

Using imprecise estimates for weights.

Alan Jessop
Durham Business School
Mill Hill Lane
Durham
DH1 3LB
UK

tel: (0)191 334 5403

email: a.t.jessop@durham.ac.uk

Abstract

In multiattribute decision problems the decision to differentiate between alternatives will be affected by the precision with which weights are specified. Specifications are imprecise because of the uncertainty characteristic of the judgements on which weights are based. Uncertainties are from two sources, the accuracy with which judgements are articulated and the inconsistency when multiple judgements are made and must be reconciled. These uncertainties are modelled using probabilistic weight estimates integrated by the Dirichlet distribution. This ensures the consistency of the estimates and leads to the calculation of significance of the differences between alternatives. A simple plot of these significant differences helps in the final decision whether this is selection or ranking. The method is used to find weight estimates in the presence of both types of uncertainty acting separately and together.

Keywords: multicriteria, weights, probability, Dirichlet

Using imprecise estimates for weights.

Introduction

Developing a model to help a decision process is necessarily iterative. The forms of and relations between model, data, and parameter values change as the understanding of the user changes. They are also a vehicle for exploration and reflection by the user so that judgements are altered until the “form and content [are] sufficient to solve the problem” (Phillips 1984). This process attempts to resolve the many uncertainties inherent in the model and its use (French 2003, 1995). Sometimes the problem may be such that all uncertainties must somehow be resolved before a satisfactory decision is made but, equally, some residual uncertainty may be tolerated. For example, if the task is to make a short list of candidates for further consideration it is not necessary to discriminate between those on the short list, only to believe that they are better than the rest.

Some parameter values will be based in whole or in part on judgement. At any stage in this process there is a need both to articulate this judgemental uncertainty and also to have a means of helping to reduce it by developing a better understanding of preferences through an exploration of their implications. Sensitivity analysis helps this exploration by testing the effects of changes in parameter values so that some reduction, perhaps resolution, of uncertainty may be achieved. There are many ways of structuring sensitivity analyses to help in this (see, for instance, French 1992; Insua and French 1991). Variations in parameter values may be considered one at a time or in combination (French 2003). For example, in multiattribute problems Mustajoki, Hämäläinen and Lindstedt (2006) describe three forms of sensitivity analysis. First, a single parameter test in which one weight is varied and the effect on scores observed. The results are easily shown in a simple diagram. Second, a multiparameter test in which several weights are varied. While this enables the effects of weight interactions to be explored the depiction of the results is not easy for more than two or three weights. Thirdly, a global sensitivity analysis assesses the effects of imprecision in all weights, most often by specifying probability distributions for weights and then using Monte Carlo analysis, though in some cases an analytical approach may be feasible, and preferable. The results of global analysis may be simply shown in two-dimensional plots (Kruskal and Wish 1978), an idea applied to multicriteria problems by Clarke and Rivett (1978; Rivett 1977).

Single parameter and global sensitivity analyses have different purposes and languages and so are used in different ways in the interaction between model and user. Considering all uncertainties together, whether by simulation or an analytical model, uses probability distributions for input and so also for output. While it is fairly straightforward to specify inputs interpreting outputs in the context of a decision problem may be more difficult.

Standard reports such as confidence interval estimates of differences between performance measures provide an easy summary. This may be enough: given the uncertainties in inputs it is possible to decide that one alternative is superior to another, even though the magnitude of the difference is not known exactly. If this is not the case a modification of inputs (smaller variances) will give more discrimination in the output. This process gives a sequence of groups or clusters, starting with one undifferentiated group of all alternatives and producing increasingly more, and smaller, groups as uncertainties are reduced and discrimination increases.

The purpose of this paper is to demonstrate the feasibility of a simple analytical model for global sensitivity analysis for those comfortable with a probabilistic approach. The analysis uses techniques separately familiar elsewhere but brought together for this particular application to the simple multiattribute scoring model. Scores are the weighted sum of attributes. Values for weights are inferred from preference statements and it is this source of uncertainty which is the object of the model.

This paper is organised as follows: the sources of uncertainty are outlined; a model incorporating uncertainty is described and an example using direct rating given. Extensions to other methods involving several estimates of the same weights are shown and results discussed.

Uncertainty about weights

The multiattribute model considered is

$$y_j = \sum_i w_i x_{ij} \quad ; \quad i = 1 \dots n, j = 1 \dots m \quad (1)$$

$$\text{with} \quad \sum_i w_i = 1 \quad (2)$$

and where y_j is the score for alternative j , w_i is the weight attached to attribute i and x_{ij} is an appropriately scaled measure of the value of attribute i for alternative j . The scaling ensures that for each variable either the range is $[0,1]$ or that mean = 0 and standard deviation = 1. If the unscaled i th attribute has values q_{ij} with minimum and maximum values q_{min} and q_{max} and mean and standard deviation \bar{q} and s_q then either

$$x_{ij} = (q_{ij} - q_{min}) / (q_{max} - q_{min}) \quad (3)$$

$$\text{or} \quad x_{ij} = (q_{ij} - \bar{q}) / s_q \quad (4)$$

with obvious adjustments if small values are preferred .

There are a number of methods which may be used to derive weights from preference statements. Different methods generally give different weights, mainly because of the different modes of elicitation (Pöyhönen and Hämäläinen 2001). It is not the purpose in this paper to compare such popular alternatives as SMART, SWING and AHP, simply to show how an analytical approach may be used to model uncertainty in some different methods

Saaty and Vargas (1987) identify two types of uncertainty, that arising from uncertainty about events and that about making judgemental preference statements, attributing this second to limits of information and understanding. Lavy and Wan (1998) describe both uncertainty about the future context for the decision and of making judgements of pairwise weight ratios. Hauser and Tadikamalla (1996) cite uncertainties about facts and also a lack of agreement between decision makers. Whatever the sources of uncertainty the effect is the same; an inability to provide precise estimates of weights.

What is a sensible response to these difficulties? Barron and Barrett (1996) speculate that “the pursuit of precise weights may be an illusion”, that trying to elicit exact weights is problematic because the result is likely to depend on the method used and because the exactness of the weights obtained “imposes a precision which may be absent in the mind of the decision maker”. Just what is in the mind of the decision maker is unknown and, probably, unknowable, perhaps even by the decision maker. Questions are asked which require answers based on some mental process we call judgement. These answers are the data for a model which uses weights as a description of the judgements.

Uncertainty in judgement leads to uncertainty about weights. This may be described by specifying ranges (e.g. Mustajoki, Hämäläinen and Sahlo 2005) or probability distributions. Probabilistic models of imprecision in weight specification are usually found in studies which seek to explore the impact of uncertainty on model structure and performance (Moskowitz, Tang and Lam 2000; Fischer, Jia and Luce 2000). Similarly probabilistic models for practical decision support are harder to find, although using the cumulative probability distribution (risk function) for each score and the identification of stochastic dominance has been proposed (Moskowitz, Tang and Lam 2000) as have the modelling of a probability distribution of the rank of each alternative (Bañuelas and Antony 2007; Jessop 2002; Butler, Jia and Dyer 1997) and the probability of rank reversal (Stam and Duarte Silva 1997; Saaty and Vargas 1987).

Modelling uncertainty

When making probabilistic judgements about weights assessors will have in mind marginal distributions. Because of (2) these distributions cannot be independent. The Dirichlet

distribution provides an appropriate model of the joint distribution to integrate individual weight estimates (Hora 2007; Butler, Jia and Dyer 1997):

$$f(W) = k \prod_i w_i^{u_i-1} ; \quad 0 < w_i < 1, \sum_i w_i = 1, u_i \geq 0, \forall i \quad (5)$$

$$k = \Gamma(\sum_i u_i) / \prod_i \Gamma(u_i) \quad (6)$$

which has Beta marginal distributions with properties

$$\mu_i = u_i / v \quad (7)$$

$$\sigma_i^2 = u_i(v - u_i) / v^2(v+1) \quad (8)$$

and covariances

$$\sigma_{ij} = -u_i u_j / v^2(v+1) \quad ; \quad i \neq j \quad (9)$$

$$\text{where } v = \sum_i u_i \quad (10)$$

In Bayesian analysis (e.g. Congdon 2001; DeGroot 1970) the Dirichlet distribution is the conjugate prior for a process with a multinomial likelihood. As data are collected parameter values are updated, increasingly higher parameter values corresponding to more data and so to reduced variance. Fischer, Jia and Luce (2000) make an analogy with respect to weight estimates; that decision makers who feel themselves to have greater expertise or familiarity with the assessment model may give estimates with smaller marginal variance and that this may be seen as stored experience. Whether smaller variance represents greater experience, technical familiarity with the model or unjustified self-assurance may not be clear but, whichever it is, the effect is modelled in the same way: the smaller the variance the larger the value of v .

Marginal variances will be inconsistent with proper Dirichlet marginal probability distributions in that they will not conform to (8). A reconciliation may be found by treating v as a parameter which controls the overall level of variance and finding a compromise value which ensures that the Dirichlet conditions are met. Using the mean and variance, e_i and s_i^2 , of each weight estimate for μ_i and σ_i^2 in (7) and (8) gives an estimate for v from the i th weight:

$$v_i = [e_i(1 - e_i) / s_i^2] - 1 \quad (11)$$

The mean

$$\bar{v} = \sum_i v_i / n \quad (12)$$

gives a compromise value of v from which parameter values

$$u_i = \bar{v} \cdot e_i \quad (13)$$

can be used in (8) and (9) to give a variance/covariance matrix. The less the uncertainty in the marginal estimates the smaller will be the variances and so the higher the value of \bar{v} , which may therefore serve as an indicator for the overall uncertainty of the weight estimates.

The different estimates, v_i , have some diagnostic value. Weights for which v_i is low compared to the summary \bar{v} are those which contribute most to overall imprecision. The judgements made about these particular weights are those which might most usefully be reconsidered if more precision is needed.

Use of the Dirichlet distribution does not depend on particular procedures for weight elicitation. Estimates of mean and variance may be obtained from direct methods, matrices of weight ratios or any other means thought satisfactory for a given application. If the estimates are a summary of a number of different assessments, as when the judgements of a number of assessors are combined, marginal means and variances are available directly. If estimates are inferred from individual preference statements means and variances may be found using estimations familiar in, for example, PERT analyses. The underlying distributions are assumed to be Beta, which are the marginal distributions of the Dirichlet. Keefer and Verdini (1993) and Keefer and Bodily (1983) compare the accuracy of several estimators. The results given by Keefer and Bodily are used here. Low, central and high estimates (l , c , h) are given. If c is taken as the mode and l and h are percentiles an estimate for the mean is

$$e = ac + (1-a)(l + h)/2 \quad (14)$$

with $a = 0.32$ for a 90% interval (Perry and Greig 1975) and $a = 0.16$ for an 80% interval (Keefer and Bodily 1983). If the central estimate is interpreted as a median then $a = 0.63$ for a 90% interval (Pearson and Tukey 1965) and $a = 0.40$ for an 80% interval (Swanson in Megill 1977). In this paper modal estimates are used. Estimates of variance are given by

$$s^2 = [(h - l) / b]^2 \quad (15)$$

where $b = 3.25$ for a 90% interval (Pearson and Tukey 1965) and $b = 2.65$ for an 80% interval (Moder and Rogers 1968 as modified by Davidson and Cooper 1976).

An example

Data on MBA programmes are used for ranking and as an aid to selection. The data on full-time MBA programmes published in the *Financial Times* of 29 January 2007 are used here. Nine attributes were chosen, each given as a percentage (the percentage of the MBA cohort that were women, and so on):

- 1 Salary increase
- 2 Aims achieved
- 3 Employment at 3 months
- 4 Women faculty
- 5 Women students
- 6 Women board
- 7 International faculty
- 8 International students
- 9 International board

In the analysis each is scaled according to (4), as was done by the newspaper.

An MBA alumna was asked to provide weights using the SMART method. First she ranked the attributes then gave the most important a weight of 100. Lower ranked attributes were given smaller weights. Finally, for all but the highest reference weight high and low estimates were given. It was explained that these limits should not be absolute and would be interpreted as bounds of a 90% interval. The results are shown in Table 1. Means were calculated using (14) and scaled so that $\sum e = 1$:

$$e_i = c_i / \sum_i c_i. \quad (16)$$

The values of l and h were scaled by the same factor and were then used to calculate standard deviations s using (15). (This is also denoted by σ_a to indicate uncertainty due to inaccuracy of response, as discussed in the next section.) Dirichlet scale factors ν were found from (11). The marginal Dirichlet standard deviations, σ_D , from (8), are also shown and include an estimate for the anchor weight, w_2 . It is a useful characteristic of the method that what might be seen as missing data, probabilistic estimates for the anchor, do not mean that uncertainty estimates cannot be made.

For those weights for which $v < \bar{v}$ the uncertainty after integration in the Dirichlet distribution is less than that specified, $\sigma_D < \sigma_a$, and vice versa.

The task in either ranking or selection is to decide, first, if it is justifiable to believe that two programmes are different and only then, second, to decide which is superior. Uncertainty about the score for programme k is

$$\text{var}(y_k) = \sum_i \sum_j \sigma_{ij} x_{ki} x_{kj} \quad (17)$$

where values of σ_{ij} are from (8) and (9).

The difference in the scores of programmes k and l is statistically significant if

$$z_{k,l} = (|y_k - y_l| - \theta) / [\text{var}(y_k) + \text{var}(y_l)]^{0.5} \geq z_{\alpha/2} \quad (18)$$

where $z_{\alpha/2}$ is the critical value for a two-tailed significance test with significance level α . The parameter θ is the test value of the difference. To test whether it is justifiable to believe that there is some non-zero difference set $\theta=0$. This is common in hypothesis testing and is used here. (To identify as justifiably distinct only pairs with scores different by some larger margin set $\theta>0$.) Although significance testing has for long been the subject of dispute (Ziliak and McCloskey 2008; Morrison and Henkel 1970) z values usefully summarise the effect of uncertainty on the attribution of difference: the greater z the less the risk of unjustifiably differentiating alternatives.

Using twenty US MBA programmes from the *Financial Times* listing and the weights shown in Table 1 the resultant performance differences are shown in Figure 1. The numbers show programmes by rank. The plot is constructed so that the distances between pairs of alternatives correspond closely to their z value. This correspondence is characteristically high; $r>0.9$. The axes are arbitrary in that they are chosen just to maximise this correlation. The axis values are not shown here so that the diagram, a decision aid, has no detail not needed for this problem. Links are shown to identify those pairs which cannot sensibly be differentiated. In Figure 1 it is easy to see that there are four clusters of programmes and that the four most highly ranked are the most weakly clustered. Given the uncertainties of the weight assessments it may be that no more can be said. But this may be enough. If the purpose is to make a short list then it seems clear that the first four alternatives are that list. If the purpose is to make a final selection then some discussion about the first two programmes is needed.

Sources of uncertainty

The example shows how three point weight estimates can be used to produce a diagram showing the consequent justifiable discrimination between programmes. The same framework can be used with different sources of uncertainty. In this section a typology is given showing three general cases, each of which is illustrated using the same MBA data.

The number and type of questions asked and the style of the answers given are both possible sources of uncertainty in the elicitation process. Paulson and Zahir (1995) distinguish between inconsistency, when results are contradictory, and uncertainty arising “from doubts expressed by an individual decision maker as to the accuracy of his or her judgements.” Describe these by variances σ_c^2 , for the uncertainty arising from contradictions, and σ_a^2 for that resulting from imperfect accuracy of response.

If the elicitation requires more than one estimate for each weight (from judgements made at different times, say, or by different people) then the variance of these different estimates, σ_c^2 , is a measure of the uncertainty arising from contradictions between assessments (Kleinmuntz 1990). Alternatively, if exactly $n-1$ questions are asked to determine n weights, as in the SMART (Edwards 1977) or SWING (Edwards and Barron 1994) methods, the weight estimation problem has zero degrees of freedom and so no way of assessing σ_c^2 .

Whatever the questions, answers may take one of two forms. If single point estimates are given then no estimate of the inaccuracy of response is possible, but if answers are given probabilistically σ_a^2 can be found, as in the example above.

Presuming that these two sources of uncertainty – consistency and accuracy – are independent the variance of weight estimates is $\sigma_w^2 = \sigma_c^2 + \sigma_a^2$. Table 2 shows the situation. There are three cases depending on which source or sources of uncertainty are considered.

Case A: $\sigma_w^2 = \sigma_a^2$

Most of the direct elicitation methods use a reference point or anchor based on an initial ranking of attributes. Because these elicitation have zero degrees of freedom assessments must be made probabilistically. The $n-1$ evaluations contribute to finding the mean value of v but uncertainty estimates are found for all n weights. The example above using SMART showed this (Table 1).

Case B: $\sigma_w^2 = \sigma_c^2$

The uncertainty measured by σ_c^2 describes the distribution of a number of estimates provided by different people (Moskowitz, Tang and Lam 2000) or by different methods. There are a number of point estimates for each weight. The mean and variance of each weight estimate can be found directly. For example, when a number of assessors have each provided estimates.

Alternatively, the same assessor may provide more than one estimate for each weight, as in the specification of weight ratios $w_i/w_j = a_{ij}$. This method is closely identified with the Analytic Hierarchy Process, AHP, but can be used separately. While symmetry is commonly assumed ($a_{ji} = 1/a_{ij}$) it is not a requirement. Though this assumption of symmetry halves the work of the assessor it may mask the full effect of uncertainty. These estimates are inevitably inconsistent ($a_{ik}a_{kj} \neq a_{ij}$) and so a weight set is found which is in some sense a best compromise (best fit) to the pairwise comparisons. The most frequently cited method is the eigenvector model of Saaty (1977). A number of studies have used simulation to investigate the effects of uncertainty on the ranking of alternatives found in this way (Bañuelas and Antony 2007; Lipovetsky and Tishler 1999; Levary and Wan 1998; Stam and Duarte Silva 1997; Hauser and Tadikamalla 1996; Saaty and Vargas 1987; Vargas 1982).

There are a number of other methods for analysing pairwise judgements. Choo and Wedley (2004) divide the methods into those which seek to optimise some function of the sum of differences ($a_{ij} - w_i/w_j$) and those based on an aggregation of the columns of the matrix of a values. They recommend the use of a normalised column sum as giving good estimates for a range of problems. This simple method is found in standard management science texts (Albright and Winston 2007, Taylor 2007) recommended in its own right but also as a good approximation to Saaty's results. The method relies on the observation that each column of the table of a values provides an unscaled estimate of the weight distribution. Scaling each column to sum to 1 gives estimates

$$g_{ij} = a_{ij} / \sum_i a_{ij} \quad (19)$$

of the weight w_j with w_i as the reference. The mean of these values

$$e_i = \sum_j g_{ij} / n \quad (20)$$

is a point estimate of weight w_i . The estimation usually goes no further, thereby failing to exploit the positive degrees of freedom available. The variances of the estimates of w_i , s_i^2 , permit the calculation of v using (11).

As illustration the alumna who gave the judgements shown in Table 1 was asked to make a set of paired evaluations using the familiar nine point scale for judgemental estimates of the ratios (e.g. Vargas 1982). She made these evaluations two weeks after the first. The results are shown in Table 3 and the weight estimates in Table 4.

With a greater number of attributes evaluation fatigue may result in an incomplete set of ratios. However, the averaging used is not, in principle, affected by missing data provided that such gaps are not so numerous as to undermine the process.

$$\text{Case C: } \sigma_w^2 = \sigma_c^2 + \sigma_a^2$$

There are two circumstances where this might arise: when probabilistic estimates are made by a number of assessors and when a number of probabilistic estimates are made by the same assessor.

First, consider that there are two or more assessors. Three MBA students were each asked to make weight assessments in the manner shown in Table 1 and to use them to make a ranking of the US MBA programmes. Their individual estimates were modelled as in Case A (Table 1) and are shown in the left-hand half of Table 5. Each of the three estimates (e_1 , e_2 and e_3) provides a point estimate of the weight. The mean these three values is e and the standard deviation is σ_c , a measure of the variation due to the different assessors.

The imprecision of the assessments themselves is σ_a^2 , the mean of the three variances (s_1^2 , s_2^2 , s_3^2), a simple yet effective aggregation (Clemen and Winkler 2007). The two estimates of uncertainty are summed to give $\sigma_w^2 = \sigma_c^2 + \sigma_a^2$ which, with e , provide the Dirichlet parameters as before.

The second case is when the same assessor makes a number of estimates for each weight. This could be at different times or in different circumstances. An example is provided when weight ratios are specified not as single point estimates but as three point estimates incorporating uncertainty (as in, for instance, Bañuelas and Antony 2007). The judgemental inputs in Table 3 were extracted from just such an evaluation, shown in full as Table 6.

The resulting nine columns of three-point estimates were treated just as the different assessors' estimates in Table 5. The results are shown in Table 7.

Discussion

This paper brings together and supplements existing methods to provide a treatment of the uncertainties inevitable in the statement of preferences and one which, via a simple diagram, provides a guide to what discrimination between alternatives may, and may not, be justified. For simple models such as this weighted sum an analytical approach is more convenient than a simulation but otherwise plays the same role. None of the constituents is new: probabilistic models of weights, with and without the Dirichlet distribution; the use of three point estimates; two dimensional plots have all been used for some years. Bringing them together in this way has not been done before. The purpose is to demonstrate the feasibility of a decision aid which uses probability to describe uncertainty.

Wallsten (1990) notes that decision makers “feel best served when representations of uncertainty are as precise as possible, but no more precise than warranted”. It is in this spirit that a probabilistic approach is offered. For some users the language of probability may be

unfamiliar and so inappropriate for them: they will prefer single parameter sensitivity tests. However, the probabilistic approach takes account of all uncertainties simultaneously so that the user may have confidence in the results which can convincingly be communicated to decision makers (Mustajoki, Hämäläinen and Lindstedt 2006). This communication should be couched in terms of justifiable discrimination between alternatives. The simple diagram (Figure 1) helps.

The application was illustrated using two popular methods for weight elicitation, SMART and weight ratios, as well as the aggregation of different judgements. There exist strongly held views about the different methods used for multiattribute modelling. No such views are offered in this paper. These methods were used as illustrations only.

A number of applications of probabilistic models for weights have been concerned with the likelihood of rank reversal. This occurs only if it is possible that the relative scores of two alternatives might be reversed. If it is unlikely that the difference between the scores is zero then it is correspondingly unlikely that there will be rank reversal. In looking at significant differences, as in Figure 1, it is implicit that the stability of the ranks is also addressed.

In both examples of Case C (Tables 5 and 7) it is notable that $\sigma_c > \sigma_a$. The results are given as illustrations of the feasibility of the proposed method of modelling weights and not as part of an argument about the relative importance of different sources of uncertainty. But if it were generally the case that uncertainty in articulation was the smaller this would argue in favour of elicitation methods with positive degrees of freedom. While elicitation methods with no degrees of freedom (Case A) make life easier for the user they necessarily cannot afford the means to estimate uncertainty due to contradictions, σ_c .

It was assumed that the sources of uncertainty are independent, that the differences between assessments is unrelated to the precision with which those assessments are expressed. This was certainly the case when more than one assessor was used (the correlation between σ_c and σ_a in Table 5 is $r = 0.07$) but for the multiple assessments of a single assessor it was not (for the results shown in Table 7 $r = 0.94$). Positive correlations will increase the overall uncertainty, σ_w , and, were these dependencies shown to be generally characteristic, the calculation of σ_w should take them into account.

While none of the constituents used in the model is problematic the use of a probabilistic model for weights requires a difference in approach when compared with the use of single parameter sensitivity analysis. The forms of interaction are different. Whether it is more difficult, and less useful, to see the effects of altering three-point estimates, even if just one at a time, than changing point estimates is not resolved in this paper. The four MBA students (a small and particular sample, to be sure) who provided the data for the illustrations reported no difficulties. Global sensitivity analysis is, by definition, comprehensive. It speaks

probability and so the flavour of the argument is about what level of discrimination is justifiable. The price paid is that some may find this a too abstract language. In that weights are found as much by interaction with a problem as by contemplation of some inner dispositions it may well be that some users will not find the global model helpful. Both the nature of what constitutes justification in a particular case and the differences in the language used in the interactions will determine which approach a user will prefer.

It was the object of this paper to establish the feasibility of the probabilistic model of uncertainty about weights and this has been done. Testing the utility of the approach on a wide range of problems, and users, remains to be done.

References

- Albright SC and Winston WL (2007). *Management Science Modeling*. Thomson: Mason, OH.
- Bañuelas R and Antony J (2007). Application of stochastic analytic hierarchy process within a domestic appliance manufacturer. *Journal of the Operational Research Society*. **58**: 29–38.
- Barron FH and Barrett BE (1996). Decision quality using ranked attribute weights. *Management Science* **42**: 1515–1523.
- Butler J, Jia J and Dyer J (1997). Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research* **103**: 531–546.
- Choo EU and Wedley WC (2004). A common framework for deriving preferences from pairwise comparison matrices. *Computers & Operations Research* **31**: 893–908.
- Clarke D and Rivett BHP (1978). A structural mapping approach to complex decision-making. *Journal of the Operational Research Society* **29**: 113–128.
- Clemen RT and Winkler RL (2007). Aggregating probability distributions. In: Edwards W, Miles RF and von Winterfeldt D (eds.) *Advances in Decision Analysis: From Foundations to Applications*. Cambridge University Press: New York.
- Congdon P (2001). *Bayesian Statistical Modelling*. Wiley: Chichester.
- Davidson LB and Cooper DO (1976). A simple way of developing a probability distribution of present value. *Journal of Petroleum Technology* (September): 1069–1078.
- DeGroot MH (1970). *Optimal Statistical Decisions*. McGraw-Hill: New York.
- Edwards W (1977). How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems, Man and Cybernetics* **SMC-7**: 326–340.
- Edwards W and Barron FH (1994). SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes* **60**: 306–325.
- Fischer GW, Jia J and Luce MF (2000). Attribute conflict and preference uncertainty: the RandMAU model. *Management Science* **46**: 669–684.
- French S (2003). Modelling, making inferences and making decisions: the roles of sensitivity analysis. *TOP* **11**: 229–251.
- French S (1995). Uncertainty and imprecision: modelling and analysis. *Journal of the Operational Research Society* **46**: 70–79.
- French S (1992). Mathematical programming approaches to sensitivity calculations in decision analysis. *Journal of the Operational Research Society* **43**: 813–819.
- Hauser D and Tadikamalla P (1996). The Analytic Hierarchy Process on an uncertain environment: a simulation approach. *European Journal of Operational Research* **91**: 27–37.

- Hora SC (2007). Eliciting probabilities from experts. In: Edwards W, Miles RF and von Winterfeldt D (eds.) *Advances in Decision Analysis: From Foundations to Applications*. Cambridge University Press: New York.
- Insua DR and French S (1991). *European Journal of Operational Research*. **54**: 176–190.
- Jessop A (2002). Prioritisation of an IT budget within a local authority. *Journal of the Operational Research Society* **53**: 36–46.
- Keefer DL and Bodily SE (1983). Three-point approximations for continuous random variables. *Management Science* **29**: 595–609.
- Keefer DL and Verдини WA (1993). Better estimates of PERT activity time parameters. *Management Science* **39**: 1086–1091.
- Kleinmuntz DA (1990). Decomposition and the control of error in decision-analytic models. In: Hogarth RM (ed.) *Insights in Decision Making: A Tribute to Hillel J. Einhorn*. University of Chicago Press: Chicago.
- Kruskal JB and Wish M (1978). *Multidimensional Scaling*. Sage: Newbury Park CA and London.
- Levary RR and Wan K (1998). A simulation approach for handling uncertainty in the analytic hierarchy process. *European Journal of Operational Research* **106**: 116–122.
- Lipovetsky S and Tishler A (1999). Interval estimation in the AHP. *European Journal of Operational Research* **114**: 153–164.
- Megill RE (1977). *An Introduction to Risk Analysis*. Petroleum Publishing Company: Tulsa.
- Moder JJ and Rodgers EG (1968). Judgement estimates of the moments of PERT type distributions. *Management Science* **15**: B76–B83.
- Morrison DE and Henkel RE (1970). *The Significance Test Controversy: A Reader*. Aldine: Chicago.
- Moskowitz H, Tang J and Lam P (2000). Distribution of aggregate utility using stochastic elements of additive multiattribute utility models. *Decision Sciences* **31**: 327–360.
- Mustajoki J, Hämäläinen RP and Sahlo A (2005). Decision support by interval SMART/SWING – incorporating imprecision in the SMART and SWING methods. *Decision Sciences* **36**: 317–339.
- Mustajoki J, Hämäläinen RP and Lindstedt MRK (2006). Using intervals for global sensitivity and worst-case analyses in multiattribute value trees. *European Journal of Operational Research* **174**: 278–292.
- Paulson D and Zahir S (1995). Consequences of uncertainty in the analytic hierarchy process: a simulation approach. *European Journal of Operational Research* **87**: 45–56.
- Pearson ES and Tukey JW (1965). Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika* **52**: 533–546.
- Phillips LD (1984). The theory of requisite decision models. *Acta Psychologica* **56**: 29–48.

- Pöyhönen M and Hämäläinen RP (2001). On the convergence of multiattribute weighting methods. *European Journal of Operational Research* **129**: 569–585.
- Perry C and Greig ID (1975). Estimating the mean and variance of subjective distributions in PERT and decision analysis. *Management Science* **21**: 1477–1480.
- Rivett BHP (1977). Policy selection by structural mapping. *Proceeding of the Royal Society A* **354**: 407–423.
- Saaty TL (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* **15**: 234–281.
- Saaty TL and Vargas LG (1987). Uncertainty and rank order in the analytic hierarchy process. *European Journal of Operational Research* **32**: 107–117.
- Stam A and Duarte Silva AP (1997). Stochastic judgements in the AHP: the emasurement of rank reversal probabilities. *Decision Sciences* **28**: 655–688.
- Taylor BW (2007). *Intoduction to Management Science (9th edn.)*. Prentice Hall: Upper Saddle River, NJ.
- Vargas LG (1982). Reciprocal matrices with random coefficients. *Mathematical Modeling* **3**: 69–81.
- Wallsten TS (1990). The costs and benefits of vague information. In: Hogarth RM (ed.) *Insights in Decision Making: A Tribute to Hillel J. Einhorn*. University of Chicago Press: Chicago.
- Ziliak ST and McCloskey DN (2008). *The Cult of Statistical Significance*. University of Michigan Press: Ann Arbor.

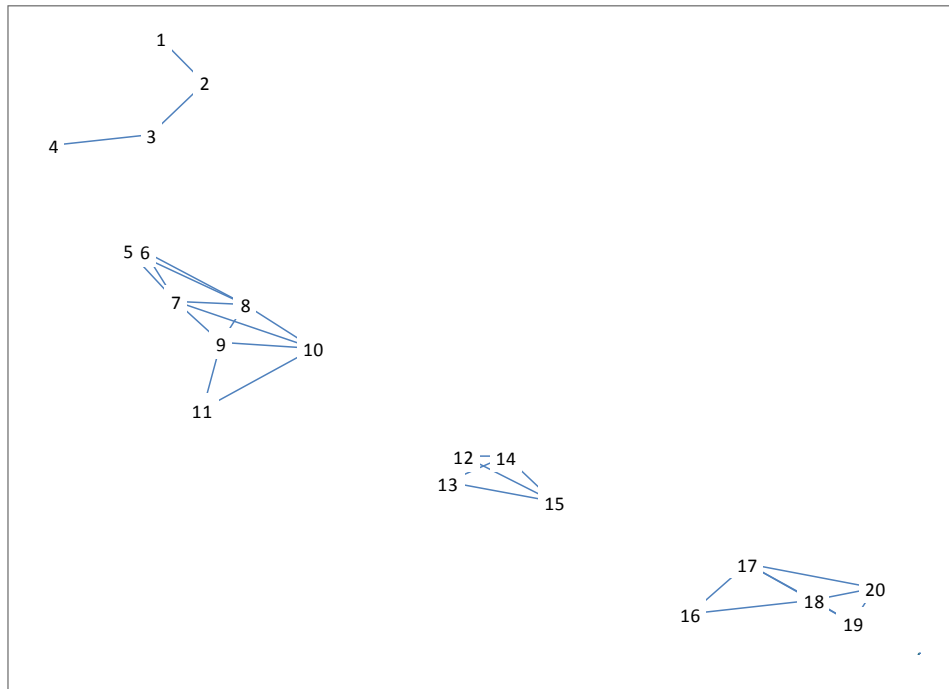


Figure 1. *Plot of z distances. Lines show insignificantly different pairs: $\alpha=0.05$, $\theta=0$.*

Attribute	user estimates			scaled values			$s = \sigma_a$	v	σ_D
	l	c	h	l	e	h			
1	70	90	95	0.144	0.175	0.196	0.016	572.4	0.020
2		100			0.206				0.021
3	70	90	100	0.144	0.179	0.206	0.019	403.4	0.020
4	10	40	50	0.021	0.069	0.103	0.025	97.9	0.013
5	40	50	60	0.083	0.103	0.124	0.013	572.7	0.016
6	5	10	15	0.010	0.021	0.031	0.006	500.2	0.008
7	10	30	40	0.021	0.055	0.083	0.019	142.0	0.012
8	60	80	90	0.124	0.158	0.186	0.019	365.7	0.019
9	5	20	25	0.010	0.034	0.052	0.013	204.1	0.010

Table 1. Case A: $\sigma_w^2 = \sigma_a^2$. Explicit uncertainty estimates. $\bar{v} = 357.3$.

		<i>type of answer</i>	
		single point	three-point
<i>degrees of freedom</i>	0	none available	σ_a^2
<i>implicit in questions</i>	>0	σ_c^2	$\sigma_c^2 + \sigma_a^2$

Table 2: *Classes of weight estimator and estimates of σ_w^2*

Attribute i								j
1	3	8	5	4	7	9	6	
3	3	7	5	9	9	9	9	
	3	9	6	7	5	9	9	
		3	5	5	9	9	5	
			3	5	5	5	5	
				4	1	5	5	
					1	5	5	
						3	5	7
							5	9

Table 3. Estimates of $a_{ij} = w_i/w_j$ (attributes ordered by importance).

Attribute	e	$s = \sigma_a$	v	σ_D
1	0.239	0.124	17.7	0.099
2	0.318	0.099	13.1	0.124
3	0.159	0.068	28.1	0.068
4	0.043	0.052	30.1	0.037
5	0.063	0.041	33.9	0.041
6	0.016	0.037	91.4	0.013
7	0.043	0.028	51.0	0.028
8	0.092	0.031	30.0	0.052
9	0.027	0.013	26.5	0.031

Table 4. Case B: $\sigma_w^2 = \sigma_c^2$. Weight estimates from Table 3. $\bar{v} = 41.6$.

	calculations as for Case A (Table 1)						effect of					
	student 1		student 2		student 3		different assessors		σ_a	σ_w	ν	σ_D
	e_1	s_1	e_2	s_2	e_3	s_3	e	σ_e				
1	0.180	0.025	0.096	0.011	0.162	0.018	0.146	0.044	0.019	0.048	53.0	0.035
2	0.246	0.028	0.160	0.014	0.146	0.017	0.184	0.054	0.020	0.058	43.9	0.038
3	0.205	0.026	0.192	0.015	0.192	0.019	0.196	0.007	0.020	0.022	335.5	0.039
4	0.029	0.011	0.075	0.010	0.069	0.012	0.058	0.025	0.011	0.027	71.6	0.023
5	0.094	0.019	0.128	0.013	0.088	0.014	0.104	0.022	0.015	0.026	133.0	0.030
6	0.004	0.004	0.053	0.008	0.058	0.011	0.038	0.030	0.008	0.031	37.7	0.019
7	0.090	0.018	0.153	0.013	0.092	0.014	0.112	0.036	0.015	0.039	64.3	0.031
8	0.139	0.022	0.100	0.011	0.108	0.015	0.116	0.021	0.017	0.027	142.2	0.032
9	0.012	0.007	0.043	0.008	0.085	0.013	0.047	0.036	0.010	0.038	30.4	0.021

Table 5. Case C: $\sigma_w^2 = \sigma_c^2 + \sigma_a^2$. Estimates from three assessors. $\bar{\nu} = 101.3$.

Attribute								
1	3	8	5	4	7	9	6	
(3,3,4)	(3,3,4)	(7,7,9)	(5,5,7)	(8,9,9)	(8,9,9)	(8,9,9)	(8,9,9)	2
	(3,3,5)	(8,9,9)	(5,6,7)	(7,7,9)	(5,5,7)	(8,9,9)	(8,9,9)	1
		(3,3,5)	(5,5,7)	(5,5,7)	(8,9,9)	(8,9,9)	(4,5,6)	3
			(3,3,4)	(5,5,7)	(5,5,7)	(4,5,6)	(5,5,7)	8
				(3,4,5)	(1,1,3)	(5,5,7)	(4,5,6)	5
					(1,1,2)	(5,5,7)	(5,5,7)	4
						(3,3,5)	(4,5,5)	7
							(5,5,6)	9

Table 6. *Paired comparisons using three point estimates.*

Attribute	e	standard deviation			v	σ_D
		σ_c	σ_a	σ_w		
1	0.236	0.099	0.031	0.104	15.8	0.076
2	0.319	0.137	0.029	0.140	10.1	0.083
3	0.155	0.068	0.016	0.070	26.1	0.065
4	0.047	0.042	0.007	0.043	23.3	0.038
5	0.064	0.042	0.012	0.044	30.6	0.044
6	0.016	0.014	0.001	0.014	77.6	0.023
7	0.040	0.029	0.007	0.029	43.7	0.035
8	0.096	0.059	0.013	0.061	22.6	0.053
9	0.027	0.033	0.002	0.033	23.0	0.029

Table 7. Case C: $\sigma_w^2 = \sigma_c^2 + \sigma_a^2$. Estimates from a single assessor. $\bar{v} = 30.3$.

Figure 1. *Plot of z distances. Lines show insignificantly different pairs: $\alpha=0.05$, $\theta=0$.*

Table 1. *Case A: $\sigma_w^2 = \sigma_a^2$. Explicit uncertainty estimates. $\bar{v} = 357.3$.*

Table 1: *Classes of weight estimator and estimates of σ_w^2*

Table 3. *Estimates of $a_{ij} = w_i/w_j$ (attributes ordered by importance).*

Table 4. *Case B: $\sigma_w^2 = \sigma_c^2$. Weight estimates from Table 3. $\bar{v} = 41.6$.*

Table 5. *Case C: $\sigma_w^2 = \sigma_c^2 + \sigma_a^2$. Estimates from three assessors. $\bar{v} = 101.3$.*

Table 6. *Paired comparisons using three point estimates.*

Table 7. *Case C: $\sigma_w^2 = \sigma_c^2 + \sigma_a^2$. Estimates from a single assessor. $\bar{v} = 30.3$.*