Soc Choice Welfare (2008) 30:419–425
DOI 10.1007/s00355-007-0239-y

ORIGINAL PAPER

A distance measure for choice functions

Christian Klamler

Received: 31 October 2006 / Accepted: 28 March 2007 / Published online: 16 May 2007

Abstract This paper discusses and characterizes a distance function on the set of quasi choice functions. The derived distance function is in the spirit of the widely used Kemeny metric on binary relations but extends Kemeny's use of the symmetric difference distance to set functions and hence to a more general model of choice.

1 Introduction

The aggregation of individual preferences into a group preference is of major interest in areas such as social choice theory and computer science. Certain aggregation rules and many of the comparisons between different aggregation rules rely on the idea of measuring distances or the similarity between such preferences. In mathematics many different distance functions have been devised to measure distances between pairs of objects in many different domains. A well-known distance in that respect which is also commonly used in social choice theory is the symmetric difference distance, i.e. the cardinality of the symmetric difference of two sets of objects. As preferences are often represented by binary relations and those are nothing else than sets of ordered pairs, a natural way of measuring distance between preferences is to use the symmetric difference distance on the set of binary relations. This is what actually has been done and characterized by Kemeny (1959). Intuitively the Kemeny distance counts the minimal number of (pairwise) "inversions" of alternatives necessary to transform one binary relation into the other. Kemeny's original focus was on complete and transitive binary relations, i.e. weak orders. However, soon other authors applied the symmetric difference distance to other preference structures. In particular, Bogart (1973) characterized a distance on partial orderings, i.e. transitive and irreflexive relations, whereas

C. Klamler (✉)
Institutes of Economics and Public Economics, University of Graz, Graz, Austria
e-mail: christian.klamler@uni-graz.at

Bogart (1975) extended this even further to not necessarily transitive relations. Other applications of the symmetric difference distance can be found in Mirkin and Chernyi (1970) who applied it to equivalence relations and in Margush (1982) who applied it to tree structures.[1]

Distances are also explicitly used in aggregation procedures which try to find median relations. One example is the Kemeny procedure whose median relation is the relation that minimizes the sum of the Kemeny distances to the individual preferences. A detailed discussion of this "median principle" has been provided by Barthelemy and Monjardet (1981).

The focus in this paper is on a more general model of choice by using choice functions instead of binary relations to represent preferences. We will provide a natural extension of the symmetric difference distance to set functions and provide a characterization of a distance function which measures the distance between pairs of quasi choice functions.

This is of interest insofar as choice functions are seen as an attractive way of dealing with aggregation problems whenever little structure is imposed on individual and/or group preferences (Xu 1996). As in certain situations the transitivity and/or completeness of preferences seems to be a strong assumption, using choice functions as primitives seems to be a reasonable alternative. A detailed discussion of the importance and the advantages of choice functions can be found in Aizerman and Aleskerov (1995) and Aleskerov and Monjardet (2002). The idea of measuring distance between choice functions has been introduced already in connection to convexity issues and the aggregation of individual choice functions by Albayrak and Aleskerov (2000) and Ilyunin et al. (1988). Distance aspects in social choice theory have also been discussed in a broader context by Nurmi (2002, 2004). In a more applied form, Brams et al. (2006) used a distance approach to analyse the 2003 Game Theory Society council election.

2 The characterization result

Let X be a finite set of m alternatives. The set of all non-empty subsets of X is denoted by K. A *quasi choice function* is a function $C : K \to K \cup \emptyset$ such that for all $S \in K$, $C(S) \subseteq S$, i.e. it assigns to any set $S \in K$ a (possibly empty) subset $C(S) \subseteq S$. The set of all quasi choice functions on domain K is denoted by $\hat{\mathcal{C}}$. A certain idea of a choice function lying "between" other choice functions or, based on this "betweenness", choice functions being "on a line" will be used. Following Albayrak and Aleskerov (2000) we give the following definitions:

Definition 1 For any $C, C', C'' \in \hat{\mathcal{C}}$, we say that C' lies between C and C'', written $[C, C', C'']$ if for all $S \in K$, $C(S) \cap C''(S) \subseteq C'(S) \subseteq C(S) \cup C''(S)$.

Definition 2 The quasi choice functions $C_1, C_2, \ldots, C_n \in \hat{\mathcal{C}}$ are on a line if for all $i < j < k \leq n$, C_j lies between C_i and C_k.

A distance function on set $\hat{\mathcal{C}}$ is a function $d : \hat{\mathcal{C}} \times \hat{\mathcal{C}} \to \mathbb{R}_+$. Consider the following properties for distance functions on $\hat{\mathcal{C}}$.

[1] Further applications and references can be found in Bogart (1982).

A1.1 $d(C, C') \geq 0$ where equality holds if and only if $C = C'$
A1.2 $d(C, C') = d(C', C)$
A1.3 $d(C, C'') \leq d(C, C') + d(C', C'')$ and equality holds if and only if C' is between C and C''
A2 If $\tilde{C}, \tilde{C}'$ result from C, C' by a permutation of the alternatives, then $d(C, C') = d(\tilde{C}, \tilde{C}')$
A3 If two choice functions $C, C' \in \hat{\mathcal{C}}$ agree except for a set $\bar{K} \subset K$ which is part of the domain in both choice functions, then the distance $d(C, C')$ is determined exclusively from the choice sets over $\bar{K}$.
A4 Let four choice functions $C, C', \tilde{C}, \tilde{C}' \in \hat{\mathcal{C}}$ disagree only on set $T \in K$ such that for some $S \subseteq T$, $C(T) = \tilde{C}(T) \cup S$, and $C'(T) = \tilde{C}'(T) \cup S$. Then the distance between C and C' should be equal to the distance between $\tilde{C}$ and $\tilde{C}'$.
A5 The minimal positive distance is 1.

Axioms A1.1 to A1.3 are the usual metric axioms with the addition of A1.3 using the idea of "betweenness" (as previously defined). Axiom A2 is a usual neutrality condition saying that if we permute the alternatives in any pair of quasi choice functions, the distance between those quasi choice functions does not change. Axiom A3 is some kind of separability condition as the distance between quasi choice functions is only based on sets on which they differ. Moreover, A4 is a sort of translation invariance condition, i.e. distances are invariant under special types of parallel translations. Finally, A5 can be seen as choosing a unit of measurement.

Our aim now is to show that the above axioms uniquely determine a distance between quasi choice functions. The first lemma shows that if n quasi choice functions are on a line, then the distance between the first and the nth quasi choice function is equal to the sum of the distances between all pairs of adjacent quasi choice functions.

Lemma 1 *If $C_1, C_2, \ldots, C_n \in \hat{\mathcal{C}}$ are on a line, then $d(C_1, C_n) = d(C_1, C_2) + d(C_2, C_3) + \cdots + d(C_{n-1}, C_n)$.*

Proof The lemma is proved by induction. Therefore we will repeatedly apply Axiom A1.3. For $n = 2$ it is of course trivially true that $d(C_1, C_2) = d(C_1, C_2)$. Hence assume that it is true for $n = k$ and let us show that it is also true for $k+1$. As the quasi choice functions are on a line it is the case that $[C_1, C_k, C_{k+1}]$. Therefore from Axiom A1.3 we know that $d(C_1, C_{k+1}) = d(C_1, C_k) + d(C_k, C_{k+1})$. However, applied over the first k quasi choice functions, we know that $d(C_1, C_k) = d(C_1, C_2) + d(C_2, C_3) + \cdots + d(C_{k-1}, C_k)$ and thus the lemma follows by induction. □

Lemma 2 *If three quasi choice functions $C, C', C'' \in \hat{\mathcal{C}}$ only differ in the choice over a set $S \subseteq X$ such that $C(S) = \emptyset$, $C'(S) = \{x\}$ and $C''(S) = \{y\}$, then $d(C, C') = d(C, C'')$.*

Proof Let us slightly abuse the notation and instead of using quasi choice functions use only the choice sets on the set S, i.e. if $C(S) = \emptyset$ and $C'(S) = \{x\}$ then $d(C, C')$ will equivalently be written as $d(\emptyset, \{x\})$. Axiom A3 implies that the distance between the quasi choice functions is determined exclusively from the choice sets on S. Then, permuting x and y, we get—from A2—that $d(\emptyset, \{x\}) = d(\emptyset, \{y\})$. □

Now, since $d(\emptyset, \{x\})$ does not depend on x, we can set $d(\emptyset, \{x\})$ equal to a constant $u = 1$ (by A5). We now show that the distance between any two quasi choice functions that only differ in the choice over one subset—such that one chooses the empty set and the other chooses the whole set—will be a multiple of u.

Lemma 3 *If two quasi choice functions $C, C' \in \hat{\mathcal{C}}$ only differ in the choice over a set $S \subseteq X$ such that $|S| = n$, $C(S) = \emptyset$ and $C'(S) = S$, then $d(C, C') = n \cdot u$.*

Proof Using the same notational abuse as in lemma 2 we know that $d(\{x\}, \emptyset) = d(\{y\}, \emptyset) = u$. Using axiom A4 we get $d(\{x, y\}, \{y\}) = d(\{x\}, \emptyset) = u$. As $\{x, y\}$, $\{y\}, \emptyset$ are on a line it follows from lemma 1 that $d(\{x, y\}, \emptyset) = d(\{x, y\}, \{y\}) + d(\{y\}, \emptyset) = 2u$. Assume that this holds for $|S| = k$, then we need to show that it holds for $|S| = k+1$. From lemma 1, $d(\{x_1, x_2, \ldots, x_k, x_{k+1}\}, \emptyset) = d(\{x_1, x_2, \ldots, x_k\}, \emptyset) + d(\{x_1, x_2, \ldots, x_{k+1}\}, \{x_1, x_2, \ldots, x_k\})$. By A4, $d(\{x_1, x_2, \ldots, x_{k+1}\}, \{x_1, x_2, \ldots, x_k\}) = d(\{x_{k+1}\}, \emptyset) = u$. Hence, if $d(\{x_1, x_2, \ldots, x_k\}, \emptyset) = k \cdot u$ then $d(\{x_1, x_2, \ldots, x_{k+1}\}, \emptyset) = (k + 1) \cdot u$. □

The final lemma generalizes lemma 3 in the sense that there now is no restriction whatsoever on the choice sets of the two quasi choice functions for the one subset on which they differ.

Lemma 4 *If two quasi choice functions in $C, C' \in \hat{\mathcal{C}}$ only differ in the choice over a set $S \subseteq X$ then their distance is based on the cardinality of their symmetric difference, i.e. $d(C, C') = \left|C(S)\Delta C'(S)\right| \cdot u$.*

Proof Let $C(S) = \emptyset$ and consider that $C'(S) \cap C''(S) = \emptyset$. Then, by the definition of betweenness, we know that C', C, C'' are on a line. Now let $C'(S) = \{x_1, x_2, \ldots, x_k\}$ and $C''(S) = \{y_1, y_2, \ldots, y_h\}$ Then by lemma 3 it is the case that $d(C, C') = k$ and $d(C, C'') = h$. Hence by A3 and lemma 1, $d(C', C'') = (k+h) \cdot u = \left|C'(S)\Delta C''(S)\right| \cdot u$. If $C'(S) \cap C''(S) = T \neq \emptyset$ then by A4 it follows that for any $\tilde{C}', \tilde{C}''$ such that $C'(S) = \tilde{C}'(S) \cup T$ and $C''(S) = \tilde{C}''(S) \cup T$, $d(C', C'') = d(\tilde{C}', \tilde{C}'')$ and therefore the lemma is true. □

Based on the symmetric difference of choice sets on different sets $S \in K$ we will now define the following distance function[2] on the set of all quasi choice functions, $\hat{\mathcal{C}}$.

Definition 3 For any $C, C' \in \hat{\mathcal{C}}$, $d_F(C, C') = \sum_{S \in K} \left|C(S)\Delta C'(S)\right|$.

Hence, d_{F} measures the distance between two quasi choice functions as the sum of the cardinalities of the symmetric differences over all subsets S in K. As turns out and is shown in the following theorem, d_{F} is the only distance function that satisfies the reasonable axioms discussed above.

Theorem 1 *A distance function d on $\hat{\mathcal{C}}$ is equal to d_F if and only if it satisfies the axioms* A1–A5.

[2] This distance function has previously been used in papers by Ilyunin et al. (1988) and Albayrak and Aleskerov (2000).

Table 1 Binary relations

R_1	R_2	R_3
x	y	x
y	x	z
z	z	y

Proof Consider two different quasi choice functions $C, C' \in \hat{\mathcal{C}}$. We can transform C into C' via a sequence of quasi choice functions $C_1, C_2, C_3, \ldots, C_{k-1}, C_k \in \hat{\mathcal{C}}$ with $C_1 \equiv C$ and $C_k \equiv C'$, where for each pair C_i, C_{i+1}, $i \in \{1, \ldots, k-1\}$, there is a different set $S_i \in K$ such that for all $T \in K \backslash S_i$, $C_i(T) = C_{i+1}(T)$, i.e. the choice functions only differ in the choice over one particular subset of X. This, however, means that $C_1, C_2, \ldots, C_k$ are on a line and hence, from lemma 1 it is the case that $d(C_1, C_k) = d(C_1, C_2) + d(C_2, C_3) + \cdots + d(C_{k-1}, C_k)$. However, for any pair C_i, C_{i+1}, the distance $d(C_i, C_{i+1})$ is determined by lemma 4, i.e. $d(C_i, C_{i+1}) = |C_i(S_i) \Delta C_{i+1}(S_i)| \cdot u$. Therefore $d(C_1, C_k) = \sum_{S \in K} |C_1(S) \Delta C_k(S)| \cdot u$. As this shows that all distance values are multiples of u, axiom A5 gives u=1. Hence, $d = d_{\mathrm{F}}$ and obviously d_{F} satisfies axioms A1–A5. □

3 The relationship between d_{F} and the Kemeny metric

As d_{F} is defined on the very general class of quasi choice functions it allows to measure distance between preferences with very little structure. Of course, using the concept of "rationalizability"[3] of choice functions we can transform preferences represented by binary relations into preferences represented by choice functions. The question now arises whether d_{F} provides the same distance information when applied on choice functions as the Kemeny metric does when applied on binary relations that rationalize those choice functions. It turns out, that this is not the case. Consider the following example (Klamler 2006):

Example 1 Let $X = \{x, y, z\}$ and R_1, R_2, R_3 represent three binary relations as stated in Table 1 (where less preferred alternatives are in lower rows).

The corresponding choice functions C_1, C_2, C_3 are given in Table 2 (for notational convenience the values of C on singletons are ommitted).

The Kemeny distance between R_1 and R_2 is the number of inversions necessary to transform R_1 into R_2 multiplied by 2. As to do so it takes exactly the inversion of alternatives x and y in R_1, the Kemeny distance between R_1 and R_2 is 2. The same distance is derived between R_1 and R_3 where we need to invert alternatives y and z. However, if we consider the distance between the choice functions that are rationalized by those binary relations we get $d_{\mathrm{F}}(C_1, C_2) = 4$ and $d_{\mathrm{F}}(C_1, C_3) = 2$.

In a nutshell, this difference in distance information is based on the fact that d_{F} attaches more "weight" to changes in the preference relation with respect to more

[3] On rationalizability of choice functions see Sen (1986).

Table 2 Choice functions rationalized by the linear orders in Table 1

	$C_1(\cdot)$	$C_2(\cdot)$	$C_3(\cdot)$
X	x	y	x
xy	x	y	x
xz	x	x	x
yz	y	y	z

preferred alternatives. A higher ranked alternative is contained in a larger number of choice sets than a lower ranked alternative and therefore any change of such an alternative in the relation would lead to more changes in choice sets than would occur in the case of a lower ranked alternative. This is represented in d_F. The Kemeny metric, on the other hand, is neutral with respect to where in the relation such changes occur.

Given this fact, d_F does indeed sound reasonable as in many situations changes in more preferred alternatives seem to be considered more relevant in both, an individual and social view, than changes in less preferred alternatives.

Of course, the Kemeny metric on binary relations could be adapted in such a way that those "weights" are taken into account. Such a distance function on binary relations has been presented in Klamler (2006).

4 Conclusion

In this paper we have characterized a distance function on the set of quasi choice functions based on the symmetric difference distance. In addition we have shown in what sense it differs from the well known Kemeny metric on binary relations.

Finally, the question arises in what sense such a distance function can lead to further insight into non-binary aggregation rules. Rules such as those devised by Dodgson (1876), Kemeny (1959) and Slater (1961) are based on binary relations and all explicitly depend on the Kemeny metric.[4] The distance function characterized in this paper seems to open the possibility to transfer the above aggregation rules into the space of quasi choice functions and therefore enables distance-based aggregation in more general models of choice.

Acknowledgments I am grateful to Fuad Aleskerov, Hannu Nurmi, the participants of the "11th Osnabrück Seminar on Individual Decision and Social Choice" and especially to the corresponding editor Bernard Monjardet and two anonymous referees for their comments that substantially improved the paper. All remaining errors are exclusively mine.

References

Aizerman M, Aleskerov F (1995) Theory of choice. Elsevier, Amsterdam

Albayrak SR, Aleskerov F (2000) Convexity of choice function sets. In: Bogazici University Research Paper, ISS/EC-2000-01 (2000)

Aleskerov F, Monjardet B (2002) Utility maximization, choice and preference. Springer, Berlin

[4] The relationship between those aggregation rules has been discussed in Klamler (2004).

Barthelemy JP, Monjardet B (1981) The median procedure in cluster analysis and social choice theory. Math Soc Sci 1:235–268

Bogart KP (1973) Preference structures I: distances between transitive perference relations. J Math Soc 3:49–67

Bogart KP (1975) Preference structures II: distances between asymmetric relations. SIAM J Appl Math 29:254–262

Bogart KP (1982) Some social science applications of ordered sets. In: Rival I, Reidel D (eds) Ordered sets, Dordrecht, pp 759–787

Brams SJ, Kilgour DM, Sanver MR (2006) A minimax procedure for electing committees. mimeo NYU

Dodgson C (1876) A method of taking votes on more than two issues. In: Black D (ed) (1958). The Theory of Committees and Elections. Cambridge University Press, London

Ilyunin OK, Popov BV, El'kin LN (1988) Majority functional operators in voting theory. Autom Remote Control 7:137–145

Kemeny J (1959) Mathematics without numbers. Daedalus 88:571–591

Klamler C (2004) The Dodgson ranking and its relation to Kemeny's method and Slater's rule. Soc Choice Welf 23:91–102

Klamler C (2006) On some distance aspects in social choice theory. In: Simeone B, Pukelsheim F (eds), Mathematics and democracy: recent advances in voting systems and collective choice. Springer, Berlin

Margush T (1982) Distances between trees. Discr Appl Math 4:281–290

Mirkin BG, Chernyi LB (1970) Measurement of the distance between distinct partitions of a finite set of objects. Autom Tel 5:120–127

Nurmi H (2002) Voting procedures under uncertainty. Springer, Berlin

Nurmi H (2004) A comparison of some distance-based choice rules in ranking environments. Theory Decis 57:5–24

Sen A (1986) Social choice theory. In: Arrow KJ, Intrilligator MD (eds) Handbook of mathematical economics, vol III, Chap 22. North Holland, Amsterdam, pp 1073–1181

Slater P (1961) Inconsistencies in a schedule of paired comparisons. Biometrica 48:303–312

Xu Y (1996) Non binary social choice: a brief introduction. In: Schofield N (ed) Collective decision-making: social choice and political economy. Kluwer, Boston