LEHIGH UNIVERSITY

ISE Industrial and Systems Engineering

Optimization of Additional Information Acquisition in Decision Making Problems: Main Framework

David Grace
Lehigh University

Eugene Perevalov
Lehigh University

Report: 12T-004

Optimization of Additional Information Acquisition in Decision Making Problems: Main Framework

E. Perevalov and D. Grace

Abstract

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. We propose a framework that relates information characteristics of a source to solution quality characteristics of the problem and formulate 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.

1 Introduction

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 and a complex structure of the feasible region optimization methods are usually used (stochastic [5], robust [3, 2] or, more recently, risk-averse [11, 35, 4]). 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.

The main motivation for the approach proposed here is the need to be able 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.

The approach proposed in this paper 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. It appears reasonable to assume that a source would find it easier to answer some questions compared to others meaning that it would be able to answer easier questions more accurately. On the other hand, information contained in an answer to any question carries a certain value of *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* [18] 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.

In order to realize the overall program sketched in the previous paragraph one would need to (i) develop 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. Item (i) was addressed in [32, 31, 33] where, in particular, the *question difficulty* and *answer depth* functionals were defined and several information source models were proposed. The present article aims to address item (ii) by developing a framework for properly measuring the value of information associated with various answers and relating it to the corresponding information related quantities (i.e. question difficulty and answer depth).

1.1 Related work

The idea of obtaining additional information to improve quality of decision in situations characterized with uncertainty is obviously not entirely new and it has been pursued, for instance, in the area of statistical decision making. Applications to innovation adoption [29], [22], fashion decisions [14] and vaccine composition decisions for flu immunization [26] 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 [13], [12] introduced various models (e.g. effective information model) for accounting for the actual, or effective, amount of information contained in the received observations (see Section 3 for more details). 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 [26], 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 in 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 [19, 27, 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 [37] since had great success in a number of fields – besides communications itself which it revolutionized – that include statistical physics [20, 21], computer vision [39], climatology [30, 38], physiology [23] and neurophysiology [6]. The relatively new field of Generalized Information Theory (see e.g. [24]) is concerned with problems of characterizing uncertainty in frameworks that are more general than classical probability such as Dempster-Shafer theory [36]. There it was shown, for example, In particular, it was shown in [28, 17] 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. On a related note, the proposed approach is based on a theory of information exchange between the decision maker/analyst and information source(s) that is developed in [32, 31, 33]. The latter can be thought of as a development of a general theory of inquiry that goes back to the work of Cox [9, 10]. This line of work received more attention recently resulting in a formulation of the calculus of inquiry [25] that, in particular, constructs a distributive lattice of questions dual to the Boolean lattice of logical assertions. The definition of questions adapted in [32] corresponds to the particular subclass of questions – the partition questions – defined in [25]. Our work in [32, 31, 33] 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 modeling of information sources that lies at the base of the proposed methodology is similar in spirit to analyzing and using information provided by human experts. In many practically relevant applications, the role of information sources will likely be played by human experts. In existing research literature, the problem of optimal usage of information obtained from experts has been addressed mostly in the form of updating the decision maker's beliefs given probability assessment from multiple experts [15, 16, 7, 8] and optimal combining of expert opinions, including experts with incoherent and missing outputs [34]. In the present and related papers [32, 31, 33], the emphasis is on *optimizing* on the particular type of information for the given expert(s) and decision making problem.

1.2 Outline

In the next section, we briefly describe a general formulation of the problem of additional information acquisition as it's interpreted in this paper. In Section 3, we give necessary preliminary information concerning mostly partitions of the parameter space of the problem. Section 4 contains a summary of main results of [32, 31, 33] concerning the process of information exchange between the decision maker/analyst and information source(s). In Section 5, 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 6, 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 7 presents example illustrating the results obtained in the earlier sections. Finally, Section 8 contains a conclusion.

2 Overall Problem Formulation

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

$$\min_{x\in X}\mathbb{E}_P f(\omega,x)=\int_\Omega f(\omega,x)P(d\omega). \tag{1}$$

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) and an information source capable of providing answers to our questions, in obtaining the best possible solution to problem (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) 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

$$L(P,Q,A(Q))=\sum_a \Pr(A(Q)=a)\left(\int_\Omega f(\omega,x^*_{P_a})P_a(d\omega)-\int_\Omega f(\omega,x^*_\omega)P_a(d\omega)\right), \tag{2}$$

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) and a given information source, the question(s) $Q\in\mathcal{Q}$ that would make the corresponding expected loss (2) as small as possible:

$$\min_{Q\in\mathcal{Q}}L(P,Q,A(Q)). \tag{3}$$

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).

3 Preliminaries: partitions of parameter space

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

$$P_C(D) = \frac{P(D \cap C)}{P(C)}, \tag{4}$$

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.

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 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. 2).

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: $\mathbf{C} = \{C'_1, \ldots, C'_r, C''_1, \ldots, C''_s\}$.

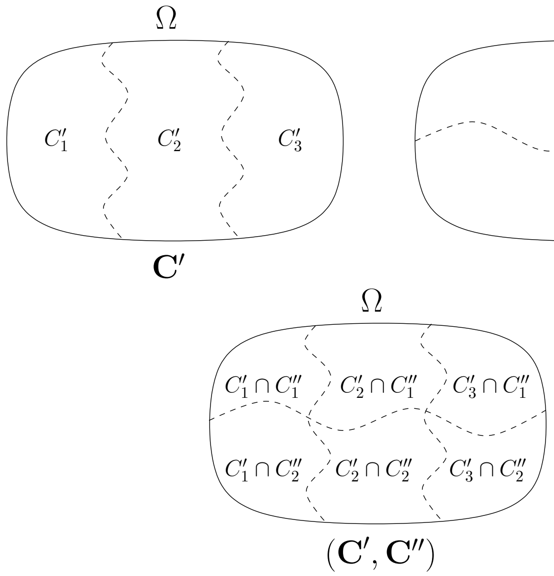

Figure 1: Two partitions of Ω and the corresponding joint partition.

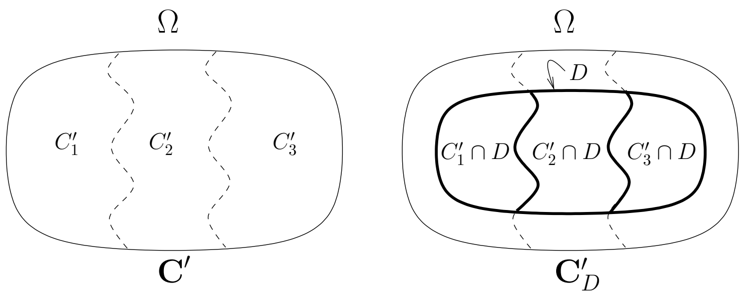

Figure 2: Partition $\mathbf{C}_D$ of set $D \subset \Omega$ induced by a partition $\mathbf{C}$ of Ω.

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.

$$P = \sum_{j=1}^{r} P(C_j) P_{C_j}. \tag{5}$$

4 Information exchange

The model of information exchange between an information source and a decision maker determines the degree of accuracy of information the source can provide to the decision maker as a function of the type of information requested by the latter. Generally speaking, it is reasonable to believe that, for wide class of information sources, the accuracy of additional information they are capable of providing will depend on both the amount of detail contained in that information and the specific content. The latter dependence reflects the source-specific “expertise”: the given information source can “know” more about certain aspects of the (likely values of input data for) problem the decision maker is facing relative to other aspects of the same problem. In other words, the information exchange model has to account for source-specific *cognitive complexity.* Such a model was developed in [32, 31, 33]. It includes three main components: decision maker questions, information source answers and the information source model that relates the quality of answers to difficulty of questions. Below, we describe the main results of [32, 31, 33] that will be needed for later developments presented in this paper.

4.1 Questions and question difficulty functional

A general theory of inquiry was developed in [9, 10, 25]. In particular, in [25] questions were constructed in a general setting from the ordered set of down-sets of assertions, starting from a partially ordered set (lattice) of logical assertions. The class of *partition questions* was considered as a subset of all possible questions that can be given a particularly well-structured answers. The definition of questions adapted in [32] is closely related to partition question of [25]. It identifies a question with a partition $\mathbf{C}$ of the parameter space Ω.

A *difficulty functional* $G(\Omega, \mathbf{C}, P)$ can be associated with any question $\mathbf{C}$. The particular form of $G(\Omega, \mathbf{C}, P)$ can be determined if some reasonable requirements, or, equivalently, *postulates*, are imposed. This was done in [32] where a particular system of postulates that expressed *linearity* and *isotropy* properties of the difficulty functional was proposed. The main theorem proved in [32] derives the general form of the difficulty functional that is required to satisfy such postulates.

Theorem 1 *Let the functional $G(\Omega, \mathbf{C}, P)$ where $\mathbf{C} = \{C_1, \ldots, C_r\}$ satisfy Postulates 1 through 6 (see [32]). 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* $u\colon \Omega \to \mathbb{R}$ *is an integrable nonnegative function on the parameter space* Ω.

In particular, the difficulty of the given question $\mathbf{C}$ depends on, besides the initial probability measure P, the function $u(\cdot)$ defined on the parameter space Ω. This function may be called the *pseudo-temperature* using parallels with thermodynamics (see [32] for more details). The question difficulty then can be interpreted as the amount of *pseudo-energy* associated with question $\mathbf{C}$.

If $\tilde{\mathbf{C}}$ is an arbitrary refinement of $\mathbf{C}$ then the difficulty of the more detailed question $\tilde{\mathbf{C}}$ can be decomposed as ([32])

$$G(\Omega, \tilde{\mathbf{C}}, P) = G(\Omega, \mathbf{C}, P) + G(\Omega, \tilde{\mathbf{C}}_{\mathbf{C}}, P), \tag{6}$$

where the expected *residual difficulty* of $\tilde{\mathbf{C}}$ given a perfect answer to $\mathbf{C}$ is defined as

$$G(\Omega, \tilde{\mathbf{C}}_{\mathbf{C}}, P) = \sum_{C \in \mathbf{C}} P(C) G(C, \tilde{\mathbf{C}}_C, P_C). \tag{7}$$

4.2 Answers and answer depth functional

Given a question $\mathbf{C}$ on Ω, an answer to $\mathbf{C}$ was defined in [31] to be a message $V(\mathbf{C})$ taking values in the set $\{s_1, \ldots, s_m\}$ such that the reception of the value s_k modifies (updates) the initial measure P on Ω to the measure $P^k \equiv P^{V(\mathbf{C})=s_k}$ such that $P^k_{C_j} = P_{C_j}$ for $k = 1, \ldots, m$ and $j = 1, \ldots, r\}$. The latter condition ensures that the answer $V(\mathbf{C})$ is indeed an answer to the question $\mathbf{C}$.

It follows from the above definition 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

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j}, \tag{8}$$

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 (8) gets slightly modified to account for the set $\overline{\hat{C}} = \Omega \setminus \hat{C}$ and takes the form

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j} + \bar{p}_k P_{\overline{\hat{C}}}, \tag{9}$$

where $\sum_{j=1}^{r} p_{kj} + \bar{p}_k = 1$. For pure free-response questions, one has to set $r = 1$ in (9).

The answer *depth* functional $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ for the answer $V(\mathbf{C})$ to question $\mathbf{C}$ measures the amount of *pseudo-energy* that is conveyed by $V(\mathbf{C})$ in response to question $\mathbf{C}$. The general form of $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ can be established if certain (reasonable) requirements it has to satisfy are imposed. This was done in [31] where such requirements – called postulates – were discussed. The particular set of postulates used in [31] was chosen to impose *linearity* and *isotropy* conditions on the answer depth functional. Under these conditions, the following result was obtained (formulated in [31] as a corollary).

Theorem 2 *The answer depth functional $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ 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 $u(C_j) = \frac{1}{P(C_j)} \int_{C_j} u(\omega) dP(\omega)$ and the function $u\colon \Omega \to \mathbb{R}$ is the same function that is used in the question difficulty functional $G(\Omega, \mathbf{C}, P)$.

It can be shown (see [31] for details) that if $V(\mathbf{C})$ is any answer to the question $\mathbf{C}$ then $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ with equality if and only if the answer $V(\mathbf{C})$ is *perfect*, i.e. $P^j = P_{C_j}$ for $j = 1, \ldots, r$. The difficulty of question $\mathbf{C}$ can be written as

$$G(\Omega, \mathbf{C}, P) = Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) + G(\Omega, \mathbf{C}, P_{V(\mathbf{C})}), \tag{10}$$

where

$$G(\Omega, \mathbf{C}, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{1}{P^k(C_j)} \tag{11}$$

can be termed the *residual difficulty* of $\mathbf{C}$ given the answer $V(\mathbf{C})$. Clearly, $G(\Omega, \mathbf{C}, P_{V(\mathbf{C})}) \geq 0$ with the inequality being tight for a perfect answer $V^*(\mathbf{C})$. The residual difficulty $G(\Omega, \mathbf{C}, P_{V(\mathbf{C})})$ can be expressed via coefficients p_{kj} that describe the answer $V(\mathbf{C})$:

$$G(\Omega, \mathbf{C}, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} \sum_{j=1}^{r} v_k p_{kj} u(C_j) \log \frac{1}{p_{kj}} \tag{12}$$

Consider a question $\mathbf{C} = \{C_1, \ldots, C_r\}$, its refinement $\tilde{\mathbf{C}} = \{\tilde{C}_1, \ldots, \tilde{C}_R\}$ and an answer $V(\mathbf{C})$ to $\mathbf{C}$ taking possible values in the set $\{s_1, \ldots, s_k\}$. Let us renumber the subsets in $\tilde{\mathbf{C}}$ in such a way that $\tilde{C}_{l_j} \subseteq C_j$ for $l_j = 1, \ldots, r_j$, $j = 1, \ldots, r$ and $\sum_{j=1}^{r} r_j = R$. Then the difficulty of $\tilde{\mathbf{C}}$ can be written as

$$G(\Omega, \tilde{\mathbf{C}}, P) = Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) + G(\Omega, \tilde{\mathbf{C}}, P_{V(\mathbf{C})}), \tag{13}$$

where the residual difficulty of $\tilde{\mathbf{C}}$ given the answer $V(\mathbf{C})$ to $\mathbf{C}$ has the form

$$G(\Omega, \tilde{\mathbf{C}}, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} \sum_{j=1}^{r} v_k p_{kj} \sum_{l_j=1}^{r_j} u(\tilde{C}_{l_j}) P_{C_j}(\tilde{C}_{l_j}) \log \frac{1}{p_{kj} P_{C_j}(\tilde{C}_{l_j})} \tag{14}$$

It turns out to be convenient to consider the class of imperfect answers for which the degree of imperfection is described by a single error probability α – the *quasi-perfect* answers [31]. For a quasi-perfect answer $V_\alpha(\mathbf{C})$ to a (complete) question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the coefficients p_{kj} have the form

$$p_{kj} = (1 - \alpha)\delta_{k,j} + \alpha P(C_j), \tag{15}$$

for $k = 1, \ldots, r$ and $j = 1, \ldots, r$, and the updated measure P^k is simply

$$P^k = \alpha P + (1 - \alpha) P_{C_k}. \tag{16}$$

for $k = 1, \ldots, r$. Clearly, for $\alpha = 0$ a quasi-perfect answer to $\mathbf{C}$ becomes a perfect one. It can be shown (see [31]) that the answer depth functional for a quasi-perfect answer $V_\alpha(\mathbf{C})$ to question $\mathbf{C}$ can be written as

$$\begin{aligned} Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})) = \sum_{k=1}^{r} u(C_k)P(C_k)(1 - \alpha + \alpha P(C_k)) \log \frac{1 - \alpha + \alpha P(C_k)}{P(C_k)} \\ + \alpha \log \alpha \sum_{k=1}^{r} u(C_k)P(C_k)(1 - P(C_k)), \end{aligned} \tag{17}$$

which is easily seen to reduce to $G(\Omega, \mathbf{C}, P)$ for $\alpha = 0$ and vanish for $\alpha = 1$.

4.3 Information source models

The pseudo-temperature function $u(\cdot)$ on the parameter space Ω characterizes the source specific relative difficulty of questions "located" in various regions of Ω. Knowing the pseudo-temperature function lets one calculate the difficulty of any question up to an overall scale (multiplicative constant). The goal of an information source model is to describe what is hard for the given information source and, in particular to set that overall scale. Different information source models were considered in [33]. Specifically, information source models are based on the following hypothesis.

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

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

where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-decreasing function of its argument that's bounded from above.

An information source model simply specifies the form of function $h(\cdot)$. In principle the latter can be either source specific or universal, with the latter option likely being more convenient in applications. Assuming the form of $h(\cdot)$ is universal, one would then postulate it and use the observed information sources' performance to estimate the parameters of function $h(\cdot)$ as well as the source specific pseudo-temperature functions.

The simplest information source model considered in [33] is the so called *simple capacity model* given by

$$h(x) = \begin{cases} x & \text{if } x \leq Y_s \\ Y_s & \text{if } x > Y_s. \end{cases} \tag{18}$$

which is fully characterized by the single parameter Y_s which has the meaning of the information source capacity.

The most apparent drawback of model (18) is that it predicts that the source would provide a perfect answer to any question whose difficulty does not exceed the source capacity. The *linear modified capacity model* described by

$$h(x) = \begin{cases} bx & \text{if } x \leq \frac{Y_s}{b} \\ Y_s & \text{if } x > \frac{Y_s}{b} \end{cases} \tag{19}$$

removes this drawback at the expense of one extra parameter $b \leq 1$ that has to be estimated. Several slightly more complicated models were proposed in [33].

The values of model parameters as well as pseudo-temperature functions for information sources can be estimated from the observed sources' performance on some set of sample questions. Optimization based formulations for such estimation were also proposed in [33].

It is easy to see that multiplying the pseudo-temperature function $u(\cdot)$ has the effect of multiplying both the question difficulty and the answer depth by the same number and is equivalent to a choice of units of pseudo-energy. It turns out to be convenient to use two different conventions in this regard.

- The convention in which $\int_\Omega u(\omega)\, d\omega = 1$. Here the units of pseudo-energy are chosen in such a way that, for constant $u(\omega)$, the pseudo-energy coincides with entropy making it convenient to make use of the standard intuition about entropy and information (i.e. 1 unit (bit) of information is equivalent to being able to perfectly distinguish between two equally likely alternatives).
- The convention in which each source has unit capacity ($Y_s = 1$). This choice of units of pseudo-energy makes it convenient to compare the "depth of knowledge" of different information sources to each other by directly comparing their respective pseudo-temperature values at the same points of the parameter space.

5 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). 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 $g_{P,\mathbf{C}}$: $g_{P,\mathbf{C}}(\omega) = x^*_{P_{C_j}}$ for all $\omega \in C_j$, $j = 1, \ldots, r$. (Here $x^*_{P_{C_j}}$ is an optimal solution of problem (1) 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 $g_{P,\mathbf{C}}$ 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:

$$S(x,P) = \mathbb{E}_P f(\omega,x) - \mathbb{E}_P f(\omega,x^*_P) = \int_\Omega (f(\omega,x) - f(\omega,x^*_P))P(d\omega), \tag{20}$$

i.e. suboptimality of x w.r.t. P is the difference in objective values of problem (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

$$L(g,P) = \mathbb{E}_P f(\omega,g(\omega)) - \mathbb{E}_P f(\omega,x^*_\omega) = \int_\Omega (f(\omega,g(\omega)) - f(\omega,x^*_\omega))P(d\omega). \tag{21}$$

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:

$$B(g,P) = \mathbb{E}_P f(\omega,x^*_P) - \mathbb{E}_P f(\omega,g(\omega)) = \int_\Omega (f(\omega,x^*_P) - f(\omega,g(\omega)))P(d\omega). \tag{22}$$

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 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

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

□

The statement of Lemma 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 takes 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 $\{P_{C_j}\}$, $j = 1, \ldots, r$:

$$f(P_{\mathbf{C}}) = \sum_{j=1}^{r} P(C_j) f(P_{C_j}), \tag{23}$$

and by $f(P_{V(\mathbf{C})})$ – the expected value of $f(\mathbf{C})$ over the set of updated measures $\{P^k\}$, $k = 1, \ldots, m$:

$$f(P_{V(\mathbf{C})}) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) f(P^k) = \sum_{k=1}^{m} v_k f(P^k), \tag{24}$$

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 (23) and (24).

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

$$S(x, P_{\mathbf{C}}) = \sum_{i=1}^{s} P(C_j) S(x, P_{C_j}), \tag{25}$$

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

$$S(x, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k S(x, P^k). \tag{26}$$

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

$$L(g, P_{\mathbf{C}}) = \sum_{j=1}^{r} P(C_j) L(g, P_{C_j}), \tag{27}$$

and

$$B(g, P_{\mathbf{C}}) = \sum_{j=1}^{r} P(C_j) B(g, P_{C_j}), \tag{28}$$

respectively.

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

$$L(g, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k L(g, P^k), \tag{29}$$

and

$$B(g, P_{V(\mathbf{C})}) = \sum_{k=1}^{m} v_k B(g, P^k), \tag{30}$$

respectively.

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

Lemma 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 (27) 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) can be decomposed in a convenient way.

Lemma 3 *For any map $g_{\mathbf{C},P} \in \mathcal{C}$, the EVPI $L(g_P, P)$ of the problem (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 (23) for functionals of measures, and (d) follows from Lemma 2. □

6 Pseudo-energy-loss efficient frontier and optimizing additional information acquisition

6.1 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

$$
\begin{aligned}
&\underset{g \in \mathcal{G}}{\text{minimize}} \quad L(g, P) \\
&\text{subject to} \quad G(\Omega, \mathbf{C}(g), P) \le \gamma
\end{aligned}
\tag{31}
$$

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 1 $\mathcal{O} \subset \mathcal{C}$

Proof: Let $g = (\mathbf{C}, I)$ where $I = \{x_1, x_2, \ldots, x_r\}$. 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 $I' = \{x^*_{P_{C_1}}, x_2, \ldots, x_r\}$. 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 $g \notin \mathcal{O}$. □

It follows from Proposition 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.

6.2 Optimal information acquisition

Let us now address the optimal information acquisition problem (3): what question(s) need to be asked the given information source in order to obtain the minimum possible loss for (1). Given a question $\mathbf{C} = \{C_1, \ldots, C_r\}$ to an information source and its answer $V(\mathbf{C})$ taking values in the set $\{s_1, \ldots, s_m\}$, we denote by $\mathcal{L}(s_k)$, $k = 1, \ldots, m$ 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

$$\mathcal{L}(V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k)\mathcal{L}(s_k), \tag{32}$$

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.

6.2.1 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 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_k 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

$$\mathcal{L}(s_j) = \int_{C_j} (f(\omega, x^*_{P_{C_j}}) - f(\omega, x^*(\omega)))\, dP_{C_j}(\omega). \tag{33}$$

Taking the expectation of (33) 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 (27), and (d) follows from Lemma 2. □

Combining the result of Proposition 2 with Lemma 2 (valid for any $g \in \mathcal{G}$) and Lemma 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 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 $\mathbf{C}$.

6.2.2 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 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 (**??**), (b) follows from the fact that measure P_{C_j} vanishes outside of C_j, (c) follows from (**??**), (d) follows from the definitions (20) and (22) of suboptimality and gain, and (e) follows from the definitions (26) and (30). □

Combining the result of Lemma 4 with that of Lemma 3, we obtain a useful decomposition of the EVPI of the original problem which we formulate as a corollary.

Corollary 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 3 *Let $\mathbf{C} = \{C_1, \dots, C_r\}$ 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

$$\mathcal{L}(s_k) = L(g_{P^k}, P^k), \tag{34}$$

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 (34) into (32):

$$\mathcal{L}(V(\mathbf{C})) = \sum_{k=1}^{m} v_k L(g_{P^k}, P^k). \tag{35}$$

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

$$\begin{aligned} L(g_P, P) &= \int_\Omega (f(\omega, x^*_P) - f(\omega, x^*_\omega))P(d\omega) = \sum_{k=1}^{m} v_k \int_\Omega (f(\omega, x^*_P) - f(\omega, x^*_\omega))P^k(d\omega) \\ &= \sum_{k=1}^{m} v_k \int_\Omega (f(\omega, x^*_P) - f(\omega, x^*_\omega) + f(\omega, x^*_{P^k}) - f(\omega, x^*_{P^k}))P^k(d\omega) \\ &= \sum_{k=1}^{m} v_k \int_\Omega (f(\omega, x^*_P) - f(\omega, x^*_{P^k}))P^k(d\omega) + \sum_{k=1}^{m} v_k \int_\Omega (f(\omega, x^*_{P^k}) - f(\omega, x^*_\omega))P^k(d\omega) \\ &= \sum_{k=1}^{m} v_k S(x^*_P, P^k) + \sum_{k=1}^{m} v_k L(g_{P^k}, P^k) = S(x^*_P, P_{V(\mathbf{C})}) + \sum_{k=1}^{m} v_k L(g_{P^k}, P^k) \end{aligned} \tag{36}$$

Comparing (35) with (36) we can obtain

$$\mathcal{L}(V(\mathbf{C})) = L(g_P, P) - S(x^*_P, P_{V(\mathbf{C})}). \tag{37}$$

Finally, using the decomposition of EVPI of Corollary 2 in (37) 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 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 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 3 with that of Corollary 2. The result is formulated as a corollary.

Corollary 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*

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

6.3 Pseudo-energy-loss correspondence

Comparing results obtained in this section with the corresponding pseudo-energy values discussed in Section 4 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 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-energy functional 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

$$\begin{aligned} \underset{\mathbf{C}}{\text{minimize}} \quad & \mathcal{L}(V(\mathbf{C})) \\ \text{subject to} \quad & Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = h(G(\Omega, \mathbf{C}, P)) \end{aligned} \tag{38}$$

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 2 (for perfect answers) or Proposition 3 (for imperfect answers).

If a source is capable of perfect answers (for instance, in the simple linear model) solution of problem (38) 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 (38) 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 1: Correspondence between pseudo-energy and loss related quantities.

The correspondence between pseudo-energy and loss quantities shown in Table 1 can be illustrated by comparing decompositions of the exhaustive question difficulty $G(\Omega, \mathbf{C}_f(P), P)$ (expression (39)) and the EVPI $L(g_P, P)$ (expression (40)) on the other hand. It is also shown in Fig. 3.

$$\underbrace{Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) + \overbrace{G(\Omega, \mathbf{C}, P_{V(\mathbf{C})})}^{}}_{G(\Omega,\mathbf{C},P)} \overbrace{+ G(\Omega, \mathbf{C}_f(P)_{\mathbf{C}}, P)}^{G(\Omega,\mathbf{C}_f(P),P_{V(\mathbf{C})})} = G(\Omega, \mathbf{C}_f(P), P) \tag{39}$$

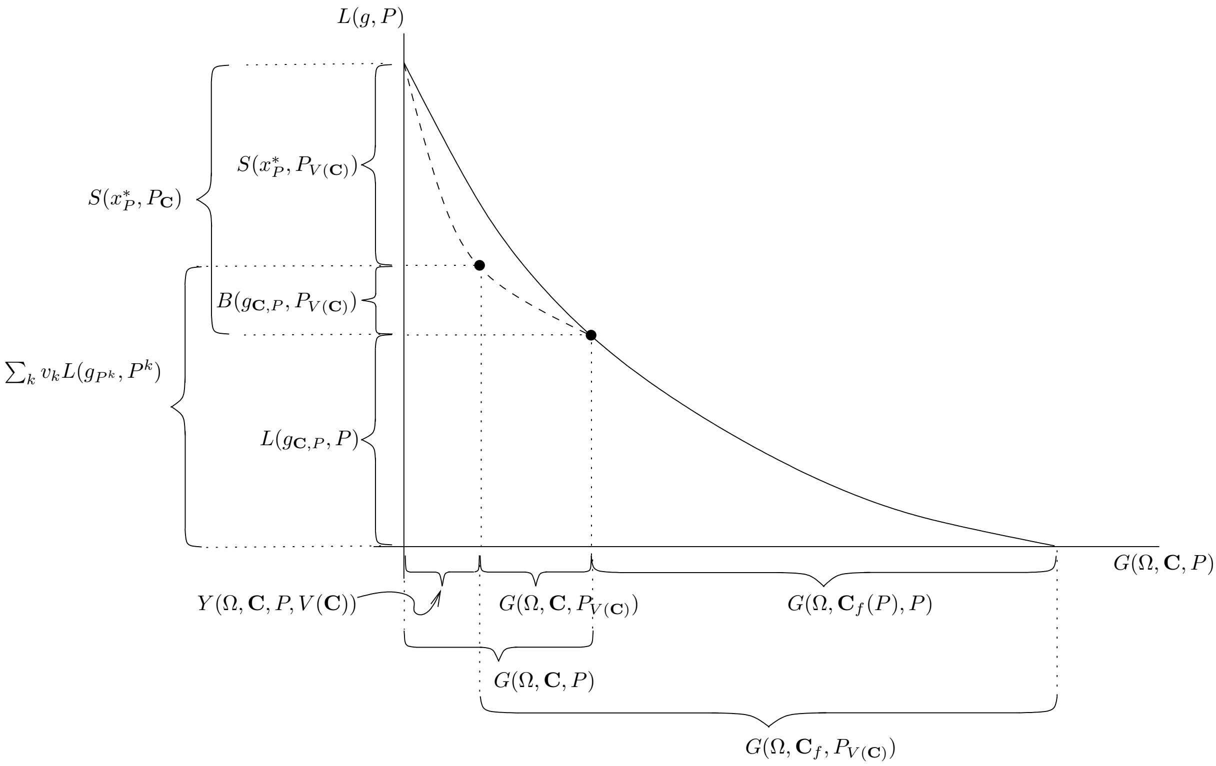

Figure 3: The efficient frontier and correspondence between pseudo-energy and objective function (loss) quantities. A Pareto-optimal map $g \in \mathcal{O}$ on the efficient frontier is shown.

$$\underbrace{S(x^*_P, P_{V(\mathbf{C})}) + \overbrace{B(g_{\mathbf{C},P}, P_{V(\mathbf{C})})}^{\sum_{k=1}^m v_k L(g_{P^k}, P^k)}}_{S(x^*_P, P_{\mathbf{C}}) = B(g_{\mathbf{C},P}, P)} \overbrace{+ L(g_{\mathbf{C},P}, P)} = L(g_P, P) \tag{40}$$

7 Examples

7.1 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) 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. 4 for an illustration).

- $g_1 = \big(\mathbf{C}, \{\frac{a}{4}, \frac{3a}{4}\}\big) = g_{\mathbf{C},P}$. 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 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: $g_2 \notin \mathcal{C}$.
- $g_3 = \big(\mathbf{C}', \{\frac{3a}{8}, \frac{5a}{8}\}\big) = g_{\mathbf{C}',P}$. 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 $g_3 \in \mathcal{C}$.

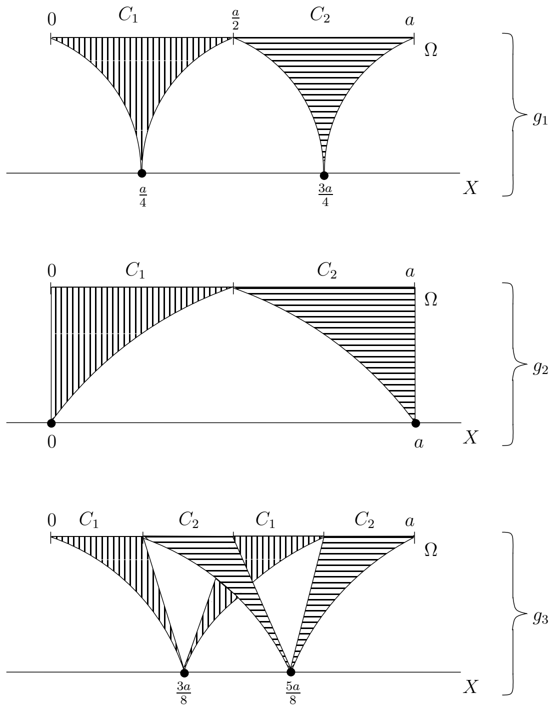

Figure 4: Maps g_1, g_2 and g_3. The partitions for g_1 and g_2 consist of connected sets only. Each element of the partition for g_3 consists of two connected sets.

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,

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

Fig. 5 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.

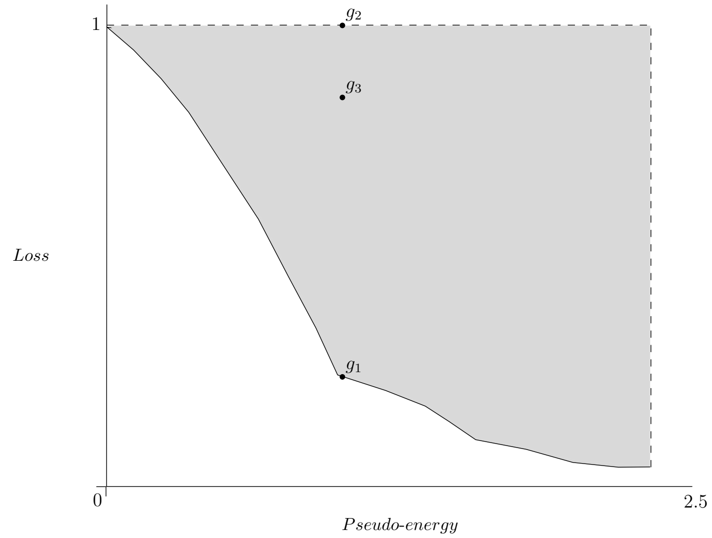

Figure 5: Maps g_1, g_2 and g_3 on (G, L) coordinate plane. All possible maps for this problem lie in the shaded region, at or above the efficient frontier.

Since $g_1, g_3 \in \mathcal{C}$ we have (as Lemma 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) $f(\omega, x)$ 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 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.

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{r} w_j^3 \\ \text{subject to} \quad & -\sum_{j=1}^{r} u(C_j) w_j \log w_j \leq h \\ & \sum_{j=1}^{r} w_j = 1 \\ & w_j \geq 0, \quad j = 1, \ldots, r, \end{aligned} \tag{41}$$

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, (41) 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,

$$u(C_j) = 2\tilde{w}_j + w_j, \tag{42}$$

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. 6.

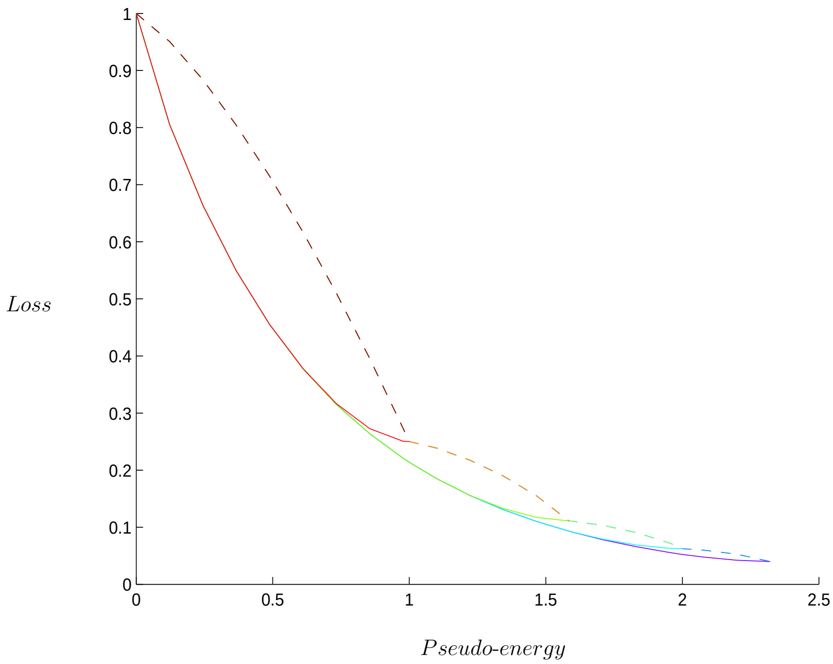

Figure 6: Efficient frontier for the toy example: constant pseudo-temperature case (dotted line) and linear pseudo-temperature case (solid line).

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 (16).

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 **??**, is

$$L(g_P, P) - S(x^*_P, P_{V(\mathbf{C})}) = \frac{a^2}{12} - \frac{a^2}{12}(1 - w_1^3 - w_2^3)(1-\alpha)^2 = \frac{a^2}{12}\left(1 - (1 - w_1^3 - w_2^3)(1-\alpha)^2\right) \quad (43)$$

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. 7 shows the dependence of the expected loss (43) 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.

7.2 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 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

$$f(\omega, x) = cx + b[x - \omega]_+ + h[\omega - x]_+, \quad (44)$$

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

$$\min_x \mathbb{E}_P f(\omega, x), \quad (45)$$

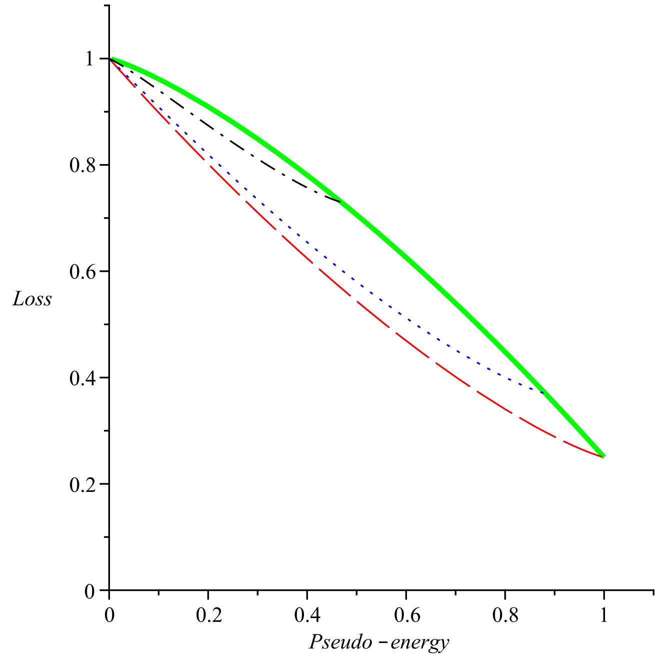

Figure 7: Dependence of the expected loss on the added information for $r = 2$ partitions. The solid curve corresponds to the error-free message case with w_1 varying from 0 to 0.5. The dashed line shows the $w_1 = w_2 = 0.5$ case with α varying from 1 to 0 (from left to right on the figure). The dotted 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.

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

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

The efficient frontier, just like in the previous example can be found by solving the optimization problem (41). Fig. 8 shows the efficient frontier for the case of constant pseudo-temperature function

Figure 8: Efficient frontier for the inventory example: constant pseudo-temperature case (dotted line) and linear increasing pseudo-temperature case (solid line).

which leads to $u(C_j) = 1$ for $j = 1, \dots, r$ and for the case of linear increasing pseudo-temperature function $u(\omega) = \frac{2}{a}\omega$ which leads to $u(C_j) = 2\tilde{w}_j + w_j$, $j = 1, \dots, r$.

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 optimal solutions to (45) with the original measure P replaced with P^k can be shown to be

$$x^*_{P1} = \begin{cases} \frac{w_1 a}{1-\alpha w_2} \cdot \frac{b-c}{b+h} & \text{if } \alpha < \frac{1}{w_2} \cdot \frac{c+h}{b+h} \\ \frac{a}{\alpha}\left(\alpha - \frac{c+h}{b+h}\right) & \text{if } \alpha \geq \frac{1}{w_2} \cdot \frac{c+h}{b+h}, \end{cases} \tag{46}$$

and

$$x^*_{P2} = \begin{cases} a\left(1 - \frac{w_2}{1-\alpha w_1} \cdot \frac{c+h}{b+h}\right) & \text{if } \alpha < \frac{1}{w_1} \cdot \frac{b-c}{b+h} \\ \frac{a}{\alpha} \cdot \frac{b-c}{b+h} & \text{if } \alpha \geq \frac{1}{w_1} \cdot \frac{b-c}{b+h}. \end{cases} \tag{47}$$

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

$$B(g_{\mathbf{C},P}, P_{V(\mathbf{C})}) + L(g_{\mathbf{C},P}, P) = L(g_P, P) - S(x^*_P, P_{V(\mathbf{C})}), \tag{48}$$

and the pseudo-energy content of answer $V_\alpha(\mathbf{C})$ is simply $Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C}))$ given by (17). 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 (48) 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. 9 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.

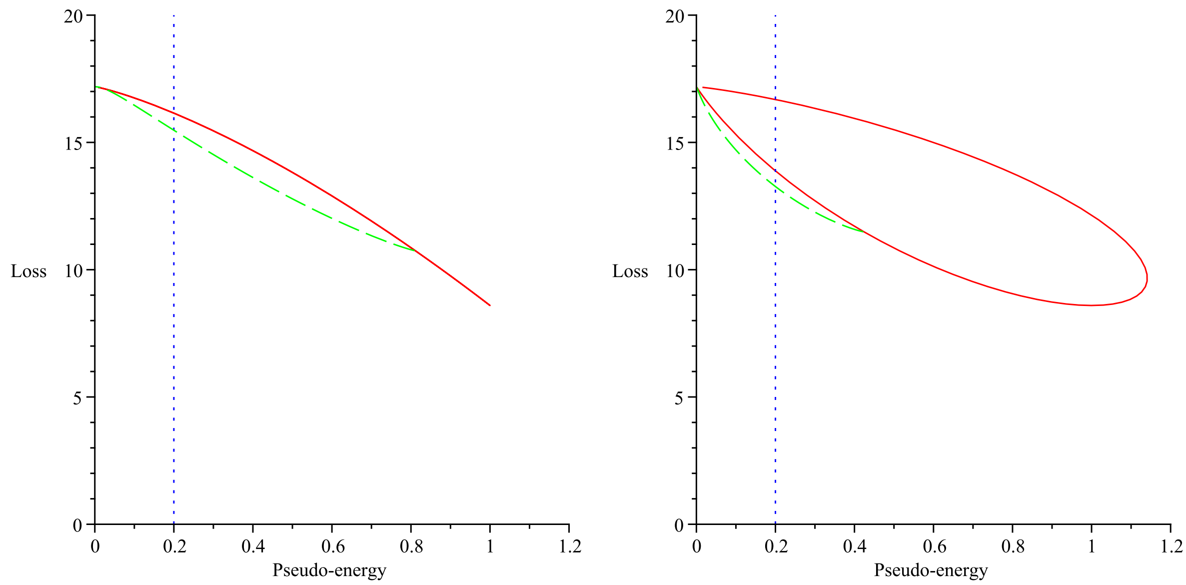

Figure 9: Loss vs. pseudo-energy (for $r = 2$ questions only) for a source with constant pseudo-temperature (left) and a source with linear increasing pseudo-temperature (right). On both plots, the solid line is obtained by varying the parameter w_1 from 0 to 1. The dashed line is obtained by fixing 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 $\mathcal{L}(V(\mathbf{C}))$.

8 Conclusion

In this article, we built on results obtained in [32, 31, 33] 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 Depth-Minimum loss coordinate plane and the problem of optimal information acquisition can in principle be solved by finding – among all answer to all possible questions – the answer (and the corresponding question) that would yield the smallest 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 Difficulty-Minimum 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 smallest 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 Difficulty-Minimum 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. This is the main subject of a companion paper [**?**].

References

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

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

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

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

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

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

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

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

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

[10] — 1961. *The Algebra of Probable Inference*. Johns Hopkins Press, Baltimore.

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

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

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

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

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

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

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

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

[19] Huh, W. T. and P. Rusmevichientong 2009. A nonparametric asymptotic analysis of inventory planning with censored demand. *MOOR* **34**(1), 103–123.

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

[21] — 1957. Information theory and statistical mechanics II. *Phys. Rev.* **108**, 171–190.

[22] Jensen, R. 1988. Information cost and innovation adoption policies. *Management Sci.* **34**(2), 230–239.

[23] Katura, T., N. Tanaka, A. Obata, H. Sato, and A. Maki 2006. Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics. *NeuroImage* **31**(4), 1592–1600.

[24] Klir, G. J. and D. Harmanec 1996. Generalized information theory: Recent developments. *Kybernetes* **25**(7/8), 50–66.

[25] Knuth, K. H. 2005. Lattice duality: The origin of probability and entropy. *Neurocomputing* **67**, 245–274.

[26] Kornish, L. J. and R. L. Keeney 2008. Repeated commit-or-defer decisions with a deadline: The influenza vaccine composition. *Oper. Res.* **56**(3), 527–541.

[27] Levi, R., R. O. Roundy, and D. B. Shmoys 2007. Provably near-optimal sampling-based policies for stochastic inventory control models. *MOOR* **32**(4), 821–839.

[28] Maeda, Y. and H. Ichihashi 1993. An uncertainty measure with monotonicity under the random set inclusion. *Int. J. Gen. Syst.* **21**(4), 379–392.

[29] McCardle, K. F. 1985. Information acquisition and the adoption of new technology. *Management Sci.* **31**(11), 1372–1389.

[30] Mokhov, I. I. and D. A. Smirnov 2006. El Niño-Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices. *Geophys. Res. Lett.* **33**. L03708.

[31] Perevalov, E. and D. Grace 2011. Information acquisition for decision making: Answer depth. Submitted for publication.

[32] — 2011. Information acquisition for decision making: Question difficulty. Submitted for publication.

[33] — 2011. Information acquisition for decision making: Source models. Submitted for publication.

[34] Predd, J. B., D. N. Osherson, S. R. Kulkarni, and H. V. Poor 2008. Aggregating probabilistic forecasts from incoherent and abstaining experts. *Decision Anal.* **5**(4), 177–189.

[35] Ruszczyński, A. and A. Shapiro 2006. Optimization of convex risk functions. *Mathematics of Operations Research* **31**, 433–452.

[36] Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.

[37] Shannon, C. E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* **27**, 379–423, 623–659.

[38] Verdes, P. F. 2005. Assessing causality from multivariate time series. *Phys. Rev. E* **72**. 026222.

[39] Viola, P. A. 1995. Alignment by maximization of mutual information. A.I. Technical Report 1548, Massachusetts Institute of Technology.