A NOTE ON INFORMATION VALUE THEORY FOR EXPERIMENTS DEFINED IN EXTENSIVE FORM

Jean-Pierre Ponssard

June 1974

WP-74-21

A NOTE ON INFORMATION VALUE THEORY FOR EXPERIMENTS DEFINED IN EXTENSIVE FORM

by

Jean-Pierre Ponssard

International Institute for Applied Systems Analysis
Laxenburg, Austria

(On leave from the Centre d'Enseignement Supérieur du Management Public, 94112, Arcueil, and from Groupe de Gestion des Organisations École Polytechnique, 75005, Paris, France)

May 1974

A b s t r a c t

An experiment is defined as a random variable which may take some posterior probability distributions according to a marginal probability. Elementary properties of this definition with respect to information value theory are derived as well as their practical implications.

1. Introduction

The concept of the value of information is one of the cornerstones of Decision Analysis [H, R]. It is ordinarily presented as a consequence of Bayes' theorem. Now, experiments may indeed be presented in terms of conditional probabilities, thus the use of Bayes' theorem, or directly as a random variable which may take some posterior probability distributions according to a marginal probability. Equivalence between the two approaches have long been recognized in the statistics literature (see [B-G]) however the second approach does not seem to have attracted much theoretical attention from decision analysts in spite of some practical advantages (see example 1.4.3 in [R-S]).

The objective of the paper is to investigate some elementary properties of this second definition of experiments with respect to information value theory. The practical significence of these properties is also studied.

2. The Value of Information Revisited

2.1 Definitions

Let us first define what shall be referred to as the classical decision problem.

Definition 2.1.1 The classical decision problem, consists in the selection of an action among a set of feasible actions $A = \{a\}$ given a set of possible states of nature $S = \{s\}$, a prior probability distribution on S,

$$p_0 = \{p_0^s\}_{s\varepsilon S} (p_0^s > 0, \sum_{s\varepsilon S} p_0^s = 1)$$

and a utility function $u(\cdot,\cdot)$ defined on $A\times S$, with values on the real line. (A and S are assumed finite).

Experiments with respect to this classical problem may now be defined in two alternative ways.

Definition 2.1.2 An experiment E, defined in normal form, consists of a finite set of possible events $E = \{e\}$ and a matrix of conditional probabilities $Q = \{q_{es} = \text{Prob } \{e|s\}\}_{e\varepsilon E,\, s\varepsilon S}$.

Definition 2.1.3 An experiment E, defined in extensive form, consists of a finite set of possible events $E = \{e\}$, a set of posterior probability distributions on S, $\{p_e = \{p_e^s\}_{s\varepsilon S}\}_{e\varepsilon E}$, and a marginal probability distribution on E,

$$\lambda^* = \{\lambda_e^*\}_{e\varepsilon E} \qquad \lambda_e^* > 0, \sum_{e\varepsilon E} \lambda_e^* = 1) \quad ,$$

which satisfy for all $s\varepsilon S$,

$$\sum_{e\varepsilon E} \lambda_e^* p_e^s = p_0^s \quad .$$

Both definitions are equivalent in the sense that one may go from one to the other by means of Bayes' theorem.

A classical decision problem and an experiment for this problem generate what might be called a "derived problem" (see Chapter 6 in [S]), in which one is interested in selecting the best strategy, namely an action for each possible event. Comparing certainty equivalents in both problems and the cost of the experiment, one then decides whether or not to carry out the experiment. These practical considerations lead to the concept of the value of information.

$$\text{Let } P = \{p = \{p^s\}_{s \in S} | p^s \geq 0, \sum_{s \in S} p^s = 1\} \quad .$$

P represents the set of all probability distributions on S. For all $p \in P$, let $u^*(\cdot)$ be the maximal expected utility associated with the classical decision problem, that is:

$$\text{for all } p \in P, \qquad u^*(p) = \operatorname*{Max}_{a \in A} \sum_{s \in S} p^s u(a,s) \quad .$$

<u>Proposition 2.1.4</u> The <u>expected value of information</u>, EVI, associated with an experiment E defined in extensive form may be expressed as:

$$EVI(p_0) = \sum_{e \in E} \lambda_e^* u^*(p_e) - u^*(p_0) \quad .$$

<u>Proof</u>: This is a standard result in Decision Analysis. ||

Assuming a linear utility for money, the EVI may be interpreted as the maximal price at which one should be willing to buy the experiment.

2.2 Comparing Experiments Defined in Extensive Form

Denote by P_E the smallest convex subset of P which contains the vectors $\{p_e\}_{e\varepsilon E}$ and for any real valued continuous function f (·) on P, let $\underset{P_E}{\text{Cav}}$ f (·) be the minimal concave function* greater or equal to f (·) on P_E. Let EVI $(p_0|E)$ be the expected value of information associated with the classical decision problem and an experiment E defined in extensive form.

Proposition 2.2.1

$$\text{EVI }(p_0|E) \leq \underset{P_E}{\text{Cav}}\ u^*(p_0) - u^*(p_0) \quad .$$

Proof: Since u* (·) is a convex function on P as being the point wise maximum of a set of hyperplanes, its concavification depends only on the values taken on the boundary of P_E, the concavification of which in turns, depends only on the values taken on $\{p_e\}_{e\varepsilon E}$. Denote by

$$\Lambda = \{\lambda = \{\lambda_e\}_{e\varepsilon E} | \lambda_e \geq 0,\ \sum_{e\varepsilon E} \lambda_e = 1,\ \sum_{e\varepsilon E} \lambda_e p_e = p_0\} \quad ,$$

then there exists some $\lambda^0 \varepsilon \Lambda$ such that

$$\underset{P_E}{\text{Cav}}\ u^*(p_0) = \sum_{e\varepsilon E} \lambda_e^0\ u^*(p_e)$$

---

* g( ) is a concave function on P if and only if for all p_1 and p_2 in P and all $\lambda\varepsilon(0,1)$:

$$g(\lambda p_1 + (1-\lambda)p_2) \geq \lambda g(p_1) + (1-\lambda)g(p_2) \quad .$$

and for all $\lambda \varepsilon \Lambda$, $\sum_{e\varepsilon E} \lambda^{0}_{e} u^*(p_e) \geq \sum_{e \varepsilon E} \lambda_e u^*(p_e)$.

By definition 2.1.3 $\lambda^* \varepsilon \Lambda$, hence the proposition holds. ||

We shall now characterize the experiments for which (2.2.1) is in general an equality. Define the P-class of classical decision problems as all problems for which S and p_0 in P remain fixed whereas A and $u(\cdot;\cdot)$ are allowed to vary.

<u>Definition 2.2.2</u> The experiment E is said to be <u>efficient</u> if and only if (2.2.1) is an equality for all problems in the P-class.

Note that the definition is meaningful since in order to define an experiment associated with a classical decision problem we need only know S and p_0 that is, the P-class.

<u>Proposition 2.2.3</u> An experiment E, defined in extensive form, is efficient if and only if the vectors $\{p_e\}_{e\varepsilon E}$ are linearily independent.

<u>Proof</u>: Assume that E is inefficient then there exists a classical decision problem in the P-class such that (2.2.1) is a strict inequality. Hence the particular λ^0 in Λ defined in proposition 2.2.1 and λ^*, which is also in Λ, are different. Subtracting $\sum_{e\varepsilon E} \lambda^0_e p_e = p_0$ and $\sum_{e\varepsilon E} \lambda^*_e p_e = p_0$ we obtain a linear dependence relation between the $\{p_e\}_{e\varepsilon E}$.

Reciprocally, since p_0 belongs to the convex hull of $\{p_e\}_{e\varepsilon E}$ it may be expressed as a convex combination of linearily independent vectors $\{p_e\}_{e\varepsilon E}$ (using Caratheodory's theorem), so that if the set $\{p_e\}_{e\varepsilon E}$ is linearily dependent Λ, contains at least two points. It is now a simple matter to construct a classical decision problem for which (2.2.1) is a strict inequality. ||

Corollary 2.2.4 An experiment defined in extensive form, is inefficient if and only if at least one of the following conditions hold

(i) there exists some $e_1 \varepsilon E$ such that $p_{e_1} \varepsilon P_{E - \{e_1\}}$,

(ii) there are more points in E than in S.

Proof: This is an immediate equivalence of the linear dependency of the vectors $\{p_e\}_{e \varepsilon E}$. ||

A typical illustration of the first condition is the case in which for some $e_1 \varepsilon E$, $p_{e_1} = p_0$. Then it is intuitive that the experiment is inefficient since we may very well end up with the same posterior probability distribution as our prior distribution. If p_{e_1} is not too different from p_0 then the experiment will remain inefficient. How close it has to be for inefficiency is made precise by the corollary.

The second condition is more difficult to interpret, essentially it is a question of dimensionality brought in by the finiteness of the set S.

Eventually, experiments should be compared in terms of EVI's. This comparison is easily facilitated for efficient experiments since then they may be partially ordered independently of the particular decision problem in the P-class.

Definition 2.2.5 An experiment E_1 is said to be more informative than an experiment E_2 if and only if for all problems in the P-class,

$$EVI(p_0|E_1) \geq EVI(p_0|E_2).$$

Proposition 2.2.6 For an efficient experiment E_1 to be more informative than an experiment E_2, a necessary and sufficient condition is that $P_{E_2} \subset P_{E_1}$.

Proof: As a simple property of the Cav operator, $P_{E_2} \subset P_{E_1}$ is equivalent to

$$\underset{P_{E_2}}{\text{Cav}}\, f(p_0) \leq \underset{P_{E_1}}{\text{Cav}}\, f(p_0)$$

for all convex functions $f(\cdot)$ on P. Since (2.3.1) is an equality for efficient experiments the proposition follows. ||

We shall conclude this section showing how the comparison of experiments in extensive form is related to their comparison in normal form. The parallel of this presentation with Blackwell and Girshick's study on the subject [B-G] will become apparent.

Proposition 2.2.7 For any experiment E, the vectors $\{p_e\}_{e\varepsilon E}$ are linearly independent if and only if the vectors $(q_e = (q_{es})_{e\varepsilon S})_{e\varepsilon E}$ are linearly independent.

Proof: Denote by R the matrix $\{p_e^s\}_{e\varepsilon E, s\varepsilon S}$ and by T the matrix $\{t_e^s\}_{e\varepsilon E, s\varepsilon S}$ in which $t_e^s = p_e^s/\lambda_e$ for all (e,s) in E×S.

According to Bayes theorem $q_{es} = t_e^s p_0^e$. Since for all (e,s) in E×S $\lambda_e^* > 0$ and $p_0^s > 0$, the vectors $\{p_e\}_{e \epsilon E}$ are independent if and only if the vectors $\{t_e\}_{e \epsilon E}$ are independent and the vectors $\{t_e\}_{e \epsilon E}$ are independent if and only if the vectors $\{q_e\}_{e \epsilon E}$ are independent. ||

We may thus replace the set $\{p_e\}_{e \epsilon E}$ by the set $\{q_e\}_{e \epsilon E}$ in our development. In particular we obtain that an experiment E_1 is more informative than an experiment E_2 if the vectors $\{q_e\}_{e \epsilon E_2}$ are linearly dependent on the vectors $\{q_e\}_{e \epsilon E_1}$. This result was derived directly by Blackwell and Girshick for experiments in normal form, hence the equivalence of the two approaches.

3. Practical Implications

The study of experiments in extensive fors lead us to the derivation of some elementary properties. These properties may now be used to somewhat simplify the decision analysis of practical situations in the following way:

(i) if one has to select one and only one experiment from a given set of equally costly experiments then proposition 2.2.6 may be used as a dominance criterium (see section 6-4 in [S] for general comments on the subject) to delete less informative experiments,

(ii) if one has to design an experiment then efficient experiments have clearly some advantages (in principle one may "redesign" an inefficient experiment so as to obtain an efficient one by modifying the

the marginal probabilities), then corollary 2.2.4 offers guidelines; moreover the marginal probability distribution need not be specified for efficient experiments since it is uniquely determined by the requirement

$$\sum_{e \in E} \lambda^*_e p_e = p_0 \quad ;$$

(iii) if one has to evaluate an inefficient experiment then proposition 2.2.1 gives an upperbound for the EVI (in this sense it is an improvement over the well known inequality EVI $\leq$ EVPI (perfect information), this upperbound may be derived with less computation than the EVI : the branches such that $p_e \varepsilon P_{E - \{e\}}$ need not be evaluated, for instance in the example 1.4.3 in [R-S] the experiment is inefficient since $p_{z_2} \varepsilon P_{\{z_1, z_3\}}$, it may be seen that

$$\max_{\lambda \varepsilon \Lambda} \Sigma \lambda_z u^*(p_z) = \frac{5}{6} u^*(p_{z_1}) + \frac{1}{6} u^*(p_{z_3}) = 35.83$$

whereas the actual EVI is 25.25 and EVPI is 70); then the knowledge of an upperbound for the EVI may enable the analyst to cut off some branch in a decision tree.

A c k n o w l e d g e m e n t s

The author wishes to express his thanks to Ralph Keeney, Howard Raiffa and Robert Winkler for many helpful discussions on a preliminary version of the paper.

References

[B-G] Blackwell, D. and Girshick, M.A. "Theory of Games and Statistical Decisions." Wiley, 1954.

[H] Howard, R.A. "Information Value Theory." IEEE Trans. on Systems Science and Cybernetics, Vol. SSC-4, June 1967, pp. 54-60.

[R] Raiffa, H. "Decision Analysis, Introductory Lectures." Addison-Wesley, 1967.

[R-S] Raiffa, H. and Schlaifer, R. "Applied Statistical Decision Theory." The M.I.T. Press, April 1968.

[S] Savage, L.J. "Foundations of Statistics." Wiley, 1954.