(A)}{\sum_{i=1}^{n}\alpha(1-\alpha)^{n-1}} &= \frac{\sum_{i=1}^{n}\alpha(1-\alpha)^{n-1}\,\mathrm{m}_i(A)}{\sum_{i=1}^{n}\alpha(1-\alpha)^{n-1}} \\ &= \sum_{i=1}^{n}\frac{\alpha(1-\alpha)^{n-1}}{n\alpha(1-\alpha)^{n-1}}\,\mathrm{m}_i(A) = \sum_{i=1}^{n}\tfrac{1}{n}\,\mathrm{m}_i(A).\end{aligned}$$

We have thus generalized our previous observation for combining two totally conflicting belief functions with equal discounting weights.

In the case where m_1 and m_2 are not totally conflicting, we need to factor in the extent of agreement. Thus, we have, for $A \neq \Omega$,

$$\hat{\mathrm{m}}_{1,2}(A) = \frac{\alpha\,\mathrm{m}_1(A)(1-\beta) + (1-\alpha)\beta\,\mathrm{m}_2(A) + \alpha\beta\sum_{X\cap Y=A}\mathrm{m}_1(X)\,\mathrm{m}_2(Y)}{1-\alpha\beta+\alpha\beta\sum_{X\cap Y\neq\emptyset}\mathrm{m}_1(X)\,\mathrm{m}_2(Y)} \tag{3.11}$$

and, for $A = \Omega$,

$$\hat{\mathrm{m}}_{1,2}(\Omega) = \frac{(1-\alpha)(1-\beta)}{1 - \alpha\beta + \alpha\beta \sum\limits_{X \cap Y \neq \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}. \tag{3.12}$$

Setting $\mathrm{m}_{1,2}(\Omega) = 0$ and normalizing we get:

$$\mathrm{m}_{1,2}(A) = \frac{\alpha(1-\beta)\,\mathrm{m}_1(A) + (1-\alpha)\beta\,\mathrm{m}_2(A) + \alpha\beta \sum\limits_{X \cap Y = A} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}{\alpha - 2\alpha\beta + \beta + \alpha\beta \sum\limits_{X \cap Y \neq \emptyset} \mathrm{m}_1(X)\,\mathrm{m}_2(Y)}. \tag{3.13}$$

Applying our modified discount & combine method to the technician example from Section 3.2.1 with $\alpha = \beta = 0.5$, we get the combined bpa:

$\mathrm{m}_{1,2}(\text{bad RAM}) = 0.48985$, $\mathrm{m}_{1,2}(\text{Drive failure}) = 0.48485$.

$\mathrm{m}_{1,2}(\text{CPU failure}) = 0.02529$,

which places support nearly evenly split between bad RAM and drive failure. Table 3.2 gives the resulting belief function and plausibility function values for both the unmodified and our modified discount & combine method. We note that in this case, the bpa produced by our method gives nonzero values to only singletons (unlike the unmodified method, which leaves some weight assigned to Ω). Hence, this belief function is additive, so Corollary 3.6 applies and the belief values are equal to the plausibility values. Moreover, unlike either Yager's rule or Dubois and Prade's rule, this method of combination assigns significant belief to both technicians' preferred causes for the malfunction, and, unlike the unmodified approach, it retains the dogmatic nature of the belief functions being combined.

Table 3.2: Discount & combine belief and plausibility values for the technician example.

	Unmodified		Modified	
	Bel	Plaus	Bel	Plaus
bad RAM	0.32660	0.65987	0.48985	0.48985
CPU failure	0.01686	0.35013	0.02529	0.02529
Drive failure	0.32327	0.65654	0.48459	0.48459
bad RAM or CPU failure	0.34346	0.67673	0.51514	0.51514
bad RAM or Drive failure	0.64987	0.98314	0.97470	0.97470
CPU or Drive failure	0.34013	0.67340	0.51014	0.51014
bad RAM or CPU or Drive failure	1.00000	1.00000	1.00000	1.00000

4. Using Persistence as Evidence

The intuitive notion of persistence is a compelling force for robust, decision-making. Given a list of "acceptable" solutions, an individual decision, such as selecting or rejecting a particular asset, that persists in appearing in the solution set is appealing to select in our final solution. This persistence conveys a form of robustness in that we feel more confident that this selection or rejection is a good decision even when our estimates of the returns are inexact.

In this chapter, we build a formal foundation using the Dempster-Shafer Theory of Evidence by treating persistence as *evidence*, which we in turn use to build belief functions. Our first approach builds belief functions by considering the total global persistence of solutions. Following that, we investigate an alternative approach that builds its belief functions by treating solutions as sets of decisions and considering the persistence of subsets of these sets.

4.1 Building Belief from Persistence of Solutions

In this section, we develop our first approach to using persistence as evidence of a solution's robustness. We define the frame of discernment, Ω, for our belief functions to be the set of 2^n possible solutions, expressed as vectors. This corresponds to the set of selections:

$$S(x) = \{j \,:\, x_j = 1\} \quad \text{and} \quad x_j(S) = \begin{cases} 1 & \text{if } j \in S; \\ 0 & \text{otherwise.} \end{cases}$$

Using vector notation, let $X^K = \{x^1, \dots, x^K\}$ denote the set of K x-vectors used to compute the *persistence vector*, $p = [\, \gamma(x_1) \ \cdots \ \gamma(x_n) \,]^\mathsf{T}$ (hence, $1 - p$ is a vector of the global persistence of rejection values). Then, for a given feasible solution space, X, we define a bpa based on the total combined (selections and rejections) global persistence value of a solution:

$$
\begin{aligned}
\mathrm{m}(\{x\}) &= 0 && \text{for } x \notin X && \text{(4.1a)}\\
\mathrm{m}(\{x\}) &= \frac{1}{\kappa}\big(p^\mathsf{T} x + (\vec{\mathbf{1}} - p)^\mathsf{T}(\vec{\mathbf{1}} - x)\big) && \text{for } x \in X && \text{(4.1b)}\\
\mathrm{m}(Y) &= 0 && \text{for } Y \subseteq X \,:\, |Y| > 1, && \text{(4.1c)}
\end{aligned}
$$

where $\kappa = \sum_{y \in X} \big(p^\mathsf{T} y + (\vec{\mathbf{1}} - p)^\mathsf{T}(\vec{\mathbf{1}} - y)\big)$ and $\vec{\mathbf{1}}$ is an appropriately sized vector of ones. Note that setting $\mathrm{m}(\{x\}) = 0$ for $x \notin X$ defines the set of focal elements to be exactly the set X. Moreover, since $\mathrm{m}(Y) = 0$ for $|Y| > 1$, we can simplify our notation, writing $\mathrm{m}(\{x\})$ as $\mathrm{m}(x)$.

We call (4.1) the *singleton-only* bpa, as it does not assign a 'probability mass' to subsets of X with more than one vector. We also consider replacing (4.1c) with the following:

$$\text{disjunctive bpa: } \mathrm{m}(Y) = \sum_{x \in Y} \mathrm{m}(x) \tag{4.2}$$

for $Y \subseteq X$, and re-defining κ in (4.1b) accordingly. The disjunctive bpa, interprets $\mathrm{m}(Y)$ as the 'probability' that *some* member of Y is the most robust, as defined in the following.

Definition 4.1. The *mode* of the bpa is the solution with the belief, which we define to be the *most robust*.

This interpretation carries over to the belief function that we now develop for the singleton-only bpa. The total persistence of the selections is given by $p^\mathsf{T} x = \sum_{j \in S(x)} p_j$, and the total persistence of the rejections is given by $(\vec{\mathbf{1}} - p)^\mathsf{T}(\vec{\mathbf{1}} - x) = \sum_{j \notin S(x)} (1 - p_j)$. This induces the belief function:

$$\begin{aligned} \mathrm{Bel}(x) &= \mathrm{m}(x) \quad \text{for } x \in X \\ \mathrm{Bel}(x) &= 0 \qquad\quad \text{for } x \notin X. \end{aligned} \tag{4.3}$$

Since $\mathrm{Bel}(x) = \mathrm{m}(x)$ for singletons and $\mathrm{m}(Y) = 0$ for all other subsets of Ω, this belief function is an example of the special case of additive belief functions that are equivalent to probability measures.

Depending on how we generate our initial list of "acceptable" solutions, the list may include solutions of questionable quality. For example, consider the ranked list of solutions from Table 2.1, which was produced by solving the knapsack problem using the method of successive exclusions discussed in Section 2.1 (see p. 7). For many of the generated solutions there exist higher-ranking solutions with better objective item-values and less total item-weight. We call such solutions *dominated*.

Definition 4.2. The solution vector x *dominates* y for the knapsack problem, if $cx > cy$ and $ax \leq ay$.

Counting dominated solutions may cause a distortion of the persistence of a decision, which we can address by modifying the definition of the persistence vector so that p_j denotes the *undominated* persistence of $x_j = 1$ (that is, we define X^K to include only the undominated solutions and compute the persistence values from this new set). The persistence of a rejection is $1 - p_j$.

As an example, Table 4.1a shows the belief and plausibility values resulting from using the solution set from Table 2.1 and setting $X = X^K$. Notice that each of the solutions with $cx = 108$ is dominated by the solution with $cx = 109$ because $ax = 101 \leq 103 \leq 105 \leq 107$. Similarly, both solutions with $cx = 107$ dominate all the lower-valued solutions. Removing these dominated solutions, we get Table 4.1b, which shows the adjusted belief values for the undominated solutions. Since this belief function is additive, we observe a characteristic of such functions: $\text{Bel}(x) = \text{Plaus}(x)$, for all solutions. (We include the extra column here for later comparison.)

Table 4.1: Singleton-only belief and plausibility values.

(a) with dominated solutions

cx	ax	Bel(x)	Plaus(x)
112	105	0.0750	0.0750
110	104	0.0766	0.0766
109	101	0.0782	0.0782
108	103	0.0735	0.0735
108	105	0.0671	0.0671
108	107	0.0671	0.0671
107	100	0.0750	0.0750
107	105	0.0656	0.0656
107	104	0.0687	0.0687
107	97	0.0766	0.0766
106	107	0.0671	0.0671
105	104	0.0687	0.0687
105	101	0.0703	0.0703
105	103	0.0703	0.0703

$p^d = [4\ 4\ 4\ 10\ 5\ 13\ 6\ 6\ 11\ 0]^{\mathsf{T}}/14$

(b) without dominated solutions

cx	ax	Bel(x)	Plaus(x)
112	105	0.2010	0.2010
110	104	0.1907	0.1907
109	101	0.2113	0.2113
107	100	0.2010	0.2010
107	97	0.1959	0.1959

$p^u = [1\ 1\ 1\ 5\ 1\ 5\ 3\ 2\ 5\ 0]^{\mathsf{T}}/5$

4.1.1 Making a Partial Decision — *Conditional Beliefs*

Suppose we make an a priori decision to select or reject some particular item. Define $X_j = \{x \in X : x_j = 1\}$ and $\overline{X}_j = \{x \in X : x_j = 0\}$. Let $\mathrm{Bel}(x \mid Y)$, where $Y \subseteq X$, denote Dempster's conditional belief function, which is given by Dempster's rule of conditioning:

$$\mathrm{Bel}(x \mid X_j) = \frac{\mathrm{Bel}(x \vee \neg X_j) - \mathrm{Bel}(\neg X_j)}{1 - \mathrm{Bel}(\neg X_j)} = \frac{\mathrm{Bel}(x \vee \overline{X}_j) - \mathrm{Bel}(\overline{X}_j)}{1 - \mathrm{Bel}(\overline{X}_j)}.$$

Theorem 4.3. *For the singleton-only bpa, the belief for a solution x given we know $x_j = 1$ is*

$$\mathrm{Bel}(x \mid X_j) = \begin{cases} \dfrac{\mathrm{Bel}(x)}{\mathrm{Bel}\,(X_j)} & \text{if } x \in X_j \\ 0 & \text{if } x \in \overline{X}_j. \end{cases}$$

Proof: For $x \in \overline{X}_j$, $\mathrm{Bel}(x) = 0$ and

$$\mathrm{Bel}(x \vee \overline{X}_j) = \mathrm{Bel}(\overline{X}_j) \quad \Longrightarrow \quad \mathrm{Bel}(x \mid X_j) = 0.$$

For $x \in X_j$, since the singleton-only belief is additive,

$$\mathrm{Bel}(x \vee \overline{X}_j) = \mathrm{Bel}(x) + \mathrm{Bel}(\overline{X}_j).$$

Then, from (3.1) and Corollary 3.5,

$$1 - \mathrm{Bel}(\overline{X}_j) = \mathrm{Plaus}(X_j) = \mathrm{Bel}(X_j) \quad \Longrightarrow \quad \mathrm{Bel}(x \mid X_j) = \frac{\mathrm{Bel}(x)}{\mathrm{Bel}(X_j)}. \quad \blacksquare$$

Intuition suggests that belief increases as "support" for that belief increases, meaning that conditional belief in x should increase with conditions that it satisfies; otherwise, if the evidence does not support x, the belief in x drops to zero. That is, intuition suggests that, if $x \notin X_i \cup X_j$ (so $x \notin X_i$ and $x \notin X_j$),

$$\mathrm{Bel}(x) \geq \mathrm{Bel}(x \mid X_i \cup X_j) = \mathrm{Bel}(x \mid X_i \cap X_j) = 0$$

and, if $x \notin X_i$ (or $x \notin X_j$),

$$\mathrm{Bel}(x \mid X_i) = \mathrm{Bel}(x \mid X_i \cap X_j) = 0 \quad (\text{or } \mathrm{Bel}(x \mid X_j) = \mathrm{Bel}(x \mid X_i \cap X_j) = 0);$$

otherwise, if $x \in X_i \cap X_j$,

$$\mathrm{Bel}(x) \leq \mathrm{Bel}(x \mid X_i \cup X_j) \leq \mathrm{Bel}(x \mid X_i) \leq \mathrm{Bel}(x \mid X_i \cap X_j).$$

Although this intuition is not valid for all bpas, it does hold for our singleton-only bpa. The first two cases follow directly from Theorem 4.3 and we confirm the third in the following theorem.

Theorem 4.4. *For the singleton-only bpa, $x \in X_i \cap X_j$ implies:*

$$\mathrm{Bel}(x) \leq \mathrm{Bel}(x \mid X_i \cup X_j) \leq \mathrm{Bel}(x \mid X_i) \leq \mathrm{Bel}(x \mid X_i \cap X_j).$$

Proof: Since $X_i \cap X_j \subseteq X_i \subseteq X_i \cup X_j \subseteq \Omega$,

$$\frac{\text{Bel}(x)}{\text{Bel}(X_i \cap X_j)} \geq \frac{\text{Bel}(x)}{\text{Bel}(X_i)} \geq \frac{\text{Bel}(x)}{\text{Bel}(X_i \cup X_j)} \geq \frac{\text{Bel}(x)}{\text{Bel}(\Omega)} = \text{Bel}(x),$$

and, by Theorem 4.3,

$$\text{Bel}(x) \leq \text{Bel}(x \,|\, X_i \cup X_j) \leq \text{Bel}(x \,|\, X_i) \leq \text{Bel}(x \,|\, X_i \cap X_j). \qquad \blacksquare$$

Table 4.2a shows the revised (conditional) beliefs for some selections from Table 4.1a, and Table 4.2b shows the associated beliefs for solutions from Table 4.1b. Observe that in the undominated case (Table 4.2b), since $x_1 = 1$ in only the first solution, the belief in the first solution jumps to 1, while the beliefs in all others drop to zero. Moreover, since $x_4 = x_6 = x_9 = 1$ in *all* undominated solutions, fixing these values leaves the beliefs unchanged. In these cases, where $\text{Bel}(x \,|\, X_j) = \text{Bel}(x)$, the knowledge that $x_j = 1$ does not affect our belief that x is the most robust solution.

Finally, we note that since $x_{10} = 0$ for all x, the conditional belief, $\text{Bel}(x \,|\, X_{10})$, is not defined in either the dominated or undominated case. As in Bayesian conditional probability, $\text{P}\left(A \,\middle|\, B\right) = \frac{\text{P}(A \cap B)}{\text{P}(B)}$ fails when $\text{P}(B) = 0$, Dempster's rule also fails when we attempt to define a belief conditioned on an event that does not occur.

4.2 Building Belief from Persistence of Decisions

In our second approach we treat the solutions in our solution list as sets of decisions and construct a belief function by examining the persistence of each of their subsets. Consider the collection of belief functions induced by each S^k

Table 4.2: Singleton-only conditional beliefs.

(a) with dominated solutions

Bel(x)	Bel($x \vert X_1$)	Bel($x \vert X_4$)	Bel($x \vert X_6$)	Bel($x \vert X_9$)
0.0750	0.2699	0.1033	0.0803	0.0945
0.0766	0	0.1054	0.0820	0.0965
0.0782	0	0.1076	0.0837	0.0985
0.0735	0	0.1011	0.0786	0.0925
0.0671	0.2415	0	0.0719	0.0846
0.0671	0.2415	0.0924	0.0719	0
0.0750	0	0.1033	0.0803	0.0945
0.0656	0	0.0902	0	0.0826
0.0687	0.2472	0.0946	0.0735	0
0.0766	0	0.1054	0.0820	0.0965
0.0671	0	0	0.0719	0.0846
0.0687	0	0	0.0735	0.0866
0.0703	0	0	0.0752	0.0886
0.0703	0	0.0967	0.0752	0

(b) without dominated solutions

Bel(x)	Bel($x \vert X_1$)	Bel($x \vert X_4$)	Bel($x \vert X_6$)	Bel($x \vert X_9$)
0.2010	1	0.2010	0.2010	0.2010
0.1907	0	0.1907	0.1907	0.1907
0.2113	0	0.2113	0.2113	0.2113
0.2010	0	0.2010	0.2010	0.2010
0.1959	0	0.1959	0.1959	0.1959

from defining each bpa, m_k, such that each nonempty subset of S^k is equally likely:

$$\mathrm{m}_k(S) = \begin{cases} \frac{1}{2^{|S^k|-1}} & \text{if } S \subseteq S^k, S \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

We then aggregate the evidence from the full solution list by defining a final bpa as a weighted sum of each of these K bpas so that,

$$\mathrm{m}(S) = \frac{1}{K}\sum_{k=1}^{K}\mathrm{m}_k(S) = \frac{1}{K}\sum_{k\,:\,S\subseteq S^k}\frac{1}{2^{|S^k|}-1},$$

and the belief for a given solution is

$$\mathrm{Bel}(x) = \frac{1}{K}\sum_{k=1}^{K}\sum_{S\subseteq S(x)}\mathrm{m}_k(S) = \frac{1}{K}\sum_{k=1}^{K}\mathrm{Bel}_k\big(S(x)\big). \tag{4.4}$$

We call (4.4) the *decision-sets* belief function. Note that, since this belief function considers only selections, there is nothing restricting infeasible solutions from having large belief values (in fact, for $x = \vec{\mathbf{1}}$ we have $\mathrm{Bel}(x) = 1$). Hence, we must explicitly require that the solutions under consideration first satisfy any constraints from the underlying problem. Because it does not affect the relative rankings this requirement is not onerous. Moreover, depending on the nature of the problem, we may feel that some solutions should have high belief values despite their infeasibility (for instance, we might want to represent solutions we "wish" were available — if only we had a larger knapsack).

Table 4.3 illustrates this approach using the solution set from Table 2.1. Mirroring Tables 4.1a and 4.1b, Table 4.3a gives the beliefs resulting from using the persistence values from all fourteen solutions, while Table 4.3b considers only the undominated solutions. Note that in considering subsets, the second solution (with $cx = 110$) has the highest belief when all fourteen solutions are considered — unlike in the singleton-only bpa where the third solution has the highest belief. When considering only undominated solutions both approaches assign the greatest belief to the third solution (with $cx = 109$).

Table 4.3: Decision-sets belief values, $\mathrm{Bel}(x) = \sum_{k=1}^{K} \frac{\mathrm{Bel}_k\left(S(x)\right)}{K}$.

(a) with dominated solutions

cx	ax	Bel(x)	Plaus(x)
112	105	0.4292	0.9109
110	104	0.4378	0.9152
109	101	0.4372	0.9246
108	103	0.3807	0.9009
108	105	0.2433	0.8106
108	107	0.2476	0.7916
107	100	0.3991	0.9055
107	105	0.2286	0.7820
107	104	0.2565	0.8057
107	97	0.3088	0.8760
106	107	0.3333	0.8441
105	104	0.3330	0.8628
105	101	0.2614	0.8192
105	103	0.2608	0.8143

(b) without dominated solutions

cx	ax	Bel(x)	Plaus(x)
112	105	0.5320	0.9544
110	104	0.4804	0.9415
109	101	0.5837	0.9673
107	100	0.5320	0.9544
107	97	0.3807	0.9226

This belief function is very closely related to the global selection persistence factor and adjusted global selection persistence factor from Section 2.1. Recall (p. 14) that for any selection set $S \subseteq \mathcal{I}$, the global selection persistence factor is defined as

$$\Gamma(S) = \frac{\left|\{k \,:\, S \subseteq S^k \text{ for } 1 \leq k \leq K\}\right|}{K} \text{ for } S \neq \emptyset \quad \text{with} \quad \Gamma(\emptyset) = 0$$

and (c.f. p. 17) the adjusted global selection persistence factor is defined as

$$\widehat{\Gamma}(V) = \frac{\displaystyle\sum_{\{1 \le k \le K \,:\, S \subseteq S^k\}} (2^{|S^k|} - 1)^{-1}}{K} \quad \text{for } S \neq \emptyset \quad \text{with} \quad \widehat{\Gamma}(\emptyset) = 0.$$

When all optimal selection sets have the same cardinality (that is, $|S^k| = m$ for all k in $1, \ldots, K$), maximizing total, global selection persistence factor is equivalent to maximizing total belief. This is shown with the following.

Theorem 4.5 (Restricted Equivalence)**.** *If* $|S^k| = |S^1|$ *for all* $k \in \{1, \ldots, K\}$,

$$\operatorname*{argmax}_{x \in X} \sum_{S' \subseteq S(x)} \Gamma(S') \quad = \quad \operatorname*{argmax}_{x \in X} \sum_{k} \mathrm{Bel}_k \left(S(x) \right).$$

Proof: Using the indicator function,

$$\mathrm{I}\left(a, A\right) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{otherwise,} \end{cases}$$

we can write

$$\begin{aligned} \operatorname*{argmax}_{x \in X} \sum_{S' \subseteq S(x)} \Gamma(S') &= \operatorname*{argmax}_{x \in X} \sum_{\substack{S' \subseteq S(x) \\ S' \neq \emptyset}} \frac{\left|\{k \,:\, S \subseteq S^k \text{ for } 1 \le k \le K\}\right|}{K} \\ &= \operatorname*{argmax}_{x \in X} \frac{1}{K} \sum_{\substack{S' \subseteq S(x) \\ S' \neq \emptyset}} \sum_{k=1}^{K} \mathrm{I}\left(S', \mathcal{P}\left(S^k\right)\right) \\ &= \operatorname*{argmax}_{x \in X} \sum_{\substack{S' \subseteq S(x) \\ S' \neq \emptyset}} \sum_{k=1}^{K} \mathrm{I}\left(S', \mathcal{P}\left(S^k\right)\right). \end{aligned}$$

Similarly, since $|S^k| = |S^1|$ for all k in $1, \ldots, K$, we have

$$\begin{aligned}
\operatorname*{argmax}_{x\in X} \sum_{k=1}^{K} \mathrm{Bel}_k\big(S(x)\big) &= \operatorname*{argmax}_{x\in X} \sum_{k=1}^{K} \sum_{S'\subseteq S(x)} m_k(S') \\
&= \operatorname*{argmax}_{x\in X} \sum_{\substack{S'\subseteq S(x)\\ S'\neq\emptyset}} \sum_{k=1}^{K} \frac{\mathrm{I}\left(S', \mathcal{P}\left(S^k\right)\right)}{2^{|S^k|}-1} \\
&= \operatorname*{argmax}_{x\in X} \frac{1}{2^{|S^1|}-1} \sum_{\substack{S'\subseteq S(x)\\ S'\neq\emptyset}} \sum_{k=1}^{K} \mathrm{I}\left(S', \mathcal{P}\left(S^k\right)\right) \\
&= \operatorname*{argmax}_{x\in X} \sum_{\substack{S'\subseteq S(x)\\ S'\neq\emptyset}} \sum_{k=1}^{K} \mathrm{I}\left(S', \mathcal{P}\left(S^k\right)\right).
\end{aligned}$$

Thus, the two formulations differ only by a constant multiplier and are optimized by the same solution set. ■

In words, Theorem 4.5 says that the following are equivalent:

1. V is a *most-robust* selection set.
2. V is a selection set with maximum total belief.
3. V is a selection set with maximum total global selection persistence factor.

The condition that all $x \in X^K$ have the same sum holds for any problem with a constant-sum constraint, $\sum_j x_j = N$, such as sensor assignment without slack.

4.2.1 Conditional Beliefs

As with the singleton-only belief function, if we make an a priori decision to select or reject some item, we can conditionally revise our beliefs. Recall from

Section 4.1.1 that the general form of Dempster's conditional belief function is given by

$$\mathrm{Bel}(x \,|\, X_j) = \frac{\mathrm{Bel}(x \vee \neg X_j) - \mathrm{Bel}(\neg X_j)}{1 - \mathrm{Bel}(\neg X_j)} = \frac{\mathrm{Bel}(x \vee \overline{X}_j) - \mathrm{Bel}(\overline{X}_j)}{1 - \mathrm{Bel}(\overline{X}_j)},$$

where we have defined $X_j = \{x \in X : x_j = 1\}$ and $\overline{X}_j = \{x \in X : x_j = 0\}$. (Note that $\mathrm{Bel}(x \,|\, X_j)$ is *not* the same as $\mathrm{Bel}(x \,|\, S(x) = \{j\})$.) We specify this rule for the decision-sets belief function in the following theorem.

Theorem 4.6. *For the decisions-sets bpa,*

$$\mathrm{Bel}(x \,|\, X_j) = \frac{\sum_{y\,:\,S(y)\subseteq S(x)\wedge y_j=1} \mathrm{m}\big(S(y)\big)}{\sum_{y\,:\,y_j=1} \mathrm{m}\big(S(y)\big)}.$$

where $\mathrm{Bel}(x \,|\, X_j)$ *is the belief in* x *given we know* $x_j = 1$.

Proof: We have

$$\begin{aligned}
\mathrm{Bel}(x \,|\, X_j) &= \frac{\mathrm{Bel}(x \vee \overline{X}_j) - \mathrm{Bel}(\overline{X}_j)}{1 - \mathrm{Bel}(\overline{X}_j)} \\
&= \frac{\sum_{y\,:\,S(y)\subseteq S(x)\vee y_j=0} \mathrm{m}\big(S(y)\big) - \sum_{y\,:\,y_j=0} \mathrm{m}\big(S(y)\big)}{1 - \sum_{y\,:\,y_j=0} \mathrm{m}\big(S(y)\big)} \\
&= \frac{\sum_{y\,:\,S(y)\subseteq S(x)\wedge y_j=1} \mathrm{m}\big(S(y)\big)}{1 - \sum_{y\,:\,y_j=0} \mathrm{m}\big(S(y)\big)} = \frac{\sum_{y\,:\,S(y)\subseteq S(x)\wedge y_j=1} \mathrm{m}\big(S(y)\big)}{\sum_{y\,:\,y_j=1} \mathrm{m}\big(S(y)\big)}.
\end{aligned}$$ ■

In Table 4.4, we illustrate some revised belief values for the belief function shown in Table 4.3. Fixing $x_1 = 1$ gives the most dramatic results since very few of the considered solutions select x_1. Additionally, the final column shows the results of conditioning on two selections, $x_4 = 1$ and $x_6 = 1$, and thus is nonzero for only those solutions that make both selections.

Table 4.4: Decision-sets conditional beliefs.

$\mathrm{Bel}(x)$	$\mathrm{Bel}(x\vert X_1)$	$\mathrm{Bel}(x\vert X_4)$	$\mathrm{Bel}(x\vert X_6)$	$\mathrm{Bel}(x\vert X_9)$	$\mathrm{Bel}(x\vert X_4 \cap X_6)$
0.4292	0.7480	0.5254	0.4527	0.4207	0.5283
0.4378	0.0000	0.5000	0.4811	0.4311	0.5283
0.4372	0.0000	0.4992	0.4609	0.4768	0.4991
0.3806	0.0000	0.4492	0.4223	0.4079	0.4717
0.2433	0.4390	0.0000	0.2813	0.2987	0.0000
0.2476	0.5020	0.3398	0.2955	0.0000	0.3641
0.3991	0.0000	0.4738	0.4217	0.4536	0.4708
0.2286	0.0000	0.3016	0.0000	0.2923	0.0000
0.2565	0.5630	0.3516	0.2948	0.0000	0.3631
0.3088	0.0000	0.3516	0.3289	0.3668	0.3349
0.3333	0.0000	0.0000	0.3835	0.3851	0.0000
0.3330	0.0000	0.0000	0.3734	0.4079	0.0000
0.2614	0.0000	0.0000	0.2999	0.3323	0.0000
0.2608	0.0000	0.3389	0.3090	0.0000	0.3631

4.3 Multiple Objectives

Another source of evidence comes when considering solutions from multiple objectives. In multi-objective optimization, though, the various objectives are usually at least partially conflicting. In this section, we investigate a couple of approaches for integrating evidence produced by multiple objectives. In both cases, each objective is associated with a distinct solution set. In the first, we treat each solution set and associated persistence values as a distinct pool of evidence, construct a belief function from each, and combine them using one of the rules for combining evidence (we consider both Dempster's rule and the modified discount & combine method). The second takes an alternative approach of attempting to merge the solution sets prior to computing persistence values.

4.3.1 Combining Multiple Objectives as Pools of Evidence

Let $X^{K_1} \subseteq X^1$, $X^{K_2} \subseteq X^2$ denote two distinct solution sets produced by considering two objectives, and let m_1 and m_2 be their associated bpas. We combine these with Dempster's Rule:

$$\mathrm{m}_{12}(\emptyset) = 0,$$

$$\mathrm{m}_{12}(Y) = \frac{\sum\limits_{U \cap V = Y} \mathrm{m}_1(U)\,\mathrm{m}_2(V)}{1 - c},$$

where

$$c = \sum_{U \cap V = \emptyset} \mathrm{m}_1(U)\,\mathrm{m}_2(V).$$

Since Dempster's rule fails when the sets to be combined do not share any focal elements (in other words, when $c = 1$), we require that $X^1 \cap X^2 \neq \emptyset$ for this approach.

Theorem 4.7. *For the singleton-only bpa,*

$$\begin{aligned}
\mathrm{m}_{12}(x) &= 0 && \textit{for } x \notin X^1 \cap X^2 \\
\mathrm{m}_{12}(x) &= \frac{\mathrm{m}_1(x)\,\mathrm{m}_2(x)}{\sum\limits_{u \in X^1 \cap X^2} \mathrm{m}_1(u)\,\mathrm{m}_2(u)} && \textit{for } x \in X^1 \cap X^2 \\
\mathrm{m}_{12}(Y) &= 0 && \textit{for } Y \,:\, |Y| > 1
\end{aligned}$$

and

$$\mathrm{Bel}_{12}(Y) = \sum_{y \in Y} \mathrm{m}_{12}(y) = \frac{\sum\limits_{y \in Y} \mathrm{m}_1(y)\,\mathrm{m}_2(y)}{\sum\limits_{u \in X^1 \cap X^2} \mathrm{m}_1(u)\,\mathrm{m}_2(u)}.$$

Proof: Since $U \cap V = Y$ with $|Y| > 1$ implies $|U| > 1$ and $|V| > 1$, and $\mathrm{m}_1(U)\,\mathrm{m}_2(V) = 0$ for $|U| > 1$ or $|V| > 1$, Dempster's rule becomes:

$$\begin{aligned}
\mathrm{m}_{12}(x) &= 0 && \text{for } x \notin X^1 \cap X^2 \\
\mathrm{m}_{12}(x) &= \frac{\mathrm{m}_1(x)\,\mathrm{m}_2(x)}{1-c} && \text{for } x \in X^1 \cap X^2 \\
\mathrm{m}_{12}(Y) &= 0 && \text{for } Y \,:\, |Y| > 1,
\end{aligned}$$

where

$$\begin{aligned}
c &= \sum_{U \cap V = \emptyset} \mathrm{m}_1(U)\,\mathrm{m}_2(V) \\
&= \sum_{\substack{u \in X^1, v \in X^2 \\ u \neq v}} \mathrm{m}_1(u)\,\mathrm{m}_2(v) \\
&= \left(\sum_{u \in X^1} \mathrm{m}_1(u)\right)\left(\sum_{v \in X^2} \mathrm{m}_2(v)\right) - \sum_{u \in X^1 \cap X^2} \mathrm{m}_1(u)\,\mathrm{m}_2(u) \\
&= 1 - \sum_{u \in X^1 \cap X^2} \mathrm{m}_1(u)\,\mathrm{m}_2(u).
\end{aligned}$$

Further,

$$\mathrm{Bel}_{12}(Y) = \sum_{U \subseteq Y} \mathrm{m}_{12}(U) = \sum_{y \in Y} \mathrm{m}_{12}(y) = \frac{\sum\limits_{y \in Y} \mathrm{m}_1(y)\,\mathrm{m}_2(y)}{\sum\limits_{u \in X^1 \cap X^2} \mathrm{m}_1(u)\,\mathrm{m}_2(u)}.$$

■

For the purpose of illustration, consider if we add a second objective to the treasure-room knapsack problem. This could be the relative historical values of the various treasures as assigned by the curator of the local museum for instance (for this example, though, we use a random perturbation of the original values).

Let our second objective be:

$$\max \ \hat{c}x,$$

where

$$\hat{c} = [\ 46\ 40\ 41\ 36\ 10\ 11\ 11\ 44\ 5\ 22\ ].$$

Table 4.5: Second-objective treasure-room knapsack problem solution sets.

k	Objective Value	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\mathrm{Bel}_2^{\mathrm{so}}(x)$	$\mathrm{Bel}_2^{\mathrm{ds}}(x)$
		Knapsack Assignment											
1	112	0	0	1	0	0	1	1	1	1	0	0.0733	0.4740
2	111	0	0	1	0	1	1	0	1	1	0	0.0717	0.4178
3	109	1	0	1	0	0	1	1	0	0	0	0.0634	0.2836
4	109	1	0	0	1	0	1	1	0	1	0	0.0799	0.4178
5	107	0	0	1	0	0	1	1	1	0	0	0.0651	0.2968
6	107	0	0	0	1	0	1	1	1	1	0	0.0815	0.4163
7	106	1	0	0	0	0	1	0	1	1	0	0.0750	0.3161
8	106	0	0	1	0	1	1	0	1	0	0	0.0634	0.2544
9	106	1	0	0	0	0	0	1	1	1	0	0.0618	0.2691
10	106	0	0	0	1	1	1	0	1	1	0	0.0799	0.3508
11	105	1	0	0	0	1	0	0	1	1	0	0.0601	0.2363
12	104	0	0	1	1	0	1	1	0	1	0	0.0799	0.3883
13	104	1	0	0	1	0	1	1	0	0	0	0.0717	0.2544
14	103	1	0	1	0	0	1	0	0	1	0	0.0733	0.2879
	Total	4	4	4	10	5	13	6	6	11	0		

In Table 4.5, we show the solutions generated by applying the method of successive exclusions to the treasure-room knapsack problem with this second objective along with the resulting singleton-only, $\mathrm{Bel}^{\mathrm{so}}$, and decision-sets, $\mathrm{Bel}^{\mathrm{ds}}$, belief values. We illustrate the use of Dempster's rule to combine the singleton-only belief functions, $\mathrm{Bel}_1^{\mathrm{so}}$ and $\mathrm{Bel}_2^{\mathrm{so}}$, in Table 4.6, which gives the post-combination beliefs and solutions for the situation in which the bpas are constructed by limiting the solution spaces to the respective solution lists (that is, $X^1 = X^{K_1}$ and $X^2 = X^{K_2}$). Since, in this case, we are combining singleton-only bpas, only the common solutions, $X^1 \cap X^2 = X^{K_1} \cap X^{K_2}$, have non-zero belief values.

Table 4.6: Combined singleton-only beliefs for the common solutions.

Objective Values		Knapsack Assignment												
		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\mathrm{Bel}_1^{\mathrm{so}}(x)$	$\mathrm{Bel}_2^{\mathrm{so}}(x)$	$\mathrm{Bel}_{12}^{\mathrm{so}}(x)$
107	104	1	0	0	1	0	1	1	0	0	0	0.0687	0.0717	0.1254
112	109	1	0	0	1	0	1	1	0	1	0	0.0750	0.0799	0.1527
107	104	0	0	1	1	0	1	1	0	1	0	0.0750	0.0799	0.1527
106	111	0	0	1	0	1	1	0	1	1	0	0.0671	0.0717	0.1225
110	106	0	0	0	1	1	1	0	1	1	0	0.0766	0.0799	0.1559
105	112	0	0	1	0	0	1	1	1	1	0	0.0687	0.0733	0.1283
109	107	0	0	0	1	0	1	1	1	1	0	0.0782	0.0815	0.1624

For the decision-sets belief function the necessity for $c < 1$ entails a more stringent requirement: it is not sufficient that the acceptable solution domains intersect (that is, $X^1 \cap X^2 \neq \emptyset$), we must further require that the respective selection sets intersect (that is, $S(X^{K_1}) \cap S(X^{K_2}) \neq \emptyset$). Moreover, Dempster's rule assigns non-zero beliefs only to subsets of this intersection. Thus, when working with objectives that produce largely disjoint selection sets, the results of combining decision-sets belief functions with Dempster's rule is of dubious quality. In these cases, using a rule like the modified discount & combine method is more appropriate.

We illustrate the combination of decision-sets belief functions in Table 4.7, where we show the results of both Dempster's rule and the modified dis-

Table 4.7: Combined decision-sets belief function values for solution sets union.

Objective Values		Knapsack Assignment										Dempster's	mod-D&C
		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\mathrm{Bel}_{12}^{\mathrm{ds}}(x)$	$\mathrm{Bel}_{12}^{\mathrm{ds}}(x)$
112	109	1	0	0	1	0	1	1	0	1	0	0.4104	0.3144
110	106	0	0	0	1	1	1	0	1	1	0	0.4043	0.2982
109	107	0	0	0	1	0	1	1	1	1	0	0.4395	0.3232
108	103	1	0	0	1	1	1	0	0	0	0	0.2431	0.1651
108	102	1	1	0	0	0	1	0	0	1	0	0.2899	0.1793
108	103	0	0	1	1	1	1	0	0	1	0	0.3675	0.2608
107	104	1	0	0	1	0	1	1	0	0	0	0.2768	0.1969
107	92	0	1	0	1	0	1	0	0	1	0	0.3118	0.1914
107	92	0	1	0	1	0	0	1	0	1	0	0.1867	0.1331
107	104	0	0	1	1	0	1	1	0	1	0	0.4055	0.2982
106	111	0	0	1	0	1	1	0	1	1	0	0.3795	0.2827
105	101	0	0	0	1	1	1	0	1	0	0	0.2681	0.1803
105	100	0	1	0	0	0	1	0	1	1	0	0.3192	0.1933
105	112	0	0	1	0	0	1	1	1	1	0	0.4108	0.3045
103	109	1	0	1	0	0	1	1	0	0	0	0.2495	0.1714
101	106	0	0	1	0	1	1	0	1	0	0	0.2466	0.1691
101	105	1	0	0	0	1	0	0	1	1	0	0.1959	0.1415
100	107	0	0	1	0	0	1	1	1	0	0	0.2735	0.1840
100	106	1	0	0	0	0	1	0	1	1	0	0.3543	0.2234
100	106	1	0	0	0	0	0	1	1	1	0	0.2226	0.1586
98	103	1	0	1	0	0	1	0	0	1	0	0.3215	0.2010

count & combine method (with $\alpha_1 = \alpha_2 = 0.5$). Since the intersection of the two selection sets, $S(X^{K_1}) \cap S(X^{K_2}) = \{1, 3, 4, 5, 6, 7, 8, 9\}$, is substantial, either method produces "reasonable" results. The resulting beliefs differ in that the bpa produced by the modified discount & combine method assigns weight to subsets of the union of the selection sets, while Dempster's rule gives weight only to subsets of the intersection of the selections sets.

4.3.2 Computing Persistence from Multiple Objectives

Let X^{K_1}, X^{K_2} be solution sets for two objectives from which their persistences are computed. One could use $X^K = X^{K_1} \cup X^{K_2}$ (where K is the number in their union), in which case nothing new is needed. However, one may intuit that solutions in their intersection are more robust against uncertainties.

Definition 4.8. Two solution sets are in *total conflict* if $X^{K_1} \cap X^{K_2} = \emptyset$. Otherwise, they are *compatible*.

If they are compatible, we could reduce to their common solutions, $X^{K_1} \cap X^{K_2}$, from which persistence values can be computed to yield a robust solution. Or, we could use the multiset union, letting $K = K_1 \uplus K_2$, so that more weight is given to common solutions and compute persistence using the combined set. However, if the solution sets are in total conflict, the multiset union simply appends one list onto the other, and we must seriously question whether the resulting persistences are meaningful. Moreover, when the solution sets are in total conflict, the approach used in Section 4.3.1 is generally not appropriate:

Dempster's rule fails because $1 - c = 0$.

Yager's rule fails because $\mathrm{m}(A) = 0$ for all A.

Dubois & Prade's rule and the **discount & combine rule** do not produce meaningful results for belief functions, such as the singleton-only bpa, whose focal elements are solution vectors (but may give meaningful results for the decision-sets belief function, since its focal elements are individual decisions).

For the remainder of this section, we suppose at least some subset of the objectives are in total conflict and investigate how to combine the evidence their solution sets provide. For two objectives, let $X^K = X^{K_1} \cup X^{K_2}$ and compute the persistences from X^K, where $K = K_1 + K_2$ since $X^{K_1} \cap X^{K_2} = \emptyset$.

For M objectives, where $M > 2$, our first phase is to partition the collection of solution sets (1 set per objective) into groups such that each group has a non-empty intersection. (In particular, two objectives in conflict result in two groups.) Among such partitions, choose a minimum number of groups. The idea is to represent each of the groups by the solutions in its intersection, and focus on the situation where the collection is in total conflict — every inter-group intersection is empty.

First, we must consider how to allocate groups. To help fix ideas, suppose $M = 3$ and

$$X^{K_1} = \{x^1, x^2\},\ X^{K_2} = \{x^2, x^3\},\ X^{K_3} = \{x^3, x^4\}. \tag{4.5}$$

There are two possible allocations that minimize the number of groups:

$$G_1^1 = \{X^{K_1}, X^{K_2}\}, \qquad G_2^1 = \{X^{K_3}\} \tag{4.6a}$$

$$G_1^2 = \{X^{K_1}\}, \qquad G_2^2 = \{X^{K_2}, X^{K_3}\} \tag{4.6b}$$

We require that the groups satisfy the following properties:

(P1) **Assignment.** Every solution set is assigned to exactly one group.

(P2) **Inter-group conflict.** $G \neq G' \implies X(G) \cap X(G') = \emptyset$.

(P3) **Intra-group compatibility.** $X(G) \neq \emptyset$.

We could use the assignment model with conflict constraints to form groups that satisfy the first two conditions. However, the intra-group compatibility condition is more difficult because we could have all pairs of sets compatible but not have a complete non-empty intersection.

Let $\alpha_{mg} = 1$ if objective set X^{K_m} is assigned to group g, so the constraints of a conflict assignment model with N_g groups are

$$\sum_g \alpha_{mg} = 1 \quad \text{for} \quad m \in \{1, \ldots, M\}, \tag{4.7a}$$

$$\sum_{i=1}^{\ell} \alpha_{m_i g} \leq \ell - 1 \quad \text{for} \quad \begin{array}{l} g \in \{1, \ldots, N_g\}, \text{ and} \\ (m_1, \ldots, \mathrm{m}_\ell) \in \mathcal{C}_\ell \text{ with } \ell = 2, \ldots, M-1, \end{array} \tag{4.7b}$$

where $\mathcal{C}_\ell = \{(\mathrm{m}_1, \ldots, \mathrm{m}_\ell) \, : \, X^{K_{m_1}} \cap \cdots \cap X^{K_{m_\ell}} = \emptyset\}$. The first constraint (4.7a) ensures that every solution set is assigned to exactly one group (P1), while the second (4.7b) enforces the requirement that every group have a non-empty intersection (P3).

The requirement that every inter-group intersection is empty (P2) is enforced not by the model constraints, but instead results from optimization. If we minimize N_g for $\alpha \in \{0,\, 1\}^{M \times N_g}$ by seeking a solution for $N_g = 1, 2, \ldots, M$, and stopping the first time a feasible solution is found, the resulting assignment must be such that every inter-group intersection is empty, otherwise we could find an assignment of $N_g - 1$ groups that would be feasible.

To illustrate this assignment model, consider the $M = 3$ solution sets from (4.5). In this example, we have only one conflict set, $\mathcal{C}_2 = \{(1, 3)\}$. So, for $N_g = 1$, (4.7b) requires that $\alpha_{11} + \alpha_{31} \leq 1$, but (4.7a) gives $\alpha_{11} = \alpha_{21} = \alpha_{31} = 1$.

Hence, the assignment is infeasible for $N_g = 1$ and we must increase N_g. When $N_g = 2$, the assignment constraints are:

$$(4.7\text{a}) \begin{cases} \alpha_{11} + \alpha_{12} = 1 \\ \alpha_{21} + \alpha_{22} = 1 \\ \alpha_{31} + \alpha_{32} = 1 \end{cases} \quad \text{and} \quad (4.7\text{b}) \begin{cases} \alpha_{11} + \alpha_{31} \leq 1 \\ \alpha_{12} + \alpha_{32} \leq 1 \end{cases}.$$

In this case, there are two distinct feasible assignments. First, in order to break symmetries, we fix $\alpha_{11} = 1$. This forces $\alpha_{12} = 0$ and $\alpha_{31} = 0$, then $\alpha_{31} = 0$ forces $\alpha_{32} = 1$ giving $X^{K_1} \in G_1^{\bullet}$ and $X^{K_3} \in G_2^{\bullet}$. Finally, the only restriction we have on α_{21} and α_{22} is that $\alpha_{21} + \alpha_{22} = 1$, so either $\alpha_{12} = 1$ ($\implies \alpha_{22} = 0$) or $\alpha_{22} = 1$ ($\implies \alpha_{12} = 0$) is feasible. Hence, both groupings, (4.6a) and (4.6b), represent optimal solutions to the assignment model.

Once we have assigned the objectives to groups, the second phase is to let each group be represented by the intersection of its members:

$$X(G) = \bigcap_{X^{K_m} \in G} X^{K_m}.$$

Thus, in our previous example we have:

$$X(G_1^1) = \{x^2\}, \qquad X(G_2^1) = \{x^3, x^4\} \tag{4.8a}$$

$$X(G_1^2) = \{x^1, x^2\}, \qquad X(G_2^2) = \{x^3\} \tag{4.8b}$$

Finally, we construct a merged solution set by taking the union of the groups and compute persistence values from resulting list. Thus, in our example, we would compute persistence values from either:

$$X^1 = \{x^2, x^3, x^4\} \qquad \text{or} \qquad X^2 = \{x^1, x^2, x^3\}.$$

4.4 Persistence and Generalized Information Theory

In the rest of this chapter, we discuss some relations between persistence as evidence and information theory. We begin by showing a relationship between optimizing with respect to global persistence and minimizing a Hamming distance. We build upon these results by further relating persistence to generalized information theory and investigating a relationship between the size of the solution set, X^K, and *entropy*. Finally, we conclude this section with a discussion of a counterintuitive, and seemingly paradoxical, relationship between uncertainty and the quantity of evidence considered that is suggested by our results.

4.4.1 Hamming Distance

In (2.7) and (2.8), we generated the global persistence values for selections and rejections by counting the number of elements in the respective multisets. An alternative method of generation, which offers some notational convenience, constructs the persistence vector directly from the K acceptable solutions as 0-1 vectors. Hence, we observe that

$$p = \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{bmatrix} = \tfrac{1}{K}\sum_{k=1}^{K} x^k \text{ and } \widetilde{p} = \begin{bmatrix} \widetilde{\gamma}(1) \\ \vdots \\ \widetilde{\gamma}(n) \end{bmatrix} = \tfrac{1}{K}\sum_{k=1}^{K} \left(\vec{\mathbf{1}} - x^k\right), \tag{4.9}$$

where $\widetilde{p} = \vec{\mathbf{1}} - p$ is the persistence vector of *rejections*. Using this notation, we can then write the total global persistence of selections and rejections for a vector $x \in \{0,1\}^n$ as

$$\sum_{s\,:\,x_s=1} \gamma(s) = p^{\mathsf{T}} x \quad \text{and} \quad \sum_{r\,:\,x_r=0} \widetilde{\gamma}(r) = \widetilde{p}^{\mathsf{T}}(\vec{\mathbf{1}} - x). \tag{4.10}$$

Since solutions are 0-1 vectors, we can relate global persistence to Hamming distance:

$$H(x,y) = \sum_i |x_i - y_i| \, .$$

Define the total combined global persistence of a vector $x \in \{0,1\}^n$ as:

$$\psi(x) = \sum_{s\,:\,x_s=1} \gamma(s) + \sum_{r\,:\,x_r=0} \widetilde{\gamma}(r) = p^{\mathsf{T}} x + \widetilde{p}^{\mathsf{T}}(\vec{\mathbf{1}} - x). \tag{4.11}$$

Then, Theorem 4.10 shows that a vector, x, that maximizes total global persistence also minimizes the sum of Hamming distances,

$$\mathcal{H}(x, X^K) = \sum_{k=1}^{K} H(x, x^k).$$

Lemma 4.9. $H(x,y) = x^{\mathsf{T}}(\vec{\mathbf{1}} - y) + (\vec{\mathbf{1}} - x)^{\mathsf{T}} y.$

Proof:
$$\begin{aligned} H(x,y) &= |\{i \,:\, x_i = 1 \wedge y_i = 0\}| + |\{i \,:\, x_i = 0 \wedge y_i = 1\}| \\ &= x^{\mathsf{T}}(\vec{\mathbf{1}} - y) + (\vec{\mathbf{1}} - x)^{\mathsf{T}} y. \end{aligned}$$
■

Theorem 4.10. *Given a collection of solution vectors, $\{x^k\}$, for $k = 1, \ldots, K$,*

$$\underset{x \in X}{\text{argmax}}\ \psi(x) = \underset{x \in X}{\text{argmin}}\ \sum_{k=1}^{K} H(x, x^k)$$

for $\emptyset \neq X \subseteq \{0,1\}^n$.

Proof: Using Lemma 4.9, we have,

$$\begin{aligned}
\underset{x \in X}{\text{argmin}} \sum_{k=1}^{K} H(x, x^k) &= \underset{x \in X}{\text{argmin}} \sum_{k=1}^{K} \left[x^\mathsf{T}(\vec{\mathbf{1}} - x^k) + (\vec{\mathbf{1}} - x)^\mathsf{T} x^k \right] \\
&= \underset{x \in X}{\text{argmin}} \sum_{k=1}^{K} \left[x^\mathsf{T}(\vec{\mathbf{1}} - x^k) + \vec{\mathbf{1}}^\mathsf{T} x^k - x^\mathsf{T} x^k \right] \\
&= \underset{x \in X}{\text{argmax}} \sum_{k=1}^{K} \left[x^\mathsf{T} x^k - \underbrace{\vec{\mathbf{1}}^\mathsf{T} x^k}_{\text{constant}} - x^\mathsf{T}(\vec{\mathbf{1}} - x^k) \right] \\
&= \underset{x \in X}{\text{argmax}} \sum_{k=1}^{K} \left[x^\mathsf{T} x^k - x^\mathsf{T}(\vec{\mathbf{1}} - x^k) \right].
\end{aligned}$$

Substituting the summations from (4.9) into (4.11), we have

$$\begin{aligned}
\underset{x \in X}{\text{argmax}}\, \psi(x) &= \underset{x \in X}{\text{argmax}} \left[\sum_{s\,:\,x_s=1} \gamma(s) + \sum_{r\,:\,x_r=0} \widetilde{\gamma}(r) \right] \\
&= \underset{x \in X}{\text{argmax}} \left[x^\mathsf{T} \left(\tfrac{1}{K} \sum_{k=1}^{K} x^k \right) + (\vec{\mathbf{1}} - x)^\mathsf{T} \left(\tfrac{1}{K} \sum_{k=1}^{K} \left(\vec{\mathbf{1}} - x^k \right) \right) \right] \\
&= \underset{x \in X}{\text{argmax}}\, \tfrac{1}{K} \left[\sum_{k=1}^{K} x^\mathsf{T} x^k + \sum_{k=1}^{K} (\vec{\mathbf{1}} - x)^\mathsf{T} \left(\vec{\mathbf{1}} - x^k \right) \right] \\
&= \underset{x \in X}{\text{argmax}}\, \tfrac{1}{K} \sum_{k=1}^{K} \left[x^\mathsf{T} x^k + (\vec{\mathbf{1}} - x)^\mathsf{T} \left(\vec{\mathbf{1}} - x^k \right) \right]
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname*{argmax}_{x\in X} \sum_{k=1}^{K} \left[x^\top x^k + \underbrace{\vec{\mathbf{1}}^\top \left(\vec{\mathbf{1}} - x^k\right)}_{\text{constant}} - x^\top \left(\vec{\mathbf{1}} - x^k\right) \right] \\
&= \operatorname*{argmax}_{x\in X} \sum_{k=1}^{K} \left[x^\top x^k - x^\top \left(\vec{\mathbf{1}} - x^k\right) \right] \\
&= \operatorname*{argmin}_{x\in X} \sum_{k=1}^{K} H(x, x^k). \qquad \blacksquare
\end{aligned}
$$

(Aside: The only property we use for the L_1 norm is that $|x - y| = 0, 1$ for $x, y = 0, 1$. This extends to $|x - y|^p$, so the Theorem and Corollary extend to maximizing any L_p norm. Since we do not use the other norms explictly, we continue with H defined by L_1.)

A special case of the relation between global persistence and Hamming distances arises when we have a constant-sum constraint, $\sum_i x_i = N$. First, since $\widetilde{p} = \vec{\mathbf{1}} - p$, we can write total global persistence (4.11) as:

$$
\begin{aligned}
\psi(x) &= p^\top x + \widetilde{p}^\top (\vec{\mathbf{1}} - x) \\
&= p^\top x + (\vec{\mathbf{1}} - \widetilde{p})^\top (\vec{\mathbf{1}} - x) \\
&= p^\top x + \vec{\mathbf{1}}^\top \vec{\mathbf{1}} - \vec{\mathbf{1}}^\top x - p^\top \vec{\mathbf{1}} + p^\top x \\
&= 2p^\top x - \vec{\mathbf{1}}^\top x + \vec{\mathbf{1}}^\top \vec{\mathbf{1}} - \vec{\mathbf{1}}^\top p \\
&= (2p - \vec{\mathbf{1}})^\top x + n - \|p\|_1.
\end{aligned}
$$

Note that the last two terms are constant with respect to x. Moreover, under the constant-sum constraint, $\vec{\mathbf{1}}^{\mathsf{T}} x = N$, we have:

$$\begin{aligned}
\underset{x \in X}{\operatorname{argmax}}\ \psi(x) &= \underset{x \in X}{\operatorname{argmax}}\ 2p^{\mathsf{T}} x - \vec{\mathbf{1}}^{\mathsf{T}} x + n - \|p\|_1 \\
&= \underset{x \in X}{\operatorname{argmax}}\ 2p^{\mathsf{T}} x - \underbrace{N + n - \|p\|_1}_{\text{constant}} \\
&= \underset{x \in X}{\operatorname{argmax}}\ p^{\mathsf{T}} x \\
&= \underset{x \in X}{\operatorname{argmax}} \sum_{s \,:\, x_s = 1} \gamma(s).
\end{aligned}$$

This is simply the total global persistence of selections. We have thus shown the following.

Corollary 4.11. *Given a collection of solution vectors, $\{x^k\}$, for $k = 1, \ldots, K$,*

$$\underset{x \in X}{\operatorname{argmax}} \sum_{s \,:\, x_s = 1} \gamma(s) = \underset{x \in X}{\operatorname{argmin}} \sum_{k=1}^{K} H(x, x^k)$$

for $\emptyset \neq X \subseteq \left\{ x \in \{0,1\}^n \,:\, x^{\mathsf{T}} \vec{\mathbf{1}} = N \right\}$.

4.4.2 Entropy

Yager [33] generalized Shannon's measure of information in the DST (also see Lamata and Moral [21] and Klir [17, 18, 19]):

$$E(m) = -\sum_{x \in X} m(x) \log_2 \operatorname{Plaus}(x),$$

where $\mathrm{Plaus}(x) = 1 - \mathrm{Bel}(\neg x)$ is the plausibility of x, and $\mathrm{m}(x) \log_2 \mathrm{Plaus}(x) \stackrel{\mathrm{def}}{=} 0$ when $\mathrm{Plaus}(x) = 0$. Using any Bayesian structure (notably, (4.1)), Yager showed

$$E(m) = -\sum_{x \in X} \mathrm{m}(x) \log_2 \mathrm{m}(x), \tag{4.12}$$

where $0 \log_2 0 \stackrel{\mathrm{def}}{=} 0$. It is well known that the entropy function is strictly concave. This follows from the strict convexity of each term in the sum:

$$\frac{\mathrm{d}^2(v \log_2 v)}{\mathrm{d}v^2} = \frac{1}{v} > 0.$$

Let m_s be the bpa for considering only selections, and let m_r be the bpa for considering only rejections. Henceforth, we assume: $x_j = 1$ for some decision j and solution $x \in X^K$; $x_j = 0$ for some decision j and solution $x \in X^K$; and, $X \supseteq X^K$. (We could have $X \supset X^K$, for instance X could be equal to the entire feasible region, when we consider choosing a solution that minimizes the sum of Hamming distances to X^K, or, equivalently, maximizes total persistence.) Then,

$$\kappa_s = \sum_{x \in X} p^\mathsf{T} x > 0 \qquad\qquad \kappa_r = \sum_{x \in X} (\vec{\mathbf{1}} - p)^\mathsf{T} (\vec{\mathbf{1}} - x) > 0,$$

so for $x \in X$:

$$m_s(x) = \frac{p^\mathsf{T} x}{\kappa_s} \qquad\qquad m_r(x) = \frac{(1 - p)^\mathsf{T}(1 - x)}{\kappa_r}.$$

Theorem 4.12. *For* $X = X^K = \{x^1, \ldots, x^K\}$*,*

$$m_s(x) = \frac{p^\mathsf{T} x}{K \|p\|^2}, \qquad m_r(x) = \frac{(1-p)^\mathsf{T}(1-x)}{K \|1-p\|^2}.$$

Proof: Since $p = \frac{1}{K}\sum_{k=1}^{K} x^k$, we have

$$\begin{aligned}
\kappa_s &= \sum_{x \in X^K} p^\mathsf{T} x = \sum_{k=1}^{K} p^\mathsf{T} x^k = p^\mathsf{T} K p = K \|p\|^2, \\
\kappa_r &= \sum_{x \in X^K} (1-p)^\mathsf{T} x = \sum_{k=1}^{K} (1-p)^\mathsf{T} x^k = (1-p)^\mathsf{T} K(1-p) \\
&= K \|1-p\|^2.
\end{aligned}$$

■

Let $m = \alpha m_s + (1-\alpha) m_r$ for some $\alpha \in [0, 1]$. In particular, the bpa defined in (4.1) corresponds to choosing $\alpha = \dfrac{\kappa_s}{\kappa_s + \kappa_r}$. The belief function is the same convex combination: $\mathrm{Bel}(x) = \alpha\,\mathrm{Bel}_s(x) + (1-\alpha)\,\mathrm{Bel}_r(x)$. However, the entropy is generally not the same linear combination, as shown in the following.

Theorem 4.13. *Suppose* $E(m) = \beta E(m_s) + (1-\beta) E(m_r)$*. Then,*

$$E(m_s) > E(m_r) \implies \beta > \alpha \quad \textit{and} \quad E(m_s) < E(m_r) \implies \beta < \alpha.$$

Proof: From the strict concavity of E,

$$\begin{aligned}
\beta E(m_s) + (1-\beta)E(m_r) &= E(m) \\
&= E(\alpha m_s + (1-\alpha)m_r) \\
&> \alpha E(m_s) + (1-\alpha)E(m_r) \\
\implies \beta\big(E(m_s) - E(m_r)\big) &> \alpha\big(E(m_s) - E(m_r)\big).
\end{aligned}$$

This yields the desired result. ∎

Consider the simple example:

$$X^K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The maximum-persistence solution is $x = x^3 = 0$ with minimum sum of Hamming distances, $\mathcal{H}(0, X^K) = 2$. (Note that $\mathcal{H}(x^k, X^K) = 3$ for $k = 1, 2$.) How much evidence is there for the added solution, $x = 0$? There is clear evidence that $x_3 = 0$, and there is some evidence that $x_2 = 0$ (by x^1) and that $x_1 = 0$ (by x^2). The "evidence" is persistence.

We have $p = (1/2, 1/2, 0)^\mathsf{T}$, so $m_s(X) = (1/2, 1/2, 0)^\mathsf{T}$ and $m_r(X) = (0.3, 0.3, 0.4)^\mathsf{T}$. The entropies are $E(m_s) = 1$ and $E(m_r) = 1.571$. This suggests that there is more uncertainty measuring by rejections.

A way to think of this is to consider the situation where $x = (?, ?, ?)$ as we gain the information necessary to progressively fill in the ?s. If we learn

$x_j = 1$, we are done — the other coordinates must be zero. However, if we learn $x_j = 0$, we are not done because the other coordinates could still be 0 or 1. If we ask, “$x_1 = 1$?” first, then ask “$x_2 = 1?''$, the average number of questions to determine x is $1/2 \times 1 + 1/2 \times 2 = 1.5$. On the other hand, if we ask, “$x_1 = 0$?” first, then ask if $x_2 = 0$?”, the average number of questions is $0.3 \times 1 + 0.7 \times 2 = 1.7$. Figure 4.1 shows the situation.

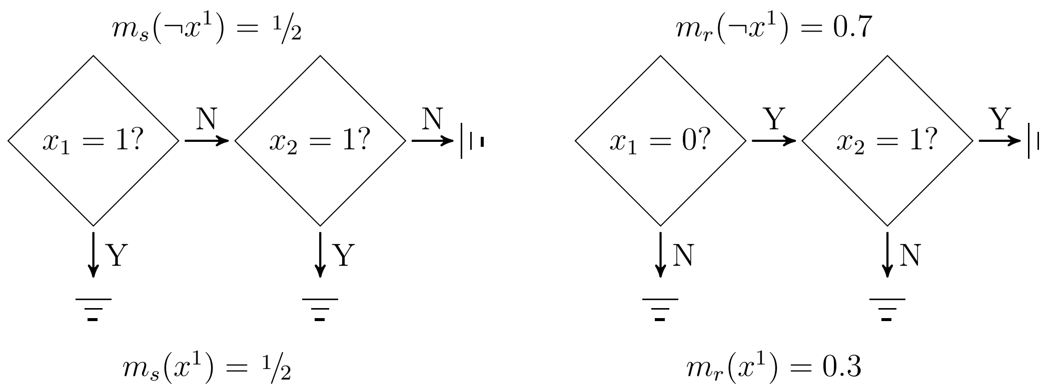


Figure 4.1: Question flow charts and information gain.

Consider the knapsack example with and without dominated solutions, shown in Table 2.1, and let $X = X^K$ in each case. We show the resulting bpa and belief values for both selections only and rejections only as well as the combined values in Table 4.8 (along with the associated α, β, and entropy scores). We see that across all three belief functions, the set with dominated solutions has more uncertainty (that is, greater entropy values) than its subset of undominated solutions.

Let $\mathcal{H}^*(X, X^K)$ be the minimum sum of Hamming distances from X to X^K:

$$\mathcal{H}^*(X, X^K) = \min_{x \in X} \mathcal{H}(x, X^K) = \min_{x \in X} \sum_{k=1}^{K} H(x, x^k).$$

Theorem 4.14. $Y \subset X^K \implies \mathcal{H}^*(X, Y) \leq \mathcal{H}^*(X, X^K)$, *with equality only if* $H(x^*, x^k) = 0$ *for all* $x^k \in X^K \setminus Y$ *and* $x^* \in \operatorname*{argmin}_{x \in X} \mathcal{H}(x, X^K)$.

Proof: Since $Y \subset X^K$, $\mathcal{H}(x^*, X^K) = \mathcal{H}(x^*, Y) + \mathcal{H}(x^*, X^K \setminus Y)$. Thus, $\mathcal{H}(x^*, X^K) \geq \mathcal{H}(x^*, Y) \geq \mathcal{H}(x^Y, Y)$, where the first inequality is strict if $\mathcal{H}(x^*, X^K \setminus Y) > 0$. ■

Theorem 4.14 suggests that it is better to use undominated solutions to obtain a robust solution that maximizes persistence (for any $\alpha \in [0, 1]$). This is consistent with using X^K with lesser entropy, meaning less uncertainty. More generally, any reduction in the set of candidates leads to a more robust maximum-persistence solution.

4.4.3 Evidence Paradox

We have seen that having a subset to determine persistence reduces the uncertainty, but the larger set has more evidence. This is the paradox we address by considering $X^K = \{x^1, \ldots, x^K\}$ for $K = 1, \ldots, K^{\max}$. Using X^1 results in a lower sum of Hamming distances than using X^2, and the entropy, visàvis the uncertainty, is less. Does this make the solution more robust? Adding x^2 as evidence gives us greater belief in a solution than using just x^1, so does X^2 yield a more robust solution? Therein lies the *evidence paradox*.

To resolve this paradox, we consider reduction of dominated solutions. Intuition suggests that this is better, consistent with the first property: fewer is better. But, that is because we used objective values in the dominance, which uses a value of merit not included in our computation of persistence. If we remove the optimum solution from X^K, we would believe the conflicting property: fewer is worse. Using objective values, however, does not reveal fully the paradox with respect to robustness. At issue is whether entropy is synonymous with uncertainty, and whether reduced uncertainty is synonymous with greater robustness.

This question can be partially settled by recalling that the purpose in considering a list of solutions is to address uncertainties in the underlying model or its parameters. The hope is that by considering a larger solution set, we have greater confidence in discovering the ex-post optimum (that is, the true optima once uncertainties have been resolved). In this context, the evidence paradox says that the solution set should be large enough, but not too large, and should reflect the uncertainties in the underlying model.

Table 4.8: Sample bpa and belief values with associated entropy scores.

(a) with dominated solutions

	$m_s^d(x^k)$	$\mathrm{Bel}_s^d(x^k)$	$m_r^d(x^k)$	$\mathrm{Bel}_r^d(x^k)$	$m^d(x^k)$	$\mathrm{Bel}^d(x^k)$
	0.0822	0.0822	0.0698	0.0698	0.0750	0.0750
	0.0841	0.0841	0.0711	0.0711	0.0766	0.0766
	0.0860	0.0860	0.0725	0.0725	0.0782	0.0782
	0.0804	0.0804	0.0684	0.0684	0.0735	0.0735
	0.0598	0.0598	0.0725	0.0725	0.0671	0.0671
	0.0598	0.0598	0.0725	0.0725	0.0671	0.0671
	0.0822	0.0822	0.0698	0.0698	0.0750	0.0750
	0.0579	0.0579	0.0711	0.0711	0.0656	0.0656
	0.0617	0.0617	0.0739	0.0739	0.0687	0.0687
	0.0710	0.0710	0.0807	0.0807	0.0766	0.0766
	0.0729	0.0729	0.0629	0.0629	0.0671	0.0671
	0.0748	0.0748	0.0643	0.0643	0.0687	0.0687
	0.0636	0.0636	0.0752	0.0752	0.0703	0.0703
	0.0636	0.0636	0.0752	0.0752	0.0703	0.0703
					$\alpha = 0.3317$	
$E(m_\bullet^d)$	3.7935		3.8253		3.8050 ($\beta = 0.639$)	

(Recall that $p^d = [\,4\ 4\ 4\ 10\ 5\ 13\ 6\ 6\ 11\ 0\,]^\mathsf{T}/14.$)

(b) without dominated solutions

	$m_s^u(x^k)$	$\mathrm{Bel}_s^u(x^k)$	$m_r^u(x^k)$	$\mathrm{Bel}_r^u(x^k)$	$m^u(x^k)$	$\mathrm{Bel}^u(x^k)$
	0.2111	0.2111	0.2000	0.2000	0.2053	0.2053
	0.1889	0.1889	0.1800	0.1800	0.1842	0.1842
	0.2222	0.2222	0.2100	0.2100	0.2158	0.2158
	0.2111	0.2111	0.2000	0.2000	0.2053	0.2053
	0.1667	0.1667	0.2100	0.2100	0.1895	0.1895
					$\alpha = 0.4737$	
$E(m_\bullet^u)$	2.3146		2.3289		2.3195 ($\beta = 0.657$)	

(Recall that $p^u = [\,1\ 1\ 1\ 5\ 1\ 5\ 3\ 2\ 5\ 0\,]^\mathsf{T}/5.$)

5. Applications

In the preceding chapters, we developed an approach to optimization under uncertainty that treats the persistence of individual decisions across a collection of "acceptable" solutions as evidence of the robustness of those decisions. Thus far, our development has focused on a relatively simple 0-1 knapsack problem. In this chapter, we demonstrate how the new paradigm of persistence yields robust solutions for substantive applications.

5.1 Sensor Placement

In this section, we investigate the problem of placing sensors in municipal water networks to detect contaminants that are maliciously or accidentally injected. Following Watson et al. [31] and Carr et al. [7], we are given a water network and N sensors and must decide to which placements the sensors should be assigned. Once a sensor is assigned to a placement, it cannot be moved, so the decision has no recourse.

Let v denote the point at which the contaminant is injected, our primary source of uncertainty. Given v, each sensor placement assignment x determines the effect of the contamination relative to that injection point and assignment as illustrated in Figure 5.1. The actual impact of the contamination is measured by multiple criteria:

Population Exposed (PE). Any node, i, in the network that is downstream

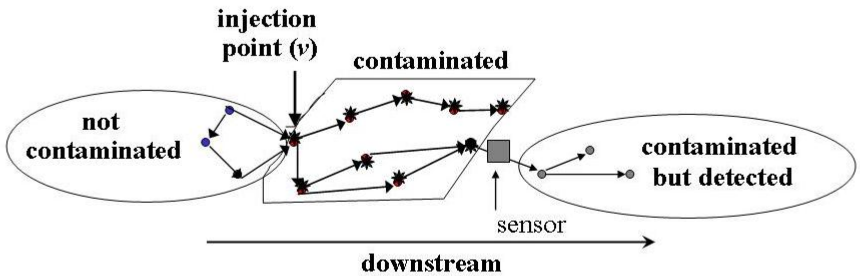


Figure 5.1: Network flow contamination relative to injection point.

of v is exposed if there is a path with no sensor from v to i. The population of i is a parameter in the model.

Extent of Contamination (EC). The contaminant flows through the system, contaminating each placement in its path until it reaches a sensor. The contaminated pipe lengths determine the extent of this contamination.

Volume Consumed (VC). The population continues to drink the water until the contaminant is detected. The demand rates and the time to detection determine how much contaminated water is consumed.

Number of Failed Detections (NF). Since, generally, there will not be enough sensors to completely cover the water network, there will be some number of attacks that will fail to be detected by any sensors. (Measured as a proportion of all attacks.)

Time Until Detection (TD). The time between the attack and detection of the contaminate by a sensor, where the time for water (or a contaminate) to pass from one node to another is a parameter in the model. (Contamination events are presumed to be detected within 24 hours with or without sensors, so any sensor detection times greater than 24 hours are considered undetected by the sensors.)

The primary decision variable is the binary assignment:

$$x_i = \begin{cases} 1 & \text{if a sensor is assigned to placement } i; \\ 0 & \text{otherwise.} \end{cases}$$

These are subject to the budget constraint:

$$\sum_i x_i \leq N,$$

where N is less than the number of placements.

Since there is no advantage to not placing a sensor, it is optimal to assign them all, so the budget constraint always holds with equality at optimality. For the purpose of this discussion we assume the number of sensors, N, is given. Other than the budget constraint, there is no issue of feasibility; we want our decision to be robust in the sense that the objective values are as insensitive as possible to the injection point.

Although the model uses secondary variables to measure each of the objectives, we can simply write each objective as a function of any assignment [32]. Thus, the sensor placement model is given by:

$$\min_{x \in X(N)} \ \Phi\big(\mathrm{PE}(x,v), \mathrm{EC}(x,v), \mathrm{VC}(x,v), \mathrm{NF}(x,v), \mathrm{TD}(x,v)\big),$$

where

$$X(N) = \left\{ x \in \{0,\, 1\}^n \;:\; \sum_i x_i \le N \right\},$$

and Φ is an appropriate function of the objectives, such as a weighted sum. In this form, the model is not yet well posed since the objective value depends upon the uncertain point of injection, v.

One model assumes no information about the injection point and prepares for the worst-case:

$$\min_{x \in X(N)} \max_{v} \; \Phi\big(\mathrm{PE}(x,v), \mathrm{EC}(x,v), \mathrm{VC}(x,v), \mathrm{NF}(x,v), \mathrm{TD}(x,v)\big).$$

However, analysts have some information about the injection point, which takes the form of *injection weights*, $w \ge 0$, where $\sum_v w_v = 1$. One could think of w as a probability and the objective as an "expected case" but it is not necessary to attribute any probability meaning to these weights.

Since, in the general case, the admissible weights include $w_{\bar{v}} = 1$ for some $\bar{v}$ and $w_v = 0$ for $v \ne \bar{v}$, we can formulate the worst-case model, where we have no information, as:

$$\min_{x \in X(N)} \max_{w \in S_n} \sum_v w_v \Phi\big(\mathrm{PE}(x,v), \mathrm{EC}(x,v), \mathrm{VC}(x,v), \mathrm{NF}(x,v), \mathrm{TD}(x,v)\big),$$

where S_n is an n-simplex, $\{w \in \mathbb{R}^n_+ \,:\, \sum_v w_v = 1\}$.

Alternatively, as in Carr et al. [7], the information about the weights may be in the form of bounds, so that w is restricted to

$$W = \{w \in S_n \,:\, \underline{w} \le w \le \overline{w}\}, \tag{5.1}$$

where we assume $\sum_v \underline{w} \le 1 \le \sum_v \overline{w}$. Tighter bounds reflect more confidence in analyst judgment regarding the injection point. In this case, the model is:

$$\min_{x \in X(N)} \max_{w \in W} \sum_v w_v \Phi\big(\mathrm{PE}(x,v), \mathrm{EC}(x,v), \mathrm{VC}(x,v), \mathrm{NF}(x,v), \mathrm{TD}(x,v)\big).$$

In order to simplify our discussion and keep the emphasis on investigating the notion of persistence, we focus on the central-value model, where the injection weights are fixed at their *central* values. The model is then:

$$\min_{x \in X(N)} \sum_v \widehat{w}_v \Phi\big(\mathrm{PE}(x,v), \mathrm{EC}(x,v), \mathrm{VC}(x,v), \mathrm{NF}(x,v), \mathrm{TD}(x,v)\big),$$

where each $\hat{w}_v$ is a fixed injection weight. (The central value could be a statistical mean or mode; we do not give it a precise meaning here.)

In the following, we investigate applying some of the approaches to using persistence as evidence discussed in earlier chapters to the sensor placement problem. We start with a discussion of using the method of successive exclusions to generate a list of "acceptable" solutions to a single objective and consider both the "persistence of solutions" and the "persistence of decisions" approaches. We follow that discussion with an investigation of using these approaches to deal with multiple objectives and finish with a discussion of grouping objectives, where the solution sets are highly or totally conflicting.

Table 5.1: Ten sensor placements for minimizing population exposed.

	Central value of Population Exposed: $\sum_v \hat{w}_v \mathrm{PE}(x,v)$										
Sensor	83.38	83.57	83.58	83.58	83.64	83.68	83.68	83.70	83.74	83.77	$\gamma(v)$
17	0	0	0	1	0	0	0	0	1	1	$^3/_{10}$
50	1	1	1	1	1	1	1	1	1	1	$^{10}/_{10}$
115	0	0	1	0	0	0	0	0	0	1	$^2/_{10}$
123	1	0	0	1	1	1	1	1	0	0	$^6/_{10}$
162	1	1	1	0	1	1	1	1	0	0	$^7/_{10}$
176	0	0	0	0	1	0	0	0	0	0	$^1/_{10}$
199	1	1	1	1	1	1	1	1	1	1	$^{10}/_{10}$
204	0	0	0	0	0	0	1	0	0	0	$^1/_{10}$
226	1	1	1	1	1	1	1	1	1	1	$^{10}/_{10}$
290	1	1	1	1	1	1	0	0	1	1	$^8/_{10}$
307	1	1	1	1	1	1	1	1	1	1	$^{10}/_{10}$
310	1	1	1	1	0	1	1	1	1	1	$^9/_{10}$
315	0	0	0	0	0	0	0	1	0	0	$^1/_{10}$
326	1	1	1	1	1	1	1	1	1	1	$^{10}/_{10}$
385	0	1	0	0	0	0	0	0	1	0	$^2/_{10}$
426	0	0	0	0	0	1	0	0	0	0	$^1/_{10}$
453	1	1	1	1	1	0	1	1	1	1	$^9/_{10}$
$\sum_v \gamma(v)$	$^{89}/_{10}$	$^{85}/_{10}$	$^{85}/_{10}$	$^{85}/_{10}$	$^{81}/_{10}$	$^{81}/_{10}$	$^{82}/_{10}$	$^{82}/_{10}$	$^{81}/_{10}$	$^{81}/_{10}$	

5.1.1 Robustness Using Successive Exclusions

As in the preceding chapters, our first step in investigating the persistence behavior of the sensor placement problem is to generate a list of "acceptable" solutions. For this discussion, we generate a list of optimal and near-optimal solutions using the method of successive exclusions discussed in Chapter 2. We begin by analyzing evidence from solutions for a single objective: population exposed, where

$$\Phi(\mathrm{PE}, \mathrm{EC}, \mathrm{VC}, \mathrm{NF}, \mathrm{TD}) = \mathrm{PE}.$$

Table 5.1 shows persistence values for the solution set formed by applying the method of successive exclusions, starting with the (unique) optimal set of placements, to the problem of minimizing population exposed with 10 sensors. Five of those sensor placements remain in the consensus set, making their global persistence values, γ, $^{10}/_{10}$ each. We also note that the first solution (column 1) maximizes total persistence.

For comparison, we look at solution lists for the problems of minimizing the extent of contamination (Table 5.2) and minimizing total volume consumed (Table 5.3). When the objective is minimizing the extent of contamination, the consensus set contains seven sensor placements: 50, 199, 226, 295, 307, 310, 344, representing $^{7}/_{16}$ of the placement selected by any solution. The total persistence values steadily decrease from $^{89}/_{10}$ to $^{82}/_{10}$ with the maximum total persistence occurring with the first solution. In the case of minimizing volume consumed, we see that consensus set also contains seven sensor placements: 26, 67, 124, 253, 262, 289, and 294, representing half of the placements chosen. Also, we see that the total persistence values of the solutions is relatively flat, with four solutions sharing the maximum, $^{85}/_{10}$, while the remaining six have a total persistence of $^{83}/_{10}$.

Finally, in Tables 5.4 and 5.5 we illustrate the persistence values for the problems of minimizing the number of failed detections and minimizing time until detection. In the case of minimizing failed detections, the consensus set contains seven of the fourteen placements: 67, 145, 249, 253, 289, 308, and 348, while the total persistence begins at $^{85}/_{10}$ for the first four solutions and drops to $^{83}/_{10}$ for the remaining six. For the problem of minimizing time to detection,

Table 5.2: Sensor placements for minimizing extent contaminated.

	Central value of Extent Contaminated: $\sum_v \hat{w}_v \mathrm{EC}(x, v)$										
Sensor	55148	55203	55232	55282	55298	55330	55331	55353	55382	55413	$\gamma(x_i)$
23	0	1	0	0	0	0	0	1	0	0	2/10
45	1	0	0	0	1	1	0	0	0	1	4/10
50	1	1	1	1	1	1	1	1	1	1	10/10
159	1	1	1	1	0	1	1	0	0	1	7/10
162	0	0	0	0	1	0	0	1	1	0	3/10
199	1	1	1	1	1	1	1	1	1	1	10/10
204	0	0	1	0	0	1	0	0	1	0	3/10
226	1	1	1	1	1	1	1	1	1	1	10/10
228	0	0	0	0	0	0	1	0	0	0	1/10
239	0	0	0	1	0	0	0	0	0	0	1/10
290	0	0	0	0	0	0	0	0	0	1	1/10
295	1	1	1	1	1	1	1	1	1	1	10/10
307	1	1	1	1	1	1	1	1	1	1	10/10
310	1	1	1	1	1	1	1	1	1	1	10/10
315	1	1	1	1	1	0	1	1	1	0	8/10
344	1	1	1	1	1	1	1	1	1	1	10/10
$\sum_i \gamma(x_i)$	89/10	87/10	88/10	86/10	85/10	84/10	86/10	83/10	84/10	82/10	

Table 5.3: Sensor placements for minimizing volume consumed.

	Central value of Volume Consumed: $\sum_v \hat{w}_v \mathrm{VC}(x, v)$										
Sensor	928.7	933.2	938.8	943.3	949.6	949.8	954.1	954.3	969.7	971.8	$\gamma(v)$
26	1	1	1	1	1	1	1	1	1	1	10/10
67	1	1	1	1	1	1	1	1	1	1	10/10
74	0	0	1	1	0	1	0	1	0	1	5/10
112	1	0	1	0	1	1	0	0	1	1	6/10
124	1	1	1	1	1	1	1	1	1	1	10/10
180	1	1	0	0	1	0	1	0	1	0	5/10
183	0	0	0	0	1	1	1	1	0	0	4/10
216	1	1	1	1	0	0	0	0	0	0	4/10
241	0	0	0	0	0	0	0	0	1	1	2/10
253	1	1	1	1	1	1	1	1	1	1	10/10
262	1	1	1	1	1	1	1	1	1	1	10/10
289	1	1	1	1	1	1	1	1	1	1	10/10
294	1	1	1	1	1	1	1	1	1	1	10/10
348	0	1	0	1	0	0	1	1	0	0	4/10
$\sum_v \gamma(v)$	85/10	83/10	85/10	83/10	85/10	85/10	83/10	83/10	83/10	83/10	

Table 5.4: Sensor placements for minimizing failed detections.

Central value of Number of Failed Detections: $\sum_v \hat{w}_v \mathrm{NF}(x,v)$											
Sensor	0.1232	0.1232	0.1244	0.1244	0.1244	0.1244	0.1255	0.1255	0.1255	0.1255	$\gamma(v)$
26	0	0	1	1	0	0	1	1	0	0	$4/10$
61	1	1	1	1	0	0	0	0	1	1	$6/10$
67	1	1	1	1	1	1	1	1	1	1	$10/10$
145	1	1	1	1	1	1	1	1	1	1	$10/10$
233	1	1	0	0	1	1	0	0	0	0	$4/10$
249	1	1	1	1	1	1	1	1	1	1	$10/10$
253	1	1	1	1	1	1	1	1	1	1	$10/10$
254	0	1	0	1	0	1	0	1	1	0	$5/10$
289	1	1	1	1	1	1	1	1	1	1	$10/10$
294	0	0	0	0	1	1	1	1	0	0	$4/10$
308	1	1	1	1	1	1	1	1	1	1	$10/10$
348	1	1	1	1	1	1	1	1	1	1	$10/10$
354	1	0	1	0	1	0	1	0	0	1	$5/10$
418	0	0	0	0	0	0	0	0	1	1	$2/10$
$\sum_v \gamma(v)$	$85/10$	$85/10$	$85/10$	$85/10$	$83/10$	$83/10$	$83/10$	$83/10$	$83/10$	$83/10$	

Table 5.5: Sensor placements for minimizing time to detection.

Central value of Time to Detection: $\sum_v \hat{w}_v \mathrm{TD}(x,v)$											
Sensor	953.4	953.6	954.3	954.4	954.5	954.6	954.8	955.3	955.5	955.6	$\gamma(v)$
26	1	1	0	1	0	0	0	0	1	0	$4/10$
45	1	1	1	0	1	1	1	0	1	0	$7/10$
124	1	1	1	1	1	1	1	1	1	1	$10/10$
174	0	1	0	1	1	0	1	1	1	1	$7/10$
228	1	1	1	1	1	1	1	1	1	1	$10/10$
233	0	0	1	0	1	0	0	1	0	0	$3/10$
253	1	1	1	1	1	1	1	1	1	1	$10/10$
262	1	0	1	1	0	1	0	1	1	1	$7/10$
289	1	1	1	1	1	1	1	1	1	1	$10/10$
294	1	1	1	1	1	1	1	1	1	1	$10/10$
346	1	1	1	1	1	1	1	1	0	1	$9/10$
348	1	1	1	1	1	1	1	1	1	1	$10/10$
418	0	0	0	0	0	1	1	0	0	1	$3/10$
$\sum_v \gamma(v)$	$87/10$	$87/10$	$86/10$	$87/10$	$86/10$	$86/10$	$86/10$	$86/10$	$85/10$	$86/10$	

six of the thirteen placements are in the consensus set 124, 228, 253, 289, 294, and 348. The total persistence ranges between $^{87}/_{10}$ (for the first, second, and fourth solutions) and $^{85}/_{10}$ (for the ninth solution).

5.1.1.1 Persistence of Solutions

Following the model from Section 4.1, let X denote the set of x-vectors used to compute the persistence vector, $p = [\,\gamma(x_1) \ \cdots \ \gamma(x_N)\,]^\mathsf{T}$ and the bpa:

$$\mathrm{m}(x) = 0 \qquad \text{for } x \notin X \tag{5.2a}$$

$$\mathrm{m}(x) = \frac{1}{\kappa}\big(p^\mathsf{T}x + (\vec{\mathbf{1}} - p)^\mathsf{T}(\vec{\mathbf{1}} - x)\big) \qquad \text{for } x \in X \tag{5.2b}$$

$$\mathrm{m}(Y) = 0 \qquad \text{for } Y \subseteq X \ : \ |Y| > 1, \tag{5.2c}$$

where $\kappa = \sum_{y \in X} \big(p^\mathsf{T}y + (\vec{\mathbf{1}} - p)^\mathsf{T}(\vec{\mathbf{1}} - y)\big)$. Recall that we call (5.2) the *singleton-only* bpa because (5.2c) implies that we assign positive belief, denoted $\mathrm{Bel}^{\mathrm{so}}$, to only singleton sets (that is, those consisting of only one solution vector) and, moreover, this implies that $\mathrm{Bel}^{\mathrm{so}}(x) = \mathrm{m}(x)$.

Table 5.6: Singleton-only beliefs for minimizing population exposed.

Objective Value	Sensor Placement Selection-Set	$\mathrm{Bel}^{\mathrm{so}}(x)$
83.38	{ 50, 123, 162, 199, 226, 290, 307, 310, 326, 453 }	0.1085
83.57	{ 50, 162, 199, 226, 290, 307, 310, 326, 385, 453 }	0.1026
83.58	{ 50, 115, 162, 199, 226, 290, 307, 310, 326, 453 }	0.1026
83.58	{ 17, 50, 123, 199, 226, 290, 307, 310, 326, 453 }	0.1026
83.64	{ 50, 123, 162, 176, 199, 226, 290, 307, 326, 453 }	0.0968
83.68	{ 50, 123, 162, 199, 226, 290, 307, 310, 326, 426 }	0.0968
83.68	{ 50, 123, 162, 199, 204, 226, 307, 310, 326, 453 }	0.0982
83.70	{ 50, 123, 162, 199, 226, 307, 310, 315, 326, 453 }	0.0982
83.74	{ 17, 50, 199, 226, 290, 307, 310, 326, 385, 453 }	0.0968
83.77	{ 17, 50, 115, 199, 226, 290, 307, 310, 326, 453 }	0.0968

Table 5.7: Singleton-only beliefs for minimizing extent contaminated.

Objective Value	Sensor Placement Selection-Set	$\text{Bel}^{\text{so}}(x)$
55148	{ 45, 50, 159, 199, 226, 295, 307, 310, 315, 344 }	0.1055
55203	{ 23, 50, 159, 199, 226, 295, 307, 310, 315, 344 }	0.1024
55232	{ 50, 159, 199, 204, 226, 295, 307, 310, 315, 344 }	0.1040
55282	{ 50, 159, 199, 226, 239, 295, 307, 310, 315, 344 }	0.1009
55298	{ 45, 50, 162, 199, 226, 295, 307, 310, 315, 344 }	0.0994
55330	{ 45, 50, 159, 199, 204, 226, 295, 307, 310, 344 }	0.0979
55331	{ 50, 159, 199, 226, 228, 295, 307, 310, 315, 344 }	0.1009
55352	{ 23, 50, 162, 199, 226, 295, 307, 310, 315, 344 }	0.0963
55382	{ 50, 162, 199, 204, 226, 295, 307, 310, 315, 344 }	0.0979
55413	{ 45, 50, 159, 199, 226, 290, 295, 307, 310, 344 }	0.0948

Table 5.8: Singleton-only beliefs for minimizing volume consumed.

Objective Value	Sensor Placement Selection-Set	$\text{Bel}^{\text{so}}(x)$
928.7	{ 26, 67, 112, 124, 180, 216, 253, 262, 289, 294 }	0.1022
933.2	{ 26, 67, 124, 180, 216, 253, 262, 289, 294, 348 }	0.0985
938.8	{ 26, 67, 74, 112, 124, 216, 253, 262, 289, 294 }	0.1022
943.3	{ 26, 67, 74, 124, 216, 253, 262, 289, 294, 348 }	0.0985
949.6	{ 26, 67, 112, 124, 180, 183, 253, 262, 289, 294 }	0.1022
949.8	{ 26, 67, 74, 112, 124, 183, 253, 262, 289, 294 }	0.1022
954.1	{ 26, 67, 124, 180, 183, 253, 262, 289, 294, 348 }	0.0985
954.3	{ 26, 67, 74, 124, 183, 253, 262, 289, 294, 348 }	0.0985
969.7	{ 26, 67, 112, 124, 180, 241, 253, 262, 289, 294 }	0.0985
971.8	{ 26, 67, 74, 112, 124, 241, 253, 262, 289, 294 }	0.0985

In Tables 5.6 - 5.10, we show the resulting *singleton-only* belief values for each of the five objectives, where in each case we have set the solution domain, X, to be exactly the solution list, X^K, for that objective. As long as all solution exclusions leave at least N placements unassigned, all solutions satisfy the constant-sum condition (p. 63). Hence, these sets have no dominated solutions.

In both the cases of minimizing population exposed and of minimizing the

Table 5.9: Singleton-only beliefs for minimizing failed detections.

Objective Value	Sensor Placement Selection-Set	$\text{Bel}^{\text{so}}(x)$
0.1232	{ 61, 67, 145, 233, 249, 253, 289, 308, 348, 354 }	0.1022
0.1232	{ 61, 67, 145, 233, 249, 253, 254, 289, 308, 348 }	0.1022
0.1244	{ 26, 61, 67, 145, 249, 253, 289, 308, 348, 354 }	0.1022
0.1244	{ 26, 61, 67, 145, 249, 253, 254, 289, 308, 348 }	0.1022
0.1244	{ 67, 145, 233, 249, 253, 289, 294, 308, 348, 354 }	0.0985
0.1244	{ 67, 145, 233, 249, 253, 254, 289, 294, 308, 348 }	0.0985
0.1255	{ 26, 67, 145, 249, 253, 289, 294, 308, 348, 354 }	0.0985
0.1255	{ 26, 67, 145, 249, 253, 254, 289, 294, 308, 348 }	0.0985
0.1255	{ 61, 67, 145, 249, 253, 254, 289, 308, 348, 418 }	0.0985
0.1255	{ 61, 67, 145, 249, 253, 289, 308, 348, 354, 418 }	0.0985

Table 5.10: Singleton-only beliefs for minimizing time to detection.

Objective Value	Sensor Placement Selection-Set	$\text{Bel}^{\text{so}}(x)$
953.4	{ 26, 45, 124, 228, 253, 262, 289, 294, 346, 348 }	0.1016
953.6	{ 26, 45, 124, 174, 228, 253, 289, 294, 346, 348 }	0.1016
954.3	{ 45, 124, 228, 233, 253, 262, 289, 294, 346, 348 }	0.0996
954.4	{ 26, 124, 174, 228, 253, 262, 289, 294, 346, 348 }	0.1016
954.5	{ 45, 124, 174, 228, 233, 253, 289, 294, 346, 348 }	0.0996
954.6	{ 45, 124, 228, 253, 262, 289, 294, 346, 348, 418 }	0.0996
954.8	{ 45, 124, 174, 228, 253, 289, 294, 346, 348, 418 }	0.0996
955.3	{ 124, 174, 228, 233, 253, 262, 289, 294, 346, 348 }	0.0996
955.5	{ 26, 45, 124, 174, 228, 253, 262, 289, 294, 348 }	0.0977
955.6	{ 124, 174, 228, 253, 262, 289, 294, 346, 348, 418 }	0.0996

extent of contamination, the solution with the greatest total persistence, the *mode* of the bpa, is uniquely the first solution on the respective solution list. By contrast, the bpas for each of the remaining objectives (minimize volume consumed, minimize failed detections, and minimize time until detection) have multiple modes, though, for all three objectives, the first solution is one of the modes. This strongly suggests the structure of the example sensor placement problem is such that the optimal solution is also a *robust* solution, which should increase our confidence in choosing it.

Table 5.11: Example singleton-only conditional beliefs.

(a) Population exposed

Bel(x)	Bel($x\|X_{162}$)	Bel($x\|X_{204}$)	Bel($x\|X_{226}$)	Bel($x\|X_{315}$)	Bel($x\|X_{162} \cap X_{204}$)
0.1085	0.1542	0.0000	0.1085	0.0000	0.1353
0.1026	0.1458	0.0000	0.1026	0.0000	0.1280
0.1026	0.1458	0.0000	0.1026	0.0000	0.1280
0.1026	0.0000	0.0000	0.1026	0.0000	0.0000
0.0968	0.1375	0.0000	0.0968	0.0000	0.1207
0.0968	0.1375	0.0000	0.0968	0.0000	0.1207
0.0982	0.1396	1.0000	0.0982	0.0000	0.2450
0.0982	0.1396	0.0000	0.0982	1.0000	0.1225
0.0968	0.0000	0.0000	0.0968	0.0000	0.0000
0.0968	0.0000	0.0000	0.0968	0.0000	0.0000

(b) Extent of contamination

Bel(x)	Bel($x\|X_{162}$)	Bel($x\|X_{204}$)	Bel($x\|X_{226}$)	Bel($x\|X_{315}$)	Bel($x\|X_{162} \cap X_{204}$)
0.1055	0.0000	0.0000	0.1055	0.1307	0.0000
0.1024	0.0000	0.0000	0.1024	0.1269	0.0000
0.1040	0.0000	0.3469	0.1040	0.1288	0.1753
0.1009	0.0000	0.0000	0.1009	0.1250	0.0000
0.0994	0.3385	0.0000	0.0994	0.1231	0.1675
0.0979	0.0000	0.3265	0.0979	0.0000	0.1649
0.1009	0.0000	0.0000	0.1009	0.1250	0.0000
0.0963	0.3281	0.0000	0.0963	0.1193	0.1624
0.0979	0.3333	0.3265	0.0979	0.1212	0.3299
0.0948	0.0000	0.0000	0.0948	0.0000	0.0000

In Table 5.11, we look at a sample of conditional beliefs on the singleton-only belief function. For purposes of comparison we examine the problems of minimizing population exposed and minimizing extent of contamination and look at conditioning on a few placements common to both. Recall that, in Section 4.1.1, we first defined the sets $X_j = \{x \in X : x_j = 1\}$ and $\overline{X}_j = \{x \in X : x_j = 0\}$. Then, by Theorem 4.3, the belief for a solution x given we know

$x_j = 1$ is

$$\mathrm{Bel}(x \mid X_j) = \begin{cases} \dfrac{\mathrm{Bel}(x)}{\mathrm{Bel}\,(X_j)} & \text{if } x \in X_j \\ 0 & \text{if } x \in \overline{X}_j. \end{cases}$$

Looking at Table 5.11a, we see that when we make an a priori decision to select sensor placement 204, all belief moves to the seventh solution; and, similarly, when we choose sensor placement 315, all belief is allocated to the eighth solution, since these solutions uniquely satisfy the respective preconditions. By contrast, since all the solutions from both the population-exposed list and the extent-of-contamination list set $x_{226} = 1$ (that is, $\gamma(x_{226}) = {}^{10}\!/_{10}$), this condition has no effect on our beliefs and, hence, the fourth columns in both Table 5.11a and Table 5.11b are identical to the unconditioned first columns. The final column contains the beliefs resulting from selecting placements 162 and 204 simultaneously, and allocates belief so that only those solutions containing both have non-zero belief values.

5.1.1.2 Persistence of Decisions

Alternatively, we can follow the *decision-sets* belief function model developed in Chapter 4.2. Recall that this model is constructed by first defining a bpa, m_k, for each solution, S^k, in the solution list such that each nonempty subset of S^k is equally likely:

$$\mathrm{m}_k(S) = \begin{cases} \frac{1}{2^{|S^k|-1}} & \text{if } S \subseteq S^k, S \neq \emptyset \\ 0 & \text{otherwise,} \end{cases}$$

and then defining a final bpa as a weighted sum of these K bpas:

$$\mathrm{m}(S) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{m}_k(S) = \frac{1}{K}\sum_{k\,:\,S\subseteq S^k} \frac{1}{2^{|S^k|}-1} = \widehat{\Gamma}(S).$$

Thus, the belief value for a solution x is given by

$$\begin{aligned}\mathrm{Bel}(x) &= \frac{1}{K}\sum_{S'\subseteq S(x)}\sum_{k=1}^{K} \mathrm{m}_k(S') \\ &= \frac{1}{K}\sum_{\emptyset\neq S'\subseteq S(x)}\sum_{k\,:\,S'\subseteq S^k} \frac{1}{2^{|S^k|}-1} = \sum_{S'\subseteq S(x)} \widehat{\Gamma}(S'),\end{aligned}$$

and, hence, can be computed as the sum of adjusted global selection persistence factors over subsets of the given solution's selection set.

Moreover, unlike the knapsack problem, the sensor placement budget constraint always allows exactly N placements (unless all N-placement solutions are excluded), so Theorem 4.5 applies. Thus the problem of finding the selection with maximum *feasible* belief is equivalent to global selection persistence factor maximization. Mathematically,

$$\operatorname*{argmax}_{x\in X(N)} \mathrm{Bel}(x) \;=\; \operatorname*{argmax}_{x\in X(N)} \sum_{S'\subseteq S(x)} \widehat{\Gamma}(S') \;=\; \operatorname*{argmax}_{x\in X(N)} \sum_{S'\subseteq S(x)} \Gamma(S')$$

where

$$X(N) = \left\{ x \in \{0,\,1\}^n \;:\; \sum_i x_i \leq N \right\}.$$

We illustrate the decision-sets belief function values for each of the five objectives in Tables 5.12–5.16. In all cases, except that of minimizing population exposed (Table 5.12) where the first solution has the unique maximum belief of

Table 5.12: Decisions-sets beliefs for minimizing population exposed.

Objective Value	Sensor Placement Selection-Set	$\mathrm{Bel}^{\mathrm{ds}}(x)$	$\mathrm{Plaus}^{\mathrm{ds}}(x)$
83.38	{ 50, 123, 162, 199, 226, 290, 307, 310, 326, 453 }	0.4995	0.9987
83.57	{ 50, 162, 199, 226, 290, 307, 310, 326, 385, 453 }	0.3994	0.9979
83.58	{ 50, 115, 162, 199, 226, 290, 307, 310, 326, 453 }	0.3994	0.9979
83.58	{ 17, 50, 123, 199, 226, 290, 307, 310, 326, 453 }	0.3994	0.9979
83.64	{ 50, 123, 162, 176, 199, 226, 290, 307, 326, 453 }	0.3243	0.9968
83.68	{ 50, 123, 162, 199, 226, 290, 307, 310, 326, 426 }	0.3243	0.9968
83.68	{ 50, 123, 162, 199, 204, 226, 307, 310, 326, 453 }	0.3494	0.9970
83.70	{ 50, 123, 162, 199, 226, 307, 310, 315, 326, 453 }	0.3494	0.9970
83.74	{ 17, 50, 199, 226, 290, 307, 310, 326, 385, 453 }	0.3494	0.9964
83.77	{ 17, 50, 115, 199, 226, 290, 307, 310, 326, 453 }	0.3494	0.9964

Table 5.13: Decisions-sets beliefs for minimizing extent contaminated.

Objective Value	Sensor Placement Selection-Set	$\mathrm{Bel}^{\mathrm{ds}}(x)$	$\mathrm{Plaus}^{\mathrm{ds}}(x)$
55148	{ 45, 50, 159, 199, 226, 295, 307, 310, 315, 344 }	0.4995	0.9987
55203	{ 23, 50, 159, 199, 226, 295, 307, 310, 315, 344 }	0.4495	0.9983
55232	{ 50, 159, 199, 204, 226, 295, 307, 310, 315, 344 }	0.4745	0.9985
55282	{ 50, 159, 199, 226, 239, 295, 307, 310, 315, 344 }	0.4244	0.9981
55298	{ 45, 50, 162, 199, 226, 295, 307, 310, 315, 344 }	0.3994	0.9979
55330	{ 45, 50, 159, 199, 204, 226, 295, 307, 310, 344 }	0.3869	0.9976
55331	{ 50, 159, 199, 226, 228, 295, 307, 310, 315, 344 }	0.4244	0.9981
55352	{ 23, 50, 162, 199, 226, 295, 307, 310, 315, 344 }	0.3744	0.9972
55382	{ 50, 162, 199, 204, 226, 295, 307, 310, 315, 344 }	0.3869	0.9976
55413	{ 45, 50, 159, 199, 226, 290, 295, 307, 310, 344 }	0.3494	0.9970

Table 5.14: Decisions-sets beliefs for minimizing volume consumed.

Objective Value	Sensor Placement Selection-Set	$\mathrm{Bel}^{\mathrm{ds}}(x)$	$\mathrm{Plaus}^{\mathrm{ds}}(x)$
928.7	{ 26, 67, 112, 124, 180, 216, 253, 262, 289, 294 }	0.4119	0.9978
933.2	{ 26, 67, 124, 180, 216, 253, 262, 289, 294, 348 }	0.3744	0.9972
938.8	{ 26, 67, 74, 112, 124, 216, 253, 262, 289, 294 }	0.4119	0.9978
943.3	{ 26, 67, 74, 124, 216, 253, 262, 289, 294, 348 }	0.3744	0.9972
949.6	{ 26, 67, 112, 124, 180, 183, 253, 262, 289, 294 }	0.4119	0.9978
949.8	{ 26, 67, 74, 112, 124, 183, 253, 262, 289, 294 }	0.4119	0.9978
954.1	{ 26, 67, 124, 180, 183, 253, 262, 289, 294, 348 }	0.3744	0.9972
954.3	{ 26, 67, 74, 124, 183, 253, 262, 289, 294, 348 }	0.3744	0.9972
969.7	{ 26, 67, 112, 124, 180, 241, 253, 262, 289, 294 }	0.3744	0.9972
971.8	{ 26, 67, 74, 112, 124, 241, 253, 262, 289, 294 }	0.3744	0.9972

Table 5.15: Decisions-sets beliefs for minimizing failed detections.

Objective Value	Sensor Placement Selection-Set	$\mathrm{Bel}^{\mathrm{ds}}(x)$	$\mathrm{Plaus}^{\mathrm{ds}}(x)$
0.1232	{ 61, 67, 145, 233, 249, 253, 289, 308, 348, 354 }	0.4119	0.9978
0.1232	{ 61, 67, 145, 233, 249, 253, 254, 289, 308, 348 }	0.4119	0.9978
0.1244	{ 26, 61, 67, 145, 249, 253, 289, 308, 348, 354 }	0.4119	0.9978
0.1244	{ 26, 61, 67, 145, 249, 253, 254, 289, 308, 348 }	0.4119	0.9978
0.1244	{ 67, 145, 233, 249, 253, 289, 294, 308, 348, 354 }	0.3744	0.9972
0.1244	{ 67, 145, 233, 249, 253, 254, 289, 294, 308, 348 }	0.3744	0.9972
0.1255	{ 26, 67, 145, 249, 253, 289, 294, 308, 348, 354 }	0.3744	0.9972
0.1255	{ 26, 67, 145, 249, 253, 254, 289, 294, 308, 348 }	0.3744	0.9972
0.1255	{ 61, 67, 145, 249, 253, 254, 289, 308, 348, 418 }	0.3744	0.9972
0.1255	{ 61, 67, 145, 249, 253, 289, 308, 348, 354, 418 }	0.3744	0.9972

Table 5.16: Decisions-sets beliefs for minimizing time to detection.

Objective Value	Sensor Placement Selection-Set	$\text{Bel}^{\text{ds}}(x)$	$\text{Plaus}^{\text{ds}}(x)$
953.4	{ 26, 45, 124, 228, 253, 262, 289, 294, 346, 348 }	0.4495	0.9983
953.6	{ 26, 45, 124, 174, 228, 253, 289, 294, 346, 348 }	0.4495	0.9983
954.3	{ 45, 124, 228, 233, 253, 262, 289, 294, 346, 348 }	0.4244	0.9981
954.4	{ 26, 124, 174, 228, 253, 262, 289, 294, 346, 348 }	0.4495	0.9983
954.5	{ 45, 124, 174, 228, 233, 253, 289, 294, 346, 348 }	0.4244	0.9981
954.6	{ 45, 124, 228, 253, 262, 289, 294, 346, 348, 418 }	0.4244	0.9981
954.8	{ 45, 124, 174, 228, 253, 289, 294, 346, 348, 418 }	0.4244	0.9981
955.3	{ 124, 174, 228, 233, 253, 262, 289, 294, 346, 348 }	0.4244	0.9981
955.5	{ 26, 45, 124, 174, 228, 253, 262, 289, 294, 348 }	0.3994	0.9979
955.6	{ 124, 174, 228, 253, 262, 289, 294, 346, 348, 418 }	0.4244	0.9981

0.4995, the maximum belief value is attained by multiple solutions. Hence, as with the singleton-only bpa, the first solution in each of the solution lists is one of those whose belief is the maximum for the list. More generally, we see that in all five cases there are only a few distinct belief values, shared by multiple solutions.

Comparing these results with the total persistence results (Tables 5.1–5.5), we see that, for the sensor placement problem, both models produce similar groupings and relative orderings. For example, in the case of population exposed, the first solution is uniquely maximal for both models, and, in the case of number of failed detections, both models produce exactly the same grouping of solutions (giving the maximal score to the first four solution while the rest are assigned a second, lower score). This similarity between the models is unsurprising when we consider that, for all five objectives, the consensus set contained roughly half of the selections with non-zero persistence and represented at least half of the selections made by any solution (since $N = 10$ and, in each case, the size of of the consensus set was at least five).

We end this section by investigating a small sample of conditional beliefs. In Table 5.17, we examine the problems of minimizing volume consumed, minimizing the number of failed detections, and minimizing the time until detection and consider the results of conditioning on some placements common to all three solution lists. By Theorem 4.6 (p 64), the conditional belief for the decisions-sets belief function is

$$\mathrm{Bel}(x \,\big|\, X_j) = \frac{\sum\limits_{y\,:\,S(y)\subseteq S(x)\wedge y_j=1} \mathrm{m}\left(S(y)\right)}{\sum\limits_{y\,:\,y_j=1} \mathrm{m}\left(S(y)\right)}.$$

where $\mathrm{Bel}(x \,\big|\, X_j)$ is the belief in x given we know $x_j = 1$.

The first thing to note is that, unlike the singleton-only belief function, with the decision-set belief function conditioning on a placement from the consensus set does not leave the beliefs unchanged, though the increase in belief is small. For example, in the problem of minimizing population exposed (Table 5.17a), sensor placement 26 is in the consensus set, but making an a priori decision to select it (column 2) results in a small increase in the belief values compared to the unconditioned values in the first column. However, conditioning simultaneously on both a selection from the consensus set and one that appears in only some of the solutions produces results identical to conditioning on the selection that appears in only a few solutions (for example, consider the second and final columns of Table 5.17b. Finally, we note that, as we should expect from any method of conditioning beliefs, only those solutions that satisfy the condition have non-zero beliefs (see, for example, the second column of Table 5.17c).

Table 5.17: Example decisions-sets conditional beliefs.

(a) Volume consumed

Bel(x)	Bel($x \mid X_{26}$)	Bel($x \mid X_{253}$)	Bel($x \mid X_{289}$)	Bel($x \mid X_{294}$)	Bel($x \mid X_{348}$)	Bel($x \mid X_{26} \cap X_{348}$)
0.4119	0.4125	0.4125	0.4125	0.4125	0.0000	0.0000
0.3744	0.3750	0.3750	0.3750	0.3750	0.5625	0.5625
0.4119	0.4125	0.4125	0.4125	0.4125	0.0000	0.0000
0.3744	0.3750	0.3750	0.3750	0.3750	0.5625	0.5625
0.4119	0.4125	0.4125	0.4125	0.4125	0.0000	0.0000
0.4119	0.4125	0.4125	0.4125	0.4125	0.0000	0.0000
0.3744	0.3750	0.3750	0.3750	0.3750	0.5625	0.5625
0.3744	0.3750	0.3750	0.3750	0.3750	0.5625	0.5625
0.3744	0.3750	0.3750	0.3750	0.3750	0.0000	0.0000
0.3744	0.3750	0.3750	0.3750	0.3750	0.0000	0.0000

(b) Number of failed detections

Bel(x)	Bel($x \mid X_{26}$)	Bel($x \mid X_{253}$)	Bel($x \mid X_{289}$)	Bel($x \mid X_{294}$)	Bel($x \mid X_{348}$)	Bel($x \mid X_{26} \cap X_{348}$)
0.4119	0.0000	0.4125	0.4125	0.0000	0.4125	0.0000
0.4119	0.0000	0.4125	0.4125	0.0000	0.4125	0.0000
0.4119	0.5625	0.4125	0.4125	0.0000	0.4125	0.5625
0.4119	0.5625	0.4125	0.4125	0.0000	0.4125	0.5625
0.3744	0.0000	0.3750	0.3750	0.5625	0.3750	0.0000
0.3744	0.0000	0.3750	0.3750	0.5625	0.3750	0.0000
0.3744	0.5625	0.3750	0.3750	0.5625	0.3750	0.5625
0.3744	0.5625	0.3750	0.3750	0.5625	0.3750	0.5625
0.3744	0.0000	0.3750	0.3750	0.0000	0.3750	0.0000
0.3744	0.0000	0.3750	0.3750	0.0000	0.3750	0.0000

(c) Time until detection

Bel(x)	Bel($x \mid X_{26}$)	Bel($x \mid X_{253}$)	Bel($x \mid X_{289}$)	Bel($x \mid X_{294}$)	Bel($x \mid X_{348}$)	Bel($x \mid X_{26} \cap X_{348}$)
0.4495	0.6250	0.4500	0.4500	0.4500	0.4500	0.6250
0.4495	0.6250	0.4500	0.4500	0.4500	0.4500	0.6250
0.4244	0.0000	0.4250	0.4250	0.4250	0.4250	0.0000
0.4495	0.6250	0.4500	0.4500	0.4500	0.4500	0.6250
0.4244	0.0000	0.4250	0.4250	0.4250	0.4250	0.0000
0.4244	0.0000	0.4250	0.4250	0.4250	0.4250	0.0000
0.4244	0.0000	0.4250	0.4250	0.4250	0.4250	0.0000
0.4244	0.0000	0.4250	0.4250	0.4250	0.4250	0.0000
0.3994	0.6250	0.4000	0.4000	0.4000	0.4000	0.6250
0.4244	0.0000	0.4250	0.4250	0.4250	0.4250	0.0000

5.1.2 Multiple Objectives

Ideally we would like to find a solution that optimizes all five objectives simultaneously. In general, though, these objectives are partially conflicting and no such solution exists — instead we seek a solution that performs reasonably well with respect to at least some of the criteria. In this section, we investigate the application of the approaches from Section 4.3 to the sensor placement problem. First, we look at combining evidence from pairs of objectives using both Dempster's rule and the modified discount & combine method on their decision-sets belief functions. Following that, we investigate the application of the grouping approach of Section 4.3.2 to combine the persistence evidence from all five objectives simultaneously.

5.1.2.1 Combining Decision-Sets Belief Functions

As discussed in Section 4.3.1, one approach to dealing with multiple objectives is to treat each objective's solution set as a distinct pool of evidence, construct a belief function for each objective, and combining them using an approach such as Dempster's rule or the modified discount & combine. For Dempster's rule the results of combining objectives whose selection sets, $S(X^{K_1})$ and $S(X^{K_2})$, are largely disjoint will be of dubious quality and the modified discount & combine method is preferable. However, for those objectives with selection sets that substantially overlap either approach will give reasonable results.

In Tables 5.18–5.21, we show the results of using Dempster's rule and the modified discount & combine method as given in (3.13) to combine the decision-sets belief function for those pairs of objectives whose selection sets have sub-

Table 5.18: Combined belief values for PE and EC.

(a) "Best" PE solutions

PE	EC	D-S	D&C
83.38	69580	0.9647	0.4658†
83.57	70862	0.9647	0.4298
83.58	69894	0.9647	0.4298
83.58	69745	0.9005	0.4084
83.64	69864	0.6846	0.3163
83.68	69617	0.9647	0.4028
83.68	69498	0.9486	0.4107
83.70	69316	0.9647	0.4230
83.74	71018	0.9005	0.3904
83.77	70055	0.9005	0.3904

(b) "Best" EC solutions

EC	PE	D-S	D&C
55148	107.00	0.8981	0.4432
55202	104.69	0.8981	0.4252
55232	105.02	0.9069	0.4378
55282	104.83	0.8981	0.4162
55297	106.67	0.9647	0.4343
55330	107.49	0.8836	0.3986
55330	107.00	0.8981	0.4162
55352	104.36	0.9647	0.4253
55381	104.69	0.9743	0.4347
55412	106.65	0.9005	0.3971

stantive intersections. Table 5.18 shows the results from combining population exposed and extent of contamination, Table 5.19 combines volume consumed and number of failed detections, Table 5.20 combines the problems of minimizing volume consumed and minimizing time until detection, and Table 5.21 combines minimizing failed detections with minimizing time until detection. In each case, the results are given for the solution lists from each of the paired objectives along with the objective values for both objectives. It is interesting to note that, using Dempster's rule the resulting combined belief score for many of the solutions is 1.0. This results from the fact that Dempster's rule only assigns non-zero bpa values to those subsets contained in the intersection of the two selection sets. Hence, after normalization, any solution that contains the entire intersection as a subset is assigned the full belief.

Table 5.19: Combined belief values for VC and NF.

(a) "Best" VC solutions

VC	NF	D-S	D&C
928.7	1.0000	0.6070	0.3052
933.2	1.0000	1.0000	0.3885
938.8	1.0000	0.6070	0.3052
943.3	1.0000	1.0000	0.3885
949.6	1.0000	0.6070	0.3052
949.8	1.0000	0.6070	0.3052
954.1	1.0000	1.0000	0.3885
954.3	1.0000	1.0000	0.3885
969.7	1.0000	0.6070	0.2908
971.8	1.0000	0.6070	0.2908

(b) "Best" NF solutions

NF	VC	D-S	D&C
0.1232	58113.7	0.7839	0.3446
0.1232	58113.7	0.7839	0.3446
0.1244	58113.7	0.8416	0.3623
0.1244	58113.7	0.8416	0.3623
0.1244	58113.7	0.9342	0.3695
0.1244	58113.7	0.9342	0.3695
0.1255	58113.7	1.0000	0.3933
0.1255	58113.7	1.0000	0.3933
0.1255	58113.7	0.7839	0.3303
0.1255	58113.7	0.7839	0.3303

Table 5.20: Combined belief values for VC and TD.

(a) "Best" VC solutions

VC	TD	D-S	D&C
928.7	1440.0	0.6070	0.3135
933.2	1440.0	1.0000	0.4052
938.8	1440.0	0.6070	0.3135
943.3	1440.0	1.0000	0.4052
949.6	1440.0	0.6070	0.3135
949.8	1440.0	0.6070	0.3135
954.1	1440.0	1.0000	0.4052
954.3	1440.0	1.0000	0.4052
969.7	1440.0	0.6070	0.2992
971.8	1440.0	0.6070	0.2992

(b) "Best" TD solutions

TD	VC	D-S	D&C
953.4	58113.7	1.0000	0.4388
953.6	58113.7	1.0000	0.4220
954.3	58113.7	0.9342	0.3971
954.4	58113.7	1.0000	0.4388
954.5	58113.7	0.9342	0.3887
954.6	58113.7	0.9342	0.3971
954.8	58113.7	0.9342	0.3887
955.3	58113.7	0.9342	0.3971
955.5	58113.7	1.0000	0.4196
955.6	58113.7	0.9342	0.3971

Table 5.21: Combined belief values for NF and TD.

(a) "Best" NF solutions

NF	TD	D-S	D&C
0.1232	1440.0	0.7701	0.3484
0.1232	1440.0	0.7701	0.3484
0.1244	1440.0	0.7835	0.3520
0.1244	1440.0	0.7835	0.3520
0.1244	1440.0	0.9157	0.3740
0.1244	1440.0	0.9157	0.3740
0.1255	1440.0	0.9310	0.3783
0.1255	1440.0	0.9310	0.3783
0.1255	1440.0	0.7529	0.3300
0.1255	1440.0	0.7529	0.3300

(b) "Best" TD solutions

TD	NF	D-S	D&C
953.4	1.0000	0.9310	0.4043
953.6	1.0000	0.9310	0.4043
954.3	1.0000	0.9157	0.3910
954.4	1.0000	0.9310	0.4043
954.5	1.0000	0.9157	0.3910
954.6	1.0000	0.8927	0.3842
954.8	1.0000	0.8927	0.3842
955.3	1.0000	0.9157	0.3910
955.5	1.0000	0.9310	0.3854
955.6	1.0000	0.8927	0.3842

5.1.2.2 Grouping Objectives

We begin our discussion of building persistence from multiple objectives by grouping with the development of some necessary extensions to the notation introduced in Chapter 2. First, let X^{K_P}, X^{K_C}, X^{K_V}, X^{K_N}, and X^{K_T} be the solution sets for the problems of minimizing population exposed, extent of contamination, volume consumed, failed detections, and time until detection, respectively, where:

$$X^{K_P} = \{x^{1_P}, \ldots, x^{10_P}\}, \qquad X^{K_C} = \{x^{1_C}, \ldots, x^{10_C}\},$$

$$X^{K_V} = \{x^{1_V}, \ldots, x^{10_V}\}, \qquad X^{K_N} = \{x^{1_N}, \ldots, x^{10_N}\},$$

$$X^{K_T} = \{x^{1_T}, \ldots, x^{10_T}\}.$$

Then, we define the selection multisets:

$$\mathcal{S}^P = \biguplus_{i=1}^{10} S(x^{i_P}), \qquad \mathcal{S}^C = \biguplus_{i=1}^{10} S(x^{i_C}), \qquad \mathcal{S}^V = \biguplus_{i=1}^{10} S(x^{i_V}),$$
$$\mathcal{S}^N = \biguplus_{i=1}^{10} S(x^{i_N}), \qquad \mathcal{S}^T = \biguplus_{i=1}^{10} S(x^{i_T}).$$

We can then write the global persistence of selection s for a given objective A in terms of these multisets. For instance,

$$\gamma^A(s) = \frac{\mathcal{N}(s, \mathcal{S}^A)}{K}$$

Moreover, we can further define global persistence of selection s for unions and intersections of two objectives, A and B, as

$$\gamma^{A\cup B}(s) = \frac{\mathcal{N}(s, \mathcal{S}^A \cup \mathcal{S}^B)}{K} \quad \text{and} \quad \gamma^{A\cap B}(s) = \frac{\mathcal{N}(s, \mathcal{S}^A \cap \mathcal{S}^B)}{K},$$

equivalently, $\gamma^{A\cup B}(s) = \max\{\gamma^A(s), \gamma^B(s)\}$ and $\gamma^{A\cap B}(s) = \min\{\gamma^A(s), \gamma^B(s)\}$.

Following the grouping approach of Section 4.3.2, we must first allocate the objectives to groups satisfying the properties: assignment, inter-group conflict, and intra-group compatibility. Examining the selection multisets from each solution list, we find that the only groups that satisfy the intra-group compatibility requirement are: (P, C), (C, T), (V, N), (V, T), (N, T), and (V, N, T). The group assignment requiring the smallest number of groups allocates the solution sets into two groups:

$$G_1 = \{X^{K_P}, X^{K_C}\} \quad \text{and} \quad G_2 = \{X^{K_V}, X^{K_N}, X^{K_T}\}.$$

Next, we find the combined global persistence values by first computing the intra-group values,

$$\gamma^{G_1}(s) = \gamma^{P \cap C}(s) = \min\{\gamma^P(s), \gamma^C(x)\},$$

and

$$\gamma^{G_2}(s) = \gamma^{V \cap N \cap T}(s) = \min\{\gamma^V(s), \gamma^N(x), \gamma^T(x)\}$$

and then computing the final inter-group values,

$$\gamma^{\text{All}}(s) = \gamma^{G_1 \cup G_2}(s) = \max\{\gamma^{G_1}(s), \gamma^{G_2}(x)\}.$$

This results in the following values for selections and rejections:

s	50	162	199	204	226	290	307	310	315	26	253	289	294	348	other
$\gamma^{\text{All}}(s)$	$10/10$	$3/10$	$10/10$	$1/10$	$10/10$	$1/10$	$10/10$	$9/10$	$1/10$	$4/10$	$10/10$	$10/10$	$4/10$	$4/10$	0
$\widetilde{\gamma}^{\text{All}}(s)$	0	$7/10$	0	$9/10$	0	$9/10$	0	$1/10$	$9/10$	$6/10$	0	0	$6/10$	$6/10$	$10/10$

We see that placing sensors 50, 199, 226, 307, 253, and 289 is a very robust decision, and not placing sensors on all but the 14 listed above is equally robust. We are thus left with eight potential placements for the remaining four of the 10 sensors available. As decision support, we have greatly reduced the uncertainty and can not only choose the four with greatest persistence, we also show some sensitivity the associated impacts on all objectives.

From this perspective, our decision support recommendations are then:

1. place sensors on $S^* = \{s : \gamma(s) = 1\}$ (note that $|S^*| < N$);

Table 5.22: Robust sensor placements based on 10 sensors.

N	Sensor Placement Selection-Set	$\sum_v \hat{w}_v \Phi(x,v)$ PE	EC	VC	NF	TD
6	{ 50, 199, 226, 253, 289, 307 }	165.50	108975	16249.2	1.0000	1207.3
7	{ 50, 199, 226, 253, 289, 307, 310 }	130.10	90867	14810.7	1.0000	1194.9
8	{ 50, 199, 226, 253, 289, 307, 310, 26 }	128.92	90867	6739.5	1.0000	1132.3
9	{ 50, 199, 226, 253, 289, 307, 310, 26, 294 }	127.60	90666	5027.8	1.0000	1103.7
10	{ 50, 199, 226, 253, 289, 307, 310, 26, 294, 348 }	127.20	90666	4657.2	1.0000	1075.9
11	{ 50, 199, 226, 253, 289, 307, 310, 26, 294, 348, 162 }	107.91	79946	4581.3	1.0000	1073.6
12	{ 50, 199, 226, 253, 289, 307, 310, 26, 294, 348, 162, 204 }	101.45	73570	4581.3	1.0000	1068.0
13	{ 50, 199, 226, 253, 289, 307, 310, 26, 294, 348, 162, 204, 209 }	98.57	70904	4581.2	1.0000	1062.4
14	{ 50, 199, 226, 253, 289, 307, 310, 26, 294, 348, 162, 204, 209, 315 }	98.18	70523	4580.9	1.0000	1056.8

2. do not place sensors on $R^* = \{r : \widetilde{\gamma}(r) = 1\} = \{r : \gamma(r) = 0\}$;
3. re-evaluate placements in the (reduced) network with $x_i = 0$ for $i \in R^*$ and $x_i = 1$ for $i \in S^*$.

Table 5.22 illustrates the use of this procedure to produce robust solutions as the sensor budget is varied. Alternately, this re-evaluation can be done with new persistence evaluations, where preference is given to some objectives — for instance, minimizing population exposed may be the most important.

Otherwise, our final step is to use these global persistence values to produce a combined belief function, which we can then maximize. As an example, in Table 5.23a, we list the ten "best" (that is, highest belief) solutions, along with their objective values, for the singleton-only belief function constructed from these global persistence vales using the approach discussed in Sections 4.1 and 5.1.1.1. We contrast the performance of these solutions with the performance of the top solution from each objective's solution list.

Table 5.23: Comparing solution performance across all objectives.

(a) "Best" Persistence-by-grouping solutions

Sensor Placement Selection-Set	$\sum_v \hat{w}_v \Phi(x,v)$ PE	EC	VC	NF	TD
{ 26, 50, 199, 226, 253, 289, 294, 307, 310, 348 }	127.20	90666	4657.2	1.0000	1075.9
{ 26, 50, 162, 199, 226, 253, 289, 294, 307, 310 }	108.32	79946	4952.0	1.0000	1101.3
{ 26, 50, 162, 199, 226, 253, 289, 307, 310, 348 }	109.23	80147	6284.3	1.0000	1102.1
{ 50, 162, 199, 226, 253, 289, 294, 307, 310, 348 }	109.10	79946	12622.5	1.0000	1136.0
{ 26, 50, 199, 204, 226, 253, 289, 294, 307, 310 }	121.14	84290	5027.8	1.0000	1098.1
{ 26, 50, 199, 204, 226, 253, 289, 307, 310, 348 }	122.06	84491	6368.8	1.0000	1099.0
{ 26, 50, 199, 226, 253, 289, 290, 294, 307, 310 }	120.84	84373	5027.7	1.0000	1098.1
{ 26, 50, 199, 226, 253, 289, 290, 307, 310, 348 }	121.76	84574	6368.8	1.0000	1099.0
{ 26, 50, 199, 226, 253, 289, 294, 307, 310, 315 }	121.16	84108	5027.4	1.0000	1097.7
{ 26, 50, 199, 226, 253, 289, 307, 310, 315, 348 }	122.08	84309	6368.5	1.0000	1098.5

(b) Single-objective optimal solutions

Sensor Placement Selection-Set	$\sum_v \hat{w}_v \Phi(x,v)$ PE	EC	VC	NF	TD
{ 50, 123, 162, 199, 226, 290, 307, 310, 326, 453 }	83.38	69580	58113.7	1.0000	$\geq$ 24hrs.
{ 45, 50, 159, 199, 226, 295, 307, 310, 315, 344 }	107.00	55148	58113.7	1.0000	$\geq$ 24hrs.
{ 26, 67, 112, 124, 180, 216, 253, 262, 289, 294 }	259.24	162633	928.7	1.0000	$\geq$ 24hrs.
{ 61, 67, 145, 233, 249, 253, 289, 308, 348, 354 }	259.21	162639	58113.7	0.1232	$\geq$ 24hrs.
{ 26, 45, 124, 228, 253, 262, 289, 294, 346, 348 }	259.27	162360	58113.7	1.0000	953.4min

5.2 Portfolio Selection

In this section, we explore a portfolio tracking problem, where the goal is to reproduce as closely as possible a given financial index or benchmark. This can be done either by constructing a portfolio with the same assets and weights as the index itself (exact replication), or by constructing a smaller subset of assets while minimizing some measure of *tracking error* for some given criteria (partial replication) [25]. For instance, we might wish to hold a small selection of securities to track a broad-market index (for instance, the NASDAQ) that matches as closely as possible the performance of the overall stock market or market sector. In this case, some measure of tracking error must be selected as an objective function to be minimized subject to constraints on the character of the tracking portfolio, such as bounds on the weights of individual assets.

Although there are many possible measures of tracking error, we limit our discussion to the linear measure, *mean absolute deviation* (MAD), where the object is to minimize the sum of absolute deviations between the benchmark returns and the tracking portfolio returns over a training period, $t = t_0, \ldots, T$. Mathematically, given a selection of n securities and a benchmark index, let B_t be the benchmark return and r_{it} be the observed return for security i for time $t = t_0, \ldots, T$. The MAD formulation of tracking error for a portfolio, P, is then given by:

$$TE_{\mathrm{MAD}}(P) = \sum_{t=t_0}^{T} \left| P_t - B_t \right|, \tag{5.3}$$

where P_t, the portfolio return for time t, is a function of asset selections and observed returns, r_{it}.

5.2.1 Base Portfolio Tracking Model

We consider a basic tracking problem where the benchmark portfolio is constructed by taking all the securities under consideration in equal portions. The benchmark returns are then:

$$B_t = \sum_{i=1}^{n} \tfrac{1}{n} r_{it}, \text{ for } t = t_0, \ldots, T. \tag{5.4}$$

Additionally, we constrain our tracking portfolio to consist of a fixed number of securities $N < n$, so the tracking portfolio returns can be written:

$$P_t(x) = \tfrac{1}{N}\sum_{i=1}^{n} x_i r_{it}, \text{ for } t = t_0, \ldots, T, \tag{5.5}$$

where

$$x_i = \begin{cases} 1 & \text{if asset } i \text{ is selected} \\ 0 & \text{otherwise.} \end{cases}$$

By substituting (5.5) into (5.3) and adding the budget constraint, we formulate our *base portfolio tracking problem*:

$$\begin{aligned} \min_{x} \quad & \sum_{t=t_0}^{T} \left| \left(\tfrac{1}{N} \sum_{i=1}^{n} x_i r_{it} \right) - B_t \right| \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = N, \\ & x \in \{0,\, 1\}^n. \end{aligned} \tag{5.6}$$

The tracking portfolio is then constructed by selecting an appropriate training period, $t = t_0, \ldots, T$, and solving (5.6) to determine the portfolio constituents for the coming period, repeating the process at regular *rebalance* periods.

Figure 5.2 shows a graph the actual returns of an example tracking portfolio compared to its benchmark over a sixteen month period. The benchmark returns in this example are constructed by taking the time series of returns of the 30 constituents of the Dow Jones Industrial Average (as of September 2009) and constructing a benchmark portfolio using (5.4). We then construct the

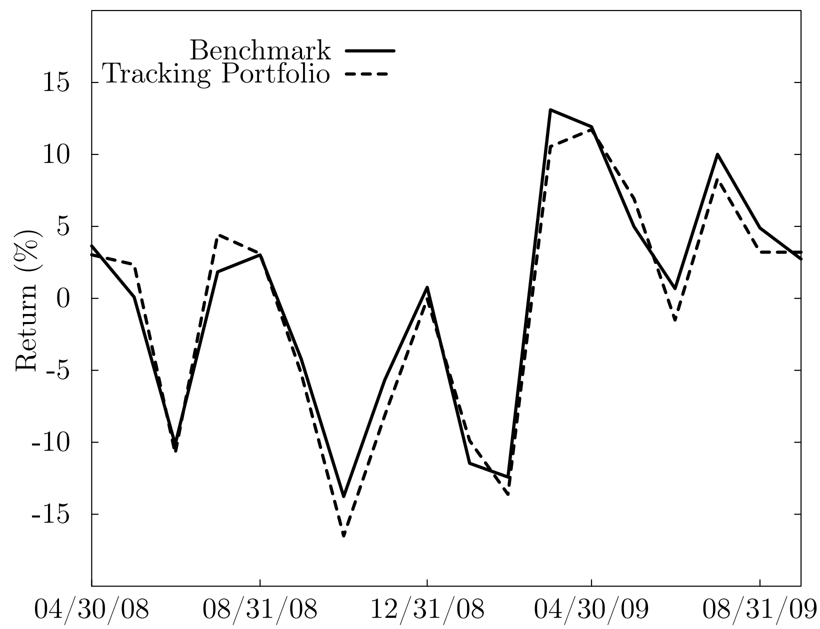


Figure 5.2: Comparing benchmark and tracking portfolio returns.

tracking portfolio rebalancing each month by using the prior month's returns as our training set, setting t_0 to the start and T to the last trading day of the prior month, and solving (5.6) with $N = 10$ to determine the tracking portfolio constituents for month to come calculating realized returns at the end of the month.

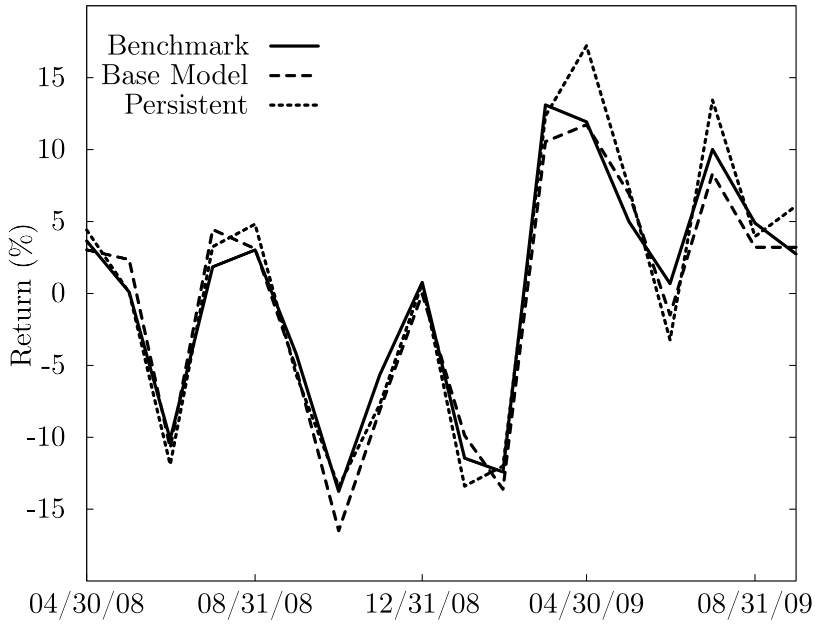
5.2.2 Persistent Portfolio Tracking Model

A common concern when constructing portfolios with regular rebalancing periods is that of preventing excessive *turnover*. This can be motivated by the need to minimize transaction cost, mitigate liquidity problems, or address tax issues. Typically, turnover is addressed by placing a penalty term on some measure of distance from the previous portfolio, but selecting the appropriate penalty can be difficult and even then fails to fully capture the cost of turnover. Instead, our approach uses persistence between time periods to reduce turnover.

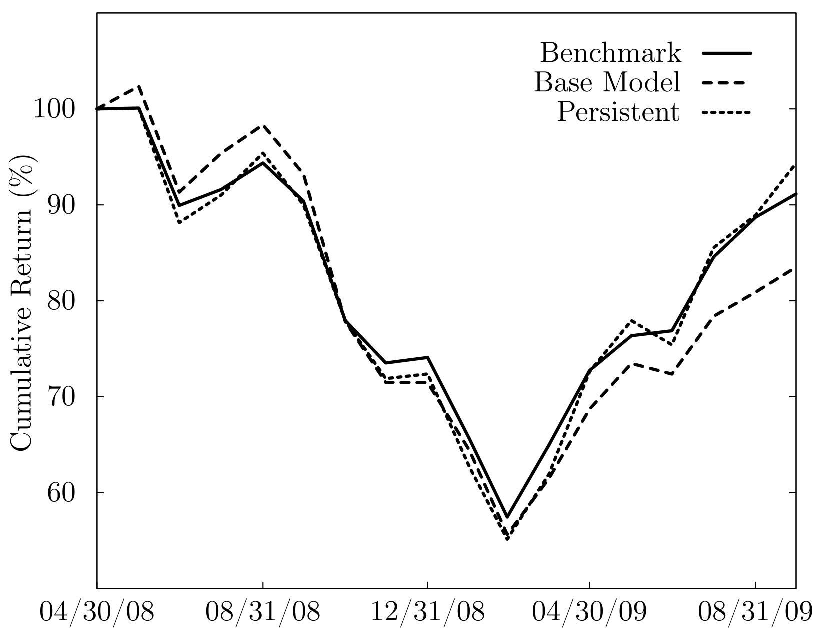
In our base tracking portfolio we determine the portfolio constituents for the coming month by minimizing the tracking error for the daily returns of the previous month. The persistent tracking portfolio, by contrast, selects the securities to invest in by computing the maximum persistence solution over the previous six months of monthly minimum tracking error solution. Thus, to construct the persistent tracking portfolio for a given month, we first solve (5.6) for each of the prior six months to generate a solution list and compute the corresponding global persistence of selection, $\gamma(i)$, for each investable security $i \in \{1, \dots, n\}$. We then determine the portfolio constituents by solving the maximum-persistence model:

$$\begin{aligned} \max_{x} \quad & \sum_{s\,:\,x_s=1} \gamma(s) \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = N, \\ & x \in \{0,\, 1\}^n. \end{aligned} \tag{5.7}$$

In Figure 5.3, we compare the performance of the persistent and base-model tracking portfolios together with the benchmark. Figure 5.3a repeats the graph of portfolio returns from Figure 5.2 adding to it the persistent tracking portfolio's returns. Figure 5.3b presents the same time series of returns represented by the cumulative returns to give an alternate view of the comparison. We note that although, in terms of percent return, the persistent portfolio noticeably



(a) Portfolio Returns



(b) Cumulative Portfolio Returns

Figure 5.3: Comparing persistent and base-model tracking portfolio returns.

overshoots the benchmark in April of 2009, but that, in terms of cumulative returns, the persistent portfolio is actually tracking the benchmark very well. These results suggest that, at least in this example, the persistent tracking portfolio tracks the benchmark every bit as well as the base-model tracking portfolio.

More to the point, the persistent tracking portfolio tracks the benchmark with greatly reduced turnover, as illustrated in Table 5.24. Table 5.24a shows the selections from the base-model tracking portfolio along with the Hamming distance between successive rebalance periods, which gives a measure of turnover. We contrast this with the selections and inter-period Hamming distances for the persistent tracking portfolio, shown in Table 5.24b. Note that turnover, as measured by the Hamming distance between periods, is significantly less in the persistent portfolio than the base-model portfolio.

The low turnover of the persistent model is not surprising. Due to the presence of the constant-sum budget constraint,

$$\sum_{i=1}^{n} x_i = N,$$

Corollary 4.11 from Section 4.4.1 (p. 79) applies. Thus, solving the persistent tracking model is equivalent to choosing a portfolio that minimizes the sum of Hamming distances between the selected portfolio and the last six months' solutions to the base-model (5.6). Though not the same as simply adding a term to the objective function that penalizes distance from the previous solution, this approach dampens turnover. It does so particularly as the number of past

Table 5.24: Comparing persistent and base-model tracking portfolio selections.

(a) Base Tracking Model

Date	Portfolio Assignment																														$H(x_t, x_{t-1})$
2008-04-30	1	1	1			1							1							1				1	1	1	1				n/a
2008-05-31		1	1			1	1	1	1		1						1						1						1		14
2008-06-30			1							1		1		1		1	1		1	1								1	1		14
2008-07-31	1					1					1	1				1			1				1					1	1	1	10
2008-08-31	1			1				1	1									1			1	1	1	1				1			14
2008-09-30	1					1			1	1				1	1	1			1	1									1		16
2008-10-31		1					1	1	1				1			1						1	1					1	1		14
2008-11-30			1	1	1		1													1	1	1					1		1	1	14
2008-12-31					1		1	1					1			1		1			1	1		1				1			12
2009-01-31			1		1					1		1		1		1		1			1	1		1							8
2009-02-28		1					1	1				1				1	1	1		1	1				1						12
2009-03-31			1		1			1	1					1		1	1	1								1			1		12
2009-04-30	1					1				1						1	1	1		1	1			1			1				14
2009-05-31		1			1				1									1		1	1			1	1		1		1		10
2009-06-30	1								1			1		1		1		1	1					1			1	1			12
2009-07-31						1		1		1			1		1			1				1	1			1				1	18
2009-08-31						1		1		1	1				1	1						1	1		1				1		8
2009-09-30	1					1							1		1	1			1	1	1	1			1						10

(b) Persistent Tracking Model

Date	Portfolio Assignment																														$H(x_t, x_{t-1})$
2008-04-30				1			1					1	1		1	1	1							1			1	1			n/a
2008-05-31				1		1	1					1	1		1									1			1	1		1	4
2008-06-30			1	1		1	1		1		1				1		1							1			1				8
2008-07-31			1	1		1	1				1	1					1			1				1				1			6
2008-08-31	1	1	1			1	1				1	1								1								1	1		6
2008-09-30	1	1	1	1		1						1											1	1				1	1		6
2008-10-31	1		1			1			1							1			1	1			1					1	1		8
2008-11-30	1	1				1		1	1							1			1				1					1	1		4
2008-12-31	1		1						1							1			1	1		1	1					1	1		6
2009-01-31	1						1	1	1							1					1	1	1					1	1		6
2009-02-28					1		1	1	1							1		1			1	1		1				1			6
2009-03-31		1			1		1	1								1		1		1	1	1							1		6
2009-04-30		1	1		1		1	1								1		1			1	1							1		2
2009-05-31			1		1		1	1								1	1	1		1	1	1									4
2009-06-30		1	1		1			1								1	1	1		1	1			1							4
2009-07-31					1				1			1		1		1	1	1		1	1			1							6
2009-08-31	1							1	1							1	1	1		1	1			1			1				6
2009-09-30	1					1		1	1	1						1		1						1			1		1		6

periods considered grows, because the solution list used to generate $\gamma(i)$ will only change very slightly (as one period, the oldest, is dropped and the most recent period's allocations added).

6. Concluding Remarks and Avenues for Future Research

Our goal in this thesis is simply to lay the foundations for a new framework for thinking about persistence of decisions and robustness, and in this sense its aim is to open more doors for further research than it closes. For this reason there are many open questions and possible directions for future research. One of the most basic open questions is, "Given a specific problem, a list of acceptable solutions, and persistence results, how do we choose an appropriate belief model?" Prior to answering this question we must first consider how a prescient resolution of uncertainties would change our problem. For instance, in the sensor placement problem, if we knew with certainty where an attack would occur, we could simply place a solitary sensor at that location.

In essence the sensor placement problem is similar to that of a fisherman needing to catch a fish, who first asks various experts where to best drop a line. Each expert responds by giving a list of one or more locations (multiple locations could represent a sampling strategy — that is, "if I have three rods, I'd place them here, here, and here"). Using the decisions-set model, the fisherman constructs a decision-sets belief function from the given selection advice and maximizes his confidence that he will succeed in catching a fish by placing lines at those locations that maximize the resulting function subject to any constraints (for example a limit on how many lines he can set). In the case of the knapsack problem, however, prescience still leaves us with a deterministic knapsack problem (a non-trivial problem in itself) since, even with perfect knowledge of their values we would still select as many items as possible in order to maximize our take, so a belief function that considers solutions has a whole (like the singleton-only bpa) would better serve our purpose. Furthermore, the belief functions developed in Chapter 4 are not exhaustive, and we must consider whether some other belief function would better suit our problem — perhaps one based on the combined global persistence factor, $\Psi(x)$.

Another avenue worthy of further consideration is the treatment of multiple objectives. In the approaches developed in Section 4.3, the evidence from each of the objectives is treated equally, which is not appropriate for all problems. Instead, we might consider assigning weights to the evidence provided by each objective. Alternatively, we can consider variations on single objective optimization. For instance, we might consider how persistence relates to robustness and stability as we vary constraints (such as a budget constraint) or how fixing variables in the original problem relates to conditional belief values (in other words, "How does a list of the 'best' solutions generated using persistence and conditional beliefs relate to solving the original problem with fixed decision variables?").

Additionally, there are conceptual ties and connections with other approaches that need further investigation. In Sections 4.4.1 and 4.4.2, we explored some relations to Hamming distance and information theory that should be explored further. Most importantly, in Section 4.4.3, we discussed the Evidence Paradox and the question of whether there is a relation between the entropy in the belief function and the magnitude of the uncertainties in the underlying model, which is most certainly a good avenue for further research. Besides these connections, there is the connection with the work of Brown et al. [5], which emphasizes the managerially preferable aspect of decision persistence. Or, the possibility that allowing managerial input into the creation of the "acceptable solutions" list could enable us to capture "expert" experience and knowledge that might be missing from the model (perhaps by means of a hybrid approach that combines successive exclusions with expert advice).

Finally, there is the consideration of additional applications for our approach. As developed, our approach should be readily applicable to most combinatorial problems: assignment and matching, network problems, facility location, and many others. In fact, since our primary input is the solutions list, our approach is largely indifferent to the structure of the underlying model and we can handle difficult, non-linear models as easily as simpler linear models. One complication, however, is that as it is currently developed our model of persistence does not readily translate to non-combinatorial optimization problems, where the decision variables are not naturally binary. For instance, although integer programming problems can generally be reformulated to use binary variables, the process of doing so will likely distort the meaning of persistence with respect to the actual decisions and some thought will be needed to develop formulations that do preserve this meaning.

NOTATION

$\vec{\mathbf{1}}$ $\equiv$ vector of ones.

x^k $\equiv$ kth "optimal" solution.

z^k $\equiv$ kth "optimal" value (that is, $z^k = c^{\mathsf{T}} x^k$).

x_i^k $\equiv$ value of ith element, $x_i \in \{0, 1\}$, in x^k.

$\mathcal{I}$ $\equiv$ Decision set, $\{1, \ldots, n\}$.

X^K $\equiv$ Solution list, $X^K = \{x^1, \ldots, x^K\}$.

S^k $\equiv$ kth Selection set, $\{i \in \mathcal{I} : x_i^k = 1\}$.

R^k $\equiv$ kth Rejection set, $\{i \in \mathcal{I} : x_i^k = 0\}$.

C^k $\equiv$ kth Consensus set, $C^k = S^0 \cap S^1 \cap \cdots \cap S^k$.

$\mathcal{C}^k$ $\equiv$ kth Consensus multiset, $\mathcal{C}^k = \mathcal{S}^1 \cap \cdots \cap \mathcal{S}^k$ for $k = 1, \ldots, K$.

b_i $\equiv$ ith breakpoint for $i = 0, \ldots, m$.

$\mathcal{S}^i$ $\equiv$ ith breakpoint multiset, $\mathcal{S}^i = \biguplus_{b_i \leq k < b_{i+1}} S^k$ for $i = 0, \ldots, m$.

$\mathcal{N}(x, \mathcal{X})$ $\equiv$ number of occurrences of element x in multiset $\mathcal{X}$.

$\rho_i(s)$ $\equiv$ local persistence of selection s in the ith breakpoint.

$\gamma(s)$ $\equiv$ global persistence of selection s.

$\widetilde{\gamma}(s)$ $\equiv$ global persistence of rejection r.

$\Gamma(S)$ $\equiv$ global selection persistence factor for selection set S.

$\widetilde{\Gamma}(R)$ $\equiv$ global rejection persistence factor for rejection set R.

$\widehat{\Gamma}(S)$ $\equiv$ adjusted global selection persistence factor for selection set S.

$\Psi(S, R)$ $\equiv$ combined global persistence factor.

$\underline{\Gamma}(S) \equiv$ normalized global selection persistence factor for selection set S.

$\widetilde{\underline{\Gamma}}(R) \equiv$ normalized global rejection persistence factor for rejection set R.

$\underline{\Psi}(S, R) \equiv$ normalized combined global persistence factor.

$\Omega \quad \equiv$ universal set (or "frame of discernment").

$\mathcal{P}(\Omega) \equiv$ power set of the universal set (or "frame of discernment").

$\mathrm{Bel}(x) \equiv$ belief of x.

$\mathrm{Plaus}(x) \equiv$ plausibility of x.

$\mathrm{m}(x) \equiv$ bpa value for x.

$\mathrm{Bel}^{\mathrm{so}}(x) \equiv$ singleton-only belief of x.

$\mathrm{Bel}^{\mathrm{ds}}(x) \equiv$ decision-sets belief of x.

$\mathrm{Bel}(x \,|\, \bullet) \equiv$ conditional belief function.

$p \quad \equiv$ persistence vector, $p = [\,\gamma(x_1) \;\cdots\; \gamma(x_n)\,]^{\mathsf{T}}$.

$H(x, y) \equiv$ Hamming distance, $H(x, y) = \sum_i |x_i - y_i|$.

GLOSSARY

Persistence

breakpoint a value, $k = b_0, b_1, \ldots$, in the sequence of solutions generated by the method of successive exclusions for which the objective value changes, that is, where $z^k < z^{k-1}$.

complement a decisions whose inclusion includes another.

consensus set the set of selections common to all the selection sets: $C^k = S^1 \cap S^2 \cap \cdots \cap S^k$.

consensus multiset the multiset of selections common to all the selection multisets: $\mathcal{C}^i = \mathcal{S}^1 \cap \cdots \cap \mathcal{S}^i$ for $i = 1, \ldots, m$.

dominated a situation where there exist higher-ranking solutions with better objective values and less total weight (that is, a solution vector y is *dominated* by x if $cx > cy$ and $ax \leq ay$).

global persistence the frequency of occurrence for a selection s for the full horizon of K solutions:

$$\gamma(s) = \frac{\mathcal{N}(s, \uplus_{1 \leq k \leq K} S^k)}{K}.$$

global selection persistence factor the frequency of occurrence for a set of selections S for the full horizon of K solutions:

$$\Gamma(S) = \frac{\left|\{k \,:\, S \subseteq S^k \text{ for } 1 \leq k \leq K\}\right|}{K} \text{ for } S \neq \emptyset \quad \text{and} \quad \Gamma(\emptyset) = 0.$$

adjusted global selection persistence factor an alternative form of the global selection persistence factor that normalizes the weight contributed by each solution:

$$\widehat{\Gamma}(S) = \frac{\displaystyle\sum_{\{1 \leq k \leq K \,:\, S \subseteq S^k\}} (2^{|S^k|} - 1)^{-1}}{K} \text{ for } S \neq \emptyset \quad \text{and} \quad \widehat{\Gamma}(\emptyset) = 0.$$

global rejection persistence factor the frequency of occurrence for a set of rejections R for the full horizon of K solutions:

$$\widetilde{\Gamma}(R) = \frac{\left|\{k\,:\,R \subseteq R^k \text{ for } 1 \le k \le K\}\right|}{K} \text{ for } R \neq \emptyset \quad \text{and} \quad \widetilde{\Gamma}(\emptyset) = 0.$$

combined global persistence factor a function that combines the global persistence factors for a selection set S and a rejection set R as a simple sum:

$$\Psi(S, R) = \Gamma(S) + \widetilde{\Gamma}(R) \quad \text{for} \quad S \cap R = \emptyset.$$

In particular, $\Psi(S, \emptyset) = \Gamma(S)$ and $\Psi(\emptyset, R) = \widetilde{\Gamma}(R)$.

local persistence the frequency of a selection s is its frequency within a breakpoint region i:

$$\rho_i(s) = \frac{\mathcal{N}(s, \mathcal{S}^i)}{b_i - b_{i-1}},$$

where

$$\mathcal{S}^i = \biguplus_{b_{i-1} \le k < b_i} S^k \text{ for } i = 1, \ldots, m.$$

mode the solution with maximum total persistence.

multiset a set for which repeated elements are considered.

persistence a measure of the continued presence of a selection s in a collection of "acceptable" solution sets.

persistence vector a vector of global persistence values: $p = [\,\gamma(x_1)\ \cdots\ \gamma(x_N)\,]^\mathsf{T}$.

substitute a decisions whose inclusion excludes another.

successive exclusions a method for generating a collection of optimal (and near-optimal) solutions, where we first obtain an initial optimal solution and then re-solve the problem with additional constraints to exclude the previously found solutions. This process repeats, first finding any alternative optima, and then continues generating solutions, allowing the objective value to decrease ensuring that no solutions are regenerated.

Dempster-Shafer Theory

additive belief function a belief function that satisfies the strict additive property:

$$\mathrm{Bel}(A \cup B) = \mathrm{Bel}(A) + \mathrm{Bel}(B).$$

These belief functions are also called Bayesian belief functionss.

basic probability assignment (bpa) a function $\mathrm{m} : \mathcal{P}(\Omega) \to [0,1]$ such that

$$\mathrm{m}(\emptyset) = 0$$
$$\sum_{A \subseteq \Omega} \mathrm{m}(A) = 1.$$

belief function a function, $\mathrm{Bel} : \mathcal{P}(\Omega) \to [0,1]$, satisfying the conditions:

$$\mathrm{Bel}(\emptyset) = 0$$
$$\mathrm{Bel}(\Omega) = 1$$
$$\mathrm{Bel}(A) = \sum_{B \subseteq A} \mathrm{m}(A).$$

discount & combine a method for combining belief functions, $\mathrm{Bel}_i(A)$, where the beliefs are first discounted by multiplying all proper subsets A of Ω by a discount factor $0 < \alpha_i < 1$, so that

$$\mathrm{Bel}_i^{\alpha}(A) = \alpha_i \, \mathrm{Bel}_i(A) \text{ for } A \neq \Omega \quad \text{and} \quad \mathrm{Bel}_i^{\alpha}(\Omega) = 1,$$

before being combined using Dempster's Rule.

dogmatic in Dempster-Shafer theory, a belief function whose underlying bpa assigns support only to proper subsets of the *frame of discernment* (hence, $\mathrm{m}(\Omega) = 0$).

focal element any subset A of the frame of discernment assigned non-zero support by the bpa (that is, any $A \subseteq \Omega$ such that $\mathrm{m}(A) > 0$).

frame of discernment the set of all the possibilities under consideration, or the universe of discourse.

plausibility function a function, Plaus : $\mathcal{P}(\Omega) \to [0,1]$, that satisfies the conditions:

$$\begin{aligned} \text{Plaus}(\emptyset) &= 0 \\ \text{Plaus}(\Omega) &= 1 \\ \text{Plaus}(A) &= \sum_{B \cap A \neq \emptyset} \text{m}(A). \end{aligned}$$

A plausibility function is related to it's companion belief function by the formula: $\text{Plaus}_m(A) = 1 - \text{Bel}_m(\neg A)$.

subadditivity a function μ on Ω such that $\mu(A \cup B) \leq \mu(A) + \mu(B)$ when $A, B \subseteq \Omega$.

superadditivity a function μ on Ω such that $\mu(A \cup B) \geq \mu(A) + \mu(B)$ when $A, B \subseteq \Omega$.

Applications

injection weight in the sensor placement problem, information about the injection point, which takes the form of weights, $w \geq 0$, where $\sum_v w_v = 1$.

mean absolute deviation (MAD) a linear measure of tracking error, where the object is to minimize the sum of absolute deviations between the benchmark, B_t, and the tracking portfolio, P_t, over a training period, $t = t_0, \ldots, T$, given by:

$$TE_{\text{MAD}}(P) = \sum_{t=t_0}^{T} |P_t - B_t|.$$

tracking error in the portfolio tracking problem, a measure of the difference between the benchmark returns and the tracking portfolio returns.

turnover in the portfolio problem, the extent to which the portfolio allocations change between successive periods. Typically, excessive turnover is something to be avoided in order to minimize transaction costs and avoid negative consequences such as increased tax rates and liquidity problems.

REFERENCES

[1] W. P. Adams, J. B. Lassiter, and H. D. Sherali. Persistency in 0-1 polynomial programming. *Mathematics of Operations Research*, 23(2):359–389, 1998.

[2] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[3] A. Ben-Tal and A. Nemirovski. Robust optimization — methodology and applications. *Mathematical Programming , Series B*, 92:453–480, 2002.

[4] D. Bertsimas, K. Natarajan, and C.-P. Teo. Persistence in discrete optimization under data uncertainty. *Mathematical Programming, Series B*, 108:251–274, 2006.

[5] G. G. Brown, R. F. Dell, and R. K. Wood. Optimization and persistence. *Interfaces*, 27(5):15–37, 1997.

[6] G. G. Brown, C. E. Goodman, and R. K. Wood. Annual scheduling of atlantic fleet naval combatants. *Operations Research*, 38(2):249–259, 1990.

[7] R. D. Carr, H. J. Greenberg, W. E. Hart, G. Konjevod, E. Lauer, H. Lin, T. Morrison, and C. A. Phillips. Robust optimization of contaminant sensor placement for community water systems. *Mathematical Programming*, Series B, 107:337–356, 2006.

[8] A. Charnes and W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.

[9] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3/4):197–206, 1955.

[10] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. 38(2):325–329, 1967.

[11] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988.

[12] F. Glover. Heuristics for integer programming using surrogate constraints. *Decision Sciences*, 8:156–166, 1977.

[13] H. J. Greenberg. Uncertainty: An OR frontier. Bibliography, University of Colorado Denver, 2006. Addendum to INFORMS Presentation (available upon request from author).

[14] H. J. Greenberg. Representing uncertainty in decision support. *OR/MS Today*, 34(3):14–16, 2007.

[15] H. J. Greenberg and T. Morrison. Robust optimization. In A. R. Ravindran, ed., *Operations Research and Management Science Handbook*, The Operations Research Series, chap. 14. CRC Press, Boca Raton, FL, 2008.

[16] A. Holder, ed. *Mathematical Programming Glossary*. INFORMS Computing Society, http://glossary.computing.society.informs.org, 2006–10. Originally authored by Harvey J. Greenberg, 1999-2006.

[17] G. J. Klir. Generalized information theory. *Fuzzy Sets and Systems*, 40(1):127–142, 1991.

[18] G. J. Klir. An update on generalized information theory. In *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, pp. 321–334. held in Lugano, Switzerland, 2003. Available at http://www.carleton-scientific.com/isipta/PDF/024.pdf.

[19] G. J. Klir. *Uncertainty and Information: Foundations of Generalized Information Theory*. Wiley-Interscience, Hoboken, NJ, 2006.

[20] P. Kouvelis and G. Yu. *Robust Discrete Optimization and Its Applications*. Nonconvex Optimization and Its Applications. Kluwer Academic Press, Norwell, MA, 1997.

[21] M. T. Lamata and S. Moral. Measures of entropy in the theory of evidence. *International Journal of General Systems*, 14(4):297–305, 1987.

[22] S. Manandhar, S. Man, A. Tarim, and T. Walsh. Scenario-based stochastic constraint programming. In *Proceedings of International Joint Conference On Artificial Intelligence (IJCAI-2003)*. Morgan Kaufmann, 2003.

[23] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[24] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios. Robust optimization of large-scale systems. 43(2):264–281, 1995.

[25] J.-L. Prigent. *Portfolio Optimization and Performance Analysis*. Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2007.

[26] K. Sentz and S. Ferson. Combination of evidence in Dempster-Shafer theory. Sandia Report SAND2002-0835, Sandia National Laboratories, 2002.

[27] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ, 1976.

[28] J. Slaney and T. Walsh. Backbones in optimization and approximation. In *Proceedings of International Joint Conference On Artificial Intelligence (IJCAI-2001)*, pp. 254–259. Morgan Kaufmann, 2001.

[29] A. Wald. Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, 10(4):299–326, 1939.

[30] T. Walsh. Stochastic constraint programming. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI)*. IOS Press, 2002.

[31] J.-P. Watson, H. J. Greenberg, and W. E. Hart. A multiple-objective analysis of sensor placement optimization in water networks. In G. Sehike, D. F. Hayes, and D. K. Stevens, eds., *Proceedings of the World Water and Environmental Resources Conference*. 2004.

[32] J.-P. Watson, W. E. Hart, and R. Murray. Formulation and optimization of robust sensor placement problems for contaminant warning systems. In S. G. Buchberger, R. M. Clark, W. M. Grayman, and J. G. Uber, eds., *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*. ASCE, 2006.

[33] R. R. Yager. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9(4):249–260, 1983.

[34] R. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41(2):93–137, 1987.