Profile Ranking Adaptive Choice-Based Conjoint Analysis: A Complementary Approach to Utility-Based Analysis for Small Populations

Abstract

To analyze Adaptive Choice-Based Conjoint (ACBC) survey samples from small populations, a new methodology called profile ranking based ACBC (PR-ACBC) is proposed as a complement to utility based ACBC. PR-ACBC offers a form of validation especially useful for small survey data with high variances in partworth utilities. Without requiring knowledge of partworth utilities, PR-ACBC deduces from sample data the maximum likelihood population ranking for known population sizes using a multivariate hypergeometric distribution, and for unknown population sizes using the Lagrange multiplier based optimization. A population ranking interval can easily be computed for each profile as an interval in which the population profile ranking must lie. Various distance measures from statistical ranking theory are used to analyze profile decomposition (importance of attribute levels in choice tasks), and multidimensional scaling (MDS) for visual representation of attribute importances of particular value in comparing sample sub-groups. The methodology of PR-ACBC is introduced using a toy survey, and its application by a recent survey administered to faith-based and non-faith based disaster relief organizations belonging to the National Voluntary Organizations Active in Disaster (VOAD).

1 Introduction

Adaptive Choice-Based Conjoint (ACBC) analysis surveys are a widely-utilized, well-developed and highly effective type of conjoint analysis (Orme and Chrzan, 2017). While the Max-Diff approach to select the best and worst among several profiles has generated a great deal of research interest, we focus on ACBC surveys whose choice tasks are designed with the simplest choice between just two concepts. Following Sawtooth's Lighthouse survey creation tool, we structure our choice-task stage as a tournament beginning with 16 profiles close to the respondent's #1 profile called "Build Your Own" (BYO). For large samples, use of a sophisticated statistical method such as hierarchical-Bayesian Markov-Chain Monte-Carlo (HB MCMC) simulation (Rossi et. al. 2005) is extremely effective to estimate partworth utilities and their variances. In the case of very small samples (eg. $n \leq 15$) from a small population (eg. $N \leq 50$), the variances in partworth utility may hamper both accurate prediction of choice experiments and ranking of attribute importances. Profile ranking (PR-) ACBC is therefore introduced as a method serving as a validity check for profile choice prediction based on partworth levels, as well as attribute importance rankings derivable from the partworth utilities. PR-ACBC utilizes distance measures from statistical ranking theory for attribute decomposition as an alternative to partworth utilities for choice predictions, and multi-dimensional scaling (MDS) as a means to assess attribute importances (Alvo and Yu 2014).

In Section 2, we introduce basic PR-ACBC methodology by means of a generic survey with only 4 profiles constructed from 2 attributes each with 2 levels. We begin with a fundamental observation that the exact sample profile rankings directly obtainable from choice tournament data can not be predicted by multiple linear regression of part-worth utilities. PR-ACBC then proceeds to analyze survey tournament data without requiring any knowledge of partworth utilities. Maximum likelihood estimate (MLE) population rankings for known population sizes are obtainable by discrete multivariate hypergeometric distribution (Oberhofer and Kaufman 1987), and for unknown population sizes by multivariable calculus optimization using Lagrange multipliers (Stewart 2016). Population ranking intervals (PRIs) are easily computed from sample tournament data as one-dimensional intervals which must contain the population profile rankings. Statistical ranking theory distance measures are then used for profile attribute decomposition (akin to partworth levels) and multi-dimensional scaling (MDS) for attribute importances. These are important for comparing and contrasting sample subgroups. In Section 3 we illustrate PR methodology using a recent ACBC survey deployed to both faith-based and non-faith based disaster relief organizations. The context motivating this methodological study is a sequel to a novel application of ACBC in disaster-response research (Gralla et. al. 2014).

2 PR-ACBC Methodology

2.1 Simple Example

Consider a generic ACBC survey with just 2 attributes each having 2 levels. We designate the 4 possible profiles $A = 11, B = 12, C = 21, D = 22$, where $X = x_1 x_2$ designates that profile X has level x_1 for the first attribute and level x_2 for the second attribute. Suppose we have obtained by anonymous survey choice tournament results for $n = 4$ respondents from a population of size $N > 4$. A sample tournament outcome is shown in Figure 1
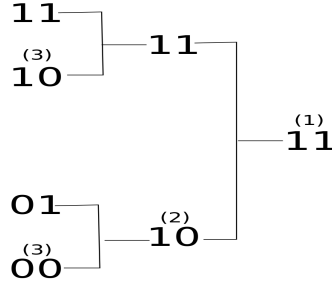

Figure 1: One possible tournament outcome ranks profile A=11 first, B=12 second, C=21 is ranked third and D=22 fourth.

For this respondent, profile A=11 is ranked 1 (tournament winner), profile B=10 is ranked 2 (runner-up), profiles C=01 third and D=00 ranked 4. The rationale for the latter 2 rankings is that assuming transitivity in match outcomes, the highest C could be ranked if all profiles were paired is 2, while D could only be ranked as high as 3. Note that there are essentially 24 different possible tournament outcomes, each specifying a ranking or permutation of A,B, C and D.(Table 1):

	Rank					Rank			
Outcome	1	2	3	4	Outcome	1	2	3	4
O_1	A	B	C	D	O_{13}	C	A	B	D
O_2	A	B	D	C	O_{14}	C	A	D	B
O_3	A	C	B	D	O_{15}	C	B	A	D
O_4	A	C	D	B	O_{16}	C	B	D	A
O_5	A	D	B	C	O_{17}	C	D	A	B
O_6	A	D	C	B	O_{18}	C	D	B	A
O_7	B	A	C	D	O_{19}	D	A	B	C
O_8	B	A	D	C	O_{20}	D	A	C	B
O_9	B	C	A	D	O_{21}	D	B	A	C
O_{10}	B	C	D	A	O_{22}	D	B	C	A
O_{11}	B	D	A	C	O_{23}	D	C	A	B
O_{12}	B	D	C	A	O_{24}	D	C	B	A

Table 1: The 24 possible tournament outcomes or rankings of the profiles A,B,C, and D. Note the lexicographic ordering of these rankings.

Each of the $n = 4$ sample respondents will have a tournament outcome, which we can compile into an *Individual Ranking Table (IRT)*, an example of which is shown in Table 2.

Respondent	Rank				Ranking
	1	2	3	4	
1	A	B	C	D	O_1
2	A	C	B	D	O_3
3	B	C	A	D	O_9
4	A	C	B	D	O_3

Table 2: Profile rankings for a sample of 4 respondents.

From this data, it is easy to compute the *Sample Profile Ranking (SPR)* table which indicates how many respondents assigned a particular rank to each profile (Table 3).

Profile	Rank				SPR
	1	2	3	4	
A	3	0	1	0	1.5
B	1	1	2	0	2.25
C	0	3	1	0	2.25
D	0	0	0	4	4.0

Table 3: Sample Profile Ranking (SPR) Table. The i, j entry of the 4x4 matrix forming the body of the table shows the number of respondents who assigned to the profile in the margin of row i, the rank shown by the header in column j. The sample profile ranking (SPR) is the arithmetic average of the sample respondent rankings.

Given any profile X and a sample of size $n = 4$, the *sample profile ranking (SPR)* $\rho_n(X) = \rho_4(X)$ is obtained as an arithmetic average of the $n = 4$ respondents. For example, for A=11, $\rho_4(A) = [3(1) + 1(3)]/4 = 1.5$, and for C=10 it is $\rho_4(C) = [3(2) + 1(3)]/4 = 2.25$. In this case, based on the sample, profile A=11 is ranked first, profiles B=10 and C=01 are tied for second, and profile D=00 comes in fourth.

In a similar way, we can use our sample to obtain rankings for each individual attribute level as shown in Table 4.

Attribute	Level	Rank				SALR
		1	2	3	4	
1	1	4	1	3	0	1.875
1	0	0	3	1	4	3.125
2	1	3	3	2	0	1.875
2	0	1	1	2	4	3.125

Table 4: Sample Attribute Level Ranking (SAL) Table.

The main questions we will develop in the sequel are:

- POPULATION RANKING INFERENCE *what can we infer from the SRT and SAL tables about the population profile ranking (PPR) and population attribute level (PAL) tables?*, the latter two meaning the tables for all population respondents; and

- CHOICE TASK PREDICTION AND ATTRIBUTE IMPORTANCE *for the purpose of choice task predictions and attribute importance rankings, how does information obtainable from the SRT and SAL compare with that obtainable from partworth utilities?*

2.2 A Fundamental Observation

In this section we show that least squares multiple regression (LSRM) will not in general give exact sample profile rankings. In other words, part-worth utilities can only approximate sample rankings.

Least squares multiple linear regression (LSRM) can be used to predict sample profile rankings based on individual sample outcomes as we will now explain using our simple 2 attribute, 2 level survey and generic sample data for 4 respondents. Table 5 gives the dataset $\{(U_i, V_i, Y_i)\}$ $(i = 1, ..., 16)$ where

$$
\begin{aligned}
U_i &= \quad 1 \text{ if attribute 1 has level 1, and 0 if it has level 2} \\
V_i &= \quad 1 \text{ if attribute 2 has level 1, and 0 if it has level 2} \\
Y_i &= \quad \text{Respondent's ranking of a profile with } U = U_i, V = V_i.
\end{aligned}
$$

Output Ranking Data

U	V	Respondent 1	Respondent 2	Respondent 3	Respondent 4
1	1	Y_1	Y_2	Y_3	Y_4
1	0	Y_5	Y_6	Y_7	Y_8
0	1	Y_9	Y_{10}	Y_{11}	Y_{12}
0	0	Y_{13}	Y_{14}	Y_{15}	Y_{16}

Table 5: Dataset ranking structure by respondents.

This dataset has certain properties:

- Each column consists of a respondent's profile rankings, and hence contains the values $1, 2, 3, 4$ in any order.

- Table 5 can also be represented in the form of Table 6, by which we see that

$$
\sum U_i = \sum V_i = \sum U_i^2 = \sum V_i^2 = 8,
$$

and

$$
\sum U_i V_i = 4,
$$

where the symbol $\sum$ represents $\sum_{n=1}^{16}$.

Using least squares multiple linear regression (LSMR) on the dataset in Table 6, we estimate each sample profile ranking Y_i as $\hat{Y}_i$:

$$
\hat{Y}_i = c_0 + c_1 U_i + c_2 V_i,
$$

where the regression coefficients c_0, c_1, c_2 are determined by minimizing the sum of squared residuals (SSR):

$$
SSR = \sum_{n=1}^{16}(Y_i - \hat{Y}_i)^2 = \sum_{n=1}^{16}(Y_i - (c_0 + c_1 U_i + c_2 V_i))^2.
$$

U	V	Rank
1	1	Y_1
1	1	Y_2
1	1	Y_3
1	1	Y_4
1	0	Y_5
1	0	Y_6
1	0	Y_7
1	0	Y_8
0	1	Y_9
0	1	Y_{10}
0	1	Y_{11}
0	1	Y_{12}
0	0	Y_{13}
0	0	Y_{14}
0	0	Y_{15}
0	0	Y_{16}

Table 6: Dataset's ranking structure with respondents combined.

To minimize the SSR, we set the partial derivatives with respect to c_0, c_1 and c_2, equal to zero:

$$\frac{\partial SSR}{\partial c_0} = \frac{\partial SSR}{\partial c_1} = \frac{\partial SSR}{\partial c_2} = 0.$$

This yields the linear system:

$$\begin{cases} nc_0 + c_1 \sum U_i + c_2 \sum V_i = \sum Y_i \\ c_0 \sum U_i + c_1 \sum U_i^2 + c_2 \sum U_i V_i = \sum U_i Y_i \\ c_0 \sum V_i + c_1 \sum U_i V_i + c_2 \sum V_i^2 = \sum V_i Y_i \end{cases},$$

which is equivalent to the matrix equation:

$$\begin{bmatrix} n & \sum U_i & \sum V_i \\ \sum U_i & \sum U_i^2 & \sum U_i V_i \\ \sum V_i & \sum U_i V_i & \sum V_i^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum U_i Y_i \\ \sum V_i Y_i \end{bmatrix}.$$

Simplifying the sums and using Cramer's Rule we obtain the regression coefficients c_0, c_1 and c_2:

$$c_0 = \frac{\begin{bmatrix} \sum Y_i & 8 & 8 \\ \sum U_i Y_i & 8 & 4 \\ \sum V_i Y_i & 4 & 8 \end{bmatrix}}{256} = \frac{\begin{bmatrix} \sum Y_i & 2 & 2 \\ \sum U_i Y_i & 2 & 1 \\ \sum V_i Y_i & 1 & 2 \end{bmatrix}}{16},$$

$$c_1 = \frac{\begin{bmatrix} 16 & \sum Y_i & 8 \\ 8 & \sum U_i Y_i & 4 \\ 8 & \sum V_i Y_i & 8 \end{bmatrix}}{256} = \frac{\begin{bmatrix} 4 & \sum Y_i & 2 \\ 2 & \sum U_i Y_i & 1 \\ 2 & \sum V_i Y_i & 2 \end{bmatrix}}{16}, \text{and}$$

$$c_2 = \frac{\begin{bmatrix} 16 & 8 & \sum Y_i \\ 8 & 8 & \sum U_i Y_i \\ 8 & 4 & \sum V_i Y_i \end{bmatrix}}{256} = \frac{\begin{bmatrix} 4 & 2 & \sum Y_i \\ 2 & 2 & \sum U_i Y_i \\ 2 & 1 & \sum V_i Y_i \end{bmatrix}}{16}.$$

The LSMR predicted profile rankings are given by:

$$\hat{Y}_A = c_0 + c_1 + c_2 = \frac{2\sum U_i Y_i + 2\sum V_i Y_i - \sum Y_i}{16},$$

$$\hat{Y}_B = c_0 + c_1 = \frac{-6\sum V_i Y_i + \sum U_i Y_i + \sum Y_i}{16},$$

$$\hat{Y}_C = c_0 + c_2 = \frac{-2\sum U_i Y_i + 2\sum V_i Y_i + \sum Y_i}{16}, \text{ and}$$

$$\hat{Y}_D = c_0 = \frac{-2\sum U_i Y_i - 2\sum V_i Y_i + 3\sum Y_i}{16}.$$

The corresponding actual sample profile rankings obtained by averaging the respondent rankings are:

$$\bar{Y}_A = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4},$$

$$\bar{Y}_B = \frac{Y_5 + Y_6 + Y_7 + Y_8}{4},$$

$$\bar{Y}_C = \frac{Y_9 + Y_{10} + Y_{11} + Y_{12}}{4}, \text{ and}$$

$$\bar{Y}_D = \frac{Y_{13} + Y_{14} + Y_{15} + Y_{16}}{4}.$$

We thus have the following theorem: *The sample profiles' LSMR predicted and actual rankings are the same if and only if the following four equations hold*

$$(\text{profile } 11) : 2\sum U_i Y_i + 2\sum V_i Y_i - \sum Y_i = 4(Y_1 + Y_2 + Y_3 + Y_4), \quad (1)$$

$$(\text{profile } 10) : -6\sum V_i Y_i + \sum U_i Y_i + \sum Y_i = 4(Y_5 + Y_6 + Y_7 + Y_8), \quad (2)$$

$$(\text{profile } 01) : -2\sum U_i Y_i + 2\sum V_i Y_i + \sum Y_i = 4(Y_9 + Y_{10} + Y_{11} + Y_{12}), \text{and} \quad (3)$$

$$(\text{profile } 00) : -2\sum U_i Y_i - 2\sum V_i Y_i + 3\sum Y_i = 4(Y_{13} + Y_{14} + Y_{15} + Y_{16}). \quad (4)$$

Furthermore, by adding these equations, we obtain a corollary: *The following is a necessary condition for the LSMR predicted sample profile rankings to equal the sample profile rankings:*

$$\sum U_i Y_i = \sum V_i Y_i. \quad (5)$$

Returning to our small sample (Table 2), we see that the average profile rankings and LSRM predicted profile rankings are different for each of the four profiles

		Respondents						
U	V	R 1	R 2	R 3	R 4	Sample Rank	Predicted Sample Rank	|Residual Error|
1	1	1	1	3	1	1.5	3.75-1.25(1)-1.25(1)=1.25	.25
1	0	2	3	1	3	2.25	3.75-1.25(1)-1.25(0)=2.5	.25
0	1	3	2	2	2	2.25	3.75-1.25(0)-1.25(1)=2.5	.25
0	0	4	4	4	4	4	3.75-1.25(0)-1.25(0)=3.75	.25

Table 7: Predicted rankings for the sample outcomes in Table 2.

(Table 7). In this case, the regression coefficients are $c_0 = 3.75$, $c_1 = -1.25$, $c_2 = -1.25$. The equality (5) in the corollary does not hold.

The conditions (12)-(15) for whether or not the LSMR profile ranking predictions are error-free may be understood geometrically by considering points in an $x_1 x_2 x_3$ coordinate system in which the $x_1 x_2$ coordinates represent the profile and the x_3 coordinate the ranking. For our simple generic survey, if the four points representing the sample profile rankings are co-planar, there is no error; otherwise, the LSMR predicted profile rankings will have a residual error (Figure 2). Such a geometric interpretation is not possible for surveys involving more than 2 attributes, in which case standard LSMR residual analysis indicates the error in sample profile rankings using regression coefficients.
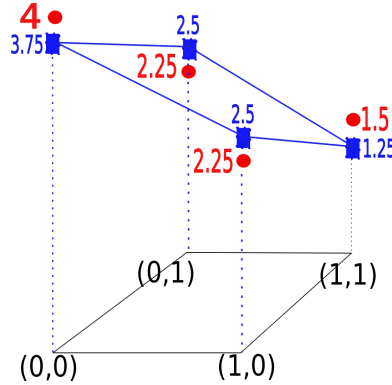

Figure 2: The LSRM predicted sample profile rankings ($\hat{Y} = 3.75 - 1.25U - 1.25V$ (square vertices) have residual errors as the actual sample rankings (dots) do not belong to the plane of regression.

2.3   Maximum Likelihood Estimation of the PRT

We will proceed to develop PR-ACBC methodology without assuming any knowledge of partworth utilitites.

2.3.1   Known Population Size

Let us assume the population has size $N = 7$, and that our sample was a random selection of 4 out of these 7. The number $N = 7$ is for simplicity of illustrating the relevant conmputations only, and could be any number $N > n = 4$. We will now use a multivariate hypergeometric distribution to obtain the maximum

likelihood estimate (MLE) PPR, meaning the population which was must likely to have yielded the observed sample outcomes.

To begin, we create what we shall call a *factor table* (Table 8)

Choose the $k = 3$ largest factors f_{ij} for a population size $N = n + k = n + 3$.				
+3	$f_{13} = 4/3$	$f_{23} = 5/3$	$f_{33} = 4/3$	
+2	$f_{12} = 3/2$	$f_{22} = 2$	$f_{32} = 3/2$	
+1	$f_{11} = 2$	$f_{21} = 3$	$f_{31} = 2$	
Number observed in sample:	$n_1 = 1$	$n_2 = 2$	$n_3 = 1$	Sample size: $n = 4$
Ranking:	$1 = O_1$	$2 = O_3$	$3 = O_9$	

Table 8: Probability factor table. The $k = 3$ largest factors $f_{ij} = 1 + (n_i/j)$ are used to determine the MLE PRT for a population size $N = n + k = 4 + 3 = 7$.

Let N_j denote the number of rankings O_j in the PRT. The number of factors a_j chosen from column j indicates that a PRT with $N_j = n_j + a_j$ is a MLE. In our case, the latter is not unique. One choice of 3 largest factors is $f_{11} = 2, f_{21} = 3, f_{22} = 2$ so that $a_1 = 1$, $a_2 = 2$ and $a_3 = 0$. A population Y in which $N_1 = n_1 + a_1 = 2$, $N_2 = n_2 + a_2 = 4$, and $N_3 = n_3 + a_3 = 1$ is a MLE. Another choice of 3 largest factors is $f_{11} = 2, f_{21} = 3, f_{31} = 1$ so that $a_1 = 1$, $a_2 = 1$ and $a_3 = 0$. Population Z in which $N_1 = n_1 + a_1 = 2$, $N_2 = n_2 + a_2 = 3$, and $N_3 = n_3 + a_3 = 2$ is also MLE. We can verify this is so by computing the respective probabilities p_Y and p_Z that our observed sample arises from erspective populations Y and Z :

$$p_Y = \frac{C(2,1)C(4,2)C(1,1)}{C(7,4)} = \frac{f_{11} \cdot f_{21} f_{22}}{C(7,4)} \tag{6}$$

and

$$p_Z = \frac{C(2,1)C(3,2)C(2,1)}{C(7,4)} = \frac{f_{11} \cdot f_{21} \cdot f_{31}}{C(7,4).} \tag{7}$$

This procedure generalizes to any number of rankings m which appear in a sample of sample n (Oberhofer and Kaufman (1987)). Let $n = \sum_{j=1}^{m} n_j$ where n_j is the number of ranking j appearing in the sample. Form the probability factor table with $f_{ij} = 1 + \frac{n_i}{j}$ ($i = 1, ..., r$, $j = 1, ..., m$ and $N = n + r$.) Choose the r largest factors in the latter table and let a_j be the number of factors chosen in column j. Then a population with a PRT such $N_j = n_j + a_j$ ($j = 1, ..., m$) is a MLE.

2.3.2 Unknown Population Size

Let us assume now that rather than having a specified size, the population size is an unknown value N. In this case, we wish to determine the probability p_i that a member of the population has ranking O_i. As before, we assume that our sample is drawn randomly from the population. The probability p that the observed sample consisting of one O_1, two O_3's and one O_9 is

$$p = f(p_1, p_3, p_9) = \frac{4!}{1!2!1!}[p_1 p_3^2 p_9], \tag{8}$$

where $g(p_1, p_3, p_9) = p_1 + p_3 + p_9 = 1$. The values of p_1^*, p_3^*, and p_9^* which maximize $H(p_1, p_3, p_9) = \ln(f(p_1, p_3, p_9))$ (and hence also maximizes $p = f(p_1, p_3, p_9)$) are obtained using Lagrange multipliers:

$$\nabla H(p_1^*, p_3^*, p_9^*) \;=\; \lambda \nabla g(p_1^*, p_3^*, p_9^*),$$

and therefore

$$\frac{1}{p_1^*} \;=\; \lambda$$
$$\frac{2}{p_2 3^*} \;=\; \lambda$$
$$\frac{1}{p_9^*} \;=\; \lambda.$$

(The scalar quantity λ is called a Lagrange multiplier.) Using $p_1^* + p_3^* + p_9^* = 1$ gives $\frac{1}{\lambda} + \frac{2}{\lambda} + \frac{1}{\lambda} = 1$ and so $\lambda = 4$. Hence, the values $p_1^* = \frac{1}{4}, p32^* = \frac{1}{2}, p_9^* = \frac{1}{4}$ maximize the probability of the observed sample outcomes.

In general, let n_k be the number of sample outcomes O_k ($k = 1, 2, ..., K$) and let p_k be the probability that a respondent in the population has outcome O_k ($k = 1, 2, ..., K$). The likelihood function $f(p_1, p_2, ..., p_K)$ giving the probability of observing the sample values $n_1, ..., n_K$ is given by

$$f(p_1, ..., p_K) = \frac{n!}{n_1! n_2! \cdots n_K!} \prod_{k=1}^{K} p_k^{n_k}, \tag{9}$$

with $\sum_{k=1}^{K} n_k = n$ and $\sum_{k=1}^{K} p_k = 1$. We seek to find the values $p_1^*, ..., p_K^*$ which maximize the likelihood function f, or equivalently, the log-likelihood function

$$H(p_1, ..., p_K) = \ln f = \ln(n!) - \sum_{k=1}^{K} n_k! + \sum_{k=1}^{K} n_k \ln(p_k), \tag{10}$$

subject to the constraint $g(p_1, ..., p_K) = p_1 + p_2 + ... + p_K = 1$. Properties of gradients imply that the optimal values p_i^* must satisfy

$$\nabla H(p_1^*, ..., p_K^*) = \lambda \nabla g(p_1^*, ..., p_K^*). \tag{11}$$

It follows that for $k = 1, ..., K$,

$$\frac{n_k}{p_k^*} = \lambda. \tag{12}$$

Hence, $n = \sum_{k=1}^{K} n_k = \lambda \sum_{k=1}^{K} p_k^* = \lambda$, and so the probabilities $p_k^* = \frac{n_k}{n}$ give the maximum likelihood of the observed sample outcomes n_k $(k = 1, 2, ..., K)$. For any sample of size n and number n_k of observed outcomes O_k $(k = 1, 2, ...K)$, the maximum likelihood probabilities $p_k^* = \frac{n_k}{n}$ indicate that for a population of size N, the expected number N_k of outcomes O_k is given by $E(N_k) = p_k N$. A maximum-likelihood population could be simulated by augmenting the observed n sample outcomes, where the probability of outcome O_k at each draw is given by p_k. For a large number of such randomly constructed populations of size N, for each k the average number of population outcomes O_k is approximately $p_k N$.

2.4 Population Ranking Intervals

Maximum likelihood provides a point estimate into the population profile rankings, meaning the PRT most likely to have produced the observed sample. In this section, given ranking data for a sample consisting of n survey respondents selected at random from a population with $N > n$ respondents, we will show how to construct for each profile ranking intervals which are guaranteed to include the population profile rankings.

Returning to our simple example, in which each profile is ranked 1, 2, 3 or 4, let $\rho_N(X)$ denote the ranking of profile X based on tournament results for N respondents. The following inequality must hold:

$$1 + \frac{n}{N}(\rho_n(X) - 1) \leq \rho_N(X) \leq 4 - \frac{n}{N}(4 - \rho_n(X)) \tag{13}$$

This interval containing $\rho_N(X)$ is obtained by either (i) assigning the rank 1 to X for all $N - n$ members of the population not in the sample (lower bound for $\rho_N(X)$); or (ii) assigning the rank 4 to X for all $N - n$ non-sample population members (upper bound for $\rho_k(X)$). Let $\lambda = \frac{N-n}{N}$ be the fraction of the population not included in the sample. It is easy to show that

$$\rho_n(X) - \lambda(\rho_n(X) - 1) \leq \rho_N(X) \leq \rho_n(X) + \lambda(4 - \rho_n(X)). \tag{14}$$

We call (14) the population profile ranking interval (PRI) for profile X. The length of this interval is 3λ, where the 3 arises algebraically as the difference between the extreme rankings 1 and 4.

The PRI can be used to quantify sample bias. Assume that, out of a total population $N = 8$, two respondents have refused to take the survey, four have completed it, and the other two have not yet replied. In this case, the length of the PRI is $\frac{3(8-4)}{8} = 3/2$. If one of the non-respondents is convinced to participate, the interval length is reduced to $\frac{3(8-5)}{8} = \frac{9}{8}$, and if both non-respondents participate, then the interval is further reduced to $\frac{3(8-6)}{8} = \frac{3}{4}$. In other words, the two non-respondents cause the length of the PRI interval to be twice as large, an important consideration in seeking to elicit survey response.
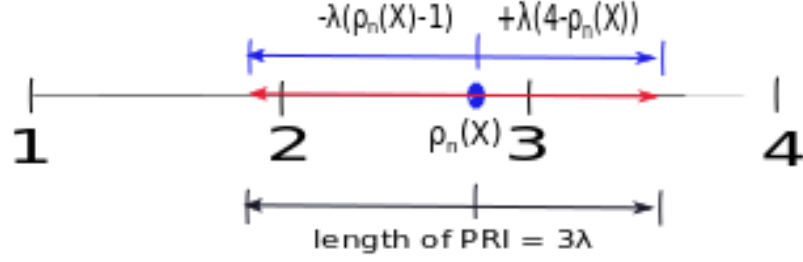
Figure 3: Given a profile X and its sample ranking $\rho_n(X)$, the length of the population profile ranking interval (PRI) is determined by $\lambda = \frac{k-n}{k}$, the proportion of the population who have not taken the survey.

2.4.1 Population Attribute Ranking Intervals

In a similar way, for any given attribute level, we can form a population *attribute level ranking interval* (ALRI). For example, for our simple survey with 2 attributes each with 2 levels, and for a sample size n if the sample attribute level ranking of level χ is $r_n(\chi)$, then for a population of size $N > n$, the population ALRI $r_N(\chi)$ must satisfy

$$r_n(\chi) + 2(N-n)r_n(\chi) \leq r_N(\chi) \leq r_n(\chi) + 8(N-n)r_n(\chi), \qquad (15)$$

which is equivalent to

$$r_n(X) - \lambda(r_n(X) - 1) \leq r_N(X) \leq r_n(X) + \lambda(4 - r_n(X)). \qquad (16)$$

with $\lambda = \frac{N-n}{N}$. Note that (14) and (16) have the same form, so the length of both intervals is 3λ.

2.5 Multimensional Scaling

One further type of analysis of ranking data is a 2-dimensional geometric representation known as multidimensional scaling (MDS) (Alvo and Yu 2014). Fundamental to MDS is use of a distance measure $d(\mu, \nu)$ in which the more similar (resp. dissimilar) are a pair of rankings μ and ν, the smaller (resp. larger). If the objects being rated (in our case survey profiles) are numberd $1, 2, ..., t$, then a ranking is simply a permutation of the set $S = \{1, 2, ..., t\}$. For example, if we number $A = 1, B = 2, C = 3, D = 4$, the ranking BACD may be represented by the permutation $\mu : S \to S$ given by $\mu(1) = 2, \mu(2) = 1, \mu(3) = 3$ and $\mu(4) = 4$. Thus, there are $t!$ rankings of t objects.
A variety of distance measures have been used for MDS. For example, Hamming distance (from coding theory) is defined as

$$d_H(\mu, \nu) = t - \sum_{i=1}^{t} I(\mu(i) = \nu(i)). \qquad (17)$$

The indicator function $I(\cdot)$ equals 1 if the statement inside parenthesis is true and 0 otherwise. Hamming distance counts the number of positions where the permutations are different.

Another example is Spearman distance, which is akin to usual Euclidean distance

$$d_S()\mu,\nu) = \frac{1}{2}\sum_{i=1}^{t}(\mu(i) - \nu(i))^2. \tag{18}$$

Note that Hamming distance formula satisfies the three required metric properties:

- NON-NEGATIVITY $d_S(\mu,\nu) \geq 0$ for all μ, ν, with equality holding if and only if $\mu = \nu$;

- SYMMETRY $d_S(\mu,\nu) = d_S(\nu,\mu)$

- TRIANGLE INEQUALITY $d(\mu,\nu) + d(\nu,\sigma) \geq d(\mu,\sigma)$.

Spearman distance, however, only satisfies the first two metric properties (Alvo and Yu).

For PR-ACBC, we would like our distance measure to detect

- small subsets of profiles with approximately the same rankings (eg. mu ranks A first and B second, and ν ranks B first and A second);

- how similar or different are profiles with the same ranking (eg. A=11 and B=10 have the same level for the first attribute, but differ in the second attribute).

To achieve this, we let (i,j) denote level j of attribute i, and let $f(i,j)$ be the average rank of level (i,j) in a respondent's ranking. For example, for the profile ranking BACD, $f(1,1) = (1+2)/2 = 1.5$ since B=10 is ranked first and A=11 is ranked second; $f(20) = (2+4)/3$ since B=10 is ranked second and $D = 00$ is ranked 4th.

For our survey, each respondent X will have four average profile level rankings $f_X(1,1), f_X(1,0), f_X(2,1), f_X(2,0)$. Given 2 respondents X and Y, we consider the squared Euclidean distance between their rankings defined as

$$d_S(X,Y) = \sum_{i,j}[f_X(i,j) - f_Y(i,j)]^2. \tag{19}$$

Note that this distance measure can be used for an ACBC survey with any number of attributes and levels.

Once a distance measure is defined, a 2-dimensional MDS is such that rankings are represented by points in an xy Cartesian coordinate system, and the Euclidean distance between these points reflects the relative distances between rankings.

TASK 1: MAKE AN MDS OF TABLE 1 USING THE SQUARED EUCLIDEAN DISTANCE (19). IS THIS AKIN TO MAKING AN MDS FROM INTER-CITY DISTANCES? DO THE THREE METRIC PROPERTIES HOLD FOR THIS DISTANCE MEASURE?
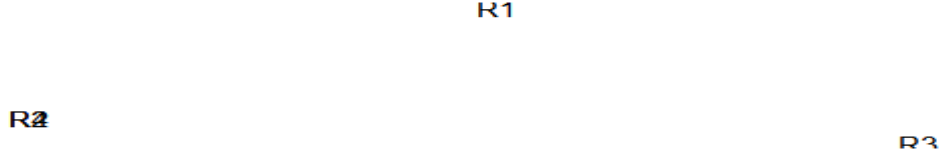
R1

R2

R3

Figure 4: MDS using the distance function defined in (19) .

3 ACBC Survey of Humanitarian Disaster Relief Organizations

PR-based ACBC methodology for analyzing small populations has many possible applications. One application involves a study of a small population ($N \approx 50$) of international, Christian faith-based disaster relief organizations headquartered in the U.S. as well as their non faith-based counterparts, will all organizations belonging to the National VOAD.

TASK 2: MAKE A TABLE OF THE N-VOAD ORGANIZATIONS CONTACTED FOR THE SURVEY

In disaster relief, effectiveness of a response may depend on the quality of collaboration between organizations with a broad diversity of religious and ideological perspectives. For effective coordination of relief, it is important that humanitarian organizations understand the unique traits and characteristics that shape their disaster response decisions. Through comparison of these factors, it is possible to design optimal partnerships and joint endeavors between organizations that may fulfill distinct, yet complimentary, humanitarian roles. Our research focuses on a few key attributes affecting our population group's decision whether or not to respond to an international humanitarian disaster. To this end, we designed an ACBC survey that creates disaster profiles with attributes and levels for this survey are displayed in Figure 5. Different disaster scenarios are paired off in the choice task (single elimination tournament) stage, beginning with 16 profiles close to the Build-Your-Own (BYO) or ideal scenario. The tournament data is the basis for profile ranking. The survey was deployed and tournament data collected using Sawtooth's Lighthouse platform.

3.1 Survey Data

As shown in Figure 4, our humanitarian survey consists of four attributes, each with three levels. Thus, the number of possible profiles is $3^4 = 81$. These are identified by four digit numbers $X = x_1 x_2 x_3 x_4$ where profile X has level x_1 for the first attribute, level x_2 for the second attribute, level x_3 for the third attribute, and level x_4 for the fourth attribute. In the tournament stage of the competition, there are four rounds, in which sixteen profiles face off against each

What is the focus of this survey?

This survey asks you to identify the type of international disaster relief scenarios your organization would respond to. Your decision may be based either on prior experience, or on the degree to which the scenario alligns with your organization's core values. With input from several managers of international relief organizations, we have framed your "go/no go" decision according to the following factors:

Factor	Description	Levels Considered
EXTERNAL FUNDING (Unrestricted money donated for the response.)	What amount of external funding is available for your response?	a) At least 75% of what is required b) About 50% of what is required c) Less than 25% of what is required
RESPONSE SCALE	Has the disaster been declared an IASC Level 1,2 or 3 emergency?	a) Large INGOs are responding in clusters (Level 3) b) Some INGOs are responding (Level 2) c) Few or no INGOs are responding (Level 1 or undeclared)
NEED ASSESSMENT	What is the level of need of the vulnerable for your organization's resources and/or capabilities?	a) Clear need for your organization's particular contribution b) Contribution could be of assistance but is not indispensable c) Unknown need of contribution
ACCESS TO AFFECTED COMMUNITY	What partnership provides connection to the community?	a) No pre-existing partnership is available b) A local partner has requested help c) An outside party has invited participation

Figure 5: An ACBC survey with 4 attributes consisting of 3 levels each.

other in head to head match-ups, much like the FIFA World Cup Round of 16. The competing profiles are selected from the 81 possible profiles based on the respondent's BYO preferences. We assign a ranking of 1 to the tournament winner, 2 to the runner up, 3 to the semifinal losers, 4 to the quarterfinal losers, 5 to the profiles that are eliminated in the first round, and 6 to those that do not appear in the tournament. The survey was first deployed to FBOs, with the results shown in Table 9.

Profile	1	2	3	4	5	6	7	8	9	10	11	12	13	PR
1212	2	3	7	3	7	7	1	7	7	7	4	1	7	4.846
1112	7	1	7	7	1	2	7	5	7	3	7	7	4	5.000
1312	7	7	7	7	5	5	7	1	1	1	7	7	3	5.000
3212	7	7	7	5	5	1	2	7	4	7	7	7	1	5.154
1122	1	7	7	4	7	7	5	7	4	7	3	3	7	5.308
2212	7	5	5	7	7	7	7	2	7	4	2	2	7	5.308
1222	7	7	4	7	4	5	7	4	3	2	7	7	7	5.462
1232	5	7	7	5	5	5	4	5	7	5	7	7	2	5.462
2312	7	7	7	1	5	3	4	7	2	7	7	7	7	5.462
2112	3	7	7	4	5	4	3	7	7	7	7	7	5	5.615

Table 9: Top 10 ranked profiles for FBOs ($n = 13, N = 50$)

TASK 3 ADD SIMILAR TABLE FOR NON FBO DATA

3.2   Survey Analysis

TASK 4 MLE
TASK 5 Population Profile Rank Intervals
Task 6 MDS of Profile Attributes

3.3   Comparison with Part-worth Utilities

Task 7 Choice task prediction accuracy
Task 8 Attribute Importances

4   Conclusion

Unlike conjoint analysis of survey data where the target populations are large and more suitable for conventional statistical tools, we have introduced a simple, intuitive approach to a small population's profile rankings based on sample data.
Task 9 COMPLETE CONCLUSION
Major areas open to further research include analysis of different ranking systems for various types of choice tournaments and application of PR ACBC methodology to other small population studies.

Acknowledgements

Mark Nussbaum, Zach Oslund, Matthew Rueger,, Nick Varberg, Mike Veatch, Joyce Yan

References

Alvo, M., and Yu, P.L.H. 2014. *Statistical Methods for Ranking Data.* Springer.

Gralla, E., Goentzel, J., and Fine G. 2014. Assessing trade-offs among multiple objectives for humanitarian aid delivery using expert preferences. *Production and Operations Management*, Springer-Verlag Berlin 23(6), 978-989.

Oberhofer, W. and Kaufman, H.1987. Maximum Likelihood Estimation of a Multivariate Hypergeometric Distribution. *Sankhya: The Indian Journal of Statistics, Series B (1960-2002)*, Indian Statistical Institute, 49(2), 188-191.

Orme, B.K., and Chrzan, K. 2017. *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros.* Sawtooth Software.

Rao, V. R. 2014. *Applied Conjoint Analysis.* Springer.

Stewart, J. 2016. *Calculus, Early Transcendentals (8E).* Cengage Learning.

Rossi, P., Allenby, G. and McCulloch R. 2005. *Baysian Statistics and Marketing.* John Wiley & Sons, Ltd.