European Journal of Operational Research 289 (2021) 595–610

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Probabilistic sensitivity measures as information value

Emanuele Borgonovo[a,*], Gordon B. Hazen[b], Victor Richmond R. Jose[c], Elmar Plischke[d]

[a] Bocconi University and BIDSA, Milan, Italy
[b] Northwestern University, Evanston, IL, USA
[c] Georgetown University, Washington, DC, USA
[d] Clausthal University of Technology, Clausthal-Zellerfeld, Germany

ARTICLE INFO

Article history:
Received 12 November 2019
Accepted 6 July 2020
Available online 12 July 2020

Keywords:
Decision support systems
Information value
Probabilistic sensitivity analysis
Renyi's postulates

ABSTRACT

Decision makers increasingly rely on forecasts or predictions generated by quantitative models. Best practices recommend that a forecast report be accompanied by a sensitivity analysis. A wide variety of probabilistic sensitivity measures have been suggested; however, model inputs may be ranked differently by different sensitivity measures. Is there some way to reduce this disparity by identifying what probabilistic sensitivity measures are most appropriate for a given reporting context? We address this question by postulating that importance rankings of model inputs generated by a sensitivity measure should correspond to the information value for those inputs in the problem of constructing an optimal report based on some proper scoring rule. While some sensitivity measures have already been identified as information value under proper scoring rules, we identify others and provide some generalizations. We address the general question of when a sensitivity measure has this property, presenting necessary and sufficient conditions. We directly examine whether sensitivity measures retain important properties such as transformation invariance and compliance with Renyi's Postulate D for measures of statistical dependence. These results provide a means for selecting the most appropriate sensitivity measures for a particular reporting context and provide the analyst reasonable justifications for that selection. We illustrate these ideas using a large scale probabilistic safety assessment case study used to support decision making in the design and planning of a lunar space mission.

1. Introduction

Forecasts or predictions generated by quantitative models support decision makers in areas ranging from business planning (Baucells & Borgonovo, 2013) to climate change modeling (Stehfest et al., 2019). Frequently, these models are built to estimate a key quantity of interest (Y), which is one of the inputs to a panel where representative agents conduct the decision making process following a pre-established protocol (French & Argyris, 2018). The analyst who develops or implements the simulation is expected to provide a forecast of Y, which can be a point estimate, a quantile, or a cumulative distribution function of Y. Best practices recommend that such a report be accompanied by a sensitivity analysis that provides a description of the level of uncertainty in the forecast and that identifies what model inputs are the drivers of the forecast, and are therefore candidates for additional information acquisition.

Common approaches to sensitivity analysis study the deterministic variation of the quantity of interest about a base value or best estimate. This type of analysis is at the basis of popular tools such as tornado diagrams (Howard, 1988) or spider plots (Eschenbach, 1992). Analysts also have the option of assigning probability distributions to uncertain model inputs, and of using these distributions to construct numerical measures of sensitivity, which we synonymously refer to as *probabilistic* or *probabilistic* sensitivity measures. However, the analyst may find a variety of such sensitivity measures to use, e.g., variance-based, (Saltelli & Tarantola, 2002; Wagner, 1995), quantile-based (Browne, Fort, Iooss, & Le Gratiet, 2017), distribution-based (Gamboa, Klein, & Lagnoux, 2018), and some authors (Felli & Hazen, 1998; 1999; Oakley, 2009; Strong, Oakley, & Brennan, 2014) advocate the use of value of information. The very variety of available sensitivity measures could be a stumbling block for the analyst: Which is the right one to use?

One of the difficulties associated with complex decision making problems is that analysts may be required to work in contexts in which there is no specified objective function or set of alternatives, perhaps due to a decision having already been made (Eschenbach, 1992). In these cases, because there is no explicit comparison of

* Corresponding author.
E-mail addresses: emanuele.borgonovo@unibocconi.it (E. Borgonovo), gbh305@northwestern.edu (G.B. Hazen), vrj2@georgetown.edu (V.R.R. Jose), elmar.plischke@tu-clausthal.de (E. Plischke).

https://doi.org/10.1016/j.ejor.2020.07.010
0377-2217/

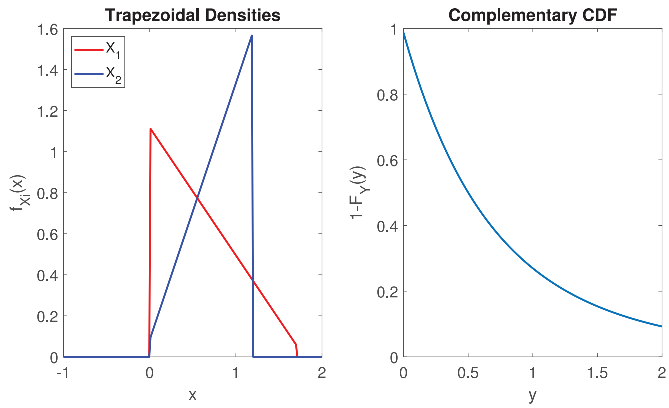

Fig. 1. Densities for the contributing rates X_1, X_2 and the resulting complementary cumulative distribution $1 - F_Y(y)$ for the forecast Y. The rate X_1 has a trapezoidal distribution on [0, 1.7] with mean 0.595, and X_2 has a trapezoidal distribution on [0,1.2] with mean 0.78.

alternatives and no objective, a direct calculation of value of information seems precluded. However, this impression is mistaken because, as noted by Bernardo and Smith (2000), one can take the alternatives to be all possible reports (e.g., all point estimates, quantiles, densities functions or cumulative distribution functions), and the objective function to be an expected proper scoring rule over such reports. Scoring rules, which we discuss further in the next section, measure the accuracy of forecast reports given data. As we show in this paper, information value under a proper scoring rule is in fact a probabilistic sensitivity measure as defined by Borgonovo, Hazen, and Plischke (2016). Here we make two key suggestions: Not only (a) can analysts justify their use of a particular probabilistic sensitivity measure by pointing out that it is information value under a scoring rule, but also (b) which scoring rule it derives from (or whether it derives from one at all) can be used to support or undermine its use compared to other measures. This is illustrated by the following example.

Example 1. Let Y denote the time until a critical event occurs and suppose that Y is a random variable exponentially distributed with rate parameter $\lambda = \lambda_0 + X_1 + X_2$, where $\lambda_0 = 0.05$ is the base rate and X_1, X_2 are independent rate additions from two possible contributing factors. We suppose that data in combination with expert opinions allow the analyst to assess the support and cumulative distribution function of such a random vector. Specifically, the cumulative distribution functions for X_1 and X_2 are reported in Fig. 1. The analyst then wishes to report an estimate $\widehat{F}_Y$ for the cumulative distribution of Y. Note $\widehat{F}_Y$ is not elicited from the experts - it is analytically determined from the elicited probability measures of X_1 and X_2. The resulting cumulative distribution estimate $\widehat{F}_Y$ is also shown. (Appendix A reports all analytical details.)

Investigating the sensitivity of this report to the inputs X_1, X_2, the analyst might use a sensitivity measure based, for example, on the expected reduction in the variance of Y associated with learning each input. The calculations show that learning X_2 reduces variance by 19.1% more than learning X_1, leading to the apparent conclusion that the problem is more sensitive to X_2 than X_1. In fact, it is known (Bernardo & Smith, 2000) that variance reduction is the value of information in the problem of choosing a *point estimate* $\hat{Y}$ under a quadratic scoring rule. From this perspective, there is a mismatch with the problem context, where the goal is to accurately estimate F_Y. Instead, we ask what is more valuable in accurately estimating F_Y, and this requires a scoring rule that can evaluate the estimate $\widehat{F}_Y$ rather than an estimate $\hat{Y}$ that that will never be made. One such rule is the CRPS score (Gneiting, 2011). Using the sensitivity measure arising as information value under this proper scoring rule, it turns out that the value of learning X_2 is only 88.5% of the value of learning X_1, leading to the opposite conclusion, that the distribution-reporting problem is really more sensitive to X_1, not X_2. Measuring sensitivity by variance reduction is therefore not only inappropriate in this problem context, but also leads to misleading conclusions.

The point of this example is not just that different probabilistic sensitivity measures may rank uncertain inputs differently – they may – but in addition that such conflicts can be resolved by examining the appropriateness of any underlying scoring rule for the given reporting context.

The reader should avoid misunderstanding the distinctions we make here, which for clarity we continue to state in terms of this simple example:

1. We are not assuming or necessarily advocating that the elicitation procedures for distributions of inputs X_1 and X_1 involves the use of any scoring rule for expert training or feedback. We merely assume that sound, established procedures for probability elicitation are followed (e.g., Garthwaite, Kadane, & O'Hagan (2005)), which may include or not the use of a scoring rule.
2. Rather, our recommended use of scoring rules occurs only implicitly, and at the next higher level, and not for elicitation of $\widehat{F}_Y$ - recall it is not elicited but determined analytically from the elicited distributions of X_1, X_2. Instead, scoring rules enter in the evaluation of the sensitivity of the cumulative distribution F_Y to the uncertain inputs X_1, X_2. We are advocating that whatever probabilistic sensitivity measure the analyst chooses, it be information value under some scoring rule appropriate for the reporting context. To achieve this, the analyst might first pick an appropriate such scoring rule, and derive the sensitivity measure that is information value under this rule; or more likely might just examine whatever sensitivity measure was chosen, to determine whether it arises as information value from a scoring rule, and if so, whether that rule is appropriate for the reporting context. To do so, the analyst might examine the literature on this topic, including the present paper. If multiple sensitivity measures are considered, they might be compared on these grounds. Measures with no underlying scoring rule or inappropriate underlying rules should be reconsidered.

To facilitate this process, this paper provides some fundamental results relating scoring rules with probabilistic sensitivity measures that arise (or do not arise) as information value from such rules. As already mentioned, we prove that, under any posited scoring rule, the corresponding information value is a probabilistic sensitivity measure falling into the common rationale of probabilistic sensitivity measures studied recently in Borgonovo et al. (2016). This provides a wide class of measures that can be used in practice and have a decision-theoretic interpretation. In addition to variance reduction, as mentioned above, we point out other popular sensitivity measures already in use that are information value under some scoring rule. However, some popular methods already in use do not fall under this paradigm, opening the question of whether they still possess an information value interpretation. We then investigate whether any particular probabilistic sensitivity measure for model inputs arises as information value under some proper scoring rule. We provide sufficient conditions for this to happen in the form of a convexity condition under a particular but still quite general type of reporting context. We also supply necessary conditions in arbitrary reporting contexts.

In our development, we obtain also additional results. In particular we show that, when the analyst's report is the cumulative distribution function, then information value under the CRPS score coincides with Szekely's energy statistic (Szekely & Rizzo, 2017). This quantity is a topic of notable interest in recent statistical and

Table 1
Examples of probabilistic sensitivity measures encompassed by the rationale in (1). Here, f_Y denotes the density, F_Y the cumulative distribution function of random variable Y, possibly conditional on X, denoted by $f_{Y|X}$ and $F_{Y|X}$, respectively.

	Name	Symbol	Definition
1	Variance-based	η_X	$\mathbb{E}[(\mathbb{E}[Y]-\mathbb{E}[Y\|X])^2]$
2	δ-importance	δ_X	$\frac{1}{2}\mathbb{E}[\int_{\mathbb{R}}\|f_Y(y)-f_{Y\|X}(y)\|dy]$
3	Kullback-Leibler	ε_X^{KL}	$\mathbb{E}[\int_{\mathbb{R}} f_{Y\|X}(y)\left(\log\frac{f_{Y\|X}(y)}{f_Y(y)}\right)dy]$
4	Kolmogorov-Smirnov	β_X^{KS}	$\mathbb{E}[\sup_{y\in\mathbb{R}}\{\|F_Y(y)-F_{Y\|X}(y)\|\}]$

machine learning studies (Lyons, 2013; Sejdinovic, Sriperumbudur, Gretton, & Fukumizu, 2013). We also provide necessary and sufficient conditions for a sensitivity measure to comply with Renyi (1959)'s Postulate D for measures of statistical dependence, showing that sensitivity measures based on strictly proper scoring rules achieve it for distribution reports. Finally, we illustrate the practical implications of the findings in the context of a complex decision, namely the planning and risk assessment of a lunar space mission.

2. Review and basic concepts

2.1. Related literature

This paper connects information value, sensitivity analysis, and forecasting with scoring rules. Each of these research streams is vast in itself and, moreover, investigations have proceeded on parallel tracks. We can then only propose a synthetic overview for positioning our work with respect to the extant literature, without any exhaustiveness claim. Information value has found widespread application in the literature — see Heath, Manolopoulou, and Baio (2017); Hilton (1981); Keisler, Collier, Chu, Sinatra, and Linkov (2014); Samson, Wirth, and Rickard (1989) for reviews, Mehrez (1985) for an application in project management, Ketzenberg (2009); Ketzenberg, Rosenzweig, Marucheck, and Metters (2007) in supply chain and Kao and Steuer (2016); Pfeifer, Schredelseker, and Seeber (2009) in portfolio selection.

Over the years, the notion of information value has been formalized in alternative ways. Raiffa and Schlaifer (1961) define information value as the increase in expected utility (EUI, henceforth) of the decision problem before and after receiving perfect information. Howard (1966) defines information value as the buying price for receiving perfect information. Hazen and Sounderpandian (1999) offer a thorough comparison of the main formulations and show that when the decision maker's utility function exhibits constant absolute risk aversion, the buying price and the EUI formulations are equivalent. The EUI definition has been employed in several studies. Bernardo and Smith (2000) demonstrate its use in Bayesian statistics. Avriel and Williams (1970) use this formulation to introduce information value in optimization problems. Pflug (2006) shows that information value formulated as EUI satisfies the axioms of coherent risk measures of Artzner, Delbaen, Eber, and Heath (1999). Gilboa and Lehrer (1991) identify the features that make a set function an information value function, moving from an EUI formulation of information value. Gilboa and Lehrer (1991)'s axiomatization provides the starting point for our discussion about whether a probabilistic sensitivity measure can be interpreted as information value. Relevant to our work is also the fact that the EUI formulation is used in the works that introduce information value as a probabilistic sensitivity measure (Felli & Hazen, 1998; 1999; Oakley, 2009; Strong et al., 2014).

In quantitative modeling, researchers have introduced several probabilistic sensitivity measures (see Borgonovo and Plischke (2016) for a review). Table 1 lists variance-based sensitivity measures (Saltelli & Tarantola, 2002; Wagner, 1995), distribution-based measures using the L^1-norm (Borgonovo, 2007), the Kullback-Leibler divergence between densities (Critchfield & Willard, 1986) and the Kolmogorov-Smirnov distance between cumulative distribution functions (Baucells & Borgonovo, 2013). Recent works study sensitivity measures based on the family of f-divergences (Rahman, 2016), on a transformation invariant version of the Cramér-von Mises distance (Gamboa et al., 2018), on copulas (Plischke & Borgonovo, 2019) and on quantiles (Browne et al., 2017). However, while probabilistic sensitivity measures have been thoroughly studied from a computational viewpoint, they remain much less understood from a decision-analytic viewpoint. In particular, their relationship to information value has not yet been established.

Finally, scoring rules have become over the years an essential tool for guiding and assessing forecasts. Gneiting and Raftery (2007), Jose, Nau, and Winkler (2008), and more recently, Gneiting and Katzfuss (2014) provide comprehensive reviews on this topic. Scoring rules can provide the link between information value and probabilistic sensitivity measures. For instance, Bernardo and Smith (2000) show that, under a quadratic or a logarithmic score, information value coincides with variance-reduction and with the Kullback-Leibler information, respectively. We discuss this connection in detail in Section 3.

2.2. Basic definitions and relevant properties

To support the decision making process at hand, the analyst creates or relies on a model that predicts the uncertain quantity Y of interest, whose value may be dependent on a set of uncertain exogenous variables $X_1, X_2, \ldots, X_n$. We let $Y, X_1, X_2, \ldots, X_n$ be real valued random variables on the measure space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is a Borel σ-algebra and P a reference probability measure.

Now, to introduce probabilistic sensitivity measures we need two ingredients: First we need the input and output probability measures and, second, a separation measure between probability distributions. Regarding the probability measures, we have the following: Throughout this work, we discuss the case in which the analyst has the possibility of gathering information $X = g(X_1, \ldots, X_n)$ about one or more of the inputs $X_1, \ldots, X_n$. Here g is any function $g : \mathbb{R}^n \to \mathbb{R}^k$ and k can be any integer $1 \le k \le n$. For instance, the analyst might gather the information $X = X_2$, the information $X = (X_1, X_2)$, the information $X = \sum_{i=1}^n X_i$, or complete information $X = (X_1, \ldots, X_n)$. We then let $\mathbb{P}_Y$ and $\mathbb{P}_X$ denote the probability measures of Y and X respectively, $F_Y(y)$ and $F_X(x)$ denote marginal cumulative distribution functions of Y and X, respectively, and $f_Y(y)$ and $f_X(x)$ the corresponding densities. The expected value of Y is denoted by μ_Y or $\mathbb{E}[Y]$. Let $\mathbb{P}_{Y|X}$ be the conditional probability measure of Y given X. $\mathbb{P}_{Y|X}$ is a random probability measure over $\mathbb{R}$ that depends on X. In the remainder, $f_{Y|X}$, $F_{Y|X}$, and $\mu_{Y|X}$ are respectively the density, the cumulative distribution function and the expected value corresponding to the conditional probability measure $\mathbb{P}_{Y|X}$.

Regarding the separation measure, let $\mathcal{P}$ be the set of all probability measures on $(\Omega, \mathcal{F})$. Consider a function $\zeta : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$. We call $\zeta(\cdot,\ \cdot)$ a separation measure if, for any pair of probability measures $(\mathbb{P}, \mathbb{Q}) \in \mathcal{P} \times \mathcal{P}$, it satisfies the conditions: a) $\zeta(\mathbb{P}, \mathbb{Q}) \ge 0$ and b) $\zeta(\mathbb{P}, \mathbb{Q}) = 0$ if $\mathbb{P} = \mathbb{Q}$. Given this setup, we now come to the definition of probabilistic sensitivity measure.

Definition 2 (Probabilistic Sensitivity Measure)**.** We call the quantity

$$\xi_X^Y = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})] \tag{1}$$

the probabilistic sensitivity measure of Y with respect to X based on the separation measure $\zeta(\cdot)$.

Definition 2 forms the so-called common rationale (Borgonovo et al., 2016) that encompasses several probabilistic sensitivity measures defined in previous works. (Because we will always be considering the importance of several different X with respect to the same Y, we shall drop the superscript Y in the remainder, for notational convenience.)

Example 3. Consider Table 1. The variance-based sensitivity measure of X (Homma & Saltelli, 1996; Iman & Hora, 1990) is given by

$$\eta_X = \mathbb{E}\Big[(\mathbb{E}[Y] - \mathbb{E}[Y|X])^2\Big]$$

based on

$$\zeta^{\text{Quad}}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = (\mathbb{E}[Y] - \mathbb{E}[Y|X])^2. \quad (2)$$

The $\delta-$importance measure (line 2 in Table 1) is given by

$$\delta_X = \frac{1}{2}\mathbb{E}\Big[\int_{\mathbb{R}} \big|f_Y(y) - f_{Y|X}(y)\big| dy\Big]$$

based on

$$\zeta^{L_1}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \frac{1}{2}\int_{\mathbb{R}} \big|f_Y(y) - f_{Y|X}(y)\big| dy, \quad (3)$$

and is derived from the L_1-norm between the marginal and conditional densities of Y (Borgonovo, 2007). The $\varepsilon_X^{\text{KL}}$ importance measure (line 3 in Table 1) is given by

$$\varepsilon_X^{\text{KL}} = \mathbb{E}\left[\int_{\mathbb{R}} f_{Y|X}(y) \log\left(\frac{f_{Y|X}(y)}{f_Y(y)}\right) dy\right]$$

based on

$$\zeta^{\text{KL}}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \int_{\mathbb{R}} f_{Y|X}(y) \log\left(\frac{f_{Y|X}(y)}{f_Y(y)}\right) dy. \quad (4)$$

This separation measure is obtained from the Kullback-Leibler divergence between the marginal and conditional densities of Y (Critchfield & Willard, 1986). Finally, the sensitivity measure in line 4 of Table 1,

$$\beta_X^{\text{KS}} = \mathbb{E}\left[\sup_{y\in\mathbb{R}} \big|F_Y(y) - F_{Y|X}(y)\big|\right]$$

based on

$$\zeta^{\text{KS}}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \sup_{y\in\mathbb{R}} \big|F_Y(y) - F_{Y|X}(y)\big| \quad (5)$$

uses the Kolmogorov-Smirnov distance between the marginal and conditional cumulative distribution functions of Y (Baucells & Borgonovo, 2013).

Definition 4. A probabilistic sensitivity measure ξ_X possesses the *nullity-implies-independence* property if $\xi_X = 0$ implies that X and Y are independent.

This property has a relevant decision-analytic implication. Only when nullity-implies-independence holds is the analyst reassured that a null value of the sensitivity measure guarantees that Y is independent of X. When nullity-implies independence fails, an indication of zero sensitivity can be misleading, as potentially X could still influence Y. Historically, this property characterizes measures of of dependence since the seminal work of Renyi (1959), where it is stated as Postulate D.

Moreover, monotonic transformation invariance has emerged as a convenient property in estimation. Transformations are used by analysts to accelerate numerical convergence and are particularly relevant when the simulator output is sparse. If a probabilistic sensitivity measure is transformation invariant, results on the transformed and on the original scales coincide. This allows analysts to fully exploit the accelerated numerical convergence, while avoiding the problem of interpreting results back on the original scale — see (Baucells & Borgonovo, 2013) among others for greater details.

Scoring rules are introduced in the following context. The analyst wishes to provide a forecast report $a \in \mathcal{A}$ of some characteristic θ_Y of an uncertain quantity Y of interest. The set $\mathcal{A}$ of possible values of θ_Y is dictated by the overall decision context, and could be a subset of the reals (if θ_Y is the mean, median or quantile) or of d-dimensional Euclidean space (if θ_Y is a d-dimensional vector of quantiles) or a space of probability measures (if θ_Y is a density or a distribution function).

A scoring rule $S : \mathbb{R} \times \mathcal{A} \to \mathbb{R}$ evaluates the quality of the report by assigning a real-valued score $S(y, a)$ to the combination of the report a and the outcome $Y = y$. We consider the collection of all scoring rules S whose conditional expectation $\mathbb{E}[S(Y, a)|X]$ and maximum conditional expectation $\max_{a\in\mathcal{A}} \mathbb{E}[S(Y, a)|X]$ exist for all information vectors $X = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})$. A scoring rule S is proper if $\mathbb{E}[S(Y, a)]$ and $\max_{a\in\mathcal{A}} \mathbb{E}[S(Y, a)]$ exist for every random variable Y on $\mathbb{R}$, and one of the maximizers $a \in \mathcal{A}$ of $\mathbb{E}[S(Y, a)]$ is $a = \theta_Y$. It is strictly proper if θ_Y is the unique maximizer.

To illustrate, two well-known examples of proper scoring rules are the quadratic (Brier, 1950) and the logarithmic scores (Kullback & Leibler, 1951). Suppose that the analyst is interested in a point estimate; then $\mathcal{A} \subset \mathbb{R}$ and $a \in \mathbb{R}$. The quadratic score for report a is

$$S^{\text{Quad}}(y, a) = -(y - a)^2. \quad (6)$$

The value of a maximizing the expected score $-\mathbb{E}[(Y - a)^2]$ is the mean $a = \mu_Y$. Suppose, instead, that the forecast of interest is the density of an absolutely continuous Y. In this case, $\mathcal{A}$ is a set of suitable densities $a : \mathbb{R} \to \mathbb{R}$ and the logarithmic score for report a is

$$S^{\log}(y, a) = \log(a(y)). \quad (7)$$

The density a maximizing the expected score $\mathbb{E}[\log(a(Y))]$ can be shown to be the density $a = f_Y$ of Y.

3. Probabilistic sensitivity measures and information value

This section is divided into three parts. After showing that information value under a scoring rule S is always a probabilistic sensitivity measure in the form of Definition 2, we derive the conditions under which there exists a reporting problem for which a probabilistic sensitivity measure can be interpreted as information value. We also examine the conditions that are necessary for a sensitivity measure to be information value under a given scoring rule.

3.1. Information value for scoring rules

Given a proper scoring rule S, we define information value in the standard way as

$$\epsilon_X^S = \mathbb{E}\left[\max_{a\in\mathcal{A}} \mathbb{E}[S(Y, a)|X]\right] - \max_{a\in\mathcal{A}} \mathbb{E}[S(Y, a)] \quad (8)$$

representing the greatest score improvement one would achieve by learning X. In order for expression (8) to be well defined, the argument of the outer expectation in (8) must be measurable. This will hold if the function $x \mapsto \mathbb{E}[S(Y, a)|X = x]$ is continuous in x for all $a \in \mathcal{A}$ — see Appendix B for details.[1] If Y has probability measure $\mathbb{P}$, define $S(\mathbb{P}, a) = \mathbb{E}[S(Y, a)]$. It follows that if $\mathbb{Q}$ is the probability measure of Y given X, then $\mathbb{E}[S(Y, a)|X] = S(\mathbb{Q}, a)$. Let $a_{\mathbb{P}}^S$ denote the set of reports a that maximizes the expected score[2] $\mathbb{E}[S(Y, a)]$. Letting $\mathbb{P}$ be the probability measure of Y and $\mathbb{Q}$ the (uncertain)

[1] To the authors' knowledge, the conditions under which the argument of the outer expectation in (8) is measurable have not been studied before. These conditions ensure that information value is well posed.

[2] We write the maximum expected score as $\mathbb{E}\big[S(Y, a_{\mathbb{P}}^S)\big]$ even though $a_{\mathbb{P}}^S$ may be a *collection* of reports. $\mathbb{E}\big[S(Y, a_{\mathbb{P}}^S)\big]$ is defined to be the common value $\mathbb{E}[S(Y, a)]$ for $a \in a_{\mathbb{P}}^S$.

conditional probability measure of Y given X, we have

$$\begin{aligned}\epsilon_X^S &= \mathbb{E}\left[\max_{a\in\mathcal{A}}\mathbb{E}[S(Y,a)|X]\right] - \mathbb{E}\left[S(Y,a_{\mathbb{P}}^S)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[S(Y,a_{\mathbb{Q}}^S)|X\right]\right] - \mathbb{E}\left[\mathbb{E}\left[S(Y,a_{\mathbb{P}}^S)|X\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[S(Y,a_{\mathbb{Q}}^S) - S(Y,a_{\mathbb{P}}^S)|X\right]\right] = \mathbb{E}\left[S(\mathbb{Q},a_{\mathbb{Q}}^S) - S(\mathbb{Q},a_{\mathbb{P}}^S)\right]. \quad (9)\end{aligned}$$

The quantity inside the last expectation is nonnegative, and because S is proper, it equals zero when $\mathbb{P}=\mathbb{Q}$. It is therefore a separation measure by Definition 2. This demonstrates the following result.

Proposition 5. *Information value ϵ_X^S under a proper scoring rule S is a probabilistic sensitivity measure (1) with separation measure*

$$\zeta^S(\mathbb{P},\mathbb{Q}) = S(\mathbb{Q},a_{\mathbb{Q}}^S) - S(\mathbb{Q},a_{\mathbb{P}}^S). \quad (10)$$

It follows that ζ^S is the so-called *divergence* function for scoring rule S (Dawid, 2007; Gneiting & Raftery, 2007).

Proposition 5 implies that a sensitivity measure not constructed in the form of (1) cannot be information value in a reporting problem under any scoring rule. The literature has proposed sensitivity measures resulting from the machine learning notion of contrast (Fort, Klein, & Rachdi, 2016), and from the so-called PAWN method (Pianosi & Wagener, 2015). Then, the sensitivity measures in Fort et al. (2016) are information value, because a contrast is the machine learning analog of a scoring rule. Conversely, the PAWN method substitutes the expectation in (1) with a generic operator — i.e., the max or the median, defining the sensitivity measure such as $\xi_X^{\max} = \max_x \zeta(\mathbb{P}_Y,\mathbb{P}_{Y|X=x})$ or $\xi_X^{\text{median}} = \text{median}\ \zeta(\mathbb{P}_Y,\mathbb{P}_{Y|X})$. Then, a PAWN-based sensitivity measure does not share an information value interpretation unless the operator is the mean.

It is easy to see that, for Y and X independent, any probabilistic sensitivity measure ξ_X including ϵ_X^S, is null. However, as it is well known, the converse is not true, i.e., a null information value does not guarantee that Y and X are independent. For this, however, the following holds (see Appendix B for all proofs) for a strictly proper scoring rule.

Proposition 6. *Consider a strictly proper scoring rule $S(y, a)$ in which the report a is a probability density or cumulative distribution function, and let ϵ_X^S be the information value of X under scoring rule S. Then $\epsilon_X^S = 0$ if and only if Y and X are independent.*

Therefore, if the analyst is reporting the entire distribution and anticipating evaluation under a strictly proper scoring rule S, then she is reassured that a null information value under S implies that Y is independent of X. Thus, one way to guarantee the equivalence of nullity and independence under a strictly proper scoring rule is to report the entire distribution. The nullity property that exists when Y and X are independent directly stems from the convexity of the expected scoring function of strictly proper scoring rules. This geometric property has been studied by Dawid (2007) and described later in the information theory literature (Ebrahimi, Soofi, & Soyer, 2010; Shoja & Soofi, 2017).

3.2. Probabilistic sensitivity measures as information value

In the previous section, we studied information value ϵ_X^S under scoring rule S and showed in Proposition 5 that it qualifies as a sensitivity measure under Definition 2. On the other hand, in Section 2.2, we gave several examples of sensitivity measures ξ_X that arise in other ways. In the current section and the next, we consider an arbitrary sensitivity measure ξ_X and ask whether it could be information value.

The difficulty of addressing this question is emphasized by the following fact: There is a broad class of sensitivity measures ξ_X which are in fact information value under in a generalized reporting problem that we are to define. We illustrate how this may occur, prove this result below and in Appendix B, and follow with a discussion.

The setting of the prior section can be represented as reporting problems (one for each event B)

$$\text{maximize}_{a\in\mathcal{A}}\mathbb{E}[S(Y,a)|B] \quad (11)$$

where a is some report about the distribution of Y. To obtain the information value of an uncertain quantity X under scoring rule S, one must consider both the null information state $B=\Omega$, and the full information states $B=\{X=x\}$ for each possible value of X. The required technical assumptions are that there is an underlying σ-algebra $\mathcal{B}$ over a probability space Ω with respect to which X is measurable, and that B is a non-null member of $\mathcal{B}$. For instance, if S is the quadratic scoring rule (6), then $\mathcal{A}$ is the set of reals and $\mathcal{B}$ could be the σ-algebra generated by X.

The new reporting problem we have in mind is one in which (i) the report a concerning Y is a full distribution $\mathbb{P}$; (ii) instead of a scoring rule $S(Y, a)$, we allow a general real-valued $\mathcal{B}$-measurable scoring function $U(\omega,\mathbb{P})$ for $\omega\in\Omega$; and (iii) the set $\mathcal{A}$ of possible reports about Y is replaced by the set of conditional distributions $\mathcal{P}_{\mathcal{B}} = \left\{\mathbb{P}_{Y|A}|A\in\mathcal{B}, A\neq\emptyset\right\}$, where $\mathbb{P}_{Y|A}(dy) = \mathbb{P}(Y\in dy|A)$. In other words, we consider the family of reporting problems depending on $B\in\mathcal{B}$

$$\text{maximize}_{\mathbb{P}\in\mathcal{P}_{\mathcal{B}}}\mathbb{E}[U(\mathbb{P})|B] \quad (12)$$

where, as is conventional notation, $U(\mathbb{P})$ is the random variable that takes on value $U(\omega,\mathbb{P})$ should outcome $\omega\in\Omega$ occur.

We give the following simplified example to illustrate this setup.

Example 7. Suppose that the probability of successful intervention is an uncertain value Y uniformly distributed on the interval [0,1]. A trial with sample size $n=3$ interventions is conducted to estimate what Y might be. Let X be the number of successful interventions in the sample. Then the distribution of X given Y is Binomial$(n=3, p=Y)$, and the marginal distribution of X is $P_X(k)=\frac{1}{4}$, for $k=0,1,2,3$. Standard conjugate Bayesian results indicate that the distribution of Y given X is Beta with shape parameters $\alpha=1+X$ and $\beta=1+n-X=4-X$. Suppose we use the KS sensitivity measure (5) as a measure of the sensitivity of Y to X. We calculate numerically

$$\begin{aligned}\beta_X^{KS} &= \sum_{k=0}^{3}\mathbb{P}_X(k)\sup_{y\in\mathbb{R}}\left|F_{Y|X=k}(y) - F_Y(y)\right| \\ &= \sum_{k=0}^{3}\frac{1}{4}\sup_{y\in[0,1]}\left|F_{Y|X=k}(y) - F_Y(y)\right| \\ &= \frac{1}{4}(0.472+0.224+0.224+0.427) = 0.348.\end{aligned}$$

Consider another random variable W equal to the indicator (1 or 0) of the event $\{X\geq 2\}$. What is the KS-sensitivity β_W^{KS} of Y to W? We use Bayes' rule to calculate the densities of Y given $W=k$ as shown in Fig. 2, and derive the cumulative distributions of Y given $W=k$. From there, we obtain numerically

$$\begin{aligned}\beta_W^{KS} &= \sum_{k=0}^{1}\mathbb{P}_W(k)\sup_{y\in\mathbb{R}}\left|F_{Y|W=k}(y) - F_Y(y)\right| = \frac{1}{2}(0.312+0.312) \\ &= 0.312.\end{aligned}$$

Not surprisingly, sensitivity to W is less than sensitivity to X.

Could these sensitivity values for X and W be information values under the reporting problem (12) for some utility function $U(\omega,\mathbb{P})$? In that setup, let $\mathcal{B}$ be the σ-algebra generated by X. Then $\mathcal{B}=\{B_K|K\subset\{0,1,2,3\}\}$ where B_K is the event $\{X\subset K\}$. This

Table 2
Utility function U and conditional probabilities for X of Example 7.

B	$P(B_0\|B)$	$P(B_1\|B)$	$P(B_2\|B)$	$P(B_3\|B)$	$U(\omega, \mathbb{P}_{Y\|B})$			
					$\omega \in B_0$	$\omega \in B_1$	$\omega \in B_2$	$\omega \in B_3$
B_0	1	0	0	0	0.945	0.503	0.250	0.193
B_1	0	1	0	0	0.280	0.696	0.441	0.473
B_2	0	0	1	0	0.280	0.441	0.696	0.473
B_3	0	0	0	1	0.193	0.250	0.503	0.945
B_{01}	1/2	1/2	0	0	0.945	0.625	0.249	0.071
B_{02}	1/2	0	1/2	0	0.691	0.249	0.504	0.446
B_{03}	1/2	0	0	1/2	0.665	0.280	0.473	0.473
B_{12}	0	1/2	1/2	0	0.280	0.665	0.473	0.473
B_{13}	0	1/2	0	1/2	0.254	0.696	0.441	0.499
B_{23}	0	0	1/2	1/2	0.000	0.320	0.696	0.874
B_{012}	1/3	1/3	1/3	0	0.753	0.695	0.443	0.000
B_{013}	1/3	1/3	0	1/3	0.665	0.504	0.249	0.473
B_{023}	1/3	0	1/3	1/3	0.665	0.249	0.504	0.473
B_{123}	0	1/3	1/3	1/3	0.000	0.443	0.695	0.753
B_{0123}	1/4	1/4	1/4	1/4	0.473	0.473	0.473	0.473

Table 3
Calculations for information X in Example 7. Each entry is a value of $\mathbb{E}[U(\mathbb{P}_{Y|A})|B]$.

B / A	B_0	B_1	B_2	B_3	B_{01}	B_{02}	B_{03}	B_{12}	B_{13}	B_{23}	B_{012}	B_{013}	B_{023}	B_{123}	B_{0123}	Max$_A$
B_0	0.945	0.280	0.280	0.193	0.945	0.691	0.665	0.280	0.254	-	0.753	0.665	0.665	-	0.473	0.945
B_1	0.503	0.696	0.441	0.250	0.625	0.249	0.280	0.665	0.696	0.320	0.695	0.504	0.249	0.443	0.473	0.696
B_2	0.250	0.441	0.696	0.503	0.249	0.504	0.473	0.473	0.441	0.696	0.443	0.249	0.504	0.695	0.473	0.696
B_3	0.193	0.473	0.473	0.945	0.071	0.446	0.473	0.473	0.499	0.874	-	0.473	0.473	0.753	0.473	0.945
B_{0123}	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473

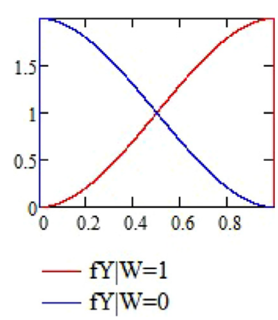

Fig. 2. Posterior densities $f_{Y|W=k}(y)$ of Y given $W = k$ for $k = 0, 1$.

σ-algebra contains 16 sets and has as a basis the four sets B_0, B_1, B_2, B_3 (where we reduce clutter by omitting set braces in subscripts, i.e., 0 means {0}, 1 means {1}, 0123 means {0, 1, 2, 3}, and so on). Notice that B_{0123} is the entire sample space Ω. The set $\mathcal{P}_\mathcal{B}$ is then the set of posterior distributions $\mathbb{P}_{Y|B_K}$ for $\emptyset \neq K \subset \{0, 1, 2, 3\}$, and the reporting problem (12) given $B = B_{K'}$ is to choose the "best" posterior distribution $\mathbb{P}_{Y|B_K}$, that is, the $\mathbb{P}_{Y|B_K}$ maximizing $\mathbb{E}\big[U(\mathbb{P}_{Y|B_K})|B_{K'}\big]$. In calculating the value of information X, we would maximize, for each possible value k of X, the expected utility $\mathbb{E}\big[U(\mathbb{P}_{Y|B_K})|B_k\big]$, and average the results using the marginal $\mathbb{P}_X(k)$ of X. The result would be the expected utility *with* information X. We would compare this with the expected utility *without* information, the maximum over $\mathbb{P}_{Y|B_K}$, of the expected utility $\mathbb{E}\big[U(\mathbb{P}_{Y|B_K})\big] = \mathbb{E}\big[U(\mathbb{P}_{Y|B_K})|\Omega\big] = \mathbb{E}\big[U(\mathbb{P}_{Y|B_K})|B_{0123}\big]$. The net would be the expected utility *of* information X. We carry this procedure out for the utility function $U(\omega, \mathbb{P})$ given by Table 2 where we also list on the left the conditional probabilities we will need.

The needed expected utilities

$$\mathbb{E}\Big[U\Big(\mathbb{P}_{Y|A}\Big)|B\Big] = \sum_{i=0}^{3} P(B_i|B)U\Big(\omega \in B_i, \mathbb{P}_{Y|B}\Big)$$

are given for all relevant A, B in Table 3.

Each row of this table corresponds to an instance of reporting problem (12), and the resulting maxima are listed to the right. We have, for example, $\max_A \mathbb{E}\big[U(\mathbb{P}_{Y|A})|B_0\big] = 0.945$, and the expected utility *without* information is $\max_A \mathbb{E}\big[U(\mathbb{P}_{Y|A})|B_{0123}\big] = 0.473$. The expected utility *with* information X is $\frac{1}{4} \cdot 0.945 + \frac{1}{4} \cdot 0.696 + \frac{1}{4} \cdot 0.696 + \frac{1}{4} \cdot 0.945 = 0.821$. The expected utility *of* information X is $0.821 - 0.473 = 0.348$, exactly equal to the KS-sensitivity β_X^{KS}.

This is no coincidence, and happens not just with X, but with any $\mathcal{B}$-measurable random variable W in this problem. For instance, with W being the indicator of $\{X \geq 2\}$ mentioned above, the event $\{W = 1\}$ is B_{23}, and $\{W = 0\}$ is B_{01}, both events having probability $\frac{1}{2}$. The relevant calculations are in Table 4. The expected utility *with* information W is $\frac{1}{2} \cdot 0.785 + \frac{1}{2} \cdot 0.785 = 0.785$, and the expected utility *of* information W is $0.785 - 0.473 = 0.312$, exactly equal to the KS-sensitivity β_W^{KS}.

As another example, the variable $R = 3(X - 1)^+$, that gives a \$3 reward for every success in excess of 1, is $\mathcal{B}$-measurable, and its possible values {6, 3, 0} correspond to the events B_3, B_2, B_{01} with respective probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$. What is its information value? The relevant calculations are found in Table 5. The expected utility *with* information R is $\frac{1}{4} \cdot 0.696 + \frac{1}{4} \cdot 0.945 + \frac{1}{2} \cdot 0.785 = 0.803$ and the expected utility *of* information R is $0.803 - 0.473 = 0.330$. One can check as above that this is equal to β_R^{KS}.

In summary, this example makes it plausible that, for any $\mathcal{B}$-measurable X (not just the X above), the quantity β_X^{KS} will equal the information value of X under the utility function $U(\omega, \mathbb{P})$ above. We haven't checked this for every $\mathcal{B}$-measurable X, but can be sure it is true because we have explicitly identified a utility function U guaranteed to exist by our Theorem 8 below. By that theorem, the β_X^{KS} sensitivity measure will act this way not just in the Beta-binomial probability space of this example, but in all probability spaces $(\Omega, \mathcal{B}, P)$. However, the needed utility function will to our knowledge have no closed form, and will be different for different probability spaces.

This leads to the more general question: Can *any* sensitivity measure ξ_X be expressed as information value in the reporting

Table 4
Calculations for information W in Example 7. Each entry is a value of $\mathbb{E}[U(\mathbb{P}_{Y|A})|B]$.

B / A	B_0	B_1	B_2	B_3	B_{01}	B_{02}	B_{03}	B_{12}	B_{13}	B_{23}	B_{012}	B_{013}	B_{023}	B_{123}	B_{0123}	Max_A
B_{01}	0.724	0.488	0.361	0.221	0.785	0.470	0.473	0.473	0.475	0.160	0.724	0.584	0.457	0.221	0.473	0.785
B_{23}	0.598	0.361	0.488	0.348	0.597	0.598	0.569	0.376	0.348	0.348	0.598	0.457	0.584	0.348	0.473	0.598
B_{0123}	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473

Table 5
Calculations for information R in Example 7. Each entry in the Table is a value of $\mathbb{E}[U(\mathbb{P}_{Y|A})|B]$.

B / A	B_0	B_1	B_2	B_3	B_{01}	B_{02}	B_{03}	B_{12}	B_{13}	B_{23}	B_{012}	B_{013}	B_{023}	B_{123}	B_{0123}	Max_A
B_2	0.250	0.441	0.696	0.503	0.249	0.504	0.473	0.473	0.441	0.696	0.443	0.249	0.504	0.695	0.473	0.696
B_3	0.193	0.473	0.473	0.945	0.071	0.446	0.473	0.473	0.499	0.874	0	0.473	0.473	0.753	0.473	0.945
B_{01}	0.724	0.488	0.361	0.221	0.785	0.470	0.473	0.473	0.475	0.160	0.724	0.584	0.457	0.221	0.473	0.785
B_{0123}	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473	0.473

problem (12) for some utility function $U(\omega, \mathbb{P})$? The answer is not necessarily, as we now explain.

In direct analogy to the notion of proper scoring rule, say that function U is proper if an optimal solution to (12) is $\mathbb{P} = \mathbb{P}_{Y|B}$, that is, for all non-null $A, B \in \mathcal{B}$,

$$\mathbb{E}\big[U(\mathbb{P}_{Y|A})|B\big] \leq \mathbb{E}\big[U(\mathbb{P}_{Y|B})|B\big].$$

The utility function of the example above is proper, as the reader may partially check in the tables of relevant calculations given there.

Let $\zeta_B = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|B})$, and say that U is *consistent* with ζ if there is a constant u_0 such that for all $B \in \mathcal{B}$

$$\mathbb{E}[U(\mathbb{P}_{Y|B})|B] = \zeta_B + u_0. \tag{13}$$

Whether there exists such a scoring function U consistent with ζ is a question that we treat shortly. But if such U does exist, then it follows that the information value of X under U must be $\xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})]$. This can be seen as follows[3]: Consider first the reporting problem (12) given $B = \Omega$. Because U is proper, the optimal report is $\mathbb{P} = \mathbb{P}_{Y|\Omega} = \mathbb{P}_Y$, and because $\zeta_\Omega = \zeta(\mathbb{P}_Y, \mathbb{P}_Y) = 0$, and U is consistent with ξ, we have $\mathbb{E}[U(\mathbb{P}_Y)] = u_0$. Second, consider the reporting problem (12) given $B = \{X = x\} \in \mathcal{B} \setminus \emptyset$. Because U is proper, the optimal report is $\mathbb{P}_{Y|X=x}$ with $\mathbb{E}[U(\mathbb{P}_{Y|X=x})|X = x] = \zeta_{\{X=x\}} + u_0 = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X=x}) + u_0$. The expected value of the reporting problem in anticipation of learning X is therefore $\mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})] + u_0 = \xi_X + u_0$. Information value is the excess of this over the unconditional value of the reporting problem, u_0, which excess is ξ_X. Therefore we have shown that if there exists a scoring function U that is proper and consistent with ζ, then the information value under U of any $\mathcal{B}$-measurable X is ξ_X.

Concerning the existence of such scoring functions U, we have the following result.

Theorem 8. *Consider the case in which the σ-algebra $\mathcal{B}$ contains finitely many sets B. Suppose the separation measure $\zeta(\mathbb{P}, \mathbb{Q})$ of ξ is convex in $\mathbb{Q}$ for each $\mathbb{P}$. Then*

1. *there is a proper scoring function U consistent with ζ, and consequently*
2. *for any such U, and any $\mathcal{B}$-measurable X, the value of information X in the family (12) of reporting problems is equal to the sensitivity measure $\xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})]$.*

We suspect that the restriction that the σ-algebra $\mathcal{B}$ contains finitely many sets B is not crucial to this result, because the infinite-$\mathcal{B}$ case is in some sense the limit of the finite-$\mathcal{B}$ case. Unfortunately, at this time we do not have a proof for $\mathcal{B}$ infinite[4].

[3] See Gilboa and Lehrer (1991) for the origin of this argument.

[4] Note because $\mathcal{B}$ may be only a sub-algebra of the full σ-algebra defining the probability space Ω, the finite $\mathcal{B}$ restriction does *not* restrict X_1, ..., X_n, Y to be discrete random variables.

The separation measures of the sensitivity measures η_X, $\varepsilon_X^{\mathrm{KL}}$, δ_X and β_X^{KS} (Table 1) are all convex in their second argument $\mathbb{Q}$, as we show in Appendix B.

Corollary 9. *The sensitivity measures η_X, $\varepsilon_X^{\mathrm{KL}}$, δ_X and β_X^{KS} are information value under the corresponding reporting problem (12).*

As we show in Section 4, the sensitivity measure η_X is also information value in problem (11) under a quadratic scoring rule. Moreover, $\varepsilon_X^{\mathrm{KL}}$ is information value under a log scoring rule, as we note in Section 5.1.

Theorem 8 unfortunately guarantees only the existence of a reporting problem consistent with ξ, and the scoring function U obtained in the proof of Theorem 8 has no apparent closed form. Moreover, U *depends on the model*, that is, on the joint distribution of $Y, X_1, \ldots, X_n$, through the quantities $\zeta_B = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|B})$ that appear as inputs into the consistency conditions (13). In contrast, a scoring rule S is model independent. Theorem 8 then provides reassurance that measures such as δ_X and β_X^{KS}, that are desirable for computational or other properties, do in fact behave as information value if not under a proper scoring rule then at least under some proper scoring function.

3.3. Relationship between scoring rules and separation measures

Previous literature has shown that a probabilistic sensitivity measure is uniquely determined by its separation measure. However the converse is not true. If any linear real-valued function of $\mathbb{P} - \mathbb{Q}$ is added to an separation measure $\zeta(\mathbb{P}, \mathbb{Q})$, the resulting separation measure generates the same sensitivity measure. For example the separation measure

$$\zeta'(\mathbb{P}, \mathbb{Q}) = (\mu_{\mathbb{P}} - \mu_{\mathbb{Q}} + k)^2 - k^2, \tag{14}$$

where k is a real number, differs from the separation measure for η_X in Table 1 by a multiple of $(\mu_{\mathbb{P}} - \mu_{\mathbb{Q}})$, and also generates the probabilistic sensitivity measure η_X. And, letting $t^+ = \frac{1}{2}(t + |t|)$ denote the positive part of $t \in \mathbb{R}$, the separation measure

$$\zeta^+(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}} (f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y))^+ dy, \tag{15}$$

differs from the separation measure of δ_X in Table 1 by half the integral of $f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y)$, but it still yields δ_X as sensitivity measure — see Appendix B for details.

Suppose sensitivity measure ξ_X has separation measure ζ and is always equal to the information value of X under some scoring rule $S(y, a)$, that is $\xi_X = \epsilon_X^S$ for all X. Because a sensitivity measure does not determine its separation measure, it does not follow that $\zeta = \zeta^S$ uniquely. If not, then what kind of relationship must exist between ζ and S? The following result provides a partial answer.

Say that a separation measure ζ *distinguishes between* two probability measures $\mathbb{P}$, $\mathbb{Q}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ if either $\zeta(\mathbb{P}, \mathbb{Q}) > 0$ or

$\zeta(\mathbb{Q},\mathbb{P}) > 0$. Analogously, say that a scoring rule $S(y, a)$ distinguishes between $\mathbb{P}$, $\mathbb{Q}$ if optimal report $a_{\mathbb{P}}$ when Y has distribution $\mathbb{P}$ differs from optimal report $a_{\mathbb{Q}}$ when Y has distribution $\mathbb{Q}$. Say that a sensitivity measure ξ_X is information value under scoring rule S if, for all X, the information value of X under S is equal to ξ_X. Then we have the following result.

Proposition 10. *Let $\mathbb{Q}_1$ and $\mathbb{Q}_0$ be probability measures on $(\Omega,\mathcal{F})$. Suppose the map $\alpha \mapsto \zeta((1-\alpha)\mathbb{Q}_0+\alpha\mathbb{Q}_1,\mathbb{Q}_0)$ is a continuous function over $\alpha\in[0,\ 1]$ for all measures $\mathbb{Q}_1,\mathbb{Q}_0$. A necessary condition for the sensitivity measure ξ_X with separation measure ζ to be information value under scoring rule $S(y, a)$ is that for all $\mathbb{P}$, $\mathbb{Q}$ on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, the scoring rule S distinguishes between $\mathbb{P}$, $\mathbb{Q}$ whenever the separation measure ζ does.*

As a simple example consider the sensitivity measures δ_X and β_X^{KS}. They fall under Proposition 10 because their separation measures satisfy its continuity requirement. If their separation measures ζ distinguish between $\mathbb{P}$ and $\mathbb{Q}$, then $F_{\mathbb{P}}\neq F_{\mathbb{Q}}$. But note, for example, that the quadratic scoring rule in (6) has optimal report a^* equal to the mean of Y (see Section 4). Therefore, S^{Quad} does not distinguish two distributions with the same mean, even through ζ does. It follows that δ_X and β_X^{KS} cannot be information value under S^{Quad}, or for that matter, under any scoring rule that reports only summary distribution statistics. What is needed is a scoring rule whose optimal report is a distribution (or from which a distribution can be derived).

A related argument shows that the condition of Proposition 10 cannot be sufficient. Consider any strictly proper scoring rule S whose optimal report is a distribution (e.g. the CRPS score we discuss in Section 5.2). Because it is strictly proper, S must distinguish any two distinct distributions. Therefore S distinguishes any two distributions that ζ does, and this for arbitrary separation measures ζ. If this condition were sufficient, it would follow that an arbitrary sensitivity measure ξ_X is information value under S, a clear impossibility because different ζ can give non-equivalent values of ξ_X.

4. Sensitivity measures and point estimate reports

4.1. Bregman scores: reporting mean values

A well-known result in the scoring rules literature states that under weak regularity conditions, a scoring function that induces the truthful reporting of a distribution Y's mean μ_Y must be a Bregman function (Savage, 1971 and Theorem 7 of Gneiting, 2011). We say that a scoring rule S^B over numerical reports $a\in\mathbb{R}$ is a *Bregman function* if there exists a twice-differentiable strictly convex function $\psi(\cdot)$ over $\mathbb{R}$ such that

$$S^B(y,a)=\psi(a)+\psi'(a)(y-a)-\psi(y). \tag{16}$$

The score S^B being strictly proper implies that $\mathbb{E}[S^B(Y,a)]$ is maximized when $a=\mu_Y$.

Proposition 11. *The information value of X under a Bregman scoring rule is given by*

$$\epsilon_X^B=\mathbb{E}[\psi(\mu_{Y|X})-\psi(\mu_Y)], \tag{17}$$

with inner statistic

$$\zeta^B(\mathbb{P}_Y,\mathbb{P}_{Y|X})=\psi(\mu_{Y|X})-(\psi(\mu_Y)+\psi'(\mu_Y)(\mu_{Y|X}-\mu_Y)). \tag{18}$$

By strict convexity of ψ, we have $\epsilon_X^B>\psi(\mathbb{E}\mu_{Y|X})-\psi(\mu_Y)=\psi(\mu_Y)-\psi(\mu_Y)=0$ as long as $\mu_{Y|X}\neq\mu_Y$ for some values of X. The only way ϵ_X^B can be zero is when $\mu_{Y|X}=\mu_Y$ for all values of X. Thus, ϵ_X^B does not possess the nullity-implies-independence property, because nullity implies only $\mu_{Y|X}=\mu_Y$ for all values X, which can occur even when X and Y are dependent random variables. However, in some special cases, nullity automatically implies independence, e.g., if Y is a binary random variable or when the joint distribution of Y, X is multivariate normal.

The most popular example of a Bregman scoring function is the quadratic score. Setting $\psi(y)=y^2$, we obtain $S^B(a,y)=-(y-a)^2$, regaining the quadratic score in (6). The information value under this scoring rule is found in Example 3.

4.2. Piecewise linear score functions and quantile reports

The quantile of the distribution of Y is sometimes the preferred point estimate in some applications. A notable example is in finance, where a popular risk measure called value at risk (VaR) typically reports the 95*th* or 99*th* quantile of an investment risk profile (see Fermanian & Scaillet, 2005; Gourieroux, Laurent, & Scaillet, 2000). Suppose the analyst's p-quantile will be evaluated by a scoring rule, then Thomson (1979)'s result shows that the only possible proper rules lie in the family of generalized piece-wise linear loss functions associated with the p-quantile (also known as the check function in quantile regression and the newsvendor function in operations management):

$$S_p^Q(y,a)=k+h\cdot\left[p(t(y)-t(a))^+ +(1-p)(t(a)-t(y))^+\right], \tag{19}$$

where $p\in(0,\ 1)$, $t:\mathbb{R}\to\mathbb{R}$ is a non-decreasing function, $k\in\mathbb{R}$ and $h<0$ are constants.

The optimal report under $S_p^Q(y,a)$ is the p-quantile[5] of Y (Cervera & Muñoz, 1996; Gneiting & Raftery, 2007; Jose, Nau, & Winkler, 2009), which we denote by $Q_Y(p)$. Let $\mathbb{E}^{a,b}[Z]=\int_a^b z dF_Z(z)$ denote the partial expectation of a random variable over interval $(a,\ b)$.

Proposition 12. *The information value of X when the choice problem is to report the p-quantile is given by:*

$$\epsilon_X^Q=h\cdot\mathbb{E}_X\Big\{\mathbb{E}^{-\infty,Q_{Y|X}(p)}[t(Y)|X]+t(Q_{Y|X}(p))\big(p-F_{Y|X}(Q_{Y|X}(p))\big) \tag{20}$$
$$-\mathbb{E}^{-\infty,Q_Y(p)}[t(Y)]+t(Q_Y(p))(p-F_Y(Q_Y(p)))\Big\}.$$

If Y is a continuous r.v. then (20) reduces to $\epsilon_X^Q=h\cdot(\mathbb{E}_X\mathbb{E}^{-\infty,Q_{Y|X}(p)}[t(Y)|X]-\mathbb{E}^{-\infty,Q_Y(p)}[t(Y)])$.

Example 13 [Quantiles of Linear Models]**.** In many applications, Y can be expressed as a linear combination of n exogenous X_i's, i.e., $Y=\sum_{i=1}^n a_iX_i$. For example, Y can be the portfolio return with X_i's being the returns of the assets in the portfolio or Y could be the net present value for an investment with X_i's as potential cash flows.

If normal distributions are appropriate to characterize beliefs about the X_i's, then the quantile information value can easily be expressed in closed form as $\epsilon_X^Q=h\varphi(\Phi^{-1}(p))\{\sigma_Y-\mathbb{E}[\sigma_{Y|X}]\}$, where φ and Φ are, respectively, the standard normal density and cumulative distribution function and σ_Y and $\sigma_{Y|X}$ are, respectively, the portfolio standard deviation and conditional standard deviation. Thus, the quantile information value of X is proportional to the decrease in standard deviation associated with the additional information provided by the variable X under a normality assumption. Of course, this relationship may not hold for more general distributions.

Proposition 12 can easily be extended to interval estimates. Consider a symmetric $(1-\alpha)\times100\%$ interval estimate for Y, where $\alpha\in(0,\ 1)$. This confidence interval $(p_L,\ p_U)$ has the $\alpha/2$-quantile as its lower bound and the $1-(\alpha/2)$-quantile as its upper

[5] Formally, the quantile function is defined as $Q_Y(p)=\inf\{y\in\mathbb{R}:p\le F(y)\}$. However, when Y is absolutely continuous and strictly increasing, then $Q_Y(p)=F_Y^{-1}(p)$.

bound. With $a = (a_L, a_U)$, the scoring function then is simply the sum $\mathcal{S}^{Q}_{\alpha/2}(y, a_L) + \mathcal{S}^{Q}_{1-(\alpha/2)}(y, a_U)$. Because the overall score is additively separable in a_L and a_U, we can express information value for a confidence interval (for an absolutely continuous r.v. Y) as

$$\epsilon_X^{\text{Interval}} = h\Big(\mathbb{E}_X\big\{\mathbb{E}^{-\infty,Q_{Y|X}(\alpha/2)}[t(Y)|X] + \mathbb{E}^{-\infty,Q_{Y|X}(1-(\alpha/2))}[t(Y)|X]\big\} \\ -\mathbb{E}^{-\infty,Q_Y(\alpha/2)}[t(Y)] - \mathbb{E}^{-\infty,Q_Y(1-(\alpha/2))}[t(Y)]\Big).$$

It also follows that the interval information value is additive over the two endpoints. In Example 13, interval information value still remains proportional to the decrease in standard deviation $\sigma_Y - \mathbb{E}[\sigma_{Y|X}]$ under the normality assumption. This holds also when we generalize this concept to a set of quantiles for the set of probabilities $P = \{p_k | k = 1, 2, \ldots, r\}$, with the scoring function $\mathcal{S}^{Q}_{P}(y, a) = \sum_{k=1}^{r} \mathcal{S}^{Q}_{p_k}(y, a)$.

5. Reporting distributions

Often an analyst is interested in the entire distribution (or density) of Y. When moving from point estimates to distributions, some clear differences exist, e.g., by Proposition 6, all the sensitivity measures derived will possess the nullity-implies-independence property.

5.1. Density forecasts

Suppose the density of Y is the quantity of interest to the analyst. (In this section, Y is assumed absolutely continuous.) Then the choice set A is represented by some space of probability densities on $\mathbb{R}$ and a possible report is $a \in A$. Dawid (2007) generalizes the Bregman function to the case of density function reporting, introducing the generalized Bregman score

$$S(y, a) = \psi'(a(y)) + \int_{\mathbb{R}} [\psi(a(s)) - a(s)\psi'(a(s))]ds. \tag{21}$$

This family of scores is theoretically appealing and contains many well-known examples of existing proper scoring rules. Information value under the generalized score (21) encompasses several probabilistic sensitivity measures.

For example, when $\psi(a) = a - a\log a$, we obtain the log scoring rule in (7). Applying the definition of information value and taking the expectation of $\zeta^{\mathcal{D}}(\mathbb{P}_Y, \mathbb{P}_{Y|X})$, we re-obtain the well known (Bernardo & Smith, 2000):

$$\varepsilon_X^{\text{KL}} = \mathbb{E}_X\Big[\int_{\mathbb{R}} f_{Y|X}(y)(\log f_{Y|X}(y) - \log f_Y(y))dy\Big], \tag{22}$$

which is the sensitivity measure $\varepsilon_X^{\text{KL}}$ introduced in Table 1 with the separation measure $\zeta^{KL}(\mathbb{P}, \mathbb{Q})$ being the well-known Kullback-Leibler divergence between f_Y and $f_{Y|X}$ (Kullback & Leibler, 1951). Here, we note that, in information theory, the mutual information is a well known measure of the statistical dependence between Y and X (Ebrahimi, Jalali, & Soofi, 2014). It is easy to see that $\varepsilon_X^{\text{KL}}$ is, in fact, the mutual information of Y and X. Thus, the mutual information can be regarded as a probabilistic sensitivity measure in the common rationale of Eq. (1) (See Pichler and Schlotter (2019) for additional details on information measures and recent advances in such a subject).

As another example, consider the power-based Bregman function $\psi^{\text{Power}}(a) = \omega a^s$ parameterized by $s \notin \{0, 1\}$, where $\omega = 1$ for $0 < s < 1$ and $\omega = -1$ for other permissible values of s. Substituting this into (21), we obtain the power scoring function (Dawid, 2007):

$$S^{\text{Power}}(y, a) = \omega\Big[s a^{s-1}(y) - (s-1)\int_{\mathbb{R}} a^s(t)dt\Big]. \tag{23}$$

The corresponding information value for an analyst reporting the density of Y is

$$\epsilon_X^{\text{Power}} = \omega\mathbb{E}_X\Big[\int_{\mathbb{R}} \Big(f^s_{Y|X}(y) - s f_{Y|X}(y) f_Y^{s-1}(y) + (s-1) f_Y^s(y)\Big)dy\Big]. \tag{24}$$

The separation measure of (24) measures the distance between $f_{Y|X}$ and f_Y. To illustrate, when $s = 2$, $\epsilon_X^{\text{Power}}$ becomes a probabilistic sensitivity measure whose separation measure $\int_{\mathbb{R}} \big(f_{Y|X}(y) - f_Y(y)\big)^2 dy$ is the L^2-norm between densities.

5.2. Distribution forecasts

Analysts may also be interested in the entire cdf of the key variable Y. A clear example is business planning, where decision makers consider the so-called risk-profile, namely, the cdf of a project net present value (Baucells & Borgonovo, 2013). The report is now a cdf $F \in \mathcal{A}$, where $\mathcal{A}$ is an ensemble of candidate cdfs for Y on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. A popular score used in practice is CRPS given by

$$S^{CRPS}(y, F) = -\int_{\mathbb{R}} (F(z) - \mathbf{1}\{z \geq y\})^2 dz, \tag{25}$$

where $\mathbf{1}\{z \geq y\}$ is the indicator variable of $z \geq y$.

This scoring rule has several interesting properties such as strict properness and sensitivity to distance (see Jose et al. (2009) also for a generalization).[6]

Proposition 14. *Information value for X under CRPS is given by:*

$$\epsilon_X^{\text{CRPS}} = \mathbb{E}\Big[\int_{\mathbb{R}} \big(F_Y(y) - F_{Y|X}(y)\big)^2 dy\Big]. \tag{26}$$

Thus, information value in this case is a probabilistic sensitivity measure based on the Cramér-von Mises divergence. The corresponding separation measure $\zeta^{CRPS}(\mathbb{P}_Y, \mathbb{P}_{Y|X}) = \int \big(F_Y(y) - F_{Y|X}(y)\big)^2 dy$ is equal to one-half of the energy statistic applied to comparing F_Y with $F_{Y|X}$. Szekely and Rizzo (2017) highlight that the energy statistic is gaining increasing interest in applied statistics and machine learning as a measure of dependence. Diebold and Shin (2017) recently proposed a related distance of the form $\int |F_Y(y) - F_{Y|X}(y)|^p w(y)dy$ with weighting function w and parameter $p > 0$ that generalizes and connects measures such as Szekelyâs energy statistic, the Cramer-von Mises and Kolmogorov Smirnov metrics.

Example 15 [Example 13 Continued]**.** Consider an analyst who is interested in learning which of the asset returns is more informative when the entire distribution of the portfolio return in Example 13 is of interest and scored with CRPS. Because Y is normally distributed, we can write (Gneiting & Raftery, 2007):

$$S^{CRPS}(y, F(\cdot; \mu, \sigma)) \\ = \sigma\Big[\frac{1}{\sqrt{\pi}} - 2\varphi\Big(\frac{y-\mu}{\sigma}\Big) - \frac{y-\mu}{\sigma}\Big(2\Phi\Big(\frac{y-\mu}{\sigma}\Big) - 1\Big)\Big]. \tag{27}$$

Here the set of forecasts consists of Gaussian distributions $F(\cdot; \mu, \sigma)$ with mean μ and standard deviation σ. The information value of asset return X_i is then given by:

$$\epsilon_{X_i}^{CRPS} = \int_{\mathbb{R}} S^{CRPS}(y, F(\cdot; \mu_Y, \sigma_Y))\nu(y; \mu_Y, \sigma_Y)dy \\ - \int\int_{\mathbb{R}^2} S^{CRPS}(y, F(\cdot; \mu_{Y|X_i}, \sigma_{Y|X_i})) \\ \times \nu(y - a_i x_i; \mu_{Y|X_i}, \sigma_{Y|X_i})\nu(x_i; \mu_{X_i}, \sigma_{X_i})dydx_i, \tag{28}$$

[6] Sensitivity to distance roughly implies that forecasts that place more weight to states that are closer to the outcome that materializes receive a higher score.

Table 6
Summary of the sensitivity measures analyzed in this work. (TI: transformation invariance, NII: nullity-implies-independence property).

Sensitivity Measure	Report	Scoring Rule	TI	NII
η_X (2)	Mean	Quadratic	No	No
ϵ_X^Q (20)	Quantile	Piecewise Linear	No	No
δ_X (3)	Density	N/A	Yes	Yes
ϵ_X^{KL} (22)	Density	Log	Yes	Yes
ϵ_X^{Power} (24)	Density	Power	No	Yes
β_X^{KS} (5)	cdf	N/A	Yes	Yes
ϵ_X^{CRPS} (26)	cdf	CRPS	No	Yes

where $\nu(y;\ \mu,\ \sigma)$ denotes the normal density with mean μ and standard deviation σ. To illustrate, for a portfolio of three standard normally distributed asset returns with relative weights $a_1 = 4/7$, $a_2 = 2/7$ and $a_3 = 1/7$ the information value of X_i expressed as percentage improvement over the expected score is equal to 51% for the first asset, to 10% for second asset and to 2.4% for the third asset. Note that for the linear combination of normal random variables the same expected percentage improvements would be obtained if we were to consider information value for reporting any quantile, because CRPS can be interpreted as a weighted average of quantile scores.

6. Choosing an appropriate sensitivity measure

Table 6 lists seven of the probabilistic sensitivity measures discussed earlier, classified according to the type of report and its corresponding scoring rule (if any). What sensitivity measure to use? Our analysis has shown that the analyst should first consider type of report (mean, quantile(s), or distribution) to produce. This may depend on the analyst's anticipated audience, or there may be a specific requirement. Should the desired report include some measure of central tendency and/or a prediction interval, then one of the first two sensitivity measures η_X or ϵ_X^Q in Table 6 would be appropriate. Although these are not transformation invariant and do not obey nullity-implies-independence, the analyst may nevertheless support their use as sensitivity measures by noting they are equal to information value should the report quality be evaluated respectively by the quadratic or piecewise-linear scoring rules.

Should a *distribution* report be requested or allowed, the analyst can then choose from one of the last five sensitivity measures in the table. All satisfy nullity-implies-independence. Should transformation invariance be desired, there are several possibilities, δ_X, ϵ_X^{KL} and β_X^{KS}, but only ϵ_X^{KL} is known to be information value under a scoring rule. If transformation invariance is not crucial, then ϵ_X^{Power} and ϵ_X^{CRPS} could be justified as information value under a corresponding scoring rule as well. If instead it is desired to use one of the others such as δ_X and β_X^{KS}, perhaps for computational convenience, then Theorem 8 provides partial reassurance that these are also information value, although not under any model-independent scoring rule. And much as we have already noted, there would be no rationale to use the first two sensitivity measures from the table, since they are information value under scoring rules whose optimal report is not a distribution.

Although these rationales do not uniquely identify the appropriate sensitivity measure under all reporting circumstances, they do narrow the field and provide reasonable justification for the measure or measures eventually chosen. In the same manner, if a new scoring rule comes to mind, a sensitivity measure can be derived using the framework presented in the earlier section that will allow the analyst to see whether it can be interpreted as information value as well as to check if properties such as transformation invariance and nullity-implies-independence hold. To illustrate, suppose for a moment that the new scoring rule is $S^{Power}(y,\ q_Y)$ in (23) with $s = 2$. By Proposition 5, we know that the information value for this scoring rule is a probabilistic sensitivity measure that follows Definition 2. Because the report is the density of Y, by Proposition 6 we know that the resulting sensitivity measure possesses the nullity-implies-independence property, although it is not transformation invariant.

Table 7
Summary of the most important inputs for the probability of loss of mission, across eight sensitivity measures. The symbols $\sim$, $>$ and $\gg$ denote differences in sensitivity between 0% and 10%, between 10% and 50%, and larger than 50%, respectively.

Report	Scoring Rule	Measure	Important Variables
0.05-Quantile, Q_{05}	Pw. Linear	ϵ_X^{05}	$X_{748} \gg X_{145} > X_{177} \sim X_{152}$
Median, Q_{50}	Pw. Linear	ϵ_X^{50}	$X_{748} \sim X_{152} > X_{143} \gg X_{145}$
0.95-Quantile, Q_{95}	Pw. Linear	ϵ_X^{95}	$X_{152} > X_{143}$
Distribution	CRPS	ϵ_X^{CRPS}	$X_{152} > X_{143} > X_{748}$
Mean	Quadratic	$\epsilon_X^{Quad} = \eta_X$	$X_{152} > X_{143}$
Density	Logarithmic	$\epsilon_X^{KL} = \theta_X$	$X_{748} \gg X_{145}$
Density	N/A	δ_X	$X_{748} > X_{152} > X_{143}$
Distribution	N/A	β_X^{KS}	$X_{748} > X_{152} > X_{143} \gg X_{145}$

7. Application: The NASA space probabilistic risk assessment case study

The case study we present is part of a wide class of decisions in which the report is used by a panel of experts/advisors in a complex decision making process. The problem relates to the design and planning of a lunar space mission. The analysis follows the guidelines in Dezfuli et al. (2011) that foresee the development of a decision support model in the form of a probabilistic safety assessment. The insights from the model are used for two main purposes: (i) to determine whether the mission complies with the agency's safety goals, and (ii) to gain insights on options for improving risk management and mission performance.

A detailed description of the systems involved can be found in NASA (2005). Structurally, the model of interest is a complex Bayesian network. The output of interest, Y, is the probability of loss of mission calculated based on the occurrence of several hundred related events. The model considers about 4,446 scenarios. Mathematically, Y is a function of 872 uncertain inputs. Most of these inputs are either failure probabilities or failure rates of relevant equipment. Uncertainty in these parameters has been assessed by engineers who have assigned the distributions detailed in Borgonovo and Smith (2011). Uncertainty in Y is then assessed through Monte Carlo simulations.

The uncertainty quantification then needs to be complemented by a sensitivity analysis displaying the key drivers of uncertainty. From the available Monte Carlo sample, we estimate eight sensitivity measures, that yield information about alternative properties of the model output distribution (Table 7): the probabilistic sensitivity measures ϵ_X^{05}, ϵ_X^{50}, ϵ_X^{95} consider quantiles as forecast of interest, η_X considers the mean, ϵ_X^{KL} and δ_X the model output density, and ϵ_X^{CRPS}, β^{KS} the output cdf. Also, the ensemble includes sensitivity measures that have a direct value of information interpretation in the sense of Proposition 5 (ϵ_X^{05}, ϵ_X^{50}, ϵ_X^{95}, η_X, ϵ_X^{KL}, ϵ_X^{CRPS}), or in the more general sense of Theorem 8 (ϵ_X^{KL}, δ_X) and that are transformation invariant ($\epsilon_X^{KL}, \delta_X, \beta_X^{KS}$).

The graphs in Fig. 3 display the importance of input parameters on the vertical axis, and the number of the parameter on the horizontal axis. The first three graphs refer to the 0.05-, 0.50-, and 0.95-quantiles, the fourth to CRPS, the fifth to the quadratic score, and the sixth to the logarithmic score for densities. The seventh and eighth graphs report results for the β_X^{KS} and δ_X sensitivity measures.

A visual inspection of these graphs suggests that the selection of the most informative inputs depends notably on the forecast elicited from the analyst. Table 7 provides the impression an analyst may receive from these graphs. In particular, the fourth

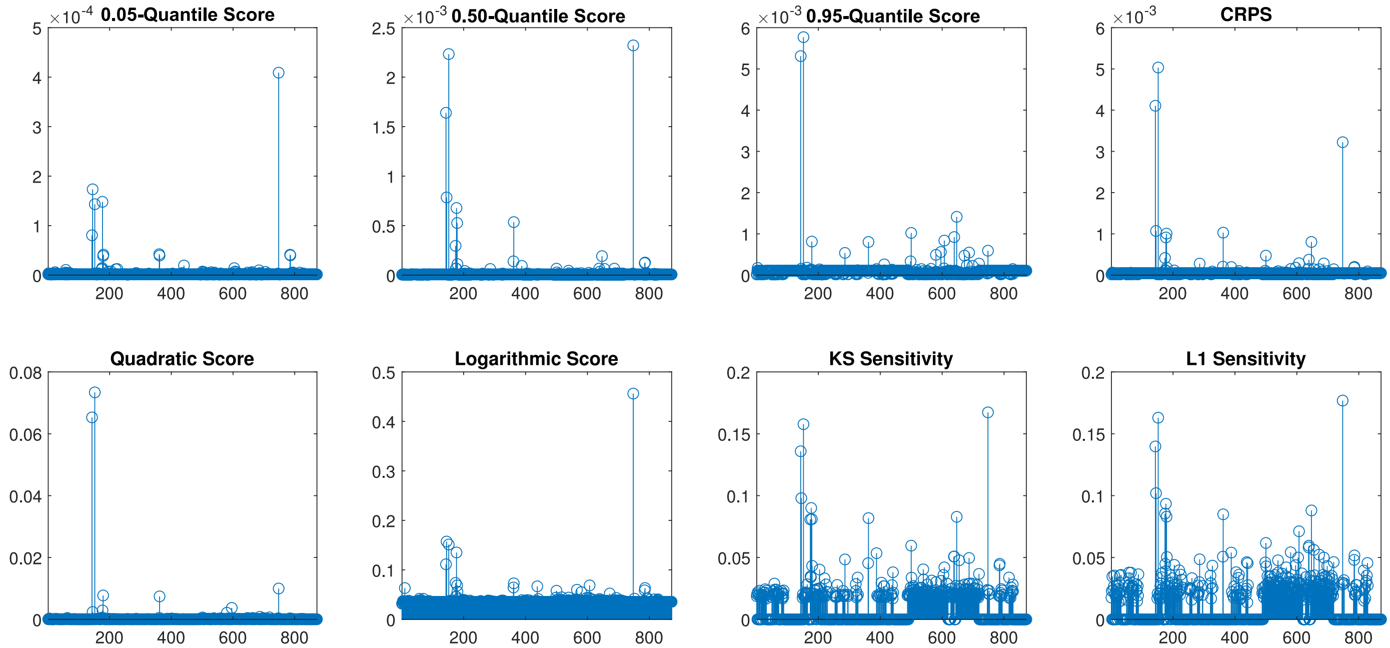

Fig. 3. Sensitivity measures for the probability of loss of mission.

column reports the most important inputs for each forecast report. Specifically, we list the group of inputs whose sensitivity measure is at least 33% of the sensitivity measure value for the most important input.

For instance, consider an analyst that defaults to the commonly used variance-based sensitivity measure. She would report X_{152} and X_{143} as the most important inputs. This report would be consistent if the analyst is asked to report a mean. However, if the analyst is reporting the 0.05-quantile, then X_{748} would the most important basic event probability, followed by X_{145}, X_{157}, and X_{152}. Now, if the report of interest is the density of the output using a logarithmic score, then X_{748} and X_{145} are the most important inputs. If the analyst is reporting the cumulative distribution function using CRPS, then she would report X_{152}, X_{143} and X_{748} as most important inputs.

These results do confirm the suggestions of Section 6: Randomly picking a sensitivity measure exposes the analyst to the risk of miscommunicating the most important uncertainties. For instance, if one used variance-based sensitivity measures, one would pinpoint a component of the crew module (X_{152}) and a failure parameter related to the orbiting phase of the missions (X_{143}). Conversely, if the report is a density, and ϵ_X^{KL} is considered, risk management would have to prioritize a component of the lunar surface access module (X_{748}) and a different component of the crew model (X_{145}).

However, in several situations the analyst or the decision maker may not feel confident enough to rely on a single forecast/score. In this case, an analysis based on an ensemble of sensitivity measures can be advantageous in pinpointing the key drivers of uncertainty. Table 7 would suggest that X_{748} and X_{152} are two important inputs, with X_{748} most influential when the report concerns the density ($\epsilon_X^{KL}, \delta_X$) and 0.05-quantile, and X_{152} most influential when the report is the 0.95-quantile, the mean, or the entire distribution under CRPS. It would also suggest X_{143} and X_{145} as relevant. Finally, one can also answer question concerning the relevance of specific components/pieces of equipment. For instance, X_{177} is an input related to the failure of the parachute deployment system, which is a crucial system in landing operations. This input is important for the 0.05-quantile of the distribution, but its relevance is lower when other types of forecasts are considered.

In either case (if attention is focused on a given report or if a holistic view on the sensitivity measures is adopted), the analyst has a way to provide solid recommendations about which variables are more relevant for further information collection. This has the potential of reducing uncertainty in predictions and consequently of making the decision and risk management process better informed.

8. Conclusions

In this paper, we investigated the decision-theoretic aspects of probabilistic sensitivity analysis through a synthesis of information value, probabilistic sensitivity analysis, and scoring rules. This synthesis yields a variety of results and practical insights. First, it allows one to better characterize the conditions under which a probabilistic sensitivity measure can be interpreted as information value. Second, it permits us to identify the value-of-information sensitivity measure consistent with a specified reporting problem. Third, it provides a better understanding of various properties of probabilistic sensitivity measures such as nullity-implies-independence and transformation invariance. We also introduced new probabilistic sensitivity measures that retain a value-of-information interpretation. Some interesting new connections to existing measures have been established. For example, we prove that Szekely's popular energy statistic is information value consistent with the CRPS score. Some of our results give rise to broad characterizations, such as the fact that well-known sensitivity measures based on metrics such as the L^1-norm and Kolmogorov-Smirnov are information value under some model-specific scoring function. We believe that the understanding of these connections and the application of these insights can aid analysts in the selection of appropriate sensitivity measures that are meaningful, insightful, and relevant to the end users of these models.

Acknowledgments

The authors thank Prof. Roman Slowinski for the editorial attention. We also wish to thank the two anonymous reviewers for many constructive comments that have greatly helped us improve the manuscript.

Appendix A. Details on the introductory example

The trapezoidal densities of the uncertain rates are:

$$f(x,b,\mu)=\begin{cases}(1-\frac{x}{b})K_0(b,\mu)+\frac{x}{b}K_1(b,\mu) & \text{if} \quad 0\le x\le b\\ 0 & \text{otherwise}\end{cases}$$

with $K_0(b,\mu)=\frac{2}{b^2}(2b-3\mu)$, and $K_1(b,\mu)=\frac{2}{b^2}(-b+3\mu)$. The conditional distribution of Y given $X_1=\lambda_1$ and $X_2=\lambda_2$ is

$$F_Y(y|\lambda_1,\lambda_2)=1-e^{-(\lambda_0+\lambda_1+\lambda_2)y}.$$

The unconditional distribution is therefore

$$F_Y(y)=1-\mathbb{E}_{X_1,X_2}[e^{-(\lambda_0+X_1+X_2)y}].$$

As we assume independence, we obtain

$$F_Y(y)=1-e^{-\lambda_0 y}L_1(y)L_2(y),$$

where $L_i(y)=\mathbb{E}_{X_i}[e^{-X_i y}]$, $i=1,2$. The conditional distributions of Y given X_1 or X_2 are

$$F_Y(y|X_1=\lambda_1)=1-e^{-(\lambda_0+\lambda_1)y}L_2(y) \quad \text{and} \quad F_Y(y|X_2=\lambda_2)=1-e^{-(\lambda_0+\lambda_2)y}L_1(y).$$

Let us consider that the analyst chooses a quadratic scoring rule, i.e., a variance based sensitivity measure, as we are to see. Then as we show in Section 4, value of information is given by

$$\varepsilon_{X_i}^{\text{Quad}}=\mathbb{V}_{X_i}[\mathbb{E}[Y|X_i]]. \tag{29}$$

We then need to compute $\mathbb{E}[Y|X_i]$, $i=1,2$. This conditional expectation is given for X_1 by

$$\mathbb{E}[Y|X_1=\lambda_1]=\mathbb{E}_{X_2}[(\lambda_0+\lambda_1+X_2)^{-1}]. \tag{30}$$

Let now $K_2(v)=\mathbb{E}_{X_2}[(v+X_2)^{-1}]$. By substituting (30) into (29) one obtains

$$\varepsilon_{X_1}^{\text{Quad}}=\mathbb{E}_{X_1}[K_2(\lambda_0+X_1)^2]-\mathbb{E}_{X_1}[K_2(\lambda_0+X_1)]^2.$$

For the second contributing rate X_2 one proceeds analogously.

For $\varepsilon_{X_1}^{\text{CRPS}}$ we obtain:

$$\varepsilon_1^{\text{CRPS}}=\mathbb{E}[\int_0^\infty(F_Y(y)-F_Y(y|X_1))^2dy=\int_0^\infty\mathbb{E}[(F_Y(y)-F_Y(y|X_1))^2]dy$$
$$=\int_0^\infty\mathbb{E}[\mathbb{V}\{F_Y(y|X_1)\}]dy.$$

The last term of the previous equality is given by

$$\begin{aligned}\mathbb{V}\{F_Y(y|X_1)\}&=\mathbb{E}[F_Y(y|X_1)^2]-F_Y(y)^2\\&=e^{-2\lambda_0 y}\mathbb{E}[e^{-2X_1 y}]L_2(y)^2-e^{-2\lambda_0 y}L_2(y)^2L_1(y)^2\\&=e^{-2\lambda_0 y}L_2(y)^2[L_1(2y)-L_1(y)^2],\end{aligned}$$

so that

$$\varepsilon_1^{\text{CRPS}}=\int_0^\infty e^{-2\lambda_0 y}L_2(y)^2[L_1(2y)-L_1(y)^2]dy. \tag{31}$$

The expressions in (30) and (31) are easily implemented in a software such as Mathcad, Matlab, Mathematica (the first two are used by the authors).

With the following parameterization, $X_1\sim$ trapezoidal(·, 1.7, 0.595), $X_2\sim$ trapezoidal(·, 1.2, 0.78), one obtains the marginal input and output distributions in Fig. 1, and the values of the sensitivity measures in Table 8.

Table 8
Sensitivity measures for the example.

Sensitivity Measure	Sensitivity to λ_1	Sensitivity to X_2
Variance-based (quadratic scoring)	0.0854	0.1018
Distribution-based (CRPS scoring)	0.0186	0.0164

Appendix B. Proofs and derivations

Proof of Proposition 6

We can write $\epsilon_X^{\mathcal{S}}=\mathbb{E}_X[\zeta^{\mathcal{S}}(\mathbb{P}_Y,\mathbb{P}_{Y|X})]$, where

$$\zeta^{\mathcal{S}}(\mathbb{P}_Y,\mathbb{P}_{Y|X})=\mathbb{E}_Y[S(Y,a^*(X))-S(Y,a^*)|X].$$

We also know that $\zeta^{\mathcal{S}}(\mathbb{P}_Y,\mathbb{P}_{Y|X})$ is greater than or equal to zero for all values of X. Then, if the scoring rule is strictly proper, $a^*(X)=F_{Y|X}$ and $a^*=F_Y$ must maximize the expected score. Therefore,

$$\zeta^{\mathcal{S}}(\mathbb{P}_Y,\mathbb{P}_{Y|X})=\mathbb{E}_Y[S(Y,F_{Y|X})-S(Y,F_Y)|X].$$

We already know $\epsilon_X^{\mathcal{S}}=0$ if Y, X are independent, because $\epsilon_X^{\mathcal{S}}$ is a sensitivity measure. Conversely, suppose that $\epsilon_X^{\mathcal{S}}=0$. Then, the nonnegativity of $\zeta^{\mathcal{S}}(\mathbb{P}_Y,\mathbb{P}_{Y|X})$ forces $\zeta^{\mathcal{S}}(\mathbb{P}_Y,\mathbb{P}_{Y|X})$ to be zero for almost all X. That is, for almost all X, $\mathbb{E}_Y[S(Y,F_{Y|X})|X]=\mathbb{E}_Y[S(Y,F_Y)|X]$. Now, because S is strictly proper, the distribution $F_{Y|X}$ is, for each value X, the unique maximizer of $\mathbb{E}_Y[S(Y,F_{Y|X})|X]$. Therefore $F_{Y|X}=F_Y$ for almost all X, which shows that Y and X are independent.

Proof of Theorem 8

For brevity, let u_B be the random variable $U(\mathbb{P}_{Y|B})$ for $B\in\mathcal{B}$, so that $\mathbb{E}[u_B|A]=\mathbb{E}[U(\mathbb{P}_{Y|B})|A]$. Because the σ-algebra $\mathcal{B}$ contains only finitely many sets, it has a finite basis $B_1,\ldots,B_m$ of mutually exclusive and collectively exhaustive event sets. Because u_B is $\mathcal{B}$-measurable, it follows that u_B has only finitely many values, one for each B_i, and can be regarded as a vector. Let π_A be the vector of conditional probabilities given A, having in the same way one value $p_{i|A}$ for each B_i. Then $\mathbb{E}[u_B|A]=\pi_A\cdot u_B$. To reduce notation clutter, we take all statements $A\in\mathcal{B}$ below to mean $A\in\mathcal{B}\setminus\{\emptyset\}$. Then the assumptions that U is proper and consistent with ζ can be written

$$\begin{aligned}&\pi_A\cdot u_B\le\zeta_A+u_0, \qquad A\in\mathcal{B},A\neq B,\\&\pi_B\cdot u_B=\zeta_B+u_0.\end{aligned} \tag{32}$$

Our goal is to show that this system has a solution u_B, implying the existence of a proper scoring function consistent with ζ.

Following Ch. 1 in Stoer and Witzgall (1970), we use the Kuhn-Fourier Theorem to write down necessary and sufficient conditions for this system to have a solution. The possible legal linear combinations of the system (32) are

$$\sum_{\substack{A\in\mathcal{B}\\A\neq B}}V_A\pi_A u_B+W\pi_B u_B\le\sum_{\substack{A\in\mathcal{B}\\A\neq B}}V_A(\zeta_A+u_0)+W(\zeta_B+u_0) \tag{33}$$

where $V_A\ge 0$ for all A, $V_A>0$ for some A and $W\in\mathbb{R}$ and

$$W\pi_B=W(\zeta_B+u_0). \tag{34}$$

A legal linear combination of a system of equations and inequalities is a *legal linear dependence* if its left side is always zero but not all V_A and W are zero. The Kuhn-Fourier Theorem states that the system (32) has a solution if and only if every legal linear dependence is always true. For (33), this means that if $V_A\ge 0$ all A, $V_A>0$ for some A, and $W\in\mathbb{R}$ then

$$\sum_{\substack{A\in\mathcal{B}\\A\neq B}}V_A\pi_A+W\pi_B=0\Rightarrow\sum_{\substack{A\in\mathcal{B}\\A\neq B}}V_A(\zeta_A+u_0)+W(\zeta_B+u_0)\ge 0. \tag{35}$$

For (34), this means that for $W\neq 0$

$$W\pi_B=0\Rightarrow W(\zeta_B+u_0)=0. \tag{36}$$

The last implication is vacuously true, since π_B is never zero for $B\neq\emptyset$. Consider then (35). Note that $\sum_{A\in\mathcal{B}, A\neq B}V_A\pi_A>0$, because $V_A>0$ for some A. Therefore, in the premise of (35), W must be strictly negative. Replace W by its negative and solve on both sides of (35) to get the following equivalent version of (35):

$$\pi_B = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\pi_A \Rightarrow \zeta_B + u_0 \le \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A(\zeta_A + u_0), \tag{37}$$

where now we are using new V_A equal to the old $V_A/(-W)$. Multiply each side of the premise to (37) by a vector of ones to obtain $1 = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A$. Therefore, the scalar u_0 on the right side of (37) cancels, and (37) is equivalent to

$$\pi_B = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\pi_A \Rightarrow \zeta_B \le \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\zeta_A, \tag{38}$$

for every collection $\{V_A | A\in\mathcal{B}, A\neq B\}$ with $V_A \ge 0$ and $\sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A = 1$. If we can demonstrate this, then it follows that the system (32) has a solution.

So suppose $V_A \ge 0$ and $\sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A = 1$, and the premise of (38) holds. In terms of the variables $p_{i|B}$ mentioned at the beginning of this proof, this premise is equivalent to

$$p_{i|B} = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A p_{i|A} \quad i = 1,\ldots,m$$

Therefore

$$\mathbb{P}_{Y|B}(dy) = \sum_i \mathbb{P}_{Y|B_i}(dy)p_{i|B} = \sum_i \mathbb{P}_{Y|B_i}(dy)\sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A p_{i|A}$$
$$= \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A \sum_i \mathbb{P}_{Y|B_i}(dy)p_{i|A} = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\mathbb{P}_{Y|A}(dy)$$

Then invoking the convexity of ζ in its second argument, we have

$$\xi_B = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|B}) = \zeta(\mathbb{P}_Y, \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\mathbb{P}_{Y|A}) \le \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|A}) = \sum_{\substack{A\in\mathcal{B}\\A\neq B}} V_A\xi_A.$$

Therefore the conclusion of (38) holds, and we have demonstrated (38). Therefore, the system (32) has a solution.

Proofs of convexity claims for corollary 9

For the sensitivity measure η_X, we have

$$\zeta^{\eta}(\mathbb{P}, \mathbb{Q}) = (\mu_{\mathbb{P}} - \mu_{\mathbb{Q}})^2 = \left(\mu_{\mathbb{P}} - \int_{\mathbb{R}} y\mathbb{Q}(dy)\right)^2.$$

This is a convex quadratic function of a linear function $\mathbb{Q} \mapsto \int_{\mathbb{R}} y\mathbb{Q}(dy)$, hence is convex.

For the sensitivity measure δ_X, we have

$$\zeta^{L1}(\mathbb{P}, \mathbb{Q}) = \frac{1}{2}\int_{\mathbb{R}} |f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y)|dy.$$

This is a composition of mappings $f_{\mathbb{Q}} \mapsto |f_{\mathbb{P}} - f_{\mathbb{Q}}| \mapsto \frac{1}{2}\int_{\mathbb{R}} |f_{\mathbb{P}}(y) - f_{\mathbb{Q}}(y)|dy$, which is a linear functional (on a space of functions) following a convex function (from a space of densities to a function space) of $f_{\mathbb{Q}}$. Hence the composition is convex in $f_{\mathbb{Q}}$.

For the sensitivity measure $\varepsilon_X^{\mathrm{KL}}$, we have

$$\zeta^{KL}(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y)(\ln f_{\mathbb{Q}}(y) - \ln f_{\mathbb{P}}(y))dy$$
$$= \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y)\ln f_{\mathbb{Q}}(y)dy - \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y)\ln f_{\mathbb{P}}(y)dy.$$

The second term in this difference is linear in $f_{\mathbb{Q}}$, so the overall function will be convex in $f_{\mathbb{Q}}$ if the first term is. Note that the first term is a composition $f_{\mathbb{Q}} \mapsto f_{\mathbb{Q}}\cdot \ln g \mapsto \int_{\mathbb{R}^+} f_{\mathbb{Q}}(y)\ln f_{\mathbb{Q}}(y)dy$, which is a linear function following the transformation $f_{\mathbb{Q}} \mapsto f_{\mathbb{Q}}\cdot\ln f_{\mathbb{Q}}$, and the latter is convex because its pointwise analog $y \mapsto y\cdot\ln y$ is convex, as may be verified by checking the second derivative. Therefore the overall transformation is convex in $f_{\mathbb{Q}}$.

For the sensitivity measure β_X^{KS}, we have the following. Note that the Kolmogorov-Smirnov distance, is given by

$$\zeta^{KS}(\mathbb{P}, \mathbb{Q}) = \sup_{y\in\mathbb{R}} |F_{\mathbb{P}}(y) - F_{\mathbb{Q}}(y)|,$$

which is a composition $F_{\mathbb{Q}} \mapsto |F_{\mathbb{P}} - F_{\mathbb{Q}}| \mapsto \sup_{y\in\mathbb{R}} |F_{\mathbb{P}}(y) - F_{\mathbb{Q}}(y)|$, that is, a linear function following a convex function of $F_{\mathbb{Q}}$.

Calculations for Eqs. (14) and (15) in Section 3.3

For the separation measure in (14), we have:

$$\eta'_X = \mathbb{E}_X\left[\zeta'^{\eta}(\mathbb{P}_Y, \mathbb{P}_{Y|X})\right]$$
$$= \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X} + k)^2 - k^2\right] = \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X})^2 + 2k(\mu_Y - \mu_{Y|X})\right]$$
$$= \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X})^2\right] + 2k\mathbb{E}_X\left[\mu - \mu_{Y|X}\right] = \mathbb{E}_X\left[(\mu_Y - \mu_{Y|X})^2\right] + 0 = \eta_X.$$

for arbitrary Y, X.

For the separation measure in (15), because $z^+ = \frac{1}{2}(z + |z|)$, we have for any Y, X

$$\mathbb{E}_X[\zeta^+(\mathbb{P}_Y, \mathbb{P}_{Y|X})] = \mathbb{E}_X\left[\int_{\mathbb{R}} \left(f_Y(y) - f_{Y|X}(y)\right)^+ dy\right]$$
$$= \mathbb{E}_X\left[\int_{\mathbb{R}} \left(\frac{1}{2}\left|f_Y(y) - f_{Y|X}(y)\right| + \frac{1}{2}\left(f_Y(y) - f_{Y|X}(y)\right)\right)dy\right]$$
$$= \mathbb{E}_X\left[\frac{1}{2}\int_{\mathbb{R}} \left|f_Y(y) - f_{Y|X}(y)\right|\right]dy + 0 = \delta_X.$$

Proof of Proposition 10

To prove Proposition 10, we first prove the following result.

Proposition 16. *Consider a scoring rule $S(y, a)$ and suppose $a_{\mathbb{P}}^S$ is the set of optimal reports whenever Y has distribution $\mathbb{P}$. A necessary condition for the sensitivity measure ξ_X with separation measure ζ to be, for all X, the value ϵ_X^S of information X under scoring rule S is that for all distributions $\mathbb{Q}_1, \mathbb{Q}_0$, and all α with $0 < \alpha < 1$,*

$$a_{\mathbb{Q}_0}^S = a_{\mathbb{Q}_1}^S \Rightarrow \zeta((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_0)$$
$$= \zeta((1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1, \mathbb{Q}_1) = 0. \tag{39}$$

The proof of the above proposition is divided into two steps. First, we state and prove the following lemma:

Lemma 17. *If a_X^S does not depend on X, then $\epsilon_X^S = 0$.*

Proof of Lemma 17. Suppose $a_X^S = a_0$ for all X, that is, $a = a_0$ optimizes $\mathbb{E}[S(Y, a)|X]$ regardless of X. Then $a = a_0$ must also optimize $\mathbb{E}[\mathbb{E}[S(Y, a)|X]]$. But the latter is equal to $\mathbb{E}[S(Y, a)]$. Therefore $a = a_0$ and $a = a^S$ are both optimizers of $\mathbb{E}[S(Y, a)]$, so that $\mathbb{E}[S(Y, a_0)] = \mathbb{E}[S(Y, a^S)]$. Consequently $\epsilon_X^S = \mathbb{E}[S(Y, a_0)] - \mathbb{E}[S(Y, a^S)] = 0$. □

We now prove Proposition 16. As in the statement of the proposition, let $\mathbb{Q}_0$ and $\mathbb{Q}_1$ be arbitrary distributions over the possible values of Y such that $a_{\mathbb{Q}_0}^S = a_{\mathbb{Q}_1}^S$, and let

$$\mathbb{P} = (1-\alpha)\mathbb{Q}_0 + \alpha\mathbb{Q}_1.$$

for $0 < \alpha < 1$. Let X be a binary variable with $\alpha = P(X = 1) = 1 - P(X = 0)$, and suppose Y has conditional distributions $\mathbb{P}_{Y|X=0} = \mathbb{Q}_0$, and $\mathbb{P}_{Y|X=1} = \mathbb{Q}_1$. Then $\mathbb{P}_Y = \mathbb{P}$. Under scoring rule S, the optimal report set given $X = 1$ is $a_{\mathbb{Q}_1}^S$ and the optimal report set given $X = 0$ is $a_{\mathbb{Q}_0}^S$. Because these two sets are by hypothesis the same, the information value of X under score S must be zero according to the lemma.

Suppose the information value of X under S is equal to the sensitivity measure ξ_X. We therefore have

$$0 = \xi_X = \mathbb{E}[\zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X})] = (1-\alpha)\zeta(\mathbb{P}, \mathbb{Q}_0) + \alpha\zeta(\mathbb{P}, \mathbb{Q}_1)$$

for $0<\alpha<1$, as desired.

We can then prove Proposition 10. The necessary condition of the proposition is equivalent to a simplified version of (39), namely to

$$a^S_{\mathbb{Q}_0}=a^S_{\mathbb{Q}_1}\Rightarrow \zeta(\mathbb{Q}_0,\mathbb{Q}_1)=\zeta(\mathbb{Q}_1,\mathbb{Q}_0)=0 \tag{40}$$

which we now demonstrate. The continuity hypothesis of the proposition implies

$$\lim_{\alpha\uparrow 1}\zeta((1-\alpha)\mathbb{Q}_0+\alpha\mathbb{Q}_1,\mathbb{Q}_0)=\zeta(\mathbb{Q}_1,\mathbb{Q}_0).$$

Then from (39), because $\zeta((1-\alpha)\mathbb{Q}_0+\alpha\mathbb{Q}_1,\mathbb{Q}_0)=0$ for all $\alpha\in(0,1)$, we obtain $\zeta(\mathbb{Q}_1,\mathbb{Q}_0)=0$. Similarly, the continuity hypothesis of the proposition implies

$$\lim_{\alpha\downarrow 0}\zeta((1-\alpha)\mathbb{Q}_0+\alpha\mathbb{Q}_1,\mathbb{Q}_1)=\zeta(\mathbb{Q}_0,\mathbb{Q}_1),$$

whence we obtain for a similar reason $\zeta(\mathbb{Q}_0,\mathbb{Q}_1)=0$.

Proof of Proposition 11

By definition and because the optimal action under Bregman scoring is the mean,

$$\begin{aligned}\epsilon^B_X=&\mathbb{E}_X\big[\mathbb{E}[\psi(\mu_{Y|X})+\psi'(\mu_{Y|X})(Y-\mu_{Y|X})-\psi(Y)|X]\\&-\mathbb{E}[\psi(\mu_Y)+\psi'(\mu_Y)(Y-\mu_Y)-\psi(Y)]\big]\\=&\mathbb{E}_X\big[\mathbb{E}[\psi(\mu_{Y|X})]+\psi'(\mu_{Y|X})\mathbb{E}[(Y-\mu_{Y|X})]-\mathbb{E}[\psi(Y)|X]-\mathbb{E}[\psi(\mu_Y)]\\&-\psi'(\mu_Y)\mathbb{E}[(Y-\mu_Y)]+\mathbb{E}[\psi(Y)]\big]\\=&\mathbb{E}_X\big[\mathbb{E}[\psi(\mu_{Y|X})]+0-\mathbb{E}[\psi(Y)|X]-\mathbb{E}[\psi(\mu_Y)]-0+\mathbb{E}[\psi(Y)]\big]\\=&\mathbb{E}_X[\psi(\mu_{Y|X})]-\mathbb{E}[\psi(Y)]-\mathbb{E}[\psi(\mu_Y)]+\mathbb{E}[\psi(Y)]=\mathbb{E}_X[\psi(\mu_{Y|X})]-\psi(\mu_Y).\end{aligned}$$

Also from Proposition 5, we have

$$\begin{aligned}\zeta^B(\mathbb{P}_Y,\mathbb{P}_{Y|X})=&\mathbb{E}\big[S^B(Y,\mu_{Y|X})-S^B(Y,\mu_Y)\big|X\big]\\=&\mathbb{E}\big[\psi(\mu_{Y|X})+\psi'(\mu_{Y|X})(Y-\mu_{Y|X})-\psi(Y)\\&-(\psi(\mu_Y)+\psi'(\mu_Y)(Y-\mu_Y)-\psi(Y))\big|X\big]\\=&\mathbb{E}\big[\psi(\mu_{Y|X})-(\psi(\mu_Y)+\psi'(\mu_Y)(Y-\mu_Y))\big|X\big]\\=&\psi(\mu_{Y|X})-(\psi(\mu_Y)+\psi'(\mu_Y)(\mu_{Y|X}-\mu_Y))\end{aligned}$$

as desired.

Proof of Proposition 12

Without loss of generality, we set $k=0$. By definition, we have:

$$\begin{aligned}\mathbb{E}[S^Q_p(Y,a)]&=h\int_{-\infty}^{\infty}\big[p(t(y)-t(a))^+ +(1-p)(t(a)-t(y))^+\big]dF_Y(y)\\&=h\left[p\int_a^{\infty}(t(y)-t(a))dF_Y(y)+(1-p)\int_{-\infty}^{a}(t(a)-t(y))dF_Y(y)\right]\\&=h\big[p\big(\mathbb{E}^{a,\infty}[t(Y)]-t(a)(1-F_Y(a))\big)+(1-p)\big(t(a)F_Y(a)-\mathbb{E}^{-\infty,a}[t(Y)]\big)\big]\\&=h\big[p\mathbb{E}^{a,\infty}[t(Y)]-(1-p)\mathbb{E}^{-\infty,a}[t(Y)]+t(a)(F_Y(a)-p)\big].\end{aligned}$$

Substituting $a^*=Q_Y(p)$ and using the identity $\mathbb{E}^{b,\infty}(g(Z))=\mathbb{E}(g(Z))-\mathbb{E}^{-\infty,b}(g(Z))$ yields

$$\begin{aligned}\mathbb{E}[S^Q_p(Y,a^*)]=&h\big[(1-p)\mathbb{E}_Y^{-\infty,Q_Y(p)}[t(Y)]-p\mathbb{E}^{Q_Y(p),\infty}[t(Y)]\\&+t(Q_Y(p))(p-F_Y(Q_Y(p)))\big]=h\big[\mathbb{E}^{-\infty,Q_Y(p)}[t(Y)]-p\mathbb{E}[t(Y)]\\&+t(Q_Y(p))(p-F_Y(Q_Y(p)))\big].\end{aligned}$$

An analogous expression holds for the conditional r.v. $Y|X$. Inserting these into (8), we get

$$\begin{aligned}\epsilon^Q_X=&h\mathbb{E}_X\Big\{\mathbb{E}^{-\infty,Q_{Y|X}(p)}[t(Y)|X]-p\mathbb{E}[t(Y)|X]+t(Q_{Y|X}(p))\cdot\\&\big(p-F_{Y|X}(Q_{Y|X}(p))\big)\Big\}\\&-h\Big\{\mathbb{E}^{-\infty,Q_Y(p)}[t(Y)]-p\mathbb{E}[t(Y)]+t(Q_Y(p))(p-F_Y(Q_Y(p)))\Big\}\\=&h\mathbb{E}_X\Big\{\mathbb{E}^{-\infty,Q_{Y|X}(p)}[t(Y)|X]+t(Q_{Y|X}(p))\big(p-F_{Y|X}(Q_{Y|X}(p))\big)\Big\}\\&-h\Big\{\mathbb{E}^{-\infty,Q_Y(p)}[t(Y)]+t(Q_Y(p))(p-F_Y(Q_Y(p)))\Big\},\end{aligned}$$

which completes the proof.

Proof of Proposition 14

By definition of information value, we have

$$\epsilon^{CRPS}_X=\mathbb{E}\Big\{\max_{F\in A}\mathbb{E}\big[S^{CRPS}(Y,F)|X\big]\Big\}-\max_{F\in A}\mathbb{E}\big[S^{CRPS}(Y,F)\big].$$

and by strict properness, the two maximizers are $F=F_{Y|X}$ and $F=F_Y$, respectively. Then

$$\begin{aligned}\epsilon^{CRPS}_X&=\mathbb{E}\Big\{\mathbb{E}\big[S^{CRPS}(Y,F_{Y|X})|X\big]\Big\}-\mathbb{E}\big[S^{CRPS}(Y,F_Y)\big]\\&=\mathbb{E}\Big\{\mathbb{E}\big[S^{CRPS}(Y,F_{Y|X})-S^{CRPS}(Y,F_Y)|X\big]\Big\}\\&=\mathbb{E}\Big\{\mathbb{E}\Big[\int_{\mathbb{R}}(F_Y(z)-\mathbf{1}\{z\ge Y\})^2-\big(F_{Y|X}(z)-\mathbf{1}\{z\ge Y\}\big)^2dz|X\Big]\Big\}\\&=\mathbb{E}\Big\{\mathbb{E}\Big[\int_{\mathbb{R}}\big(F^2_Y(z)-F^2_{Y|X}(z)-2\big(F_Y(z)-F_{Y|X}(z)\big)\cdot\mathbf{1}\{z\ge Y\}\big)dz|X\Big]\Big\}\\&=\mathbb{E}\Big\{\int_{\mathbb{R}}\big(F^2_Y(z)-F^2_{Y|X}(z)-2\big(F_Y(z)-F_{Y|X}(z)\big)\cdot F_{Y|X}(z)\big)dz\Big\}\\&=\mathbb{E}\Big\{\int_{\mathbb{R}}\big(F_Y(z)-F_{Y|X}(z)\big)^2dz\Big\}.\end{aligned}$$

Conditions for information value to be well-defined

We consider proper scoring rules $S:\mathbb{R}\times\mathcal{A}\to\mathbb{R}$ and suppose there is potential information available that is uncertain and represented by a measurable random quantity X with possible values in a set $\mathcal{X}$. We seek conditions under which the random quantity $\max_{a\in\mathcal{A}}\mathbb{E}[S(Y,a)|X]$ is measurable, in order to guarantee that information value ε^S_X in (8) is well-defined. We do so by showing that under suitable topologies, the function $x\mapsto\max_{a\in\mathcal{A}}\mathbb{E}[S(Y,a)|X=x]$ is continuous, and invoking the fact that every continuous function of a measurable function is measurable under the corresponding Borel σ-algebra.

The following two results derive immediately from results in Berge (1963, Chapter VI.3), specialized to the case in which the nonempty constraint set $\mathcal{A}$ does not depend on $x\in\mathcal{X}$, as is the case here.

1. If φ is a real-valued lower semi-continuous function in $\mathcal{X}\times\mathcal{A}$ then the function $x\mapsto\sup_{a\in\mathcal{A}}\varphi(x,a)$ is lower semi-continuous.
2. If φ is a real-valued upper semi-continuous function in $\mathcal{X}\times\mathcal{A}$ then the function $x\mapsto\max_{a\in\mathcal{A}}\varphi(x,a)$ is defined and upper semi-continuous.

Noting now that real-valued continuous functions are both lower and upper semi-continuous, we conclude:

3. If φ is a real-valued continuous function in $\mathcal{X}\times\mathcal{A}$ and $\max_{a\in\mathcal{A}}\varphi(x,a)$ exists, then the function $x\mapsto\max_{a\in\mathcal{A}}\varphi(x,a)$ is continuous.

These results hold under arbitrary topologies for $\mathcal{X}\times\mathcal{A}$ and the usual topology for the reals. Let the topology on $\mathcal{X}$ remain arbitrary, let the topology on $\mathcal{A}$ be the discrete topology, and let the topology on $\mathcal{X}\times\mathcal{A}$ be the product of these two topologies.

Lemma 18. *If $\mathcal{A}$ has the discrete topology, a function $f:\mathcal{X}\times\mathcal{A}\to\mathbb{R}$ is continuous if and only if for each $a\in\mathcal{A}$ the function $f_a:\mathcal{X}\to\mathbb{R}$ given by $f_a(x)=f(x,a)$ is continuous.*

Proof. In the following, the invoked topological properties may be found in Dugundji (1966) and other standard sources. Suppose that f is continuous. We will show that f_a is continuous for each $a \in \mathcal{A}$. We use the fact that the projection $p_X(x, a) = x$ is a continuous open map from $\mathcal{X} \times \mathcal{A}$ onto $\mathcal{X}$. Let $V \subset \mathbb{R}$ be an open set, and suppose $x \in f_a^{-1}(V)$, that is, $f(x, a) \in V$. Then because f is continuous there is an open $U \subset \mathcal{X} \times \mathcal{A}$ containing (x, a) such that $f(U) \subset V$. In particular, $(y, a) \in U \Rightarrow f(y, a) \in V$; that is, $(y, a) \in U \cap (\mathcal{X} \times \{a\}) \Rightarrow f(y, a) \in V$; that is, $y \in p_X(U) \Rightarrow f(y, a) \in V \Rightarrow f_a(y) \in V$; that is, $p_X(U) \subset f_a^{-1}(V)$. Note that $p_X(U)$ is an open set because U is open and p_X is an open map. So supposing that $x \in f_a^{-1}(V)$, we have shown that there is an open $U' = p_X(U)$ containing x such that $U' \subset f_a^{-1}(V)$. Therefore $f_a^{-1}(V)$ is an open set for every open $V \subset \mathbb{R}$. This means that f_a is continuous.

Conversely, suppose that f_a is continuous for each $a \in A$. We will show that f is continuous. We use the fact that $U \times \{a\} \subset \mathcal{X} \times \mathcal{A}$ is open for any open $U \subset \mathcal{X}$. Let $V \subset \mathcal{X}$ be an open set, and suppose that $(x, a) \in f^{-1}(V)$, that is $f(x, a) \in V$, that is, $f_a(x) \in V$, that is, $x \in f_a^{-1}(V)$. Then by continuity of f_a, there is an open set $U \subset \mathcal{X}$ containing x such that $f_a(U) \subset V$, that is, $f(U, a) \subset V$, that is, $f(U \times \{a\}) \subset V$. So we have shown that if $(x, a) \in f^{-1}(V)$, then there is an open set $U' = U \times \{a\}$ such that $f(U') \subset V$. This means that f is continuous. □

By this lemma applied to point 3 above, we conclude

4. If φ is a real-valued function on $\mathcal{X} \times \mathcal{A}$ that is continuous in x for each $a \in \mathcal{A}$, and $\max_{a \in \mathcal{A}} \varphi(x, a)$ exists, then the function $x \mapsto \max_{a \in \mathcal{A}} \varphi(x, a)$ is continuous.

Applying point 4 to $\varphi : \mathcal{X} \times \mathcal{A} \to \mathbb{R},\ (x, a) \mapsto \mathbb{E}[S(Y, a)|X = x]$, we can establish the following needed result. Recall we are considering only scoring rules S whose conditional expectation $\mathbb{E}[S(Y, a)|X]$ and maximum conditional expectation $\max_{a \in \mathcal{A}} \mathbb{E}[S(Y, a)|X]$ exist for all information vectors X.

Theorem 19. *If $S : \mathbb{R} \times \mathcal{A} \to \mathbb{R}$ is a proper scoring rule and the function $x \mapsto \mathbb{E}[S(Y, a)|X = x]$ from $\mathcal{X}$ to the reals is continuous in x for each $a \in \mathcal{A}$, then the random variable*

$$\max_{a \in \mathcal{A}} \mathbb{E}[S(Y, a)|X]$$

is well-defined and also P-measurable.

Proof. Because $\varphi(x, a)$ is continuous in x for each $a \in \mathcal{A}$, point 4 implies that the function $x \mapsto \mathbb{E}[S(Y, a)|X = x]$ is continuous. A continuous function of the P-measurable function $\omega \mapsto X(\omega)$ is P-measurable in the Borel σ-algebra generated from the topology on $\mathcal{X}$, so the random variable $\max_{a \in \mathcal{A}} \mathbb{E}[S(Y, a)|X]$ is P-measurable, as desired. □

References

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9*, 203–228.
Avriel, M., & Williams, A. (1970). The value of information in stochastic programming. *Operations Research, 18*(5), 947–954.
Baucells, M., & Borgonovo, E. (2013). Invariant probabilistic sensitivity analysis. *Management Science, 59*(11), 2536–2549.
Berge, C. (1963). *Topological Spaces*. New York, NY, USA: The Macmillan Company.
Bernardo, J., & Smith, A. (2000). *Bayesian Theory*. New York, NY, USA, second edition: Wiley&Sons.
Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering and System Safety, 92*(6), 771–784.
Borgonovo, E., Hazen, G., & Plischke, E. (2016). A common rationale for global sensitivity measures and their estimation. *Risk Analysis, 36*(10), 1871–1895.
Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research, 3*(1), 869–887.
Borgonovo, E., & Smith, C. (2011). A study of interactions in the risk assessment of complex engineering systems: An application to space PSA. *Operations Research, 59*(6), 1461–1476.
Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1–3.
Browne, T., Fort, J. C., Iooss, B., & Le Gratiet, L. (2017). Estimate of quantile-oriented sensitivity indices. *HAL Repository, (Work in Progress)*, 1–24.
Cervera, J., & Muñoz, J. (1996). Proper scoring rules for fractiles. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian statistics: 5* (pp. 513–519). Oxford, UK.: Oxford University Press.
Critchfield, G. G., & Willard, K. E. (1986). Probabilistic analysis of decision trees using monte carlo simulation. *Medical Decision Making, 6*(2), 85–92.
Dawid, A. (2007). The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics, 59*(1), 77–93.
Dezfuli, H., Benjamin, A., Everett, C., Maggio, G., Stamatelatos, M., & Youngblood, R. (2011). NASA risk management handbook. *NASA SP 2011 3422*.
Diebold, F. X., & Shin, M. (2017). Assessing point forecast accuracy by stochastic error distance. *Econometric Reviews, 36*(6–9), 588–598.
Dugundji, J. (1966). *Topology*. Boston, MA, USA: Allyn and Bacon.
Ebrahimi, N., Jalali, N. Y., & Soofi, E. S. (2014). Comparison, utility, and partition of dependence under absolutely continuous and singular distributions. *Journal of Multivariate Analysis, 131*, 32–50.
Ebrahimi, N., Soofi, E. S., & Soyer, R. (2010). Information measures in perspective. *International Statistical Review, 78*(3), 383–412.
Eschenbach, T. G. (1992). Spiderplots versus tornado diagrams for sensitivity analysis. *Interfaces, 22*(6), 40–46.
Felli, J. C., & Hazen, G. B. (1998). Sensitivity analysis and the expected value of perfect information. *Medical Decision Making, 18*, 95–109.
Felli, J. C., & Hazen, G. B. (1999). A bayesian approach to sensitivity analysis. *Health Economics, 8*, 263–268.
Fermanian, J.-D., & Scaillet, O. (2005). Sensitivity analysis of var and expected shortfall for portfolios under netting agreements. *Journal of Banking & Finance, 29*(4), 927–958.
Fort, J. C., Klein, T., & Rachdi, N. (2016). New sensitivity analysis subordinated to a contrast. *Communications in Statistics - Theory and Methods, 45*(15), 4349–4364.
French, S., & Argyris, N. (2018). Decision analysis and political processes. *Decision Analysis, 15*(4), 208–222.
Gamboa, F., Klein, T., & Lagnoux, A. (2018). Sensitivity analysis based on cramer von mises distance. *SIAM/ASA Journal on Uncertainty Quantification, 6*(2), 522–548.
Garthwaite, P., Kadane, J., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association, 100*(470), 680–700.
Gilboa, I., & Lehrer, E. (1991). The value of information: an axiomatic approach. *Journal of Mathematical Economics, 20*, 443–459.
Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association, 106*(74), 746–762.
Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annnual Reviews of Statistics and Its Applications, 1*, 25–51.
Gneiting, T., & Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association, 102*(477), 359–377.
Gourieroux, C., Laurent, J., & Scaillet, O. (2000). Sensitivity analysis of values at risk. *Journal of Empirical Finance, 7*(3–4), 225–245.
Hazen, G., & Sounderpandian, J. (1999). Lottery acquisition versus information acquisition: Prices and preference reversals. *Journal of Risk and Uncertainty, 18*(2), 125–136.
Heath, A., Manolopoulou, I., & Baio, G. (2017). A review of methods for analysis of the expected value of information. *Medical Decision Making, 37*(7), 747–758.
Hilton, R. (1981). The determinants of information value: synthesizing some general results. *Management Science, 27*(1), 57–64.
Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety, 52*(1), 1–17.
Howard, R. A. (1966). Decision analysis: Applied decision theory. In *Proceedings of the international conference operations research*. New York: Wiley-Interscience.
Howard, R. A. (1988). Decision analysis: Practice and promise. *Management Science, 34*(6), 679–695.
Iman, R., & Hora, S. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis, 10*, 401–406.
Jose, V., Nau, R., & Winkler, R. (2008). Scoring rules, generalized entropy, and utility maximization. *Operations Research, 56*(5), 1146–1157.
Jose, V., Nau, R., & Winkler, R. (2009). Sensitivity to distance and baseline distributions in forecast evaluation. *Management Science, 55*(4), 582–590.
Kao, C., & Steuer, R. E. (2016). Value of information in portfolio selection, with a taiwan stock market application illustration. *European Journal of Operational Research, 253*(2), 418–427.
Keisler, J., Collier, Z., Chu, E., Sinatra, N., & Linkov, I. (2014). Value of information analysis: The state of application. *Environment Systems and Decisions, 34*(1), 3–23.
Ketzenberg, M. E. (2009). The value of information in a capacitated closed loop supply chain. *European Journal of Operational Research, 198*(2), 491–503.
Ketzenberg, M. E., Rosenzweig, E. D., Marucheck, A. E., & Metters, R. D. (2007). A framework for the value of information in inventory replenishment. *European Journal of Operational Research, 182*(3), 1230–1250.
Kullback, S., & Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*(1), 79–86.
Lyons, R. (2013). Distance covariance in metric spaces. *Annals of Probability, 41*(5), 3284–3305.
Mehrez, A. (1985). A note on the analysis of the expected value of perfect information with respect to a class of r and d projects. *European Journal of Operational Research, 19*(2), 217–221.
NASA (2005). NASA's exploration systems architecture study. Technical report, nasa Washington, DC.
Oakley, J. (2009). Decision-theoretic sensitivity analysis for complex computer models. *Technometrics, 51*(2), 121–129.
Pfeifer, C., Schredelseker, K., & Seeber, G. U. H. (2009). On the negative value of information in informationally inefficient markets: Calculations for large number of traders. *European Journal of Operational Research, 195*(1), 117–126.

Pflug, G. (2006). A value-of-information approach to measuring risk in multi-period economic activity. *Journal of Banking & Finance, 30*(2), 695–715.
Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling and Software, 67*, 1–11.
Pichler, A., & Schlotter, R. (2019). Entropy based risk measures. *European Journal of Operational Research*. (to appear)
Plischke, E., & Borgonovo, E. (2019). Copula theory and probabilistic sensitivity analysis: Is there a connection? *European Journal of Operational Research, 277*(3), 1046–1059.
Rahman, S. (2016). The f-sensitivity index. *SIAM/ASA Journal on Uncertainty Quantification, 4*(1), 130–162.
Raiffa, H., & Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Boston: Harvard University Press.
Renyi, A. (1959). On measures of statistical dependence. *Acta Mathematica Academiae Scientiarum Hungarica, 10*, 441–451.
Saltelli, A., & Tarantola, S. (2002). On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. *Journal of the American Statistical Association, 97*(459), 702–709.
Samson, D., Wirth, A., & Rickard, J. (1989). The value of information from multiple sources of uncertainty in decision analysis. *European Journal of Operational Research, 39*(3), 254–260.
Savage, L. (1971). *Elicitation of Personal Probabilities and Expectations, 66*(336), 783–801.
Sejdinovic, D., Sriperumbudur, B., Gretton, A., & Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics, 41*(5), 2263–2291.
Shoja, M., & Soofi, E. S. (2017). Uncertainty, information, and disagreement of economic forecasters. *Econometric Reviews, 36*(6–9), 796–817.
Stehfest, E., van Zeist, W. J., Valin, H., Havlik, P., Popp, A., Kyle, P., … Wiebe, K. (2019). Key determinants of global land-use projections. *Nature Communications, 10*(1).
Stoer, J., & Witzgall, C. (1970). *Convexity and optimization in finite dimensions I*. Springer.
Strong, M., Oakley, J., & Brennan, A. (2014). Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: A nonparametric regression approach. *Medical Decision-Making, 34*, 311–326.
Szekely, G., & Rizzo, M. (2017). The energy of data. *Annual Reviews, 4*, 447–479.
Thomson, W. (1979). Eliciting production possibilities from a well-informed manager. *Journal of Economic Theory, 20*(3), 360–380.
Wagner, H. M. (1995). Global sensitivity analysis. *Operations Research, 43*(6), 948–969.