Discrete Choice with Probabilistic Inversion:
Application to energy policy choice and wiring failure

Roger Cooke
Resources for the Future
Dept. Math. TU Delft
Jolanta Misiewicz
Dept. Math. T.U.Delft
Dept Math. Inf. And Econ. U. of Zielona Góra
April 14, 2007

Abstract: Discrete choice models normally represent the utility values of a finite set of mutually exclusive and exhaustive choice alternatives in terms of an expected value and an error. Strong assumptions are introduced on the error term. A different approach is sketched here. We assume that each member of a population of decision makers has a utility function, and that these functions can be scaled to have the same 0 and 1 values. Starting with a uniform distribution over the set of such utility functions, we apply probabilistic inversion techniques to acquire a distribution over the set of utility functions which optimally complies with discrete choice data. More generally, we may have some model explaining the discrete choice, where the parameters of the model are distributed over the population of actors. Probabilistic inversion finds an optimal distribution over the model parameters, given a starting distribution.

Key words: Discrete choice, paired comparisons, probit model, logistic model, Thurstone model, Bradley Terry model, probabilistic inversion, minimal information, iterative proportional fitting, PARFUM, energy policies, wiring failure.

1. Introduction

Probabilistic inversion denotes the operation of inverting a function at a set of random variables, rather than at a set of points. More precisely, let $\Re^m$ denote the set of real m-vectors, let G be a measurable function from $\Re^n$ to $\Re^m$, and let $\mathcal{C}$ be a set of random vectors in $\Re^m$. A random vector X in $\Re^n$ is called a probabilistic inverse of G at $\mathcal{C}$ if $G(X) \in \mathcal{C}$. Suppose we have a number of real functions $G_1,\ldots G_m$ defined on $\Re^n$, and we would like to find a distribution on $\Re^n$ such that the distribution of $(G_1,\ldots G_m)$ satisfies given constraints. This is an example of a probabilistic inversion problem. Numerical algorithms for solving such problems are available. The algorithms used here start with an initial distribution and adapt this via an iterative procedure so as to satisfy the imposed constraints, if possible. Iterative Proportional Fitting (IPF) (Kruithof 1937, Deming and Stephan 1940) converges to a solution which is minimally informative relative to the starting distribution, if the inversion problem is feasible. If the problem is not feasible, little is known about its behavior. A new algorithm (Du et al 2006) finds a 'minimal unlikely' distribution in case of infeasibility.

In this article probabilistic inversion will be applied to problems in the area of discrete choice. Each member of a population of decision makers has a utility function; hence the population of decision makers may be described as a distribution over the set of utility functions. Utility functions of rational decision makers are invariant up to a positive affine transformation; only ratio's of utility differences are meaningful. If alternative i is preferred to alternative j, then i can be assigned the value "1"and j the value "0". We assume that our utility functions are standardized in the following way: Two alternatives, heaven and hell, not necessarily belonging to the set {1, 2,…n}; are selected such that the utility of each of the n alternatives is between the utility of heaven and hell, for all members of the population. Heaven and hell are then assigned the scores 1 and 0 respectively. Hence, all utility functions on the n alternatives take values in the interval [0,1] and the set *U* of possible utility functions becomes the set $\Omega = [0, 1]^n$. Lebesgue measure on Ω may be defined in the obvious way: the measure of the set of utility functions taking values in the interval $[a_j, b_j]$ for alternative j = 1,..n is $\prod_{j=1..n} (b_j-a_j)$.

The discrete choice data may be of many forms (Anderson et al 1996, Train, 2003). Subjects can choose one preferred element from a discrete set, or they may choose k out of n elements, or they may rank some or all of the elements. The simplest type of discrete choice is paired comparisons, where each subject chooses one element of a pair of elements, for each possible pair of n elements. The set of individuals who prefer alternative i to alternative j (written "{i > j}") may be represented as the set of utility functions which assign i a higher utility than j. Assigning a probability to this set is equivalent to assigning a distribution to the indicator function of {i > j}. We seek a distribution over the set of utility functions which satisfy a given set of constraints of this form.

When a distribution has been found, this distribution will in general make the utilities for the alternatives dependent. There is no need to introduce assumptions on "error terms"; these emerge from the data and from the starting distribution.

A full formal treatment of probabilistic inversion is found in (Kraan and Bedford, 2006), a full discussion of algorithms deployed here is in (Kurowicka and Cooke, 2006). The relation between IPF and maximal likelihood is explored in Fienberg (1970), and Haberman (1974, 1984). Applications are found in Kraan and Cooke (2000, 2000a). Section 2 discusses paired comparison, section 3 elaborates on probabilistic inversion for paired comparison data. Section 4 looks at more general discrete choice problems; section 5 treats examples.

2. Paired Comparisons.

We consider alternatives 1,2,...n; and we assume that we have data of the form:

$\text{Prob}\{i > j\} = p_{ij}$; i,j = 1…n; $i \neq j$; $p_{ij} = 1 - p_{ji}$.

We say that a utility $u \in \Omega$ *expresses a preference* {i > j} if u(i) > u(j). A Probability distribution P on Ω *expresses the set* $\{p_{ij}\}$ *of pairwise preferences* if:

1) $\forall i,j;\ P\{ u \in \Omega \mid u(i) > u(j)\} = p_{ij}$.

Similarly, a probability distribution P over the set n! of permutations of {1,…n} *expresses a set of pairwise preferences* if

2) $\forall\ i, j,\ P\{ \pi \in n! \mid \pi(i) > \pi(j)\} = p_{ij}$.

We consider two problems:

Problem 1
Find a probability measure on the set Ω that expresses the pairwise preferences $\{p_{ij}\}$. If there is more than one, or none, find the "best".

Problem 2
Find a probability measure on the set n! that expresses the pairwise preferences $\{p_{ij}\}$ If there is more than one, or none, find the "best".

Clearly, a solution to the first problem yields a solution to the second and conversely. In fact, if we distribute the probability for a given ranking $\pi \in n!$ evenly over all $u \in \Omega$ which express that ranking, then this will yield the minimum information distribution with respect to Lebesgue measure which satisfies (1).

2.1 Feasibility

The question of feasibility is explored in the following lemmata.. We write *n-choose-2* for the number of distinct unordered pairs of n objects: n (n-1) / 2.

Lemma 1:
1. For any $\pi_1, \pi_2 \in n!$, if π_1 and π_2 agree on all (n-choose-2) paired comparisons, then $\pi_1 = \pi_2$;
2. For (n-choose-2) – 1 pairs, there exist $\pi_1, \pi_2 \in n!$ such that π_1, π_2 agree on paired comparisons for these pairs, but $\pi_1 \neq \pi_2$

Proof: (1) The top ranked alternative according to π_1 is preferred to every alternative according to π_1; by assumption the same holds for π_2. Therefore this alternative is also top ranked in π_2. Proceeding in this way, we show $\pi_1 = \pi_2$. (2) Suppose only pair (1,2) is excluded, Let π_1, π_2 rank 1 resp. 2 in the first position, and 2 resp. 1 in the second position, all other rankings being the same. Then π_1, π_2 agree on all paired comparisons except (1,2).

This shows that if we wish to learn a ranking by asking pairwise preferences, we may need to query all pairs. .

Lemma 2: If $p_{ij} + p_{jk} + p_{ki} > 2$, for some i,j,k; then there is no probability over n! satisfying (2).

Proof: Suppose P was such a probability. Let A_{ij} be the set of rankings satisfying $\pi(i) > \pi(j)$, and similarly for A_{jk}, A_{ki}. Then $P(A_{ij}) = p_{ij}$ etc. Any $\pi \in A_{ij} \cap A_{jk} \cap A_{ki}$ would have a circular triad, so $P(A_{ij} \cap A_{jk} \cap A_{ki}) = P(\emptyset) = 0$. We have
$P(A_{ik}) = 1\text{-}p_{ki} \geq P(A_{ij} \cap A_{jk}) = p_{ij} + p_{jk} - P(A_{ij} \cup A_{jk}) \geq p_{ij} + p_{jk} - 1$, which violates the assumption. □

Lemma 2 shows that the problems 1 and 2 above may not be feasible. For n = 3 we can get a necessary and sufficient condition for the existence of a probability over n! expressing the pairwise comparisons. The proof is a bit fussy and is placed in the appendix. It helps to understand the complexity of the feasibility questions in general.

Lemma 3. If n = 3 and $\forall i,j,k = 1,2,3;\ i \neq j \neq k \neq i,\ \ p_{ij} + p_{jk} + p_{ki} \leq 2$, then there exists a distribution over 3! that expresses the pairwise preferences.

Proof: See appendix.

Recall the Condorcet voting paradox: 1/3 of the population prefer Mozart > Bach > Hayden, 1/3 prefer Bach > Hayden > Mozart, and 1/3 prefer Hayden > Mozart > Bach. Each person is consistent, but the majority preference is intransitive. Note that $p_{MB} + p_{BH} + p_{HM} = 2$. This example is realizes the upper bound of intransitive majority preference with transitive voters.

Let μ be Lebesgue measure on $\Omega = [0, 1]^n$. We may formulate our problem as a constrained optimization problem as follows:

9) Find a density f on $[0, 1]^n$ minimizing $\int_\Omega \ln(f(x))\, d\mu(x)$, such that for all {i, j}

$$\int_{A_{ij}} f(x)\, d\mu(x) = p_{ij};\ \ A_{ij} = \{u \mid u(i) > u(j)\}.$$

Since f(x) is the Radon Nikodym derivative of a measure ν with respect to μ, we can write this as $\int_\Omega \ln(d\nu(x)/d\mu(x))\, d\mu(x)$ which is recognized as the Kullback Leibler relative information of ν with respect to μ.

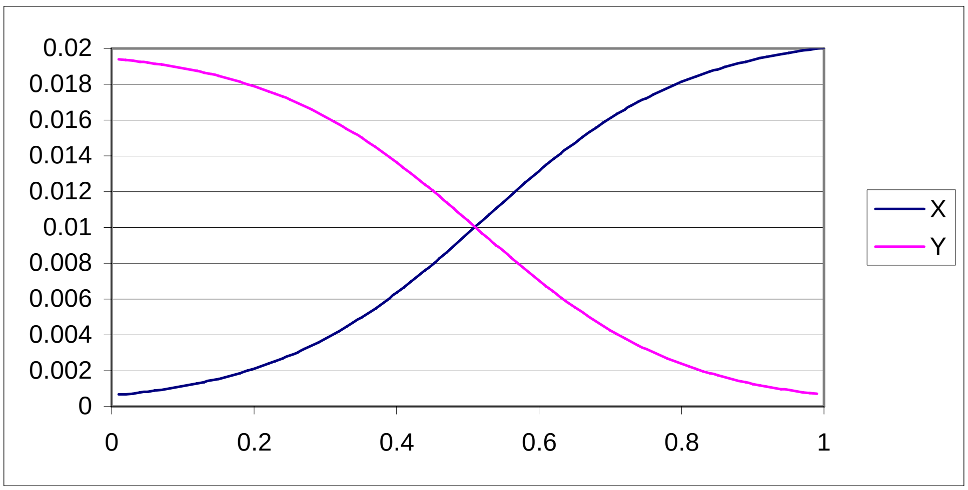
Remark A distribution on n! expressing a set of pairwise preferences is not in general unique. It suffices to consider $p_{12} = p_{23} = p_{13} = 0.5$. This set of pairwise preferences is expressed by the uniform distribution over all 6 permutations of {1, 2, 3} and also by the uniform distribution over (1,2,3) and (3,2,1).

Remark Solving (9) will induce correlations. This is easy to see in the case of 2 alternatives. Let u(1) and u(2) be independent and uniformly distributed on [0, 1]. The distribution which minimizes relative information relative to this starting measure and which satisfies p(u(1) > u(2)) = 1, is a uniform density concentrated on one side of the main diagonal. The induced correlation of u(1) and u(2) is 0.5. Writing p = p(u(1) > u(2)); $p' = 1 - p$, one can calculate the correlation between u(1) and u(2) in the minimal information distribution given p = p(u(1) > u(2)) as:

$$\rho(u(1),u(2)) = \frac{\tfrac{1}{4} - (2/3 - p/3)(2/3 - p'/3)}{[(1/2 + p/9 - p^2/9 - 4/9)\,(1/2 + p'/9 - p'^2/9 - 4/9)]^{0.5}} \qquad (10)$$

For example, $p(u(1) > u(2)) = 0.9$ induces a correlation of 0.271. We shall see that with more alternatives, paired comparison data sometimes induce much stronger correlations.

Starting again with the uniform distribution on the unit square, we can ask which distribution has minimum relative information with respect tos this measure, satisfies $P(X>Y)=0.9$, and makes X and Y independent. The solution on a discrete grid of the unit square can easily be found with the method of Lagrange multipliers. The marginal distributions of X and Y are shown below:



3. Probabilistic Inversion for paired comparisons

As we might guess from Lemma 3, determining the feasibility of (9) is hard. We can also look at this as a problem of probabilistic inversion.

Define a mapping $C_{ij}: \Omega \rightarrow \{0,1\}$ as $C_{ij}(u) = 1$ if $u(i) > u(j)$, and $= 0$ otherwise.
Define C: $\Omega \rightarrow \{0,1\}^{n\text{-choose-}2}$ as the mapping $C(u) = (C_{12}(u)\ldots C_{n-1,n}(u))$. Let M denote the set of probability measures on $\{0,1\}^{n\text{-choose-}2}$ whose marginal probability for "1" on the $\{i,j\}$-th coordinate is p_{ij}.

We define a probabilistic inversion problem as:

10) Find a measure λ on Ω, minimally informative with respect to Lebesgue measure, such that $C(\lambda) \in M$; where $C(\lambda)$ is the "push through" of λ onto $\{0,1\}^{n\text{-choose-}2}$: $C(\lambda)(B) = \lambda(C^{-1}(B))$ for all $B \subseteq \{0,1\}^{n\text{-choose-}2}$.

In other words, we want to invert the function C at the set M. If an inverse exists, we want the inverse which is minimally informative with respect to Lebesgue measure.

To solve these problems we restrict to a (large) finite subset Ω' of Ω. Formulation (9) then solves for the probability mass on each element of Ω'. If Ω' is small it is less representative of Ω and less likely to be feasible. Making Ω' large makes the problem numerically intractable.

Formulation (10) admits 2 algorithms based on sample re-weighting. There are no constraints on sample size (other than hardware). A large sample from Ω is drawn and re-weighted to (try to) satisfy the constraints. When we re-sample this sample using these weights, we are sampling a distribution whose Radon Nikodym derivative with respect to the original sample distribution is given by the set of weights.

There are two algorithms at our disposal. Iterative Proportional Fitting (IPF) (Kruithof 1937, Deming and Stefan 1940, Ciszar, 1975, Feinberg 1970, Haberman 1974, 1984, Bishop 1967, Brown 1967) and PARFUM (Parameter Fitting For Uncertain Models) (Cooke, 1994, Du et al 2006).

A brief description of these algorithms is as follows. Start with a sample from Ω of size N, and assign each sample weight $w_o(k) = 1/N$; $k = 1,\ldots N$. Let Σ_{ij} denote summation over the samples u_k for which $u_k(i) > u_k(j)$. For IPF, we adapt a current set of weights w_s for each (ij) ∈ (n-choose-2) according to

$$w_s(k) \rightarrow w_{s+1}(k) = (p_{ij}\, w_s(k) / \Sigma_{ij}\, w_s(k)) \quad \text{if } u_k(i) > u_k(j)$$

$$w_s(k) \rightarrow w_{s+1}(k) = ((1-p_{ij}) w_s(k) / \Sigma_{ji}\, w_s(k)) \quad \text{if } u_k(i) < u_k(j).$$

We cycle through the constraints in this manner, repeatedly. Table 1 below illustrates the first three iterations in a simple case with n = 3 alternatives and N = 5 samples from Ω.

sample	u(1)	u(2)	u(3)	w0	w1		w2		w3	
1	0.1	0.6	0.2	0.2	0.4*0.2 / (3*0.2)=	0.133	0.4*0.133/0.2666=	0.200	0.7*0.2 /0.755=	0.185
2	0.4	0.3	0.9	0.2	0.6*0.2 / (2*0.2) =	0.300	0.6*0.3/0.7333=	0.245	0.7*0.245/0.755=	0.227
3	0.6	0.8	0.7	0.2	0.4*0.2 / (3*0.2)=	0.133	0.4*0.133/0.2666=	0.200	0.7*0.2 /0.755=	0.185
4	0.3	0.4	0.5	0.2	0.4*0.2 / (3*0.2)=	0.133	0.6*0.1333/0.7333=	0.109	0.7*0.109/0.755=	0.101
5	0.9	0.1	0.3	0.2	0.6*0.2 / (2*0.2) =	0.300	0.6*0.3/0.7333=	0.245	0.3=	0.300

Table 1 3 IPF iterations, $p_{12} = 0.6$, $p_{23} = 0.4$, $p_{13} = 0.3$

If the inversion problem is feasible, then the IPF algorithm converges to the distribution whose likelihood ratio relative to the starting measure is maximal. This is equivalent to minimizing the Kullback Leibler information relative to the starting measure (Csiszar 1975). If the problem is not feasible, little is known about the behavior of IPF (Csiszar and Tusnady (1984) discuss a special case).

For PARFUM we simply average the adaptations for one complete cycle. For the example in Table 1, the first PARFUM iteration would be $(w_1 + w_2 + w_3)/3$. Formally, letting 1_A denote the indicator function for the set A:

$$w_s(k) \rightarrow w_{s+1}(k) =$$

$$\frac{\sum_{\{i,j\} \in \text{n-choose-2}} 1_{\{uk(i) > uk(j)\}} p_{ij} [w_s(k) / \Sigma_{ij} w_s(k)] + 1_{\{uk(j)>uk(i)\}} (1-p_{ij})[w_s(k) / \sum_{ji} w_s(k).]}{\text{(n-choose-2)}}$$

The relative information of w_{s+1} with respect to w_s converges to a constant; that constant is zero if and only if the problem is feasible. In that case PARFUM's stationary points are solutions. If the problem is not feasible, then PARFUM's stationary points minimize an information functional. Whereas IPF finds a 'maximally likely' distribution among all feasible distributions, if this set is non empty, PARFUM finds a "minimally infeasible distribution" in case of infeasibility (Du et al 2006). Infeasibility is quite common in these problems.

4. Probabilistic inversion for other discrete choice problems

Probabilistic inversion techniques can be applied more generally. A few possibilities will be illustrated. If there are a large number of alternatives, say 20 or more, then a full paired comparison study becomes tedious for the subjects. An alternative is to let subjects rank their top 5 and bottom 5 alternatives. From the subjects' assessments, each alternative is assigned a probability of being ranked first, second, etc. Subjects may be offered smaller groups of alternatives and be asked to indicate a first or first and second ranked alternative from each set. The only requirement is a given utility function uniquely determines the choice behavior. We can then find a distribution over utility functions which mimics the discrete choice data.

Another option arises if we wish to explain the choice behavior in terms the alternative's attributes. Suppose we have discrete choice data on N automobiles, and our goal is to explain the preferences in terms of the attributes Price (P), Reliability (R), Acceleration (A), Roominess (Rm) and Milieu Friendliness (MF). We know how each car scores on each attribute but we don't know how these attributes combine to shape the preferences.

One possibility is to derive utility values from the discrete choice data and regress these values on the attributes. This would have the advantage placing a well developed theory at our disposal. It has the disadvantage of requiring a number of alternatives that is large relative to the number of attributes. In situations where that is not possible, a different approach could be explored. We adopt a functional form for the utility of automobile x, for example:

$$U(x) = B_1P(x) + B_2R(x) + B_3A(x) + B_4Rm(x) + B_5MF(x).$$

where P(x) is the price of automobile x, etc. We then choose a starting distribution over $(B_1,\ldots B_5)$ and adapt this such that the resulting distribution over the utilities of automobiles expresses the discrete choice preferences. We may also impose constraints on the B's, for example that they are non negative and sum to one. These options are explored in a forthcoming publication.

5. Example
5.1 Energy Policies

A small paired comparison study was conducted at Resources for the Future regarding preferences for energy policies. The policies are:.

1: Tax@pump: 1$ per gallon gasoline surcharge, to be used for research in renewables
2. Tax Break: (a) No sales tax for purchase of new hybrid or electric car; (b)First car owners can deduct purchase cost from their income tax; (c) No sales tax on bio-diesel or ethanol (c) Tax credits for energy efficiency home improvements (insulation, double glass windows, solar panels)
3. **Road Tax:** Annual road tax 1$ per lb on all light duty vehicles, no tax rebate for driving to work or parking, to be used for research in fuel efficient vehicles and bio fuels.
4. CO_2 cap CO_2 emissions cap on electricity generation.
5. Subsidies for clean coal Give subsidies for clean coal with carbon sequestration to make coal competitive with natural gas.
6. Do Nothing

The software packaged EXCALIBUR is used to solve the Thurstone models B (constant variance and correlation) and C (constant variance and independence) (Thurstone 1927). These are called probit models in the economics literature. When scaled to the [0, 1] interval there is scarcely any difference between these two. The Bradley Terry model is similar to the logit models for discrete choice, except that it is solved using maximal likelihood instead of least squares (Bradely 1952, Bradley and Terry 1953). For more background see (Torgerson 1958, David 1963, McFadden 1974 1974, Mosteller 1952) .

	Item name	Bradley-Terry	Thurstone C (scaled [0,1])	Thurstone B (scaled [0,1])	Prob. inv. Mean values scaled to [0, 1]
1	tax@pump	0.3426	1	1	1
2	TaxBreak	0.181	0.8003	0.7912	0.7581
3	RoadTax	0.1408	0.675	0.6714	0.6433
4	CO2Cap	0.2054	0.8393	0.8386	0.8296
5	CleanCoal	0.1008	0.5531	0.5508	0.5087
6	DoNothing	0.0294	0	0	0

Table 2: Bradley Terry, Thurstone models, and probabilistic inversion means values compared, energy Policies.

The average probabilistic inversion utility scores are shown in Table 2, after scaling to [0, 1]. They agree reasonably well with the Thurstone scores, suggesting that the Thurstone assumptions are reasonably compliant with the paired comparison data. The Bradely Terry scores are not scaled to [0, 1], as they are not invariant under affine transformations.

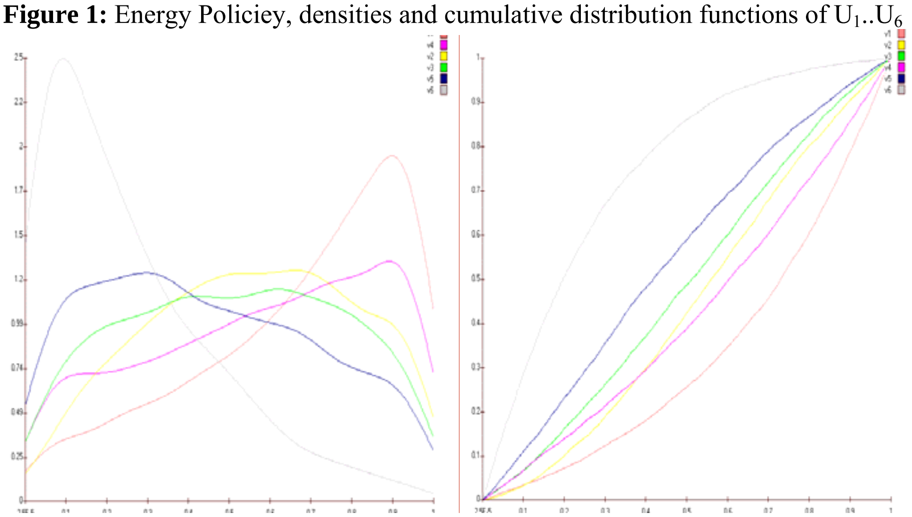
The utility scores are not independent, but the correlation structure is weak, as shown in Table 3.

	1	2	3	4	5	6
1	1.0000	0.0531	0.0938	0.0136	0.0130	0.0447
2	0.0531	1.0000	0.0333	0.0378	0.0569	0.2567
3	0.0938	0.0333	1.0000	0.0181	0.0619	0.0898
4	0.0136	0.0378	0.0181	1.0000	-0.0055	0.0741
5	0.0130	0.0569	0.0619	-0.0055	1.0000	0.0989
6	0.0447	0.2567	0.0898	0.0741	0.0989	1.0000

Table 3: Correlation matrix of IPF utilities

The densities and cumulative distribution functions of U1…U6 are shown in Figure 1 below:

Figure 1: Energy Policiey, densities and cumulative distribution functions of $U_1..U_6$



The small correlation values does not mean that the joint distribution is independent. The Cobweb plots shown in Figure 2 enable us to see higher order conditional dependences. The utilities for the six alternatives are plotted as vertical lines. On each sample one value is realized on each line; a jagged line connecting these represents one sample. Figure 1 shows 1066 samples. Note that U_1 and U_6 are concentrated on higher and lower values respectively. The "cross densities" above the midline between variables i and i+1 are the densities of $U_i + U_{i+1}$. If U_i and U_{i+1} are uniform and independent, the cross density is triangular. Figure 2 shows the percentiles instead of the natural scale values of Figure 1. The marginal distributions are now uniform by definition, and the cross densities are nearly triangular. The dependence becomes visible when we conditionalize. If the variables were independent, the distributions of U_5 and U_6 would not be affected by conditionalizing on the first four utilities lying below their median. The result of this conditionalization is shown in Figure 3. Note that the cross density is far from triangular, showing that U_5 and U_6 are not independent given that U_1, … U_4 are beneath their medians.

Figure 1 Cobweb plots Energy Policies

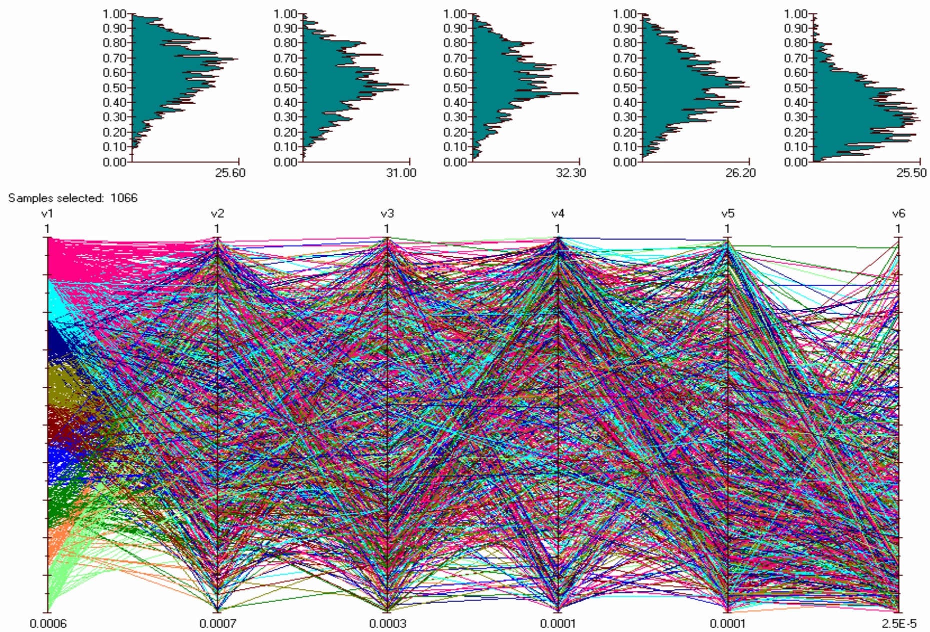
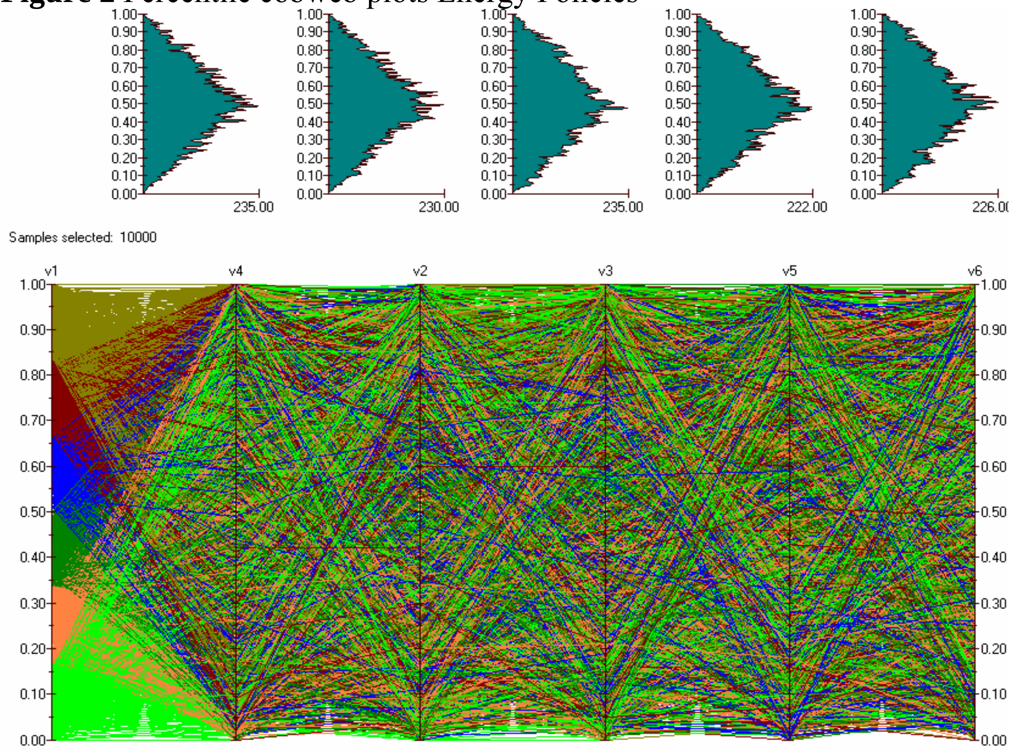


Figure 2 Percentile cobweb plots Energy Policies



5.2 Open wiring failure: IPF

A study of open wiring failure (Mazzuchi et al 2007) involved 9 experts and 15 items. This case is of interest, as the IPF algorithm did not converge. The percentile cobweb for the IPF solution after 100 iterations is shown below.

Figure 3 Conditional percentile cobweb plot, Energy Policies

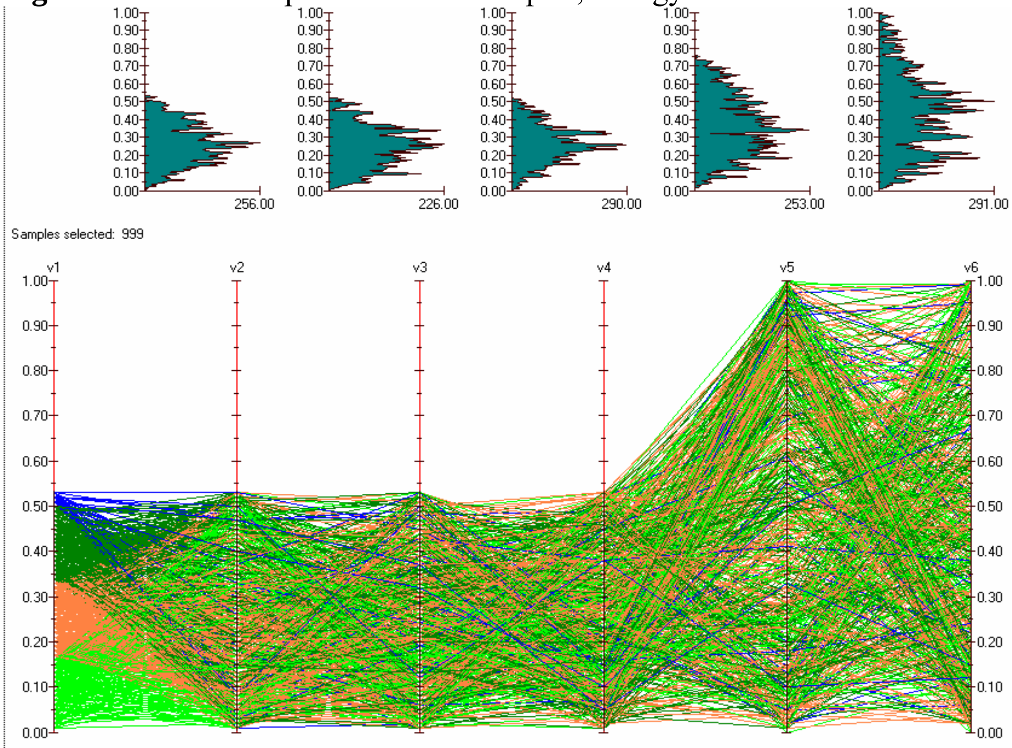
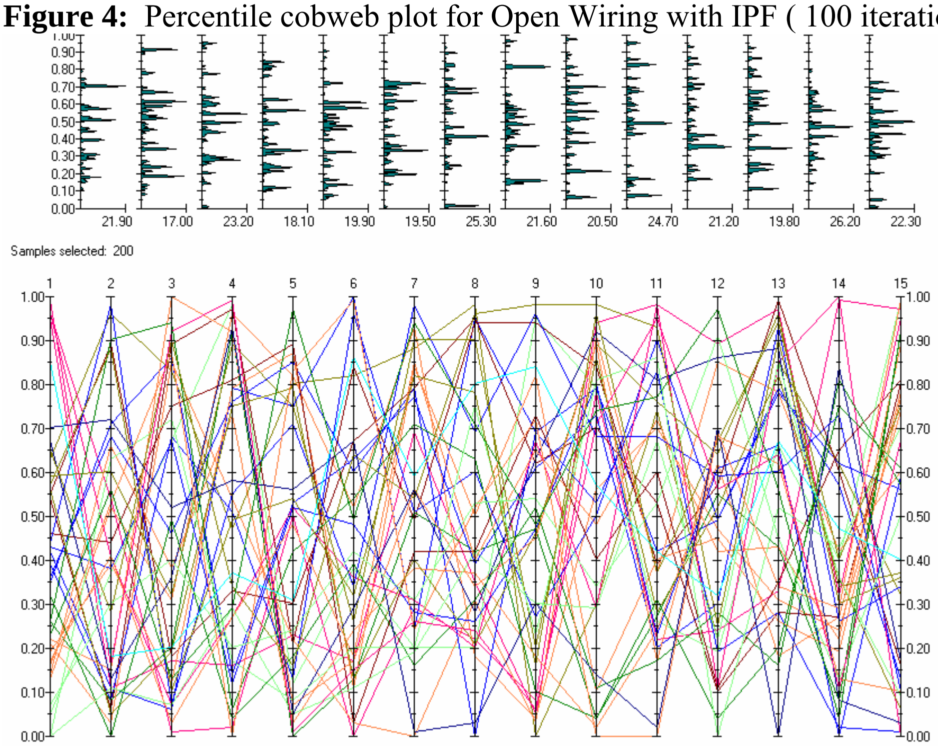


Figure 4: Percentile cobweb plot for Open Wiring with IPF (100 iterations)



Note that the distribution is quite sparse, and that the cross densities indicate high positive correlation. Non-convergence is shown by the proportions matrix, giving the target and realized proportions of preferring alternative i to j (only part of the matrix is shown). Large differences between the target and realized proportions are emphasized. The rather high correlations are shown Figure 6, where large correlations are emphasized.

Figure 5: Proportion Matrix for Open wiring, IPF (100 iterations)

Proportions matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0,3333 0,3386	0,4444 0,6078	0,5000 0,6152	0,2222 0,2783	0,2778 0,2383	0,9444 0,7266	0,2222 0,5468	0,4444 0,5525	0,3333 0,6107	0,7222 0,7722	0,8889 0,8178	0,0556 0,1241	0,8889 0,8066	0,4444 0,4692
2	0,6667 0,6614	0	0,9444 0,8993	0,8889 0,7183	0,5556 0,3837	0,3333 0,2878	1,0000 1,0000	0,6667 0,6334	0,7778 0,6067	0,7778 0,7767	0,8889 0,8158	0,7778 0,7754	0,3333 0,4810	0,8889 0,8331	0,6111 0,6218
3	0,5556 0,3922	0,0556 0,1007	0	0,1111 0,3485	0,2222 0,2622	0,1111 0,1673	0,3889 0,5631	0,3333 0,3883	0,3333 0,5168	0,3333 0,4949	0,5556 0,6483	0,5556 0,7043	0,2222 0,1660	0,8889 0,7105	0,7778 0,5005
4	0,5000 0,3848	0,1111 0,2817	0,8889 0,6515	0	0,0556 0,1195	0,2778 0,1192	0,7222 0,6324	0,4444 0,5004	0,4444 0,4770	0,3889 0,5386	0,6667 0,5961	0,8333 0,8490	0,2222 0,2721	0,8889 0,8866	0,5556 0,4734
5	0,7778 0,7217	0,4444 0,6163	0,7778 0,7378	0,9444 0,8805	0	0,2222 0,3578	1,0000 1,0000	0,6667 0,7313	0,7778 0,7756	0,7222 0,8607	0,7778 0,9290	0,8333 0,8857	0,2222 0,5325	1,0000 1,0000	0,8889 0,8128
6	0,7222 0,7617	0,6667 0,7122	0,8889 0,8327	0,7222 0,8808	0,7778 0,6422	0	1,0000 1,0000	1,0000 1,0000	0,8889 0,8126	0,8333 0,9404	0,9444 0,9889	0,9444 0,9362	0,8333 0,7241	1,0000 1,0000	0,6667 0,6298
7	0,0556 0,2734	0,0000 0,0000	0,6111 0,4369	0,2778 0,3676	0,0000 0,0000	0,0000 0,0000	0	0,1667 0,1975	0,2778 0,3369	0,1111 0,1834	0,3889 0,5863	0,6111 0,5874	0,1111 0,1427	0,7778 0,6725	0,2778 0,3354
8	0,7778 0,4532	0,3333 0,3666	0,6667 0,6117	0,5556 0,4996	0,3333 0,2687	0,0000 0,0000	0,8333 0,8025	0	0,4444 0,5032	0,7222 0,7607	0,8889 0,7752	0,5556 0,6292	0,0556 0,0723	0,9444 0,8370	0,4444 0,4464
9	0,5556 0,4475	0,2222 0,3933	0,6667 0,4832	0,5556 0,5230	0,2222 0,2244	0,1111 0,1874	0,7222 0,6631	0,5556 0,4968	0	0,5556 0,5251	0,6667 0,5752	0,3889 0,5848	0,3333 0,4273	0,7778 0,7746	0,6667 0,5889
10	0,6667 0,3893	0,2222 0,2233	0,6667 0,5051	0,6111 0,4614	0,2778 0,1393	0,1667 0,0596	0,8889 0,8166	0,2778 0,2393	0,4444 0,4749	0	0,7778 0,7121	0,8333 0,7096	0,0000 0,0000	0,7778 0,7170	0,7778 0,5420
11	0,2778 0,2278	0,1111 0,1842	0,4444 0,3517	0,3333 0,4039	0,2222 0,0710	0,0556 0,0111	0,6111 0,4137	0,1111 0,2248	0,3333 0,4248	0,2222 0,2879	0	0,5556 0,6936	0,0000 0,0000	0,7222 0,7163	0,6667 0,4576
12	0,1111 0,1822	0,2222 0,2246	0,4444 0,2957	0,1667 0,1510	0,1667 0,1143	0,0556 0,0638	0,3889 0,4126	0,4444 0,3708	0,6111 0,4152	0,1667 0,2904	0,4444 0,3064	0	0,2222 0,1772	0,5556 0,5759	0,4444 0,3420
13	0,9444 0,8759	0,6667 0,5190	0,7778 0,8340	0,7778 0,7279	0,7778 0,4675	0,1667 0,2759	0,8889 0,8573	0,9444 0,9277	0,6667 0,5727	1,0000 1,0000	1,0000 1,0000	0,7778 0,8228	0	0,8333 0,8867	0,5556 0,5683
14	0,1111 0,1934	0,1111 0,1669	0,1111 0,2895	0,1111 0,1134	0,0000 0,0000	0,0000 0,0000	0,2222 0,3275	0,0556 0,1630	0,2222 0,2254	0,2222 0,2830	0,2778 0,2837	0,4444 0,4241	0,1667 0,1133	0	0,3333 0,3333
15	0,5556 0,5308	0,3889 0,3782	0,2222 0,4995	0,4444 0,5266	0,1111 0,1872	0,3333 0,3702	0,7222 0,6646	0,5556 0,5536	0,3333 0,4111	0,2222 0,4580	0,3333 0,5424	0,5556 0,6580	0,4444 0,4317	0,6667 0,6667	0

Figure 6: Correlation matrix, Open Wiring. IPF (100 iterations)

Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	-0,1176	0,0112	-0,0463	0,0467	-0,0507	-0,1271	0,2473	-0,3175	0,0307	0,3201	-0,2442	0,2950	-0,3513	0,0817
2	-0,1176	1	0,0901	0,2116	-0,0224	-0,2450	0,5832	-0,2292	0,1503	0,4345	-0,3805	0,0497	-0,1192	0,0249	-0,0352
3	0,0112	0,0901	1	-0,1586	-0,0480	-0,1375	-0,1305	-0,0352	0,1484	0,3194	-0,0925	-0,0573	0,2638	0,0520	0,0530
4	-0,0463	0,2116	-0,1586	1	0,2097	0,3398	0,0037	0,0334	0,1550	0,0584	-0,0142	0,4789	-0,2121	0,4439	-0,4436
5	0,0467	-0,0224	-0,0480	0,2097	1	0,0697	0,2506	0,0786	0,0987	-0,2961	-0,1414	0,0381	-0,0821	0,1439	0,0456
6	-0,0507	-0,2450	-0,1375	0,3398	0,0697	1	-0,2067	0,6069	-0,1544	0,0989	0,2588	0,2527	0,1793	0,5137	-0,1636
7	-0,1271	0,5832	-0,1305	0,0037	0,2506	-0,2067	1	-0,1946	0,1470	0,0304	-0,4244	0,0274	-0,3240	-0,0813	0,2419
8	0,2473	-0,2292	-0,0352	0,0334	0,0786	0,6069	-0,1946	1	-0,5720	0,2233	0,2141	-0,0598	0,4408	0,0843	0,1233
9	-0,3175	0,1503	0,1484	0,1550	0,0987	-0,1544	0,1470	-0,5720	1	-0,3503	-0,2804	0,3660	-0,5610	0,2787	-0,3898
10	0,0307	0,4345	0,3194	0,0584	-0,2961	0,0989	0,0304	0,2233	-0,3503	1	0,0902	0,0927	0,5198	0,0498	0,1381
11	0,3201	-0,3805	-0,0925	-0,0142	-0,1414	0,2588	-0,4244	0,2141	-0,2804	0,0902	1	-0,2639	0,4880	-0,0198	-0,0258
12	-0,2442	0,0497	-0,0573	0,4789	0,0381	0,2527	0,0274	-0,0598	0,3660	0,0927	-0,2639	1	-0,2893	0,5488	-0,3747
13	0,2950	-0,1192	0,2638	-0,2121	-0,0821	0,1793	-0,3240	0,4408	-0,5610	0,5198	0,4880	-0,2893	1	-0,1836	0,2686
14	-0,3513	0,0249	0,0520	0,4439	0,1439	0,5137	-0,0813	0,0843	0,2787	0,0498	-0,0198	0,5488	-0,1836	1	-0,3074
15	0,0817	-0,0352	0,0530	-0,4436	0,0456	-0,1636	0,2419	0,1233	-0,3898	0,1381	-0,0258	-0,3747	0,2686	-0,3074	1

5.2.1 Open wiring failure: PARFUM

In such cases, PARFUM does converge to a distribution over Ω that minimizes the lack of fit on each pairwise comparison, in the sense of relative information. The cobweb from the PARFUM solution is shown in Figure 7, the marginal utility densities are shown in Figure 8.

Figure 7: Percentile cobweb plot Open Wiring, PARFUM (100 iterations)

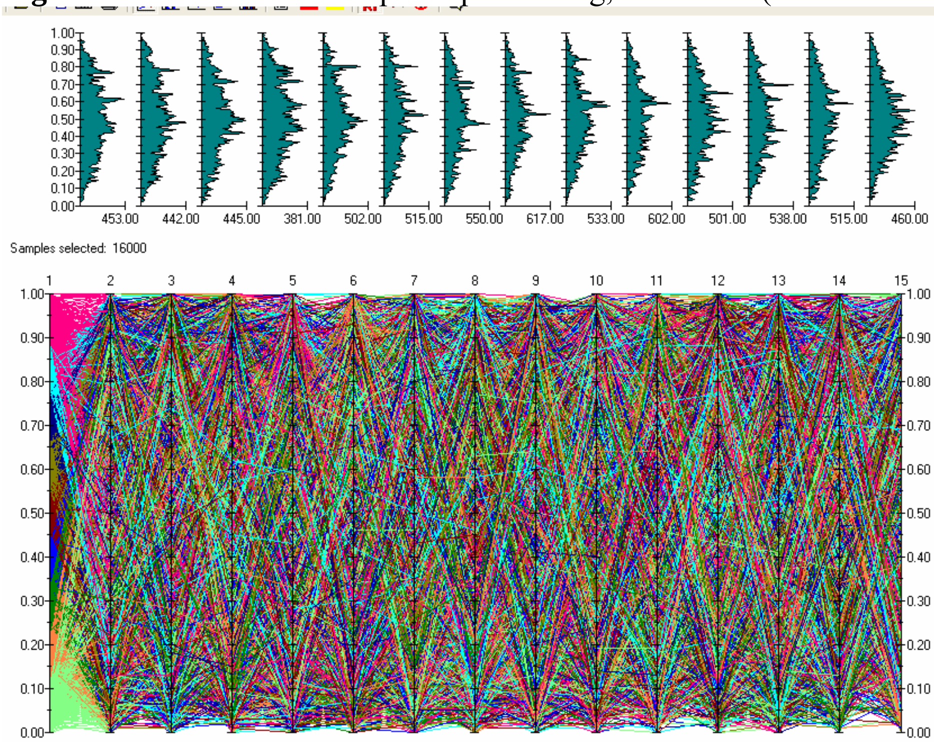


Figure 8: Densities, Open Wiring, PARFUM (100 iterations)

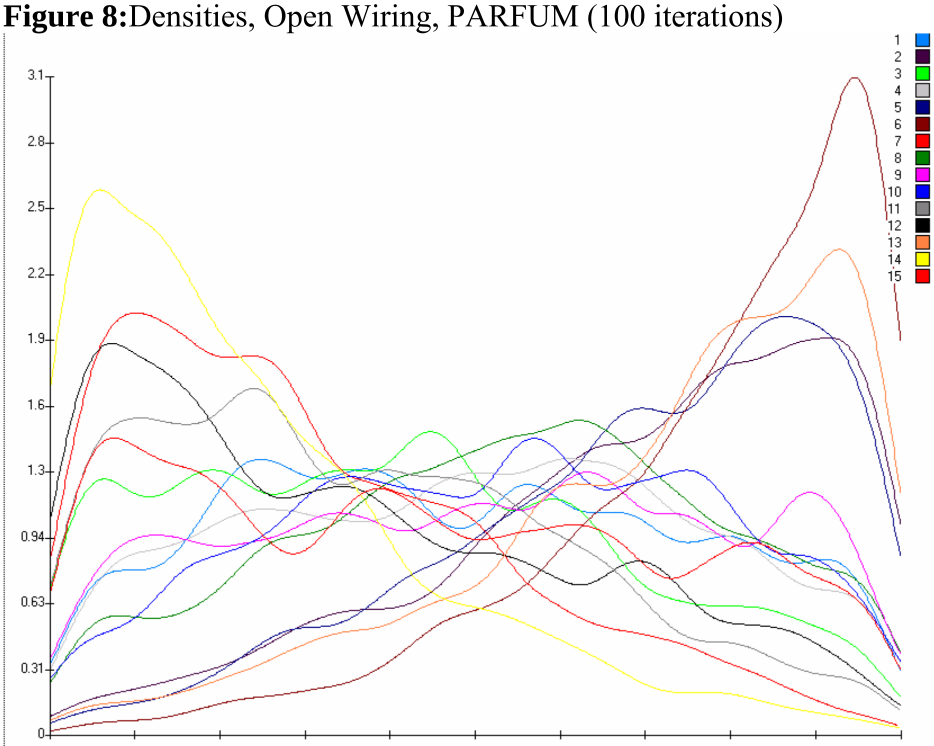


Table 4: Scale values for Thurstone models, IPF and PARFUM, scaled to [0,1], for Open Wiring.

Item Nr	Thurston C (scaled to 0,1)	Thurston B (scaled to 0,1)	IPF Mean (Scaled to 0,1)	PARFUM Mean (scaled to 0,1)
1	0,469	0,480	0,536	0,463
2	0,807	0,777	0,681	0,825
3	0,321	0,352	0,316	0,313
4	0,476	0,482	0,388	0,489
5	0,803	0,748	0,932	0,825
6	1,000	1,000	1,000	1,000
7	0,120	0,194	0,109	0,129
8	0,531	0,507	0,485	0,560
9	0,472	0,466	0,416	0,495
10	0,513	0,509	0,332	0,531
11	0,232	0,268	0,178	0,235
12	0,221	0,228	0,090	0,216
13	0,880	0,808	0,846	0,901
14	0,000	0,000	0,000	0,000
15	0,374	0,369	0,411	0,371

Now there are strong differences between Thutrstone B and C, and with the mean values from the PARFUM solution. IPF is somewhat eccentric.

Figure 9: Proportion matrix Open Wiring (PARFUM 100 iterations)

1	0	0.3333 0.2919	0.4444 0.5725	0.5000 0.4916	0.2222 0.2823	0.2778 0.1897	0.9444 0.7292	0.2222 0.4175	0.4444 0.4531	0.3333 0.4408	0.7222 0.6308	0.8889 0.6766	0.0556 0.2320	0.8889 0.8082	0.4444 0.5396
2	0.6667 0.7081	0	0.9444 0.8071	0.8889 0.7309	0.5556 0.4856	0.3333 0.3560	1.0000 0.8899	0.6667 0.6459	0.7778 0.6932	0.7778 0.6866	0.8889 0.8364	0.7778 0.8190	0.3333 0.4179	0.8889 0.9034	0.6111 0.7352
3	0.5556 0.4275	0.0556 0.1929	0	0.1111 0.3650	0.2222 0.2360	0.1111 0.1413	0.3889 0.5660	0.3333 0.3666	0.3333 0.4081	0.3333 0.3723	0.5556 0.5332	0.5556 0.5391	0.2222 0.2116	0.8889 0.7023	0.7778 0.5148
4	0.5000 0.5084	0.1111 0.2691	0.8889 0.6350	0	0.0556 0.2909	0.2778 0.2337	0.7222 0.7289	0.4444 0.4756	0.4444 0.4869	0.3889 0.4757	0.6667 0.6503	0.8333 0.6840	0.2222 0.2709	0.8889 0.7787	0.5556 0.5611
5	0.7778 0.7177	0.4444 0.5144	0.7778 0.7640	0.9444 0.7091	0	0.2222 0.3694	1.0000 0.8840	0.6667 0.6371	0.7778 0.7254	0.7222 0.6674	0.7778 0.7915	0.8333 0.8056	0.2222 0.4455	1.0000 0.9165	0.8889 0.7497
6	0.7222 0.8103	0.6667 0.6440	0.8889 0.8587	0.7222 0.7663	0.7778 0.6306	0	1.0000 0.9350	1.0000 0.8053	0.8889 0.8117	0.8333 0.8013	0.9444 0.9006	0.9444 0.8941	0.8333 0.6029	1.0000 0.9606	0.6667 0.8044
7	0.0556 0.2708	0.0000 0.1101	0.6111 0.4340	0.2778 0.2711	0.0000 0.1160	0.0000 0.0650	0	0.1667 0.2394	0.2778 0.2826	0.1111 0.2236	0.3889 0.4212	0.6111 0.4780	0.1111 0.1053	0.7778 0.6128	0.2778 0.3566
8	0.7778 0.5825	0.3333 0.3541	0.6667 0.6334	0.5556 0.5244	0.3333 0.3629	0.0000 0.1947	0.8333 0.7606	0	0.4444 0.5429	0.7222 0.5435	0.8889 0.7320	0.5556 0.6634	0.0556 0.2654	0.9444 0.8205	0.4444 0.5848
9	0.5556 0.5469	0.2222 0.3068	0.6667 0.5919	0.5556 0.5131	0.2222 0.2746	0.1111 0.1883	0.7222 0.7174	0.5556 0.4571	0	0.5556 0.5041	0.6667 0.6703	0.3889 0.6411	0.3333 0.2917	0.7778 0.7825	0.6667 0.5958
10	0.6667 0.5592	0.2222 0.3134	0.6667 0.6277	0.6111 0.5243	0.2778 0.3326	0.1667 0.1987	0.8889 0.7764	0.2778 0.4565	0.4444 0.4959	0	0.7778 0.7111	0.8333 0.7099	0.0000 0.2527	0.7778 0.7924	0.7778 0.6076
11	0.2778 0.3692	0.1111 0.1636	0.4444 0.4668	0.3333 0.3497	0.2222 0.2085	0.0556 0.0994	0.6111 0.5788	0.1111 0.2680	0.3333 0.3297	0.2222 0.2889	0	0.5556 0.5237	0.0000 0.1235	0.7222 0.6793	0.6667 0.4660
12	0.1111 0.3234	0.2222 0.1810	0.4444 0.4609	0.1667 0.3160	0.1667 0.1944	0.0556 0.1059	0.3889 0.5220	0.4444 0.3366	0.6111 0.3589	0.1667 0.2901	0.4444 0.4763	0	0.2222 0.1694	0.5556 0.6059	0.4444 0.4439
13	0.9444 0.7680	0.6667 0.5821	0.7778 0.7884	0.7778 0.7291	0.7778 0.5545	0.1667 0.3971	0.8889 0.8947	0.9444 0.7346	0.6667 0.7083	1.0000 0.7473	1.0000 0.8765	0.7778 0.8306	0	0.8333 0.9213	0.5556 0.7545
14	0.1111 0.1918	0.1111 0.0966	0.1111 0.2977	0.1111 0.2213	0.0000 0.0835	0.0000 0.0394	0.2222 0.3872	0.0556 0.1795	0.2222 0.2175	0.2222 0.2076	0.2778 0.3207	0.4444 0.3941	0.1667 0.0787	0	0.3333 0.3061
15	0.5556 0.4604	0.3889 0.2648	0.2222 0.4852	0.4444 0.4389	0.1111 0.2503	0.3333 0.1956	0.7222 0.6434	0.5556 0.4152	0.3333 0.4042	0.2222 0.3924	0.3333 0.5340	0.5556 0.5561	0.4444 0.2455	0.6667 0.6939	0

Conclusions

We have seen that probabilistic inversion enables us to obtain utility values without introducing assumptions on the error term. In fact, we obtain a distribution over a population of utility functions which optimally reproduces discrete choice data, given a starting distribution. The starting distribution may be chosen to be minimally informative, or may impose constraints, for example that the utility values add to 1. The IPF algorithm converges very quickly IF it converges. In the case of infeasibility, the PARFUM algorithm is shown to distribute the lack of fit more reasonably, and yields better distributions.

Figure10: Correlation matrix Open Wiring, PARFUM (100 iterations)

Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0,0267	-0,0109	0,0045	0,0227	0,0541	0,0529	0,0178	-0,0289	-0,0022	0,0249	0,0086	0,0572	0,0795	-0,0070
2	0,0267	1	0,0600	0,0694	0,0514	0,0142	0,0853	0,0248	0,0271	0,0590	0,0555	0,0036	0,0019	0,0432	-0,0329
3	-0,0109	0,0600	1	0,0278	0,0286	0,0398	0,0226	0,0318	0,0187	0,0208	0,0150	-0,0157	-0,0116	0,0424	-0,0075
4	0,0045	0,0694	0,0278	1	0,0802	0,0344	0,0293	0,0270	-0,0051	0,0385	0,0341	0,0126	0,0175	0,0402	-0,0065
5	0,0227	0,0514	0,0286	0,0802	1	0,0367	0,1075	0,0269	0,0124	0,0458	0,0143	0,0086	0,0274	0,0798	0,0497
6	0,0541	0,0142	0,0398	0,0344	0,0367	1	0,1139	0,0770	0,0290	0,0356	0,0709	0,0685	0,0203	0,1077	-0,0025
7	0,0529	0,0853	0,0226	0,0293	0,1075	0,1139	1	0,0194	-0,0026	0,0260	0,0265	0,0037	0,0282	0,0289	-0,0138
8	0,0178	0,0248	0,0318	0,0270	0,0269	0,0770	0,0194	1	0,0167	0,0295	0,0451	-0,0329	0,0572	0,0680	-0,0190
9	-0,0289	0,0271	0,0187	-0,0051	0,0124	0,0290	-0,0026	0,0167	1	-0,0071	0,0010	-0,0088	0,0115	0,0320	0,0185
10	-0,0022	0,0590	0,0208	0,0385	0,0458	0,0356	0,0260	0,0295	-0,0071	1	0,0416	0,0452	0,0354	0,0431	-0,0002
11	0,0249	0,0555	0,0150	0,0341	0,0143	0,0709	0,0265	0,0451	0,0010	0,0416	1	0,0124	0,0734	-0,0006	-0,0470
12	0,0086	0,0036	-0,0157	0,0126	0,0086	0,0685	0,0037	-0,0329	-0,0088	0,0452	0,0124	1	0,0140	0,0091	-0,0030
13	0,0572	0,0019	-0,0116	0,0175	0,0274	0,0203	0,0282	0,0572	0,0115	0,0354	0,0734	0,0140	1	0,0288	-0,0036
14	0,0795	0,0432	0,0424	0,0402	0,0798	0,1077	0,0289	0,0680	0,0320	0,0431	-0,0006	0,0091	0,0288	1	0,0002
15	-0,0070	-0,0329	-0,0075	-0,0065	0,0497	-0,0025	-0,0138	-0,0190	0,0185	-0,0002	-0,0470	-0,0030	-0,0036	0,0002	1

Appendix

Lemma 3. If n = 3 and $\forall i,j,k = 1,2,3;\ i \neq j \neq k \neq i,\ \ p_{ij} + p_{jk} + p_{ki} \leq 2$, then there exists a distribution over 3! that expresses the pairwise preferences.

Proof: The permutation (1,2,3) gives rise to the inequality $p_{12} + p_{23} + p_{31} \leq 2$ as a necessary condition (lemma 2). The permutations (2,3,1) and (3,1,2) give rise to the same condition. The remaining permutations, (1,3,2), (3,2,1) and (2,1,3) give rise to the condition $p_{13} + p_{32} + p_{21} \leq 2$. Using $p_{ij} = 1\text{-}p_{ji}$, the latter condition can be re-written as $1 \leq p_{12} + p_{23} + p_{31}$. Hence we must prove that

3) $\quad 1 \leq p_{12} + p_{23} + p_{31} \leq 2$

is necessary and sufficient for the existence of a distribution over 3! expressing these preferences. Where $a,b,c \in \{1,2,3\}$ consider the "potential circular triad" formed by the string:

4) ab, bc, ca.

Where "ab" corresponds to "a better than b". If we reverse the meaning of "better" and "worse" the following pattern is obtained:

5) ba, cb, ac

This same pattern can be obtained by writing the potential circular triad (4) in reverse order, ca, bc, ab, and switching the labels "b"and "c". The probabilities of ab, bc, ca may be in any of six orders; the reader may verify that by order reversal and re-labeling the permutations can be brought into either format (4) or (5). Let us assume that

6) $p_{12} \geq p_{23} \geq p_{31}$.

which corresponds to (4). We want that $1 \geq p_{12} + p_{31}$; if this does not hold then $1 > p_{21} + p_{13}$. Switching the meaning of "better", and "worse", our pattern would switch from (4) to (5):

7) $p_{13} \geq p_{32} \geq p_{21}$ with $1 \geq p_{13} + p_{21}$.

Similarly, if we started with pattern (5), we could always arrange that either pattern (4) or (5) emerged with the sum of the most and least probable pair less than or equal to 1.

Therefore, without loss of generality, we may assume that either (7) holds or:

8) $p_{12} \geq p_{23} \geq p_{31}$ with $1 \geq p_{12} + p_{31}$..

The table below gives weights to the permutations which satisfy the constraints in these two cases, where $A = p_{12} + p_{23} + p_{31} - 1 \geq 0$:

	(123)	(132)	(312)	(213)	(231)	(321)
(7)	A	p_{12} –A	0	$1-p_{12} - p_{31}$	0	p_{31}
(8)	p_{12}	0	0	$1-p_{12} - p_{31}$	$A - p_{12}$	$1 - p_{13}$

The reader may check that the constraints are satisfied, eg in case (7);

$$p_{13} = p(123) + p(132) + p(213) = p_{12} + 1 - p_{12} - p_{31} = 1 - p_{31}. \quad \square$$

References

1. Anderson, S.P., A. de Palma and J-F Thisse (1996) Discrete Choice Theory of Product Differentiation, MIT Press, Cambridge.
2. Bradley, R. (1953) " Some statistical methods in taste testing and quality evaluation" *Biometrica,* vol. 9, 22-38.
3. Bradley, R. and Terry, M. (1952) "Rank analysis of incomplete block designs *Biometrica,* vol. 39, 324-345.
4. Csiszar. (1975) I-divergence geometry of probability distributions and minimization problems. Ann.of Probab., 3:146–158,.

5. Csiszar, I. and Tusnady, G.,(1984) "Information geometry and alternating minimization procedures", Statistics & Decisions, vol 1, pp=205-237.
6. David, H.A. (1957) The Method of Paired comparisons, Charles griffin, London, 1963.
7. Deming W.E. and Stephan F.F.. (1940)On a last squares adjustment of sampled frequency table when the expected marginal totals are known. Ann. Math. Statist., 11:427–444.
8. Du C., Kurowicka D., Cooke R.M.,(2006) Techniques for generic probabilistic inversion, Comp. Stat. & Data Analysis (50), 1164-1187.
9. Fienberg S.E. (1970). An iterative procedure for estimation in contingency tables. Ann. of Math. Statist., 41:907–917.
10. Haberman, S.J.(1974). An Analysis of Frequency Data. Univ. Chicago Press.
11. Haberman, S.J.(1984). Adjustment by minimum discriminant information. Ann. of Statist., 12:971–988.
12. Kraan, B., Bedford, 2004. Probabilistic inversion of expert judgements in the quantification of model uncertainty Management Sci.,
13. Kraan, B.C.P., Cooke, R.M., 2000. Processing expert judgements in accident consequence modeling. Radiat.Prot. Dosim. 90 (3), 311–315.
14. Kraan, B.C.P., Cooke, R.M., 2000a. Uncertainty in compartmental models for hazardous materials—a case study. J. Hazard. Mater. 71, 253–268.
15. Kruithof, J. (1937). Telefoonverkeersrekening. De Ingenieur, 52(8):E15–E25.
16. Kullback, S., 1968. Probability densities with given marginals. Ann. Math. Statist. 39 (4), 1236–1243.
17. Kullback, S., 1971. Marginal homogeneity of multidimensional contingency tables. Ann. Math. Statist. 42 (2), 594–606.
18. Kurowicka, D. and Cooke, R.M. Uncertainty Analysis with High Dimensional Dependence Modeling, Wiley, 2006.
19. Mazzuchi, T.A., Linzey, W.G. and Brunin, A. (2007) "An Expert Judgment Paired Comparison Analysis for Aircraft Wiring Risk Assessment", appearing in *Reliability Engineering and System Safety*.
20. McFadden, D. (1974) "Conditional logit analysis of qualitative choice behavior" in Zarembka, P (ed) *Frontiers in Econometrics*, New York Academic Press, USA, 105-142.
21. Mosteller, F. (1952) “Remarks on the method of paired comparisons: the least squares solution assuming equal standard deviations and equal correlations” *Psychometrica,* vol. 16, no. 1 3-9.
22. Thurstone, L. (1927) “A law of comparative judgment” *Pschyl. Rev*. vol. 34, 273-286.
23. Torgerson, W. (1958) *Theory and Methods of Scaling*, Wiley, New York, 1958.
24. Train K.E. (2003)“Discrete Choice Methods with Simulation” Cambridge University Press.