The optimality of coarse categories in decision-making and information storage

Michael Mandler
Department of Economics Royal Holloway College, University of London Egham, United Kingdom

An agent who lacks preferences and instead makes decisions using criteria that are costly to create should select efficient sets of criteria, where the cost of making a given number of choice distinctions is minimized. Under mild conditions, efficiency requires that binary criteria with only two categories per criterion are chosen. When applied to the problem of determining the optimal number of digits in an information storage device, this result implies that binary digits (bits) are the efficient solution, even when the marginal cost of using additional digits declines rapidly to 0. This short paper pays particular attention to the symmetry conditions entailed when sets of criteria are efficient.

1 Introduction

Suppose that agents, rather than forming a separate preference judgment for each pair of alternatives, make decisions using *criteria*. A criterion orders a small number of categories, each of which consists of many alternatives. The potential of a criterion to order alternatives within another criterion's categories allow sets of criteria to generate large numbers of choice distinctions. If an agent has objective preferences that can be inferred from a large set of sufficiently discriminating criteria, the agent will be better off if more of the criterion orderings are discovered: the agent will then be able to determine the optimal allocation from more choice sets. The uncovering of more criterion discriminations is costly, however, and we therefore consider *efficient* points where the cost of making a given number of choice distinctions is minimized.

This optimization problem seems to lead to a trade-off. Given a number of choice distinctions, an agent could either use a large set of coarse criteria (criteria with only a small number of categories) or a small set of finer, more discriminating criteria. We show under mild conditions that large sets of coarse criteria always lead to reductions in decision-making costs. Binary criteria with only two categories per criterion therefore provide the only efficient arrangement. Under mild restrictions on how criteria are aggregated into decisions, binary criteria lead to *rational* choice functions, where decisions are determined by a complete and transitive binary relation.

We apply our model to the problem of determining the optimal number of digits in an information storage device. We show that, even if the marginal cost of additional digits declines rapidly to 0, binary digits (bits) offer the efficient solution.

In this short paper, we pay particular attention to the symmetry conditions that are entailed when sets of criteria are efficient. A full working paper [2] is available on-line.

2 An outline of the model

A *criterion* C_i is an asymmetric binary relation on a domain of alternatives X and a *set of criteria* is denoted $\mathscr{C} = \{C_1, \ldots, C_N\}$. Two alternatives x and y are deemed C_i-equivalent if x and y share the same set

R. Ramanujam (Ed.): TARK 2015
EPTCS 215, 2016, pp. 227–230, doi:10.4204/EPTCS.215.16

of C_i-superior alternatives and the same set of C_i-inferior alternatives (see [1]). A C_i*-category* is a maximal set of C_i-equivalent alternatives and $e(C_i)$ denotes the number of C_i-categories. The *discrimination vector* of $\mathscr{C}$ is $(e(C_1),...,e(C_N))$. A C_i is *coarser* than C_i' if $e(C_i) < e(C_i')$.

Let c be a choice function on a domain of finite subsets of X. Two alternatives x and y are in the same *choice class* of c if c treats them interchangeably: first, when x is chosen and y is available then y is chosen too, and second, if x but not y is available then x is chosen if and only if, when y is available and not x, y is chosen.

A choice function c *uses* $\mathscr{C}$, denoted $(\mathscr{C},c)$, if c does not make distinctions that are not already present in the criteria: for each set of alternatives A that contains only alternatives that are in the same C_i-category, $i = 1,...,N$, there is a choice class of c that contains A.

Let $\kappa(C_i)$ denote the *cost of criterion* C_i. We assume $\kappa(C_i)$ is determined by the number of C_i-categories and therefore also write $\kappa(e)$ to denote the cost of a C_i with e categories. We assume that the *cost of a set of criteria*, $\kappa[\mathscr{C}]$, equals the sum of the costs of the criteria in $\mathscr{C}$. Letting $n(c)$ be the number of choice classes in c, a pair $(\mathscr{C},c)$ is *more efficient* than the pair $(\mathscr{C}',c')$ if $n(c) \geq n(c')$ and $\kappa[\mathscr{C}] \leq \kappa[\mathscr{C}']$, with at least one strict inequality, and $(\mathscr{C},c)$ is *efficient* if there does not exist a $(\mathscr{C}',c')$ that is more efficient than $(\mathscr{C},c)$.

The fundamental advantage of criteria is that each criterion can discriminate within the categories of other criteria. Given constraints that specify that criterion C_i can have no more than e_i categories (and assuming that $|X|$ is sufficiently large), we can find a $(\mathscr{C},c)$ such that (i) there is a partition of X with $\prod_{i=1}^{N} e_i$ cells such that x and y are in distinct cells if and only if they lie in different C_i-categories for at least one i and (ii) each cell of this partition forms a choice class of c. Subject to the e_i constraints, this $(\mathscr{C},c)$ maximizes $n(c)$ and accordingly we define $(\mathscr{C},c)$ to *maximally discriminate* if $n(c) = \min\left[\prod_{i=1}^{N} e(C_i), |X|\right]$.

3 Main results

(1) Since criteria with only one category make no discriminations and require no decisions, we assume they are costless. To compare a $(\mathscr{C},c)$ and $(\mathscr{C}',c')$ that have the same number of costly categories, suppose that $\sum_{i=1}^{N}(e(C_i)-1) = \sum_{i=1}^{N'}(e(C_i')-1)$. Assume also that either (i) the marginal cost of categories is increasing and the smaller of $n(c)$ and $n(c')$ is less than the cardinality of X or (ii) the marginal costs of categories is strictly increasing. We show that if $\mathscr{C}$ has greater proportions of coarser criteria than does $\mathscr{C}'$ and if $(\mathscr{C},c)$ maximally discriminates, then $(\mathscr{C},c)$ is more efficient than $(\mathscr{C}',c')$.

(2) Fix a set of domains that, for each finite m, contains a X with m elements and call a domain *admissible* if it is drawn from this set. Then, every efficient $(\mathscr{C},c)$ where the domain is admissible has a $\mathscr{C}$ that contains only binary criteria if and only if $\kappa(e) > \kappa(2)\lceil \log_2 e \rceil$ for all integers $e > 2$.

Thus the cost of e categories can rise as slowly as $\log_2 e$ – in which case the marginal cost of categories descends to 0 – and still the only efficient arrangement is for all criteria to be binary.

(3) We apply the result in (2) to information storage. Suppose we wish to store some integer between 1 and n using N k-ary digits and that the cost of storage equals $\kappa(k)N$. We show that, for all positive integers n, binary digits are the minimum-cost storage method if and only if $\kappa(k) > \kappa(2)\lceil \log_2 k \rceil$ for all integers $k > 2$.

(4) We specify axioms for how to aggregate sets of criteria into choice functions that generalize weighted voting. Suppose that the choice function c in the pair $(\mathscr{C},c)$ satisfies these axioms, that $\mathscr{C}$ contains only binary criteria, and that c satisfies the following Condorcet rule: if there is a x in a choice set A that is chosen by c from all $\{x,y\}$ with $y \in A$ then x is chosen from A too. Then c makes selections that maximize a complete and transitive binary relation. Given (2), we conclude that in a broad range of

cases, efficient decision-making is rational.

4 Symmetry and maximal categorization

Maximal discrimination is necessary for decision-making efficiency since otherwise $n(c)$ could be increased without an increase in costs. The key feature required for a $(\mathscr{C}, c)$ to maximally discriminate is that the following property of $\mathscr{C}$, called *maximal categorization*, is satisfied: the *discrimination partition* $\mathscr{P}$ of X that places x and y in distinct cells if and only if x and y lie in different C_i-categories for at least one i must have $\prod_{i=1}^{N} e(C_i)$ cells.

We will now see that if X is a product of attributes and each criterion orders a distinct attribute, then maximal categorization is satisfied and conversely if maximal categorization is satisfied then we can label alternatives so that X becomes a product of attributes. By joining this conclusion to result (2), that only binary criteria are efficient, we can describe efficient decision-making concisely: to be efficient agents must be able to describe the alternatives in X so that they form a product of attributes and each criterion must divide a distinct attribute into exactly two categories.

The simplest way to achieve maximal categorization is for X to be formed by a product of attributes and for each C_i to divide X into categories based only on attribute i. The domain X might be a set of cars, and the attributes might be colors, top speeds, and prices. A 'speed' C_i would then order cars based on the ranges of top speeds that C_i deems to be equivalent.

Formally, an *attribute* is a set X_i and N attributes define the domain of alternatives $X = \prod_{i=1}^{N} X_i$. We will say that a set of criteria $\mathscr{C}$ *is based on a product of attributes* if for each C_i there is a set X_i and a partition of $\{X_i^1, \ldots, X_i^{e(C_i)}\}$ of X_i such that the categories of C_i are the sets $X_i^j \times \left(\prod_{k \neq i} X_k\right)$, $j = 1, \ldots, e(C_i)$. So, if C_i is an ordering of cars by color then each X_i^j would represent a color and x and y would be placed into distinct C_i-categories if and only if the ith coordinates of x and y indicate different colors: $x_i \in X_i^j$ and $y_i \in X_i^k$ where $j \neq k$. The cells of the discrimination partition $\mathscr{P}$ would then be the $\prod_{i=1}^{N} e(C_i)$ sets $X_1^{j_1} \times \cdots \times X_N^{j_N}$ where, for each i, j_i is an integer between 1 and $e(C_i)$. Maximal categorization thus obtains.

This treatment assumes that X is a product space: for each possible combination of attributes (each possible color-speed-price combination), there is a corresponding element of X. But for maximal categorization it is enough that there is *some* alternative in X for each combination of attribute ranges specified by the criteria, that is, it is sufficient for X to be a subset of $\prod_{i=1}^{N} X_i$ such that each $X_1^{j_1} \times \cdots \times X_N^{j_N}$ intersects X.

A set of criteria $\mathscr{C}$ that is based on a product of attributes enjoys a wide-ranging symmetry property. Fix some C_i in $\mathscr{C}$, and consider a set $\mathscr{E}_{-i}$ formed by an arbitrary union of the categories of the remaining criteria C_j, $j \neq i$. Given the product structure of $\mathscr{C}$, any such $\mathscr{E}_{-i}$ must intersect each of the C_i-categories. To continue the car example, the set of cars $\mathscr{E}_{-i}$ defined by a certain range of top speeds and prices can be partitioned into all the possible color subsets, say blue, red, and yellow. If we use C_i to order the cells of the color partition of the cars in $\mathscr{E}_{-i}$, the ordering will have the same 'shape' as – be order isomorphic to – the original color ordering C_i of X. If, for example, C_i on X is a cycle – blue is better than red which is better than yellow which is better than blue – then the C_i ordering of any set of cars defined by a range of speeds and prices will also form a cycle. We conclude that any two sets of cars Y and Z defined by selections of non-color attributes will be order isomorphic to each other when each is endowed with the color ordering C_i (or rather the restrictions of C_i to Y and Z).

This symmetry property may seem to be of limited value since it appears to apply only to products of attributes. But in fact the symmetry property characterizes any $\mathscr{C}$ that maximally categorizes. If for an arbitrary (possibly nonproduct) $\mathscr{C}$, we apply C_i to some $\mathscr{E}_{-i}$ and it defines fewer than $e(C_i)$ C_i-category subsets then the discrimination partition $\mathscr{P}$ would have to contain fewer than $\prod_{i=1}^N e(C_i)$ cells. And the only way that $\mathscr{E}_{-i}$ and $\mathscr{E}'_{-i}$ can each define $e(C_i)$ C_i-category subsets is for the C_i ordering of these subsets to be order isomorphic.

Moreover, if an arbitrary (possibly nonproduct) $\mathscr{C}$ enjoys the symmetry property we can relabel the elements of the domain X so that $\mathscr{C}$ is then based on a product of attributes. To do this, we associate each C_i with an attribute (e.g., color) and identify each C_i-category with an arbitrary value X_i^j for that attribute (e.g., blue): each cell of $\mathscr{P}$ is thus identified with a vector of attribute values. So, although a product of attributes looks special, it provides a model for any set of criteria that maximally categorizes.

The following definitions and theorem make these claims precise. We use $E_i^1,\ldots,E_i^{e(C_i)}$ to denote the categories of criterion C_i.

Given the set of criteria $\{C_1,\ldots,C_N\}$, $\mathscr{E}_{-i}$ is a *union of C_{-i}-categories* if $\mathscr{E}_{-i} = \bigcup_j E_k^j$ for some collection of criterion categories $\{E_k^j\}$ such that $k \neq i$ for each j. Let $C_i^{\mathscr{E}_{-i}}$ denote the binary relation defined by $E\,C_i^{\mathscr{E}_{-i}}E'$ if and only if there are C_i-categories E_i and E_i' such that $E = E_i \cap \mathscr{E}_{-i}$, $E' = E_i' \cap \mathscr{E}_{-i}$, and $xC_i\,y$ for $x \in E$ and $y \in E'$. We then say $\mathscr{C}$ satisfies the *order-isomorphism property* if for any i and any two unions of C_{-i}-categories, $\mathscr{E}_{-i}$ and $\mathscr{E}'_{-i}$, the binary relations $C_i^{\mathscr{E}_{-i}}$ and $C_i^{\mathscr{E}'_{-i}}$ are order-isomorphic.

The set of criteria $\mathscr{C}$ has a *product representation* if (i) for each i, there is a nonempty set Y_i and a partition $\{Y_i^1,\ldots,$
$Y_i^{e(C_i)}\}$ of Y_i, (ii) there is a set of criteria $\widehat{\mathscr{C}} = \{\widehat{C}_1,\ldots,\widehat{C}_N\}$ defined on $Y = \prod_{i=1}^N Y_i$ where the categories of each $\widehat{C}_i$ are the sets $Y_i^j \times \left(\prod_{k\neq i} Y_k\right)$, and (iii) for each i, there is a order-preserving bijection f between the categories of C_i and $\widehat{C}_i$, that is, $E_i^j C_i E_i^{j'}$ if and only if $f\left(E_i^j\right)\widehat{C}_i\, f\left(E_i^{j'}\right)$.

We then have the following result.

Theorem. For a set of criteria $\mathscr{C}$, the following statements are equivalent: (i) $\mathscr{C}$ maximally categorizes, (ii) $\mathscr{C}$ satisfies the order-isomorphism property, (iii) $\mathscr{C}$ has a product representation.

5 Bibliography

References

[1] Mandler, M., 2015, Rational agents are the quickest. J. Econ. Theory 155: 206-233, doi:10.1016/j.jet.2014.10.003.

[2] Mandler, M., 2014, Coarse, efficient decision-making. Available at SSRN http://ssrn.com/abstract=2494600.