Cumulative Dominance and Heuristic Performance in Binary Multi-attribute Choice

Manel Baucells[*], Juan A. Carrasco[†], and Robin M. Hogarth[‡]

September 29, 2005

Abstract

Several studies have reported high performance of simple decision heuristics in multi-attribute decision making. In this paper, we focus on situations where attributes are binary and analyze the performance of Deterministic-Elimination-By-Aspects (DEBA) and similar decision heuristics. We consider non-increasing weights and two probabilistic models for the attribute values: one where attribute values are independent Bernoulli random variables; the other one where they are binary random variables with inter-attribute positive correlations. Using these models, we show that good performance of DEBA is explained by the presence of cumulative as opposed to simple dominance. We therefore introduce the concepts of cumulative dominance compliance and fully cumulative dominance compliance and show that DEBA satisfies those properties. We derive a lower bound with which cumulative dominance compliant heuristics will choose a best alternative and show that, even with many attributes, this is not small. We also derive an upper bound for the expected loss of fully cumulative compliance heuristics and show that this is moderate even when the number of attributes is large. Both bounds are independent of the values of the weights.

Keywords: Multi-attribute decision making. Binary attributes. DEBA. Cumulative dominance. Performance bounds.

[*]IESE Business School, Universidad de Navarra, Avda. Pearson 21, 08034 Barcelona, Spain, mbaucells@iese.edu

[†]Departament d'Enginyeria Electrònica, Universitat Politècnica de Catalunya, Diagonal 647, plta. 9, 08028 Barcelona,Spain, carrasco@eel.upc.edu

[‡]ICREA & Universitat Pompeu Fabra,Department of Economics and Business, Ramón Trias Fargas, 25–27, 08005 Barcelona, Spain, robin.hogarth@upf.edu

1 Introduction

We consider a standard multi-attribute choice problem having m alternatives i, $1 \leq i \leq m$, each characterized by k attributes $x_{i,r}$, $1 \leq r \leq k$. The utility of the $i\text{th}$ alternative, $x_i = (x_{i,1}, x_{i,2}, ..., x_{i,k})$, is defined as

$$U_i = w_1 x_{i,1} + w_2 x_{i,2} + ... + w_k x_{i,k} \, , \tag{1}$$

where the w_r are positive weighting parameters subject to the constraint $w_1+w_2+...+w_k = 1$. The problem is to identify which of the m alternatives is best, i.e., has the largest value of U_i. This is a classical decision problem (cf. Keeney and Raiffa 1993). We make the assumption that the decision maker can order the weights by size such that, without loss of generality, $w_1 \geq w_2 \geq ... \geq w_k \geq 0$ but that the exact values of the weights are unknown. This assumption is realistic in many scenarios. Consider, for instance, a situation in which a committee has to choose one of several candidates to fill a job opening. Typically, members of the committee will agree on which attributes of the candidates are relevant and may easily agree to take the decision using a linear utility function where each attribute is given a positive weight. Moreover, whereas committee members might disagree as to what values should be given to the weights, they can agree on their relative importance.

Since the exact values of the weights unknown, a reasonable approach is to use a heuristic. In this paper, we will obtain results regarding the performance of a class of heuristics to solve this decision problem. We will make the assumption that the $x_{i,r}$ are, non-necessarily independent, random variables with support $[0, 1]$. While some of our results are general and do not require additional assumptions on $x_{i,r}$, most assume that the $x_{i,r}$ are binary random variables taking only the values 0 and 1. That more particular setting has interest on its own. For example, it is common to have alternative features that are either present or absent (e.g., the candidate has good knowledge or not of a given foreign language), or that take two values (e.g., the candidate is male or female). Even if the attribute is multi-valued, the decision-maker may be able to distinguish between zero and non-zero values, but be insensitive to the actual magnitude of the attribute (Hsee and Rottenstreich 2004). Also, in order to simplify the decision, the decision-maker may use a cut-off to partition the range between high and low regions. Here, several choices are available depending on the cutoff values chosen to separate between high ($x_{i,r} = 1$) and low ($x_{i,r} = 0$) values. One could use a low cutoff representing a minimum acceptable level. Alternatively, one could assign a value 1 only to those attribute values with the best level on that attribute. Those two choices yield, respectively, the *LEX* and the *EBA* heuristics discussed by Payne et al. (1993).

A possible decision rule we will consider would make use of the attribute ordering in a lexicographic fashion. Specifically, at the first stage, alternatives with non greatest value in the first attribute would be eliminated (unless all alternatives had the same value for the first attribute). If a single alternative remains, it would be chosen. Otherwise, the values of the second attribute would be examined and alternatives with non greatest value in that second attribute would be eliminated. This procedure would continue until only one alternative remains or all attributes have been examined. If only one alternative remains, that alternative would be chosen. If several alternatives remain after all attributes have been examined, then the choice between them would be made at random. This model is a deterministic variant of the *EBA* (Elimination-By-Aspects) heuristic proposed by Tversky (1972). We therefore call it *DEBA* (Deterministic-Elimination-By-Aspects). It differs from *EBA* in that the attributes (aspects) used to eliminate alternatives at each stage of the process are selected by a deterministic as opposed to a probabilistic procedure. As a procedure, *DEBA* generalizes —to more than two alternatives— the lexicographic binary-choice model Take-The-Best (*TTB*) proposed by Gigerenzer and Goldstein (1996). There is a small difference, however: in *TTB*, the attributes are ordered by their validities, which are computed using a database of previous instances of alternatives, while in *DEBA* the ordering of the attributes by decreasing weights is assumed known.

The *DEBA* heuristic is easy to use. In many situations, for example, there is no need to look beyond the first one or first two attributes to make a decision. Several studies have shown *DEBA* to be effective in relation to alternative simple decision heuristics (Gigerenzer and Goldstein 1996; Czerlinski et al. 1999; Martignon and Hoffrage 1999, 2002) as well as having desirable properties for both binary and multivariate choice (Hogarth and Karelaia 2003; Katsikopoulos and Martignon 2003; Katsikopoulos and Fasolo, in press). Even when attributes are continuous variables, the model can be quite effective under some circumstances (Gigerenzer, Todd et al. 1999; Hogarth and Karelaia 2005a). Most of these studies are restricted to the case of two or three alternatives. Finally, there is empirical evidence that people do sometimes use *DEBA*-like strategies in decision making (see, e.g., Bröder 2000; Newell and Shanks 2003; Newell et al. 2003).

Our goal is to understand the observed good performance of *DEBA* and other related heuristics. The effectiveness of the decision heuristic can be measured using two metrics: 1) the probability that the heuristic will select a best alternative and, 2) the expected loss of the heuristic, i.e. the expected difference between the utility of a best alternative and the utility of the alternative chosen by the heuristic. The exact values of those metrics depend, of course, on the exact values of the weights w_r, $1 \leq r \leq k$, and on the probabilistic model underlying the values of the attributes $x_{i,r}$, $1 \leq i \leq m$, $1 \leq r \leq k$. We will explain the good performance of DEBA and other related heuristics by deriving a lower bound for the probability that the heuristic will choose a best alternative and an upper bound for the expected loss *independent of the weights*. Moreover, we show that, even with many attributes, the former is large and the latter small. This will be done for two probabilistic models for the attributes: one in which the attribute values $x_{i,r}$ are assumed to be binary independent Bernoulli random variables with a common parameter p and one in which the attributes $x_{i,r}$ are assumed to be binary random variables with positive inter-attribute correlation, i.e. in which the values of the attributes of a given alternative are positively correlated.

The use of the simple dominance concept is a first, trivial trial. An alternative i simply dominates alternative j if each attribute value of i is non-smaller than each attribute value of j. It is clear that, irrespective of the values of the weights and, therefore, not depending on the values of the weights being non-increasing, whenever an alternative simply dominates all other alternatives both that alternative will have the largest utility and *DEBA* will choose that alternative. Then, the probability that an alternative simply dominates all other alternatives provides a lower bound on the probability that *DEBA* will choose a best alternative. However, as we shall show, that probability can be very small when the number of attributes is large. Thus, simple dominance does not explain the observed good performance of *DEBA*.

The approach we will follow to justify theoretically the effectiveness of *DEBA* and other related heuristics is the use of the use of the concept of *cumulative dominance* (Kirkwood and Sarin 1985). An alternative i is said to cumulative dominate alternative j if the accumulated values of the attributes of i are non-smaller than the accumulated values of the attributes of j. To illustrate, consider alternatives $x_1 = (1, 0, 1)$ and $x_2 = (0, 1, 1)$. Then, alternative x_1 cumulative dominates alternative x_2 because $x_{1,1} \geq x_{2,1}$, $x_{1,1}+x_{1,2} \geq x_{2,1}+x_{2,2}$, and $x_{1,1}+x_{1,2}+x_{1,3} \geq x_{2,1}+x_{2,2}+x_{2,3}$. As we will show, since the weights are non-increasing, an alternative which cumulative dominates another alternative alternative necessarily has a non-smaller utility than the cumulative dominated alternative. We observe next that *DEBA* complies with cumulative dominance, i.e. in the event that some alternative cumulative dominates all other alternatives, *DEBA* is guaranteed to choose one of those alternatives. Then, the probability that some alternative cumulative dominates all other alternatives is a lower bound to the probability with which *DEBA* will choose a "best" alternative. Contrary to simple dominance, the probability that some alternative exhibits cumulative dominance over all other alternatives is not small even when the number of attributes is large. This provides a first justification of the observed good performance of *DEBA*. The approach we take to provide an upper bound for the expected loss of *DEBA* is to compute an upper bound for the loss of *DEBA* conditioned on the maximum attribute index for which some alternative cumulative dominates all others. That upper bound is computed using the fact that *DEBA* will necessarily choose one of the alternatives in the set of alternatives that cumulative dominate all other alternatives up to the highest possible attribute index, a property which is called *fully cumulative dominance compliance*. That upper bound does not depend on the attributes being binary: it only depends on the attribute values having support $[0, 1]$. Those upper bounds, combined with the computation of the probability distribution of the maximum attribute index for which some alternative cumulative dominates all others, allows the computation of an upper bound for the expected loss of *DEBA*. As the computation of the lower bound for the probability that *DEBA* will choose a best alternative, our computation of that probability distribution is particular for the assumed probabilistic models underlying the attribute values. We show that the upper bound for the expected loss remains reasonable even when the number of attributes is large, providing a second justification for the observed good performance of *DEBA*.

The performance justifications just exposed are not restricted to the *DEBA* heuristic. It applies as well to any heuristic that complies/fully complies with cumulative dominance. For instance, it applies (partially) to the *EWn/DEBA* heuristic, which is cumulative dominance compliant but not fully cumulative dominance compliant. The *EWn/DEBA* heuristic first chooses the alternatives with the highest total sum of attributes up to attribute n, and then breaks ties using *DEBA*. The results given in the paper regarding the performance of *DEBA* and any other cumulative/fully cumulative dominance compliant heuristics are, however, restricted to the assumed probabilistic models underlying the attribute values. It is an open problem to justify the good performance of *DEBA* and other cumulative/fully cumulative dominance compliant heuristics under other probabilistic models, in particular when the attributes are continuous random variables.

The rest of the paper is organized as follows. In Section 2 we define the two probabilistic models underlying the attribute values which will be used throughout the paper. In Section 3, we obtain, for the two probabilistic models under consideration, the probability of simple dominance and show that the presence of that kind of dominance does not justify the observed good performance of *DEBA*. In Section 4, we introduce the concepts of cumulative dominance compliance and fully cumulative dominance, show that *DEBA* satisfies both properties, give examples of other heuristics satisfying those properties, derive a lower bound for the probability that any cumulative dominance compliant heuristic will choose a best alternative, derive an upper bound for the expected loss in any fully cumulative dominance compliant heuristic, and using those metrics justify the observed good performance of *DEBA* and other related heuristics. Section 5 concludes the paper and highlights directions for future work.

2 Probabilistic Models

Two probabilistic models for the values of the attributes $x_{i,r}$, $1 \leq i \leq m$, $1 \leq r \leq k$ will be considered:

ZIAC (Zero Inter-Attribute Correlation) model: The $x_{i,r}$ are independent Bernoulli random variables with parameter p, $0 < p < 1$.

PIAC (Positive Inter-Attribute Correlation) model: The $x_{i,r}$ are obtained as $x_{i,r} = z_i y^h_{i,r} + (1 - z_i) y^l_{i,r}$, where the z_i, $y^h_{i,r}$, and $y^l_{i,r}$ are independent Bernoulli random variables with parameters p, $p_h = p + \sqrt{\rho}(1-p)$, and $p_l = p - \sqrt{\rho}p$, respectively, for some $0 < p < 1$ and some $0 \leq \rho < 1$.

The ZIAC model is a simple model without need for justification. We note that $E[x_{i,j}] = p$. Thus, the parameter p of the common Bernoulli distributions can be looked at as measuring the average quality of the attributes: higher values of p model attributes of higher average quality. The PIAC model is intuitively appealing: if there is positive correlation among the attributes of a given alternative, it is because there is some common cause shifting the average quality of the attributes of a given alternative. In the PIAC model, this is captured by the alternatives belonging to a "good" population (with averaged values for the attribute values equal to $p_h = p + \sqrt{\rho}(1-p)$) with probability p and to a "bad" population (with averaged values for the attribute values equal to $p_l = p - \sqrt{\rho}(1-p)$) with probability $1-p$. In the PIAC model $E[x_{i,j}] = p$ and the attribute values of any given alternative have positive correlation ρ. The ZIAC model can be seen as a particular case of the PIAC model with $\rho = 0$. Since $\sum_{r=1}^{k} w_r = 1$, in both models the expected value of the utility of any given alternative i is $E[U_i] = p$.

3 Simple Dominance does not justify the good performance of *DEBA*

An alternative i is said to exhibit *simple dominance* up to attribute r over alternative j, denoted by $d_r(i,j)$, if and only if $x_{i,s} \geq x_{j,s}$, $1 \leq s \leq r$. An alternative i is said to exhibit simple dominance over alternative j if and only if $d_k(i,j)$, i.e. if and only if alternative i exhibits simple dominance up to attribute k over alternative j. For $1 \leq r \leq k$, let D_r denote the set of alternatives that exhibit simple dominance over any other alternative up to attribute r, i.e.

$$D_r = \{1 \leq i \leq m \,:\, d_r(i,j), 1 \leq j \leq m\}. \tag{2}$$

Obviously, $D_1 \supset D_2 \supset \cdots \supset D_k$. Also all alternatives i in D_r have identical attribute profiles up to attribute r, $x_{i,1}, x_{i,2}, \ldots, x_{i,r}$. Since the weights are non-negative, any alternative i which exhibits simple dominance over another alternative j will have largest utility U_i than the utility U_j of j. Then, it is clear that when $D_k \neq \emptyset$ the alternatives in D_k, with identical attribute profiles, will be best. It is also clear that when $D_k \neq \emptyset$, *DEBA* will choose an alternative from D_k. Then, when $D_k \neq \emptyset$, *DEBA* will choose a best alternative and the probability $[P_B]_{\text{lbs}} = P[D_k \neq \emptyset]$ will be a lower bound for the probability with which *DEBA* will choose a best alternative. In this section we will develop efficient computational procedures for $[P_B]_{\text{lbs}}$ for the two probabilistic models under consideration. Using these computational procedures, we will compute $[P_B]_{\text{lbs}}$ for a wide range of model parameters and will discuss the extent to which the presence of simple dominance is able to explain the observed good performance of *DEBA*.

We will start by deriving an efficient computational scheme for $[P_B]_{\text{lbs}}$ for the ZIAC model. Consider the discrete-parameter stochastic process with truncated parameter $Y = \{Y_r; 0 \leq r \leq k\}$ with state-space $\{0, 1, \ldots, m\}$ defined by $Y_0 = m$ and $Y_r = |D_r|$, $1 \leq r \leq k$. The following theorem establishes that Y is a homogeneous discrete-parameter Markov chain (with truncated parameter) and gives its one-step transition probabilities. Figure 1 gives the state transition diagram of Y for the case $m = 3$.

Theorem 1. $Y = \{Y_r; 0 \leq r \leq k\}$ *is a homogeneous discrete-parameter Markov chain (with truncated parameter) with state space* $\{0, 1, \ldots, m\}$*, initial state* m*, and one-step transition probabilities* $Q_{i,j} = P[Y_{r+1} = j \mid Y_r = i]$ *given by:*

$$Q_{0,0} = 1\,,$$

$$Q_{0,j} = 0 \quad \textit{for } 1 \leq j \leq m\,,$$

$$Q_{i,0} = (1-p)^i[1-(1-p)^{m-i}] \quad \textit{for } 1 \leq i < m\,,$$

$$Q_{i,j} = \binom{i}{j} p^j (1-p)^{i-j} \quad \textit{for } 1 \leq i \leq m, 1 \leq j < i\,,$$

$$Q_{i,i} = p^i + (1-p)^m \quad \textit{for } 1 \leq i \leq m\,,$$

$$Q_{i,j} = 0 \quad \textit{for } 1 \leq i \leq m, i < j \leq m\,.$$

Proof. See the Appendix. □

Theorem 1 allows the numerical computation for the ZIAC model of $[P_B]_{\text{lbs}} = P[D_k \neq \emptyset] = \sum_{i=1}^{m} P[Y_k = i]$ using standard discrete-parameter Markov chain analysis techniques. However,

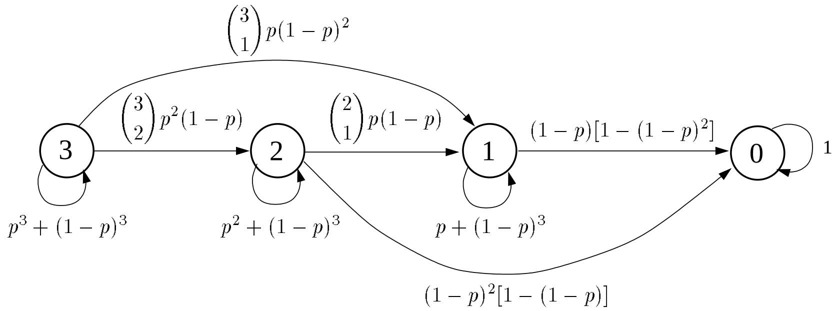

Figure 1: State transition diagram of Y for the case $m = 3$.

given the values of the one-step transition probabilities of Y, it is possible to obtain a simple closed-form expression for $[P_B]_{\text{lbs}}$. We start by deriving a closed-form expression for $P[Y_r = i]$, $1 \le r \le k$, $1 \le i \le m$:

Proposition 1. *For* $1 \le r \le k$ *and* $1 \le i \le m$*:*

$$P[Y_r = i] = \binom{m}{i} \sum_{j=i}^{m} \binom{m-i}{j-i} (-1)^{j-i} [p^j + (1-p)^m]^r \,.$$

Proof. We start by proving that the one-step transition probabilities $Q_{i,j}$ for $1 \le i \le m$, $1 \le j < i$ and $Q_{i,i}$, $1 \le i \le m$ given by Theorem 1 can be formulated in a more compact way as:

$$Q_{i,j} = \binom{i}{j} \sum_{l=j}^{i} \binom{i-j}{l-j} (-1)^{l-j} [p^l + (1-p)^m] \,, \quad 1 \le i \le m, 1 \le j \le i \,. \tag{3}$$

To make the proof, we rewrite the previous expression as:

$$\begin{aligned} &\binom{i}{j} \sum_{l=0}^{i-j} \binom{i-j}{l} (-1)^l [p^{j+l} + (1-p)^m] \\ &\quad = \binom{i}{j} p^j \sum_{l=0}^{i-j} \binom{i-j}{l} (-p)^l + \binom{i}{j} \sum_{l=0}^{i-j} \binom{i-j}{l} (-1)^l (1-p)^m \,. \end{aligned}$$

For $1 \le j < i$, the previous expression gives

$$\binom{i}{j} p^j (1-p)^{i-j} + \binom{i}{j} (1-1)^{i-j} (1-p)^m = \binom{i}{j} p^j (1-p)^{i-j} \,,$$

which is the expression for $Q_{i,j}$, $1 \le i \le m$, $1 \le j < i$ given by Theorem 1. For $j = i$, the expression gives

$$p^i + (1-p)^m \,,$$

which is the expression for $Q_{i,i}$, $1 \le i \le m$ given by Theorem 1.

Using (3), the proof of the proposition is by induction on r. For $r = 1$, using $Y_0 = m$ and (3), we obtain

$$P[Y_1 = i] = Q_{m,i} = \binom{m}{i} \sum_{j=i}^{m} \binom{m-i}{j-i} (-1)^{j-i} [p^j + (1-p)^m] \, ,$$

completing the base case. For the induction step, assume the result holds for $r = s \geq 1$ and let us prove the result for $r = s + 1$. Using Theorem 1, the induction step, (3), and the identity $\binom{m}{k}\binom{j}{i} = \binom{m}{i}\binom{m-i}{j-i}$:

$$\begin{aligned}
P[Y_{s+1} = i] &= \sum_{j=0}^{m} P[Y_s = j] \, Q_{j,i} = \sum_{j=i}^{m} P[Y_s = j] \, Q_{j,i} \\
&= \sum_{j=i}^{m} \left[\binom{m}{j} \sum_{l_1=j}^{m} \binom{m-j}{l_1-j} (-1)^{l_1-j} \, [p^{l_1} + (1-p)^m]^s \right] \\
&\qquad \left[\binom{j}{i} \sum_{l_2=i}^{j} \binom{j-i}{l_2-i} (-1)^{l_2-i} \, [p^{l_2} + (1-p)^m] \right] \\
&= \sum_{j=i}^{m} \binom{m}{j} \binom{j}{i} \left[\sum_{l_1=j}^{m} \binom{m-j}{l_1-j} (-1)^{l_1-j} \, [p^{l_1} + (1-p)^m]^s \right] \\
&\qquad \left[\sum_{l_2=i}^{j} \binom{j-i}{l_2-i} (-1)^{l_2-i} \, [p^{l_2} + (1-p)^m] \right] \\
&= \binom{m}{i} \sum_{j=i}^{m} \binom{m-i}{j-i} \left[\sum_{l_1=j}^{m} \binom{m-j}{l_1-j} (-1)^{l_1-j} \, [p^{l_1} + (1-p)^m]^s \right] \\
&\qquad \left[\sum_{l_2=i}^{j} \binom{j-i}{l_2-i} (-1)^{l_2-i} \, [p^{l_2} + (1-p)^m] \right] \, ,
\end{aligned}$$

which can be written as

$$P[Y_{s+1} = i] = \sum_{l_2=i}^{m} \sum_{l_1=l_2}^{m} C(l_1, l_2) [p^{l_1} + (1-p)^m]^s [p^{l_2} + (1-p)^m] \tag{4}$$

with

$$C(l_1, l_2) = \binom{m}{i} \sum_{j=l_2}^{l_1} \binom{m-i}{j-i} \binom{m-j}{l_1-j} \binom{j-i}{l_2-i} (-1)^{l_1-j} (-1)^{l_2-i} \, .$$

Using the identity $\binom{m-i}{j-i}\binom{m-j}{l_1-j}\binom{j-i}{l_2-i} = \binom{m-i}{l_1-i}\binom{l_1-i}{l_2-i}\binom{l_1-l_2}{j-l_2}$:

$$\begin{aligned}
C(l_1, l_2) &= \binom{m}{i} \binom{m-i}{l_1-i} \binom{l_1-i}{l_2-i} (-1)^{l_2-i} \sum_{j=l_2}^{l_1} \binom{l_1-l_2}{j-l_2} (-1)^{l_1-j} \\
&= \binom{m}{i} \binom{m-i}{l_1-i} \binom{l_1-i}{l_2-i} (-1)^{l_2-i} \sum_{j=0}^{l_1-l_2} \binom{l_1-l_2}{j} (-1)^{l_1-l_2-j} \, .
\end{aligned}$$

Then, we have

$$C(l_2, l_2) = \binom{m}{i}\binom{m-i}{l_2-i}(-1)^{l_2-i}$$

and, for $l_2 > l_1$,

$$C(l_1, l_2) = \binom{m}{i}\binom{m-i}{l_1-i}\binom{l_1-i}{l_2-i}(-1)^{l_2-i}(1-1)^{l_1-l_2} = 0\,.$$

Plugging those results into (4):

$$P[Y_{s+1} = i] = \binom{m}{i}\sum_{l_2=i}^{m}\binom{m-i}{l_2-i}(-1)^{l_2-i}[p^{l_2} + (1-p)^m]^{s+1}\,,$$

completing the induction step. □

The closed-form expression for $[P_B]_{\mathrm{lbs}}$ for the ZIAC model is given by the following theorem:

Theorem 2. *For the ZIAC model,*

$$[P_B]_{\mathrm{lbs}} = \sum_{i=1}^{m}\binom{m}{i}(-1)^{i-1}[p^i + (1-p)^m]^k\,.$$

Proof. Using $[P_B]_{\mathrm{lbs}} = \sum_{i=1}^{m} P[Y_k = i]$ and Proposition 1:

$$\begin{aligned}[P_B]_{\mathrm{lbs}} &= \sum_{i=1}^{m} P[Y_k = i] = \sum_{i=1}^{m}\binom{m}{i}\sum_{j=i}^{m}\binom{m-i}{j-i}(-1)^{j-i}[p^j + (1-p)^m]^k \\ &= \sum_{j=1}^{m}\sum_{i=1}^{j}\binom{m}{i}\binom{m-i}{j-i}(-1)^{j-i}[p^j + (1-p)^m]^k\,.\end{aligned}$$

Using the identity $\binom{m}{i}\binom{m-i}{j-i} = \binom{m}{j}\binom{j}{i}$:

$$\begin{aligned}[P_B]_{\mathrm{lbs}} &= \sum_{j=1}^{m}\binom{m}{j}[p^j + (1-p)^m]^k\sum_{i=1}^{j}\binom{j}{i}(-1)^{j-i} \\ &= \sum_{j=1}^{m}\binom{m}{j}[p^j + (1-p)^m]^k\left(\sum_{i=0}^{j}\binom{j}{i}(-1)^{j-i} - (-1)^j\right) \\ &= \sum_{j=1}^{m}\binom{m}{j}[p^j + (1-p)^m]^k((1-1)^j - (-1)^j) \\ &= \sum_{j=1}^{m}\binom{m}{j}(-1)^{j-1}[p^j + (1-p)^m]^k\,. \quad \square\end{aligned}$$

We will consider next the PIAC model. For that model we have not been able to derive a closed-form expression for $[P_B]_{\mathrm{lbs}}$ and will content ourselves with a recurrent computational scheme. Let

G be the subset of good alternatives (those whose attribute values are independent Bernoulli random variables with parameter p_h). Since each alternative is independently good with parameter p, the number of good alternatives $|G|$ has a binomial distribution with parameters m and p. Then, conditioning on $|G|$:

$$[P_B]_{\mathrm{lbs}} = P[D_k \neq \emptyset] = \sum_{g=0}^{m} \binom{m}{g} p^g (1-p)^{m-g} P[D_k \neq \emptyset \,|\, |G| = g] \,. \tag{5}$$

By symmetry, all $P[D_k \neq \emptyset | G = G']$, $|G'| = g$ are equal and, therefore, $P[D_k \neq \emptyset \,|\, |G| = g] = P[D_k \neq \emptyset | G = G']$, $|G'| = g$. Following ideas similar to the ones used for the ZIAC model we can formalize the computation of $P[D_k \neq \emptyset | G = G']$ in terms of the transient behavior of an homogeneous discrete-parameter Markov chain (with truncated parameter). Let

$$D_r^g = \{ i \in G \,:\, x_{i,s} \geq x_{j,s}, 1 \leq j \leq m, 1 \leq s \leq r \}$$

and

$$D_r^b = \{ i \in \{1, 2, \ldots, m\} - G \,:\, x_{i,s} \geq x_{j,s}, 1 \leq j \leq m, 1 \leq s \leq r \} \,,$$

i.e., D_r^g collects the good alternatives which exhibit simple dominance over any other alternative up to attribute r and D_r^b collects the bad alternatives which exhibit simple dominance over any other alternative up to attribute r. Given a set of good alternatives G, let $Y^G = \{ Y_r^G ; 0 \leq r \leq k \}$ be the discrete-parameter stochastic process (with truncated parameter) with state space $\{(i,j), 0 \leq i \leq |G|, 0 \leq j \leq m - |G|\}$ defined by $Y_0^G = (|G|, m - |G|)$ and $Y_r^G = (|D_r^g|, |D_r^b|)$, $1 \leq r \leq k$. The following theorem establishes that Y^G is a homogeneous discrete-parameter Markov chain (with truncated parameter) and gives its one-step transition probabilities. The proof of the Theorem is parallel to the proof of Theorem 1.

Theorem 3. $Y^G = \{ Y_r^G ; 0 \leq r \leq k \}$ *is a homogeneous discrete-parameter Markov chain (with truncated parameter) with state space* $\{(i,j), 0 \leq i \leq |G|, 0 \leq j \leq m - |G|\}$*, initial state* $Y_0^G = (|G|, m - |G|)$*, and one-step transition probabilities* $Q_{(i^g,i^b),(j^g,j^b)} = P[Y_{r+1}^G = (j^g, j^b) \mid Y_r^G = (i^g, i^b)]$ *given by:*

$$Q_{(0,0),(0,0)} = 1 \,,$$

$$Q_{(0,0),(j^g,j^b)} = 0 \quad \textit{for } 0 \leq j^g \leq |G|, 0 \leq j^b \leq m - |G|, (j^g, j^b) \neq (0,0) \,,$$

$$Q_{(i^g,i^b),(0,0)} = (1-p_h)^{i^g} (1-p_l)^{i^b} [1 - (1-p_h)^{|G|-i^g} (1-p_l)^{m-|G|-i^b}]$$

$$* \qquad \textit{for } (i^g, i^b) \neq (0,0), (i^g, i^b) \neq (|G|, m - |G|) \,,$$

$$Q_{(i^g,i^b),(j^g,j^b)} = \binom{i^g}{j^g} p_h^{j^g} (1-p_h)^{i^g - j^g} \binom{i^b}{j^b} p_l^{j^b} (1-p_l)^{i^b - j^b}$$

$*$ *for* $(i^g,i^b)\neq(0,0)$, $0\le j^g\le i^g$, $0\le j^b\le i^b$,$(j^g,j^b)\neq(0,0)$, $(j^g,j^b)\neq(i^g,i^b)$,

$Q_{(i^g,i^b),(i^g,i^b)}=p_h^{i^g}p_l^{i^b}+(1-p_h)^{|G|}(1-p_l)^{m-|G|}$ *for* $(i^g,i^b)\neq(0,0)$,

$Q_{(i^g,i^b),(j^g,j^b)}=0$ *for* $(i^g,i^b)\neq(0,0)$, $i^g\le j^g\le |G|$, $i^b\le j^b\le m-|G|$, $(j^g,j^b)\neq(i^g,i^b)$.

Proof. See the Appendix. □

Clearly,:

$$P[D_k\neq\emptyset\mid |G|=g]=\sum_{\substack{0\le i^g\le |G'|\\ 0\le i^b\le m-|G'|\\ (i^g,i^b)\neq(0,0)}} P[Y_k^{G'}=(i^g,i^b)]\,,|G'|=g\,. \tag{6}$$

Using standard numerical techniques for transient analysis of discrete-parameter Markov chains, we can obtain recurrent expressions for $P[Y_r^G=(i^g,i^b)]$, $1\le r\le k$, $|G|=g$, $0\le g\le m$. Those expressions together with (5) and (6) define a recurrent computational scheme for $[P_B]_{\mathrm{lbs}}$ for the PIAC model. The result is:

Theorem 4. *For the PIAC model,*

$$[P_B]_{\mathrm{lbs}}=\sum_{g=0}^{m}\binom{m}{b}p^g(1-p)^{m-g}W_g\,,$$

where

$$W_g=\sum_{\substack{0\le i^g\le g\\ 0\le i^b\le m-g\\ (i^g,i^b)\neq(0,0)}} Z_{g,k,i^g,i^b}$$

and the Z_{g,k,i^g,i^b}, $0\le g\le m$, $0\le i^g\le g$, $0\le i^b\le m-g$, $(i^g,i^b)\neq(0,0)$ *can be computed using, for increasing* r, *a set of recurrences giving* Z_{g,r,i^g,i^b}, $0\le g\le m$, $1\le r\le k$, $0\le i^g\le g$, $0\le i^b\le m-g$, $(i^g,i^b)\neq(0,0)$. *The initial values of the recurrences are:*

$$Z_{g,0,g,m-g}=1\,,\quad 0\le g\le m\,,$$

$$Z_{g,0,i^g,i^b}=0\,,\quad 0\le g\le m, 0\le i^g\le g, 0\le i^b\le m-g, (i^g,i^b)\neq(g,m-g), (i^g,i^b)\neq(0,0)\,.$$

The recurrences are:

$$\begin{aligned} Z_{g,r+1,i^g,i^b} \quad = \quad & \sum_{\substack{i^g\le j^g\le g\\ i^b\le j^b\le m-g\\ (j^g,j^b)\neq(i^g,i^b)}} \binom{j^g}{i^g}p_h^{i^g}(1-p_h)^{j^g-i^g}\binom{j^b}{i^b}p_l^{i^b}(1-p_l)^{j^b-i^b}Z_{g,r,j^g,j^b} \\ & + [p_h^{i^g}p_l^{i^b}+(1-p_h)^g(1-p_l)^{m-g}]\,Z_{g,r,i^g,i^b}\,, \\ & 0\le g\le m, 0\le r<k, 0\le i^g\le g, 0\le i^b\le m-g, (i^g,i^b)\neq(0,0)\,. \end{aligned}$$

Proof. The Z_{g,r,i^g,i^b} are $P[Y_r^G = (i^g, i^b)]$, $|G| = g$. Then, the recurrences for Z_{g,r,i^g,i^b} and their initial values follow from Theorem 3 using $Z_{g,r+1,i^b,i^b} = \sum_{j^g,j^b} Z_{g,r,j^b,j^b} \, Q_{(j^g,j^b),(i^g,i^g)}$. W_g is $P[D_k \neq \emptyset \mid |G| = g]$. Then, the expression for W_g follows from (6). The expression for $[P_B]_{\rm lbs}$ in terms of W_g follows from (5). □

Theorems 2 and 4 give computationally efficient procedures for $[P_B]_{\rm lbs}$ for, respectively, the ZIAC and the PIAC models. Using those procedures, we can obtain $[P_B]_{\rm lbs}$ for quite large values of k and m. Figure 2 plots $[P_B]_{\rm lbs}$, for values of k ranging from 2 to 10 and values of m ranging from 2 to 10, for the ZIAC model with $p =$ 0.2, 0.5, 0.8 and for the PIAC model with $p = 0.5$ and $\rho =$ 0.0, 0.2, 0.5. For a fixed number of alternatives, m, $[P_B]_{\rm lbs}$ decays, in some cases rapidly, as the number of attributes k increases. For a fixed number of attributes, k, $[P_B]_{\rm lbs}$ first decreases with the number of alternatives m up to a certain value of m, m^*, beyond which it increases with m. The explanation for that behavior is as follows. The addition of one alternative may have several effects. First, it may happen that the new alternative simply dominates all others, making the new D_k non-empty irrespectively of whether it was empty or not before. Second, the new alternative may be simply dominated by some alternative, leaving D_k unchanged. Third, it may also happen that the additional alternative neither simply dominates all others nor is simply dominated by any alternative, making empty the new D_k if it was non-empty before. The first effect would force an increase with m of $[P_B]_{\rm lbs}$, while the third effect would force a decrease. As m increases, the probability that the new alternative neither simply dominates all others nor is simply dominated by any other alternative becomes small, and for large enough m the third effect is negligible and $[P_B]_{\rm lbs}$ increases with m as a result of the first effect. In fact, as $m \to \infty$, the probability that some alternative will have all its attributes equal to 1 tends to 1, ensuring that $[P_B]_{\rm lbs} \to 1$ as $m \to \infty$. The m^* turning point seems to increase as the number of attributes k increases and as the quality of the alternatives decreases (p gets smaller). However, the more importance conclusion is that, except when the average quality of the alternatives is very good (ZIAC model, $p = 0.8$) or when the alternatives exhibit a strong positive inter-attribute correlation (PIAC model, $p = 0.5$, $\rho = 0.5$), $[P_B]_{\rm lbs}$ decays fast with k and has small values for large k. Thus, simple dominance does not explain the observed good performance of *DEBA*.

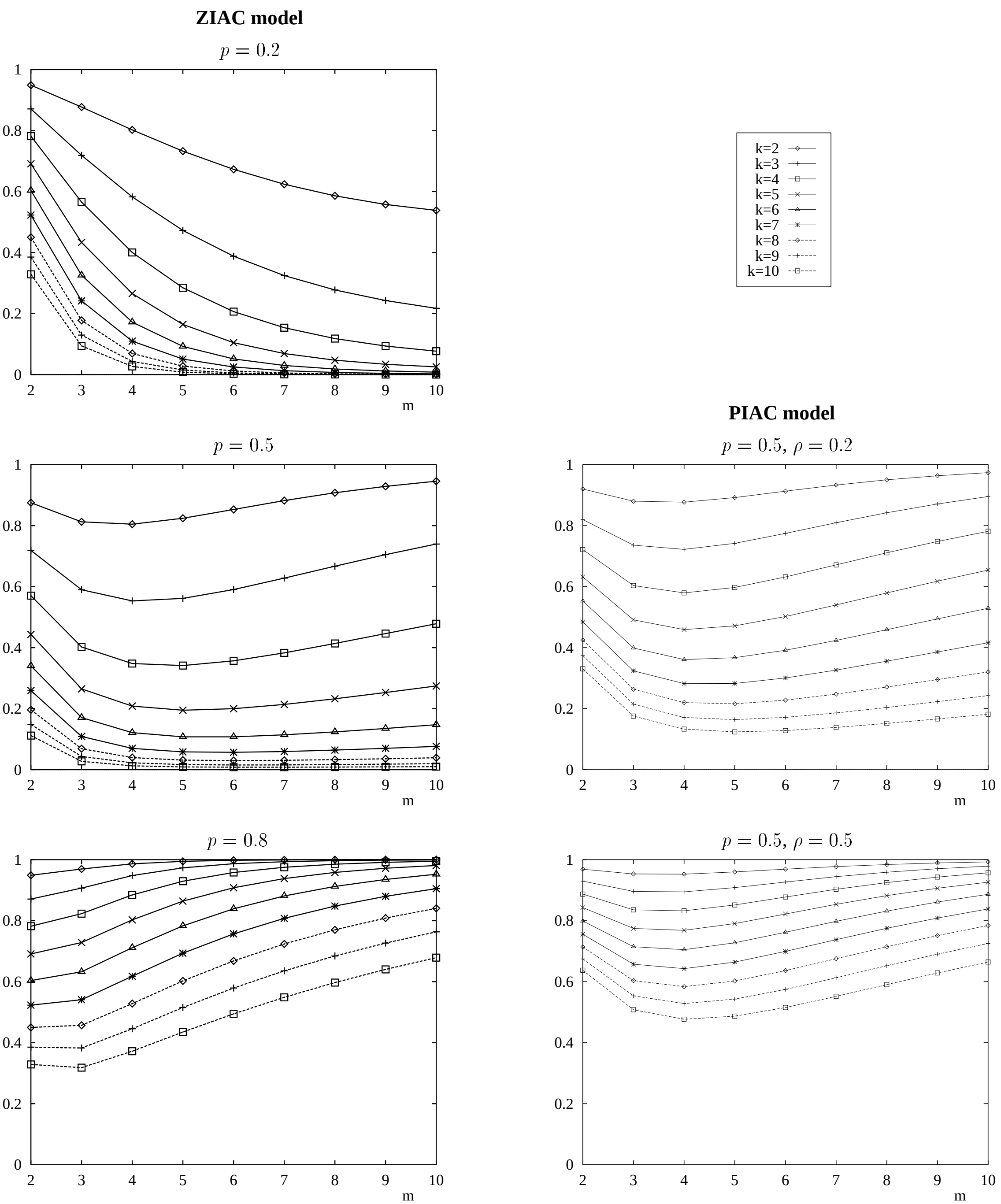

Figure 2: $[P_B]_{\text{lbs}}$ for the ZIAC model (left) for several values of p and the PIAC model (right) for $p = 0.5$ and several values of ρ.

4 Cumulative Dominance and *DEBA* Performance

As shown in the previous section, the presence of simple dominance is not enough to justify the good observed performance of *DEBA*. In this section we will review the concept of cumulative dominance and use it to explain, for the binary attribute case, the observed good performance of *DEBA*. Our results are however not restricted to the *DEBA* heuristic. They apply to classes of heuristics which we will call *cumulative dominance compliant* heuristics and *fully cumulative dominance compliant* heuristics, and examples of other heuristics belonging to those classes different from *DEBA* will be given.

4.1 Definitions and basic results

The cumulative profile of an alternative i, $1 \leq i \leq m$, is defined as $X_{i,s} = \sum_{t=1}^{s} x_{i,t}$, $1 \leq s \leq k$. Cumulative dominance is identical to simple dominance, but applied to the cumulative profile. Alternative i exhibits *cumulative dominance* over alternative j up to attribute r, denoted by $c_r(i,j)$, if and only if $X_{i,s} \geq X_{j,s}$, $1 \leq s \leq r$. Alternative i exhibits cumulative dominance over alternative j if and only if $c_k(i,j)$, i.e. if alternative i exhibits cumulative dominance over alternative j up to attribute k. Figure 3 illustrates cumulative dominance in the binary attribute case. In the figure, alternative 2 exhibits cumulative dominance over alternative 3 up to attribute 2 and alternative 1 exhibits cumulative dominance over alternatives 2 and 3. It is known that cumulative dominance characterizes optimality for non-increasing weights (Kirkwood and Sarin 1985):

Proposition 2. *$U_i \geq U_j$ for all weights $w_1 \geq w_2 \geq \cdots \geq w_k \geq 0$, $\sum_{s=1}^{k} w_s = 1$ if and only if $c_k(i,j)$.*

Proof. Notice that

$$U_i = \sum_{s=1}^{k} w_s x_{i,s} = \sum_{s=1}^{k-1} (w_s - w_{s+1}) X_{i,s} + w_k X_{i,k}$$

so that

$$U_i - U_j = \sum_{s=1}^{k-1} (w_s - w_{s+1}) \left(X_{i,s} - X_{j,s}\right) + w_k \left(X_{i,k} - X_{j,k}\right) ,$$

which is necessarily positive if alternative i cumulative dominates alternative j and weights are non-increasing. For the converse, that $\sum_{s=1}^{k} w_s x_{i,s} \geq \sum_{s=1}^{k} w_s x_{j,s}$ holds for all weights $w_1 \geq w_2 \geq$

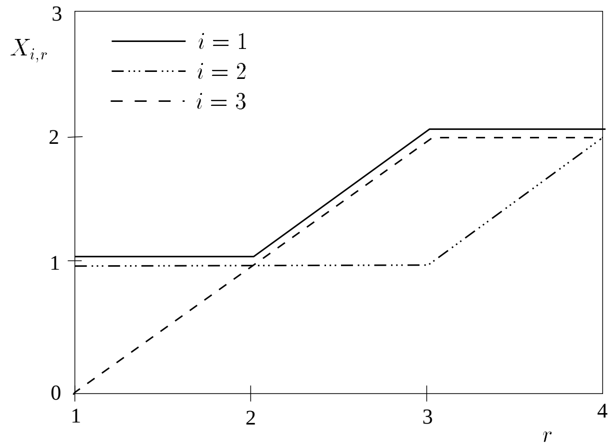

Figure 3: Alternative profiles illustrating cumulative dominance in the binary attribute case.

$\cdots \geq w_k \geq 0$ implies that it holds for the sets of weights

$$(w_1, w_2, ..., w_k) = (1, 0, 0, \cdots, 0),$$
$$(w_1, w_2, ..., w_k) = \left(\frac{1}{2}, \frac{1}{2}, 0, \cdots, 0\right),$$
$$\cdots$$
$$(w_1, w_2, ..., w_k) = \left(\frac{1}{k}, \frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k}\right),$$

yielding $c_r(i, j)$, $1 \leq r \leq k$. □

Note that Proposition 2 is not restricted to the binary attribute case.

For $1 \leq r \leq k$, let C_r denote the set of alternatives that exhibit cumulative dominance over any other alternative up to attribute r, i.e.,

$$C_r = \{1 \leq i \leq m \,:\, c_r(i, j), 1 \leq j \leq m\}. \tag{7}$$

Obviously, $C_1 \supset C_2 \supset \cdots \supset C_k$. All alternatives in C_r have identical cumulative attribute profiles up to attribute r and, therefore, they have identical attribute profiles up to attribute r. More importantly, if C_k is non-empty, then Proposition 2 guarantees that the alternatives in C_k will have the largest utility. In the example of Figure 3, $C_1 = C_2 = \{1, 2\}$ and $C_3 = C_4 = \{1\}$. C_1 will always be non-empty. In the binary attribute case, C_2 will be always non-empty also. This follows by noting that C_2 can only be empty if there exist two alternatives i, j with $x_{i,1} > x_{j,1}$ and $x_{i,1} + x_{i,2} < x_{j,1} + x_{j,2}$, which, being $x_{i,r}$ and $x_{j,r}$ binary, is impossible. In the non-binary attribute case, however, C_2 may well be empty. For $r \geq 3$, there is no guarantee even in the binary attribute case that C_r will be non-empty. Consider for instance the case of two alternatives with attribute profiles $x_{1,1} = 1$, $x_{1,2} = 0$, $x_{1,3} = 0$ and $x_{2,1} = 0$, $x_{2,2} = 1$, $x_{2,3} = 1$. In that case, we have $C_3 = \emptyset$. We say that a heuristic is *cumulative dominance compliant* if, whenever $C_k \neq \emptyset$, the heuristic chooses an alternative from C_k. Then, according to Proposition 2 we can state:

Theorem 5. *When C_k is non-empty any cumulative dominance compliant heuristic will choose a best alternative.*

Theorem 5 is not restricted to the binary attribute case.

The highest attribute index for which some alternative exhibits cumulative dominance over all other alternatives deserves careful attention. We will denote that index by r^*. Formally,

$$r^* = \max_{1 \leq r \leq k} \{1 \leq r \leq n \,:\, C_r \neq \emptyset\}. \quad (8)$$

By definition, $C_r = \emptyset$, $r^* < r \leq k$. Of course, C_k is non-empty if and only if $r^* = k$. In the binary attribute case, $r^* \geq 2$. For non-binary attributes, r^* could be equal to 1. A heuristic is said to be *fully cumulative dominance compliant* if it always chooses an alternative from C_{r^*}. Fully cumulative dominance compliance implies cumulative dominance compliance. The motivation by introducing the notion of fully cumulative dominance compliance is that results regarding the loss of those heuristics independent of the weights will be obtained for heuristics satisfying this property.

Consider the *DEBA* heuristic. Let A_r, $1 \leq r \leq k$ be the set of alternatives selected by the heuristic at its rth step. Remember that A_1 includes the alternatives i with largest $x_{i,1}$: the ones with $x_{i,1} = 1$ if some alternative has attribute 1 value 1 and all if $x_{i,1} = 0$, $1 \leq i \leq m$. A_2 includes the alternatives i in A_1 with largest $x_{i,2}$, and so on. Obviously $A_1 \supset A_2 \supset \cdots \supset A_k$. The *DEBA* heuristic selects at random any alternative in $A_k \neq \emptyset$. Informally speaking, an alternative exhibits cumulative dominance over another when it has superior values in more important attributes, possibly followed by inferior values in less important attributes. But *DEBA* eliminates those alternatives that have inferior values in the most important attributes, and hence it will never choose a cumulative dominated alternative. More formally, that *DEBA* is fully cumulative dominance compliant can be easy seen by noting the following important relation between the subsets A_r and C_r, $1 \leq r \leq r^*$:

Proposition 3. $A_r = C_r$, $1 \leq r \leq r^*$.

Proof. That $C_r \subset A_r$, $1 \leq r \leq k$, can be seen by induction on r. Obviously, $C_1 = A_1$. Assume the result holds for $r = s$ and consider the case $r = s + 1$. Let $i \in C_{s+1}$. We have $X_{i,s+1} \geq X_{j,s+1}$, $1 \leq j \leq m$, $j \neq i$. Since $C_{s+1} \subset C_s$, by the induction hypothesis $i \in A_s$. Assume $i \notin A_{s+1}$.

Then, there exists an alternative $l \in A_{s+1}$, $l \neq i$, with $x_{l,s+1} > x_{i,s+1}$ and $x_{l,u} = x_{i,u}$, $1 \leq u \leq s$. But this implies $X_{i,s+1} < X_{l,s+1}$ and, therefore, $i \notin C_{s+1}$, a contradiction. That $A_r \subset C_r$ for all r, $1 \leq r \leq r^*$ can be seen by contradiction. Take some r, $1 \leq r \leq r^*$, and an alternative i such that $i \in A_r$ and $i \notin C_r$. Since all alternatives in A_r are identical up to attribute r, this would imply $A_r \cap C_r = \emptyset$, which by $C_r \subset A_r$, implies $C_r = \emptyset$, a contradiction. Thus, $A_r = C_r$ for all r, $1 \leq r \leq r^*$. □

Since *DEBA* chooses and alternative from A_k and $A_k \subset A_{r^*} = C_{r^*}$ we have:

Theorem 6. DEBA *fully complies with cumulative dominance.*

DEBA is not alone in the classes of cumulative dominance compliant heuristics and fully cumulative dominance compliant heuristics. An example of a heuristic different from *DEBA* which is cumulative dominance compliant is the *EWn/DEBA* (Equal-Weights $n2 \leq n \leq k$. That heuristic first selects the alternatives i with largest $X_{i,n}$ and from them selects an alternative using *DEBA*. The *EWn/DEBA* heuristic has as special case ($n = k$) the *EW/DEBA* (Equal-Weights/Deterministic-Elimination-By-Aspects) heuristic and with $n = 2$ reduces to *DEBA* for the binary attribute case. Since no alternative i can cumulatively dominate all others if it does not have largest $X_{i,n}$, the first phase of *EWn/DEBA* will select a superset, A, of C_k. Assume $C_k \neq \emptyset$. Then, C_k will cumulative dominate all alternatives in A and, being *DEBA* cumulative dominance compliant, in the second phase, *EWn/DEBA* will choose an alternative from C_k, implying that *EWn/DEBA* is cumulative dominance compliant. However, *EWn/DEBA* is not fully cumulative dominance compliant. Consider, for instance, the case with three attributes and two alternatives with profiles $x_1 = (1, 0, 0)$ and $x_2 = (0, 1, 1)$. In that case, $r^* = 2$, and C_{r^*} contains only alternative 1, but *EW3/DEBA* (*EW/DEBA*) will choose alternative 2.

A heuristic different from *DEBA* which is fully cumulative dominance compliant would be the heuristic which first selects the alternatives in C_{r^*} and, then, selects among those alternatives one with largest $X_{i,k}$. We call that heuristic *CDS/EW* (Cumulative-Dominance-Selection/Equal-Weights). While more expensive to apply than *DEBA*, *CDS/EW* is intuitively appealing, since it first maximizes with certainty the part of the utility corresponding to attributes $1, 2 \ldots, r^*$, and, then, takes a more global view than *DEBA* to try to maximize the part of the utility corresponding to the attributes $r^* + 1, \ldots, k$, which might be advantageous if r^* is not close to k.

4.2 A lower bound for the probability of choosing a best alternative for cumulative dominance compliant heuristics

Consider any cumulative dominance compliant heuristic. Since alternatives in C_k have the largest utility and, by definition, when $C_k \neq \emptyset$, the heuristic will choose an alternative from C_k. Hence, $[P_B]_{\text{lbc}} = P[C_k \neq \emptyset]$ is a lower bound for the probability with which the heuristic will choose a best alternative. Since simple dominance implies cumulative dominance, $C_k \supset D_k$, $P[C_k \neq \emptyset] \geq P[D_k \neq \emptyset]$, and $[P_B]_{\text{lbc}}$ might be significantly better (tighter) than $[P_B]_{\text{lbs}}$. $[P_B]_{\text{lbc}}$ is a lower bound on the probability P_B that a cumulative dominance compliant heuristic will choose a best alternative which only depends on the weights being non-increasing. For a particular set of weights, that lower bound might not be tight. In fact, if the weights are non-compensatory ($w_r \geq \sum_{s=r+1}^{k} w_s$, $1 \leq r \leq k-2$), then it can be shown that *DEBA* (Katsikopoulos and Fasolo (in press), Martignon and Hoffrage, 1999, 2002) and *EW/DEBA* (Hogarth and Karelaia (in press)) choose the best alternative with probability one, whereas, as we will see, $[P_B]_{\text{lbc}}$ can be far from 1. However, we will show (for the two probabilistic models considered in the paper) that the lower bound for P_B does not decrease fast with m and k, implying that P_B will not decrease fast with m and k for any cumulative dominance compliant heuristic and providing a first explanation of the observed good performance of *DEBA*. On the other hand, P_B may decrease fast with both m and k for non cumulative dominance compliant heuristics. For instance, such behavior has been observed (Hogarth and Karelaia, 2003) for the *EW/RAN* (Equal-Weights/Random) heuristic, which chooses at random among the alternatives i with largest $X_{i,k}$.

In this section, we will compute $[P_B]_{\text{lbc}}$ for the two probabilistic models considered in the paper. Since, as noted, for the binary attribute case, $C_2 \neq \emptyset$, for $k = 2$, $[P_B]_{\text{lbc}} = 1$. We will therefore assume $k \geq 3$. Computation of $[P_B]_{\text{lbc}}$ seems to be significantly harder than computation of $[P_B]_{\text{lbs}}$. Essentially, this is because, in the case $x_{i,r+1} = 0$, $i \in C_r$, whether C_{r+1} is empty or not not only depends on $x_{i,r+1}$, $i \in \{1, 2, \ldots, m\} - C_r$. This prevents the use of discrete-parameter Markov chain approaches similar to the ones used in Section 3 to compute $[P_B]_{\text{lbs}}$ for the two probabilistic models considered in the paper. We have taken another approach, which profits from our binary set-up and uses ROBDDs (Reduced Ordered Binary Decision Diagrams). A ROBDD (see Bryant 1986) is a directed acyclic graph having a single root note and two terminal nodes (leaves), one labeled 0 and another labeled 1, which represents an arbitrary given Boolean function of a given set of binary variables. ROBDDs are called reduced because each node represents a different Boolean function (the root node represents the given Boolean function). They are called ordered because they depend on the ordering of the binary variables. ROBDDs are canonical (unique) representations of Boolean functions which only depend on the ordering of the binary variables. That property has given to ROBDDs many applications, e.g., formal verification of digital circuits. Given a Boolean function $F(x_1, x_2, \ldots, x_n)$ of n independent Bernoulli random variables, we can compute $P[F(x_1, x_2, \ldots, x_n) = 1]$ by building the ROBDD of $F()$ as a function of $x_1, x_2, \ldots, x_n$ and, then, traversing bottom-up the ROBDD. At each step, we obtain the probability that the Boolean function represented by a node is equal to 1 by multiplying the corresponding probability of the 0-edge node by the probability that the binary variable associated with the processed node has value 0, multiplying the corresponding probability of the 1-edge node by the probability that the binary variable has value 1, and adding up those partial results. To build the ROBDD, a Boolean expression for $F()$ as a function of $x_1, x_2, \ldots, x_n$ involving basic Boolean functions like NOT, AND, OR is required.

The Boolean function we have to consider to compute $[P_B]_{\text{lbc}}$ is the indicator function of the event $\{C_k \neq \emptyset\}$. For the ZIAC model, the Bernoulli random variables to be considered are $x_{i,s}$, $1 \leq i \leq m$, $1 \leq s \leq k$ and an expression for $F_{m,k}(x_{1,1}, , x_{1,k}, \ldots, x_{m,1}, \ldots, x_{m,l}) = \mathbf{1}_{C_k \neq \emptyset}$ is:

$$F_{m,k}(x_{1,1}, \ldots, x_{1,r}, \ldots, x_{m,1}, \ldots, x_{m,r}) = \bigvee_{i=1}^{m} \bigwedge_{\substack{j=1 \\ j \neq i}}^{m} \bigwedge_{s=1}^{k} \mathbf{1}_{X_{i,s} \geq X_{j,s}} \, ,$$

where the indicator functions $\mathbf{1}_{X_{i,s} \geq X_{j,s}}$ can be expressed in terms of the Bernoulli random variables $x_{i,t}$, $1 \leq i \leq m$, $1 \leq t \leq s$ using standard implementations of binary adders and binary comparators. For the PIAC model, the Bernoulli random variables to be considered are z_i, $1 \leq i \leq m$, and $y^0_{i,s}$, $y^1_{i,s}$, $1 \leq i \leq m$, $1 \leq s \leq k$ and an expression for $F_{m,k}(z_1, \ldots, z_m, y^0_{1,1}, \ldots, y^0_{m,r}, y^1_{1,1}, \ldots, y^1_{m,r}) = \mathbf{1}_{C_k \neq \emptyset}$ is:

$$F_{m,r}(z_1, \ldots, z_m, y^0_{1,1}, \ldots, y^0_{m,r}, y^1_{1,1}, \ldots, y^1_{m,r}) = \bigvee_{i=1}^{m} \bigwedge_{\substack{j=1 \\ j \neq i}}^{m} \bigwedge_{s=1}^{k} 1_{X_{i,s} \geq X_{j,s}} \, ,$$

$$x_{i,s} = (1 - z_i) \wedge y^0_{i,s} \vee z_i \wedge y^1_{i,s} \, ,$$

where the indicator functions $\mathbf{1}_{X_{i,s} \geq X_{j,s}}$ can be expressed in terms of the Boolean functions $x_{i,t}$, $1 \leq i \leq m$, $1 \leq t \leq s$ using standard implementations of binary adders and binary comparators.

The computational cost of the ROBDD based method is mainly determined by the size (number of nodes) of the resulting ROBDD. It is also affected by the peak number of reserved nodes. The ROBDD of the function is built (Bryant 1986) by traversing the description of the Boolean function in terms of basic Boolean functions such as NOT, AND and OR functions and combining the ROBDDs of the nodes of that description. Then, the peak number of reserved nodes is the maximum sum of the nodes in the ROBDDs which have to be held during the process. The size of the ROBDD depends on the ordering chosen for the variables on which the function depends and can be reduced by using ROBDDs with complement edges (Brace et al. 1990). The variable ordering is typically chosen using heuristics based on the Boolean description of the function . We have used the topology heuristic (Nikolskaia et al. 1998) with good results. Using that heuristic and ROBDDs with complement 0-edges, we have been able to compute the probabilities $P_C(r)$ for values of m and k as large as 10. As expected, the size of the ROBDDs increased with both m and r. For $m = 10$ and $k = 10$, the ROBDD for the ZIAC model had 320,558 nodes and its construction resulted in a peak number of reserved nodes of 5,182,179. For the PIAC model, the corresponding ROBDDs were a bit larger. For $m = 10$ and $k = 10$, the ROBDD had 681,216 nodes and its construction resulted in a peak number of reserved nodes of 11,639,367. To build the ROBDDs we used the CU Decision Diagram Package (CU 2005).

Figure 4 plots $[P_B]_{\mathrm{lbc}}$, for values of k ranging from 3 to 10 and values of m ranging from 2 to 10, for the ZIAC model for $p = $ 0.2, 0.5, 0.8 and for the PIAC model for $p = 0.5$ and $\rho =$ 0.0, 0.2, 0.5. We can note that in all cases $[P_B]_{\mathrm{lbc}}$ is significantly larger than $[P_B]_{\mathrm{lbs}}$ (Figure 2). As $[P_B]_{\mathrm{lbs}}$, for a fixed number of alternatives m, $[P_B]_{\mathrm{lbc}}$ decreases with k but, contrary to $[P_B]_{\mathrm{lbs}}$, $[P_B]_{\mathrm{lbc}}$ never decreases fast with k. As for $[P_B]_{\mathrm{lbs}}$, for fixed k, there exists a turning point, m^*, for m before which $[P_B]_{\mathrm{lbc}}$ decays with m and beyond which $[P_B]_{\mathrm{lbc}}$ increases with m. The explanation of the existence of those turning points is similar to the explanation of the corresponding turning points for $[P_B]_{\mathrm{lbs}}$ but in terms of cumulative dominance instead of in terms of simple dominance. For fixed k and m, the values of $[P_B]_{\mathrm{lbc}}$ improve (increase) with the average quality of the alternatives (higher p) and with a positive inter-attribute correlation (higher ρ). It is noteworthy that $[P_B]_{\mathrm{lbc}}$ is very close to 1 when either the alternatives have good average quality (ZIAC model, $p = 0.8$) or there exists strong positive correlation among the attribute values of a given alternative (PIAC model, $p = 0.5$, $\rho = 0.5$). In those cases, the presence of cumulative dominance is enough to explain a very good performance of any cumulative dominance compliant heuristic, including, of course, *DEBA*. It is also noteworthy that, contrary to $[P_B]_{\mathrm{lbs}}$ and contrary to intuition, $[P_B]_{\mathrm{lbc}}$ has a significant value even when the alternatives have a poor quality and there does not exist any positive correlation among the values of the attributes of a given alternative (ZIAC model, $p = 0.2$).

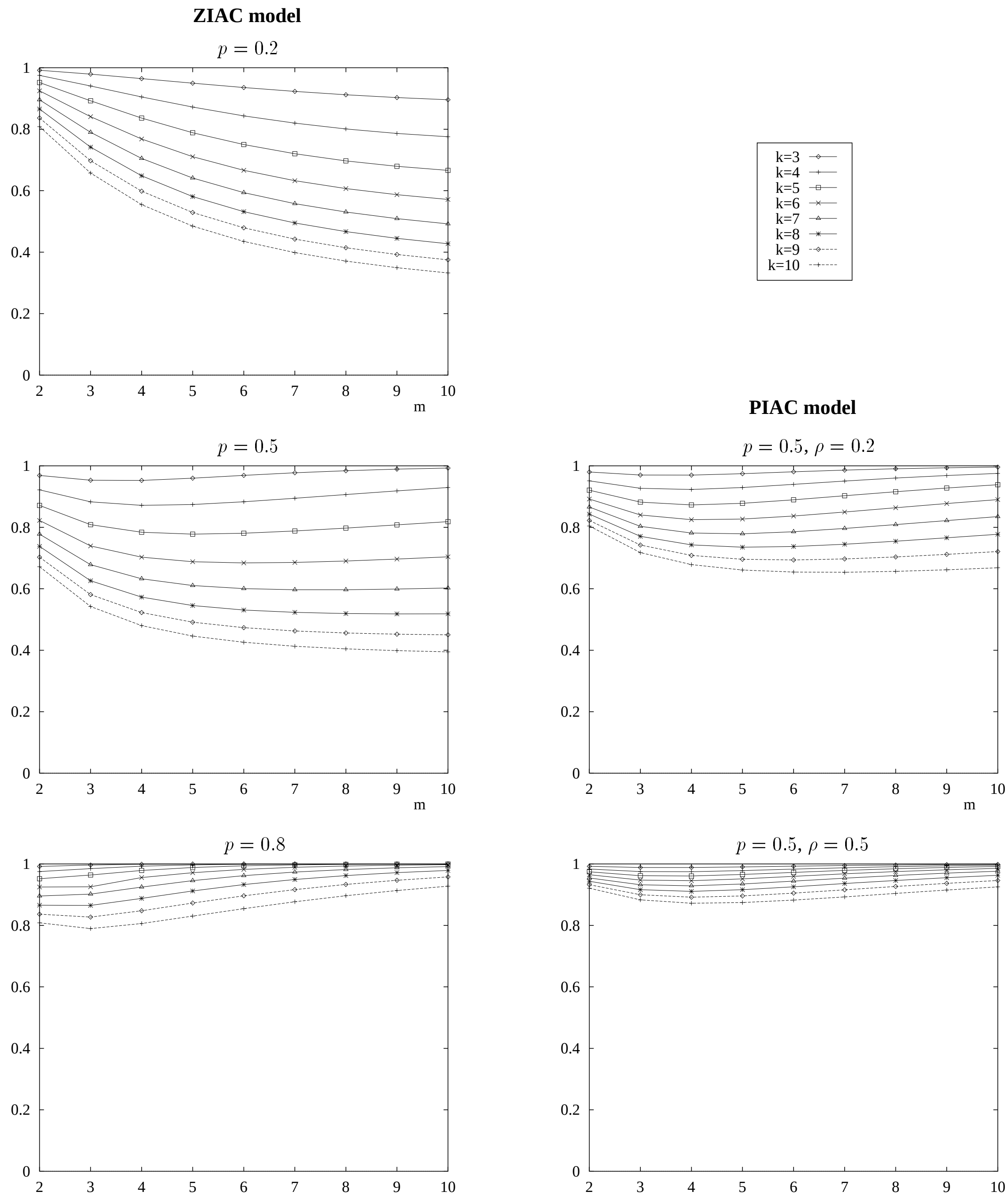

Figure 4: $[P_B]_{\mathrm{lbc}}$ for the ZIAC model (left) for several values of p and the PIAC model (right) for $p = 0.5$ and several values of ρ.

4.3 An upper bound for the expected loss of fully cumulative dominance compliant heuristics

The probability that a heuristic chooses a best alternative is an important metric of the performance of the heuristic. Guaranteeing that probability will be close to 1 certainly shows that the heuristic is a good heuristic. The expected loss of the heuristic, i.e. the expected difference between the utility of a best alternative and the utility of the alternative chosen by the heuristic is another relevant metric, which is specially useful when the probability of chosen a best alternative is not close to 1. The reason is simple: in many cases, we would be content with a non-best alternative as far as its utility is reasonably close to the utility of a best alternative. With that motivation, in this section, we will derive, for the two probabilistic models under consideration, an upper bound for the expected loss of any fully cumulative compliant heuristic, including, of course, *DEBA*. Since for $k = 2$ any fully cumulative dominance compliant heuristic will choose a best alternative with probability 1, and, therefore, the expected loss will be 0, we will assume $k \geq 3$.

Let b the alternative chosen by the heuristic. Then, the loss of the heuristic is

$$L = \max_{1 \leq i \leq m} U_i - U_b \, . \tag{9}$$

We will derive an upper bound for L as a function of r^*. Note that L is a random variable. The upper bound for the expected loss will follow by conditioning on r^* and taking expectations.

Since the heuristic is fully cumulative dominance compliant, we know that $b \in C_{r^*}$. Let i be any other alternative. Compared to b, how much better can j be? To answer that question, it is useful to consider the following formulation for the utility of an alternative $U_i = \sum_{s=1}^{k} w_s x_{i,s}$ in terms of its cumulative profile.

$$U_i = \sum_{s=1}^{k-1} (w_s - w_{s+1}) X_{i,s} + w_k X_{i,k} \, .$$

According to this formulation, given a set of weights, the highest loss occurs when the cumulative profile of i meets the following two conditions: 1) $X_{i,s} = X_{b,s}$, $1 \leq s \leq r^*$ (since $b \in C_{r^*}$, $X_{i,s} \leq X_{b,s}$, $1 \leq s \leq r^*$), 2) $X_{i,s} = X_{b,s} + (s - r^*)$, $r^* + 1 \leq s \leq k$ (which is possible, since all $x_{i,s}$, $r^* + 1 \leq s \leq k$ could be 1 and all $x_{b,s}$, $r^* + 1 \leq s \leq k$ could be 0). Thus, for a given set of weights,

$$L \leq \sum_{s=r^*+1}^{k-1} (w_s - w_{s+1})(s - r^*) + w_k(k - r^*) = \sum_{s=r^*+1}^{k} w_s \, .$$

To find an upper bound for L independent of the weights, it remains to maximize $\sum_{s=r^*+1}^{k} w_s$ subject to the restrictions which the w_s, $r^* + 1 \leq s \leq k$ have to satisfy. The restrictions are (the last one comes from $w_1 \geq w_2 \geq \cdots \geq w_{r^*+1}$ and $\sum_{s=1}^{k} w_s = 1$):

$$w_k \geq 0\,,$$

$$w_{s-1} \geq w_s\,, \quad r^* + 2 \leq s \leq k\,,$$

$$(r^*+1)w_{r^*+1} + \sum_{s=r^*+2}^{k} w_s \leq 1\,.$$

This is a linear programming problem with bounded domain and, as it is well known, the maximum occurs at some vertex of the polyhedron defined by the restrictions. The vertices of the polyhedron are

$$(w_{r^*+1}, w_{r^*+2}, w_{r^*+3}, \ldots, w_k) = (0, 0, 0, \ldots, 0)\,,$$
$$(w_{r^*+1}, w_{r^*+2}, w_{r^*+3}, \ldots, w_k) = \left(\frac{1}{r^*+1}, 0, 0, \ldots, 0\right)\,,$$
$$(w_{r^*+1}, w_{r^*+2}, w_{r^*+3}, \ldots, w_k) = \left(\frac{1}{r^*+2}, \frac{1}{r^*+2}, 0, \ldots, 0\right)\,,$$
$$\ldots$$
$$(w_{r^*+1}, w_{r^*+2}, w_{r^*+3}, \ldots, w_k) = \left(\frac{1}{k}, \frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k}\right)\,,$$

and, therefore, the maximum is

$$\max_{r^*+1 \leq s \leq k} \frac{s - r^*}{s} = \frac{k - r^*}{k}\,.$$

Then, we can state the following result:

Theorem 7. *Any heuristic that fully complies with cumulative dominance will have a loss with respect to a best alternative upper bounded by* $(k - r^*)/k$.

Note that the upper bound for the loss given by Theorem 7 is not restricted to the binary attribute case.

Recall that for $n > 3$, *EWn/DEBA* is not fully cumulative dominance compliant. Hence, the upper bound on the expected loss does not apply. Considering again the example with $k = 3$ and $m = 2$ given by $x_1 = (1, 0, 0)$ and $x_2 = (0, 1, 1)$, the maximum loss guaranteed by any heuristics that fully complies with cumulative dominance is $(k - r^*)/k = 1/3$. $DEBA$ chooses alternative 1 and, as expected, the maximum loss in the most pessimistic weight scenario ($w_1 = w_2 = w_3 = 1/3$) is given by $L = U_2 - U_1 = 1/3$. In contrast, *EW3/DEBA* chooses alternative 2, and for appropriate weights ($w_1 = 1 - 2\varepsilon$, $w_2 = w_3 = \varepsilon$), this choice may yield a loss of $L = U_1 - U_2 = 1 - 4\varepsilon \approx 1$.

As noted, in the binary attribute case $2 \leq r^* \leq k$. Let $P(r) = P[r^* = r]$, $2 \leq r \leq k$. Then, conditioning on the value of r^* and taking expectations:

$$E[L] = \sum_{r=2}^{k} P(r) E[L | r^* = r]$$

and using Theorem 7, for any fully cumulative dominance compliant heuristic:

$$E[L] \leq \sum_{r=2}^{k-1} P(r) \frac{k-r}{k} \, .$$

This is the sought upper bound for the expected loss. Let us call it $[E[L]]_{\text{ub}}$. It remains to discuss a procedure for computing $P(r)$, $2 \leq r \leq k-1$ for the two considered probabilistic models. Let $Q(r) = P[r^* \geq r]$. We have

$$P(r) = Q(r) - Q(r+1) \, , \quad 2 \leq r \leq k-1 \, .$$

Since $r^* \geq 2$, $Q(2) = 1$. The $Q(r)$, $3 \leq r \leq k$ required to compute $P(r)$, $2 \leq r \leq k-1$ can be obtained, noting that $Q(r) = P[C_r \neq \emptyset]$, using the ROBDD approaches described in Section 4.2 for the computation of $[P_B]_{\text{lbc}} = Q(k)$ for the ZIAC and the PIAC probabilistic models with the index k replaced by the index r.

Figure 5 plots $[E[L]]_{\text{ub}}$, for values of k ranging from 3 to 10 and values of m ranging from 2 to 10, for the ZIAC model for $p =$ 0.2, 0.5, 0.8 and for the PIAC model for $p = 0.5$ and $\rho =$ 0.0, 0.2, 0.5. For fixed number of alternatives m, $[E[L]]_{\text{ub}}$ increases with k, but in no case does so fast. For fixed k , there exist a turning point m^* before which $[E[L]]_{\text{ub}}$ increases with m and beyond which $[E[L]]_{\text{ub}}$ decreases with m . Not surprisingly, the value of $[E[L]]_{\text{ub}}$ is very small when either the alternatives have good average quality (ZIAC model, $p = 0.8$) or there exist strong positive inter-attribute correlation (PIAC model, $p = 0.5$, $\rho = 0.5$). The values of $[E[L]]_{\text{ub}}$ are reasonably small in the presence of a moderate positive inter-attribute correlation (PIAC model, $p = 0.5$, $\rho = 0.2$) and are moderate in all cases. Those observations complete the explanation of the observed good performance of *DEBA* and make that good performance extensible to any fully cumulative dominance compliant heuristic.

5 Final Remarks and Conclusions

Using the cumulative dominance concept we have justified, for the binary attribute case and for two probabilistic models, the observed good performance of the *DEBA* heuristic. The results obtained in

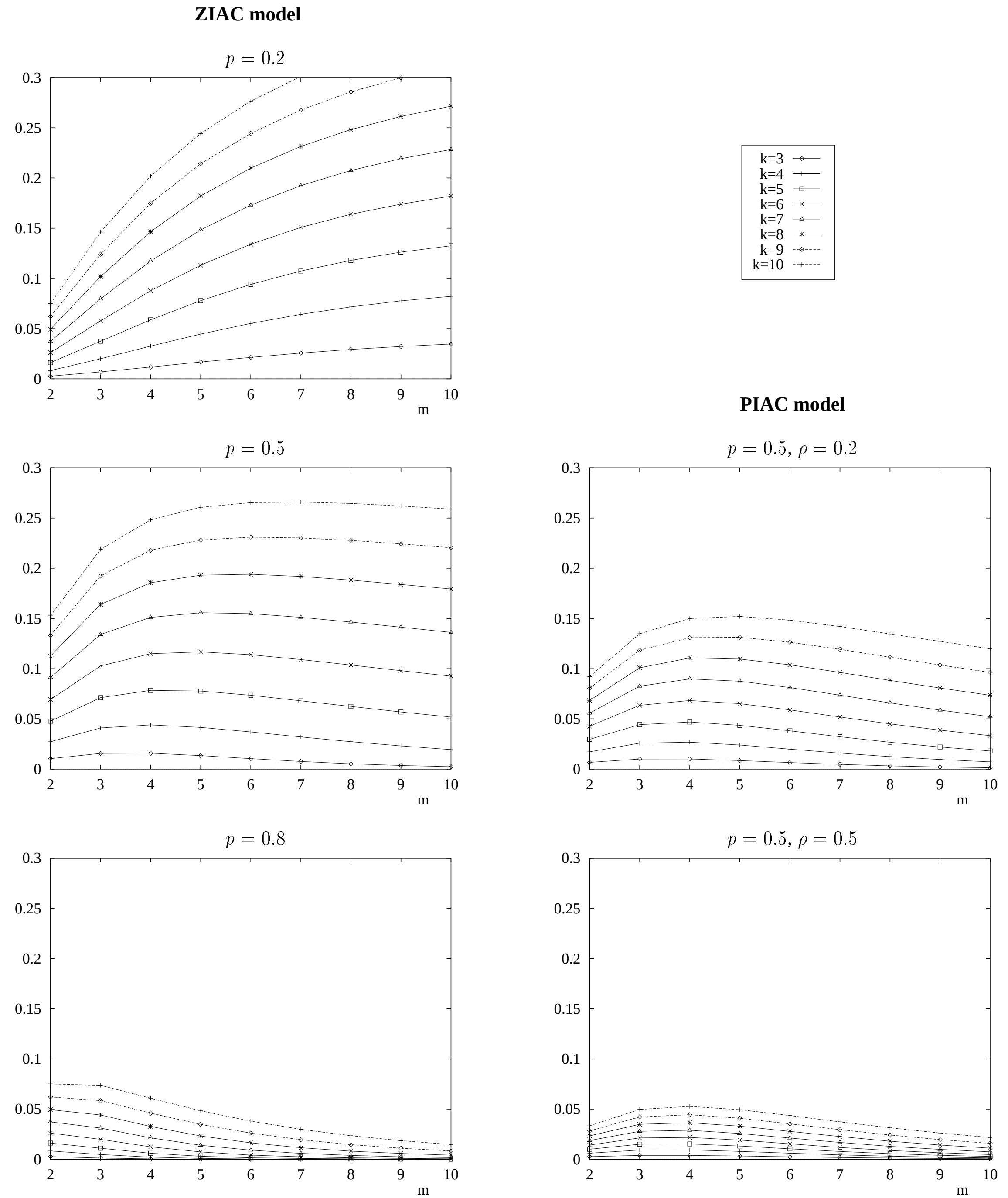

Figure 5: $[E[L]]_{\mathrm{ub}}$ for the ZIAC model (left) for several values of p and the PIAC model (right) for $p = 0.5$ and several values of ρ.

the paper are applicable to any cumulative dominance compliant heuristic and any fully cumulative dominance compliant heuristic and examples of heuristics in those classes different from *DEBA* have been given. Our results can be used to bound the performance of those heuristics independently of the particular values of the weights, which are unknown. Our computational procedures are feasible for quite large values of m and k (we have given results for m up to 10 and k up to 10). Previous studies concerning the performance of *DEBA* and *EWn/DEBA* (Hogarth and Karelaia 2003) used simpler enumeration approaches and were restricted to the ZIAC model with $p = 0.5$ and more modest values of m and k (m up to 5 and k up to 5).

Our study is one more step in the direction of reducing the descriptive–prescriptive gap in multi-attribute decision making. We have shown that *DEBA* and other related heuristics achieve a good performance in the binary attribute setting with a moderate number of attributes. This strongly supports the insight that *the key* managerial skill is to identify and rank the most relevant attributes or factors. Efforts to specify exact values of weights and/or use a informational-intensive decision procedures may have a minor return and be justified only for a small fraction of decisions (Keeney 2004). Since much may not be lost by the binary encoding of attribute values (Hogarth and Karelaia 2005b), our results can also justify good performance of *DEBA* and related heuristics when the attribute are continuous random variables.

Our analysis can be extended in several directions. First, it would be interesting to analyze the impact of a negative inter-attribute correlation. However, whereas this can be introduced in several ways, it is not a simple task. Another, obvious, direction is the consideration of probabilistic models in which attributes are continuous random variables, possibly correlated. Another possibility is the consideration of different scenarios for the available knowledge about the values of the weights w_i, $1 \leq i \leq k$ (see Barron 1992). Our analysis has been restricted to the case of non-increasing weights. A possible extension is to consider the case where the relative ranking of the first q weights is not known, i.e. $w_1, w_2, \ldots w_q \geq w_{q+1} \geq \cdots \geq w_k \geq 0$. Picking up $q = 1$ puts us in the non-increasing weights scenario assumed in the paper, which is optimally characterized by cumulative dominance. Picking up $q = k$ puts us in the non-negative weights scenario, which is optimally characterized by simple dominance. It is easy to check that the more general scenario is optimally characterized by q-dominance: an alternative i exhibits q-dominance over another alternative j if and only if $d_r(i, j)$ for all r, $1 \leq r \leq q$ and $c_r(i, j)$ for all r, $q + 1 \leq r \leq k$. Using the q-dominance concept we could derive in a similar way as it has been done in the paper performance measures for *q-dominance compliant* heuristics and *fully q-dominance compliant* heuristics. All those extensions are expected to be the subject of future work.

Appendix

Proof of Theorem 1 That $Y_0^G = (|G|, m-|G|)$ is by definition. We will compute the probabilities $P[Y_1^G = (j^g, j^b) \,|\, Y_0^G = (|G|, m-|G|)]$, $(j^g, j^b) \in \{(i,j), 0 \le i \le |G|, 0 \le j \le m-|G|\}$ and the probabilities $P[Y_{r+1}^G = (j^g, j^b) \,|\, Y_r^G = (i^g, i^b) \wedge Y_{r-1}^G = (i_{r-1}^g, i_{r-1}^b) \wedge \cdots \wedge Y_1^G = (i_1^g, i_1^b) \wedge Y_0^G = (|G|, m-|G|)]$, $(i_1^g, i_1^b), \ldots, (i_{r-1}^g, i_{r-1}^b), (i^g, i^b), (j^g, j^b) \in \{(i,j), 0 \le i \le |G|, 0 \le j \le m-|G|\}$. It will trun out that the former are equal to $Q_{(|G|, m-|G|),(j^g,j^b)}$ and the latter only depend on (i^g, i^b) and (j^g, j^b) and are equal to $Q_{(i^g,i^b),(j^g,j^b)}$, thus proving that $Y^G = \{Y_r^G; 0 \le r \le k\}$ is an homogeneous discrete-parameter Markov chain (with truncated parameter) with one-step transition probabilities $Q_{(i^g,i^b),(j^g,j^b)}$.

Since $Y_0^G = (|G|, m-|G|)$ with probability 1, $P[Y_1^G = (j^g, j^b) \mid Y_0^G = (|G|, m-|G|)] = P[Y_1^G = (j^g, j^b)]$. First, $Y_1^G = (|G|, m-|G|)$ if and only if all alternatives have same attribute 1 value. Then, $P[Y_1^G = (|G|, m-|G|) | Y_0^G = (|G|, m-|G|)] = p_h^{|G|} p_l^{m-|G|} + (1-p_h)^{|G|}(1-p_l)^{m-|G|}$. Second, $Y_1^G = (j^b, j^b)$, $0 \le j^g \le |G|$, $0 \le j^b \le m-|G|$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (|G|, m-|G|)$ if and only if j^b of the $|G|$ good alternatives have attribute 1 value 1, the remaining $|G| - j^g$ good alternatives have attribute 1 value 0, j^b of the $m-|G|$ bad alternatives have attribute 1 value 1, and the remaining $m-|G|-j^b$ bad alternatives have attribute 1 value 0. Then, $P[Y_1^G = (j^g, j^b) \,|\, Y_0^G = (|G|, m-|G|)] = \binom{|G|}{j^g} p_h^{j^g} (1-p_h)^{|G|-j^g} \binom{m-|G|}{j^b} p_l^{j^b} (1-p_l)^{m-|G|-j^b}$, $0 \le j^b \le m-|G|$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (|G|, m-|G|)$. Finally, Y_1^G cannot be $(0,0)$. Then, $P[Y_1^G = (0,0) \mid Y_0^G = (|G|, m-|G|)] = 0$.

Let $0 < r < k$. Assume $(i^g, i^b) = (0,0)$. Thus, $D_r^g = D_r^b = \emptyset$. Since $D_{r+1}^g \subset D_r^g$ and $D_{r+1}^b \subset D_r^b$, $D_{r+1}^g = D_{r+1}^b = \emptyset$, implying $P[Y_{r+1}^G = (0,0) \mid Y_r^G = (0,0) \wedge Y_{r-1}^G = (i_{r-1}^g, i_{r-1}^b) \wedge \cdots \wedge Y_1^G = (i_1^g, i_1^b) \wedge Y_0^G = (|G|, m-|G|)] = 1$ and $P[Y_{r+1}^G = (j^g, j^b) \mid Y_r^G = (0,0) \wedge Y_{r-1}^G = (i_{r-1}^g, i_{r-1}^b) \wedge \cdots \wedge Y_1^G = (i_1^g, i_1^b) \wedge Y_0^G = (|G|, m-|G|)] = 0$, $0 \le j^g \le |G|$, $0 \le j^b \le m-|G|$, $(j^g, j^b) \ne (0,0)$. Assume $(i^g, i^b) \ne (0,0)$. Thus, $|D_r^g| = i^g$ and $|D_r^b| = i^b$. The values of $|D_{r+1}^g|$ and $|D_{r+1}^b|$ depend on $|D_r^g| = i^g$ and $|D_r^b| = i^b$ and the values of the attributes $r+1$ of the alternatives as follows. First, $D_{r+1}^b \subset D_r^b$ and $D_{r+1}^g \subset D_r^b$ imply $|D_{r+1}^g| \le |D_r^g| = i^g$ and $|D_{r+1}^b| \le |D_r^b| = i^b$ and, then, $P[Y_{r+1}^G = (j^g, j^b) \mid Y_r^G = (i^g, i^b) \wedge Y_{r-1}^G = (i_{r-1}^g, i_{r-1}^b) \wedge \cdots \wedge Y_1^G = (i_1^g, i_1^b) \wedge Y_0^G = (|G|, m-|G|)] = 0$, $i^g \le j^g \le |G|$, $i^b \le j^b \le m-|G|$, $(j^g, j^b) \ne (i^g, i^b)$. Second, for $(i^g, i^b) \ne (|G|, m-|G|)$, $|D_{r+1}^g| = 0$ and $|D_{r+1}^b| = 0$ if and only if all alternatives in D_r^g have attribute $r+1$ value 0, all alternatives in D_r^b have attribute $r+1$ value 0, and some alternative in $\{1, 2, \ldots, m\} - D_r^g - D_r^b$ has attribute $r+1$ value 0. Then,

$P[Y^G_{r+1} = (0,0) | Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = (1-p_h)^{i^g}(1-p_l)^{i^b}[1-(1-p_h)^{|G|-i^g}(1-p_l)^{m-|G|-i^b}]$. Third, $|D^g_{r+1}|$ and $|D^b_{r+1}|$ will have values j^g and j^b, respectively, $0 \le j^g \le i^g$, $0 \le j^b \le i^b$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (i^g, i^b)$ if and only if j^b alternatives in D^g_r have attribute $r+1$ value 1, the remaining $i^b - j^b$ alternatives in D^g_r have attribute $r+1$ value 0, j^b alternatives in D^b_r have attribute $r+1$ value 1, and the remaining $i^b - j^b$ alternatives in D^b_r have attribute $r+1$ value 0. Then, $P[Y^G_{r+1} = (j^g, j^b) \mid Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = \binom{i^g}{j^g} p_h^{j^g}(1-p_h)^{i^g-j^g}\binom{i^b}{j^b} p_l^{j^b}(1-p_l)^{i^b-j^b}$, $0 \le j^g \le i^g$, $0 \le j^b \le i^b$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (i^g, i^b)$. Finally, $|D^g_{r+1}|$ will have value i^g and $|D^b_{r+1}|$ will have value i^b if and only if either all alternatives in $D^g_r \cup D^b_r$ have attribute $r+1$ value 1 or all alternatives have attribute $r+1$ value 0. Then, $P[Y^G_{r+1} = (i^g, i^b) \mid Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = p_h^{i^g} p_l^{i^b} + (1-p_h)^{|G|}(1-p_l)^{m-|G|}$. □

Proof of Theorem 3 That $Y^G_0 = (|G|, m-|G|)$ is by definition. We will compute the probabilities $P[Y^G_1 = (j^g, j^b) \mid Y^G_0 = (|G|, m-|G|)]$, $(j^g, j^b) \in \{(i,j), 0 \le i \le |G|, 0 \le j \le m-|G|\}$ and the probabilities $P[Y^G_{r+1} = (j^g, j^b) \mid Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)]$, $(i^g_1, i^b_1), \ldots, (i^g_{r-1}, i^b_{r-1}), (i^g, i^b), (j^g, j^b) \in \{(i,j), 0 \le i \le |G|, 0 \le j \le m-|G|\}$. It will turn out that the former are equal to $Q_{(|G|, m-|G|),(j^g, j^b)}$ and the latter only depend on (i^g, i^b) and (j^g, j^b) and are equal to $Q_{(i^g, i^b),(j^g, j^b)}$, thus proving that $Y^G = \{Y^G_r; 0 \le r \le k\}$ is an homogeneous discrete-parameter Markov chain (with truncated parameter) with one-step transition probabilities $Q_{(i^g, i^b),(j^g, j^b)}$.

Since $Y^G_0 = (|G|, m-|G|)$ with probability 1, $P[Y^G_1 = (j^g, j^b) \mid Y^G_0 = (|G|, m-|G|)] = P[Y^G_1 = (j^g, j^b)]$. First, $Y^G_1 = (|G|, m-|G|)$ if and only if all alternatives have same attribute 1 value. Then, $P[Y^G_1 = (|G|, m-|G|) | Y^G_0 = (|G|, m-|G|)] = p_h^{|G|} p_l^{m-|G|} + (1-p_h)^{|G|}(1-p_l)^{m-|G|}$. Second, $Y^G_1 = (j^b, j^b)$, $0 \le j^g \le |G|$, $0 \le j^b \le m-|G|$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (|G|, m-|G|)$ if and only if j^b of the $|G|$ good alternatives have attribute 1 value 1, the remaining $|G| - j^g$ good alternatives have attribute 1 value 0, j^b of the $m-|G|$ bad alternatives have attribute 1 value 1, and the remaining $m-|G|-j^b$ bad alternatives have attribute 1 value 0. Then, $P[Y^G_1 = (j^g, j^b) \mid Y^G_0 = (|G|, m-|G|)] = \binom{|G|}{j^g} p_h^{j^g}(1-p_h)^{|G|-j^g}\binom{m-|G|}{j^b} p_l^{j^b}(1-p_l)^{m-|G|-j^b}$, $0 \le j^b \le m-|G|$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (|G|, m-|G|)$. Finally, Y^G_1 cannot be $(0,0)$. Then, $P[Y^G_1 = (0,0) \mid Y^G_0 = (|G|, m-|G|)] = 0$.

Let $0 < r < k$. Assume $(i^g, i^b) = (0,0)$. Thus, $D^g_r = D^b_r = \emptyset$. Since $D^g_{r+1} \subset D^g_r$ and $D^b_{r+1} \subset D^b_r$, $D^g_{r+1} = D^b_{r+1} = \emptyset$, implying $P[Y^G_{r+1} = (0,0) \mid Y^G_r = (0,0) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = 1$ and $P[Y^G_{r+1} = (j^g, j^b) \mid Y^G_r =$

$(0,0) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = 0$, $0 \le j^g \le |G|$, $0 \le j^b \le m - |G|$, $(j^g, j^b) \ne (0,0)$. Assume $(i^g, i^b) \ne (0,0)$. Thus, $|D^g_r| = i^g$ and $|D^b_r| = i^b$. The values of $|D^g_{r+1}|$ and $|D^b_{r+1}|$ depend on $|D^g_r| = i^g$ and $|D^b_r| = i^b$ and the values of the attributes $r+1$ of the alternatives as follows. First, $D^b_{r+1} \subset D^b_r$ and $D^g_{r+1} \subset D^b_r$ imply $|D^g_{r+1}| \le |D^g_r| = i^g$ and $|D^b_{r+1}| \le |D^b_r| = i^b$ and, then, $P[Y^G_{r+1} = (j^g, j^b) \mid Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = 0$, $i^g \le j^g \le |G|$, $i^b \le j^b \le m - |G|$, $(j^g, j^b) \ne (i^g, i^b)$. Second, for $(i^g, i^b) \ne (|G|, m-|G|)$, $|D^g_{r+1}| = 0$ and $|D^b_{r+1}| = 0$ if and only if all alternatives in D^g_r have attribute $r+1$ value 0, all alternatives in D^b_r have attribute $r+1$ value 0, and some alternative in $\{1, 2, \ldots, m\} - D^g_r - D^b_r$ has attribute $r+1$ value 0. Then, $P[Y^G_{r+1} = (0,0) | Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = (1-p_h)^{i^g}(1-p_l)^{i^b}[1-(1-p_h)^{|G|-i^g}(1-p_l)^{m-|G|-i^b}]$. Third, $|D^g_{r+1}|$ and $|D^b_{r+1}|$ will have values j^g and j^b, respectively, $0 \le j^g \le i^g$, $0 \le j^b \le i^b$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (i^g, i^b)$ if and only if j^b alternatives in D^g_r have attribute $r+1$ value 1, the remaining $i^b - j^b$ alternatives in D^g_r have attribute $r+1$ value 0, j^b alternatives in D^b_r have attribute $r+1$ value 1, and the remaining $i^b - j^b$ alternatives in D^b_r have attribute $r+1$ value 0. Then, $P[Y^G_{r+1} = (j^g, j^b) \mid Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = \binom{i^g}{j^g} p_h^{j^g}(1-p_h)^{i^g-j^g}\binom{i^b}{j^b} p_l^{j^b}(1-p_l)^{i^b-j^b}$, $0 \le j^g \le i^g$, $0 \le j^b \le i^b$, $(j^g, j^b) \ne (0,0)$, $(j^g, j^b) \ne (i^g, i^b)$. Finally, $|D^g_{r+1}|$ will have value i^g and $|D^b_{r+1}|$ will have value i^b if and only if either all alternatives in $D^g_r \cup D^b_r$ have attribute $r+1$ value 1 or all alternatives have attribute $r+1$ value 0. Then, $P[Y^G_{r+1} = (i^g, i^b) \mid Y^G_r = (i^g, i^b) \wedge Y^G_{r-1} = (i^g_{r-1}, i^b_{r-1}) \wedge \cdots \wedge Y^G_1 = (i^g_1, i^b_1) \wedge Y^G_0 = (|G|, m-|G|)] = p_h^{i^g} p_l^{i^b} + (1-p_h)^{|G|}(1-p_l)^{m-|G|}$. □

Acknowledgements

We are grateful for helpful comments from the Decision Sciences group of INSEAD, the participants of the Barcelona Economics Decision Group, the ABC research group, Antonio Cabrales, Nathan Berg, and Konstantinos Katsikopoulus. We also appreciate the financial support received from the Spanish *Ministerio de Educación y Ciencia* under research grants SEC2003–03474 (R. M. Hogarth) and DPI2004–05077 (J. A. Carrasco) and from *Fundación BBVA* under research grant SOCIALES-02 (M. Baucells).

References

[1] Barron, F.H. 1992. Selecting a Best Multiattribute Alternative with Partial Information about Attribute Weights. *Acta Psychologica* **80** (1-3) 91–103.

[2] Brace, K. S., R. L. Rudell, and R. E. Bryant 1990. Efficient Implementation of a BDD Package. In *Proc. 27th ACM/IEEE Design Automation Conference* 40–57.

[3] Bröeder, A. 2000. Assessing the empirical validity of the 'Take-The-Best' heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning Memory, and Cognition*, **26** (5), 1332-1346.

[4] Bryant, R. E. 1986. Graph-based Algorithms for Boolean Function Manipulation. *IEEE Transactions on Computers* **C-35** (8) 677–691.

[5] CUDD: CU Decision Diagram Package Release 2.4.1. 2005. Available at http://vlsi.colorado.edu/ fabio/CUDD/cuddIntro.html.

[6] Czerlinski, J., G. Gigerenzer, and D. G. Goldstein 1999. How good are simple heuristics? In G Gigerenzer, P. M. Todd and the ABC Research Group. *Simple heuristics that make us smart* 97–118. Oxford University Press, New York.

[7] Dawes, R. M. and B. Corrigan 1974. Linear models in decision making. *Psychological Bulletin* **81** (2) 95–106.

[8] Einhorn, H. J. and R. M. Hogarth 1975. Unit Weighting Schemes for Decision Making. *Organizational Behavior and Human Performance* **13** (2) 171–192.

[9] Gigerenzer, G. and D. Goldstein 1996. Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review* **103** (4) 650–669.

[10] Gigerenzer, G., P. M. Todd, and the ABC Research Group 1999. *Simple heuristics that make us smart*. Oxford University Press, New York.

[11] Hogarth, R. M. and N. Karelaia 2003. Take-the-Best and other Simple Strategies: Why and When they Work in Binary Choice. DEE working paper no. 709. Universitat Pompeu Fabra, Barcelona.

[12] Hogarth, R. M. and N. Karelaia (in press). Simple Models for Multi-attribute Choice with Many Alternatives: When it Does and Does not Pay to Face Tradeoffs with Binary Attributes. *Management Science*.

[13] Hogarth, R. M. and N. Karelaia 2005. Ignoring Information in Binary Choice with Continuous Variables: When is less more? *Journal of Mathematical Psychology* **49** (2) 115–124.

[14] Hogarth, R. M. and N. Karelaia 2005. Regions of Rationality: Maps for Bounded Agents. DEE working paper no. 828. Universitat Pompeu Fabra, Barcelona.

[15] Hsee, C.K. and Y. Rottenstreich 2004. Music, Pandas, and Muggers: On the Affective Psychology of Value. *Journal of Experimental Psychology* **133** (1) 23–30.

[16] Katsikopoulos, K. and L. Martignon 2003. Paired comparisons: Ordering or adding attributes?. Working paper. Max-Planck Institute for Human Development, Berlin.

[17] Katsikopoulos, K. and B. Fasolo (in press). New Tools for Decision Analysts. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*.

[18] Keeney, R. L. and H. Raiffa 1993. *Decisions with multiple objectives: Preferences and value tradeoffs*. Cambridge University Press, Cambridge.

[19] Keeney, R. L. 2004. Making Better Decision Makers. *Decision Analysis* **1** (4) 193–204.

[20] Kirkwood, C. W. and R. K. Sarin 1985. Ranking with Partial Information: A Method and an Application. *Operations Research* **33** (1) 38–48.

[21] Martignon, L. and U. Hoffrage 1999. Why Does One-reason Decision Making Work? A Case Study in Ecological Rationality. In G Gigerenzer, P. M. Todd and the ABC Research Group. *Simple heuristics that make us smart* 119–140. Oxford University Press, New York.

[22] Martignon, L. and U. Hoffrage 2002. Fast, Frugal, and Fit: Simple Heuristics for Paired Comparison. *Theory and Decision* **52** 29-71.

[23] Newell, B. R. and D. R. Shanks 2003. Take the Best or Look at the Rest? Factors Influencing One-reason Decision Making. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29** (1) 53–65.

[24] Newell, B. R., N. J. Weston, and D. R. Shanks 2003. Empirical Tests of a Fast-and-Frugal Heuristic: Not Everyone Takes-the-Best. *Organizational Behavior and Human Decision Processes* **91** (1) 82–96.

[25] Nikolskaia, M., A. Rauzy, and D. J. Sherman 1998. Almana: A BDD Minimization Tool Integrating Heuristic and Rewriting Methods. In Springer-Verlag LNCS 1522, *Formal Methods in Computer-Aided Design, 2nd Int. Conf.* 100–114.

[26] Payne, J. W., Bettman, J. R. and E. J. Johnson. 1993. *The adaptive decision maker*. Cambridge University Press, New York.

[27] Tversky, A. 1972. Elimination by Aspects: A Theory of Choice. *Psychological Review* **79** (4) 281–299.