2012

Completing the Information Chain: On the Theory of Optimal Information Acquisition for Quantitative Decision Making

David Paul Grace
Lehigh University

COMPLETING THE INFORMATION CHAIN: ON THE THEORY OF OPTIMAL INFORMATION ACQUISITION FOR QUANTITATIVE DECISION MAKING

by

David Paul Grace

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Industrial Engineering

Lehigh University

May 2012

Received and approved by the Doctoral Committee directing the proposed program of study for David Grace, a Ph.D. Candidate in the Department of Industrial and Systems Engineering.

Date

Dr. Eugene Perevalov, Chairman
Industrial and Systems Engineering
Lehigh University

Dissertation Advisor

Accepted Date

Committee Members:

Dr. Eugene Perevalov, Chairman
Industrial and Systems Engineering
Lehigh University

Dr. Robert Storer
Industrial and Systems Engineering
Lehigh University

Dr. George Wilson
Industrial and Systems Engineering
Lehigh University

Dr. Vladimir Dobric
Mathematics
Lehigh University

Acknowledgments

Throughout my time spent at Lehigh and especially during the final year, I was supported by many people. They have all contributed to the successful completion of this work.

First and foremost, I would like to thank my advisor, Dr. Eugene Perevalov, for his guidance throughout my PhD. It has been a great pleasure to work with him. His intellectual depth and insight into a diverse range of topics have enlightened my experience. He supported me throughout and always had time for our many discussions on ideas on how to improve the work or new directions to explore. I also appreciate the many more discussions we had on sports, society and life in general.

I would also like to thank Dr. Vladimir Dobric, Dr. Robert Storer, and Dr. George Wilson for all their time and efforts. They were always willing and able to provide advice and commentary and their suggestions and insights helped to grow this work.

And to all my friends and family, especially my parents Paul and Sally Grace. I appreciate all that you've done for me over the years, more than any words in a document could convey.

Contents

List of Tables

List of Figures

Abstract

Classical Information Theory solves the problem of maximizing the quantity of information that can be reliably transmitted over the given (imperfect) channel, without dealing with the questions of where the information comes from and what it's going to be used for. If the information is to be used to make decisions, and the goal is to maximize the decision quality (e.g. by minimizing the properly defined loss) by making use of available information sources, then one needs to know what specific information is to be requested from a source so that, on one hand, the source would be able to fulfill the request accurately and, on the other hand, the information obtained would have a large impact on the decision quality for the specific problem at hand. It can be said that the developed methodology complements the classical Information Theory in that it deals–in the context of quantitative decision making at least–with the first and last link of the full "information chain": extracting it from the source and using it to obtain the best possible decision. The classical Information Theory describes the middle link of that chain–in case a transmission of the information obtained from the source over some channel is involved. The middle link just happens to be largely independent of the end links and can be treated separately, while the end links are rather closely connected and therefore have to be treated together. It is curious to note that a similar state of affairs can often be observed in material supply chains: for example, if some raw material has to be extracted, transported and

used to make a certain product, then the material to be extracted depends on the product that needs to be made and the given material can be extracted best from a certain source. The transportation task, however, has a more universal character and can usually be considered in abstraction from the nature of the particular material.

In classical Information Theory, the main question is two-fold: what is the maximum (theoretical) speed of accurate transmission for the given channel and how that speed can be (practically) achieved. The first part of the main question is addressed by calculating the channel capacity and the second part of the main question is addressed by designing appropriate codes for input symbols. The main question being addressed in the proposed approach is also two-fold: what is the maximum decision quality (for the given problem) that can be achieved by using the available information source(s) and what is the practical way of achieving that quality. The first part of this question is addressed by computing the pseudo-energy/loss efficient frontier for the given problem and (source specific) pseudo-temperature function, and the second part is addressed by designing appropriate questions (that lie on the efficient frontier) as means of extracting information from the source(s) optimally with respect to the decision making problem being solved.

Chapter 1

Introduction

1.1 Overview

When uncertainty is present, several approaches to decision making are used depending on the problem at hand. If the main difficulty lies in a large number of possible solutions as well as a complex structure of the feasible region then optimization methods are usually used (stochastic [6], robust [2, 4] or, more recently, risk-averse [16, 54, 3]). The information available about the unknown problem parameters is usually assumed to be fixed. If the number of possible solutions is relatively small and the main difficulty lies in the process of updating the initial information, decision theoretic methods are appropriate. In Markov decision processes and stochastic optimal control, additional assumptions (such as Markovian or Gaussian property) are made which allows one to obtain solutions with special properties making it possible to handle the dynamic aspect of the problem efficiently.

In many practically important decision making problems where uncertainty about input data is present and optimization methods are appropriate, sources of additional information are in principle available. Often, information that such sources possess fails to be taken advantage of due to its perceived and factual imprecision and to the lack of a methodology that allows for this in a controlled and regular fashion.

This is the main motivation for the approach developed here: the need to efficiently add "little pieces" of useful information to the information already present in a decision making problem formulation. A typical situation when such ability is needed arises in industrial product portfolio selection problems. An electronics manufacturing company, for example, has to choose which products to schedule for production (and which currently produced products to phase out) in the next quarter. The candidate products are characterized with the respective production costs (that are relatively well known at the time of the decision) and future demands (that are very uncertain at the same time). There are also relations between production costs and demands of various products that can be written as constraints. A stochastic optimization formulation can typically be developed with a probability measure obtained from historic data. On the other hand, decision makers know that there exists other useful information that is "spread around the organization" which consistently fails to get utilized because of the inability of decision makers and analysts to properly extract it. Moreover, the above-mentioned inability to extract additional pieces of useful information often results in decisions being made simply based on decision makers' intuition and qualitative judgement because of the perceived imprecision of the available probability measure.

In what follows, we initiate development of a unified theoretical framework for optimal information acquisition in general purpose decision making problems including those with large and complex feasible regions to address such a situation. The approach begins with the assumption that one or several information sources are available that are capable of providing potentially *various* (i.e. qualitatively different) "bits" of additional information on top of what's already contained in the initial probability measure. The assumption of having available such "multi-purpose" information sources is made to describe primarily human experts that possess a certain "picture" of the way the investigated system will likely develop in the future and capable of internally "processing" that picture to answer specific questions concerning possible future outcomes. Generally speaking, any source has finite capability that manifests itself in answering easy questions with higher accuracy than difficult ones. Difficulty of various questions is source-specific: what is easy for one source can be difficult for another and vice versa.

On the other hand, information contained in an answer to any question carries a certain *value of information* with respect to the given decision making problem. The latter measures the improvement in the value of the problem objective resulting from the information contained in the answer. The decision maker would naturally be interested in maximizing this value of information [30] and can achieve this goal by carefully choosing a question that would be sufficiently easy for the source to yield an accurate answer and, at the same time, relevant to the problem at hand so that the resulting value of information would have the highest possible value.

This naturally leads to an important question the decision maker appears to be facing: how the information source should be optimally "aligned" with the given problem, or, more precisely, what question the decision maker should ask the information source so that the respective answer would have the largest positive effect on the solution quality for the given problem. More generally, if several information sources are available the decision maker would want to know what question(s) and, possibly, in what order the sources should be asked so that the combined effect of the respective answers on the solution quality can be maximized. In other words, here the overall problem is that of optimal "alignment" of a system of information sources to the given decision making problem. What can make that latter problem more difficult is that optimal question(s) to be asked a given source might in general depend on the number and properties ("expertise") of other available sources.

If such a methodology is to be developed, it seems logical to begin with (i) a quantitative framework describing information sources, questions and answers, (ii) study relationships between questions and the value of information of answers of the given source to these questions and (iii) use the results of (i) and (ii) to develop algorithms for choosing optimal questions and thus optimizing the process of acquiring additional information from the available source(s) for the decision making problem of interest.

1.2 Related Work

The idea of obtaining additional information to improve the quality of decisions in situations characterized with uncertainty is obviously not an entirely new idea and it has been pursued, for instance, in the area of statistical decision making. Applications to innovation adoption [44], [35], fashion decisions [20] and vaccine composition decisions for flu immunization [40] can be mentioned in this regard. It's interesting to observe that the amount of information in these applications is typically measured simply as the number of relevant observations which can be either costless or costly, depending on the model. Some authors [19, 17] introduced various models (e.g. effective information model) for accounting for the actual, or effective, amount of information contained in the received observations. The common theme of this line of work is to try to find an optimal trade-off between the amount of additional information obtained and the suitably measured degree of achieving the original goal. Thus, for instance, in [40], waiting longer allows the decision makers to obtain more precise forecast of which flu virus strains are going to be predominant but leaves less time for actual vaccine production. The difference of the proposed approach is that it explicitly describes and allows to optimize over not just the quantity of additional information but also its content and is based on explicit description of properties of information sources. As another example of this overall line of research, one should mention the recent work on optimal decision making in the absence of the knowledge of the distribution shape and parameters [31, 41, 1]. Instead, the decision maker observes historic data and updates the solution according to an algorithm whose purpose is to minimize the difference in objective relative to a complete knowledge of the uncertain parameter distribution. Thus an optimal usage of the available information is also explicitly considered.

This work can also be looked upon as an attempt to make Information Theory methods useful for optimization and decision making under uncertainty. The field of Information Theory, born from Shannon's work on the theory of communications [57] has had great success in a number of fields – besides communications itself which it revolutionized – that include statistical physics [33, 34], computer vision [60], climatology [45, 59], physiology [37] and neurophysiology [10]. The relatively new field of Generalized Information Theory (see e.g. [38]) is concerned with problems of characterizing uncertainty in frameworks that are more general than classical probability such as Dempster-Shafer theory [56]. There it was shown, for example, [43, 27] that the minimal uncertainty measure satisfying consistency requirements (such as general subadditivity and additivity for combining uncertainty for independent subsystems) is obtained by maximizing Shannon entropy over all classical probability distributions consistent with the given (generalized) belief specification.

In Chapter 2, we use an axiomatic approach to determine the overall form of the question difficulty function. Chapter 3 uses a similar axiomatic approach to determine the overall form of the answer depth function. Together question difficulty and answer depth can be thought of as a logical development of the entropy concept of information theory. The axiomatic approach was first used, besides Shannon himself, in [18] to derive the most general form of the entropy function. Later, [52] used a different set of axioms to find the one-parameter family of functions (later called Rényi entropies) that included standard (Shannon) entropy as a special case. The concept of structural entropy was introduced in [28] and used for classification purposes. Also known as Havrda-Charvat entropy, it was more recently obtained by axiomatic means in [58] where axiomatization of partition entropy was discussed on rather general grounds (see also [32] for closely related work). It was shown in [58] that Shannon entropy, Havrda-Charvat entropy and Gini index all obtain as particular cases of general partition entropy that satisfies a system of reasonable axioms.

The approach developed here can be interpreted as a theory of information exchange between the decision maker/analyst and information source(s). Similarly, it can be thought of as a development of a general theory of inquiry that goes back to the work of Cox [13, 14]. This line of work received more attention recently resulting in a formulation of the calculus of inquiry [39] that constructs a distributive lattice of questions dual to the Boolean lattice of logical assertions. The definition of questions adapted in Chapter 2 corresponds to the particular subclass of questions – the partition questions – defined in [39]. The work here goes beyond that on the calculus of inquiry in that it introduces the concept of *pseudo-energy* as a measure of source specific difficulty of various questions to the given information source. One could say that it develops a quantitative theory of *knowledge* as opposed to the theory of information.

Explicit consideration of information sources that lies at the core of the proposed methodology is similar in spirit to analyzing and using information provided by human experts. In fact, in many practically relevant applications the role of multi-purpose information sources used in the proposed approach will likely be played by experts. In existing research literature, the problem of optimal usage of information obtained from human experts has been addressed mostly in the form of updating the decision maker's beliefs given probability assessment from multiple experts [22, 23, 11, 12] and, in particular, optimal combining of expert opinions, including experts with incoherent and missing outputs [47]. Closely related to the approach initiated here are the investigations on using and combining information of experts that partition the event differently [7] and on rules of updating probabilities based on outcomes of partially similar events [8]. The latter investigations essentially consider experts that provide qualitatively different information. The dependence of the quality of experts' output on the particular partition was also studied in [21]. Here, the emphasis is on *optimizing* on the particular type of information (i.e. partition) for the given expert(s) and the given decision making problem.

Evaluating a source's ability to answer various questions is closely related to the evaluation of probability forecasts by scoring rules. A scoring rule measures the accuracy of a forecast by computing a score based on how the forecast compares to the actual realization of the uncertain event. An early application of this is the Brier, or quadratic, score that evaluates probabilistic weather forecasts [9]. Scoring rules also provide an incentive for the forecaster to provide truthful probabilities and share a connection to subjective probability theory (e.g. [25, 55]. See [61, 5, 24] for a more thorough discussion of scoring rules and literature reviews. More closely aligned to the work here has been the development of scoring rules that also take into consideration the decision problem at hand, in particular [36]. They start with a decision problem and find scoring rules to fit the problem in a way that aligns interests of the expert and the decision maker. In contrast, to these scoring rules that measure the forecast with a single aggregated scalar value, our work introduces a *pseudo-temperature* function that evaluates the source over the entire state space. In this way, when there are multiple sources of information, the proper source can be chosen based on which one can more accurately answer the specified questions.

Methodologically, Chapter 6 borrows heavily from the field of probability metrics and scenario reduction in stochastic optimization. More details, along with relevant references, can be found in Appendix B and C.

1.3 Motivation: Decision Making Under Uncertainty

In decision making under uncertainty, the goal is to choose the best decision given the available information, according to a suitable criterion. One of the most widely used criteria is that of optimizing the *expected* objective function given the probability distribution that describes the available information. The problem so formulated can be formally written as

$$\min_{x\in X}\mathbb{E}_P f(\omega,x)=\int_\Omega f(\omega,x)P(d\omega). \tag{1.1}$$

Here $X \subset \mathcal{D}$ is the set of all *feasible* solutions, i.e. the set satisfying all (deterministic) constraints that are present in the problem formulation, where $\mathcal{D}$ is the space to which all solutions belong (e.g. a suitable Euclidean space). Ω has the meaning of a space of possible values of input data parameters that are not known with certainty. It is often referred to as a parameter space. P is a fixed initial probability measure (with a suitable sigma-algebra assumed) on Ω that describes the initial state of the uncertainty and that can in principle be modified by querying information sources. The function $f\colon \Omega \times \mathcal{D} \to \overline{\mathbb{R}}$ is assumed to be integrable on Ω for each $x \in X$. For example, in the context of stochastic optimization, X is the set of feasible first-stage solutions and $f(\omega, x)$ is the best possible objective value for the first stage decision x in case when the random outcome ω is observed.

We are interested, given the problem (1.1) and an information source capable of providing answers to our questions, in obtaining the best possible solution to problem (1.1), suitably modified by the source's answer(s). To make this desideratum a bit more specific, let $L(P)$ be the *expected loss* corresponding to measure P defined as follows.

$$L(P) = \int_\Omega f(\omega, x^*_P) P(d\omega) - \int_\Omega f(\omega, x^*_\omega) P(d\omega),$$

where x^*_P is a solution of (1.1) and x^*_ω is a solution of $\min_{x \in X} f(\omega, x)$ for the given ω.

Let $\mathcal{Q}$ be the set of all possible (suitably defined) questions that can be directed towards the source of information, and let $A(Q)$ be its answer to a particular question $Q \in \mathcal{Q}$. Further, let P_a be the measure on Ω conditional on reception of a particular value a of the answer A. One can think of P_a as the measure updated by the value a, from the original measure P. Then the expected loss following question Q and answer $A = A(Q)$ can be found as

$$L(P, Q, A(Q)) = \sum_a \Pr(A(Q) = a) \left(\int_\Omega f(\omega, x^*_{P_a}) P_a(d\omega) - \int_\Omega f(\omega, x^*_\omega) P_a(d\omega) \right), \qquad (1.2)$$

where the sum is over all possible values a of the answer A.

Our goal then can be stated as that of finding, for the given problem (1.1) and a given information source, the question(s) $Q \in \mathcal{Q}$ that would make the corresponding expected loss (1.2) as small as possible:

$$\min_{Q\in\mathcal{Q}} L(P, Q, A(Q)). \tag{1.3}$$

Informally speaking, the problem is about finding the question(s) that is "aligned" optimally with both the information source's "strengths" and the particular decision making problem. Changing the purely "optimization" component of the problem (the function $f(\omega, x)$ and the set X) while keeping the "information" component (the space Ω and the measure P) the same will in general change the optimal question(s) Q for the same information source. Thus the main goal can also be described as that of finding an optimal alignment between the optimization and information components of the problem (where the information source itself is included in the latter).

1.4 Preliminaries

In the following we denote by Ω the base space consisting of all possible outcomes of potential interest to the decision maker. We will often refer to it, as mentioned earlier, as parameter space. Ω can be finite or infinite, such as a closed subset of a Euclidean space $\mathbb{R}^s$. We denote by $\mathcal{F}$ a sigma-algebra on Ω. Let P be a fixed probability measure on $(\Omega, \mathcal{F})$. We will usually refer to it – and other measures – as a measure on Ω, omitting an explicit specification of $\mathcal{F}$ unless needed.

Let $C \in \mathcal{F}$ be a (measurable) subset of Ω. We denote by P_C the conditional measure on Ω

defined as

$$P_C(D) = \frac{P(D \cap C)}{P(C)}, \tag{1.4}$$

for arbitrary $D \in \mathcal{F}$.

A partition $\mathbf{C} = \{C_1, \ldots, C_r\}$ of Ω is a collection of (measurable) subsets $C_j \in \mathcal{F}$ of Ω such that $C_j \cap C_l = \emptyset$ for $j \neq l$ and $\cup_{j=1}^{r} C_j = \Omega$. A partition $\tilde{\mathbf{C}}$ is a *refinement* of $\mathbf{C}$ if every set from $\tilde{\mathbf{C}}$ is a subset of some set from $\mathbf{C}$. In such a case, $\mathbf{C}$ is a *coarsening* of $\tilde{\mathbf{C}}$. Given measure P on Ω, we call partition $\mathbf{C}_f(P)$ the *finest* partition of Ω associated with measure P if $P(C) > 0$ for all $C \in \mathbf{C}_f(P)$ and there exists at least one set of zero measure in any refinement of $\mathbf{C}_f(P)$. In case Ω is a closed subset of a Euclidean space and $\mathcal{F}$ is a Borel algebra, it is easy to see that finest partitions do not exist if measure P has a continuous support or has a component with continuous support. It is also clear that if the measure P has discrete support there exist many partitions of Ω that are finest for P.

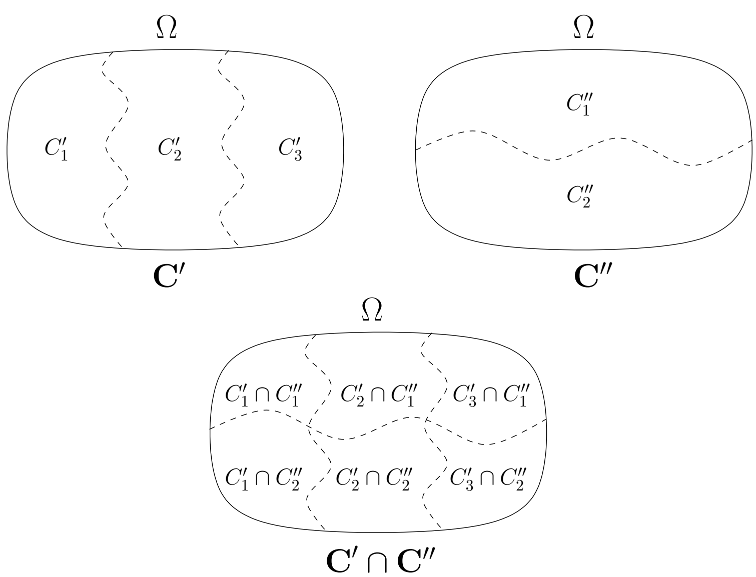

Figure 1.1: *Two partitions of* Ω *and the corresponding joint partition.*

Let $\mathbf{C}' = \{C_1', \ldots, C_r'\}$ and $\mathbf{C}'' = \{C_1'', \ldots, C_s''\}$ be two partitions of Ω. Then partition

$\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ is defined as the partition that consists of all sets of the form $C'_i \cap C''_j$: $\mathbf{C}' \cap \mathbf{C}'' = \{C'_1 \cap C''_1, C'_1 \cap C''_2, \ldots, C'_r \cap C''_s\}$ (see Fig. 1.1 for an illustration). Obviously, some of the sets constituting partition $\mathbf{C}' \cap \mathbf{C}''$ may be empty. Clearly, partition $\mathbf{C}' \cap \mathbf{C}''$ is a refinement of both $\mathbf{C}'$ and $\mathbf{C}''$.

If D is a subset of Ω and $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ is a partition of Ω, the partition $\mathbf{C}'_D = \{D \cap C'_1, \ldots, D \cap C'_r\}$ of D will be called the partition of D *induced* by the the partition $\mathbf{C}'$ of Ω (see Fig. 1.2).

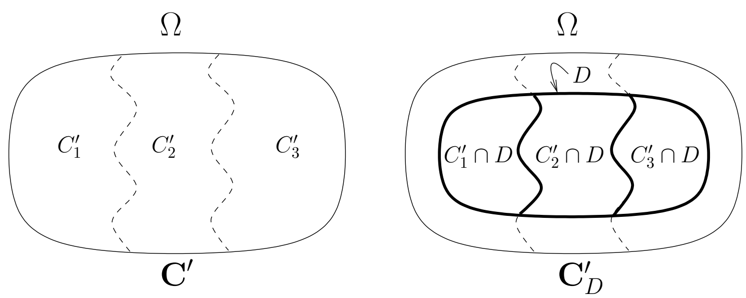

Figure 1.2: *Partition $\mathbf{C}'_D$ of set $D \subset \Omega$ induced by a partition $\mathbf{C}'$ of Ω.*

Besides standard partitions of Ω, we will also need *incomplete* partitions $\mathbf{C} = \{C_1, \ldots, C_r\}$ such that $\cup_{i=1}^r C_i \neq \Omega$. For any partition $\mathbf{C}$, we will use the notation $\hat{C} \equiv \cup_{i=1}^r C_i$. Clearly, partition $\mathbf{C}$ is complete if and only if $\hat{C} = \Omega$.

Let now $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ and $\mathbf{C}'' = \{C''_1, \ldots, C''_s\}$ be two incomplete partitions of Ω that are completely disjoint, i.e. such that $\hat{C}' \cap \hat{C}'' = \emptyset$. Then the partition $\mathbf{C} = \mathbf{C}' \cup \mathbf{C}''$ is defined as partition consisting of all subsets in the constituent partitions: $\mathbf{C} = \{C'_1, \ldots, C'_r, C''_1, \ldots, C''_s\}$. Clearly, partition $\mathbf{C}' \cup \mathbf{C}''$ may be complete or incomplete (it would be complete if and only if $\hat{C}' \cup \hat{C}'' = \Omega$). In case $\hat{C}' \cap \hat{C}'' \neq \emptyset$, the partition $\mathbf{C}' \cup \mathbf{C}''$ is not defined.

For an arbitrary complete partition $\mathbf{C} = \{C_1, \ldots, C_r\}$, it is straightforward to show that the following decomposition of the measure P into the corresponding conditional measures is valid.

$$P = \sum_{j=1}^{r} P(C_j) P_{C_j}. \tag{1.5}$$

1.5 Outline

Chapter 2 is devoted to a discussion of the question difficulty function, but starts with a discussion of the overall information exchange framework in section 2.2. In particular, the main theorem establishing the overall shape of the question difficulty function that is required to satisfy certain reasonable postulates is proved in section 2.3. Additionally, relationships between different questions are explored in section 2.4. Section 2.5 contains simple numerical examples illustrating the results obtained in the chapter. Finally, a conclusion summarizing the main results is given.

In Chapter 3, the overall form of the answer depth function is derived from a set of plausible postulates. Section 3.3 describes the main relationship between question difficulty and answer depth for main types of possible questions. In section 3.4, a special class of answers – the quasi-perfect answers – is discussed. Section 3.5 is devoted to relationship between different questions and, in particular, the relative depth of an answer to one question with respect to another question is introduced. Section 3.6 contains simple numerical example illustrating concepts and results discussed earlier in the article. Finally, a short summary of main results.

Chapter 4, the concept of an information source model is introduced and section 4.2 proposes several simple models. Section 4.3 describes the process of estimating – assuming the overall *ideal gas* question difficulty model – the pseudo-temperature function defined on the parameter space. Section 4.4 presents some numerical examples and finally a brief summary of the results.

Chapter 5 relates the informational characteristics of a source developed in Part I to solution quality characteristics of the problem and formulates the problem of optimal information acquisition. In section 5.2, we study maps from the parameter space of the problem to its solution space and some of their properties that are needed for later developments. In section 5.3, we relate the loss of a decision making/optimization problem with uncertainty to the characteristics of questions and answers, establishing, in particular, the value of minimum loss achievable with the help of a given depth answer to a particular question. Section 5.4 presents an example illustrating the results obtained in the earlier sections. Finally, a brief conclusion.

Finally, in Chapter 6, approximate solution methods based on the method of probability metrics and its application to scenario reduction in stochastic optimization are developed. Section 6.2 develops the main theoretical framework for the use of scenario reduction methods for optimization of additional information acquisition. Section 6.3 develops specific algorithms for determining the efficient frontier and optimizing information acquisition. Section 6.4 provides an example illustrating the use of methods developed in previous sections. A conclusion summarizes the main results.

Part I

Information Exchange

Chapter 2

Question Difficulty

2.1 Introduction

The problem of optimal decision making in environments characterized with both uncertainty and presence of information sources is considered in a general setting. This motivates searching for quantitative measures of question difficulty that would allow for maximizing the effect of additional information the information sources are capable of supplying. In this chapter, the concept of question difficulty for questions identified with partitions of problem parameter space is introduced and the overall form of question difficulty function is derived that satisfies a particular system of reasonable postulates. It is found that the resulting difficulty function depends on a single scalar function on the parameter space that can be interpreted – using parallels with classical thermodynamics – as a temperature-like quantity, with the question difficulty being similar to thermal energy. Quantitative relationships between different questions are also explored.

2.2 Overall Framework: Main Ingredients

The main components of the information exchange framework developed here are information sources, decision maker's questions and corresponding source's answers. Below, we discuss them in turn with some emphasis on questions which are the main subject of the current chapter.

2.2.1 Information Source

Assume that a source of additional information is available that is capable of answering specific questions concerning input data for problem (1.1). This implies that the source's answers are capable of modifying the initial measure P on Ω. The overall idea that we would like to formalize can be summarized as a set of – loosely formulated at this point – reasonable assumptions.

- The source has a finite capacity (appropriately defined).
- Questions that can be given to the source have, in general, different degrees of detalization (elaborateness) and difficulty.
- A question's degree of difficulty is related to the question degree of detalization but in general does not coincide with it.
- The quality of source's answers is directly related to the degree of difficulty of the corresponding questions.
- The source "tries equally hard" to answer any question it receives. The result is that it answers questions well (with low error probabilities) if the question difficulty does not exceed its capacity and the quality of its answers progressively degrades as the difficulty exceeds the source's capacity.

2.2.2 Questions

A question is a request for new information on top of what is already known. The latter is represented by the measure P on the parameter space Ω and is assumed to be common knowledge. Since, in the context discussed, any information is represented by some measure on Ω – with a measure concentrated on a single element of Ωcorresponding to a state of full knowledge – a question can be associated to a specific request for an updated measure on Ω. Therefore we identify a question with a (possibly incomplete) *partition* of Ω.

Definition: A question is a partition $\mathbf{C} = \{C_1, C_2, \ldots, C_r\}$ where C_j, $j = 1, \ldots, r$ are subsets of Ω such that $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\cup_{j=1}^{r} C_j \subseteq \Omega$.

Note that we allow for incomplete partitions for which $\cup_{j=1}^{r} C_j$ is a proper subset of Ω. For any partition $\mathbf{C}$, we denote the union of all subsets in $\mathbf{C}$ by $\hat{C}$:$\hat{C} \equiv \cup_{j=1}^{r} C_j$. Thus for any complete partition $\mathbf{C}$, $\hat{C} = \Omega$.

In everyday terms, a complete partition can be interpreted as a *multiple-choice* question (e.g. *"Is this apple red, green or yellow?"*. An incomplete partition consisting of a single subset can be associated with a *free-response* question, e.g. *"What color is this apple?"* Incomplete partitions consisting of several subsets can be interpreted as combinations of these two kinds – as *mixed* questions, e.g. *"What fruit is it and is it red, green or yellow?"*. In the given more narrow context – when the parameter space Ω and measure P on it are precisely known to the information source – the interpretation of an incomplete partition as a free-response (or mixed) question is not quite correct since if the source is presented with a description of a subset C of Ω the question becomes implicitly multiple-choice: *"Is the random outcome ω in C or not?"*. In order to accurately model real free-response questions (as they are usually understood), more complicated models are likely needed. From the more narrow point of view adapted here, incomplete questions are best thought of as an auxiliary construction that helps in determining the difficulty of complete questions that can be unambiguously defined and interpreted.

In the following, we will use the terms "partition" and "question" interchangeably. We will also use terms "multiple-choice question", "free-response question" and "mixed question" to mean complete partition, incomplete partition consisting of a single set and incomplete partitions consisting of more than one set, respectively.

Given a question $\mathbf{C}$, we are interested in quantifying its degree of difficulty, i.e. finding, for the given parameter space Ω and measure P on Ω, a function G: $\mathbf{C} \to \mathbb{R}$ that assigns larger values to more difficult questions. We use the notation $G(\Omega, \mathbf{C}, P)$ to emphasize the dependence of the question difficulty on Ω and P.

The particular shape of the function $G(\Omega, \mathbf{C}, P)$ could conceivably range in a fairly broad domain and would have to be approximated and estimated using experimental data such as observed performance of a source on various questions. On the other hand, due to the very fact of possibly wide range of shapes of the question difficulty function it makes sense to try to limit that range somewhat by imposing reasonable restrictions on the properties of the difficulty function. Such imposed restrictions can naturally be termed *postulates*. Then the validity of such postulates can be tested by observing a source's performance (such as empirical error probabilities) on various question of this type.

2.2.3 Perfect Answers

While a detailed discussion of answers will be given in Chapter 3, here we introduce a concept of a *perfect answer* to a question $\mathbf{C}$ as an answer that provides an exhaustive reply to $\mathbf{C}$. Specifically, we have the following definition.

Definition: Given a question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the perfect answer $V^*(\mathbf{C})$ is a message that takes one of the values in the set $\{s_1, \ldots, s_r\}$ such that the measure $P^j \equiv P^{V^*(\mathbf{C})=s_j}$ updated by a reception of value s_j of $V^*(\mathbf{C})$ is equal to the conditional measure P_{C_j}.

Informally speaking, a perfect answer to $\mathbf{C}$ completely resolves the uncertainty associated with the partition $\mathbf{C}$, i.e. places a random outcome ω in one of the subsets in $\mathbf{C}$ with certainty but otherwise does no more (since the resulting measure on the subset C_j is the conditional measure P_{C_j}). One can say that a perfect answer is the most basic type of an answer to a given question. It is convenient to think of a question difficulty $G(\Omega, \mathbf{C}, P)$ as an amount of *pseudo-energy* (the term ‘motivated by certain parallels with thermodynamics) contained in $\mathbf{C}$. Then it is natural to require that the *depth* of a perfect answer $V^*(\mathbf{C})$ be equal to the difficulty of $\mathbf{C}$. In other words, the amount of pseudo-energy contained in a perfect answer to $\mathbf{C}$ is equal to that in $\mathbf{C}$.

If question $\mathbf{C}$ is complete and $V^*(\mathbf{C})$ is the corresponding perfect answer it is reasonable to assume that $V^*(\mathbf{C})$ does not change the original measure P on average, or, in other words, that the original measure P is a “valid” one that only gets refined by the answer to $\mathbf{C}$. Formally speaking, this assumption means that

$$\sum_{j=1}^{r} \Pr(V^*(\mathbf{C}) = s_j) P_{C_j} = P, \tag{2.1}$$

from which it follows – by evoking (1.5) – that $\Pr(V^*(\mathbf{C}) = s_j) = P(C_j)$. We will call (2.1) the *consistency* condition for answer $V^*(\mathbf{C})$.

For incomplete question, the condition (2.1) has to be modified to read

$$\sum_{j=1}^{r} \Pr(V^*(\mathbf{C}) = s_j) P_{C_j} = P_{\hat{C}}, \tag{2.2}$$

from which it follows that $\Pr(V^*(\mathbf{C}) = s_j) = \frac{P(C_j)}{P(\hat{C})}$.

2.3 Question Difficulty Function

Our goal in this section is to derive a general form of the question difficulty function $G(\Omega, \mathbf{C}, P)$ and – along the way – establish the set of parameters it can depend upon. In many fields of scientific inquiry, when faced with a new phenomenon, *linear* models are often explored first both because of their simplicity and because of their role as elementary building blocks for more complicated models. We will attempt to do same in our situation. Besides linearity, we will – somewhat implicitly – assume that the parameter space is *isotropic*, i.e. the pseudo-energy amount does not depend on the orientation of subsets in $\mathbf{C}$ in the parameter space. Later on, both of these basic assumption – linearity and isotropy – can be relaxed and more general models can be obtained.

As has been mentioned earlier, in the model adapted here, incomplete (free-response and mixed) questions are to be understood as auxiliary constructions, while complete (multiple-choice) questions have a clear meaning. For a free-response question $C \subset \Omega$, the difficulty function $G(\Omega, C, P)$ can be interpreted as *conditional* difficulty of any complete question $\mathbf{C}$ containing the subset C given that the random outcome ω is in C. For example, if the subset C_1 represents apple, C_2 – pear and C_3 - peach so that $C_1 \cup C_2 \cup C_3 = \Omega$, then $G(\Omega, C_1, P)$ can be interpreted as the difficulty of the question *"Is it an apple, a pear, or a peach?"*, or, equivalently *"What kind of fruit is it?"* (since the source knows that the possible types are apple, pear and peach), provided that an apple is shown to the source.

One reasonable and almost obvious requirement that can be imposed on the question difficulty function $G(\Omega, \mathbf{C}, P)$ is that of *certainty*, i.e. the difficulty of a question should vanish if there is no new knowledge to acquire given the original state of it. Formally speaking, $G(\Omega, \mathbf{C}, P) = 0$ whenever $P(C_j) = 1$ for some value of the index j. One can say that in this case the question is already answered at the time of its formulation. These are questions of the kind *"Is this red apple red, green or yellow?"* for complete (multiple-choice) questions or *"What color is this red apple?"* for incomplete (free-response) questions. Thus we obtain

Postulate Q1 (*Certainty*). Suppose $\mathbf{C} = \{C_1, \dots, C_r\}$ and $P(C_j) = 1$ for some value of j. Then $G(\Omega, \mathbf{C}, P) = 0$.

Note that Postulate Q1 implies neither linearity nor isotropy and should be included even if these two basic assumptions are relaxed.

The second postulate we propose is of the same universal variety. It simply requires that the question difficulty function be continuous in all its arguments (which are yet to be determined).

Postulate Q2 (*Continuity*). The function $G(\Omega, \mathbf{C}, P)$ is continuous in all its arguments.

Again, it seems to be reasonable to keep Postulate Q2 even if more general models are desired.

The next postulate states that, for questions that have both free-response and multiple-choice components, i.e. for questions that are incomplete but consist of several subsets, the difficulty is additive: the overall difficulty of the question is the sum of the difficulty of the free-response part and the difficulty of the multiple-choice part given the free-response part has been answered perfectly. Formally, we obtain the following.

Postulate Q3 (*Mixed question decomposition*). Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be such that $\hat{C} = \cup_{j=1}^{r} C_j \neq \Omega$. Then

$$G(\Omega, \mathbf{C}, P) = G(\Omega, \hat{C}, P) + G(\hat{C}, \mathbf{C}, P_{\hat{C}}).$$

This postulate describes the difficulty of questions of the sort *"What kind of fruit is it and is it red, green or yellow?"*. It states that the difficulty of the overall question is additive: it is equal to the sum of difficulties of two questions: *"What fruit is it?"* and *"Is this apple red, green or yellow?"* assuming the correct answer to the first question was *"Apple"*. This postulate may likely be changed or relaxed when more general models are considered.

The next postulate states the mean value property of incomplete questions: the difficulty of the question $\mathbf{C} \cup \mathbf{C}'$ obtained by taking the union of two incomplete non-overlapping partitions $\mathbf{C}$ and $\mathbf{C}'$ is equal to the arithmetic mean value of the difficulties of the constituents questions with respect to the original measure P.

Postulate Q4 (*Mean value*). Let $\mathbf{C}$ and $\mathbf{C}'$ be two incomplete questions such that $\hat{C} \cap \hat{C}' = \emptyset$. Then

$$G(\Omega, \mathbf{C} \cup \mathbf{C}', P) = \frac{P(\hat{C}) G(\Omega, \mathbf{C}, P) + P(\hat{C}') G(\Omega, \mathbf{C}', P)}{P(\hat{C} \cup \hat{C}')}.$$

This postulate can be interpreted as follows. Let $\mathbf{C}$ and $\mathbf{C}'$ each consist of a single subset: $\mathbf{C} = \{C\}$ and $\mathbf{C}' = \{C'\}$ for $C \subset \Omega$, $C' \subset \Omega$. Assume also that $C \cup C' = \Omega$, so that $\{C, C'\}$ is a complete question. Then the statement of Postulate Q4 would read

$$G(\Omega, \{C, C'\}, P) = P(C) G(\Omega, C, P) + P(C') G(\Omega, C', P), \tag{2.3}$$

which is consistent with the interpretation of the difficulty $G(\Omega, C, P)$ of a free-response question as difficulty of a multiple-choice question containing C as one of possible answers given that C is true (that is conditioned on $\omega \in C$). For instance, let C represent an apple and C' a pear and assume these are the only two possible types of fruit. Then expression (2.3) states that the difficulty of the question *"What kind of fruit is it?"* (which, given the structure of Ω and the measure P is equivalent to the question *"Is it an apple or a pear?"*) is equal to the difficulty of the same question in case an apple is shown times the probability that an apple can be shown plus the same expression for the pear. Thus $G(\Omega, \{C, C'\}, P)$ is the expected value of the multiple-choice question difficulty where the expectation is taken over possible correct answers. From this point of view, Postulate Q4 sounds rather natural and generic. However, the real meaning of Postulate Q4 is in that it states that the conditional difficulties are *independent of the number and measures of other options (subsets).* Postulate Q4 assigns the same conditional difficulty $G(\Omega, C, P)$ to the subset $C \subset \Omega$ *regardless of the complete partition it is a member of.* For instance, if $C \subset \Omega$ represents an apple then, in the case the source is shown an apple, the difficulty of the question *"Is it an apple or not?"* would be the same as that of *"What kind of a fruit is it?"* even if the number of possible choices (types of fruit) is large. It is easy to see that this, while not unreasonable, still is a rather strong assumption which may not be true for realistic information sources. Postulate Q4 can be thought of as an expression of linearity of the difficulty function and it can be fully expected that it will be relaxed or modified in more general models.

To state the next postulate we need to introduce a new concept. We say that the parameter space Ω is *homogeneous* if the question difficulty function depends only on its subset measures for any question $\mathbf{C}$ in Ω: $G(\Omega, \mathbf{C}, P) = f(P(\mathbf{C}))$ where $P(\mathbf{C})$ stands for the vector $(P(C_1), \ldots, P(C_r))$. More generally, we say that a subset $D \subseteq \Omega$ is *homogeneous* if $G(D, \mathbf{C}, P_D) = f(P_D(\mathbf{C}))$ as long as $\hat{C} \subseteq D$. In particular, any atom (minimal set) of the sigma-algebra $\mathcal{F}$ is homogeneous. Postulate 5 then states that a free-response question can be posed in stages without changing its overall difficulty as long as all the intermediate questions lie inside a homogeneous subset of the parameter space.

Postulate Q5 (*Homogeneous free-response sequentiality*). Let $D \subset \Omega$ be a homogeneous subset of the parameter space and let $\mathbf{C}$ be a question such that $\hat{C} \subset D$. Then

$$G(\Omega, \mathbf{C}, P) = G(\Omega, D, P) + G(D, \mathbf{C}, P_D).$$

To get a little more "feel" for this postulate think of a question asking to identify a certain animal species. The gradual approach to such a question would involve asking intermediate questions about the class the animal belongs to, order, suborder, superfamily, family, and, finally, the species itself. In case the original question is of "harder than average" variety it would be easier to answer the question in stages compared to answering it right away. On the other hand, if the original question is an easy one (easier than other similar questions) it can be easier to answer it without resorting to the intermediate "guiding" questions. A good example of the latter would be a question about a domestic cat that an average person would be able to answer easily and correctly whereas the "guiding" questions about class, order etc. would likely present some difficulty. Respectively, if all such questions are equally hard (for the same measure) then it would make sense to believe that the intermediate "guiding" questions would not change the difficulty of the original question just like the postulate states.

Finally, it certainly makes sense to require that if $D \subset \Omega$ is homogeneous and $C \subset D$ then $f(P_D(C)) = G(\Omega, C, P_D)$ should be a decreasing function of its argument $P_D(C)$. Indeed, a free-response question about something "rare" should be more difficult. We thus obtain Postulate 6.

Postulate Q6 (*Homogeneous free-response monotonicity*). Suppose $D \subset \Omega$ is homogeneous and $C \subset D$. Then $f(P_D(C)) = G(\Omega, C, P_D)$ is a decreasing function of its argument $P_D(C)$.

In order to get still more insight into the proposed set of postulates for the question difficulty function consider the following alternative postulate.

Postulate Q$3'$ (*Multiple-choice sequentiality*). Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a complete question and let $\tilde{\mathbf{C}}$ be its refinement. Then

$$G(\Omega, \tilde{\mathbf{C}}, P) = G(\Omega, \mathbf{C}, P) + \sum_{j=1}^{r} P(C_j) G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j}).$$

Postulate Q$3'$ states that if a multiple-choice question is made more detailed the difficulty of the resulting question can be obtained as a sum of the difficulty of the original question and the average (with respect to the measure P) of difficulties of conditional detalizations. For instance if the original question was *"Is it an apple or a pear?"* and the detalization sounds like *"Is it an apple or a pear and is its color red, green or yellow?"* then Postulate Q$3'$ says that the difficulty of the detailed question is equal to the difficulty of the original question plus the average of difficulties of questions *"Is this apple red, green or yellow?"* and the question *"Is this pear red, green or yellow?"*. This postulate may seem to be somewhat more reasonable and grounded in experience compared to, for instance, the *Mean value* postulate. It turns out though that Postulate Q$3'$ is implied by Postulate Q3 and Postulate Q4 as the following lemma shows.

Lemma 2.1 *Suppose Postulate Q3 and Postulate Q4 hold. Then Postulate Q$3'$ holds as well.*

Proof: Let $\tilde{\mathbf{C}}$ be a refinement of $\mathbf{C} = \{C_1, \ldots, C_r\}$. Then we can write

$$\begin{aligned}
G(\Omega, \tilde{\mathbf{C}}, P) &\overset{(a)}{=} \sum_{j=1}^{r} P(C_j) G(\Omega, \tilde{\mathbf{C}}_{C_j}, P) \\
&\overset{(b)}{=} \sum_{j=1}^{r} P(C_j)(G(\Omega, C_j, P) + G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j})) \\
&= \sum_{j=1}^{r} P(C_j) G(\Omega, C_j, P) + \sum_{j=1}^{r} P(C_j) G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j}) \\
&\overset{(c)}{=} G(\Omega, \mathbf{C}, P) + \sum_{j=1}^{r} P(C_j) G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j}),
\end{aligned}$$

where (a) follows from the Postulate Q4 since $\mathbf{C} = \cup_{j=1}^{r} \tilde{\mathbf{C}}_{C_j}$, (b) follows from Postulate Q3 and (c) follows from Postulate Q4. $\square$

Thus we see that Postulates Q3 and Q4 can be regarded as a somewhat stronger version of the *Multiple-choice sequentiality* property expressed by Postulate Q3$'$.

If we now demand that Postulates Q1 through Q6 hold for the question difficulty function $G(\Omega, \mathbf{C}, P)$ the question is what form this function can possibly take. The answer is given in the following theorem.

Theorem 2.1 *Let the function $G(\Omega, \mathbf{C}, P)$ where $\mathbf{C} = \{C_1, \ldots, C_r\}$ satisfy Postulates Q1 through Q6. Then it has the form*

$$G(\Omega, \mathbf{C}, P) = \frac{\sum_{j=1}^{r} u(C_j) P(C_j) \log \frac{1}{P(C_j)}}{\sum_{j=1}^{r} P(C_j)},$$

where $u(C_j) = \frac{\int_{C_j} u(\omega)\, dP(\omega)}{P(C_j)}$ and u: $\Omega \to \mathbb{R}$ is an integrable nonnegative function on the parameter space Ω.

Proof: We can assume, without loss of generality, that there exists a (complete) partition $\mathbf{D} = \{D_1, \ldots, D_N\}$ of Ω into homogeneous subsets D_j, $j = 1, \ldots, N$.

Let $D \subset \Omega$ be a homogeneous subset of the parameter space and let $C \subset D$ be a (free-response) question lying inside of D. Furthermore, let $C' \subset C$ be another question inside of C. Then, according to Postulate Q5,

$$G(\Omega, C, P) = G(\Omega, D, P) + G(D, C, P_D), \tag{2.4}$$

and, since C is homogeneous as well,

$$G(D, C', P_D) = G(D, C, P_D) + G(C, C', P_C). \tag{2.5}$$

Using the form of $G(\cdot)$ for homogeneous subsets, we obtain from (2.5)

$$f(P_D(C')) = f(P_D(C)) + f(P_D(C')/P_D(C)),$$

from which it follows, using standard additivity arguments, monotonicity and continuity of the function $f(\cdot)$ (which follow from Postulates Q6 and Q2, respectively) that $f(x) = -c \log x$ where $c > 0$ is a constant (see [52] for details). Since the constant c may depend on the particular homogeneous subset D we can denote it by $u(D)$ and obtain that

$$G(D, C, P_D) = -u(D) \log P_D(C), \tag{2.6}$$

for any $C \subseteq D$ whenever D is homogeneous.

Substituting (2.6) into (2.4) we can obtain

$$G(\Omega, C, P) = G(\Omega, D, P) + f(P(C)/P(D)) = G(\Omega, D, P) - u(D) \log \frac{P(C)}{P(D)},$$

or, equivalently,

$$G(\Omega, C, P) - G(\Omega, D, P) = -u(D) \log P(C) - u(D) \log P(D), \tag{2.7}$$

where C is an arbitrary subset of D. Then it follows from (2.7) and continuity of the function G (Postulate Q1) that

$$G(\Omega, C, P) = -u(D) \log P(C) + v(D), \tag{2.8}$$

for any $C \subseteq D$ whenever D is a homogeneous subset of Ω. Here $v(D)$ is an arbitrary function of D. Setting $P(C) = 1$ in (2.8) and making use of Postulate Q1, we obtain that $v(D) \equiv 0$ and therefore

$$G(\Omega, C, P) = -u(D) \log P(C). \tag{2.9}$$

Now let $\mathbf{D} = \{D_1, \ldots, D_N\}$ be a complete partition of Ω into homogeneous subsets D_j, $j = 1, \ldots, N$. Let $C \subset \Omega$ be a free-response question. Then $C = \cup_{j=1}^{N} C \cap D_j$, and since D_j is homogeneous and $C \cap D_j \subseteq D_j$, we obtain using (2.9) that

$$G(\Omega, C \cap D_j, P) = -u(D_j) \log P(C \cap D_j). \tag{2.10}$$

On the other hand, by Postulate Q3,

$$G(\Omega, C, P) = G(\Omega, \mathbf{D}_C, P) - G(C, \mathbf{D}_C, P_C), \tag{2.11}$$

where

$$G(\Omega, \mathbf{D}_C, P) = -\frac{1}{P_C}\sum_{j=1}^{N} u(D_j)P(C \cap D_j)\log P(C \cap D_j), \tag{2.12}$$

(using the identity $C = \cup_{j=1}^{N} C \cap D_j$, expression (2.10) and Postulate Q4), and analogously,

$$G(C, \mathbf{D}_C, P_C) = -\sum_{j=1}^{N} u(D_j)\frac{P(C \cap D_j)}{P(C)}\log\frac{P(C \cap D_j)}{P(C)}, \tag{2.13}$$

Substituting (2.12) and (2.13) into (2.11) we obtain

$$G(\Omega, C, P) = -\sum_{j=1}^{N}\frac{P(C \cap D_j)}{P(C)}u(D_j)\log P(C). \tag{2.14}$$

We can rewrite (2.14) as

$$G(\Omega, C, P) = -u(C)P(C)\log P(C), \tag{2.15}$$

where

$$u(C) \equiv \sum_{j=1}^{N}\frac{P(C \cap D_j)u(D_j)}{P(C)} \tag{2.16}$$

can be thought of as the definition of function u: $\mathcal{F} \to \mathbb{R}$ for inhomogeneous subsets of Ω. If we define the function $u(\omega)$ on Ω by

$$u(\omega) = \sum_{j=1}^{N} u(D_j)I_{D_j}(\omega),$$

where $I_D(\omega)$ is the indicator function of a subset $D \subseteq \Omega$, then the expression (2.16) can be written as

$$u(C) = \frac{\int_C u(\omega) dP(\omega)}{P(C)}. \tag{2.17}$$

Finally, if $\mathbf{C} = \{C_1, \ldots, C_r\}$ is an arbitrary question, we can use (2.15) and Postulate Q4 to obtain

$$G(\Omega, \mathbf{C}, P) = \frac{-\sum_{j=1}^{r} u(C_j) P(C_j) \log P(C_j)}{\sum_{j=1}^{r} P(C_j)},$$

where the "weights" $u(C_j)$ of the subsets C_j are given by (2.17). □

Theorem 2.1 establishes the general form of the question difficulty function if isotropy and linearity conditions are imposed. It appears that the Postulate Q4 (*Mean value*) is the most restricting one of all. It is also the one, as mentioned above, that imposes the linearity constraint on the question difficulty function. The result depends on the measure P and and integrable function u on the parameter space Ω that can be thought of as an attribute of the parameter space. Note that while the measure is extensive, i.e. the measure of a union of two disjoint subsets of Ω is the sum of individual measures ($P(C \cup C') = P(C) + P(C')$ if $C \cap C' = \emptyset$), the function u represents an intensive quantity in that it averages for a union of two disjoint subsets ($u(C \cup C') = \frac{P(C)u(C) + P(C')u(C')}{P(C) + P(C')}$). One can say, loosely speaking, that while measure is similar to volume, u is similar to temperature if physics analogies are to be used. In fact, Appendix A describes some insightful parallels between question difficulty on one hand and thermal energy (heat) on the other. These parallels suggest that the function $u(\cdot)$ can be thought of as temperature-like quantity that is allowed to be different at different points of the parameter space. In the following, we refer to the function $u(\omega)$ as *intensity* or *pseudo-temperature*. For the same reason, as mentioned earlier in the paper, it is convenient to think of question difficulty as the amount of *pseudo-energy* associated with the question.

It is also convenient to introduce the *entropy* of question $\mathbf{C}$ as

$$H(\Omega, \mathbf{C}, P) = \frac{\sum_{j=1}^{r} P(C_j) \log \frac{1}{P(C_j)}}{\sum_{j=1}^{r} P(C_j)}, \tag{2.18}$$

which differs from the pseudo-energy (difficulty) in that it does not involve the pseudo-temperature $u(\cdot)$. It is easy to see that, for any complete question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the expression (2.18) for question entropy coincides with Shannon entropy of the probability distribution $P(\mathbf{C}) = (P(C_1), \ldots, P(C_r))$ generated by partition $\mathbf{C}$ and measure P on Ω. Moreover, for any complete question $\mathbf{C}$, the pseudo-energy $G(\Omega, \mathbf{C}, P)$ is equal to the *weighted entropy* (studied in [26]) of the same distribution $P(\mathbf{C})$ with the corresponding weights given by the subset pseudo-temperature values $u(C_j)$, $j = 1, \ldots, r$.

It is also easy to see that for a free-response question $C \subset \Omega$, the relationship between pseudo-energy and entropy is simply

$$G(\Omega, C, P) = u(C) H(\Omega, C, P),$$

that is identical to the relationship that exists between thermal energy (heat) and entropy in thermodynamics for reversible processes.

A remark on units of pseudo-energy and pseudo-temperature seems to be in order. It is clear, since the expression for question difficulty is linear in $u(\cdot)$, multiplication of pseudo-temperature function $u(\cdot)$ by any (positive) overall constant would multiply the difficulty of any question by the same constant. A particular choice of this constant corresponds to the choice of units in which pseudo-temperature and pseudo-energy is measured. If just a single information source is considered this choice seems to be largely arbitrary. It appears to be convenient to adapt the convention in which the average pseudo-temperature of the parameter space Ω is equal to 1, i.e. to set the overall scale of $u(\cdot)$ by demanding that $\int_\Omega u(\omega)dP(\omega) = 1$. If two or more information sources need to be compared a different convention turns out to be useful. This issue is discussed further in Chapter 4 where information source models are considered.

2.4 Relationships Between Questions

In this section, we assume that all questions are complete (multiple-choice). If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary (complete) questions, the expression $\sum_{C' \in \mathbf{C}'} P(C')G(C', \mathbf{C}''_{C'}, P_{C'})$ will be denoted $G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P)$ and called the *conditional difficulty* of $\mathbf{C}''$. Using this notation, the sequentiality property expressed by Postulate Q3′ can be rewritten as

$$G(\Omega, \tilde{\mathbf{C}}, P) = G(\Omega, \mathbf{C}, P) + G(\Omega, \tilde{\mathbf{C}}_{\mathbf{C}}, P), \tag{2.19}$$

where $\tilde{\mathbf{C}}$ is an arbitrary refinement of $\mathbf{C}$.

If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary (complete) questions and $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ then obviously $\mathbf{C}$ is a refinement of both $\mathbf{C}'$ and $\mathbf{C}''$. One can then write the sequentiality property (2.19) as

$$G(\Omega, \mathbf{C}, P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}_{\mathbf{C}'}, P). \tag{2.20}$$

But it is easy to see that the partition induced by $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ on any set C' in $\mathbf{C}'$ is exactly the same as the partition induced on that set by $\mathbf{C}''$. Therefore, the term $G(\Omega, \mathbf{C}_{\mathbf{C}'}, P)$ in (2.20) can be equivalently written as $G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P)$ and we arrive at the *chain rule* for the question difficulty which we formulate as a lemma.

Lemma 2.2 *If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary complete questions and P is a measure on Ω then*

$$G(\Omega, \mathbf{C}' \cap \mathbf{C}'', P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P).$$

Again, let $\mathbf{C}'$ and $\mathbf{C}''$ be two (complete) questions on Ω and let $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ be the resulting combined question. Then the *pseudo-energy overlap* $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ between $\mathbf{C}'$ and $\mathbf{C}''$ can be defined as the difference between the sum of difficulties of $\mathbf{C}'$ and $\mathbf{C}''$ and that of the combined question $\mathbf{C}' \cap \mathbf{C}''$:

$$J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}' \cap \mathbf{C}'', P) \tag{2.21}$$

The definition (2.21) can be illustrated by a Venn diagram (see Fig. 2.1). Note that $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ is symmetric with respect to $\mathbf{C}'$ and $\mathbf{C}''$.

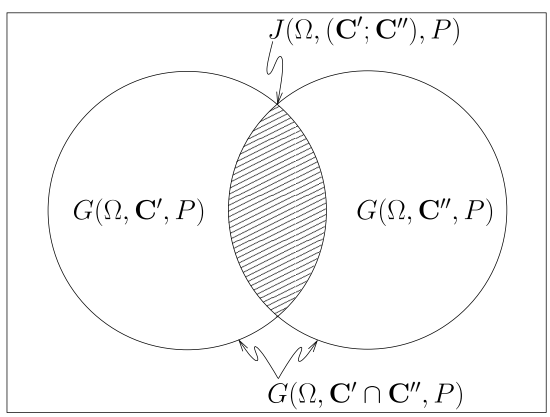

Figure 2.1: *Venn diagram for pseudo-energy overlap.*

One can make use of the sequentiality property of pseudo-energy to rewrite expression for the pseudo-energy overlap as follows.

$$\begin{aligned}
J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) &= G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, (\mathbf{C}', \mathbf{C}''), P) \\
&= G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}'_{\mathbf{C}''}, P) \\
&= G(\Omega, \mathbf{C}', P) - G(\Omega, \mathbf{C}'_{\mathbf{C}''}, P).
\end{aligned}$$

We formulate this result as a lemma.

Lemma 2.3 *If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary questions and P is a measure on Ω then the pseudo-energy overlap $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ can be found as*

$$J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = G(\Omega, \mathbf{C}', P) - G(\Omega, \mathbf{C}'_{\mathbf{C}''}, P).$$

Clearly, due to symmetry, the expression for the pseudo-energy overlap stated in Lemma 2.3 can be equivalently written as $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P)$.

If an expression for the pseudo-energy overlap as a function of the measure P and the pseudo-temperature function $u(\omega)$ is desired the definition (2.21) together with Theorem 2.1 can be used to obtain

$$J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = \sum_{i=1}^{r'} \sum_{j=1}^{r''} u(C'_i \cap C''_j) P(C'_i \cap C''_j) \log \frac{P(C'_i \cap C''_j)}{P(C'_i) P(C''_j)}. \tag{2.22}$$

We will be interested in exploring relationships between different questions: given two distinct questions, we would like to know to what degree they are similar to each other. More specifically, if a perfect answer to one question is available, how the difficulty of the other question is affected.

To answer this question, let $\mathbf{C}'$ and $\mathbf{C}''$ be two arbitrary complete questions on Ω and let $V^*(\mathbf{C}')$ be a perfect answer to $\mathbf{C}'$. We would like to find an expression for the conditional difficulty of $\mathbf{C}''$ given $V^*(\mathbf{C}')$. Clearly, since a reception of value s'_j of $V(\mathbf{C}')$ updates the measure P to $P_{C'_j}$, the difficulty of $\mathbf{C}''$ given $V(C') = s'_j$ is equal to

$$G(\Omega, \mathbf{C}'', P_{C'_j}) = G(C'_j, \mathbf{C}''_{C'_j}, P_{C'_j}), \tag{2.23}$$

since subsets of zero measure do not contribute to the difficulty function. Therefore the overall (expected) difficulty $G(\Omega, \mathbf{C}'', V^*(\mathbf{C}'))$ of question $\mathbf{C}''$ given a perfect answer $V^*(\mathbf{C}')$ to $\mathbf{C}'$ can be written as

$$\begin{aligned} G(\Omega, \mathbf{C}'', V^*(\mathbf{C}')) &= \sum_{j=1}^{r'} \Pr(V^*(\mathbf{C}') = s_j) G(\Omega, \mathbf{C}'', P_{C'_j}) \\ &\overset{(a)}{=} \sum_{j=1}^{r'} P(C'_j) G(C'_j, \mathbf{C}''_{C'_j}, P_{C'_j}) = G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P) \\ &\overset{(b)}{=} G(\Omega, \mathbf{C}'', P) - J(\Omega, (\mathbf{C}'; \mathbf{C}''), P), \end{aligned} \tag{2.24}$$

where (a) follows from (2.23) and the consistency condition (2.1) – which implies that $\Pr(V^*(\mathbf{C}') = s_j) = P(C_j)$; (b) follows from Lemma 2.3.

We see from (2.24) that the conditional difficulty of $\mathbf{C}''$ can be represented as a difference of the standard (unconditional) difficulty and the pseudo-energy overlap $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$. Thus the latter provides a measure of reduction of difficulty of a question that is due to a perfect knowledge of an answer to another question. Such a measure can naturally be termed *relative depth* of an answer $V(\mathbf{C}')$ (which in general may not be perfect) with respect to question $\mathbf{C}''$. We can formulate the result just obtained as a lemma.

Lemma 2.4 *The relative depth of a perfect answer $V^*(\mathbf{C}')$ to question $\mathbf{C}'$ with respect to question $\mathbf{C}''$ is equal to the pseudo-energy overlap between questions $\mathbf{C}'$ and $\mathbf{C}''$.*

The result of Lemma 2.4 has a clear intuitive interpretation: If two distinct questions are close, i.e. "almost about the same thing" then knowing a (perfect) answer to one of them nearly answers the other one – reduces the difficulty of it to a small value compared to the initial difficulty. The pseudo-energy overlap quantifies the notion of closeness for two arbitrary questions.

2.5 Examples

We consider an example with a finite parameter space first. Let Ω consist of 8 elements, corresponding to green, yellow and red apples (denoted GA, YA and RA, respectively), green, yellow and red pears (denoted GPr, YPr and RPr), and yellow and red peaches (denoted YPc and RPc). Let all elements be equiprobable so that $P(\cdot) = \frac{1}{8}$ for all $\omega \in \Omega$. The function $u(\omega)$ describes the relative difficulty of respective free-response questions. Let $u(GA) = u(GPr) = 1$ reflecting the observation that the green (cold) color is easier to tell from the both yellow and red (warm) colors on one hand, and an apple and a pear are also easy to distinguish from each other because of a different overall shape on the other hand. (Recall that there is no green peach that could be possibly confused with a green apple.) Let $u(YPr) = u(RPr) = 1.5$ reflecting the observation that yellow and red pears can be possibly confused with each other but not with anything else because of either their warm color (compared to green pears) or their distinct shape (compared to red or yellow apples or peaches). Finally, let $u(YA) = u(RA) = u(YPc) = u(RPc) = 2$ as these four combinations appear to be the hardest to distinguish from each other as they all possess a warm color and round shape. Normalizing the values of $u(\cdot)$ so that $\int_\Omega u(\omega)dP(\omega) = 1$ one obtains $u(GA) = u(GPr) =$

$\frac{8}{13}$, $u(YPr) = u(RPr) = \frac{12}{13}$ and $u(YA) = u(RA) = u(YPc) = u(RPc) = \frac{16}{13}$.

The difficulties of free-response questions corresponding to individual elements of Ω can be found as follows: $G(\Omega, GA, P) = G(\Omega, GPr, P) = \frac{8}{13} \cdot \log 8 = \frac{24}{13}$, $G(\Omega, YPr, P) = G(\Omega, RPr, P) = \frac{12}{13} \cdot \log 8 = \frac{36}{13}$ and $G(\Omega, YA, P) = G(\Omega, RA, P) = G(\Omega, YPc, P) = G(\Omega, RPc, P) = \frac{16}{13} \cdot \log 8 = \frac{48}{13}$. The difficulty of the exhaustive multiple choice question (that asks to determine the type and color of the fruit presented to the source) can be found as an expectation of the difficulties of all these free-response questions. Denoting the corresponding (finest) partition of Ω by $\mathbf{C}_f$ we obtain

$$G(\Omega, \mathbf{C}_f, P) = \sum_{\omega \in \Omega} P(\omega) G(\Omega, \omega, P) = 3.$$

Now let us consider difficulties of other multiple-choice questions. Let first of such questions be *"Is the fruit green or not?"*. Let $C_g = \{GA, GPr\} \subset \Omega$ be the subset consisting of all green fruit (apples and pears) and let $\overline{C}_g = \Omega \setminus C_g$ be the subset containing fruit of all other colors (red and yellow). The values $u(\cdot)$ for the sets in this partition are $u(C_g) = \frac{8}{13}$ and $u(\overline{C}_g) = \frac{1}{3} \cdot \frac{12}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{44}{39}$. The measures are $P(C_g) = \frac{1}{4}$ and $P(\overline{C}_g) = \frac{3}{4}$. Thus the difficulty of the question *"Is the fruit green or not?"* can be found as

$$G(\Omega, \{C_g, \overline{C}_g\}, P) = u(C_g) P(C_g) \log \frac{1}{P(C_g)} + u(\overline{C}_g) P(\overline{C}_g) \log \frac{1}{P(\overline{C}_g)} = 0.66$$

Consider another question with subset measures (and thus entropy) equal to those of $\{C_g, \overline{C}_g\}$. Let this question be *"Is the fruit a peach or not?"*. The corresponding partition is $\{C_{Pc}, \overline{C}_{Pc}\}$ where $C_{Pc} = \{YPc, RPc\}$ and $\overline{C}_{Pc} = \Omega \setminus C_{Pc}$. The values of function $u(\cdot)$ on these subsets are

$u(C_{Pc}) = \frac{16}{13}$ and $u(\overline{C}_{Pc}) = \frac{1}{3} \cdot \frac{8}{13} + \frac{1}{3} \cdot \frac{12}{13} + \frac{1}{3} \cdot \frac{16}{13} = \frac{12}{13}$. The measures are $P(C_{Pc}) = \frac{1}{4}$ and $P(\overline{C}_{Pc}) = \frac{3}{4}$. The difficulty of the question $\{C_{Pc}, \overline{C}_{Pc}\}$ is

$$G(\Omega, \{C_{Pc}, \overline{C}_{Pc}\}, P) = u(C_{Pc})P(C_{Pc}) \log \frac{1}{P(C_{Pc})} + u(\overline{C}_{Pc})P(\overline{C}_{Pc}) \log \frac{1}{P(\overline{C}_{Pc})} = 0.90$$

We see that this question is somewhat more difficult than the question on whether the fruit is green. The main reason for this difference is that to answer the question on whether the fruit is a peach one might need to have to distinguish a peach from an apple of similar (warm) color which is relatively difficult while answering the question on whether the fruit is green does not involve any "hard" decisions since the color itself is distinct.

Consider now the question *"What color is the given fruit?"* on one hand and *"What type is the given fruit?"* on the other. The former question can be represented as the partition $\mathbf{C}_c = \{C_g, C_y, C_r\}$ where $C_g = \{GA, GPr\}$, $C_y = \{YA, YPr, YPc\}$ and $C_r = \{RA, RPr, RPc\}$; the latter question can be identified with the partition $\mathbf{C}_t = \{C_A, C_{Pr}, C_{Pc}\}$ where $C_A = \{GA, YA, RA\}$, $C_{Pr} = \{GPr, YPr, RPr\}$ and $C_{Pc} = \{YPc, RPc\}$. The values of $u(\cdot)$ on these subsets are $u(C_g) = \frac{8}{13}$, $u(C_y) = \frac{1}{3} \cdot \frac{12}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{44}{39}$, $u(C_r) = u(C_y) = \frac{44}{39}$; $u(C_A) = \frac{1}{3} \cdot \frac{8}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{40}{39}$, $u(C_{Pr}) = \frac{1}{3} \cdot \frac{8}{13} + \frac{2}{3} \cdot \frac{12}{13} = \frac{32}{39}$, $u(C_{Pc}) = \frac{16}{13}$. The measures are $P(C_g) = \frac{1}{4}$, $P(C_y) = \frac{3}{8}$, $P(C_r) = \frac{3}{8}$; $P(C_A) = P(C_{Pr}) = \frac{3}{8}$, $P(C_{Pc}) = \frac{1}{4}$. Thus the difficulties of these two questions are

$$\begin{aligned} G(\Omega, \mathbf{C}_c, P) = u(C_g)P(C_g) \log \frac{1}{P(C_g)} + u(C_y)P(C_y) \log \frac{1}{P(C_y)} \\ + u(C_r)P(C_r) \log \frac{1}{P(C_r)} = \frac{11}{13} \log \frac{8}{3} + \frac{2}{13} \log 4 = 1.51, \end{aligned}$$

and

$$G(\Omega, \mathbf{C}_t, P) = u(C_A)P(C_A) \log \frac{1}{P(C_A)} + u(C_{Pr})P(C_{Pr}) \log \frac{1}{P(C_{Pr})}$$
$$+ u(C_{Pc})P(C_{Pc}) \log \frac{1}{P(C_{Pc})} = \frac{9}{13} \log \frac{8}{3} + \frac{4}{13} \log 4 = 1.60,$$

respectively.

The question about color turns out to be slightly easier than that about type. Qualitatively, the main reason for this difference is that the relatively rare event (that the fruit is green and that it is a peach, respectively) that gives a larger contribution to the difficulty because of the $\log \frac{1}{P(\cdot)}$ factor has smaller average value of pseudo-temperature $u(\cdot)$ in the case of the question about the fruit color.

The pseudo-energy overlap between the "color" and "type" questions can be calculated using the expression (2.22):

$$J(\Omega, (\mathbf{C}_c; \mathbf{C}_t), P) = \frac{6}{13} \log \frac{4}{3} + \frac{7}{13} \log \frac{8}{9} = 0.100,$$

indicating that while a perfect knowledge of the fruit color helps answering the question about its type, the reduction of difficulty of the "type" question due to the knowledge of color is relatively mild so the question about the fruit type remains almost as hard as it was before the color became known.

For an example with infinite parameter space, consider $\Omega = [0, 1]^2$ with uniform measure P (see Fig. 2.2 for an illustration). Let $u(\omega) = \frac{3}{2}(\omega_1^2 + \omega_2^2)$ where ω_1 and ω_2 are coordinates on Ω. Let us consider three different questions: $\mathbf{C}_i = \{C_i, \overline{C}_i\}$, where $C_1 = \{\omega : \omega_1 \in [\frac{1}{2}, 1], \omega_2 \in [\frac{1}{2}, 1]\}$,

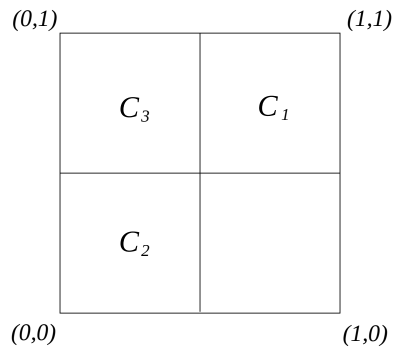

Figure 2.2: *The parameter space* $\Omega = [0,1]^2$ *and subsets* C_i*,* $i = 1, 2, 3$.

$C_2 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [0, \frac{1}{2}]\}$, $C_3 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [\frac{1}{2}, 1]\}$. It is easy to see that $P(C_i) = \frac{1}{4}$ for $i = 1, 2, 3$.

For question $\mathbf{C}_1$, we have $u(C_1) = \frac{3}{2}\int_{\frac{1}{2}}^{1} d\omega_1 \int_{\frac{1}{2}}^{1} d\omega_2(\omega_1^2 + \omega_2^2) = \frac{7}{4}$. Then, using the normalization condition $u(C_1)P(C_1) + u(\overline{C}_1)P(\overline{C}_1) = 1$, we can obtain $u(\overline{C}_1) = \frac{3}{4}$, which allows us to compute the difficulty:

$$G(\Omega, \{C_1, \overline{C}_1\}, P) = u(C_1)P(C_1)\log\frac{1}{P(C_1)} + u(\overline{C}_1)P(\overline{C}_1)\log\frac{1}{P(\overline{C}_1)}$$
$$= \frac{7}{16}\log 4 + \frac{9}{16}\log\frac{4}{3} = 1.108.$$

For question $\mathbf{C}_2$, we obtain $u(C_2) = \frac{3}{2}\int_0^{\frac{1}{2}} d\omega_1 \int_0^{\frac{1}{2}} d\omega_2(\omega_1^2 + \omega_2^2) = \frac{1}{4}$, and, making use of the normalization condition, $u(\overline{C}_2) = \frac{5}{4}$. The difficulty function value for this question becomes

$$G(\Omega, \{C_2, \overline{C}_2\}, P) = u(C_2)P(C_2)\log\frac{1}{P(C_2)} + u(\overline{C}_2)P(\overline{C}_2)\log\frac{1}{P(\overline{C}_2)} = \log\frac{4}{3} = 0.514.$$

Finally, for question $\mathbf{C}_3$, we have $u(C_3) = \frac{3}{2}\int_0^{\frac{1}{2}} d\omega_1 \int_{\frac{1}{2}}^{1} d\omega_2(\omega_1^2 + \omega_2^2) = 1$, and, obviously,

$u(\overline{C}_3) = 1$. The difficulty function is

$$G(\Omega, \{C_3, \overline{C}_3\}, P) = u(C_3)P(C_3)\log\frac{1}{P(C_3)} + u(\overline{C}_3)P(\overline{C}_3)\log\frac{1}{P(\overline{C}_3)}$$
$$= \frac{3}{4}\log\frac{4}{3} + \frac{1}{4}\log 4 = 0.811.$$

We see that, among these three questions $\mathbf{C}_1$ turns out to be the most difficult while difficulty of $\mathbf{C}_2$ is the smallest of the three. The reason is that $\mathbf{C}_1$ includes a small measure (rare) set in the region of high values of pseudo-temperature $u(\omega)$. On the other hand, the rare subset in $\mathbf{C}_2$ is located in the region of small values of $u(\omega)$. Question $\mathbf{C}_3$ is naturally placed between these two extremes: its rare subset is located in the region of moderate values of the field $u(\omega)$ so that the difficulty weight of this subset is equal to the average for the whole parameter space.

The overlaps between these questions can easily be computed using expression (2.22).

$$J(\Omega, (\mathbf{C}_1; \mathbf{C}_2), P) = \frac{1}{2}\log\frac{4}{3} + \frac{1}{2}\log\frac{8}{9} = 0.123,$$

$$J(\Omega, (\mathbf{C}_1; \mathbf{C}_3), P) = \frac{11}{16}\log\frac{4}{3} + \frac{5}{16}\log\frac{8}{9} = 0.232,$$

and

$$J(\Omega, (\mathbf{C}_2; \mathbf{C}_3), P) = \frac{5}{16}\log\frac{4}{3} + \frac{11}{16}\log\frac{8}{9} = 0.013,$$

showing that the most difficult questions – $\mathbf{C}_1$ and $\mathbf{C}_3$ – also exhibit the largest overlap which agrees with the common sense derived notion that knowledge of a perfect answer to a more difficult question can give more help in answering another question.

It is interesting to consider the limit in which the measure of the rare set approaches zero. For this purpose, let $C_1 = \{\omega : \omega_1 \in [1-a,1], \omega_2 \in [1-a,1]\}$, $C_2 = \{\omega : \omega_1 \in [0,a], \omega_2 \in [0,a]\}$ and $C_3 = \{\omega : \omega_1 \in [0,a], \omega_2 \in [1-a,1]\}$ and let $\mathbf{C}_i = \{C_i, \overline{C}_i\}$ for $i = 1,2,3$. Let $u(\omega) = \frac{n+1}{2}(\omega_1^n + \omega_2^n)$ where $n \geq 2$ is an integer and $\omega \in \Omega = [0,1]^2$. Then repeating the calculations for the previously considered example, taking the limit $a \to \infty$ and retaining only terms of the lowest order in a we obtain

$$G(\Omega, \{C_1, \overline{C}_1\}, P) \simeq (n+1)a^2 \log \frac{1}{a} + \log e \cdot a^2 \simeq (n+1)a^2 \log \frac{1}{a},$$

$$G(\Omega, \{C_2, \overline{C}_2\}, P) \simeq \log e \cdot a^2,$$

and

$$G(\Omega, \{C_3, \overline{C}_3\}, P) \simeq 2a^2 \log \frac{1}{a} + \log e \cdot a^2 \simeq 2a^2 \log \frac{1}{a}.$$

Again, we can see that the question $\mathbf{C}_1$ ends up being the most difficult one, with $\mathbf{C}_2$ being the least difficult. It's interesting to note that, to leading order in a, the difficulty of $\mathbf{C}_1$ and $\mathbf{C}_3$ behaves as $a^2 \log \frac{1}{a}$ (with only a numerical coefficient being different), while the difficulty of $\mathbf{C}_2$ behaves as a^2. A related observation is that, in this limit, the difficulty of both $\mathbf{C}_1$ and $\mathbf{C}_3$ is dominated by the rare subset while that of $\mathbf{C}_2$ is dominated by the larger subset with measure approaching 1 since the contribution of the rare subset is diminished by the low value of pseudo-temperature $u(\cdot)$ over that subset.

2.6 Conclusion

This chapter initiated development of a quantitative general framework for the description of the process of information extraction from information sources capable of providing answers to given questions. The main motivation for such a framework is the need for optimal decision making in situations characterized with incomplete information and availability of additional information sources. The framework is expected to be especially useful when the knowledge that the information sources possess is of a relatively “loose” variety, i.e. cannot be readily represented in a form admitting direct use in a mathematical formulation. A typical example of such a source would be a human expert who can express a preference for one of the two regions in the parameter space but would find it difficult to produce an accurate probability distribution over the parameter space.

The three main components of the proposed framework are questions, answers and information sources. The present chapter’s subject is questions and, in particular, question difficulty functions. The purpose of the latter is measuring the degree of accuracy the given source can achieve on various questions. The idea is that a source would answer easy questions well but its answers’ accuracy would decrease with increasing difficulty of questions. The overall form of the question difficulty function is in general determined by the constraints the difficulty function is required to satisfy. The latter constraints depend on the overall properties imposed on the difficulty function. Here, we assumed the question difficulty to be linear and isotropic on the parameter space. The resulting form was then derived from a system of postulates expressing the desired properties along with more general consistency requirements.

It turns out that the resulting question difficulty function depends on a single scalar quantity $u(\cdot)$ defined on the parameter space and can be interpreted – using parallels with thermodynamics – as an energy-like quantity while the function $u(\cdot)$ takes on the role of temperature that is allowed to take different values at different points of the parameter space. It is interesting to contrast the resulting difficulty function to the corresponding Shannon entropy that is a purely informational quantity measuring the minimum expected number of bits required to communicate a (perfect) answer to the question under consideration. Using parallels with thermodynamics, while the former is similar to thermal energy, the latter can be likened to entropy. It is also worth noting that the linear isotropic model – in thermodynamics terms – can be interpreted as that of *ideal gas*. We expect that other more involved (anisotropic, for instance) versions of question difficulty function would still allow useful interpretations in thermodynamics terms with the difficulty function being similar to thermal energy associated with an appropriate thermodynamical system.

Chapter 3

Answer Depth

3.1 Introduction

The difficulty of a decision maker's questions was considered in the previous chapter. Here, information sources' answers are investigated. In particular, the concept of answer depth is introduced that quantifies the amount of suitably defined effort required to provide an answer of a given accuracy. The overall form of the answer depth function is derived by demanding that it satisfy a particular set of postulates expressing, besides some reasonable consistency conditions, the linearity and isotropy properties. The latter properties justify calling the resulting information exchange model the "ideal gas model" making use of potentially fruitful parallels with classical thermodynamics.

3.1.1 Answers

Given a question $\mathbf{C}$, a source is capable of providing an answer. Since any information in this context can be represented by some measure on P, it is reasonable to think of an answer to question $\mathbf{C}$ as a message the reception of which implies certain changes in the initial measure P. In an extreme case, a message can change the original measure to a measure supported at a single element of Ω – this describes a complete resolution of the initial uncertainty and to the best possible answer to the corresponding (exhaustive) question.

Thus, given a question $\mathbf{C}$, it makes sense to define an answer $V(\mathbf{C})$ to it as a message that can take values in the set $\{s_1, s_2, \ldots, s_m\}$, where s_k, $k = 1, \ldots, m$ is some symbolic string the length of which does not play an important role in the present context. Then, the conditional measure $P^{V(\mathbf{C})=s_k} \equiv P^k$ is in general different from the original measure P following a reception of the value s_k of message $V(\mathbf{C})$. Additionally, care has to be taken to ensure that the answer $V(\mathbf{C})$ is indeed an answer to the specific question $\mathbf{C}$ and not some other question. To achieve this we can require that a reception of $V(\mathbf{C})$ leave the relative likelihood of the elements inside every subset in $\mathbf{C}$ unchanged. Therefore probability is only "redistributed" between the members of $\mathbf{C}$. This way, an answer can't provide more information than what was requested in the question. We arrive at the following definition.

Definition: An answer to the question $\mathbf{C} = \{C_1, \ldots, C_r\}$ is a message $V(\mathbf{C})$ that takes values in the set $\{s_1, s_2, \ldots, s_m\}$ and such that $P^k_{C_j} = P_{C_j}$ for all $k = 1, \ldots m$ and all $j = 1, \ldots, r$.

Following this definition, it is straightforward to show that for $V(\mathbf{C})$ to be an answer to a multiple-choice question $\mathbf{C}$, it is necessary and sufficient for the updated measures P^k, $k = 1, \ldots, m$, to take the form

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j}, \tag{3.1}$$

where p_{kj}, $k = 1, \ldots, m$, $j = 1, \ldots, r$ are nonnegative coefficients such that $\sum_{j=1}^{r} p_{kj} = 1$ for $k = 1, \ldots, m$.

For incomplete (free-response and mixed) questions, the expression (3.1) gets slightly modified to account for the set $\overline{\hat{C}} = \Omega \setminus \hat{C}$ and takes the form

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j} + \bar{p}_k P_{\overline{\hat{C}}}, \tag{3.2}$$

where $\sum_{j=1}^{r} p_{kj} + \bar{p}_k = 1$. For pure free-response questions, $r = 1$.

While the function $G(\Omega, \mathbf{C}, P)$ measures difficulty of questions, it would be desirable to develop a measure of the amount of difficulty in $\mathbf{C}$ that is resolved by the answer $V(\mathbf{C})$. As mentioned earlier, the question difficulty can be interpreted as the amount of *pseudo-energy* associated with the question. Therefore, it is natural to think that a perfect answer would contain an amount of pseudo-energy equal to the amount in the question it answers. Any other answer would contain somewhat less pseudo-energy, as long as it is an answer to $\mathbf{C}$ and not some other – possibly more difficult – question.

In the following we denote the amount of pseudo-energy contained in the answer $V(\mathbf{C})$ – the *depth* of $V(\mathbf{C})$ – by $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ to emphasize its dependence on Ω and the initial measure P.

3.2 Answer Depth Function

In this section, our goal is to derive the general form of the answer depth function by imposing certain plausible requirements it has to satisfy. These requirements that we call Postulates are similar to those stated in Postulates Q1 through Q6 for questions (see Chapter 2).

Information in $V(\mathbf{C})$ is conveyed by modifying the original measure P and it modifies P differently for each value of the message $V(\mathbf{C})$. Therefore, the depth function for the message $V(\mathbf{C})$ should be the weighted average of the conditional values of the depth:

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) Y(\Omega, \mathbf{C}, P, P^k), \tag{3.3}$$

where P^k is the measure modified by the reception of $V(\mathbf{C}) = s_k$ and $Y(\Omega, \mathbf{C}, P, P^k)$ is the conditional depth that depends on the modified measure P^k.

We now impose reasonable requirements on conditional depth functions $Y(\Omega, \mathbf{C}, P, P^k)$ which are formulated as postulates as before.

The first such requirement is that the conditional depth should vanish if the measure is not modified at all, i.e. if $P^k = P$. On the other hand, if the modified measure assigns larger probabilities to *all* subsets in $\mathbf{C}$ (which can happen only for incomplete – free-response and mixed – questions) then the conditional depth should be strictly positive. This is the content of Postulate A1.

Postulate A1 (*Correct direction*). Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be any question. Then $Y(\Omega, \mathbf{C}, P, P) = 0$ if $P^k(C_j) = P(C_j)$ for all $j = 1, \ldots r$ and $Y(\Omega, \mathbf{C}, P, P^k) > 0$ if $P^k(C_j) > P(C_j)$ for all $j = 1, \ldots r$.

The second part of the postulate says that, for a free-response question for instance, if upon reception of the value s_k of $V(\mathbf{C})$ the set C_k has a higher probability than before, then the value s_k has a positive amount of pseudo-energy. For example, if the original question was *"What kind of fruit is it?"* with *"Pear"* being the correct answer then in case the answer sounds like *"It looks a lot like a pear"* or *"It's either a pear or an apple"*, such an answer is assigned positive pseudo-energy as it moves "in the right direction" towards the correct answer.

The next postulate parallels Postulate Q2 for questions.

Postulate A2 (*Continuity*). The function $Y(\Omega, \mathbf{C}, P, P^k)$ is continuous in all parameters it may depend upon.

The next postulate follows from the notion that $V(\mathbf{C})$ is an answer to precisely the question $\mathbf{C}$ and therefore the depth of $V(\mathbf{C})$ cannot exceed the difficulty of $\mathbf{C}$. The property is easiest to formulate for free-response questions $C \subset \Omega$.

Postulate A3 (*Free-response complete answer*). Let C be a free-response question and suppose $P^k(C) = 1$. Then

$$Y(\Omega, C, P, P^k) = G(\Omega, C, P).$$

This postulate expresses a simple desideratum that an exhaustive correct answer to a question should convey exactly the amount of information requested by the question. For instance, if the question is *"What fruit is it?"* with *"Apple"* as a correct answer then the answer *"Apple"* should carry all the information the question was asking for.

The next three postulates parallel Postulates Q3 through Q5.

Postulate A4 (*Mixed answer decomposition*) Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a mixed question. Then

$$Y(\Omega, \mathbf{C}, P, P^k) = Y(\Omega, \hat{C}, P, P^k) + Y(\hat{C}, \mathbf{C}, P_{\hat{C}}, P^k_{\hat{C}}).$$

This postulate states that the amount of pseudo-energy contained in a particular answer (value of $V(\mathbf{C})$ to a mixed question $\mathbf{C}$) can be represented as a sum of two components: the pseudo-energy of the same answer to the (free-response) question $\hat{C}$ and that of the same answer assuming the free-response part has been answered correctly. For example, if the question $\mathbf{C}$ is *"What kind of fruit is it and is it red, green or yellow?"* then Postulate A4 says that the depth of any particular answer to $\mathbf{C}$ is equal to the sum of the depth of the same answer to the question *"What kind of fruit is it?"* and the depth of the same answer to the question *"Is this apple red, green o yellow?"* (assuming the fruit in question was indeed an apple).

Postulate A5 (*Mean value*). Let $\mathbf{C}$ and $\mathbf{C}'$ be two incomplete questions such that $\hat{C} \cap \hat{C}' = \emptyset$. Then

$$Y(\Omega, \mathbf{C} \cup \mathbf{C}', P, P^k) = \frac{P^k(\hat{C})Y(\Omega, \mathbf{C}, P, P^k) + P^k(\hat{C}')Y(\Omega, \mathbf{C}', P, P^k)}{P^k(\hat{C} \cup \hat{C}')}.$$

This postulate expresses the linearity property of the answer depth function, similarly to the analogous postulate for question difficulty. We expect that it will be modified (or dropped) when more general models of information exchange are considered.

Just as it was done for questions, we say that the subset D of Ω is *homogeneous* if the conditional depth function depends only on measures of partition $\mathbf{C}$ whenever $\hat{C} \subseteq D$, i.e. $Y(D, \mathbf{C}, P_D, P_D^k) = f(P_D(\mathbf{C}), P_D^k(\mathbf{C}))$. In particular, any atom (minimal set) of $\mathcal{F}$ is homogeneous. The next postulate concerns answers to free-response questions located inside homogeneous regions of the parameter space Ω.

Postulate A6 (*Homogeneous free-response sequentiality*). Let $D \subseteq \Omega$ be a homogeneous subset of the parameter space and let C be a free-response question such that $C \subseteq D$. Then

$$Y(\Omega, C, P, P^k) = Y(\Omega, D, P, P^k) + Y(D, C, P_D, P_D^k).$$

This postulate states that, whether a free-response question located inside a homogeneous region of the parameter space is answered in stages or right away, the overall effort required of the answerer (to achieve certain fixed accuracy) is the same. For example, let the question be *"What species does this animal belong to?"*. Instead of answering this question right away the source could answer a question about the order first, then suborder, then superfamily, family and only then about the actual species. In general, it is clear that the effort required to answer the original question right away could be more (i.e. if the animal is exotic) or less (i.e. if the animal is common like a domestic cat) than that required to answer the same question in stages to the same accuracy. Postulate A6 sates that the effort would be the same if all questions involved are located inside a homogeneous region. We expect that this postulate would be retained (perhaps in a modified form) when more general information exchange models are considered.

We can now state the main result about the possible shape of answer conditional depth function $Y(\Omega, \mathbf{C}, P, P^k)$. It is formulated as a theorem.

Theorem 3.1 *Let Postulates A1 through A6 hold. Then the conditional answer depth function $Y(\Omega, \mathbf{C}, P, P^k)$ has the following form*

$$Y(\Omega, \mathbf{C}, P, P^k) = \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)},$$

where $u(C_j) = \frac{\int_{C_j} u(\omega) dP^k(\omega)}{P^k(C_j)}$ and the integrable function u: $\Omega \to \mathbb{R}$ is the same that is used in *characterizing the question difficulty function $G(\cdot)$.*

Proof: The proof is similar to that of main theorem in sections 2.3. We can assume without loss of generality that there exists a complete partition $\mathbf{D} = \{D_1, \ldots, D_N\}$ of Ω such that every subset in $\mathbf{D}$ is homogeneous.

Let D be a homogeneous subset of Ω and let $C' \subset C \subset D$ be two subsets of D. Then, by Postulate A6,

$$Y(\Omega, C, P, P^k) = Y(\Omega, D, P, P^k) + Y(D, C, P_D, P_D^k), \tag{3.4}$$

and

$$Y(D, C', P_D, P_D^k) = Y(D, C, P_D, P_D^k) + Y(C, C', P_C, P_C^k). \tag{3.5}$$

Since D is homogeneous it follows from (3.5) that

$$f(P_D(C'), P_D^k(C')) = f(P_D(C), P_D^k(C)) + f(P_D(C')/P_D(C), P_D^k(C')/P_D^k(C)).$$

Then standard arguments using Postulates A1 and A2 (see [52] for details) lead to the conclusion that the function $f(\cdot)$ has the form

$$f(p, q) = c \log \frac{q}{p},$$

where c is a positive constant. Going back to the function Y we obtain

$$Y(D, C, P_D, P_D^k) = u'(D) \log \frac{P_D^k(C)}{P_D(C)}, \tag{3.6}$$

where $u'(D) > 0$ is a constant that can possibly depend on the particular homogeneous subset D.

Substituting (3.6) into (3.4) we arrive at

$$\begin{aligned} Y(\Omega, C, P, P^k) - Y(\Omega, D, P, P^k) = Y(D, C, P_D, P_D^k) &= u'(D) \log \frac{P_D^k(C)}{P_D(C)} \\ &= u'(D) \log \frac{P^k(C)}{P(C)} - u'(D) \log \frac{P^k(D)}{P(D)}, \end{aligned}$$

from which it follows (using continuity of Y and the fact that the subset $C \subset D$ is arbitrary) that

$$Y(\Omega, C, P, P^k) = u'(D) \log \frac{P^k(C)}{P(C)} + v'(D),$$

for any $C \subset D$ whenever D is homogeneous. Here $v'(D)$ is another constant that can possibly depend on the homogeneous subset D. We can now use Postulate A3 to conclude that $u'(D) = u(D)$ for all homogeneous sets D and that $v(D') \equiv 0$. This leads to the following expression for the conditional depth function of a free-response answer lying inside a homogeneous subset:

$$Y(\Omega, C, P, P^k) = u(D) \log \frac{P^k(C)}{P(C)}. \tag{3.7}$$

Now let $\mathbf{D} = \{D_1, \ldots, D_N\}$ be a (complete) partition of Ω such that every subset in $\mathbf{D}$ is homogeneous. Let $C \subset \Omega$ be a free-response question. Using Postulate A5, we can write

$$Y(\Omega, \mathbf{D}_C, P, P^k) = \frac{\sum_{j=1}^N u(D_j) P^k(C \cap D_j) \log \frac{P^k(C \cap D_j)}{P(C \cap D_j)}}{P^k(C)}, \tag{3.8}$$

and

$$Y(C, \mathbf{D}_C, P_C, P_C^k) = \sum_{j=1}^N u(D_j) \frac{P^k(C \cap D_j)}{P^k(C)} \log \frac{P^k(C \cap D_j)/P^k(C)}{P(C \cap D_j)/P(C)}. \tag{3.9}$$

An application of Postulate A4 now yields

$$
\begin{aligned}
Y(\Omega, C, P, P^k) &= Y(\Omega, \mathbf{D}_C, P, P^k) - Y(C, \mathbf{D}_C, P_C, P_C^k) \\
&= \sum_{j=1}^{N} u(D_j) \frac{P^k(C \cap D_j)}{P^k(C)} \log \frac{P^k(C)}{P(C)} = u(C) \log \frac{P^k(C)}{P(C)},
\end{aligned}
$$

where

$$
u(C) \equiv \sum_{j=1}^{N} \frac{P^k(C \cap D_j) u(D_j)}{P^k(C)} = \frac{\int_C u(\omega) dP^k(\omega)}{P^k(C)}. \tag{3.10}
$$

Here, the function u: $\Omega \to \mathbb{R}$ is defined as

$$
u(\omega) = \sum_{j=1}^{N} u(D_j) I_{D_j}(\omega),
$$

and therefore is the same exact function that was used to describe the question difficulty function G.

Finally, let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be an arbitrary question on Ω. An application of Postulate A5 yields

$$
Y(\Omega, \mathbf{C}, P, P^k) = \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)}, \tag{3.11}
$$

where $u(C_j)$ is given by (3.10). $\square$

Having found the expression for conditional depth function we can now use it to obtain the unconditional (expected) answer depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$. We formulate the result as a corollary.

Corollary 3.1 *The answer depth function $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ has the form*

$$
Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)},
$$

where $P^k \equiv P^{V(\mathbf{C})=s_k}$ is the measure on Ω conditioned on reception of $V(\mathbf{C}) = s_k$ and $u(C_j)$ is as defined in Theorem 3.1.

In the following, we will often use the notation $\Pr(V(\mathbf{C}) = s_k) \equiv v_k$ for the sake of brevity.

3.3 Relationship Between Difficulty and Depth

Theorems 2.1 and 3.1 (together with Corollary 3.1) establish the overall form that question difficulty and answer depth, respectively, can take. The conditional depth function $Y(\Omega, \mathbf{C}, P, P^k)$ depends, besides the original measure P, on the updated measure $P^k \equiv P^{V(\mathbf{C})=s_k}$.

3.3.1 Multiple-choice Questions

For multiple-choice (complete) questions, it makes sense to assume that the original measure P is a "valid" one in the sense that it does not change on average upon reception of the answer message $V(\mathbf{C})$. More formally speaking, for any question $\mathbf{C} = \{C_1, \ldots, C_r\}$,

$$P = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) P^k, \tag{3.12}$$

The expression (3.12) can be thought of as a condition of consistency of the answer message $V(\mathbf{C})$ with the original measure P and can be used for determining probabilities $v_k \equiv \Pr(V(\mathbf{C}) = s_k)$ of various values of the answer message $V(\mathbf{C})$.

Note that, taking into account the form (3.1) of the updated measures P^k the consistency condition (3.12) can be written as

$$\sum_{k=1}^{m} v_k p_{kj} = P(C_j), \quad j = 1, \ldots, r. \tag{3.13}$$

Let us assume the consistency condition (3.12) holds and consider the answer depth function given by Corollary 3.1. Since for a multiple-choice question $\sum_{j=1}^{r} P^k(C_j) = 1$, we can write

$$\begin{aligned}
Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) &= \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)} \\
&= \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log P^k(C_j) - \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log P(C_j) \\
&\overset{(a)}{=} \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log P^k(C_j) + G(\Omega, \mathbf{C}, P) \overset{(b)}{\le} G(\Omega, \mathbf{C}, P),
\end{aligned}$$

where (a) follows from (3.12) and Theorem 2.1, and (b) follows from the inequality $\log P^k(C_j) \le 0$. It is also clear that the inequality (b) becomes an equality if and only if, for every value s_k of the answer message, either $P^k(C_j) = 0$ or $\log P^k(C_j) = 0$ for every value of the index j. For the latter to be true it is necessary and sufficient that, for all values of k,

$$P^k(C_j) = \delta_{f(k),j}, \tag{3.14}$$

where f: $\{1, 2, \ldots, m\} \to \{1, 2, \ldots, r\}$ is a map from the set of possible values of index k to that of index j. Substituting (3.14) into (3.12) we obtain

$$P(C_j) = \sum_{k=1}^{m} v_k \delta_{f(k),j} = \sum_{k: f(k)=j} v_k. \tag{3.15}$$

It is easy to see that without loss of generality one can define an equivalent message $V'(\mathbf{C})$ such that $V'(\mathbf{C}) = s_j$ whenever $V(\mathbf{C}) = s_k$ such that $f(k) = j$. Then (3.15) becomes simply

$$P(C_j) = \Pr(V'(\mathbf{C}) = s_j). \tag{3.16}$$

A *perfect* answer to a multiple-choice question is defined $\mathbf{C} = \{C_1, \dots, C_r\}$ as the message $V(\mathbf{C}) = \{s_1, \dots, s_r\}$ such that $P^k(C_j) = \delta_{k,j}$, and, as a consequence, $\Pr(V(\mathbf{C}) = s_j) = P(C_j)$. Then we can state the result obtained above as a lemma.

Lemma 3.1 *Let $\mathbf{C}$ be a multiple-choice question and assume the condition (3.12) for any answer $V(\mathbf{C})$ holds. Then $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ with the inequality being tight if and only if the answer $V(\mathbf{C})$ is perfect (up to trivial equivalences).*

3.3.2 Free-response Questions

Let $C \subset \Omega$ be a free-response question. We can write the depth function for a corresponding answer $V(C)$ as follows.

$$\begin{aligned} Y(\Omega, C, P, V(C)) &= \sum_{k=1}^{m} \Pr(V(C) = s_k) u(C) \log \frac{P^k(C)}{P(C)} \\ &= u(C) \sum_{k=1}^{m} \Pr(V(C) = s_k) \log P^k(C) - u(C) \log P(C) \sum_{k=1}^{m} \Pr(V(C) = s_k) \\ &= u(C) \sum_{k=1}^{m} \Pr(V(C) = s_k) \log P^k(C) + G(\Omega, C, P) \overset{(a)}{\leq} G(\Omega, C, P), \end{aligned}$$

where (a) follows from that the inequality $\log P^k(C) \leq 0$. It is straightforward to see that for the inequality (a) to become an equality it is necessary and sufficient that $P^k(C) = 1$ for all values k of the answer message. Clearly, in that case, we can define an equivalent message $V'(C)$ that takes a single value s so that $P^s(C) = 1$.

A *perfect* answer $V(C)$ to a free-response question C is defined to be a message taking a single value s such that $P^s(C) = 1$.

We can again state the result obtained above as a lemma.

Lemma 3.2 *Let C be a free-response question and $V(C)$ an answer to it. Then $Y(\Omega, C, P, V(C)) \leq G(\Omega, C, P)$ with the inequality being tight if and only if the answer $V(C)$ is perfect (up to trivial equivalences).*

3.3.3 Mixed Questions

Finally, let $\mathbf{C} = \{C_1, \ldots, C_r\}$ where $\hat{C} = \cup_{i=j}^{r} C_j \subset \Omega$ be a mixed question. We define a *perfect* answer $V(\mathbf{C})$ to a mixed question as a message taking values in the set $\{s_1, \ldots, s_r\}$ such that $P^j(C_j) = 1$ for $j = 1, \ldots, r$.

For any answer to a mixed question we demand that the following *consistency with the original knowledge* condition holds

$$\sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) P^k(C_j) = \gamma P(C_j), \tag{3.17}$$

where

$$\gamma = \frac{P^s(\hat{C})}{P(\hat{C})} = \frac{\sum_{j=1}^{r} P^k(C_j)}{\sum_{j=1}^{r} P(C_j)}, \tag{3.18}$$

for all values of k characterizes the free-response component of $V(\mathbf{C})$. Then it is straightforward to prove a result analogous to that of Lemmas 3.1 and 3.2.

Lemma 3.3 *If* $\mathbf{C}$ *is a mixed question and* $V(\mathbf{C})$ *is an answer to it such that condition (3.17) holds. Then* $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ *with the inequality becoming tight if and only if the answer* $V(\mathbf{C})$ *is perfect.*

Proof: We can write the depth function for $V(\mathbf{C})$ as follows.

$$
\begin{aligned}
Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) &= \sum_{k=1}^{m} v_k \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)} \\
&\stackrel{(a)}{=} \frac{\sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\gamma P(\hat{C})} \\
&= \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P^k(C_j) \\
&- \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P(C_j) \\
&\stackrel{(b)}{=} \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P^k(C_j) - \frac{1}{P(\hat{C})} \sum_{j=1}^{r} u(C_j) P(C_j) \log P(C_j) \\
&= \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P^k(C_j) + G(\Omega, \mathbf{C}, P) \stackrel{(c)}{\leq} G(\Omega, \mathbf{C}, P),
\end{aligned}
$$

where (a) follows from (3.18), (b) follows from (3.17) and (c) follows from the inequality $\log P^k(C_j) \leq 0$. Using the same arguments as those employed for the proof of Lemma 3.1 we arrive at the statement of this lemma. □

3.4 Quasi-perfect Answers to Complete Questions

Let the question $\mathbf{C} = \{C_1, \ldots, C_r\}$ be complete (multiple-choice) and let $V(\mathbf{C})$ be an answer to $\mathbf{C}$. If $V(\mathbf{C})$ is perfect, its depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ is equal to the difficulty $G(\Omega, \mathbf{C}, P)$ of $\mathbf{C}$ as Lemma 3.1 states. Here we would like to consider some simple classes of imperfect answers. To make the form of an imperfect answer more specific let us assume such an answer to resemble a perfect one in that the number of possible values it can take is equal to r and each message s_k, $k = 1, \dots, r$ expresses a degree of preference towards the subset C_k. Let e_k be the error probability associated with s_k, i.e. $e_k = P(\bar{C}_k)$, where $\bar{C}_k = \Omega \backslash C_k$. Let us also make the additional assumption that the error associated with s_k is "proportionally distributed" between the sets C_j $j \neq k$, i.e. $P^k(C_j) = \frac{e_k P(C_j)}{P(\bar{C}_k)} = \frac{e_k P(C_j)}{1-P(C_k)}$. Obviously, both of these assumptions can be stated in the following way

$$P^k = (1 - e_k) P_{C_k} + \sum_{j \neq k} \frac{e_k P(C_j)}{1 - P(C_k)} P_{C_j},$$

implying that the coefficients p_{kj} in (3.1) have the form

$$p_{kj} = \left(1 - \frac{e_k}{1 - P(C_k)}\right) \delta_{k,j} + \frac{e_k P(C_j)}{1 - P(C_k)}. \tag{3.19}$$

To further simplify the analysis and provide more concise description of errors associated with imperfect answers, we make a further assumption: that the error probability e_k constitutes the same fraction of $P(\bar{C}_k)$ for all values of k, i.e. $e_k = \alpha(1 - P(C_k))$, $k = 1, \dots, r$, where $0 \leq \alpha \leq 1$. Under this assumption, the error associated with the answer $V(\mathbf{C})$ that we will denote by $V_\alpha(\mathbf{C})$ is fully described by a single parameter α. The coefficients p_{kj} in (3.19) become

$$p_{kj} = (1 - \alpha)\delta_{k,j} + \alpha P(C_j), \tag{3.20}$$

and the updated measure P^k becomes simply

$$P^k = \alpha P + (1 - \alpha) P_{C_k}. \tag{3.21}$$

We see that for $\alpha = 0$ the measure P^k turns into the conditional measure P_{C_k} making the answer perfect, and for $\alpha = 1$ each measure P^k becomes the original measure P thus rendering the answer $V_\alpha(\mathbf{C})$ empty, i.e. possessing vanishing depth.

Substituting (3.21) into the general expression for the answer depth and using the fact that in this case $v_k = P(C_k)$, $k = 1, \dots, r$, we can obtain

$$\begin{aligned} Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})) = \sum_{k=1}^{r} u(C_k) P(C_k)(1 - \alpha + \alpha P(C_k)) \log \frac{1 - \alpha + \alpha P(C_k)}{P(C_k)} \\ + \alpha \log \alpha \sum_{k=1}^{r} u(C_k) P(C_k)(1 - P(C_k)), \end{aligned} \tag{3.22}$$

It is easy to see that the expression (3.22) becomes $G(\Omega, \mathbf{C}, P)$ for $\alpha = 0$ and vanishes for $\alpha = 1$.

In the following we will call answers characterized by updated measures of the form (3.21) and depth functions given by (3.22) the *quasi-perfect* answers. Their advantage is that they allow to smoothly interpolate between perfect and empty answers using just a single parameter α taking values on the interval $[0, 1]$.

Substituting (3.20) into the consistency condition (3.13) it is easy to see that for quasi-perfect answers

$$v_j = P(C_j), \tag{3.23}$$

for $j = 1, \dots, r$, regardless of the value of error probability α.

3.5 Relationships Between Questions

It was also shown in section 2.4 that the pseudo-energy overlap can be interpreted as the reduction of difficulty of question $\mathbf{C}''$ due to the knowledge of a perfect answer $V^*(\mathbf{C}')$ to question $\mathbf{C}'$,

$$G(\Omega, \mathbf{C}'', V^*(\mathbf{C}')) = G(\Omega, \mathbf{C}'', P) - J(\Omega, (\mathbf{C}'; \mathbf{C}''), P), \tag{3.24}$$

where the *conditional difficulty* $G(\Omega, \mathbf{C}'', V^*(\mathbf{C}'))$ is defined (for any answer $V^*(\mathbf{C}')$ to question $\mathbf{C}'$) as

$$G(\Omega, \mathbf{C}'', V^*(\mathbf{C}')) = \sum_{k=1}^{m'} \Pr(V^*(\mathbf{C}') = s_k) G(\Omega, \mathbf{C}'', P'^k). \tag{3.25}$$

It would be interesting to find out how the relation (3.24) generalizes for the case of an arbitrary answer to question $\mathbf{C}'$. Clearly, since a reception of value s'_k of $V(C')$ updates the measure P to P'^k, the difficulty of $\mathbf{C}''$ given $V(C') = s'_k$ is equal to

$$\begin{aligned} G(\Omega, \mathbf{C}'', P'^k) &= -\sum_{j=1}^{r''} u(C''_j) P'^k(C''_j) \log P'^k(C''_j) \\ &= -\sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P'^k(C''_j), \end{aligned}$$

and therefore the overall (expected) difficulty $G(\Omega, \mathbf{C}'', V(\mathbf{C}'))$ of question $\mathbf{C}''$ given an answer

$V(\mathbf{C}')$ to $\mathbf{C}'$ can be written – denoting $\Pr(V(\mathbf{C}') = s'_k)$ by v'_k – as

$$
\begin{aligned}
&G(\Omega, \mathbf{C}'', V(\mathbf{C}')) \equiv \sum_{k=1}^{m'} v'_k G(\Omega, \mathbf{C}'', P'^k) \\
&= -\sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P'^k(C''_j) \\
&= \sum_{k=1}^{m'} v'_k \left(\sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P'^k(C''_j) \right. \\
&\left. + \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P(C''_j) - \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P(C''_j) \right) \\
&= -\sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log \frac{P'^k(C''_j)}{P(C''_j)} \\
&- \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P(C''_j) \\
&= G(\Omega, \mathbf{C}'', P) - \sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log \frac{P'^k(C''_j)}{P(C''_j)}.
\end{aligned}
\tag{3.26}
$$

We see from (3.26) that the conditional difficulty of $\mathbf{C}''$ can be represented as a difference of the standard (unconditional) difficulty and another expression that can be appropriately denoted $Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}'))$ and called the *relative depth* of the answer $V(\mathbf{C}')$ with respect to question $\mathbf{C}''$:

$$
G(\Omega, \mathbf{C}'', V(\mathbf{C}')) = G(\Omega, \mathbf{C}'', P) - Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')), \tag{3.27}
$$

where the relative depth $Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}'))$ is given by

$$
Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')) = \sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log \frac{P'^k(C''_j)}{P(C''_j)}. \tag{3.28}
$$

Using the expression (3.1) for the updated measures P'^k we find that

$$P'^k(C'_l \cap C''_j) = p_{kl} \frac{P(C'_l \cap C''_j)}{P(C'_l)} \tag{3.29}$$

and

$$P'^k(C''_j) = \sum_{l=1}^{r'} p_{kl} \frac{P(C'_l \cap C''_j)}{P(C'_l)}, \tag{3.30}$$

and, substituting (3.29) and (3.30) into (3.28) we obtain for the relative depth:

$$Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')) = \sum_{k=1}^{m'} v'_k \sum_{l=1}^{r'} \sum_{j=1}^{r''} u(C'_l \cap C''_j) p'_{kl} \cdot \frac{P(C'_l \cap C''_j)}{P(C'_l)} \log \sum_{i=1}^{r'} p'_{ki} \cdot \frac{P(C'_i \cap C''_j)}{P(C'_i) \cdot P(C''_j)}. \tag{3.31}$$

We can summarize the result just obtained as a lemma.

Lemma 3.4 *Let $\mathbf{C}'$ and $\mathbf{C}''$ be two arbitrary complete questions on Ω and let $V(\mathbf{C}')$ be an answer to $\mathbf{C}'$. Then the conditional difficulty of $\mathbf{C}''$ given the answer $V(\mathbf{C}')$ can be found as*

$$G(\Omega, \mathbf{C}'', V(\mathbf{C}')) = G(\Omega, \mathbf{C}'', P) - Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')),$$

where the relative depth of $V(\mathbf{C}')$ is given by the expression (3.31).

Suppose now that $V^*(\mathbf{C}')$ is a perfect answer to $\mathbf{C}'$ which implies that $m' = r'$ and $p'_{kl} = \delta_{k,l}$. Substituting this into (3.31) and performing the sum over k while making use of the answer consistency condition (3.13) we obtain

$$Y(\Omega, \mathbf{C}'', P, V^*(\mathbf{C}')) = \sum_{l=1}^{r'} \sum_{j=1}^{r''} u(C'_l \cap C''_j) P(C'_l \cap C''_j) \log \frac{P(C'_l \cap C''_j)}{P(C'_l) P(C''_j)}, \tag{3.32}$$

which coincides with the expression (2.22) for the pseudo-energy overlap between questions $\mathbf{C}'$ and $\mathbf{C}''$. We thus recover the result (3.24).

Now let $V_\alpha(\mathbf{C}')$ be a quasi-perfect answer to question $\mathbf{C}'$ characterized by error probability α. Substituting expressions (3.20) and (3.23) into (3.31) we obtain, after some straightforward algebra

$$Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}')) = (1-\alpha)\sum_{l=1}^{r'}\sum_{j=1}^{r''} u(C'_l \cap C''_j)P(C'_l \cap C''_j)\log\left[(1-\alpha)\frac{P(C'_l \cap C''_j)}{P(C'_l)P(C''_j)} + \alpha\right]$$
$$+\alpha\sum_{l=1}^{r'}\sum_{j=1}^{r''} u(C'_l \cap C''_j)P(C'_l \cap C''_j)\sum_{k=1}^{r'} P(C'_k)\log\left[(1-\alpha)\frac{P(C'_k \cap C''_j)}{P(C'_k)P(C''_j)} + \alpha\right]. \tag{3.33}$$

It is easy to see that for $\alpha = 0$, expression (3.33) reduces to (3.32) which is the overlap between questions $\mathbf{C}'$ and $\mathbf{C}''$, and for $\mathbf{C}''$ coinciding with $\mathbf{C}'$ the relative depth (3.33) becomes the depth $Y(\Omega, \mathbf{C}', P, V_\alpha(\mathbf{C}'))$ (given by expression (3.22)) of quasi-perfect answer to $\mathbf{C}'$ characterized by the same value of error probability α. To see that it is sufficient to set $C''_j = C'_j$ (and hence $P(C'_l \cap C''_j) = \delta_{l,j}P(C'_l)$) in (3.33), one must make use of the (obvious) identity $\sum_{k\neq j} P(C'_k) = 1 - P(C'_j)$.

3.6 Examples

Let us revisit the example with a finite parameter space from Chapter 2. The parameter space Ω consists of 8 elements, corresponding to green, yellow and red apples (denoted GA, YA and RA, respectively), green, yellow and red pears (denoted GPr, YPr and RPr), and yellow and red peaches (denoted YPc and RPc). The elements are equiprobable so that $P(\cdot) = \frac{1}{8}$ for all $\omega \in \Omega$. The function $u(\omega)$ describes the relative difficulty of respective free-response questions. The observation that the green (cold) color is easier to distinguish from both the yellow and red (warm) colors is reflected in $u(GA) = u(GPr) = 1$. On the other hand, an apple and a pear are also easy to distinguish from each other because of a different overall shape. (Recall that there is no green peach that could possibly be confused with a green apple.) The observation that yellow and red pears can possibly be confused with each other but not with anything else because of either their warm color (compared to green pears) or their distinct shape (compared to red or yellow apples or peaches) is reflected in $u(YPr) = u(RPr) = 1.5$. Finally, $u(YA) = u(RA) = u(YPc) = u(RPc) = 2$ as these four combinations appear to be the hardest to distinguish from each other as they all possess a warm color and round shape. Normalizing the values of $u(\cdot)$ so that $\int_\Omega u(\omega)dP(\omega) = 1$ we obtain $u(GA) = u(GPr) = \frac{8}{13}$, $u(YPr) = u(RPr) = \frac{12}{13}$ and $u(YA) = u(RA) = u(YPc) = u(RPc) = \frac{16}{13}$.

Consider, the question *"Is the fruit green or not?"*. Let $C_g = \{GA, GPr\} \subset \Omega$ be the subset consisting of all green fruit (apples and pears) and let $\overline{C}_g = \Omega \setminus C_g$ be the subset containing fruit of all other colors (red and yellow). The partition is $\mathbf{C}_g = \{C_g, \overline{C}_g\}$. The values $u(\cdot)$ for the sets in this partition are $u(C_g) = \frac{8}{13}$ and $u(\overline{C}_g) = \frac{1}{3} \cdot \frac{12}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{44}{39}$. The measures are $P(C_g) = \frac{1}{4}$ and $P(\overline{C}_g) = \frac{3}{4}$. The second similar question is *"Is the fruit a peach or not?"*. The corresponding partition is $\mathbf{C}_{Pc} = \{C_{Pc}, \overline{C}_{Pc}\}$ where $C_{Pc} = \{YPc, RPc\}$ and $\overline{C}_{Pc} = \Omega \setminus C_{Pc}$. The values of function $u(\cdot)$ on these subsets are $u(C_{Pc}) = \frac{16}{13}$ and $u(\overline{C}_{Pc}) = \frac{1}{3} \cdot \frac{8}{13} + \frac{1}{3} \cdot \frac{12}{13} + \frac{1}{3} \cdot \frac{16}{13} = \frac{12}{13}$. The measures are $P(C_{Pc}) = \frac{1}{4}$ and $P(\overline{C}_{Pc}) = \frac{3}{4}$. Let $V_\alpha(\mathbf{C}_g)$ and $V_\alpha(\mathbf{C}_{Pc})$ be the corresponding quasi-perfect answers. The depth functions of these answers can be computed using (3.22) as (see Fig. 3.1 for an illustration)

$$Y(\Omega, \mathbf{C}_g, P, V_\alpha(\mathbf{C}_g)) = \frac{2}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4 - 3\alpha) + \frac{11}{13}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4-\alpha}{3} + \frac{15}{52}\alpha\log\alpha,$$

and

$$Y(\Omega, \mathbf{C}_{Pc}, P, V_\alpha(\mathbf{C}_{Pc})) = \frac{4}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{9}{13}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4-\alpha}{3} + \frac{21}{52}\alpha\log\alpha.$$

Consider the question *"What color is the given fruit?"* on one hand and *"What type is the given fruit?"* on the other. The former question can be represented as the partition $\mathbf{C}_c = \{C_g, C_y, C_r\}$ where $C_g = \{GA, GPr\}$, $C_y = \{YA, YPr, YPc\}$ and $C_r = \{RA, RPr, RPc\}$; the latter question can be identified with the partition $\mathbf{C}_t = \{C_A, C_{Pr}, C_{Pc}\}$ where $C_A = \{GA, YA, RA\}$, $C_{Pr} = \{GPr, YPr, RPr\}$ and $C_{Pc} = \{YPc, RPc\}$. The values of $u(\cdot)$ on these subsets are $u(C_g) = \frac{8}{13}$, $u(C_y) = \frac{1}{3}\cdot\frac{12}{13} + \frac{2}{3}\cdot\frac{16}{13} = \frac{44}{39}$, $u(C_g) = u(C_y) = \frac{44}{39}$; $u(C_A) = \frac{1}{3}\cdot\frac{8}{13} + \frac{2}{3}\cdot\frac{16}{13} = \frac{40}{39}$, $u(C_{Pr}) = \frac{1}{3}\cdot\frac{8}{13} + \frac{2}{3}\cdot\frac{12}{13} = \frac{32}{39}$, $u(C_{Pc}) = \frac{16}{13}$. The measures are $P(C_g) = \frac{1}{4}$, $P(C_y) = \frac{3}{8}$, $P(C_r) = \frac{3}{8}$; $P(C_A) = P(C_{Pr}) = \frac{3}{8}$, $P(C_{Pc}) = \frac{1}{4}$. Let $V_\alpha(\mathbf{C}_c)$ and $V_\alpha(\mathbf{C}_t)$ be quasi-perfect answers to questions $\mathbf{C}_c$ and $\mathbf{C}_t$. The depth of these answers can be found using the expression (3.22). The results are (see Fig. 3.1)

$$Y(\Omega, \mathbf{C}_c, P, V_\alpha(\mathbf{C}_c)) = \frac{2}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{11}{13}\left(1 - \frac{5}{8}\alpha\right)\log\frac{8-5\alpha}{3} + \frac{67}{104}\alpha\log\alpha,$$

and

$$Y(\Omega, \mathbf{C}_t, P, V_\alpha(\mathbf{C}_t)) = \frac{4}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{9}{13}\left(1 - \frac{5}{8}\alpha\right)\log\frac{8-5\alpha}{3} + \frac{69}{104}\alpha\log\alpha.$$

Let us consider the second example from section 2.5. The parameter space is $\Omega = [0,1]^2 \subset \mathbb{R}^2$. Let the pseudo-temperature function be $u(\omega) = \frac{3}{2}(\omega_1^2 + \omega_2^2)$ (so that the hard questions are located towards the upper-right corner of Ω). Consider the following three subsets of Ω: $C_1 = \{\omega : \omega_1 \in$

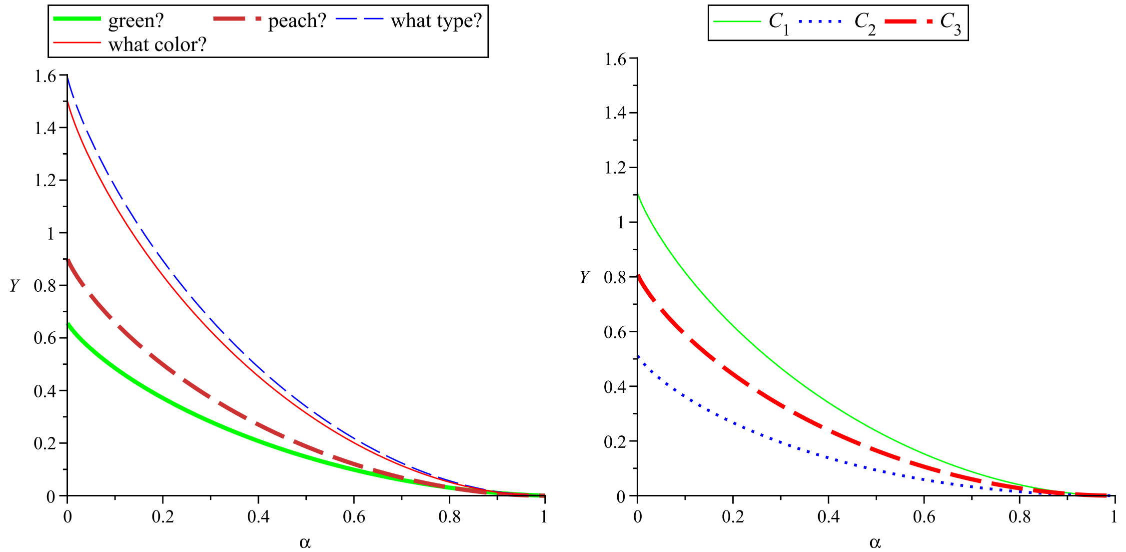

Figure 3.1: *Answer depth as function of α for quasi-perfect answers to questions on the finite parameter space (left) and infinite parameter space (right).*

$[\frac{1}{2}, 1], \omega_2 \in [\frac{1}{2}, 1]\}$, $C_2 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [0, \frac{1}{2}]\}$, $C_3 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [\frac{1}{2}, 1]\}$ and let $\mathbf{C}_i = \{C_i, \overline{C}_i\}$ for $i = 1, 2, 3$ be three complete questions on Ω. Let $V(\mathbf{C}_i)$ be a quasi-perfect answer to question $\mathbf{C}_i$, $i = 1, 2, 3$ characterized by error probability α. We can use the expression (3.22) to obtain the depth of these answers (see Fig. 3.1 for an illustration).

$$Y(\Omega, \mathbf{C}_1, P, V(\mathbf{C}_1)) = \frac{7}{16}\left(1 - \frac{3}{4}\alpha\right)\log(4 - 3\alpha) + \frac{9}{16}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4 - \alpha}{3} + \frac{15}{32}\alpha\log\alpha,$$

$$Y(\Omega, \mathbf{C}_2, P, V(\mathbf{C}_2)) = \frac{1}{16}\left(1 - \frac{3}{4}\alpha\right)\log(4 - 3\alpha) + \frac{15}{16}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4 - \alpha}{3} + \frac{9}{32}\alpha\log\alpha,$$

and

$$Y(\Omega, \mathbf{C}_3, P, V(\mathbf{C}_3)) = \frac{1}{4}\left(1 - \frac{3}{4}\alpha\right)\log(4 - 3\alpha) + \frac{3}{4}\left(1 - \frac{1}{4}\alpha_1\right)\log\frac{4 - \alpha}{3} + \frac{3}{8}\alpha\log\alpha.$$

In all these examples, we see that, as expected, the depths of answers to more difficult questions is higher for the same accuracy (value of error probability α). In other words, it takes more effort on the part of the information source to answer a more difficult question with the same accuracy. Equivalently, the same amount of effort (measured by pseudo-energy) yields a lower accuracy answer to a more difficult question. We can see from Fig. 3.1 that, for instance, a quasi-perfect answer of depth equal to 0.4 to the question *"Is the fruit green?"* has an error probability of around 0.18, but an equally deep (i.e. of the same depth) answer to the more difficult question *"Is the fruit a peach or not?"* has a larger error probability of around 0.24.

Let us turn to relative depth of answers. Consider the above example again. The relative depth $Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}'))$ of a quasi-perfect answer $V_\alpha(\mathbf{C}')$ with respect to question $\mathbf{C}''$ can be readily computed using the expression (3.33). We obtain, for questions $\mathbf{C}_1$ and $\mathbf{C}_2$,

$$
\begin{aligned}
Y(\Omega, \mathbf{C}_2, P, V_\alpha(\mathbf{C}_1)) = &\left(\frac{1}{2} - \frac{7}{32}\alpha\right) \log\left(\frac{4}{3}(1-\alpha) + \alpha\right) \\
&+ \left(\frac{1}{2} + \frac{13}{64}\alpha\right) \log\left(\frac{8}{9}(1-\alpha) + \alpha\right) + \frac{1}{64}\alpha \log \alpha,
\end{aligned}
$$

$$
\begin{aligned}
Y(\Omega, \mathbf{C}_1, P, V_\alpha(\mathbf{C}_2)) = &\left(\frac{1}{2} - \frac{1}{32}\alpha\right) \log\left(\frac{4}{3}(1-\alpha) + \alpha\right) \\
&+ \left(\frac{1}{2} - \frac{5}{64}\alpha\right) \log\left(\frac{8}{9}(1-\alpha) + \alpha\right) + \frac{7}{64}\alpha \log \alpha.
\end{aligned}
$$

Likewise, for questions $\mathbf{C}_1$ and $\mathbf{C}_3$, we have

$$
\begin{aligned}
Y(\Omega, \mathbf{C}_3, P, V_\alpha(\mathbf{C}_1)) = &\left(\frac{11}{16} - \frac{5}{16}\alpha\right) \log\left(\frac{4}{3}(1-\alpha) + \alpha\right) \\
&+ \left(\frac{5}{16} + \frac{1}{4}\alpha\right) \log\left(\frac{8}{9}(1-\alpha) + \alpha\right) + \frac{1}{16}\alpha \log \alpha,
\end{aligned}
$$

and

$$Y(\Omega, \mathbf{C}_1, P, V_\alpha(\mathbf{C}_3)) = \left(\frac{11}{16} - \frac{7}{32}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) + \left(\frac{5}{16} + \frac{7}{64}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{7}{64}\alpha\log\alpha.$$

Finally, for questions $\mathbf{C}_2$ and $\mathbf{C}_3$, expression (3.33) yields

$$Y(\Omega, \mathbf{C}_3, P, V_\alpha(\mathbf{C}_2)) = \left(\frac{5}{16} + \frac{1}{16}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) + \left(\frac{11}{16} - \frac{1}{8}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{1}{16}\alpha\log\alpha,$$

and

$$(\Omega, \mathbf{C}_2, P, V_\alpha(\mathbf{C}_3)) = \left(\frac{5}{16} - \frac{1}{32}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) + \left(\frac{11}{16} + \frac{1}{64}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{1}{64}\alpha\log\alpha.$$

These relative depth curves are shown in Fig. 3.2. We can see, in particular, that the relative depth $Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}'))$ is not in general symmetric in the two questions unless $\alpha = 0$ or $\alpha = 1$. In the former case the relative depth reduces to the overlap $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ which is symmetric and in the latter case the relative depth simply vanishes. Further, it can be seen from Fig. 3.2 that the relative depth can in fact be negative meaning that it is possible that the knowledge of an (imperfect) answer to a question may make another question more difficult. It would be interesting to establish general conditions under which relative depth is nonnegative. Another useful observation is that if for a pair of questions $\mathbf{C}'$ and $\mathbf{C}''$, question $\mathbf{C}'$ is the more difficult one of the two then it appears that

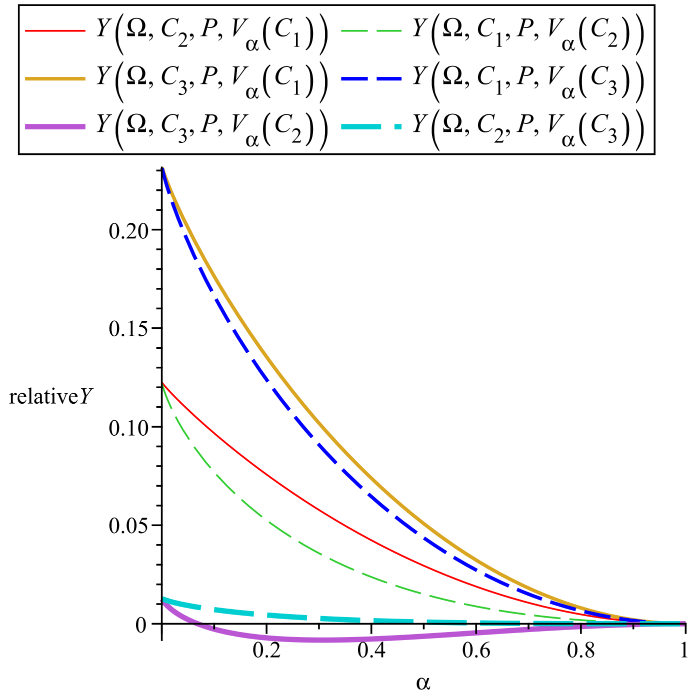

Figure 3.2: *Relative depth of quasi-perfect answers as function of* α*.*

the inequality $Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}')) > Y(\Omega, \mathbf{C}', P, V_\alpha(\mathbf{C}''))$ holds for $0 < \alpha < 1$ implying that a quasi-perfect answer to a more difficult question results in a higher reduction of difficulty of the other question. It would be of interest to see if this property holds in the general case or exceptions are possible.

3.7 Conclusion

The subject of the present chapter is answers that the information source can give in response to questions. In particular, any answer to a question can be assigned the amount of pseudo-energy that measures the answer depth, i.e. the amount of "work" the source has to do in order to answer the question to the given accuracy. Clearly, the higher the desired accuracy is the more "work" the source would have to do and the higher the answer depth is. Also, if properly defined, it makes sense that the answer depth has to be bounded by the question difficulty from above reaching that bound if and only if the answer is fully correct. In this chapter, the overall form of the answer depth function was established in the way similar to how it was done for the question difficulty in Chapter 2. Namely, reasonable postulates were formulated that the answer depth function had to satisfy. The proposed system of postulates expressed the linearity and isotropy properties of the answer depth function. One can say therefore that the latter is obtained within the "ideal gas model" that was already used in the previous chapter. It turns out that the resulting depth function is described, besides appropriate probability measures, by a scalar function on the problem parameter space that has to be the same pseudo-temperature function describing the corresponding question difficulty.

In addition to answer depth, the relative depth of an answer to one question with respect to another question was defined that can be used to determine how an answer to one question reduces the difficulty of a different question. It is expected that the relative depth will become especially useful when the optimal additional information acquisition process with multiple information sources is studied in later publications.

Chapter 4

Information Source Models

4.1 Introduction

The two previous chapters established the first two parts of a quantitative framework for the description of information exchange between the decision maker and information sources: the concepts of question difficulty and answer depth. This chapter explores the third component of the framework, i.e. quantitative models of information sources. The concept of a source model is introduced and several different models are proposed. The source model parameters and the pseudo-temperature function on the problem parameter space characterizing question difficulty and answer depth in the overall "ideal gas" information exchange model can be estimated from the observed source performance on a set of sample questions. Optimization based methods for such estimation are discussed.

4.1.1 Information Source

In addition to the knowledge of the probability measure P that embodies the original state of information available to the decision maker, an information source is assumed to be capable of answering questions of the form $\mathbf{C}$ discussed above. The answers $V(\mathbf{C})$ modify the original measure P on Ω. The questions differ from each other in the degree of difficulty that can be measured – under certain assumptions – by the question difficulty function whose general form is given in Theorem 2.1. The source's answers can be characterized by their depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ whose general form is established in Corollary 3.1. As was mentioned earlier, both the question difficulty and answer depth functions depend, besides the original and updated measures on Ω, on an integrable pseudo-temperature function u: $\Omega \to \mathbb{R}$ whose value at a point $\omega \in \Omega$ has the meaning of the "local difficulty" at that point. Therefore, if the function $u(\omega)$ is given then the difficulty of any question can be computed for any original measure P on Ω. On the other hand, in any real application, the function $u(\omega)$ cannot be known since it is not directly observable. What can be observed is the information source's actual performance: the proportion of correct answers. From that, the error probabilities can be estimated. This means that the function $u(\omega)$ has to be estimated from the knowledge of error probabilities exhibited by the information source in response to some particular questions. Informally speaking, the error probabilities tell us indirectly what questions are easy and which are hard for the information source. If we assume that the postulates discussed in Chapters 2 and 3 are valid (that is if the linear isotropic model is adequate) then the function $u(\omega)$ can be found that would reproduce – within estimation error – the observed (estimated) error probabilities. The estimated function $u(\omega)$, in turn, would allow for computation of difficulties of other questions that have not been given to the source before.

4.1. INTRODUCTION

Let us recall the general assumptions that were made about the information source:

(i) Questions that can be given to the source have different degrees of detalization and difficulty.

(ii) A question's degree of difficulty is related to the question degree of detalization but in general does not coincide with it.

(iii) The quality of source's answers is directly related to the degree of difficulty of the corresponding questions.

(iv) The source has a finite capacity.

(v) The source "tries equally hard" to answer any question it receives. Therefore, the source answers those questions well (i.e. with low error probabilities) whose difficulty does not exceed the source's capacity. As the difficulty exceeds the source's capacity the quality of its answers progressively degrades.

Assumptions (i) and (ii) are subsumed by question difficulty postulates: the degree of detalization for the question $\mathbf{C} = \{C_1, \dots, C_r\}$ can be identified with the number of subsets in the corresponding partition (in the "topological" sense) or with the expression $-\sum_{j=1}^{r} P(C_j) \log P(C_j)$ (in the "metric" sense) and its difficulty is given by $G(\Omega, \mathbf{C}, P)$. The latter is different from the "metric" degree of detalization by virtue of the presence of function $u(\omega)$ and reduces to it for the case of constant u.

Assumption (iii) implies that the source answers questions in such a way that the quality of its answers measured by the answer depth function is in direct relation to the question difficulty – measured by the question difficulty function. More precisely, for the given information source, the answer depth has to be a function of the corresponding question difficulty. Assumptions (iv) and

(v) then imply that this function is non-decreasing and is bounded from above. We formalize these observations by adapting the following main hypothesis.

Hypothesis S1. For the given information source and any question $\mathbf{C}$, the corresponding answer depth is a function of the question difficulty:

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = h(G(\Omega, \mathbf{C}, P)),$$

where $h(\cdot)$ is a non-decreasing function of its argument that's bounded from above.

The hypothesis S1 essentially states that the question difficulty and answer depth are exhaustive characterizations of the pseudo-energy content of questions and answers, respectively. If two different questions have the same difficulty, the information source will answer them equally well, i.e. the depth of answers will be the same.

It is natural to call the particular form of function $h(\cdot)$ the *model of the source*. In practice, the overall form of $h(\cdot)$ has to be postulated. Then the values of parameters needed full specification of $h(\cdot)$ and the function $u(\omega)$ can be estimated from the observed performance of the source on sample questions.

4.2 Possible Source Models

As was mentioned above, the model of the source is described by a non-decreasing function $h(\cdot)$ where the role of the argument is played by the question difficulty $G(\cdot)$. The function $h(\cdot)$ should also be bounded from above if one assumes (as we do) that a source has a finite (effective) informational capacity. Let us now describe some possible models.

4.2.1 Simple Capacity Model

In this model, the information source is characterized by a single parameter that can be called the *pseudo-energy capacity* and denoted by Y_s. Under this model, the source can provide perfect answers to questions whose difficulty does not exceed Y_s and, for questions with difficulty exceeding Y_s, the error probabilities increase in such a way that the depth of the corresponding answer stays equal to Y_s. Put slightly differently, the information source provides answers whose depth is constant unless the question is too easy for the source in which case the depth of the answer is limited by the difficulty of the question. Formally speaking, the function $h(x)$ for this model takes the following form.

$$h(x) = \begin{cases} x & \text{if } x \leq Y_s \\ Y_s & \text{if } x > Y_s. \end{cases} \tag{4.1}$$

In reality, while one wouldn't expect a perfect fit of empirical data to (4.1), large deviations could indicate either inadequacy of the linear isotropic model of question difficulty or that of the capacity model (4.1) of the information source.

4.2.2 Modified Capacity Models

The main drawback of the simple capacity model described above is that the information source is postulated to provide perfect answers to questions whose difficulty is below the source's capacity. On the other hand, in many situations, it is reasonable to expect that a source will make some error answering even relatively simple questions. The modified capacity models' goal is to allow for finite error probabilities for answers to questions with difficulties below the source capacity. This model depends on more than one parameter: besides the capacity Y_s, there is also a parameter describing the approach by the function $h(\cdot)$ of its maximum value Y_s. The simplest of such models is the *linear* modified capacity model described by

$$h(x) = \begin{cases} bx & \text{if } x \leq \frac{Y_s}{b} \\ Y_s & \text{if } x > \frac{Y_s}{b}. \end{cases} \tag{4.2}$$

where $b \leq 1$ is the second parameter. Under this model, the source makes errors even on questions with difficulties below the capacity with error probabilities gradually increasing with question difficulties. Once the question difficulty exceeds the capacity of the source, the corresponding answer depth stays equal to the capacity Y_s.

The linear modified capacity model can be naturally generalized to the *polynomial modified capacity model* in which the function $h(\cdot)$ approaches its maximum value according to a polynomial law. To describe it, let $p_q(x) = a_0 + a_1 x + \ldots + a_q x^q$ be an order q polynomial and let x_q^* be the smallest positive root of the equation $p_q(x) - Y_s = 0$. Then the polynomial modified capacity model has the form

$$h(x) = \begin{cases} p_q(x) & \text{if } x \leq x_q^* \\ Y_s & \text{if } x > x_q^*. \end{cases} \tag{4.3}$$

Demanding that $h(0) = 0$ and $h(x) \leq x$ for all $x \geq 0$ leads to $a_0 = 0$ and $0 \leq a_1 \leq 1$. For $q = 2$, the polynomial modified capacity model (4.3) reduces to the *quadratic modified capacity model* that is most conveniently written in the form

$$h(x) = \begin{cases} bx - \frac{\gamma}{Y_s} x^2 & \text{if } x \leq G_2 \\ Y_s & \text{if } x > G_2, \end{cases} \tag{4.4}$$

where $0 < b \leq 1$ and (assuming $\gamma > 0$ so that $h(x)$ is concave) $\gamma \leq \frac{b}{4}$; $G_2 = \frac{b-\sqrt{b^2-4\gamma}}{2\gamma} Y_s$. In this model Y_s has the meaning of the source capacity and the coefficients b and γ are pure numbers (i.e. their numerical values do not depend on the choice of units of pseudo-temperature $u(\cdot)$ and capacity Y_s).

Another simple model that belongs to the class of modified capacity models is the *exponential modified capacity model*

$$h(x) = Y_s(1 - e^{-\frac{\theta}{Y_s}x}) \tag{4.5}$$

that depends on two parameters: capacity Y_s and $0 < \theta \leq 1$ that controls the speed with which the function $h(x)$ approaches its upper bound Y_s. The coefficient θ is a pure number in the sense described above. One of the advantages of the exponential model (4.5) is that it is described by a single analytical function that allows the corresponding estimation problem that is discussed in the next section to avoid binary variables.

4.3 Estimation of Model Parameters and Function $u(\omega)$

First, let us note that both question difficulty and answer depth functions are linear in $u(\omega)$ and therefore multiplying $u(\omega)$ by any constant would result in both difficulty and depth being multiplied by the same constant without changing any of the coefficients p_{kj}, $k = 1, \ldots, m$, $j = 1, \ldots, r$ and therefore answer error probabilities. This means that the function $u(\omega)$ is really defined up to a single multiplicative constant the choice of which is equivalent to a choice of units in which $u(\omega)$ (and the difficulty/depth functions) are measured. We use two different conventions that turn out to be convenient.

- The normalized $u(\cdot)$ convention in which $\int_\Omega u(\omega)dP(\omega) = 1$ for every information source. This convention is convenient because if $u(\omega) \equiv 1$ the difficulty of question $\mathbf{C}$ reduces to Shannon entropy of the distribution $P(\mathbf{C}) = (P(C_1, \ldots, P(C_r))$.

- The unit source capacity convention in which the units of $u(\omega)$ are chosen in such a way that, for each information source, the capacity is unity: $Y_s = 1$. This convention is especially convenient for comparing different information sources to each other. Indeed, in this case, functions $u(\omega)$ for any two sources can be directly compared to each other showing clearly the relative degree of "expertise" of each source in various regions of Ω and also giving a sense of "absolute" quality of each source.

If the function $u(\omega)$ is known, then Theorem 2.1 gives – for the given measure P – the difficulty of any question $\mathbf{C}$. Then for any answer $V(\mathbf{C})$ to $\mathbf{C}$ the knowledge of updated measures P^k allows one to find the depth of $V(\mathbf{C})$. On the other hand, a given source model $Y = h(G)$ lets one *predict* the depth of the source's answer to any question before measures P^k can be estimated. Thus in order to be able to predict the depth of source's answer to various questions – and hence possibly solve the problem (1.3) – one needs to know (i) the function $u(\omega)$ and (ii) the source model described by the function $h(\cdot)$. Since these functions cannot be directly measured or observed, the only way to find these two functions in any realistic application is to estimate them from the source's performance on a certain set of sample questions.

Let $\mathbf{D} = \{D_1, \ldots, D_{N_d}\}$ be a partition of Ω to be used for discretizing the weight function $u(\omega)$: we assume that $u(\omega)$ takes a constant value equal to u_i on subset D_i. Let $w_i = P(D_i)$ and let $\mathcal{N}_i \subset \{1, \ldots, N_d\}$ be the index set of the subsets in $\mathbf{D}$ that are immediate neighbors of (i.e. have a common boundary with) subset D_i. We assume that the partition $\mathbf{D}$ is sufficiently fine so that any partition $\mathbf{C}$ used for estimating $u(\omega)$ can be considered a coarsening of $\mathbf{D}$.

Further, let $\mathbf{C}_1, \ldots, \mathbf{C}_K$ be a set of questions that the source has answered and its answers have been compared with actual outcomes in Ω. Let us denote by $G_1, \ldots, G_K$ the difficulties of these questions and let $Y_1, \ldots, Y_K$ be the corresponding answer depth values that were computed using the estimated error probabilities. For the sake of simplicity, we assume that the answers of the source are quasi-perfect (see (3.20) and (3.21) for the form of coefficients p_{kj} and updated measures P^k) with the corresponding (estimated) error probabilities being equal to $\alpha_1, \ldots, \alpha_K$, respectively.

Let us denote $z_i = |Y_i - h(G_i)|$, $i = 1, \ldots, K$ where the function $h(\cdot)$ is given by the suitable information source model. The quantities z_i measure the absolute values of deviations of the empirical data from the chosen source model, vanishing values of all variables z_i corresponding to a perfect fit. In addition to minimizing the sum of the deviations (i.e. maximizing the fit), it makes sense to demand that the quantities u_j, $j = 1, \ldots, N_d$, describe a reasonably smooth function $u(\omega)$. This can be achieved, for instance, by putting an upper bound on the gradient of $u(\omega)$ or, equivalently, by putting a corresponding term in the objective function. To make it more precise, let $N(\mathbf{D})$ be the set of neighbors in the partition $\mathbf{D}$ (i.e. $N(\mathbf{D}) = \{(i,j) : j \in \mathcal{N}_i, i = 1, \ldots N_d\}$) and let U be the desired upper bound on the difference of two values of u on neighboring sets of partition $\mathbf{D}$. Then if the capacity model $h(\cdot)$ is postulated, the following formulation of the estimation problem for the function $u(\omega)$ and the parameters of model $h(\cdot)$ is obtained.

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\
\text{subject to} \quad & Y_i - h(G_i) \le z_i,\ i = 1, \dots, K \\
& h(G_i) - Y_i \le z_i,\ i = 1, \dots, K \\
& u_j - u_k \le U,\ (j, k) \in N(\mathbf{D}) \\
& u_k - u_j \le U,\ (j, k) \in N(\mathbf{D})
\end{aligned} \tag{4.6}$$

The decision variables in (4.6), besides z_i, are u_j, $j = 1, \dots, N_d$ and the parameters of function $h(\cdot)$. The parameter λ controls the trade-off between the objective of maximizing the fit and that of maximizing smoothness of $u(\omega)$ (understood as minimizing the maximum gradient of $u(\omega)$). The difficulties G_i, $i = 1, \dots, K$ are expressed via the decision variables as follows

$$G_i = -\sum_{j=1}^{r_i} \log P(C_j) \sum_{\{l: D_l \subset C_j\}} u_l w_l. \tag{4.7}$$

For the values of the depth function for the corresponding answers, let us assume, for simplicity that the answers are quasi-perfect implying that their errors can be characterized with a single probability α_i, $i = 1, \dots, K$. Then the depth Y_i can be written as

$$\begin{aligned}
Y_i = & \sum_{j=1}^{r_i} (1 - \alpha_i + \alpha_i P(C_j)) \log \frac{1 - \alpha_i + \alpha_i P(C_j)}{P(C_j)} \sum_{\{l: D_l \subset C_j\}} u_l w_l \\
& + \alpha_i \log \alpha_i \sum_{j=1}^{r_i} P(C_j) \left(1 - \sum_{\{l: D_l \subset C_j\}} u_l w_l \right).
\end{aligned} \tag{4.8}$$

Note that in general, (4.6) is a potentially complex nonlinear optimization problem where nonlinearity is introduced by the function $h(\cdot)$. For the case of the simple capacity model the problem (4.6) can be written as

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\
\text{subject to} \quad & Y_i - Y_s \leq z_i + My_i,\ i = 1, \ldots, K \\
& Y_s - Y_i \leq z_i + My_i,\ i = 1, \ldots, K \\
& G_i - Y_i \leq z_i + M(1 - y_i),\ i = 1, \ldots, K \\
& u_j - u_k \leq U,\ (j,k) \in N(\mathbf{D}) \\
& u_k - u_j \leq U,\ (j,k) \in N(\mathbf{D}) \\
& y_i \in \{0, 1\},\ i = 1, \ldots, K.
\end{aligned} \tag{4.9}$$

In this formulation, M is a large number, y_i, $i = 1, \ldots K$ are auxiliary binary variables. The main decision variables in the formulation (4.9) are the values u_j, $j = 1, \ldots, N_d$ and the capacity value Y_s. Since both (4.7) and (4.8) are linear in the variables u_l, the optimization problem (4.9) is mixed-linear with K binary variables. Therefore, it can at least be solved efficiently for moderate values K of sample questions used for estimating model parameter Y_s and the (discretized) function $u(\omega)$.

The formulation (4.9) can be modified easily from the simple to the modified capacity model. The resulting formulation is as follows

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\ \text{subject to} \quad & Y_i - Y_s \le z_i + M y_i,\ i = 1, \ldots, K \\ & Y_s - Y_i \le z_i + M y_i,\ i = 1, \ldots, K \\ & bG_i - Y_i \le z_i + M(1 - y_i),\ i = 1, \ldots, K \\ & u_j - u_k \le U,\ (j,k) \in N(\mathbf{D}) \\ & u_k - u_j \le U,\ (j,k) \in N(\mathbf{D}) \\ & y_i \in \{0,1\},\ i = 1, \ldots, K. \end{aligned} \tag{4.10}$$

The additional decision variable in (4.10) is $b \le 1$. The values G_i and Y_i, $i = 1, \ldots K$ are given by expressions (4.7) and (4.8), respectively. The formulation (4.10), just like (4.9), is a mixed-linear optimization problem with K binary variables and thus can at least be solved efficiently for moderate values of the number K of sample questions.

The formulation for the quadratic modified capacity model (4.4) can be easily obtained from (4.10) by replacing the constraints $bG_i - Y_i \le z_i + M(1-y_i)$, $i = 1, \ldots, K$ with $bG_i + cG_i^2 - Y_i \le z_i + M(1 - y_i)$, $i = 1, \ldots, K$. Recalling that G_i is a linear function of the decision variables u_l, we see that the resulting problem is that of quadratic optimization with K binary variables that enter the formulation in a linear fashion. Even though such problems can't in general be solved as efficiently as mixed-linear optimization problems of equal size, they can still be solved to optimality for moderate values of parameters K and N_d.

As mentioned earlier, the exponential capacity model has the advantage over other models discussed here in that it obviates the need for binary variables even though it becomes severely nonlinear:

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\
\text{subject to} \quad & Y_i - Y_s(1 - e^{-\theta G_i}) \leq z_i,\ i = 1, \ldots, K \\
& Y_s(1 - e^{-\theta G_i}) - Y_i \leq z_i,\ i = 1, \ldots, K \\
& u_j - u_k \leq U,\ (j,k) \in N(\mathbf{D}) \\
& u_k - u_j \leq U,\ (j,k) \in N(\mathbf{D}).
\end{aligned} \tag{4.11}$$

Besides the quantities z_i, $i = 1, \ldots, K$, u_l, $l = 1, \ldots, N_d$ and the source capacity Y_s, another decision variable is the parameter $0 < \theta \leq \frac{1}{Y_s}$.

It is worth noting that in estimation of the pseudo-temperature function and model parameters the error probabilities are themselves estimated values. That introduces obvious imprecision in estimation of pseudo-temperature and source model parameters. In fact, one can think of the procedure described in this section as similar to point estimation of parameters in classical statistics. For more information about the pseudo-temperature function, confidence intervals would be needed. The width of such confidence intervals would obviously depend on the precision with which error probabilities are known and therefore on the sample size used in error probability estimation. Practically, such confidence intervals may turn out to be sufficiently wide to effectively invalidate precise estimation of the shape of pseudo-temperature function. The practical approach instead could be that of the hypothesis testing type: a null (default) hypothesis about the shape of the pseudo-temperature function would be stated (i.e. that the pseudo-temperature is constant or linear) and then tested using standard statistical methods.

Just like in probability estimation, expert opinion can be used for estimating pseudo-temperature function. Since pseudo-temperature admits a simple intuitive interpretation (as local "degree of difficulty") experts should find it easy enough to give useful estimates of pseudo-temperature. If, in addition, some data about observed source performance is available, it can be used in conjunction with expert estimates, for instance, by using these estimates as a null hypothesis and using observed data for the purpose of testing it.

4.4 Examples

To illustrate the process of estimation of the pseudo-temperature $u(\omega)$ and source model parameters, consider an example in which $\Omega = [0,1]^2 \subset \mathbb{R}^2$, and the measure P is uniform continuous on Ω. Consider the set of sample (complete) questions illustrated in Fig. 4.1. Our goal is, given the error parameters α_i for quasi-perfect answer $V_{\alpha_i}(\mathbf{C}_i)$ to question $\mathbf{C}_i$, $i = 1, \ldots, 10$, estimate the function $u(\omega)$ and the parameter(s) of the chosen information source model.

We adapt the modified linear source model and use formulation (4.10) to estimate $u(\omega)$, and parameters Y_s and b of the model. We do this for different values of error probabilities.

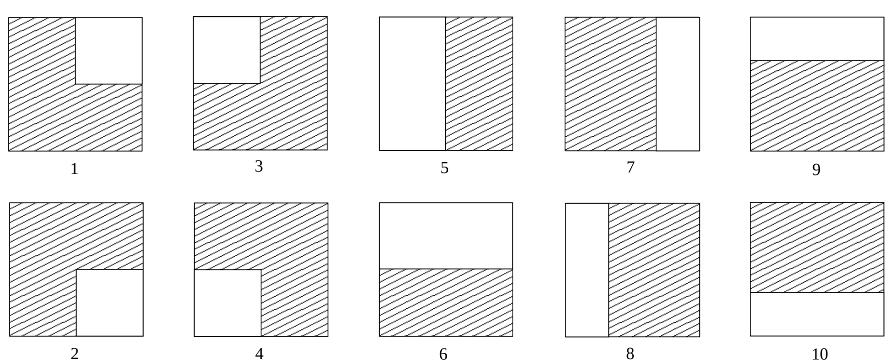

Figure 4.1: *Sample questions.*

First consider data shown in Table 4.1. In this and following tables, the first column contains the index i of question $\mathbf{C}_i$ from Fig. 4.1, the second column shows the corresponding error probability α_i, and the last two columns contain the question difficulty $G(\Omega, \mathbf{C}_i, P)$ and answer depth $Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$, respectively, obtained from the estimated values of $u(\omega)$ and parameters of the source model. In the lower part of Table 4.1, the resulting value of the objective of problem (4.10) along with the estimated values of parameters Y_s and b are shown.

The error probability values shown in Table 4.1 result in a perfect fit ($z = 0$) with the estimated pseudo-temperature function $u(\omega)$ (shown in Fig. 4.2). We can see that the resulting pseudo-temperature function increases for the larger values of coordinates ω_1 and ω_2 on Ω reflecting the fact that, for instance $\alpha_1 > \alpha_4$, implying that question $\mathbf{C}_1$ has higher difficulty (larger value of pseudo-energy) than $\mathbf{C}_4$ in spite of these two questions having same value of entropy. This means that the smaller measure subset in $\mathbf{C}_1$ has to have higher pseudo-temperature which we indeed see. It is also worth noting that questions $\mathbf{C}_5$ and $\mathbf{C}_6$ were answered with equal accuracy suggesting that these questions are of equal difficulty. This in fact is a necessary condition for a perfect fit within the ideal gas question difficulty model since in this model any complete question with all subsets of equal measure would have the same difficulty (pseudo-energy) regardless of the pseudo-temperature function form.

Consider now data shown in Table 4.2. The resulting pseudo-temperature $u(\omega)$ is shown in Fig. 4.3. We see that in this case the perfect fit could not be achieved by any pseudo-temperature function, in particular because questions $\mathbf{C}_5$ and $\mathbf{C}_6$ were answered with slightly different accuracy whereas these two questions necessarily have equal pseudo-energy content (equal difficulty) within the ideal gas question difficulty model.

Table 4.1: *Sample question error probabilities, fitted values of the difficulty and depth functions, and estimated model parameter values for the modified linear model when perfect fit is possible.*

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.265	1.106	0.516
2	0.143	0.803	0.516
3	0.143	0.803	0.516
4	0.077	0.533	0.404
5	0.210	1.000	0.516
6	0.210	1.000	0.516
7	0.253	1.102	0.516
8	0.116	0.761	0.516
9	0.253	1.102	0.516
10	0.116	0.761	0.516

$\sum_{i=1}^{N_d} z_i = 0$; $U = 0.13$; $Y_s = 0.52$; $b = 0.76$.

Table 4.2: *Sample question error probabilities, fitted values of the difficulty and depth functions, and estimated model parameter values for the modified linear model when perfect fit is not possible, with small misfit.*

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.238	1.057	0.531
2	0.157	0.856	0.531
3	0.129	0.794	0.531
4	0.084	0.538	0.399
5	0.189	1.000	0.549
6	0.230	1.000	0.484
7	0.227	1.055	0.531
8	0.127	0.806	0.531
9	0.278	1.200	0.525
10	0.127	0.806	0.531

$\sum_{i}^{N_d} z_i = 0.07$; $U = 0.43$; $Y_s = 0.53$; $b = 0.74$.

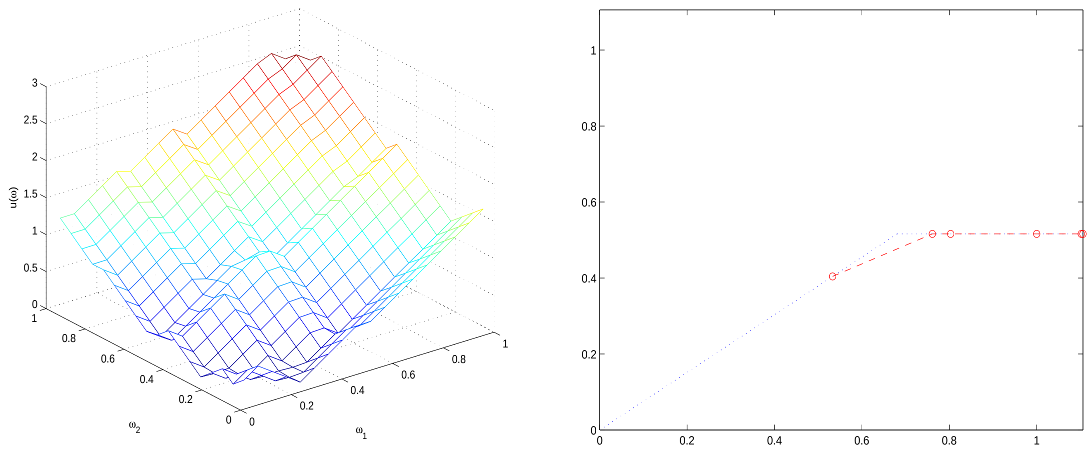

Figure 4.2: *The estimated pseudo-temperature (left) and the fitted values of difficulty and depth (right) for the data of Table 4.1.*

Now, consider the data shown in Table 4.3. As can be seen from Fig. 4.4, the fit that could be achieved to the ideal gas question difficulty model (with the linear modified information source model) is relatively (at least compared to the previous example) poor, possibly indicating that the ideal gas model may not be adequate in this case and that a different model (for example, anisotropic – to be able to model different pseudo-energy content of questions $\mathbf{C}_5$ and $\mathbf{C}_6$) may be needed.

Let us now turn to comparing different sources. Suppose $\Omega = [0, 1]$ with P being a uniform continuous measure on Ω. Let sample questions be as follows. $\mathbf{C}_1 = \{[0, 1/2], (1/2, 1]\}$, $\mathbf{C}_2 = \{[0, 1/3], (1/3, 1]\}$, $\mathbf{C}_3 = \{[0, 2/3], (2/3, 1]\}$, $\mathbf{C}_4 = \{[0, 1/4], (1/4, 1]\}$, $\mathbf{C}_5 = \{[0, 3/4], (3/4, 1]\}$. Let source 1 accuracy be described by error probabilities (assuming quasi-perfect answers as before) shown in Table 4.4. Then, using the modified capacity model and formulation (4.10), we can estimate the pseudo-temperature function $u(\cdot)$ and the model parameters Y_s and b. The results – as well as fitted values of the question difficulty and answer depth – are shown in Table 4.4.

Table 4.5 shows error probabilities achieved on the same set of sample questions by a different source 2, along with the resulting fitted values of difficulty and depth functions and the estimated

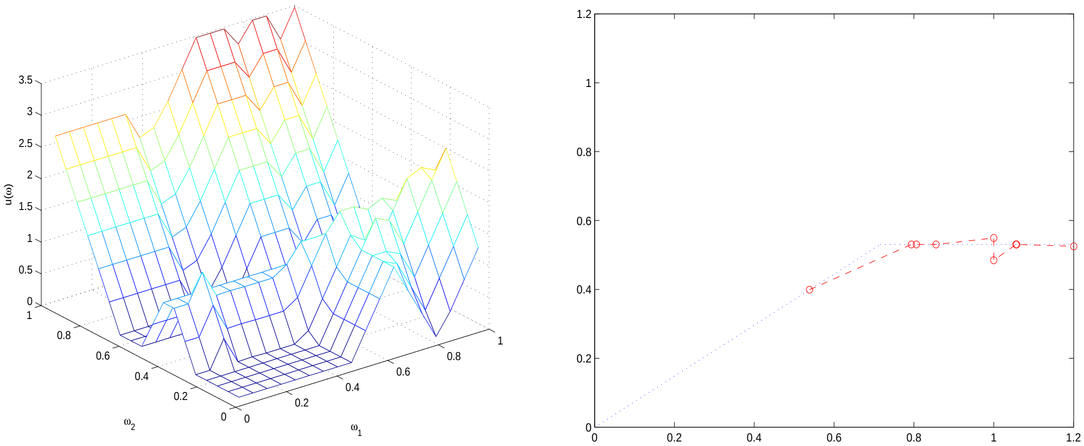

Figure 4.3: *The estimated pseudo-temperature (left) and the fitted values of difficulty and depth (right) for the data of Table 4.2.*

model parameter values. Looking at Tables 4.4 and 4.5 we can see, for example, that source 1 shows better overall performance on all questions, but there exist questions (question 5, for instance) that appear to be easier for source 2. Indeed, the estimated pseudo-temperature functions shown in Fig. 4.5 (in the unit source capacity convention) clearly demonstrate that the overall pseudo-temperature is significantly higher for source 2 thus making the majority of sample questions more difficult for it (which is reflected in higher error probabilities). On the other hand, while the pseudo-temperature function for source 1 is (mostly) increasing on the interval $[0, 1]$, it is a decreasing function on the same interval for source 2. In particular, there exist regions of $\Omega = [0, 1]$ where the pseudo-temperature for source 2 is lower than that for source 1. This means that some questions can be easier for source 2, question 5 from the sample set being an example.

Table 4.3: *Sample question error probabilities, fitted values of the difficulty and depth functions, and estimated model parameter values for the modified linear model when perfect fit is not possible, with larger misfit.*

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.371	0.418	0.118
2	0.086	0.488	0.358
3	0.200	0.589	0.312
4	0.107	1.750	1.281
5	0.126	1.000	0.661
6	0.293	1.000	0.399
7	0.354	0.585	0.180
8	0.162	1.320	0.812
9	0.354	0.590	0.182
10	0.162	1.219	0.746

$\sum_i^{N_d} z_i = 1.51$; $U = 0.56$; $Y_s = 1.28$; $b = 0.73$.

Table 4.4: *Sample question error probabilities, fitted values of the difficulty and depth functions, estimated model parameter values for the modified linear model, for information source 1.*

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.090	1.000	0.735
2	0.070	0.678	0.525
3	0.153	1.174	0.735
4	0.070	0.528	0.408
5	0.146	1.131	0.735

$\sum_i^{N_d} z_i = 0.09$; $U = 0.54$; $Y_s = 0.74$; $b = 0.77$.

Table 4.5: *Sample question error probabilities, fitted values of the difficulty and depth functions, estimated model parameter values for the modified linear model, for information source 2.*

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.300	0.933	0.386
2	0.350	1.000	0.331
3	0.170	0.415	0.229
4	0.350	1.115	0.386
5	0.080	0.585	0.434

$\sum_i^{N_d} z_i = 0.18$; $U = 0.56$; $Y_s = 0.39$; $b = 0.74$.

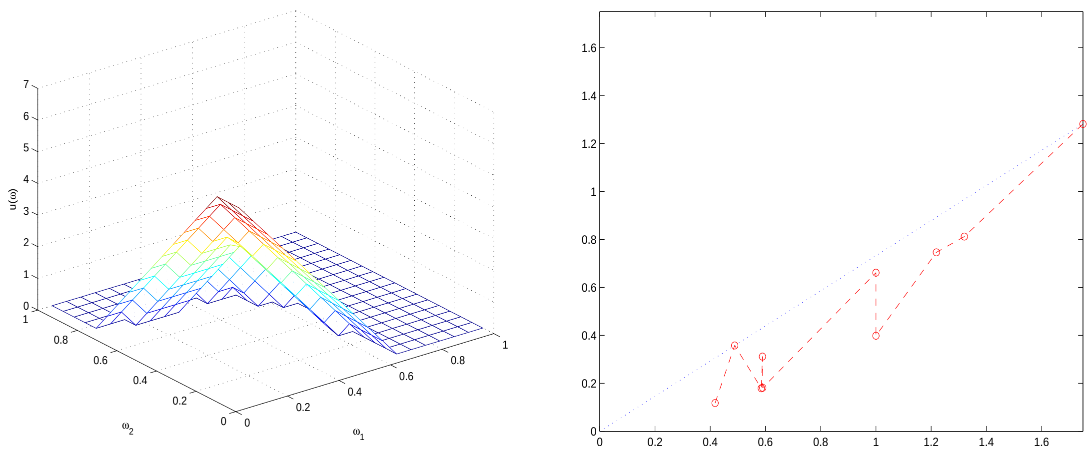

Figure 4.4: *The estimated pseudo-temperature (left) and the fitted values of difficulty and depth (right) for the data of Table 4.3.*

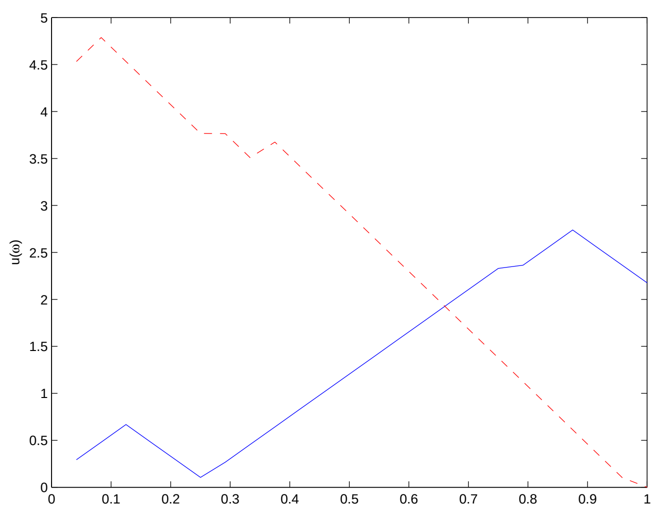

Figure 4.5: *Estimated pseudo-temperature functions for information source 1 (solid blue line) and source 2 (dashed red line).*

4.5 Conclusion

This is the third and final chapter devoted to the development of an information exchange model as part of a larger quantitative framework describing the process of additional information acquisition in decision making problems under uncertainty. The proposed framework is based on the assumption that the decision maker has access to one or more sources of information capable of answering questions concerning the problem parameter space, or, equivalently, the space of uncertain problem parameters (input data). The question difficulty function introduced and studied in Chapter 2 can serve as a quantitative measure of the degree of difficulty of various questions for the given source. The main idea is that the knowledge of this function of the source allows the decision maker to predict the degree of accuracy of the source's possible answers to various questions and thus enables the decision maker to determine the particular question(s) that need to be asked to the given source in order to maximize the answer's impact on the solution quality for the given problem. The answer depth function studied in Chapter 3 provides a quantitative measure of the "amount of work" the source has to do in order to provide an answer of given accuracy to the question at hand. Roughly speaking, the main idea here is that the source would not be able to answer difficult question accurately because the answer depth required to make the answer accurate would exceed the source's capability. And it is the latter that is the main subject of the present chapter.

The main goal of the present chpater is twofold: to study possible models of information sources and to propose methods for estimation of model parameters from the observed source's performance on sample questions. Information source models quantitatively express the idea that an information source can answer easy question more accurately than difficult ones. More precisely, the source's answer depth is limited just by question difficulty for questions that are easy enough and by the source's capability for more difficult questions. This simple and natural idea is quantified by the information source model that is a functional dependence of the answer depth on the question difficulty. It is easy to see that such a function has to be nondecreasing and has to approach a finite value for large values of the argument. In this paper, several such functions were proposed.

As was shown, both the question difficulty and answer depth functions are described, besides appropriate probability measures, by a scalar function on the problem parameter space – termed pseudo-temperature using parallels with thermodynamics. In real applications, this function needs to be estimated along with source model parameters, from the observed source performance on a set of sample questions. In this paper, optimization based algorithms for estimating the pseudo-temperature function (using a suitable discretization of the parameter space) and the chosen source model parameters were proposed.

Finally, it is worth mentioning that the developments in Part I were all based on the assumption that both the question difficulty and answer depth possess linearity and isotropy (on the problem parameter space) properties that – using parallels with thermodynamics – were referred to as the "ideal gas model". While this particular assumption leads to a concise and attractive form of the difficulty and depth functions, it is entirely possible that more general (i.e. anisotropic) models would be required for accurate description of performance of realistic information sources. Such generalizations will be the subject of future research.

Part II

Optimization of Additional Information

Chapter 5

Main Framework

5.1 Introduction

When additional information sources are available in decision making problems that allow stochastic optimization formulations, an important question is how to optimally use the information the sources are capable of providing. Here, a framework is developed that relates information characteristics of a source to solution quality characteristics of the problem and formulates the problem of optimal information acquisition. The problem is that of minimization of the expected loss of the solution subject to (pseudo-energy) capacity constraints of the information source.

5.2 Maps and their properties

In what follows, we make use of maps from Ω into X with discrete image sets. Let $\mathcal{G}$ be the set of all such maps. Since the image set of all maps from $\mathcal{G}$ is assumed to be discrete, any such map

$g \in \mathcal{G}$ can be uniquely described by the corresponding partition $\mathbf{C} = \{C_1, \ldots, C_r\}$ of Ω and the corresponding image set $I = \{x_1, \ldots, x_r\}$ such that $g(\omega) = x_j$ for all $\omega \in C_j$. We will sometimes write $g = (\mathbf{C}, I)$ whenever the components of a map (partition and image set) need to be made explicit.

The following maps from the set $\mathcal{G}$ are important special cases that will be referred to later.

- Optimal ("zero loss") map g_0: $g_0(\omega) = x^*_\omega$, where x^*_ω is the solution of $\min_{x \in X} f(\omega, x)$. It simply maps each scenario into the corresponding (deterministic) optimal solution.

- All-to-one maps g_x: $g_x(\omega) = x$ for all $\omega \in \Omega$. These map all elements of Ω into some single element of X.

- For the given measure P on Ω, the stochastic optimal map g_P: $g_P(\omega) = x^*_P$, where x^*_P is a solution of (1.1). Obviously, it is just a special case for of all-to-one maps g_x.

- For the given measure P and a (complete) partition $\mathbf{C} = \{C_1, \ldots, C_r\}$ of Ω, the stochastic subset optimal map $g_{\mathbf{C},P}$: $g_{\mathbf{C},P}(\omega) = x^*_{P_{C_j}}$ for all $\omega \in C_j$, $j = 1, \ldots, r$. (Here $x^*_{P_{C_j}}$ is an optimal solution of problem (1.1) with measure P replaced with the conditional measure P_{C_j}.) In the following, we denote by $\mathcal{C}$ the set of all maps of the form $g_{\mathbf{C},P}$ for all possible partitions $\mathbf{C}$ of Ω and will sometimes refer to maps from the set $\mathcal{C}$ as *subset optimal maps*.

Next, we define some useful functionals to be used later.

Let P be any probability measure on Ω and x an arbitrary element of the solution space X. We define the *suboptimality* of x with respect to P as follows:

$$S(x, P) = \mathbb{E}_P f(\omega, x) - \mathbb{E}_P f(\omega, x^*_P) = \int_\Omega (f(\omega, x) - f(\omega, x^*_P)) P(d\omega), \tag{5.1}$$

i.e. suboptimality of x w.r.t. P is the difference in objective values of problem (1.1) if x is used instead of the optimal solution x_P^*.

If P is an arbitrary measure on Ω and $g \in \mathcal{G}$ is an arbitrary map from Ω into X we define the *loss* of g with respect to P as

$$L(g, P) = \mathbb{E}_P f(\omega, g(\omega)) - \mathbb{E}_P f(\omega, x_\omega^*) = \int_\Omega (f(\omega, g(\omega)) - f(\omega, x_\omega^*))P(d\omega). \tag{5.2}$$

In particular if $g = g_P$ is the stochastic optimal map corresponding to the measure P, the loss $L(g_P, P)$ is the traditional *expected value of perfect information* (EVPI). If $g = g_0$ is the optimal map, the loss is equal to zero for any measure P: $L(g_0, P) = 0$.

Finally, for any measure P and map $g \in \mathcal{G}$ we define the *gain* of g with respect to P as follows:

$$B(g, P) = \mathbb{E}_P f(\omega, x_P^*) - \mathbb{E}_P f(\omega, g(\omega)) = \int_\Omega (f(\omega, x_P^*) - f(\omega, g(\omega)))P(d\omega). \tag{5.3}$$

The gain functional of a map g measures the decrease in loss that can be achieved by the map g, compared to the best all-to-one map g_P. In particular the largest possible gain obtains by an optimal map g_0, and for this map, the value of gain is equal to the loss of g_P, as it should since any optimal map has zero loss. It is also clear that, while suboptimality and loss are always nonnegative, gain can take both positive and negative values. For example the gain of any all-to-one map g_x is negative unless $x = x_P^*$ (in which case the gain vanishes).

The following lemma states an elementary but useful relationship between gain and loss for an arbitrary map g from Ω into X.

Lemma 5.1 *For any map $g \in \mathcal{G}$ and any measure P on Ω,*

$$B(g, P) + L(g, P) = L(g_P, P),$$

where g_P is the stochastic optimal map for the measure P.

Proof: Using definitions of gain and loss we can write

$$\begin{aligned} B(g, P) + L(g, P) &= \int_\Omega (f(\omega, x_P^*) - f(\omega, g(\omega)))P(d\omega) + \int_\Omega (f(\omega, g(\omega)) - f(\omega, x_\omega^*))P(d\omega) \\ &= \int_\Omega (f(\omega, x_P^*) - f(\omega, x_\omega^*))P(d\omega) = L(g_P, P) \end{aligned}$$

$\square$

The statement of Lemma 5.1 can be rewritten as $B(g, P) = L(g_P, P) - L(g, P)$ and, in fact can be used as a definition of the gain of arbitrary map $g \in \mathcal{G}$: the gain is equal to the decrease of the value of loss compared to the loss of the best all-to-one map g_P.

Let $f(\mathcal{P}) \to \mathbb{R}$ be a real-valued functional on the suitably restricted set $\mathcal{P}$ of measures on Ω. For the later developments it turns out to be convenient to introduce the following notation. Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a partition of Ω (a question), and let $V(\mathbf{C})$ be an answer to $\mathbf{C}$ that can take values in the set $\{s_1, \ldots, s_m\}$.

We denote by $f(P_\mathbf{C})$ the expected value of the functional $f(\cdot)$ over the set of conditional measures $\{P_{C_j}\}$, $j = 1, \ldots, r$:

$$f(P_\mathbf{C}) = \sum_{j=1}^{r} P(C_j) f(P_{C_j}), \tag{5.4}$$

and by $f(P_{V(\mathbf{C})})$ – the expected value of $f(\mathbf{C})$ over the set of updated measures $\{P^k\}$, $k =$

$1, \ldots, m$:

$$f(P_{V(\mathbf{C})}) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) f(P^k) = \sum_{k=1}^{m} v_k f(P^k), \tag{5.5}$$

Then we can define suboptimality, loss and gain functionals for a given question $\mathbf{C}$ and an answer $V(\mathbf{C})$ using the just introduced notational convention (5.4) and (5.5).

Namely, for an arbitrary $x \in X$, the suboptimality of solution x with respect to question $\mathbf{C}$ (and initial measure P) is given by

$$S(x, P_{\mathbf{C}}) = \sum_{i=1}^{s} P(C_j) S(x, P_{C_j}), \tag{5.6}$$

and the suboptimality of x with respect to answer $V(\mathbf{C})$ to question $\mathbf{C}$ (and initial measure P) reads

$$S(x, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k S(x, P^k). \tag{5.7}$$

Likewise, for an arbitrary map $g \in \mathcal{G}$, and question $\mathbf{C}$, the loss and gain of g with respect to $\mathbf{C}$ are given by

$$L(g, P_{\mathbf{C}}) = \sum_{j=1}^{r} P(C_j) L(g, P_{C_j}), \tag{5.8}$$

and

$$B(g, P_{\mathbf{C}}) = \sum_{j=1}^{r} P(C_j) B(g, P_{C_j}), \tag{5.9}$$

respectively.

The loss and gain functionals for a map $g \in \mathcal{G}$ with respect to answer $V(\mathbf{C})$ are defined analogously:

$$L(g, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k L(g, P^k), \tag{5.10}$$

and

$$B(g, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k B(g, P^k), \tag{5.11}$$

respectively.

The following representation for the expected loss $L(g, P)$ will be useful later.

Lemma 5.2 *For any map $g = (\mathbf{C}, I) \in \mathcal{G}$, the expected loss $L(g, P)$ can be written as*

$$L(g, P) = \sum_{j=1}^{r} P(C_j) L(g, P_{C_j}) = L(g, P_{\mathbf{C}}).$$

Proof:

$$\begin{aligned}
L(g, P) &= \int_{\Omega} (f(\omega, g(\omega)) - f(\omega, x^*_\omega)) P(d\omega) = \sum_{j=1}^{r} \int_{C_j} (f(\omega, g(\omega)) - f(\omega, x^*_\omega)) P(d\omega) \\
&= \sum_{j=1}^{r} P(C_j) \int_{C_j} \frac{1}{P(C_j)} (f(\omega, g(\omega)) - f(\omega, x^*_\omega)) P(d\omega) \\
&= \sum_{j=1}^{r} P(C_j) \int_{C_j} (f(\omega, g(\omega)) - f(\omega, x^*_\omega)) P_{C_j}(d\omega) \\
&\overset{(a)}{=} \sum_{j=1}^{r} P(C_j) L(g, P_{C_j}) \overset{(b)}{=} L(g, P_{\mathbf{C}}),
\end{aligned}$$

where (a) follows directly from the definition of the expected loss for the measure P_{C_j} and (b) follows from the definition (5.8) of $L(g, P_{\mathbf{C}})$. □

Let $g = (\mathbf{C}, I) \in \mathcal{C}$ be a subset optimal map. Then the EVPI for the problem (1.1) can be decomposed in a convenient way.

Lemma 5.3 *For any map $g_{\mathbf{C},P} \in \mathcal{C}$, the EVPI $L(g_P, P)$ of the problem (1.1) can be decomposed as*

$$L(g_P, P) = S(x^*_P, P_{\mathbf{C}}) + L(g_{\mathbf{C},P}, P).$$

Proof: We have

$$\begin{aligned}
L(g_P, P) &= \int_\Omega (f(\omega, x^*_P) - f(\omega, x^*_\omega))P(d\omega) \\
&= \int_\Omega (f(\omega, x^*_P) - f(\omega, x^*_\omega) + f(\omega, g_{\mathbf{C},P}(\omega)) - f(\omega, g_{\mathbf{C},P}(\omega)))P(d\omega) \\
&= \int_\Omega (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega)))P(d\omega) + \int_\Omega (f(\omega, g_{\mathbf{C},P}(\omega)) - f(\omega, x^*_\omega))P(d\omega) \\
&= \sum_{j=1}^r P(C_j) \int_{C_j} \frac{1}{P(C_j)} \left(f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega))\right) P(d\omega) \\
&+ \sum_{j=1}^r P(C_j) \int_{C_j} \frac{1}{P(C_j)} \left(f(\omega, g_{\mathbf{C},P}(\omega)) - f(\omega, x^*_\omega)\right) P(d\omega) \\
&\overset{(a)}{=} \sum_{j=1}^r P(C_j) \int_{C_j} \Big(f(\omega, x^*_P) - f(\omega, x^*_{P_{C_j}})\Big) P_{C_j}(d\omega) \\
&+ \sum_{j=1}^r P(C_j) \int_{C_j} \left(f(\omega, g_{\mathbf{C},P}(\omega)) - f(\omega, x^*_\omega)\right) P_{C_j}(d\omega) \\
&\overset{(b)}{=} \sum_{j=1}^r P(C_j) S(x^*_P, P_{C_j}) + \sum_{j=1}^r P(C_j) L(g_{\mathbf{C},P}, P_{C_j}) \\
&\overset{(c)}{=} S(x^*_P, P_{\mathbf{C}}) + L(g_{\mathbf{C},P}, P_{\mathbf{C}}) \overset{(d)}{=} S(x^*_P, P_{\mathbf{C}}) + L(g_{\mathbf{C},P}, P),
\end{aligned}$$

where (a) follows from the definition of the conditional measure P_{C_j}, (b) follows from the definitions of $S(x^*_P, P_{C_j})$ and $L(g, P_{C_j})$, (c) follows from the notational convention (5.4) for functionals of measures, and (d) follows from Lemma 5.2. $\square$

5.3 Effect of additional information on solution quality

5.3.1 Pseudo-energy-loss efficient frontier

Let us consider the set $\mathcal{G}$ of maps from Ω into X. Each map $g = (\mathbf{C}(g), I(g))$ from this set can be characterized by the corresponding loss $L(g, P)$ with respect to the original measure P and the value $G(\Omega, \mathbf{C}(g), P)$ – the difficulty of the corresponding question. We will be interested – for reasons that will become clear shortly – in finding the *efficient frontier* in the Euclidean plane with coordinates $(G(\Omega, \mathbf{C}(g), P), L(g, P))$. In other words, we will be looking for the set $\mathcal{O}$ of Pareto-optimal maps that can be found by solving the following parametric optimization problem

$$\begin{aligned} &\underset{g\in\mathcal{G}}{\text{minimize}} && L(g, P) \\ &\text{subject to} && G(\Omega, \mathbf{C}(g), P) \leq \gamma \end{aligned} \tag{5.12}$$

for all values of the parameter γ.

The first observation we can make is that to find the set $\mathcal{O}$ of Pareto-optimal maps it is sufficient to consider the set of subset-optimal maps $\mathcal{C}$ as the following proposition asserts.

Proposition 5.1 $\mathcal{O} \subset \mathcal{C}$

Proof: Let $g = (\mathbf{C}, I)$ where $I = \{x_1, x_2, \dots, x_r\}$. Suppose that $g \notin \mathcal{C}$. Then there exists at least one $C \in \mathbf{C}$ such that $g(C) \neq x^*_{P_C}$. Without loss of generality we can assume that $C = C_1$. Consider a different map $g' = (\mathbf{C}, I')$ such that $I' = \{x^*_{P_{C_1}}, x_2, \dots, x_r\}$. Obviously, $G(\Omega, \mathbf{C}(g')P) = G(\Omega, \mathbf{C}(g)P)$ (since $\mathbf{C}(g') = \mathbf{C}(g)$). On the other hand,

$$L(g', P) - L(g, P) = P(C_1)(L(g', P_{C_1}) - L(g, P_{C_1})) < 0,$$

since $L(g', P_{C_1})$ takes the minimum value among all maps with the same partition $\mathbf{C}$. We thus find that $L(g', P) < L(g, P)$ which means that $g \notin \mathcal{O}$. $\square$

It follows from Proposition 5.1 that one needs to look no further than the set $\mathcal{C}$ of subset optimal maps. Such maps are uniquely characterized by the corresponding partition $\mathbf{C}$ only (up to simple equivalences). Therefore the task of finding maps that belong to the set $\mathcal{C}$ is equivalent to that of finding the corresponding partitions of the set Ω.

5.3.2 Optimal information acquisition

Let us now address the optimal information acquisition problem (1.3): what question(s) need to be asked the given information source in order to obtain the minimum possible loss for (1.1). Given a question $\mathbf{C} = \{C_1, \dots, C_r\}$ to an information source and its answer $V(\mathbf{C})$ taking values in the set $\{s_1, \dots, s_m\}$, we denote by $\mathcal{L}(s_k)$, $k = 1, \dots, m$ the *minimum conditional expected loss* given that $V(\mathbf{C}) = s_k$ and by $\mathcal{L}(V(\mathbf{C}))$ the *minimum expected loss* that the decision maker can achieve given the answer $V(\mathbf{C})$. The latter can be found as

$$\mathcal{L}(V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k)\mathcal{L}(s_k), \tag{5.13}$$

i.e. as an expectation over possible values of the answer $V(\mathbf{C})$.

Clearly, if no answer was received – and the decision maker has to choose a solution $x \in X$ based on the original information only – the minimum expected loss will be equal to the EVPI of the original problem: $\mathcal{L}(\emptyset) = L(g_P, P)$.

If the decision maker poses a question $\mathbf{C} = \{C_1, \ldots, C_r\}$ to the information source and receives a particular value s_k of answer $V(\mathbf{C})$, the original measure P on Ω gets updated to $P^k \equiv P^{V(\mathbf{C})=s_k}$. Therefore in order to minimize loss for the given value s_k of answer $V(\mathbf{C})$ the decision maker needs to choose the solution $x^*_{P^k}$ – the solution minimizing the expectation $\mathbb{E}_{P^k} f(\omega, x)$ over all (feasible) values of x.

Perfect answers

First, let us assume that the information source can provide a perfect answer to $\mathbf{C}$. Then the following result can be obtained.

Proposition 5.2 *Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a complete question and $g_{\mathbf{C},P} \in \mathcal{C}$ be a corresponding subset optimal map. If the decision maker is given a perfect answer $V^*(\mathbf{C})$ to $\mathbf{C}$ then*

$$\mathcal{L}(V^*(\mathbf{C})) = L(g_{\mathbf{C},P}, P).$$

Proof: For the given value s_j of the answer, $P^j = P_{C_j}$, $j = 1, \ldots, r$. Therefore the decision maker can achieve the smallest possible loss by choosing the solution $x^*_{P_{C_j}}$. The resulting conditional loss will be

$$\mathcal{L}(s_j) = \int_{C_j} (f(\omega, x^*_{P_{C_j}}) - f(\omega, x^*(\omega)))\, dP_{C_j}(\omega). \tag{5.14}$$

Taking the expectation of (5.14) over possible values of the answer $V^*(\mathbf{C})$ we obtain

$$\begin{aligned}
\mathcal{L}(V^*(\mathbf{C})) &\overset{(a)}{=} \sum_{j=1}^{r} P(C_j)\mathcal{L}(s_j) = \sum_{j=1}^{r} P(C_j) \int_{C_j} (f(\omega, x^*_{P_{C_j}}) - f(\omega, x^*_\omega))\, dP_{C_j}(\omega) \\
&\overset{(b)}{=} \sum_{j=1}^{r} P(C_j) \int_{C_j} (f(\omega, g_{\mathbf{C},P}(\omega)) - f(\omega, x^*_\omega))\, dP_{C_j}(\omega) \\
&= \sum_{j=1}^{r} P(C_j) L(g_{\mathbf{C},P}, P_{C_j}) \overset{(c)}{=} L(g_{\mathbf{C},P}, P_{\mathbf{C}}) \overset{(d)}{=} L(g_{\mathbf{C},P}, P),
\end{aligned}$$

where (a) follows from that for a perfect answer consistent with the original measure, $\Pr(V^*(\mathbf{C}) = s_j) = P(C_j)$, (b) follows from that the map $g_{\mathbf{C},P}$ is subset optimal, (c) follows from the definition (5.8), and (d) follows from Lemma 5.2. $\square$

Combining the result of Proposition 5.2 with Lemma 5.2 (valid for any $g \in \mathcal{G}$) and Lemma 5.3 (valid for any $g \in \mathcal{C}$) we can find the value of the largest *loss reduction* due to a perfect answer to question $\mathbf{C}$. The result is formulated as a corollary.

Corollary 5.1 *Given a perfect answer to question* $\mathbf{C}$*, the largest possible reduction in expected loss a decision maker can achieve is equal to*

$$\mathcal{L}(\emptyset) - \mathcal{L}(V^*(\mathbf{C})) = B(g_{\mathbf{C},P}, P) = S(x^*_P, P_{\mathbf{C}}),$$

where $g_{\mathbf{C},P} \in \mathcal{C}$ *is a subset optimal map corresponding to question* $\mathbf{C}$*.*

Imperfect answers

Now, let us relax the assumption of availability of a perfect answer to question $\mathbf{C}$. Instead, we assume that the decision maker can obtain an answer $V(\mathbf{C})$ which is in general imperfect. First, we formulate a useful auxiliary result.

Lemma 5.4 *Let $V(\mathbf{C})$ be an answer to question $\mathbf{C}$ and let $g_{\mathbf{C},P} \in \mathcal{C}$ be a corresponding subset optimal map. Then*

$$S(x_P^*, P_{\mathbf{C}}) = S(x_P^*, P_{V(\mathbf{C})}) + B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}).$$

Proof:

$$
\begin{aligned}
S(x^*_P, P_{\mathbf{C}}) &= \sum_{j=1}^{r} P(C_j) S(x^*_P, P_{C_j}) = \sum_{j=1}^{r} P(C_j) \int_{C_j} (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega))) P_{C_j}(d\omega) \\
&\overset{(a)}{=} \sum_{j=1}^{r} \sum_{k=1}^{r} p_{kj} v_k \int_{C_j} (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega))) P_{C_j}(d\omega) \\
&\overset{(b)}{=} \sum_{j=1}^{r} \sum_{k=1}^{r} p_{kj} v_k \int_{\Omega} (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega))) P_{C_j}(d\omega) \\
&= \sum_{k=1}^{r} v_k \int_{\Omega} \sum_{j=1}^{r} p_{kj} (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega))) P_{C_j}(d\omega) \\
&\overset{(c)}{=} \sum_{k=1}^{r} v_k \int_{\Omega} (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega))) P^k(d\omega) \\
&= \sum_{k=1}^{r} v_k \int_{\Omega} (f(\omega, x^*_P) - f(\omega, g_{\mathbf{C},P}(\omega)) + f(\omega, x^*_{P^k}) - f(\omega, x^*_{P^k})) P^k(d\omega) \\
&= \sum_{k=1}^{r} v_k \int_{\Omega} (f(\omega, x^*_P) - f(\omega, x^*_{P^k})) P^k(d\omega) \\
&+ \sum_{k=1}^{r} v_k \int_{\Omega} (f(\omega, x^*_{P^k}) - f(\omega, g_{\mathbf{C},P}(\omega))) P^k(d\omega) \\
&\overset{(d)}{=} \sum_{k=1}^{r} v_k S(x^*_P, P^k) + \sum_{k=1}^{r} v_k B(g_{\mathbf{C},P}, P^k) \\
&\overset{(e)}{=} S(x^*_P, P_{V(\mathbf{C})}) + B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}),
\end{aligned}
$$

where (a) follows from (3.13), (b) follows from the fact that measure P_{C_j} vanishes outside of C_j, (c) follows from (3.1), (d) follows from the definitions (5.1) and (5.3) of suboptimality and gain, and (e) follows from the definitions (5.7) and (5.11). $\square$

Combining the result of Lemma 5.4 with that of Lemma 5.3, we obtain a useful decomposition of the EVPI of the original problem which we formulate as a corollary.

Corollary 5.2 *Let $V(\mathbf{C})$ be an answer to question $\mathbf{C}$ and $g_{\mathbf{C},P} \in \mathcal{C}$ a corresponding subset optimal* *map. Then*

$$L(g_P, P) = S(x^*_P, P_{V(\mathbf{C})}) + B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}) + L(g_{\mathbf{C},P}, P).$$

Now we can determine the minimum expected loss $\mathcal{L}(V(\mathbf{C}))$ that's obtainable with the help of an answer $V(\mathbf{C})$ to question $\mathbf{C}$. We state the result as a proposition.

Proposition 5.3 *Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a complete question and $g_{\mathbf{C},P} \in \mathcal{C}$ be a corresponding subset optimal map. If the decision maker is given a (generally imperfect) answer $V(\mathbf{C})$ to $\mathbf{C}$ then*

$$\mathcal{L}(V(\mathbf{C})) = B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}) + L(g_{\mathbf{C},P}, P).$$

Proof: The value s_k of answer $V(\mathbf{C})$ implies that the measure on Ω is equal to P^k. Therefore the the decision maker can achieve minimum loss by using the stochastic optimal solution $x^*_{P^k}$. The resulting minimum loss will be

$$\mathcal{L}(s_k) = L(g_{P^k}, P^k), \tag{5.15}$$

where g_{P^k} is the all-to-one map $g_{P^k}(\omega) = x^*_{P^k}$ for all $\omega \in \Omega$.

The minimum expected loss $\mathcal{L}(V(\mathbf{C}))$ can be obtained by substituting (5.15) into (5.13):

$$\mathcal{L}(V(\mathbf{C})) = \sum_{k=1}^{m} v_k L(g_{P^k}, P^k). \tag{5.16}$$

On the other hand, we can decompose the EVPI $L(g_P, P)$ as follows.

$$
\begin{aligned}
L(g_P, P) &= \int_\Omega (f(\omega, x_P^*) - f(\omega, x_\omega^*))P(d\omega) = \sum_{k=1}^m v_k \int_\Omega (f(\omega, x_P^*) - f(\omega, x_\omega^*))P^k(d\omega) \\
&= \sum_{k=1}^m v_k \int_\Omega (f(\omega, x_P^*) - f(\omega, x_\omega^*) + f(\omega, x_{P^k}^*) - f(\omega, x_{P^k}^*))P^k(d\omega) \\
&= \sum_{k=1}^m v_k \int_\Omega (f(\omega, x_P^*) - f(\omega, x_{P^k}^*))P^k(d\omega) \\
&+ \sum_{k=1}^m v_k \int_\Omega (f(\omega, x_{P^k}^*) - f(\omega, x_\omega^*))P^k(d\omega) \\
&= \sum_{k=1}^m v_k S(x_P^*, P^k) + \sum_{k=1}^m v_k L(g_{P^k}, P^k) \\
&= S(x_P^*, P_{V(\mathbf{C})}) + \sum_{k=1}^m v_k L(g_{P^k}, P^k) \qquad (5.17)
\end{aligned}
$$

Comparing (5.16) with (5.17) we can obtain

$$
\mathcal{L}(V(\mathbf{C})) = L(g_P, P) - S(x_P^*, P_{V(\mathbf{C})}). \qquad (5.18)
$$

Finally, using the decomposition of EVPI of Corollary 5.2 in (5.18) yields

$$
\mathcal{L}(V(\mathbf{C})) = B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}) + L(g_{\mathbf{C},P}, P).
$$

□

It is easy to see that, for perfect answer $V^*(\mathbf{C})$ to question $\mathbf{C}$, the gain $B(g_{\mathbf{C},P}, P_{V(\mathbf{C})})$ in Proposition 5.3 vanishes (since $B(g_{\mathbf{C},P}, P_{V^*(\mathbf{C})}) = B(g_{\mathbf{C},P}, P_{\mathbf{C}}) = 0$) and the result of Proposition 5.2 is recovered.

The amount of maximum reduction of loss due to answer $V(\mathbf{C})$ to question $\mathbf{C}$ can be obtained by combining the result of Proposition 5.3 with that of Corollary 5.2. The result is formulated as a corollary.

Corollary 5.3 *Given a (generally imperfect) answer to question* $\mathbf{C}$*, the largest possible reduction in expected loss a decision maker can achieve is equal to*

$$\mathcal{L}(\emptyset) - \mathcal{L}(V(\mathbf{C})) = S(x^*_P, P_{V(\mathbf{C})}).$$

5.3.3 Pseudo-energy-loss correspondence

Comparing results obtained in this section with the corresponding pseudo-energy values discussed in Chapters 2 and 3, we can make several interesting observations regarding their correspondence that reveal a rather clear picture. We assume that the measure P admits existence of a finest partition of Ω. Let $C_f(P)$ be such finest partition. We can then summarize the observations made in the previous sections as follows.

- The initial loss is equal to EVPI $L(g_P, P)$. In order to reduce it to zero, one needs to completely resolve the underlying uncertainty by answering the exhaustive question $C_f(P)$ about possible outcomes on Ω perfectly. The required pseudo-energy is equal to $G(\Omega, \mathbf{C}_f(P), P)$.
- A perfect answer to question $\mathbf{C}$ (that, as a partition, is some coarsening of $\mathbf{C}_f(P)$) requires $G(\Omega, \mathbf{C}, P)$ worth of pseudo-energy from an information source and allows the decision maker to reduce the loss by the amount equal to $S(x^*_P, P_{\mathbf{C}}) = B(g_{\mathbf{C},P}, P)$.
- If the source is able to produce only an imperfect answer $V(\mathbf{C})$ to question $\mathbf{C}$ the corresponding amount of pseudo-energy is equal to the answer depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$. Such an

answer can reduce the initial loss $L(g_P, P)$ by the amount of $S(x^*_P, P_{V(\mathbf{C})})$.

- The difference of depths (pseudo-energy contents) between a perfect and and imperfect answers to question $\mathbf{C}$ is equal to $G(\Omega, \mathbf{C}, P_{V(\mathbf{C})})$. The corresponding difference in loss reductions (values of information) is $B(g_{\mathbf{C},P}, P_{V(\mathbf{C})})$. The latter quantity can be naturally interpreted as a price the decision maker pays for imperfection of the answer he/she receives to question $\mathbf{C}$.

- Given a perfect answer to question $\mathbf{C}$, the residual pseudo-energy measuring the degree of difficulty of resolving the remaining uncertainty is equal to $G(\Omega, \mathbf{C}_f(P)_{\mathbf{C}}, P)$. The corresponding residual loss is simply $L(g_{\mathbf{C},P}, P)$.

- Given an imperfect answer to question $\mathbf{C}$, the residual pseudo-energy measuring the degree of difficulty of resolving the remaining uncertainty is equal to $G(\Omega, \mathbf{C}_f(P), P_{V(\mathbf{C})})$ – the difficulty of the exhaustive question $\mathbf{C}_f(P)$ given the answer $V(\mathbf{C})$ to question $\mathbf{C}$. The corresponding residual loss is equal to $\sum_{k=1}^{m} v_k L(g_{P^k}, P^k)$.

Table 5.1 shows the correspondence between pseudo-energy and loss related quantities discussed above. We see that for every loss related quantity there is a corresponding pseudo-energy quantity, meaning that in order to reduce the loss by a certain amount the corresponding pseudo-energy has to be made available in the form of an answer to some question. Depending on the structure of the question, the amount of loss reduction and, respectively, the amount of residual loss can vary in size. The goal of the decision maker is to find the specific question(s) that would maximize the effect of the given information source (characterized by its pseudo-temperature function and source model parameters such as capacity) on the given problem. More specifically, the decision maker would want to find the specific question $\mathbf{C}$ that would result in the smallest possible minimum expected loss $\mathcal{L}(V(\mathbf{C}))$ where $V(\mathbf{C})$ is the answer that the source can provide to question $\mathbf{C}$. Formally, this *information acquisition optimization* problem can be written as

$$\begin{aligned} &\underset{\mathbf{C}}{\text{minimize}} && \mathcal{L}(V(\mathbf{C})) \\ &\text{subject to} && Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = h(G(\Omega, \mathbf{C}, P)) \end{aligned} \tag{5.19}$$

where minimization is performed over all possible partitions of the parameter space Ω. The expression for the minimum loss $\mathcal{L}(V(\mathbf{C}))$ is given either by Proposition 5.2 (for perfect answers) or Proposition 5.3 (for imperfect answers).

If a source is capable of perfect answers (for instance, in the simple linear model) solution of problem (5.19) reduces to finding the efficient frontier: if $L^*(G)$ is the expression describing the efficient frontier (abstracting from its true discrete structure) and Y_s is the capacity of the information source, then the minimum in (5.19) is equal to $L^*(Y_s)$ and is achieved by the question $\mathbf{C}$ lying on the efficient frontier such that $G(\Omega, \mathbf{C}, P) = Y_s$.

If a source cannot provide perfect answers (likely a more realistic scenario), one would need to consider questions with difficulty exceeding the source capacity ($G(\Omega, \mathbf{C}, P) > Y_s$) in order to minimize the expected loss. The search for an optimal question in this case becomes somewhat more complicated as the error structure for the source's answers needs to be taken into account. If answers are assumed, for instance, to be quasi-perfect, optimal question(s) can be readily found approximately provided the efficient frontier is already known. An illustration is provided in the next section.

Pseudo-energy	Loss	Comments
$G(\Omega, \mathbf{C}_f(P), P)$	$L(g_P, P)$	exhaustive question difficulty/total initial loss (EVPI)
$G(\Omega, \mathbf{C}, P)$	$S(x^*_P, P_{\mathbf{C}}) = B(g_{\mathbf{C},P}, P)$	question difficulty/loss reduction due to perfect answer
$Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$	$S(x^*_P, P_{V(\mathbf{C})})$	answer depth/loss reduction due to that answer
$G(\Omega, \mathbf{C}, P_{V(\mathbf{C})})$	$B(g_{\mathbf{C},P}, P_{V(\mathbf{C})})$	residual difficulty/“price” of answer imperfection
$G(\Omega, \mathbf{C}_f(P)_{\mathbf{C}}, P)$	$L(g_{\mathbf{C},P}, P)$	residual pseudo-energy/loss given perfect answer to $\mathbf{C}$
$G(\Omega, \mathbf{C}_f(P), P_{V(\mathbf{C})})$	$\sum_{k=1}^m v_k L(g_{P^k}, P^k)$	residual pseudo-energy/loss given an imperfect answer to $\mathbf{C}$

Table 5.1: *Correspondence between pseudo-energy and loss related quantities.*

The correspondence between pseudo-energy and loss quantities shown in Table 5.1 can be illustrated by comparing decompositions of the exhaustive question difficulty $G(\Omega, \mathbf{C}_f(P), P)$ (expression (5.20)) and the EVPI $L(g_P, P)$ (expression (5.21)) on the other hand. It is also shown in Fig. 5.1.

$$\underbrace{Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) + \overbrace{G(\Omega, \mathbf{C}, P_{V(\mathbf{C})})}_{G(\Omega,\mathbf{C},P)} + G(\Omega, \mathbf{C}_f(P)_{\mathbf{C}}, P)}^{G(\Omega,\mathbf{C}_f(P),P_{V(\mathbf{C})})} = G(\Omega, \mathbf{C}_f(P), P) \tag{5.20}$$

$$\underbrace{S(x^*_P, P_{V(\mathbf{C})}) + \overbrace{B(g_{\mathbf{C},P}, P_{V(\mathbf{C})})}_{S(x^*_P,P_{\mathbf{C}})=B(g_{\mathbf{C},P},P)} + L(g_{\mathbf{C},P}, P)}^{\sum_{k=1}^m v_k L(g_{P^k},P^k)} = L(g_P, P) \tag{5.21}$$

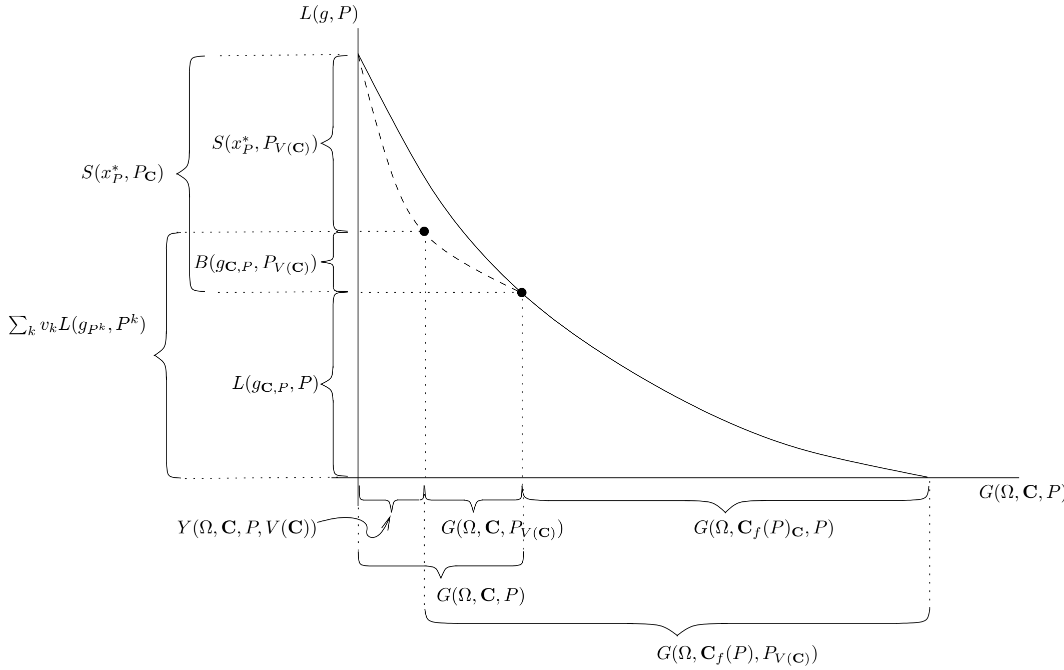

Figure 5.1: *The efficient frontier and correspondence between pseudo-energy and objective function (loss) quantities. A Pareto-optimal map $g \in \mathcal{O}$ on the efficient frontier is shown.*

5.4 Examples

5.4.1 Toy example

To illustrate the concepts introduced in previous sections, let us consider a very simple example. Let Ω be the interval $[0, a]$ and let X be the real line $\mathbb{R}$. Let the integrand $f(\omega, x)$ have the following form: $f(\omega, x) = (x - \omega)^2$ and let the original measure P be the uniform continuous distribution on $[0, a]$.

It is obvious that the optimal solution for the given realization ω is simply $x^*_\omega = \omega$. The stochastic optimal map is $g_P(\omega) = \frac{a}{2} \in X$ for all $\omega \in \Omega$. Therefore the EVPI of the problem (1.1) is

$$L(g_P, P) = \frac{1}{a}\int_0^a \left((x^*_P - \omega)^2 - (x^*_\omega - \omega)^2\right) d\omega = \frac{1}{a}\int_0^a \left(\frac{a}{2} - \omega\right)^2 d\omega = \frac{a^2}{12}.$$

Let $\mathbf{C} = \left\{\left[0, \frac{a}{2}\right), \left[\frac{a}{2}, a\right]\right\}$ and $\mathbf{C}' = \left\{\left[0, \frac{a}{4}\right) \cup \left[\frac{a}{2}, \frac{3a}{4}\right), \left[\frac{a}{4}, \frac{a}{2}\right) \cup \left[\frac{3a}{4}, a\right]\right\}$ be two $r = 2$ partitions of Ω. Let us consider several different $r = 2$ maps $g \in \mathcal{G}$ (see Fig. 5.2 for an illustration).

- $g_1 = \left(\mathbf{C}, \left\{\frac{a}{4}, \frac{3a}{4}\right\}\right) = g_{\mathbf{C},P}$. The measures P_{C_1} and P_{C_2} are uniform on C_1 and C_2 respectively. We have $x^*_{P_{C_1}} = \frac{a}{4}$ and $x^*_{P_{C_2}} = \frac{3a}{4}$. Thus $g_1 \in \mathcal{C}$. Note that in this case $g_1 \in \mathcal{O}$ as well as it lies on the efficient frontier in (G, L) coordinate plane (see Fig. 5.3 for an illustration).

- $g_2 = (\mathbf{C}, \{0, a\})$. For this map, the partition is the same as that for g_1, but the image set is different. This map is therefore not subset-optimal: $g_2 \notin \mathcal{C}$.

- $g_3 = \left(\mathbf{C}', \left\{\frac{3a}{8}, \frac{5a}{8}\right\}\right) = g_{\mathbf{C}',P}$. For this map's partition both subsets C_1' and C_2' consist of two connected components. It is easy to check that $x^*_{P_{C_1}} = \frac{3a}{8}$ and $x^*_{P_{C_2}} = \frac{5a}{8}$ and thus $g_3 \in \mathcal{C}$.

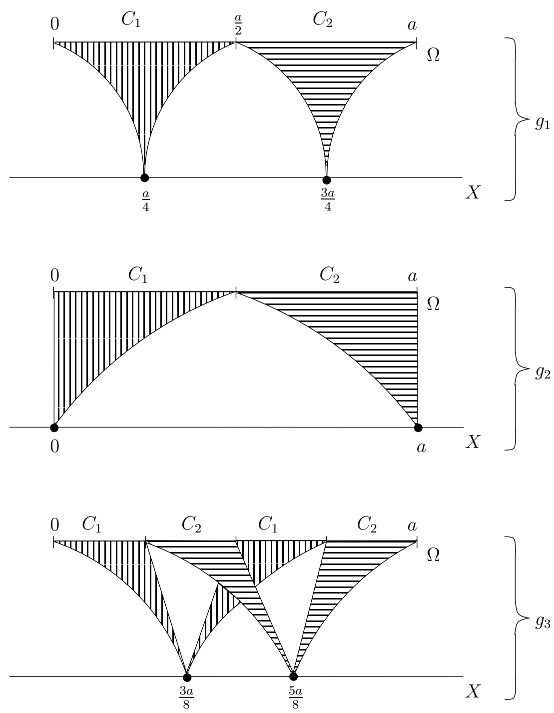

Figure 5.2: *Maps g_1, g_2 and g_3. The partitions for g_1 and g_2 consist of connected sets only. Each element of the partition for g_3 consists of two connected sets.*

The loss for these three maps can be found as follows. For g_1,

$$L(g_1, P) = \frac{1}{2} \cdot \frac{2}{a} \int_0^{a/2} \left(\frac{a}{4} - \omega\right)^2 d\omega + \frac{1}{2} \cdot \frac{2}{a} \int_{a/2}^{a} \left(\frac{3a}{4} - \omega\right)^2 d\omega = \frac{a^2}{48},$$

for g_2,

$$L(g_2, P) = \frac{1}{2} \cdot \frac{2}{a} \int_0^{a/2} (0 - \omega)^2 d\omega + \frac{1}{2} \cdot \frac{2}{a} \int_{a/2}^{a} (1 - \omega)^2 d\omega = \frac{a^2}{12},$$

and for g_3,

$$\begin{aligned} L(g_3, P) &= \frac{1}{2} \cdot \frac{2}{a} \left(\int_0^{a/4} \left(\frac{3a}{8} - \omega\right)^2 d\omega + \int_{a/2}^{3a/4} \left(\frac{3a}{8} - \omega\right)^2 d\omega \right) \\ &\quad + \frac{1}{2} \cdot \frac{2}{a} \left(\int_{a/4}^{a/2} \left(\frac{5a}{8} - \omega\right)^2 d\omega + \int_{3a/4}^{a} \left(\frac{5a}{8} - \omega\right)^2 d\omega \right) = \frac{13a^2}{192}. \end{aligned}$$

Fig. 5.3 shows the efficient frontier and maps g_1, g_2 and g_3 in (G, L) coordinate plane. We see that $g_1 \in \mathcal{O}$ lies on the efficient frontier while g_2 and g_3 are located above it.

Since $g_1, g_3 \in \mathcal{C}$ we have (as Lemma 5.3 states) $S(x_P^*, P_{\mathbf{C}}) = \frac{a^2}{12} - \frac{a^2}{48} = \frac{a^2}{16}$ for g_1 and $S(x_P^*, P_{\mathbf{C}'}) = \frac{a^2}{12} - \frac{13a^2}{192} = \frac{a^2}{64}$ for g_3. For g_2, the suboptimality is the same as that for g_1. Note that, since $g_2 \notin \mathcal{C}$, $S(x_P^*, P_{\mathbf{C}}) + L(g_3, P) = \frac{7a^2}{48} \neq L(g_P, P)$.

For this one-dimensional example it turns out to be straightforward to find maps on the efficient frontier. Indeed, it is obvious that partitions for such maps have to consist of connected sets only. It is also clear that the order in which subsets C_j appear on the interval $[0, a]$ does not matter because the integrand in (1.1) $f(\omega, x)$ depends on $|\omega - x|$ only. So, for the fixed value of r, any map $g \in \mathcal{C}$ that can lie on the efficient frontier can be uniquely characterized by the subset measures $w_j = P(C_j)$, $j = 1, \ldots, r$. Given the values w_j, the expected loss of the corresponding map can

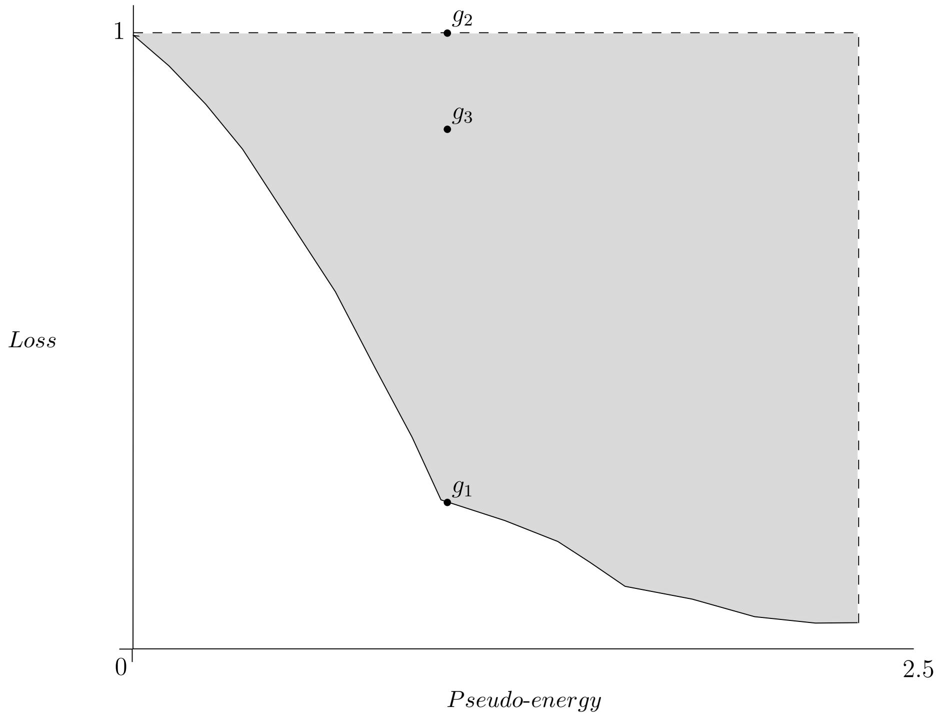

Figure 5.3: *Maps g_1, g_2 and g_3 on (G, L) coordinate plane. All possible maps for this problem lie in the shaded region, at or above the efficient frontier.*

be written as

$$L(g, P) = \sum_{j=1}^{r} w_j \frac{(w_j a)^2}{12} = \frac{a^2}{12} \sum_{j=1}^{r} w_j^3.$$

In order to find the optimal values of w_j yielding the smallest loss for the question difficulty $G(\Omega, \mathbf{C}, P)$ not exceeding h the following optimization problem needs to be solved.

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{r} w_j^3 \\ \text{subject to} \quad & -\sum_{j=1}^{r} u(C_j) w_j \log w_j \le h \\ & \sum_{j=1}^{r} w_j = 1 \\ & w_j \ge 0, \quad j = 1, \dots, r, \end{aligned} \tag{5.22}$$

where $u(C_j)$ is the pseudo-temperature of subset C_j and h is a nonnegative parameter. Since the function $-\sum_{j=1}^{r} u(C_j) w_j \log w_j$ is concave, (5.22) is a global optimization problem. However it can easily be solved to optimality for moderate values of the partition size r. We consider two cases: constant pseudo-temperature function $u(\omega) \equiv 1$ and linear pseudo-temperature $u(\omega) = \frac{2}{a}\omega$. We can assume that $C_j = [a\tilde{w}_j, a(\tilde{w}_j + w_j)]$. In the former case, $u(C_j) = 1$, $j = 1, \ldots, r$ and in the latter case,

$$u(C_j) = 2\tilde{w}_j + w_j, \tag{5.23}$$

where $\tilde{w}_j = \sum_{l=1}^{j-1} w_l$ if $j > 1$ and $\tilde{w}_1 = 0$.

The resulting efficient frontier is shown in Fig. 5.4.

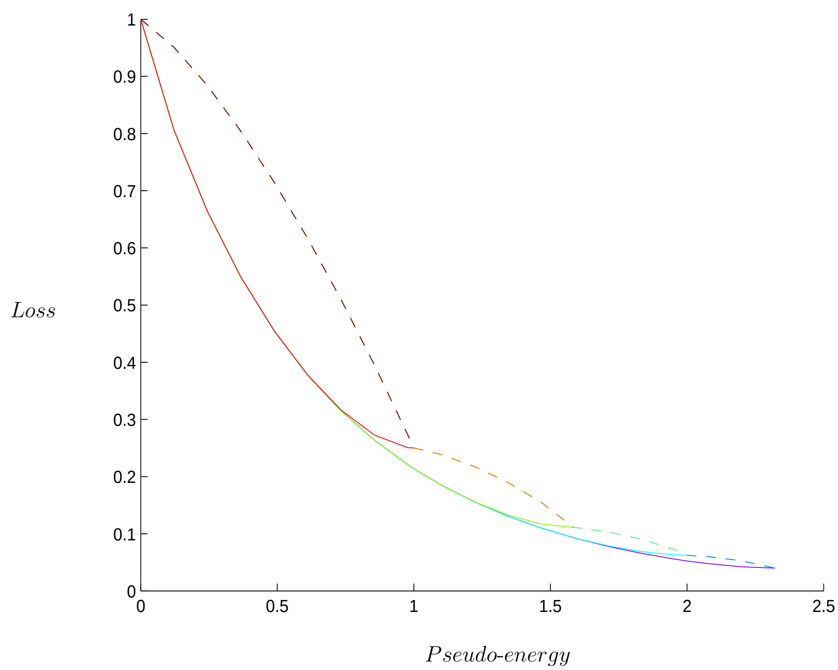

Figure 5.4: *Efficient frontier for the toy example: constant pseudo-temperature case (dotted line) and linear pseudo-temperature case (solid line).*

Let us now consider imperfect answers to questions **C** in the same example. For simplicity, we set $r = 2$ for questions and assume the pseudo-temperature to be constant on Ω. We also assume all answers to be quasi-perfect so that the updated measures P^k, $k = 1, 2$ have the form (3.21).

The stochastic optimal solutions $x^*_{P^k}$ for measures P^k can be found as

$$x^*_{P^k} = \arg\min_x \int_\Omega f(\omega, x) P^k(d\omega).$$

We have

$$\begin{aligned} x^*_{P^1} &= \arg\min_x \left(\frac{1-\alpha(1-w_1)}{w_1 a} \int_0^{w_1 a} (x-\omega)^2 d\omega + \frac{\alpha}{a} \int_{w_1 a}^a (x-\omega)^2 d\omega \right) \\ &= \frac{1}{2}(w_1 a + \alpha(1-w_1)a) = \frac{1}{2} a(w_1 + \alpha w_2), \end{aligned}$$

and, analogously,

$$x^*_{P^2} = \frac{1}{2} a(w_2 + \alpha w_1).$$

We can now find the suboptimalities:

$$\begin{aligned} S(x^*_P, P^1) &= \int_\Omega \left(f(\omega, x^*_P) - f(\omega, x^*_1(\alpha)) \right) P_1^{(\alpha)}(d\omega) \\ &= \frac{a^2}{12} \left((3 - 6w_1 + 3w_1^2)(1+\alpha^2) + \alpha(-6 + 12w_1 - 6w_1^2) \right), \end{aligned}$$

and, analogously,

$$S(x^*_P, P^2) = \frac{a^2}{12} \left((3 - 6w_2 + 3w_2^2)(1+\alpha^2) + \alpha(-6 + 12w_2 - 6w_2^2) \right).$$

The suboptimality $S(x^*_P, P_{V(\mathbf{C})})$ is then

$$
\begin{aligned}
S(x^*_P, P_{V(\mathbf{C})}) &= w_1 S(x^*_P, P_1^{(\alpha)}) + w_2 S(x^*_P, P_2^{(\alpha)}) \\
&= \frac{a^2}{12}(1 - w_1^3 - w_2^3)(1-\alpha)^2.
\end{aligned}
$$

The new value of the expected loss, according to Corollary 5.2, is

$$
\begin{aligned}
L(g_P, P) - S(x^*_P, P_{V(\mathbf{C})}) &= \frac{a^2}{12} - \frac{a^2}{12}(1 - w_1^3 - w_2^3)(1-\alpha)^2 \\
&= \frac{a^2}{12}\left(1 - (1 - w_1^3 - w_2^3)(1-\alpha)^2\right) \qquad (5.24)
\end{aligned}
$$

Note that for $\alpha = 0$ we recover the expression $L(g_{\mathbf{C},P}, P) = \frac{a^2}{12}(w_1^3 + w_2^3)$ for a perfect answer and for $\alpha = 1$ the new value of the loss is simply $L(g_P, P) = \frac{a^2}{12}$ since $\alpha = 1$ describes the case in which the answer $V(\mathbf{C})$ carries no new information and the updated measure is simply P.

Fig. 5.5 shows the dependence of the expected loss (5.24) on answer depth with the error parameter α ranging from 0 to 1 for several values of subset measures w_1 and w_2 for the $r = 2$ case. The part of the efficient frontier that can be achieved for $r = 2$ is also shown (solid bold line). It is interesting to observe that, for the same amount of pseudo-energy, lower values of the expected loss can be achieved with imperfect answers to more difficult questions.

5.4.2 Inventory example

A company has to decide on the order quantity x of a certain product and is required to satisfy an uncertain demand ω. The cost of ordering is $c > 0$ per unit of product. If the demand is larger than the ordered quantity, the shortage has to be covered by back ordering at a higher cost $b > c$. If the

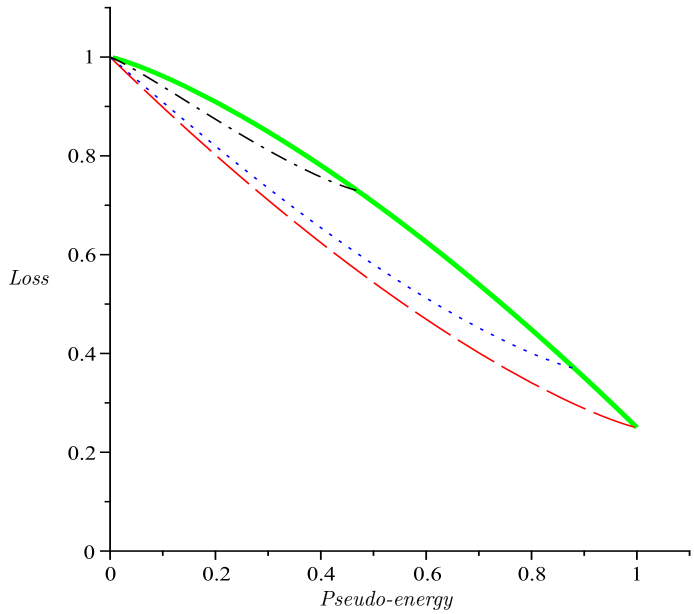

Figure 5.5: *Dependence of the expected loss on the added information for $r = 2$ partitions. The solid curve corresponds to the error-free message case with w_1 varying from 0 to 0.5. The dashed line shows the $w_1 = w_2 = 0.5$ case with α varying from 1 to 0 (from left to right on the figure). The dotted line is the same for $w_1 = 1 - w_2 = 0.7$ case, and the dash-dotted line is for $w_1 = 1 - w_2 = 0.9$ case.*

demand turns out to be lower than the ordered quantity, the extra units are held in storage at unit cost of $h > 0$. Thus the total cost has the form

$$f(\omega, x) = cx + b[x - \omega]_+ + h[\omega - x]_+, \tag{5.25}$$

where $[y]_+ = \max\{y, 0\}$ for any real y. We assume that both x and ω are continuous variables, for convenience. It is well-known that if the measure on the parameter space Ω is described by a cdf $F(\cdot)$ then the optimal solution of the problem

$$\min_x \mathbb{E}_P f(\omega, x), \tag{5.26}$$

is given by $x_P^* = F^{-1}\left(\frac{b-c}{b+h}\right)$.

Let us assume that the probability measure P is uniform on $\Omega = [0, a]$. Then, clearly, $x_P^* = a\frac{b-c}{b+h}$ (and therefore $g_P(\omega) = a\frac{b-c}{b+h}$ for all $\omega \in \Omega$). Consider partitions of Ω such that $P(C_j) = w_j$, $j = 1, \ldots, r$ and all sets C_j are connected. Just like in the previous example, we can assume, without loss of generality that $C_j = [a\tilde{w}_j, a(\tilde{w}_j + w_j)]$, where $\tilde{w}_j = \sum_{l=1}^{j-1} w_l$ if $j > 1$ and $\tilde{w}_1 = 0$.

It is straightforward to show that the EVPI of this problem is

$$L(g_P, P) = \frac{a}{2} \cdot \frac{(b-c)(c+h)}{b+h}.$$

and, for the partition $\mathbf{C} = \{C_1, \ldots, C_r\}$, $x^*_{P_{C_j}} = a\left(\tilde{w}_j + w_j\frac{b-c}{b+h}\right)$, and

$$\begin{aligned} L(g_{\mathbf{C},P}, P) = L(g_{\mathbf{C},P}, P_{\mathbf{C}}) = \sum_{j=1}^{r} P(C_j) L(g_{\mathbf{C},P}, P_{C_j}) = \sum_{j=1}^{r} w_j \frac{a w_j}{2} \cdot \frac{(b-c)(c+h)}{b+h} \\ = \frac{a}{2} \cdot \frac{(b-c)(c+h)}{b+h} \sum_{j=1}^{r} w_j^2 = \left(\sum_{j=1}^{r} w_j^2\right) L(g_P, P) \end{aligned}$$

The efficient frontier, just like in the previous example can be found by solving the optimization problem (5.22). Fig. 5.6 shows the efficient frontier for the case of constant pseudo-temperature function which leads to $u(C_j) = 1$ for $j = 1, \ldots, r$ and for the case of linear increasing pseudo-temperature function $u(\omega) = \frac{2}{a}\omega$ which leads to $u(C_j) = 2\tilde{w}_j + w_j$, $j = 1, \ldots, r$.

Let us now consider quasi-perfect answers $V_\alpha(\mathbf{C})$ to question $\mathbf{C}$ with partitions $\mathbf{C}$ as described before. Consider the case $r = 2$ only, for simplicity. Then $C_1 = [0, w_1 a]$ and $C_2 = [w_1 a, a]$. The

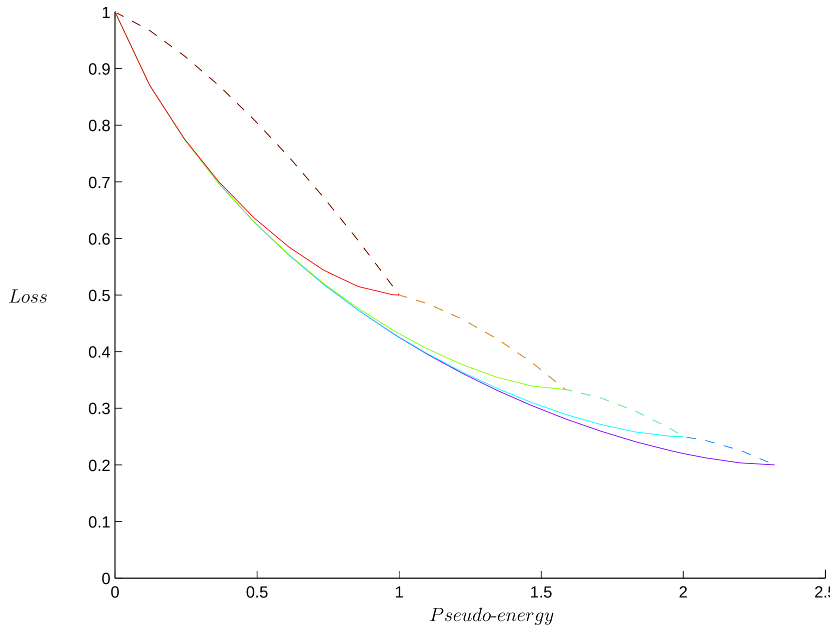

Figure 5.6: *Efficient frontier for the inventory example: constant pseudo-temperature case (dotted line) and linear increasing pseudo-temperature case (solid line).*

optimal solutions to (5.26) with the original measure P replaced with P^k can be shown to be

$$x^*_{P^1} = \begin{cases} \frac{w_1 a}{1-\alpha w_2} \cdot \frac{b-c}{b+h} & \text{if } \alpha < \frac{1}{w_2} \cdot \frac{c+h}{b+h} \\ \frac{a}{\alpha}\left(\alpha - \frac{c+h}{b+h}\right) & \text{if } \alpha \geq \frac{1}{w_2} \cdot \frac{c+h}{b+h}, \end{cases} \tag{5.27}$$

and

$$x^*_{P^2} = \begin{cases} a\left(1 - \frac{w_2}{1-\alpha w_1} \cdot \frac{c+h}{b+h}\right) & \text{if } \alpha < \frac{1}{w_1} \cdot \frac{b-c}{b+h} \\ \frac{a}{\alpha} \cdot \frac{b-c}{b+h} & \text{if } \alpha \geq \frac{1}{w_1} \cdot \frac{b-c}{b+h}. \end{cases} \tag{5.28}$$

The suboptimalities $S(x^*_P, P^k)$ for $k = 1, 2$ can then be calculated. The resulting expressions are too lengthy (and not very illuminating) to be given here. The resulting loss can be found as

$$B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}) + L(g_{\mathbf{C},P}, P) = L(g_P, P) - S(x^*_P, P_{V(\mathbf{C})}), \tag{5.29}$$

and the pseudo-energy content of answer $V_\alpha(\mathbf{C})$ is simply $Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C}))$ given by (3.22). Let us set, for definiteness, $c = 1$, $b = 1.5$, $h = 0.1$ and $a = 100$. Then the EVPI of the original problem is $L(g_P, P) = 17.19$. Let us also consider two information sources, described by the modified linear model, with equal capacity of $Y_s = 0.2$ (in the average unit pseudo-temperature calibration) and same value of parameter $b = 0.8$. The first source is characterized by a constant pseudo-temperature function $u(\omega) \equiv 1$ and the second has linear increasing pseudo-temperature $u(\omega) = \frac{2}{a} \cdot \omega$. The second source can be said to have relatively more "knowledge" about lower values of possible demand.

We are interested in finding, for each source, an $r = 2$ question $\mathbf{C} = \{C_1, C_2\}$ an answer to which would help the decision maker minimize the expected loss. This can easily be done numerically, for example, by graphing the loss (5.29) against the answer depth $Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C}))$, for different questions $\mathbf{C}$ (in this case, uniquely characterized by a single parameter w_1). It turns out (see Fig. 5.7 for an illustration) that the minimum loss at $Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})) = Y_s = 0.2$ is achieved for $w_1 = 0.25$ for the first source and $w_1 = 0.21$ for the second source. The minimum loss itself turns out to be equal to $\mathcal{L}(V(\mathbf{C})) = 15.48$ for the first source and $\mathcal{L}(V(\mathbf{C})) = 13.27$ for the second source, representing, respectively, 10% and 23% loss reduction from the original EVPI of 17.19. Clearly, the reason the second source is able to help the decision maker significantly more is that the latter is capable of utilizing the particular "expertise" of the second source by asking a question that is easy for the source and thus can be answered relatively well (with error probability $\alpha = 0.21$). On the other hand, the first source answers its "best" question with error probability of $\alpha = 0.56$ which results – expectedly – in a lower loss reduction. Note that the difficulty of the optimal question is equal to 0.80 for the first source and 0.41 for the second source, while the depth of the respective answer is equal to 0.2 (the source's capacity) in both cases. Note also that, in the modified linear model, a source can provide an answer of depth equal to capacity Y_s whenever the question difficulty exceeds the value Y_s/b, i.e. the question has to be sufficiently difficult for the source so that the latter can provide an answer of maximum depth.

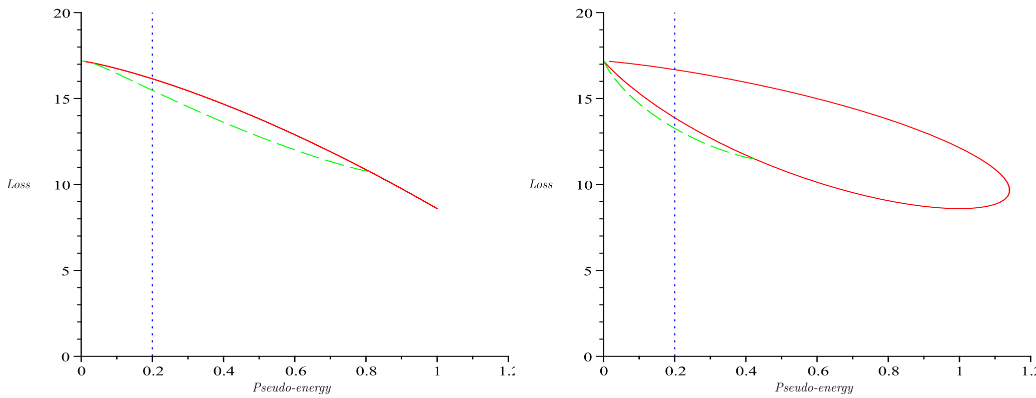

Figure 5.7: *Loss vs. pseudo-energy (for $r = 2$ questions only) for a source with constant pseudo-temperature (left) and a source with linear increasing pseudo-temperature (right). On both plots, the solid line is obtained by varying the parameter w_1 from 0 to 1. The dashed line is obtained by fixing a value of w_1 and varying α from 0 to 1. The value of w_1 (characterizing the optimal question) is chosen so that the point of intersection of the dashed line and the vertical dotted line (source capacity) has the lowest possible value of the vertical coordinate. The latter is equal to the minimum expected loss $\mathcal{L}(V(\mathbf{C}))$.*

5.5 Conclusion

In this chapter, we built on results obtained in Part I and explored the relationship between the pseudo-energy content of information sources' answers to decision maker's questions and the resulting minimum loss the decision maker can achieve for the problem being solved. For this purpose, we studied maps from the problem parameter space Ω to its set X of feasible solutions. We defined and studied several functionals of such maps, elements of the feasible solution set and probability measures on the parameter space. It was shown that the minimum loss the decision maker can achieve upon reception of an information source's answer to a certain question can be expressed via these functionals. On the other hand, the pseudo-energy content of such answers can be obtained if the source characteristics (such as pseudo-temperature function) are known. Therefore, to each answer there corresponds a point in the pseudo-energy – loss coordinate plane and the problem of optimal information acquisition can in principle be solved by finding – among all answers to all possible questions – the answer (and the corresponding question) that would yield the minimum loss but have depth not exceeding the source's capacity (so that the source can actually provide this answer). This problem appears to be rather complicated and it appears to be easier to begin from a search for a subset of Pareto-optimal questions, i.e. questions that lie on the efficient frontier in the pseudo-energy – loss coordinate plane. Put slightly differently, we imagine that a source can provide a perfect answer to each question and search for questions that would give the minimum loss value for each value of imaginary source capacity. If such efficient frontier is found, an optimal question (the answer to which for the *given* source would yield the smallest loss) can be found approximately by considering questions on the efficient frontier with difficulties of least the source capacity.

Thus the problem of additional information acquisition optimization reduces to that of finding question that lie on the efficient frontier in the pseudo-energy – loss coordinate plane. It appears that the latter problem is too complex to be solved exactly for any realistic size problem. Fortunately, it turns out that methods based on probability metrics that were used in scenario reduction approaches to stochastic optimization can be of use for approximate efficient frontier determination as well.

Chapter 6

Solution Methods

6.1 Introduction

In this chapter, approximate solution methods for the problem of optimal information acquisition are developed making use of the method of probability metrics and its application to scenario reduction in stochastic optimization.

6.2 Information Acquisition Optimization

In the following, we assume that the (initial) probability measure P is supported at a discrete set $\{\omega_1, \ldots, \omega_N\} \equiv \Omega_N \subset \Omega$:

$$P = \sum_{i=1}^{N} p_i \delta_{\omega_i}, \tag{6.1}$$

where δ_ω is a Dirac delta that puts a unit mass at ω. Points $\omega_i \in \Omega_N$ are usually referred to as *scenarios*. The *scenario reduction* methodology (see Appendix C) is often used in stochastic optimization to lower computational complexity of various practically important problems. In scenario reduction approach, the original discrete measure P given by (6.1) is said to be *reduced* to another discrete measure Q given by

$$Q = \sum_{j=1}^{M} q_j \delta_{\tilde{\omega}_j}, \tag{6.2}$$

if the support $\{\tilde{\omega}_1, \dots, \tilde{\omega}_M\}$ of Q is a subset of Ω_N.

For later convenience, we denote by $\mathcal{R}_M(\Omega_N)$ the set of all *scenario reduction maps* from the set of measures of the form (6.1) supported at Ω_N into the set of all measures of the form (6.2) supported at some subset of Ω_N of cardinality $M < N$ satisfying the additional property that we call *simplicity*. A map $\nu \in \mathcal{R}_M(\Omega_N)$ is called *simple* if there exists a partition $\{S_1, \dots, S_M\}$ of the set of scenarios Ω_N such that $\nu(\omega_i) = \tilde{\omega}_j$ for all $\omega_i \in S_j$ and $q_j = \sum_{\{i:\omega_i \in S_j\}} p_i$. In such a case we write $Q = \nu(P)$ and $S_j = \nu^{-1}(\tilde{\omega}_j)$ for $j = 1, \dots, M$.

Additionally, if c: $\Omega \times \Omega \to \mathbb{R}_+$ is some symmetric cost function, we call a map $\nu \in \mathcal{R}_M(\Omega_N)$ *c-optimal* if $c(\omega_i, \nu(\omega_i)) \leq c(\omega_i, \nu(\omega_j))$, $\forall i$, $j \neq i$. It is shown in [29] that the Monge-Kantorovich functional (see Appendix A) $\hat{\mu}_c(P, Q)$ is minimized for all measures Q supported at $\{\tilde{\omega}_1, \dots, \tilde{\omega}_M\} = \nu(\Omega_N)$ if the corresponding simple scenario reduction map is c-optimal.

In the following we call measures P and Q $\mathbf{C}$-*equivalent* for some partition $\mathbf{C}$ of Ω if $P(C) = Q(C)$ for all $C \in \mathbf{C}$. It is easy to see that measures P and Q are $\mathbf{C}$-equivalent for all possible partitions $\mathbf{C}$ if and only if $P = Q$ but two distinct measures can easily be $\mathbf{C}$-equivalent for a specific partition $\mathbf{C}$. In particular, any two measures on Ω are $\mathbf{C}$-equivalent if $\mathbf{C}$ is the trivial partition $\mathbf{C} = \{\Omega\}$.

Given measure P on Ω and some measure Q that was obtained from P by a reduction, let us denote by $\mathcal{Q}(Q|P)$ the *virtual pseudo-energy* content of measure Q relative to P. It is defined as follows

$$\mathcal{Q}(Q|P) = G(\Omega, \mathbf{C}_f(P), P) - G(\Omega, \mathbf{C}_f(Q), Q), \tag{6.3}$$

i.e. $\mathcal{Q}(Q|P)$ is the difference between the difficulties of exhaustive questions associated with measures P and Q, respectively. One can think about the virtual pseudo-energy of Q relative to P as an amount pseudo-energy a source would need to supply in order to obtain a new state in which the hardest possible question has a difficulty equal to $G(\Omega, \mathbf{C}_f(Q), Q)$. Since no question is in fact answered in going from measure P to the reduced measure Q we call this pseudo-energy virtual.

We can now introduce the *virtual difficulty* of question $\mathbf{C}$ for measure Q with respect to measure P:

$$G_P(\Omega, \mathbf{C}, Q) = \mathcal{Q}(Q|P) + G(\Omega, \mathbf{C}, Q). \tag{6.4}$$

In particular, $G_P(\Omega, \mathbf{C}, P) = G(\Omega, \mathbf{C}, P)$, i.e. the virtual difficulty of $\mathbf{C}$ for measure P relative to P reduces just to the standard difficulty of $\mathbf{C}$.

It also turns out to be useful to introduce the *relative expected loss* for partitions of Ω and measures Q obtained from the original measure P by a (simple) scenario reduction operation. In other words, we assume that there exists $\nu \in \mathcal{R}_M(\Omega_N)$ for some value of $M < N$ such that $Q = \nu(P)$. The relative (to measure P) expected loss of partition $\mathbf{C}$ and measure Q is then defined as follows.

$$L_P(\mathbf{C}, Q) = \sum_{C \in \mathbf{C}} P(C) L(g_{\mathbf{C},Q}, P), \tag{6.5}$$

where $g_{\mathbf{C},Q}$ is the subset-optimal map for partition $\mathbf{C}$ and measure Q. In particular, if $\mathbf{C}$ is the trivial partition $\mathbf{C} = \{\Omega\}$, the loss of Q relative to P is simply[1] $L_P(Q) = L(g_Q, P)$. If the measure Q coincides with P, the loss relative to P is just the standard expected loss of the corresponding subset-optimal map: $L_P(\mathbf{C}, P) = L(g_{\mathbf{C},P}, P)$.

Let us now consider the following construction. Reduce the original measure P to Q that is supported at r points: $Q = \nu(P)$, where $\nu \in \mathcal{R}_r(\Omega_N)$. Let $Q = \sum_{j=1}^{r} q_j \delta_{\tilde{\omega}_j}$ and let S_j the preimage of $\tilde{\omega}_j$ under map ν: $\nu(\omega_i) = \tilde{\omega}_j$ for all $\omega_i \in S_j$. Then let $\mathbf{C}$ be a partition of Ω such that $S_j \subset C_j$ for $j = 1, \dots, r$. We say that the partition $\mathbf{C}$ is *generated* by the map $\nu \in \mathcal{R}_r(\Omega_N)$, or, equivalently by the reduction of measure P to Q. Let $\hat{\mathbf{C}}$ be an arbitrary coarsening of $\mathbf{C}$.

We are interested in the location of points P, Q, $(\mathbf{C}, P)$, $(\mathbf{C}, Q)$, $(\hat{\mathbf{C}}, P)$ and $(\hat{\mathbf{C}}, Q)$ on the plane with coordinates $(G_P(\Omega, \cdot), L_P(\cdot))$. First of all, it is clear that $G_P(\Omega, P) = 0$ and $L_P(P) = L(g_P, P)$ where $L(g_P, P)$ is the EVPI of problem (1.1). Second, it is also clear that

$$
\begin{aligned}
G_P(\Omega, \mathbf{C}, Q) &= \mathcal{Q}(Q|P) + G(\Omega, \mathbf{C}, Q) \\
&= G(\Omega, \mathbf{C}_f(P), P) - G(\Omega, \mathbf{C}_f(Q), Q) + G(\Omega, \mathbf{C}, Q) \\
&= G(\Omega, \mathbf{C}_f(P), P)
\end{aligned} \tag{6.6}
$$

since $\mathbf{C} = \mathbf{C}_f(Q)$ by construction of Q. In words, the virtual difficulty of the question $\mathbf{C}$ for measure Q where the partition $\mathbf{C}$ was generated by a reduction of the original measure P to Q is equal to the difficulty of the exhaustive question for the original measure P.

To obtain relationships between relative expected losses the following two auxiliary lemmas are needed.

Lemma 6.1 *Let $c_{ij} = c_{ji}$, $i, j = 1, \dots, N$ be a symmetric matrix with elements c_{ij} satisfying the*

[1]Here and later we omit the trivial partition from the list of arguments of $G(\cdot)$ and $L(\cdot)$.

triangle inequality $c_{ij} \leq c_{ik} + c_{kj}$. *Let* $\{p_i\}_{i=1}^N$ *be a probability distribution. Then*

$$\sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{ij} \leq 2 \min_i \sum_{j=1}^N p_j c_{ij}.$$

Proof: Let $i^* = \arg\min_i \sum_{j=1}^N p_j c_{ij}$ (so that $\min_i \sum_{j=1}^N p_j c_{ij} = \sum_{j=1}^N p_j c_{i^*j}$). Then we can write

$$\begin{aligned}
\sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{ij} &\overset{(a)}{\leq} \sum_{i=1}^N \sum_{j=1}^N p_i p_j (c_{ii^*} + c_{i^*j}) \\
&= \sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{ii^*} + \sum_{i=1}^N \sum_{j=1}^N p_i p_j c_{i^*j} \\
&= \sum_{j=1}^N p_j \sum_{i=1}^N p_i c_{i^*i} + \sum_{i=1}^N p_i \sum_{j=1}^N p_j c_{i^*j} \\
&\overset{(b)}{=} 2 \min_i \sum_{j=1}^N p_j c_{ij},
\end{aligned}$$

where (a) follows from the triangle inequality satisfied by the elements c_{ij} and (b) follows from the definition of i^*. □

The second lemma states a useful probability metrics result. Let $P = \sum_{i=1}^N p_i \delta_{\omega_i}$ be a discrete support probability measure on Ω and let $Q = \sum_{i=1}^M q_i \delta_{\tilde{\omega}_i}$ be another such measure. Let $\zeta_c(P, Q)$ be a Fortet-Mourier metric for some cost function $c : \Omega \times \Omega \to \mathbb{R}_+$ that satisfies conditions described in Appendix B. Finally, let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a partition of Ω such that the measures P and Q are $\mathbf{C}$-equivalent.

Lemma 6.2 *Under assumptions described above,* *1.* $\zeta_c(P, Q) \leq \sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j})$*, where* $w_j = P(C_j) = Q(C_j)$*.*

2. If Q *is generated by some map* $\nu \in \mathcal{R}_r(\Omega_N)$ *that is* $\hat{c}$*-optimal, where* $\hat{c}$ *is the reduced cost function defined as in (B.8) then*

$\zeta_c(P, Q) = \sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j})$.

Proof: The first statement actually holds true for any measures $P, Q \in \mathcal{P}_c(\Omega)$ (see Appendix B for the definition of $\mathcal{P}_c(\Omega)$). Indeed, let $f^*(\omega) \in \mathcal{F}_c$ be the function that achieves the maximum of

$$\left| \int_\Omega f(\omega) P(d\omega) - \int_\Omega f(\omega) Q(d\omega) \right|.$$

Let $f_j^*(\omega)$ be the restriction of $f^*(\omega)$ to C_j. Clearly, $f_j^*(\omega) \in \mathcal{F}_c(C_j)$. We can write

$$\begin{aligned}
\zeta_c(P, Q) &= \left| \int_\Omega f^*(\omega) P(d\omega) - \int_\Omega f^*(\omega) Q(d\omega) \right| \\
&\stackrel{(a)}{=} \sum_{j=1}^{r} w_j \left| \int_{C_j} f^*(\omega) dP_{C_j}(\omega) - \int_{C_j} f^*(\omega) dQ_{C_j}(\omega) \right| \\
&\stackrel{(b)}{=} \sum_{j=1}^{r} w_j \left| \int_{C_j} f_j^*(\omega) dP_{C_j}(\omega) - \int_{C_j} f_j^*(\omega) dQ_{C_j}(\omega) \right| \\
&\stackrel{(c)}{\leq} \sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j}),
\end{aligned}$$

where (a) follows from the definition of conditional measures P_{C_j} and Q_{C_j}, (b) follows from the definition of functions $f_j^*(\omega)$, and (c) follows from that $f_j^*(\omega) \in \mathcal{F}_c(C_j)$ and definition of $\zeta_c(P_{C_j}, Q_{C_j})$.

To prove the second statement, we can use the duality result (B.5) described in Appendix B together with (B.10) that relates the values of Kantorovich-Rubinstein and Monge-Kantorovich functionals. Let $\nu \in \mathcal{R}_r(\Omega_N)$ be the map that generates partition $\mathbf{C}$, and let $\tilde{\omega}_j = \nu(\omega_i)$ for all

$\omega_i \in \mathbf{C}_j$. Note also that $q_j = \sum_{\{i:\omega_i \in C_j\}} p_i = w_j$, $j = 1, \ldots, r$. We can write

$$\sum_{j=1}^{r} w_j \zeta_c(P_{C_j}, Q_{C_j}) \overset{(a)}{=} \sum_{j=1}^{r} w_j \hat{\mu}_{\hat{c}}(P_{C_j}, Q_{C_j}) \overset{(b)}{=} \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} \frac{p_i}{w_j} \hat{c}(\omega_i, \tilde{\omega}_j)$$
$$= \sum_{j=1}^{r} \sum_{\{i:\omega_i \in C_j\}} p_i \hat{c}(\omega_i, \tilde{\omega}_j) \overset{(c)}{=} \hat{\mu}_{\hat{c}}(P, Q) \overset{(d)}{=} \zeta_c(P, Q),$$

where (a) and (d) follow from (B.5) and (B.10), (b) follows from that Q_{C_j} is supported at a single point $\tilde{\omega}_j$, (c) follows from the way measure Q was constructed as a reduction of the measure P with a $\hat{c}$-optimal map $\nu \in \mathcal{R}_r(\Omega_N)$. $\square$

Now, assume that the integrand $f(\omega, x)$ in (1.1) is in class $\mathcal{F}_c$ defined in Appendix B (expression (B.3)) for some symmetric cost function $c : \Omega \times \Omega \to \mathbb{R}_+$ that satisfies the conditions described in Appendix B. The following proposition describes a relation between relative expected losses for measures P and Q.

Proposition 6.1 *Let* $\mathbf{C}$ *be a partition of* Ω *generated by a reduction of a measure* P *with support at* $\Omega_N \subset \Omega$ *to* Q *by means of a* $\hat{c}$*-optimal map* $\nu \in \mathcal{R}_r(\Omega_N)$ *and let* $\hat{\mathbf{C}}$ *any coarsening of* $\mathbf{C}$ *(including* $\mathbf{C}$ *itself). Then*

$$L_P(\hat{\mathbf{C}}, Q) \leq L_P(\hat{\mathbf{C}}, P) + 2\zeta_c(P, Q),$$

Proof: Let $w_j = P(\hat{C}_j) = Q(\hat{C}_j)$ be the measure of subsets in $\hat{\mathbf{C}}$ and let $P_j \equiv P_{\hat{C}_j}$ and

$Q_j \equiv Q_{\hat{C}_j}$ be the corresponding subset measures

$$
\begin{aligned}
L_P(\hat{\mathbf{C}}, Q) &= \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{\omega}) P_j(d\omega) \right] \\
&= \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{\omega}) P_j(d\omega) \right. \\
&+ \left. \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) \right] \\
&\overset{(a)}{=} L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) \right] \\
&= L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left[\int_{\hat{C}_j} f(\omega, x^*_{Q_j}) P_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{P_j}) P_j(d\omega) \right. \\
&\qquad \left. + \int_{\hat{C}_j} f(\omega, x^*_{Q_j}) Q_j(d\omega) - \int_{\hat{C}_j} f(\omega, x^*_{Q_j}) Q_j(d\omega) \right] \\
&\overset{(b)}{=} L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left[v(Q_j) - v(P_j) + \int_{\hat{C}_j} f(\omega, x^*_{Q_j}) (P_j - Q_j)(d\omega) \right] \\
&\leq L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \left| v(Q_j) - v(P_j) \right| + \sum_{j=1}^{r} w_j \left| \int_{\hat{C}_j} f(\omega, x^*_{Q_j}) (P_j - Q_j)(d\omega) \right| \\
&\overset{(c)}{\leq} L_P(\hat{\mathbf{C}}, P) + \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) + \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) \\
&= L_P(\hat{\mathbf{C}}, P) + 2 \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) \\
&\overset{(d)}{=} L_P(\hat{\mathbf{C}}, P) + 2 \zeta_c(P, Q),
\end{aligned}
$$

where (a) follows from the definition of $L_P(\hat{\mathbf{C}}, P)$, (b) follows from the definition of the optimal objective values $v(P_j)$ and $v(Q_j)$, (c) follows from that the integrand $f(\omega, x)$ is in $\mathcal{F}_c$ and definition (B.4) of Fortet-Mourier metric ζ_c, and (d) follows from Lemma 6.2. □

If we use the trivial partition $\hat{\mathbf{C}} = \{\Omega\}$ (which is obviously a coarsening of any $\mathbf{C}$) in Proposition 6.1 we can obtain an upper bound on the relative loss of Q with respect to P which we formulate as a corollary.

Corollary 6.1 *The loss of reduced measure Q relative to P can be bounded from above as*

$$L_P(Q) \leq L(g_P, P) + 2\zeta_c(P, Q),$$

where $L(g_P, P) \equiv L_P(P)$ is the EVPI of the original problem (1.1).

The following proposition relates the expected loss of a subset-optimal map based on a partition generated by a reduction of the original measure P to measure Q to the Fortet-Mourier distance between P and Q.

Proposition 6.2 *Let $\mathbf{C}$ be a partition of Ω generated by a reduction of a measure P supported at the discrete set $\Omega_N \subset \Omega$ to measure Q by means of a $\hat{c}$-optimal map $\nu \in \mathfrak{R}_r(\Omega_N)$. Then*

$$L_P(\mathbf{C}, P) \equiv L(g_{\mathbf{C},P}, P) \leq 2\zeta_c(P, Q),$$

Proof: Let $w_j = P(C_j) = P(Q_j)$, $j = 1, \ldots, r$ be measures of subsets in $\mathbf{C}$ and let P_j and Q_j

be the corresponding subset measures.

$$
\begin{aligned}
L(g_{\mathbf{C},P}, P) &= \sum_{j=1}^{r} w_j L(g_{P_j}, P_j) = \sum_{j=1}^{r} w_j \int_{C_j} \Big(f(\omega, x^*_{P_j}) - f(\omega, x^*_{\omega}) \Big) P_j(d\omega) \\
&= \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} \frac{p_i}{w_j} \Big(f(\omega_i, x^*_{P_j}) - f(\omega_i, x^*_{\omega_i}) \Big) \\
&\overset{(a)}{=} \sum_{j=1}^{r} w_j \left(\sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, v(P_j) - \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, v(\delta_{\omega_i}) \right) \\
&= \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, (v(P_j) - v(\delta_{\omega_i})) \\
&\overset{(b)}{\leq} \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, \zeta_c(P_j, \delta_{\omega_i}) \\
&\overset{(c)}{=} \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, \hat{\mu}_{\hat{c}}(P_j, \delta_{\omega_i}) = \sum_{j=1}^{r} w_j \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \sum_{\{k:\omega_k \in C_j\}} (P_j)_k \, \hat{c}(\omega_i, \omega_k) \\
&\overset{(d)}{\leq} 2 \sum_{j=1}^{r} w_j \min_{\{k:\omega_k \in C_j\}} \sum_{\{i:\omega_i \in C_j\}} (P_j)_i \, \hat{c}(\omega_i, \omega_k) = 2 \sum_{j=1}^{r} w_j \min_{\{k:\omega_k \in C_j\}} \hat{\mu}_{\hat{c}}(P_j, \delta_{\omega_k}) \\
&\overset{(e)}{=} 2 \sum_{j=1}^{r} w_j \hat{\mu}_{\hat{c}}(P_j, Q_j) = 2 \sum_{j=1}^{r} w_j \zeta_c(P_j, Q_j) \overset{(f)}{=} 2\zeta_c(P, Q),
\end{aligned}
$$

where $(P_j)_i \equiv \frac{p_i}{w_j}$ for $\omega_i \in C_j$, (a) follows from the definition of optimal values $v(P_j)$ and $v(\delta_{\omega_i})$, (b) follows from the upper bound (B.11), (c) follows from the duality relation (B.5) and from the relation (B.10) between the Kantorovich-Rubinstein and Monge-Kantorovich functionals, (d) follows from Lemma 6.1 (since $\hat{c}$ is a metric and $\{(P_j)_i\}_{\{i:\omega_i \in C_j\}}$ is a probability distribution), (e) follows from that $Q = \nu(P)$, where ν is $\hat{c}$-optimal, and (f) follows from Lemma 6.2. □

Fig. 6.1 shows the locations of various points on $(G_P(\Omega, \cdot, \cdot), L_P(\cdot, \cdot))$ coordinate plane. Several useful observations can now be made.

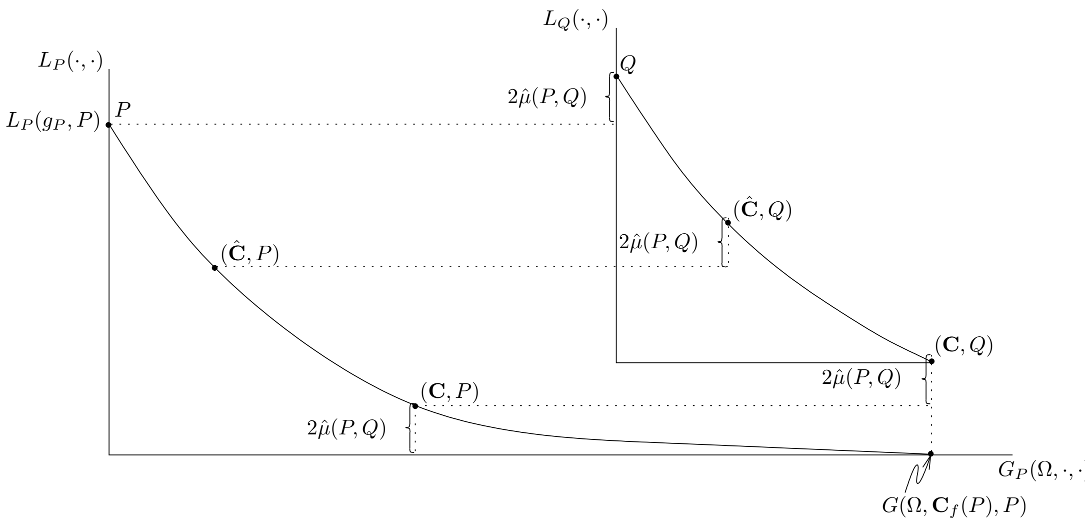

Figure 6.1: *Pseudo-energy (including virtual pseudo-energy) vs. relative loss.*

- The result of Proposition 6.2 suggests that good (near-optimal) partitions of Ω can be generated by a reduction of the original measure P to a measure Q that is (i) supported at a few points and (ii) has a low value of the Fortet-Mourier metric $\zeta_c(P,Q) = \hat{\mu}_{\hat{c}}(P,Q)$. The latter value of the Monge-Kantorovich functional $\hat{\mu}_{\hat{c}}(P,Q)$ with the reduced cost $\hat{c}$ can be readily computed as that of a minimum-cost transportation problem.

- For a wide class of linear multi-period two stage stochastic optimization problems, the relevant cost function c is given by c_p (see Appendix B, expression (B.12)) with $p = l + 1$ where l is the number of periods. The corresponding minimum cost transportation problem can be easily solved exactly for fixed support of measure Q and approximately if the support itself needs to be optimized (see Appendix C for details).

- The optimality "price" one pays for scenario reduction from the original measure P to a simpler measure Q – which can be thought of as adding information that's minimally relevant to the problem in question without actually finding it – can be estimated by the amount

$2\hat{\mu}_{\hat{c}}(P, Q)$. This implies, in particular that one could do a scenario reduction before starting the search for the efficient frontier. In fact, scenario reduction and additional information acquisition are complementary to each other in the sense of information: scenario reduction, as already mentioned can be thought of as an addition of information that's minimally relevant as opposed to information acquisition optimization, where one looks for small amounts of maximally relevant information.

6.3 Methods for Determining the Efficient Frontier

As was discussed earlier, an approximate solution to the optimal information acquisition problem (5.19) can be greatly facilitated by finding the efficient frontier $L^*(\gamma)$ of maps from Ω into X. The problem of finding the efficient frontier (5.12) appears to still be sufficiently complex to warrant a search for approximate solutions. Using Proposition 5.1, one can reduce the scope of search to subset-optimal maps $g \in \mathcal{C}$ only.

On the other hand, obviously, not all maps in the set $\mathcal{C}$ belong to the set $\mathcal{O}$ of Pareto-optimal map defining the efficient frontier. We call partition $\mathbf{C}$ *optimal* if the corresponding map $g = (\{C_1, \ldots, C_r\}, \{x^*_{P_1}, \ldots, x^*_{P_r}\}) \in \mathcal{C}$ belongs to the set $\mathcal{O}$ of Pareto-optimal maps. So the problem of finding maps in the set $\mathcal{O}$ is equivalent to that of searching for optimal partitions of the set Ω.

Proposition 6.2 provides a useful tool for approximating the efficient frontier. We can use the following algorithm (here and later we assume that the original measure P on Ω has a support at a discrete set $\Omega_N \subset \Omega$ consisting of N points).

(i) Choose an integer parameter $r \geq 2$.

(ii) Choose an appropriate cost function c: $\Omega \times \Omega \to \mathbb{R}_+$ such that $f(\omega, x) \in \mathcal{F}_c$ for all $x \in X$. Let $\hat{c}$ be the corresponding reduced cost function.

(iii) Reduce the original measure P to measure Q supported at r points in the set Ω_N, i.e. find a $\hat{c}$-optimal map $\nu \in \mathcal{R}_r(\Omega_N)$ such that $Q = \nu(P)$.

(iv) Let $\mathbf{C}$ be any partition of Ω generated by the map ν.

(v) Let the map $g_{\mathbf{C},P} \in \mathcal{C}$ be a subset-optimal map corresponding to partition $\mathbf{C}$.

Varying the value of parameter r from 2 upwards one can obtain a series of maps in the set $\mathcal{C}$ that are (approximately) Pareto-optimal. Step 2 of the above algorithm is essential for its feasibility. For example, if the problem (1.1) is a linear multi-period stochastic optimization problem, the cost function of the form (B.12) can be used. In step 3, finding the measure Q supported at r points that minimizes the value of Monge-Kantorovich functional $\hat{\mu}_{\hat{c}}(P, Q)$ is an NP-hard problem [29] but approximate algorithm such as *fast forward selection algorithm* are available (see Appendix B).

Using the algorithm described above, one can obtain one approximately Pareto-optimal map for each value of the chosen integer parameter. If more Pareto-optimal maps are needed (especially in the region with lower values of pseudo-energy) additional heuristics can be used. For instance, one could begin with the algorithm described above for some relatively high value of r and then merge some of the resulting subsets into one giving rise to a partition with a lower value of r. Clearly, this can be done in $B_r - 1$ ways, where B_n is the n-th Bell number which is just the number of all different partitions of a set consisting of n elements and that can be found from the recursive relation $B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$ and $B_0 = 1$. (For example, the Bell number for the lower values of n are $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, $B_5 = 52$, $B_6 = 203$, $B_7 = 877$, $B_8 = 4140$.)

We see that if the original chosen value of r is not very high this would lead to a manageable number of partitions. Additionally, scenario reduction can be used to reduce computational complexity of finding the values $x^*_{P_C}$ for subsets C of resulting partitions. On the other hand, if the original value of r makes evaluation of all maps that can be obtained this way computationally prohibitive, a heuristic algorithm described by the following pseudo-code can be used. It finds another partition, with a lower value of r, so that the subset merging procedure can be applied.

Algorithm 1: Approximation to Pareto-optimal boundary.

Input:
  $\mathbf{C} = \{C_1, \ldots, C_r\}$,
  $\{\omega_1, \ldots, \omega_r | \omega_i \in C_i\} \subset \Omega$,
  choose an integer n such that $1 \leq n \leq r - 2$.
Step 0**:**
  $J^{[0]} := \{1, \ldots r\}$,
  $\mathbf{C}' := \{C'_1, \ldots, C'_{n+1}\}$ such that $C'_i := \emptyset$, $\forall\, i$,
  calculate $\hat{c}_p(\omega_i, \omega_j),\ \forall\, i, j \in J^{[0]}$.
Step $k = 1, \ldots, n$**:**
  foreach $i \in J^{[k-1]}$ **do**
    $\bar{c}_p(i) := \frac{1}{|J^{[k-1]}|} \sum_{j \in J^{[k-1]}} \hat{c}_p(\omega_i, \omega_j)$,
  end
  $u_k := \arg\max_{i \in J^{[k-1]}} \bar{c}_p(i)$,
  $J^{[k]} := J^{[k-1]} \backslash \{u_k\}$,
  $C'_k := C_{u_k}$.

The goal of the algorithm represented by the pseudo-code is to identify subsets which are locally compact but as far away from one another as possible. In each step k, we find the average distance $\bar{c}_p$ of each subset center remaining in the index set $J^{[k-1]}$ to only the other remaining centers. The center, and therefore the associated subset, with the largest average distance is chosen and removed from the set $J^{[k-1]}$. The remaining subsets are then merged into a single set.

So far the pseudo-temperature function u has not been taken into account. It is clear, on the other hand, that it will in general affect the composition of the set $\mathcal{O}$ of Pareto-optimal maps. In order to properly incorporate the pseudo-temperature function into the heuristics described above, one could note that the questions difficultly is generally smaller when subsets with high pseudo-temperature values have large measures as well. In other words, if one wishes to keep the question difficulty low, one should avoid creating subsets of small measure in regions of the parameter space characterized with high pseudo-temperature values. To facilitate creation of such subsets, one could, for example modify the (reduced) cost function $\hat{c}$ in the following way

$$\hat{c}(\omega_i, \omega_j) \to \frac{\hat{c}(\omega_i, \omega_j)}{f_c(u(\omega_i), u(\omega_j))}, \tag{6.7}$$

where f_c: $\mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is some increasing function of its arguments. The specific shape of f_c can be determined experimentally, and several shapes can be tried for every given instance assuming computational resources are not a limiting factor.

6.4 Example

Let us consider an example. The original problem is a that of two-stage linear stochastic optimization with simple recourse taken from a well-known textbook [6]. The problem is for a farmer to allocate the appropriate amount of land between wheat, corn and sugar beets in order to maximize profits. The farmer knows that at least 200 tons of wheat and 240 tons of corn must be grown for cattle feed. If not enough is grown to satisfy this demand, both wheat and corn can be bought for \$238 and \$210 per ton, respectively. Any excess above the demand can be sold for \$170 and \$150 per ton of wheat and corn, respectively. It costs \$150 per acre to plant the wheat and \$230 per acre to plant the corn. The farmer can also grow sugar beets that sell for \$36 per ton. However, there is a quota of 6000 tons and any amount grown above this may only be sold at \$10 per ton. It costs \$260 per acre to plant sugar beets. The farmer has 500 acres available.

The problem can be stated as:

$$\begin{aligned} \text{minimize} \quad & 150x_1 + 230x_2 + 260x_3 + \mathbb{E}_P Q(x, \Omega) \\ \text{subject to} \quad & x_1 + x_2 + x_3 \quad \leq 500 \\ & x_1, x_2, x_3 \geq 0, \end{aligned} \tag{FP}$$

where the second stage problem for a specific scenario can be written

$$\begin{aligned} Q(x, s) = \text{minimize} & \{238y_1 - 170w_1 + 210y_2 - 150w_2 - 36w_3 - 10w_4\} \\ \text{subject to} \quad & \omega_1(s)x_1 + y_1 + w_1 \quad \geq 200 \\ & \omega_2(s)x_2 + y_2 + w_2 \quad \geq 240 \\ & w_3 + w_4 \quad \leq \omega_3(s)x_3 \\ & w_3 \quad \leq 6000 \\ & y_1, y_2, w_1, w_2, w_3, w_4 \geq 0, \end{aligned}$$

where $\omega_i(s)$ represents the yield of crop $i := 1, 2, 3$ for wheat, corn, and sugar beets, respectively, under scenario s; x_i are the acres of land to devote to each crop i; y_1, y_2, are tons of wheat and corn, respectively, purchased to meet cattle feed requirements; w_1, w_2, w_3, w_4 are tons of wheat, corn, sugar beets below quota, and sugar beets above quota, respectively, sold for profit.

The problem has been modified in order to create the illustrative example used below. In this example, only wheat and sugar beet yields are uncertain. Each is allowed to take five different values of yields resulting in 25 scenarios. For the sake of convenience, we assume that the corn yield is non-random and is equal to 3 tons per acre, while for both wheat and beets the average yield equal to 2.5 and 20, respectively, has a probability of 0.30. The yield for both of these cultures can be either higher or lower than average by 20% with probability 0.20 and also can be higher or lower than average by 30% with probability 0.15. The yields for wheat and beets are assumed to be independent.

The resulting uncertain yields are summarized below:

wheat (ω_1)	[1.75, 2.00, 2.50, 3.00, 3.25]	w.p. (0.15,0.20,0.30,0.20,0.15),
corn	[3]	w.p. (1),
sugar beets (ω_2)	[14, 16, 20, 24, 26]	w.p. (0.15,0.20,0.30,0.20,0.15).

Also, let us assume that the pseudo-temperature function $u(\omega_1, \omega_2)$ is given as

$$u(i,j) = i \cdot j^{0.5}, \forall \quad i,j \in 1,\ldots,5, \tag{6.8}$$

where $i,\ j$ are the indices referencing the uncertain yields of wheat and sugar beets, respectively (where the smallest value of the uncertain yield corresponds to $i = 1$ ($j = 1$) and the largest yield corresponds to $i = 5$ ($j = 5$)). The pseudo-temperature function is then normalized so that $\mathbb{E}_P u(ij) = 1$. Fig. 6.2 shows a plot of the pseudo-temperature function.

The efficient frontier can be approximated by using the scenario reduction based algorithm described in the previous section together with subset merging heuristics. The resulting maps are shown in Fig. 6.3 for the case of constant pseudo-temperature. The resulting approximate efficient

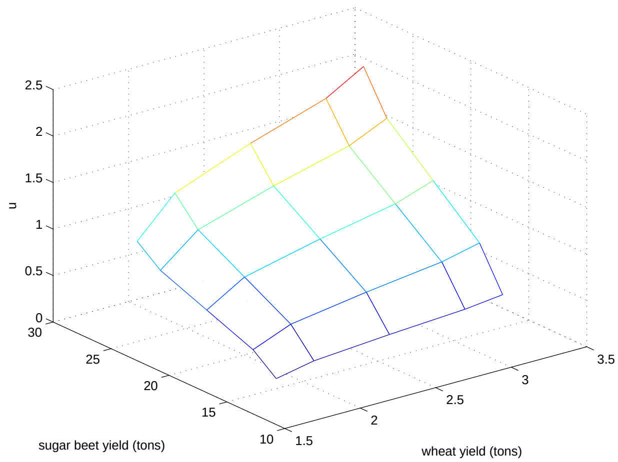

Figure 6.2: *Pseudo-temperature function given for the farmer land allocation problem with uncertainty residing in the yields of wheat and sugar beets.*

frontier both for constant pseudo-temperature function and for the pseudo-temperature given shown in Fig. 6.2 are shown in Fig. 6.4.

Now consider an information source described by the modified linear model with parameters $b = 0.8$ and $Y_s = 0.2$ (which is a rather modest capacity value). We would like to find out how much the original loss can be reduced by optimally using such an information source. In other words, we want to solve problem (5.19). For this purpose one can take questions on the (approximate) efficient frontier and plot parametric curves $(Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})), \mathcal{L}(V_\alpha(\mathbf{C})))$ where $\mathcal{L}(V_\alpha(\mathbf{C}))$ is given by Proposition 5.3. The question yielding the lowest point of intersection of such a curve with the vertical line $G = Y_s$ will give an approximate solution of problem (5.19).

Results for the case of constant pseudo-temperature are shown in Fig. 6.5. The parametric curves for three questions (all three with $r = 2$) are produced. We can see that the lowest value of the expected loss that can be obtained this way is equal to 7250 which constitutes a reduction of about 14%.

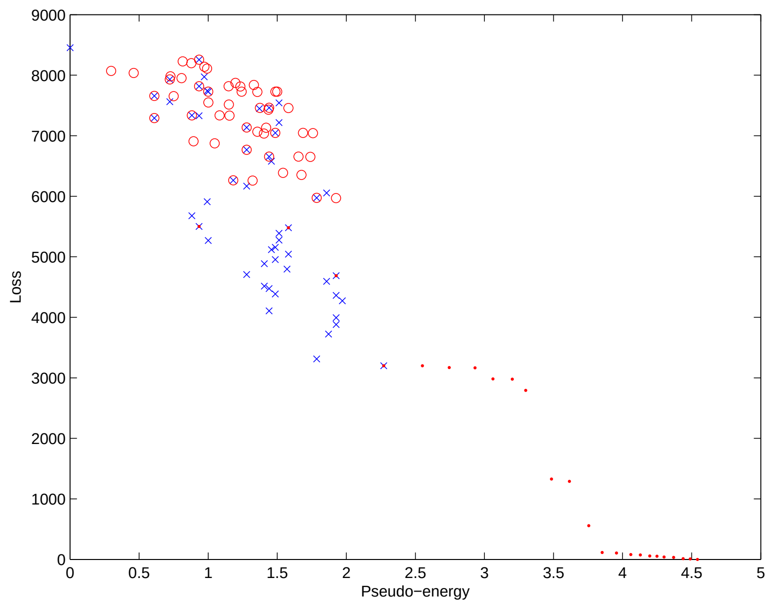

Figure 6.3: *Maps that are generated by scenario reduction for various values of r (solid dots), scenario reduction for $r = 5$ with subsequent subset merging (crosses), scenario reduction to $r = 10$, reducing to $r = 5$ using the pseudo-code and subsequent subset merging (circles). Pseudo-temperature function is set to a constant.*

For the case of non-constant pseudo-temperature are shown in Fig. 6.6. Analogously, three $r = 2$ questions were chosen on the approximate efficient frontier and the corresponding parametric curves plotted. The best curve is observed to intersect the vertical line $G = 0.2$ at the value of vertical coordinate equal to about 6900 which represents a reduction of about 18% compared to the EVPI of 8450 of the original problem.

6.5 Conclusion

This chapter develops (approximate) methods for solving the problem of optimizing additional information acquisition in decision making problems with uncertainty that are typically solved using

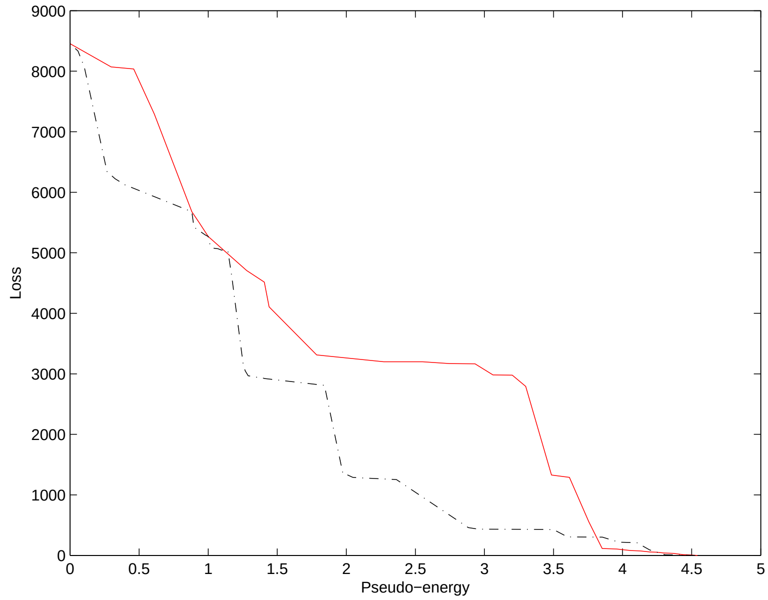

Figure 6.4: *Approximate efficient frontiers for the constant pseudo-temperature function (solid line) and pseudo-temperature shown in Fig. 6.2.*

stochastic optimization techniques. It represents a logical continuation of the developments presented in Chapter 5. The main problem that was formulated there is that of finding an efficient frontier in pseudo-energy – loss coordinate plane and to determining the question(s) that would allow to minimize the expected loss for the given (stochastic optimization) problem and a given information source.

The solution methods proposed in this paper are based on the method of probability metrics and their application for scenario reduction in stochastic optimization. The main idea is that, informally speaking, optimal scenario reduction on one hand and optimal information acquisition on the other hand are complementary. More specifically, in scenario reduction the goal is reproduce the overall shape of the original probability distribution as faithfully as possible with a small fraction of the original scenarios. In information acquisition, the goal is to identify the types of uncertainty encoded by the original probability distribution the reduction of which would have the largest effect on the

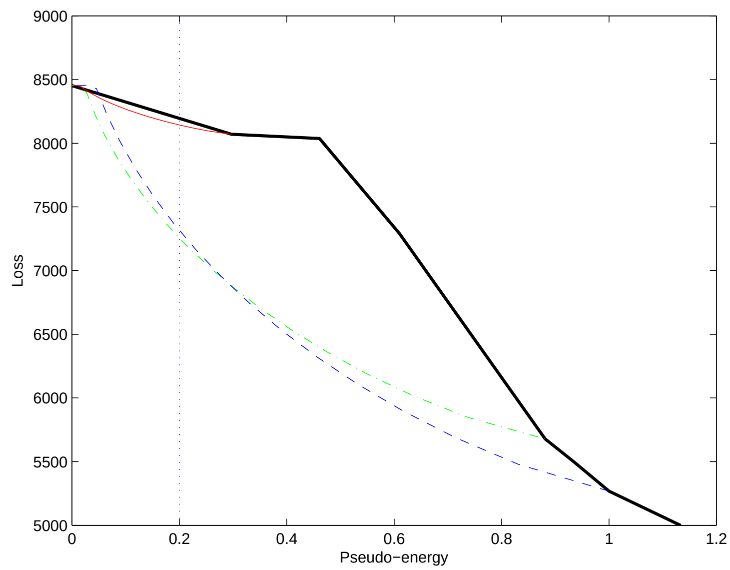

Figure 6.5: *Part of approximate efficient frontier and parametric loss curves for quasi-perfect answers to three different questions for the case of constant pseudo-temperature.*

solution quality. It turns out that these types of uncertainty are associated with the "overall shape" of the distribution (as opposed to "local details") which scenario reduction strives to preserve.

This allows us to develop simple approximate algorithms for determining the efficient frontier (and for finding optimal questions for the given information source) with the help of existing scenario reduction algorithms. The methods described in this chapter are shown to work for the class of linear multi-period two stage stochastic optimization problems and should generalize relatively easily to other problem classes for which scenario reduction based on probability metrics was shown to be possible such as chance constrained and two-stage integer stochastic optimization problems.

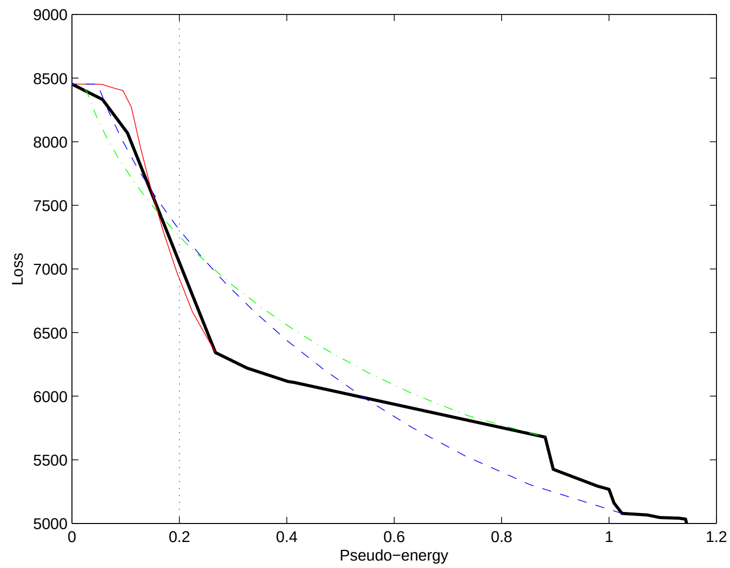

Figure 6.6: *Part of approximate efficient frontier and parametric loss curves for quasi-perfect answers to three different questions for the case of non-constant pseudo-temperature shown in Fig. 6.2.*

Bibliography

[1] A. Bassamboo and A. Zeevi. On a data-driven method for staffing large call centers. *Oper. Res.*, 57(3):514–526, May-June 2009.

[2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[3] A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: the optimized certainty equivalent. *Mathematical Finance*, 17:449–476, 2007.

[4] A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Math. Programming, Ser.B*, 107:63–89, 2006.

[5] E. Bickel. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Anal.*, 4(2):49–65, 2007.

[6] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, New York, NY, 1997.

[7] R. F. Bordley. Combining the opinions of experts who partition events differently. *Decision Anal.*, 6(1):38–46, March 2009.

[8] R. F. Bordley. Using Bayes' rule to update an event's probabilities based on the outcomes of partially similar events. *Decision Anal.*, 8(2):117–127, June 2011.

[9] G. W. Brier. Verification of forecasts expressed in terms of probabilities. *Monthly Weather Rev.*, 78(1):1–3, 1950.

[10] M. Chávez, J. Martinerie, and M. Le Van Quyen. Statistical assessment of nonlinear causality: Application to epileptic EEG signals. *J. of Neurosci. Methods*, 124(2):113–128, April 2003.

[11] R. Clemen. Combining overlapping information. *Management Sci.*, 33(3):373–380, 1987.

[12] R. Clemen and R. Winkler. Combining probability distributions from experts in risk analysis. *Risk Anal.*, 19(2):187–203, 1999.

[13] R. T. Cox. Probability, frequency, and reasonable expectation. *Am. J. Phys.*, 14:1–13, 1946.

[14] R. T. Cox. *The Algebra of Probable Inference*. Johns Hopkins Press, Baltimore, 1961.

[15] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Math. Programming, Ser.A*, 95(3):493–511, 2003.

[16] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16:69–95, 2005.

[17] G. Ellison and D. Fudenberg. Rules of thumb for social learning. *J. Political Econom.*, 101 (4):612–643, 1993.

[18] D. K. Faddeev. On the concept of entropy of a finite probabilistic scheme. *Uspekhi Mat. Nauk*, 11(1(67)):227–231, 1956.

[19] A. J. Fischer, A. J. Arnold, and M. Gibbs. Information and the speed of innovation adoption. *Amer. J. Agr. Econ.*, 78(4):1073–1081, 1996.

[20] M. L. Fisher and A. Raman. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.*, 44(1):87–99, 1996.

[21] C. Fox and R. Clemen. Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Sci.*, 51(9):1417–1432, 2005.

[22] S. French. Group consensus probability distributions: A critical survey. *Bayesian Statist.*, 2: 183–202, 1985.

[23] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statist. Sci.*, 1:114–148, 1986.

[24] T. Gneiting and A. Raftery. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.*, 102:359–378, 2007.

[25] I. J. Good. Rational decisions. *J. Roy. Statist. Soc. Ser. B*, 14(1):107–114, 1952.

[26] Silviu Guiaşu. Weighted entropy. *Reports on Mathematical Physics*, 2(3):165–179, 1971.

[27] D. Harmanec and G. J. Klir. Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *Int. J. Gen. Syst.*, 22(4):405–419, 1994.

[28] J. H. Havrda and F. Charvat. Quantification methods of classification processes: Concepts of structural α entropy. *Kybernetika*, 3:30–35, 1967.

[29] Holger Heitsch and Werner Römisch. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24(2):187–206, 2003.

[30] Ronald A. Howard. Information value theory. *IEEE Transactions on systems science and cybernetics*, 2(1):22–26, 1966.

[31] W. T. Huh and P. Rusmevichientong. A nonparametric asymptotic analysis of inventory planning with censored demand. *Math. Oper. Res.*, 34(1):103–123, February 2009.

[32] S. Jaroszewicz and D. A. Simovici. On axiomatization of conditional entropy of functions between finite sets. In *Proc. 29th ISMVL, Freiburg, Germany*, pages 24–31, 1999.

[33] E. T. Jaynes. Information theory and statistical mechanics I. *Phys. Rev.*, 106:620–630, 1957.

[34] E. T. Jaynes. Information theory and statistical mechanics II. *Phys. Rev.*, 108:171–190, 1957.

[35] R. Jensen. Information cost and innovation adoption policies. *Management Sci.*, 34(2):230–239, 1988.

[36] D. L. Johnstone, V. R. R. Jose, and R. L. Winkler. Tailored scoring rules for probabilities. *Decision Anal.*, 8(4):256–268, 2011.

[37] T. Katura, N. Tanaka, A. Obata, H. Sato, and A. Maki. Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics. *NeuroImage*, 31(4):1592–1600, July 2006.

[38] G. J. Klir and D. Harmanec. Generalized information theory: Recent developments. *Kybernetes*, 25(7/8):50–66, 1996.

[39] K. H. Knuth. Lattice duality: The origin of probability and entropy. *Neurocomputing*, 67: 245–274, 2005.

[40] L. J. Kornish and R. L. Keeney. Repeated commit-or-defer decisions with a deadline: The influenza vaccine composition. *Oper. Res.*, 56(3):527–541, 2008.

[41] R. Levi, R. O. Roundy, and D. B. Shmoys. Provably near-optimal sampling-based policies for stochastic inventory control models. *Math. Oper. Res.*, 32(4):821–839, November 2007.

[42] V. L. Levin. On the problem of mass transfer. *Sov. Math. Dokl.*, 16:1349–1353, 1975.

[43] Y. Maeda and H. Ichihashi. An uncertainty measure with monotonicity under the random set inclusion. *Int. J. Gen. Syst.*, 21(4):379–392, 1993.

[44] K. F. McCardle. Information acquisition and the adoption of new technology. *Management Sci.*, 31(11):1372–1389, 1985.

[45] I. I. Mokhov and D. A. Smirnov. El Niño-Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices. *Geophys. Res. Lett.*, 33, 2006. L03708.

[46] J. Neveu and R. M. Dudley. On Kantorovich-Rubinstein theorems. transcript, 1980.

[47] J. B. Predd, D. N. Osherson, S. R. Kulkarni, and H. V. Poor. Aggregating probabilistic forecasts from incoherent and abstaining experts. *Decision Anal.*, 5(4):177–189, December 2008.

[48] S. T. Rachev and R. M. Shortt. Duality theorems for Kantorovich-Rubinstein and Wasserstein functionals. *Dissertationes Mathematicae*, 299:647–676, 1990.

[49] Svetlozar T. Rachev and Werner Römisch. Quantitative stability in stochastic programming: The method of probability metrics. *Math. Oper. Res.*, 27(4):792–818, 2002.

[50] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass Transportation Problems, Volume I: Theory*. Springer-Verlag, New York, 1998.

[51] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass Transportation Problems, Volume II: Applications*. Springer-Verlag, New York, 1998.

[52] A. Rényi. On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 547–561. Univ. California Press, Berkeley, Calif., 1961.

[53] Werner Römisch and Roger J.-B. Wets. Stability of ε-approximate solutions to convex stochastic programs. *SIOPT*, 18(3):961–979, 2007.

[54] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Math. Oper. Res.*, 31: 433–452, 2006.

[55] L. J. Savage. Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.*, 66:783–801, 1971.

[56] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[57] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 623–659, 1948.

[58] D. A. Simovici and S. Jaroszewicz. An axiomatization of partition entropy. *IEEE Trans. Inf. Theory*, 48(7):2138–2142, July 2002.

[59] P. F. Verdes. Assessing causality from multivariate time series. *Phys. Rev. E*, 72, 2005. 026222.

[60] P. A. Viola. Alignment by maximization of mutual information. A.I. Technical Report 1548, Massachusetts Institute of Technology, June 1995.

[61] R. L. Winkler. Scoring rules and the evaluation of probabilities. *Test*, 5:1–60, 1996.

[62] V. M. Zolotarev. Probability metrics. *Theory of Probability and its Applications*, 28(2):264–287, 1983.

Appendix A

Parallels with Thermodynamics

Imagine an ideal gas contained in container A with unit volume $V = 1$ and held at temperature T (see Fig. A.1 for an illustration). There is a "marked" molecule. Let $C \subset A$ be a part of the original container that has volume $V < 1$. We are interested in whether the molecule of interest is located in C or otherwise. A "constructive" way of reducing this uncertainty is compressing the original container so that all gas – including the special molecule – is in C with certainty.

The energy conservation law reads $dQ = pdV + dU$, where dQ is the (infinitesimal) heat transferred to the gas, pdV is the work done by the gas and dU is the increment of the gas internal energy. If we insist that the gas be kept at constant temperature T then (since the gas is ideal)

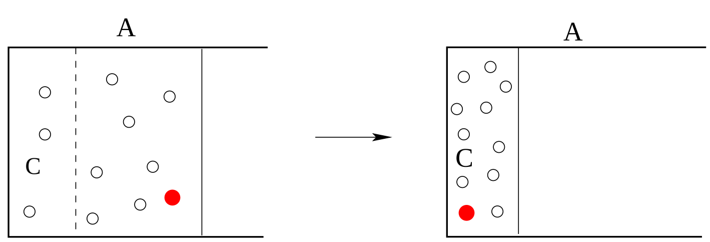

Figure A.1: *Gas in container A is compressed from original unit volume to volume $V < 1$. The marked molecule is shown as a shaded circle.*

$dU = 0$ and the energy conservation law reduces to

$$dQ = pdV. \tag{A.1}$$

The ideal gas equation of state reads $pV = \nu RT$ where p is the pressure, ν is the amount of substance (in moles) and R is the ideal gas constant. We can use the equation of state to express pressure as a function of the gas volume:

$$p(V) = \frac{\nu RT}{V}. \tag{A.2}$$

Substituting (A.2) into (A.1) we can obtain for the amount of heat transferred to the gas while its volume is educed from 1 to V at constant temperature T:

$$\Delta Q = \int_1^V p(V)dV = \nu RT \int_1^V \frac{dV}{V} = \nu RT \ln V = \nu RT (\ln 2) \log V < 0,$$

implying that the amount of heat equal to

$$-\Delta Q = \ln 2\, \nu RT \log \frac{1}{V} > 0 \tag{A.3}$$

is taken away from the gas. We can note now that $V = P(C)$ where $P(\cdot)$ is the uniform measure on A describing the initial information on the location of the marked molecule in A. Comparing expression (A.3) with that for the difficulty of a free-response question C

$$G(\Omega, C, P) = u(C) \log \frac{1}{P(C)}$$

we see that (i) the value $u(C)$ plays the role of temperature and (ii) the question difficulty can be thought of as the energy-like quantity that is similar to the thermal energy (heat) that has to be taken away from the system in order to reduce uncertainty about the microstate that can be characterized by entropy. The latter is related to heat by the relationship $dQ = TdS$, where S stands for the thermodynamic entropy. Thus, the higher temperature is the larger the amount of heat that has to be dissipated in order to reduce entropy. Therefore temperature can be interpreted as (thermal) energy per unit of entropy. In application to inquiry, respectively, the pseudo-temperature $u(\cdot)$ can be thought of as the amount of pseudo-energy (difficulty) per unit of Shannon entropy that represents the purely informational quantity measuring the minimum expected number of bits that is necessary to *communicate* a perfect answer to the given question.

Appendix B

Probability metrics and stability in stochastic optimization

Consider the problem (1.1). Let $\mathcal{P}(\Omega)$ be the set of all Borel probability measures on Ω and define

$$v(P) = \inf\left\{\int_\Omega f(\omega, x)\, dP(\omega) \,:\, x \in X\right\}$$

and

$$S(P) = \left\{x \in X \,:\, \int_\Omega f(\omega, x)\, dP(\omega) = v(P)\right\}$$

to be the optimal value and optimal solution set of (1.1), respectively.

Let's also define (as in, for example, [53])

$$\mathcal{F} = \{f(\cdot, x) : \; x \in X\}$$

and

$$\mathcal{P}_{\mathcal{F}}(\Omega) = \left\{ Q \in \mathcal{P} : \ -\infty < \int_{\Omega} \inf_{x \in X \cap \rho\mathbb{B}} f(\omega, x) Q(d\omega) \ \text{ and } \right.$$
$$\left. \sup_{x \in X \cap \rho\mathbb{B}} \int_{\Omega} f(\omega, x) Q(d\omega) < \infty, \ \text{for all}\, \rho > 0 \right\},$$

where $\mathbb{B}$ is the closed unit ball in $\mathbb{R}^n$.

Then the probability distance of the form

$$d_{\mathcal{F},\rho}(P, Q) = \sup_{x \in X \cap \rho\mathbb{B}} \left| \int_{\Omega} f(\omega, x) P(d\omega) - \int_{\Omega} f(\omega, x) Q(d\omega) \right| \tag{B.1}$$

can be defined on $\mathcal{P}_{\mathcal{F}}(\Omega)$. This distance is called *Zolotarev's pseudometric with ζ-structure* [62, 49, 50, 51]. The pseudometric (B.1) would become a metric if the class $\mathcal{F}$ were rich enough so that $d_{\mathcal{F},\rho}(P, Q) = 0$ implies $P = Q$.

Theorem 2 in [15] states that if $P, Q \in \mathcal{P}_{\mathcal{F}}$, $S(P)$ is nonempty and bounded then there exist $\rho > 0$ and $\delta > 0$ such that

$$|v(P) - v(Q)| \leq d_{\mathcal{F},\rho}(P, Q) \tag{B.2}$$

is valid for all $Q \in \mathcal{P}_{\mathcal{F}}$ such that $d_{\mathcal{F},\rho}(P, Q) < \delta$.

The distance $d_{\mathcal{F},\rho}$ in (B.2) is typically difficult to handle since the class of functions $\mathcal{F}$ is determined by the specific integrand $f(\omega, x)$ for the given instance of problem (1.1). The main idea underlying the use of the probability metrics method for the study of stability and for scenario reduction in stochastic programming is to suitably enlarge the class $\mathcal{F}$ so that it still shares its main analytical properties with functions $f(\cdot, x)$. Such properly enlarged classes are sometimes referred to as *canonical classes* and the corresponding metrics are sometimes called *canonical metrics*.

Consider, for instance the class $\mathcal{F}_c$ of continuous functions defined as

$$\mathcal{F}_c = \{f:\ \Omega \to \mathbb{R}:\ |f(\omega) - f(\tilde{\omega})| \le c(\omega, \tilde{\omega}),\ \text{for all } \omega, \tilde{\omega} \in \Omega\}, \tag{B.3}$$

where c: $\Omega \times \Omega \to \mathbb{R}_+$ is a continuous symmetric function such that $c(\omega, \tilde{\omega}) = 0$ if and only if $\omega = \tilde{\omega}$. Then the corresponding (pseudo-) metric has the form

$$\zeta_c(P, Q) \equiv d_{\mathcal{F}_c}(P, Q) = \sup_{f \in \mathcal{F}_c} \left| \int_\Omega f(\omega) P(d\omega) - \int_\Omega f(\omega) Q(d\omega) \right| \tag{B.4}$$

and is known as *Fortet-Mourier metric*. If the cost function $c(\omega, \tilde{\omega})$ satisfies additional boundedness and continuity conditions:

- $c(\omega, \tilde{\omega}) \le \lambda(\omega) + \lambda(\tilde{\omega})$ for some λ: $\Omega \to \mathbb{R}_+$ mapping bounded sets into bounded sets,
- $\sup\{c(\omega, \tilde{\omega})\ :\ \omega, \tilde{\omega} \in \mathbb{B}_\epsilon(\omega_0), ||\omega, -\tilde{\omega}|| \le \delta\} \to 0$ as $\delta \to 0$ for each $\omega_0 \in \Omega$, where $\mathbb{B}_\epsilon(\omega_0)$ is the ϵ-ball centered at ω_0,

the Fortet-Mourier metric (B.4) admits a dual representation as the *Kantorovich-Rubinstein functional* [48]:

$$\begin{aligned} \zeta_c(P, Q) = \mathring{\mu}_c(P, Q) = \inf \bigg\{ \int_{\Omega \times \Omega} & c(\omega, \tilde{\omega}) \eta(d\omega, d\tilde{\omega}), \\ & \eta \in \mathcal{P}(\Omega \times \Omega),\ \pi_1 \eta - \pi_2 \eta = P - Q \bigg\}, \end{aligned} \tag{B.5}$$

where π_1 and π_2 denote projections on first and second components, respectively. It is straightforward to show that the Kantorovich-Rubinstein functional (B.5) can be upper-bounded by the

Monge-Kantorovich functional:

$$\overset{\circ}{\mu}_c(P,Q) \le \hat{\mu}_c(P,Q) = \inf\bigg\{ \int_{\Omega\times\Omega} c(\omega,\tilde{\omega})\eta(d\omega,d\tilde{\omega}), \\ \eta \in \mathcal{P}(\Omega\times\Omega),\ \pi_1\eta = P, \pi_2\eta = Q\bigg\}, \tag{B.6}$$

and that the bounds becomes tight, (i.e. $\overset{\circ}{\mu}_c(P,Q) = \hat{\mu}_c(P,Q)$) if the cost function $c(\omega,\tilde{\omega})$ is a metric on Ω [42, 46]. The problem of finding the minimum in (B.6) is known the *Monge-Kantorovich mass transportation problem*.

Note that if measures P and Q are discrete ($P = \sum_{i=1}^{N} p_i\delta_{\omega_i}$ and $Q = \sum_{j=1}^{M} q_i\delta_{\tilde{\omega}_j}$), the Monge-Kantorovich functional (B.6) takes the following form:

$$\begin{aligned}\hat{\mu}_c(P,Q) &= \min\left\{\sum_{i=1}^{N}\sum_{j=1}^{M} c(\omega_i,\tilde{\omega}_j)\eta_{ij} : \eta_{ij}\ge 0, \sum_{i=1}^{N}\eta_{ij} = q_j, \sum_{j=1}^{M}\eta_{ij} = p_i\ \forall i,j\right\} \\ &= \max\left\{\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{M} q_j v_j : u_i + v_j \le c(\omega_i,\tilde{\omega}_j)\ \forall i,j\right\}\end{aligned} \tag{B.7}$$

Given the cost function $c(\omega,\tilde{\omega})$ one can define the *reduced cost* $\hat{c}(\omega,\tilde{\omega})$ on $\Omega\times\Omega$ by

$$\hat{c}(\omega,\tilde{\omega}) = \inf\left\{\sum_{i=1}^{m-1} c(\omega_i,\omega_{i+1}) : m\in\mathbb{N},\ \omega_i\in\Omega,\ \omega_1=\omega,\ \omega_m=\tilde{\omega}\right\}. \tag{B.8}$$

It can easily be shown that the reduced cost function $\hat{c}(\omega,\tilde{\omega})$ is a metric (since it satisfies the triangle inequality) on Ω and that $\hat{c}(\omega,\tilde{\omega}) \le c(\omega,\tilde{\omega})$ with the inequality being tight when $c(\omega,\tilde{\omega})$ is also a metric.

It can also be shown (see [50], chapter 4) that if Ω is compact with analytic sublevel sets then the Kantorovich-Rubinstein functional (B.5) with the reduced cost function $\hat{c}$ coincides with the

Kantorovich-Rubinstein functional with the original cost function c (the result referred to as the *reduction theorem*):

$$\mathring{\mu}_{\hat{c}}(P,Q) = \mathring{\mu}_c(P,Q). \tag{B.9}$$

Since the reduced cost is a metric on Ω we have $\mathring{\mu}_{\hat{c}}(P,Q) = \hat{\mu}_{\hat{c}}(P,Q)$ and, comparing with (B.9) we conclude that, for compact parameter spaces with analytic sublevel sets, the equality

$$\mathring{\mu}_c(P,Q) = \hat{\mu}_{\hat{c}}(P,Q) \leq \hat{\mu}_c(P,Q) \tag{B.10}$$

holds true.

We thus arrive at the following useful stability result. If the integrand in problem (1.1) belongs to class $\mathcal{F}_c$ for all $x \in X$ for some cost function c satisfying additional boundedness and continuity conditions described earlier in the appendix, then the estimate

$$|v(P) - v(Q)| \leq \zeta_c(P,Q) = \mathring{\mu}_c(P,Q) = \hat{\mu}_{\hat{c}}(P,Q) \tag{B.11}$$

is valid for Borel measures P and Q in $\mathcal{P}_c(\Omega)$ on compact Ω characterized with analytic sublevel sets. (Here $\mathcal{P}_c(\Omega) = \{Q \in \mathcal{P}(\Omega) \,:\, \int_\Omega c(\omega,\omega_0)dQ(\omega) < \infty\}$ for some $\omega_0 \in \Omega$.)

The particular function $c(\omega,\tilde{\omega})$ that plays an important role in the context of convex stochastic optimization has the form

$$c_p(\omega,\tilde{\omega}) = \max\{1, ||\omega - \omega_0||^{p-1}, ||\tilde{\omega} - \omega_0||^{p-1}\}||\omega - \tilde{\omega}||, \tag{B.12}$$

for some $\omega_0 \in \Omega$. The corresponding metric $\zeta_p \equiv \zeta_{c_p}$ is referred to as the *p-th order Fortet-Mourier* *metric*.

To give an example of a class of problems for which the p-th order Fortet-Mourier metric is relevant, consider linear multi-period stochastic optimization problems of the form

$$\begin{aligned} \min\Big\{ & cy_0 + \mathbb{E}_P\Big(\min \sum_{j=1}^{l} c_j(\omega) y_j\Big), \\ & y_0 \in X,\, y_j \in Y_j,\, W_{jj} y_j = b_j(\omega) - W_{jj-1}(\omega) y_{j-1},\, j = 1, \ldots, l\Big\}, \end{aligned} \tag{B.13}$$

where $Y_j \subseteq \mathbb{R}^{n_j}$ are polyhedral sets. Problem (B.13) can be written in the form (1.1) with the integrand $f(\omega, x)$ given by

$$\begin{aligned} f(\omega, x) &= cx + \inf\left\{ \sum_{j=1}^{l} c_j(\omega) y_j : y_j \in Y_j,\, W_{jj} y_j = b_j(\omega) - W_{jj-1}(\omega) y_{j-1},\, j = 1, \ldots, l \right\} \\ &= cx + \Psi_1(\omega, x), \end{aligned}$$

where the function $\Psi_1(\omega, x)$ is defined recursively:

$$\Phi_j(\omega, u_{j-1}) = \inf\{c_j(\omega) y_j + \Psi_{j+1}(\omega, y_j) : y_j \in Y_j,\, W_{jj} y_j = u_{j-1}\}$$

$$\Psi_j(\omega, y_{j-1}) = \Phi_j(\omega, b_j(\omega) - W_{jj-1}(\omega) y_{j-1})$$

for $j = l, \ldots, 1$ and $\Psi_{l+1}(\omega, y_l) \equiv 0$.

It is shown in [53] that if $b_j(\omega) - W_{jj-1}(\omega)x \in W_{jj} Y_j$ for all pairs (ω, x) (*relatively complete recourse*) and $\ker(W_{jj}) \cap Y_j^\infty = \{0\}$ for $j = 1, \ldots, l-1$ (where Y_j^∞ denotes the horizon cone[1] of

[1]The horizon cone D^∞ for the convex set $D \subseteq \mathbb{R}^m$ is defined as the set of all elements $x_d \in \mathbb{R}^m$ such that $x + \lambda x_d \in D$ for all $x \in D$ and all $\lambda \in \mathbb{R}_+$. In particular, $D^\infty = \{0\}$ if D is bounded.

Y_j) then there exists a constant $\hat{K}$ such that

$$|f(\omega, x) - f(\tilde{\omega}, x)| \leq \hat{K} \max\{1, \rho, ||\omega||^l, ||\tilde{\omega}||^l\}||\omega - \tilde{\omega}|| \tag{B.14}$$

for all ω, $\tilde{\omega} \in \Omega$ and $x \in X \cap \rho\mathbb{B}$. This implies that $\frac{1}{\hat{K}\max\{1,\rho\}} f(\omega, x) \in \mathcal{F}_{c_{l+1}}$ for all ω, $\tilde{\omega} \in \Omega$ and $x \in X \cap \rho\mathbb{B}$.

It is now straightforward to obtain the following result ([53]). Let $v(P)$ be the optimal value of problem (B.13). Assume that the relatively complete recourse condition for (B.13) is satisfied and that $\ker(W_{jj}) \cap Y_j^\infty = \{0\}$ for $j = 1, \ldots, l-1$. Then there exists a constant $K > 0$ such that the estimate

$$|v(P) - v(Q)| \leq K\zeta_{l+1}(P, Q) \tag{B.15}$$

is valid for any P, $Q \in \mathcal{P}_{l+1}(\Omega)$. (Here $\mathcal{P}_{l+1}(\Omega)$ denotes the set of Borel measures on Ω with finite $(l+1)$-th order moments.)

Specifying the general result (B.11) to the cost function of the form (B.12) with $p = l+1$ we can rewrite the estimate (B.15) for the difference in optimal objective values of problem (B.13) as

$$|v(P) - v(Q)| \leq K\mathring{\mu}_{l+1}(P, Q) = K\hat{\mu}_{\hat{c}_{l+1}}(P, Q), \tag{B.16}$$

where $K > 0$ is some constant.

Appendix C

Scenario reduction algorithms

The goal of scenario reduction algorithms is, given a stochastic optimization problem of the form (1.1) characterized by a discrete measure $P = \sum_{i=1}^{N} p_i \delta_{\omega_i}$ find the discrete measure $Q = \sum_{j=1}^{M} q_i \delta_{\tilde{\omega}_j}$ such that $M < N$ and the difference in the optimal objective values $|v(P) - v(Q)|$ is as small as possible.

If the stochastic optimization problem has the form (B.13) of a linear multi-period problem then, as discussed earlier in Appendix B, under relatively complete recourse assumption, the upper bound (B.16) can be shown to hold. This motivates searching for discrete measures Q that minimize the distance $\hat{\mu}_{l+1}(P, Q)$ (or $\mathring{\mu}_{l+1}(P, Q)$).

Thus the optimal scenario reduction problem based on the method of probability metrics can be formulated as follows [15]. Let $J \subset \{1, 2, \ldots, N\}$ and consider the measure $Q = \sum_{j \notin J} q_j \delta_{\omega_j}$ supported at points ω_j, $j \in \{1, 2, \ldots, N\} \setminus J$. The measure Q is said to be *reduced* from P by deleting scenarios ω_j, $j \in J$ and by assigning new probabilities q_j to the remaining scenarios. The optimal reduction concept proposed in [15] seeks the minimum value of the functional

$$D(J;q) = \hat{\mu}_p \left(\sum_{i=1}^{N} p_i \delta_{\omega_i}, \sum_{j \notin J} q_j \delta_{\omega_j} \right). \tag{C.1}$$

It is shown in [15] that, for set J fixed, the optimal weights q are straightforward to find:

$$q_j = p_j + \sum_{i \in J_j} p_i, \quad \text{for each } j \notin J, \tag{C.2}$$

where $J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min_{j \notin J} c_p(\omega_i, \omega_j)$ for each $i \in J$. The corresponding minimum of the functional $D(J;q)$ is

$$D_J = \min_q \{D(J;q) : q_j \geq 0, \sum_{j \notin J} q_j = 1\} = \sum_{i \in J} p_i \min_{j \notin J} c_p(\omega_i, \omega_j).$$

On the other hand, the optimal choice of the set J of given cardinality $|J| = k$

$$\min_J \{D_J = \sum_{i \in J} p_i \min_{j \notin J} c_p(\omega_i, \omega_j) : J \subset \{1, 2, \ldots, N\}, |J| = k\}$$

is a combinatorial problem, and it is unlikely that efficient solution algorithms for arbitrary value of k are available. However cases $k = 1$ and $k = N - 1$ are easy to solve to optimality and they can be used to formulate heuristic algorithms for other values of k. The *fast forward* scenario reduction algorithms proposed in [29] proceeds as follows.

Algorithm 2: Fast forward selection algorithm.

Step 1**:**

$c_{ku}^{[1]} := c_p(\omega_k, \omega_u),\ k, u = 1, \ldots, N,$

$z_u^{[1]} := \sum_{\substack{k=1 \\ k \neq u}} p_k c_{ku}^{[1]},\ u = 1, \ldots N,$

$u_1 \in \arg \min_{u \in \{1, \ldots, N\}} z_u^{[1]},\ J^{[1]} := \{1, \ldots, N\} \setminus \{u_1\}.$

Step i**:**

$c_{ku}^{[i]} := \min\{c_{ku}^{[i-1]}, c_{ku_{i-1}}^{[i-1]}\},\ k, u \in J^{[i-1]},$

$z_u^{[i]} := \sum_{k \in J^{[i-1]} \setminus \{u\}} p_k c_{ku}^{[i]},\ u \in J^{[i-1]},$

$u_i \in \arg \min_{u \in J^{[i-1]}} z_u^{[i]},\ J^{[i]} := J^{[i-1]} \setminus \{u_i\}.$

Step $n+1$**:**

Redistribution by (C.2)

- **Place of Birth:** Morristown, New Jersey
- **Date of Birth:** February 12, 1980
- **Parents:** Paul and Sally Grace
- **Degrees:** Lehigh University, Bethlehem, PA
  Ph.D. Industrial Engineering, May 2012
  B.S. Materials Science and Engineering, May 2002

TECHNICAL REPORTS

- Grace, D., E. Perevalov. "Information Acquisition for Decision Making: Information Source Models", available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.
- Perevalov, E., D. Grace. "Information Acquisition for Decision Making: Question Difficulty", available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.
- Perevalov, E., D. Grace. "Information Acquisition for Decision Making: Answer Depth", available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.
- Perevalov, E., D. Grace. "Optimization of additional information acquisition in decision making problems: Main framework", available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.
- D. Grace, Perevalov, E. "Optimization of additional information acquisition in decision making problems: Solution Methods", available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.

CONFERENCE PROCEEDINGS

- Perevalov, E., D. Grace. "Information-centric Approach to Decision Making Under Uncertainty", in *Proceedings of the International Workshop of Computer Science and Engineering*, May 15-18, 2008, Moscow, Russia.
- Perevalov, E., D. Grace. "On the Information Aspect of Optimization Under Uncertainty, in *Proceedings of the NSF CMMI Engineering Research and Innovation Conference*, January 7-10, 2008, Knoxville, Tennessee.

ENGINEERING EXPERIENCE

Lehigh University, Industrial and Systems Engineering – Bethlehem, PA
Research Assistant, Fall 2006 – Fall 2008

- Developed novel methods to assess the information content of optimization problems under uncertainty.
- Developed a framework allowing the decision maker to incorporate qualitative knowledge with quantitative solutions.
- Modeled and solved global optimization problems using both a Genetic Algorithm and Simulated Annealing Algorithm implemented in Matlab.
- Researched, designed, and coordinated the purchase of a high performance cluster computer and its software totaling $30,000.
- Developed and implemented a parallel Genetic Algorithm and applied two separate parallel tree search algorithms, CHiPPS and PEBBL, on the high performance cluster.

École Polytechnique Fédéral de Lausanne, Mathématiques – Lausanne, Switzerland
Research Assistant – MyCapital: Electronic Trading, May 2005 – September 2005

- Utilized Bloomberg software to capture equity option tick data for entire Swiss Market Index.
- Lead development of Volatility Dispersion Trading strategy and identified possible arbitrage opportunities.

Lehigh University, Industrial and Systems Engineering – Bethlehem, PA
Integrated Graduate Education and Research Training (IGERT) Fellow, Fall 2002 – Spring 2006

- Awarded the NSF doctoral fellowship in Global Manufacturing and Logistics between Lehigh University and The Wharton School of the University of Pennsylvania.
- Researched analytical and numerical approximations for the early exercise boundary for American put options.
- Implemented a recursive integration procedure in Matlab that improved the accuracy of computed put prices over long time horizons.

TEACHING EXPERIENCE

Lehigh University, Industrial and Systems Engineering – Bethlehem, PA
Instructor – Engineering Probability and Statistics, Spring 2009, Spring 2010

- Prepared course materials and delivered lectures to undergraduate engineers on different introductory topics of Probability Theory.
- Prepared, administered and evaluated students' assignments, midterms, and final.
- Maintained regularly scheduled office hours to advise and assist students.

Lehigh University, Industrial and Systems Engineering – Bethlehem, PA
Instructor – Introduction to Operations Research, Summer 2007

- Prepared course materials and delivered lectures to undergraduate students on different topics in operations research. Covered linear programming and the simplex method, other mathematical models including transportation, network, integer, and non-linear models, Markov chains and Queuing Theory.
- Prepared, administered and evaluated students' assignments, midterms, and final.
- Maintained regularly scheduled office hours to advise and assist students.

PROFESSIONAL AFFILIATIONS

- AMS – American Mathematical Society
- INFORMS – Institute for Operations Research and the Management Sciences
- SIAM – Society of Industrial and Applied Mathematics