Knowledge-Based Systems 66 (2014) 92–98

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Probability weighted means as surrogates for stochastic dominance in decision making

Ronald R. Yager [a,b,*], Naif Alajlan [c]

[a] Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States
[b] Visiting Distinguished Scientist, King Saud University, Riyadh, Saudi Arabia
[c] ALISR Laboratory, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form 14 April 2014
Accepted 15 April 2014
Available online 26 April 2014

Keywords:
Uncertainty
Decision making
Stochastic dominance
Means
OWA operators

ABSTRACT

We discuss the role of stochastic dominance as tool for comparing uncertain payoff alternatives. However, we note the fact that this is a very strong condition and in most cases a stochastic dominance relationship does not exist between alternatives. This requires us to consider the use of surrogates for stochastic dominance to compare alternatives. Here we consider a class of surrogates that are called Probability Weighted Means (PWM). These surrogates are numeric values associated with an uncertain alternative and as such comparisons can be based on these values. The PWM are consistent with stochastic dominance in the sense that if alternative A stochastically dominated alternative B then its PWM value is larger. We look at a number of different examples of probability weighted means.

1. Introduction

Decision making in situations in which there is a probabilistic uncertainty associated with the payoff that results from the selection of an alternative is a very common task. Here each alternative is characterized by an uncertain payoff profile, a probability distribution over possible payoffs. A crucial problem here is the selection of a preferred alternative from a set of possible alternatives. While the objective is clear, select the alternative that gives the biggest payoff, the comparison of these uncertainty profiles with regard to this objective is difficult. One well-regarded method for comparing two uncertainty profiles is via the idea of stochastic dominance [1–3]. Essentially alternative A stochastically dominates alternative B if for any payoff value x alternative A has a higher probability of resulting in a payoff greater then or equal x then does alternative B. While providing an intuitively reasonable paradigm for deciding which of two alternatives is preferred, stochastic dominance is a strong condition and generally a stochastic dominance relationship between two alternatives does not exist, neither one stochastically dominates the other. In order to provide operational decision tools we look for surrogates for stochastic dominance. These surrogates associate with each alternative a numeric value, the larger the value the more preferred, and hence always allows comparison between alternatives. An important feature of these surrogates is their consistency with stochastic dominance in the sense that if A stochastically dominates B then the surrogate value of A is larger then the surrogate value of B. Here we consider a class of surrogates that we refer to as Probability Weighted Means (PWM).

2. Stochastic dominance and decision making

In decision making under uncertainty we are faced with the problem of selecting a preferred alternative from among a collection of alternatives based upon each alternative's payoff profile. Assume A_i is a decision alternative consisting of a set of possible payoffs that can result from the selection of this alternative, C_{ij} for $j = 1$ to n_i, and an associated uncertainty profile over this set of payoffs. Here we shall assume all the C_{ij} are numeric values. An important example of uncertainty profile associated with a decision alternative is a probabilistic uncertainty profile. Here each C_{ij} has a probability $p_{ij} > 0$.

A fundamental task here is to decide if alternative A_1 is preferred to alternative A_2, A_1 "$>$" A_2. One commonly used method for comparing alternatives is based upon the idea of **stochastic dominance** [1–10] with the understanding that if alternative A_1 stochastically dominates alternative A_2 then A_1 is the preferred alternative. We say that A_1 stochastically dominates A_2 if for all

* Corresponding author at: Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States.

E-mail addresses: yager@panix.com (R.R. Yager), najlan@ksu.edu.sa (N. Alajlan).

http://dx.doi.org/10.1016/j.knosys.2014.04.024
0950-7051/

values C, Prob($A_1 > C$) $\geqslant$ Prob($A_2 > C$) and there exists at least one value C^* such that Prob($A_1 > C^*$) > Prob($A_2 > C^*$). The intuition of stochastic dominance is essentially that we prefer alternatives that have a larger probability of resulting in bigger payoffs.

We observe that

$$\mathrm{Prob}(A_1 > C) = \sum_{j, C_{1j} > C} p_{1j} \text{ and } \mathrm{Prob}(A_2 > C) = \sum_{j, C_{2j} > C} p_{2j}$$

We further observe that since $1 = \sum_{j, C_{ij} \leqslant C} p_{ij} + \sum_{j, C_{ij} > C} p_{ij}$ then if Prob($A_1 > C$) $\geqslant$ Prob($A_2 > C$) it follows that

$$1 - \sum_{j, C_{1j} \leqslant C} p_{1j} \geqslant 1 - \sum_{j, C_{2j} \leqslant C} p_{2j}$$

and hence $\sum_{j, C_{2j} \leqslant C} p_{2j} \geqslant \sum_{j, C_{1j} > C} p_{1j}$. Recalling that $\sum_{j, C_{ij} \leqslant C} p_{ij}$ is the cumulative distribution function, $F_i(C)$, for A_i, $F_i(C) = \mathrm{Prob}(A_j \leqslant C)$. Thus alternative A_1 stochastically dominates A_2 if $F_1(C) \leqslant F_2(C)$ for all C and for at least one C^* we have $F_1(C^*) < F_2(C^*)$.

Formally we recall that a cumulative distribution function (CDF) associates with a probability distribution a mapping $F_i: R \to [0, 1]$ with the properties

(1) Monotonicity: $F_i(a) \geqslant F_i(b)$ for $a > b$
(2) $F_i(b) = 1$ for all $b \geqslant \mathrm{Max}_{j=1 \text{ to } n_i}[C_{ij}]$
(3) $F_i(a) = 0$ for all a $< \mathrm{Min}_{j=1 \text{ to } n_i}[C_{ij}]$

In the following for notational convenience we shall use "<", quoted <, to denote any relationship $G_1(x) \leqslant G_2(x)$ for all x and their exists at least one x^* such that $G_1(x^*) < G_2(x^*)$. Using this notation we have just indicated A_1 stochastically dominates A_2 if $F_1(x)$ "<" $F_2(x)$ for all $x \in R$, that is if Prob($A_1 \leqslant x$) "<" Prob($A_2 \leqslant x$) for all x.

In the following, without loss of generality, we shall use the following more convenient structure to investigate the issues of interest. Let $\mathbf{C} = \{C_j | j = 1 \text{ to } n\}$ be a collection of relevant numeric payoffs in a decision problem. A decision alternative A_i consists of a probability distribution such that p_{ij} is the probability of obtaining payoff C_j if we choose A_i. We note that if $p_{ij} = 0$ then C_j is not a possible payoff under A_i. Furthermore, we shall assume the C_j have been indexed in ascending order $C_{j+1} > C_j$. Using this notation we see that a cumulative distribution function is expressible as

$$P(A_i \leqslant x) = F_i(x) = \sum_{j \text{s.t.} C_j \leqslant x} p_{ij}$$

We see that for any $x < C_1$ we have $F_i(x) = 0$ and for any $x \geqslant C_n$ we have $F_i(x) = 1$.

We emphasize here that while $\mathbf{C}$ is a finite subset of real numbers, $F_i(x)$ is defined over the whole real line.

As we have earlier indicated we say that alternative A_1 stochastically dominates A_2, if $F_1(x)$ "<" $F_2(x)$ for all x, $F_1(x) \leqslant F_2(x)$ for all x and $F_1(x) < F_2(x)$ for at least one x. We shall refer to this as $A_1 >_{SD} A_2$. In Fig. 1 we show a typical example of stochastic dominance. Here F_1 stochastically dominates F_2.

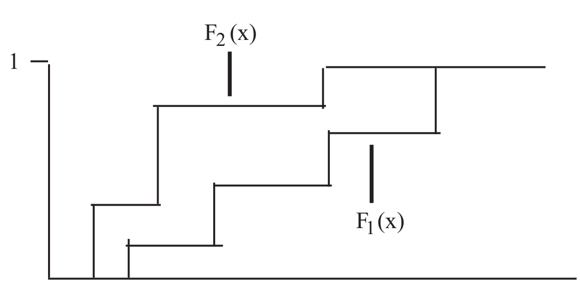

Fig. 1. Illustration of stochastic dominance.

As we have indicated if $A_j >_{SD} A_k$ then alternative A_j is preferred to A_k. Assume $\mathbf{A} = \{A_1, \ldots, A_q\}$ are a collection of alternatives then stochastic dominance can be viewed as a binary relationship on the space $\mathbf{A}$ [11]. Viewed as a binary relationship we see that it is transitive

$$A_1 >_{SD} A_2 \quad \text{and} \quad A_2 >_{SD} A_3 \quad A_1 >_{SD} A_3$$

From this transitivity it follows that if this relationship is complete, for each pair A_j and A_k either $A_j >_{SD} A_k$ or $A_k >_{SD} A_j$, then we can induce a linear ordering over the space A with respect to our preference of the alternative.

However, one important problem often arises with this agenda. The property of completeness is often lacking with respect to space $\mathbf{A}$. That is, there often exists pairs of alternatives, A_j and A_k such that neither $A_j >_{SD} A_k$ nor $A_k >_{SD} A_j$. This lack of completeness is a result of the fact that while stochastic dominance is a clear indication of preference between alternatives it is a relatively strong requirement and often does not exist between pairs of alternatives.

Because of this difficulty we must look for surrogates to stochastic dominance as way of comparing alternatives.

3. Probability weighted means

We shall now look at some properties that follow from stochastic dominance, $F_1(x)$ "<" $F_2(x)$ for all x.

We first recall the concept of median associated with a probability distribution. Given a probability distribution P over the order space, $\mathbf{C} = \{C_1, \ldots, C_n\}$ and its associated cumulative distribution function (CDF), F_i, then the median is the element C_{med} such that

$$F_i(C_{med-1}) < 0.5 \leqslant F_i(C_{med})$$

It is essentially the payoff where the CDF transitions from less the 0.5 to at least 0.5.

Assume A_1 and A_2 are two decision alternatives such that $A_1 >_{SD} A_2$, then $F_1(x) \leqslant F_2(x)$ for all x. Assume the median of A_1 occurs at $C_{med(1)}$, that is $F_1(C_{med(1)}) \geqslant 0.5$ then it is clear that $F_2(C_{med(1)}) \geqslant 0.5$. From this it follows that $C_{med(2)}$, the median of A_2 cannot occur at a value greater than $C_{med(1)}$. This allows us to conclude the following.

OBSERVATION: Assume A_1 and A_2 are such that $A_1 >_{SD} A_2$ then $Med(A_1) \geqslant Med(A_2)$

In Fig. 2 we clearly illustrate this relationship

We now shall look at the relationship between the expected value's of alternatives and their relationship with respect to stochastic dominance. First we shall look at the relationship between an alternative's expected value and its associated CDF. Consider a generic alternative A with probability p_j associated with payoff C_j where we have indexed the payoffs in increasing order, $C_{j+1} > C_j$. We recall its expected value is $E(A) = \sum_{j=1}^{n} p_j C_j$ and $F(C_j) = \mathrm{Prob}(A \leqslant C_j) = \sum_{i=1}^{j} P_i$. Here we shall by convention let $F(C_0) = 0 = \sum_{i=1}^{0} p_i$. We also note the $F(C_n) = 1$.

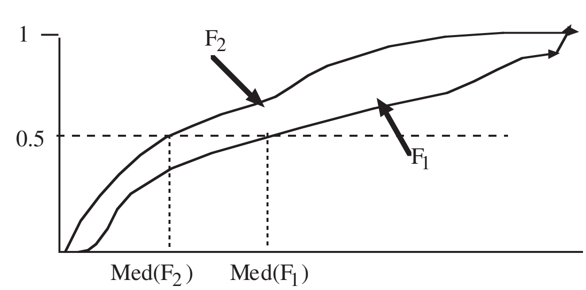

Fig. 2. Median calculation in case of stochastic dominance.