Boundary linear utility and sensitivity of decisions with imprecise utility trade-off parameters

Malcolm Farrow[1] and Michael Goldstein[2]

[1] School of Mathematics and Statistics, Newcastle University, U.K.
[2] Department of Mathematical Sciences, Durham University, U.K.

E-mail: Malcolm.Farrow@ncl.ac.uk, Michael.Goldstein@dur.ac.uk

1: Imprecise utility tradeoffs

Standard utility theory: The decision maker (DM) may state preferences between all combinations of outcomes.

Imprecise utility: DM can state preferences for some, but not all, outcomes. Imprecise utility is defined by obeying all of the constraints implied by the stated preferences.

Imprecise utility tradeoffs: We suppose that DM can make preference statements over all outcomes of each individual attribute, and so may specify precise marginal utilities, but can only make preference statements for some, but not all, combinations of the various attributes. Each such preference statement imposes constraints on the trade-off parameters which are used to combine the individual attributes into an imprecise multi-attribute utility.

- We construct a *utility hierarchy*.
- Marginal utilities, referring to individual attributes, are represented by *marginal nodes* at the lowest level.
- At each higher level, utilities, from *parent nodes*, are combined at *child nodes* until, at the highest level, we have one overall utility node.
- We choose the structure in such a way that, at each child node, we can assume that the utilities being combined are mutually utility independent.
- All utilities are on a *standard scale*
  - Worst outcome considered: $U = 0$.
  - Best outcome considered: $U = 1$.

Given mutual utility independence and the standard scale we have three kinds of nodes:

Additive node: Combining any number s of utilities

$$U(\underline{X}) = \sum_{i=1}^{s} a_i U_i(X_i)$$

where $\sum_{i=1}^{s} a_i \equiv 1$ and $a_i > 0$ for $i = 1, \ldots, s$.

Binary node: When exactly two utilities are combined

$$U = a_1 U_1 + a_2 U_2 + h U_1 U_2$$

where $0 < a_i < 1$ and $-a_i \le h \le 1 - a_i$, for $i = 1, 2$, and $a_1 + a_2 + h \equiv 1$.

Multiplicative node: When more than two utilities are combined

$$U(\underline{X}) = \frac{\prod_{i=1}^{s}[1 + k a_i U_i(X_i)] - 1}{\prod_{i=1}^{s}(1 + k a_i) - 1}$$

with $a_i \equiv 1, k > -1$ and, for $i = 1, \ldots, s$, we have $a_i > 0$ and $k a_i > -1$.

Example: Designing a new course module at a university

S_1 short term learning,
S_2 longer-term learning,
S_3 student satisfaction,
V_1 staff satisfaction,
V_2 institutional benefits,
V_3 staff development,
C financial cost.

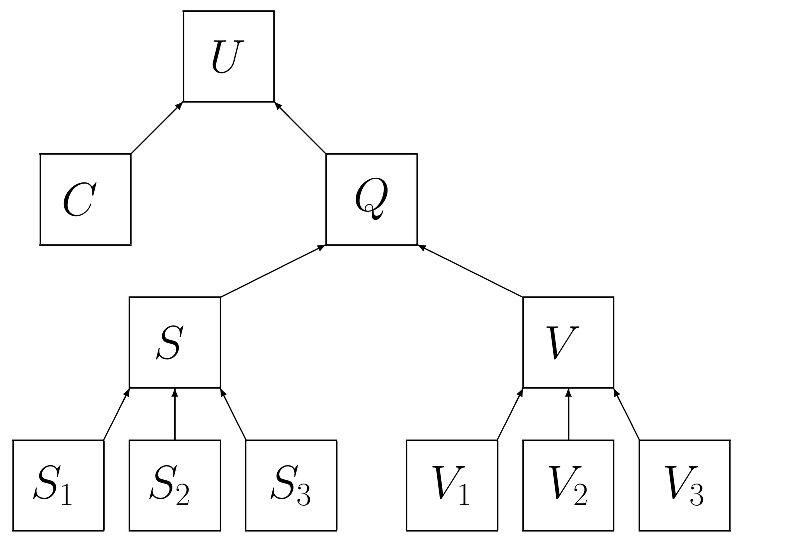

FIGURE 1: Utility hierarchy for the course design example.

Imprecise trade-off parameters
Attribute comparisons at node n determine a convex region R_n with vertices $P_n = \{\phi_n^{(1)}, \ldots, \phi_n^{(r_n)}\}$. See Figure 2.

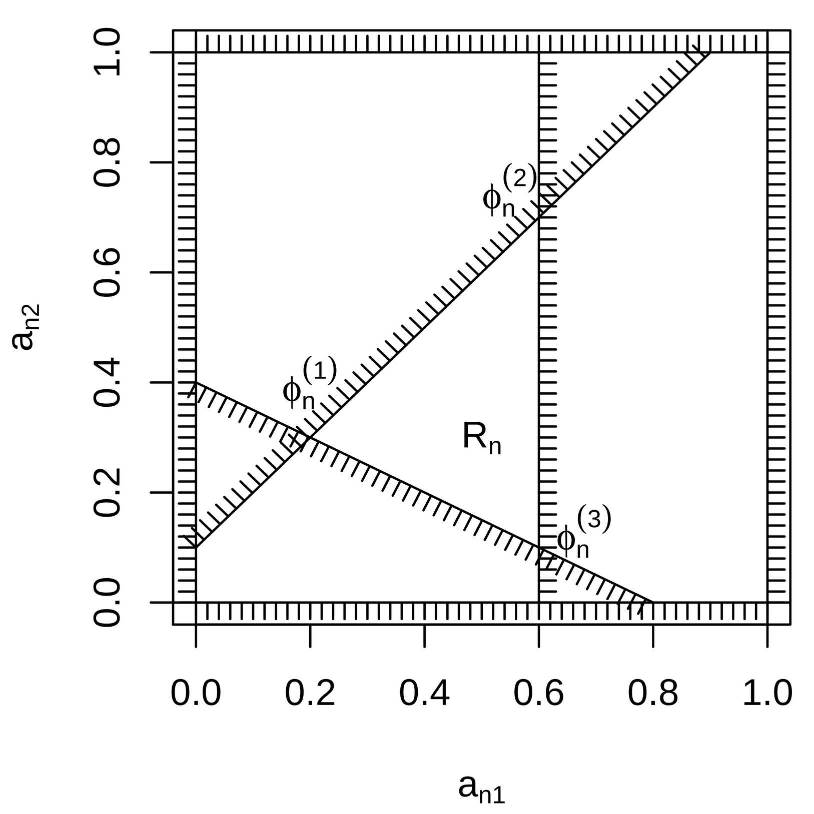

FIGURE 2: Elicitation and feasible set: Binary node.

Combining all N nodes:

$$R = R_1 \times R_2 \times \cdots \times R_N$$
$$P = P_1 \times P_2 \times \cdots \times P_N$$

Example: Choices

- The course has six units.
- Each unit can be taught by one of three methods:
  1. Lectures.
  2. Computer laboratory.
  3. "Open Learning".
- There are thus $3^6 = 729$ possible choices.
- Denote a choice $(\mu_1, \ldots, \mu_6)$.

Example: Pareto and almost-Pareto optimality

- In [1] we showed how to find the Pareto-optimal choices using only evaluations at the vertices. There are 50. Of these 37 can be eliminated because they are equivalent to other retained choices.
- In [2] we showed how to use "almost-preference" to reduce the number of choices further. This leaves us with six.

2: Boundary linear utility

Definition and motivation

- Let U_i be the utility function determined by the trade-offs $\underline{\theta}^{(i)} \in P, i = 1, \ldots, r.$ Any function of the form

$$\bar{U}_\lambda = \sum_{i=1}^{r} \lambda_i U_i$$

  where $\lambda = (\lambda_1, \ldots, \lambda_r)$ are non-negative constants such that $\sum_{i=1}^{r} \lambda_i = 1$ is a *boundary linear utility*.
- For any such $\bar{U}_\lambda$, we may find d which maximises $\bar{U}_{d,\lambda} = \sum_{i=1}^{r} \lambda_i U_{d,i}$, where $U_{d,i}$ is the utility of alternative d with trade-off $\underline{\theta}^{(i)}$.
- By varying λ we can emphasise or de-emphasize different parts of the feasible set.
- This gives us an alternative metric for the feasible set. The distance between any two vertices is now $\sqrt{2}$.
- For further motivation, see [1]

Properties
For simplicity assume here that we have only additive and binary nodes, i.e. a "SIIH" (Simple Imprecise Independence Hierarchy).

- A choice which is either (i) a unique Bayes decision for some $\bar{U}_\lambda$, or (ii) a Bayes decision for some $\bar{U}_\lambda$ with $\lambda_i > 0$ for $i = 1, \ldots, r$, is Pareto optimal over R.
- Let $\lambda(i_1, \ldots, i_N)$ be the weight applied to the vertex combination $\underline{\phi}_1^{(i_1)}, \ldots, \underline{\phi}_N^{(i_N)}$. Let $\lambda_{n,i}$ be the weight for vertex $\underline{\phi}_n^{(i)}$ at node n. If $\lambda(i_1, \ldots, i_N) = \prod_{n=1}^{N} \lambda_{n,i_n}$ we have a *multiplicative weighting*.
  - For any $\underline{\theta} \in R$, there exists a multiplicative weighting λ such that $\underline{\theta} = \bar{\underline{\theta}}_\lambda$.
  - For any multiplicative weighting λ, there exists a $\underline{\theta} \in R$ such that $\underline{\theta} = \bar{\underline{\theta}}_\lambda$.
  - If λ is a multiplicative weighting then $\bar{U}_\lambda = U(\bar{\underline{\theta}}_\lambda)$.

Example With equal λ-weights, $\mathrm{E}(\bar{U}_\lambda)$ is maximised by d_1 but how sensitive is this to choice of λ?

d_1	1, 3, 1, 1, 3, 2	d_4	1, 3, 1, 1, 1, 3
d_2	2, 3, 1, 3, 3, 2	d_5	1, 3, 1, 1, 3, 3
d_3	1, 3, 1, 3, 3, 2	d_6	2, 3, 1, 1, 3, 2

TABLE 1: Alternatives for comparison

References

[1] M. Farrow and M. Goldstein. Trade-off sensitive experimental design: a multicriterion, decision theoretic, Bayes linear approach. *Journal of Statistical Planning and Inference*, 136:498–526, 2006.

[2] M. Farrow and M. Goldstein. Almost-Pareto decision sets in imprecise utility hierarchies. *Journal of Statistical Theory and Practice*, 3:137–155, 2009.

3: Exploring sensitivity

- Set of remaining alternatives : D.
- Proposed choice : d^*.
- "Almost equivalent" utility difference : ε.
- Original λ specification : λ_0.
- λ value at vertex v : λ_v.
- $\bar{U}_\lambda(d_j) - \bar{U}_\lambda(d_k) = \delta(d_j, d_k;\ \lambda)$.

Volume sensitivity
Compute the volume of λ-space, as a proportion of the total volume within which $\sum \lambda_j = 1$, over which the difference in utility between alternative d^* and each of the other retained alternatives is at least $-\varepsilon$.
Computation: Hierarchies may be large and we may require such comparisons for many decision pairs. We exploit the formal equivalence between these calculations and those that we would make to evaluate the probabilities of certain events given a uniform probability specification over the elements of λ. This allows us to use the local computation properties of the implied graphical model to simplify the calculations that we require.

Distance in λ-space
For each $d_j \in D$, identify vertices v where $\delta(d_j, d^*;\ \lambda_v) \ge \varepsilon$. For each such v, find

$$T_v = t_v \sqrt{(\lambda_v - \lambda_0)'(\lambda_v - \lambda_0)}$$

where

$$t_v = \frac{\delta(d_j, d^*;\ \lambda_0) - \varepsilon}{\delta(d_j, d^*;\ \lambda_0) - \delta(d_j, d^*;\ \lambda_v)}.$$

Large T_v suggests robustness of $d^\star$.

Sensitivity in the θ-metric

- Original central parameter value: $\underline{\theta}_0$.
- $P = \{\underline{\theta}^{(1)}, \ldots, \underline{\theta}^{(r)}\}$.
- $P_t = \{\underline{\theta}_t^{(i)} \ :\ \underline{\theta}_t^{(i)} = \underline{\theta}_0 + t(\underline{\theta}^{(i)} - \underline{\theta}_0)\},\ (0 \le t \le 1)$.
- *Scaled range* R_t is the convex hull of P_t.
- Difference in expected utility: $\delta(d_j, d_k;\ \underline{\theta})$.
- Find $\max_{P_t}\{\delta(d_j, d^*;\ \underline{\theta}_t^{(i)})\}$.

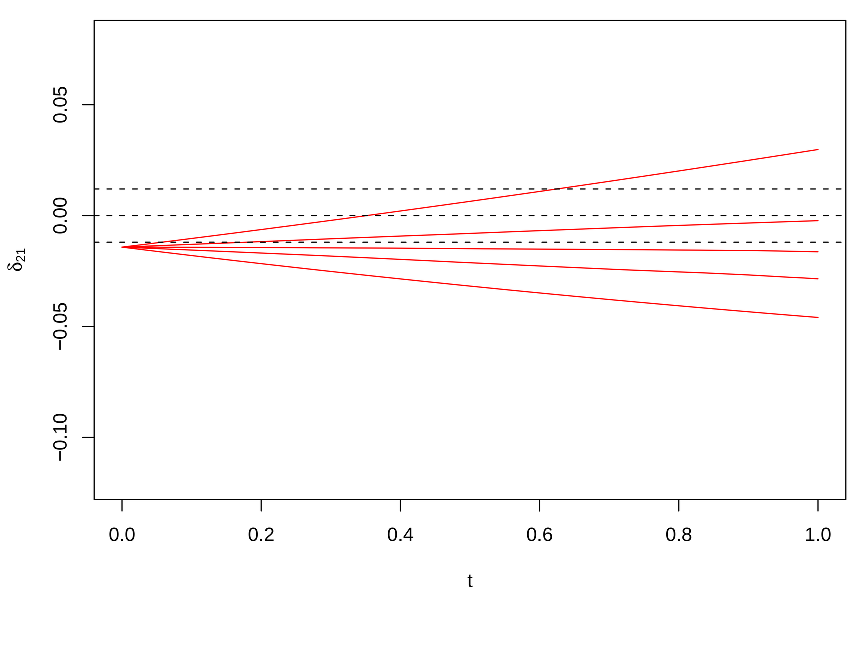

FIGURE 3: Expansion with respect to all parameters. Maximum, quartiles and minimum of the difference in expected utility between d_2 and d_1 at 144 vertices, against expansion factor t. Reference lines are given at zero and $\pm\varepsilon$.