# On confident outrankings with multiple criteria of uncertain significance

Raymond Bisdorff

University of Luxembourg, FSTC/CSC/ILIAS

raymond.bisdorff@uni.lu

Research Note 2014–02

*Rev* : 616, September 17, 2014

**Abstract**

We develop Monte Carlo simulation techniques for taking into account uncertain criteria significance weights and ensuring an *a priori* level of confidence of the Condorcet outranking digraph, depending on the decision maker. Those outranking situations that cannot be ensured at a required level of confidence are assumed to be indeterminate. This approach allows us to associate a given confidence degree to the decision aiding artifacts computed from a bipolarly-valued outranking, which accounts for the essential and unavoidable uncertainty of numerical criteria weights.

**Keywords:** Multiple criteria decision aid; Uncertain criteria weights; Stochastic outranking relations; Confidence of the Condorcet outranking digraph.

## 1 Introduction

In a social choice problem concerning a very important issue like amending a country's Constitution, the absolute majority of voters is often not seen as sufficient for supporting a convincing social consensus. A higher majority of voters, two third or even three forth of them, may be required to support the bill in order to take effective decisions. Sometimes, even unanimity is required; a condition that, however, may generate in practice many indecisive situations. A similar idea is sometimes put forward in multiple criteria decision aiding in order to model global compromise preferences when the significance of the criterion are not known with sufficient precision. In his seminal work on the Electre I method ([1], concerning a best unique choice problematique, Roy is clearly following this line of thought by proposing to choose a sufficiently qualified majority of criterial support before considering an outranking statement to be significant.

We are here proposing a different approach. The individual criteria significance weights are considered to be random variables. The bipolarly valued characteristic of the pairwise outranking situations [2, 6] appear hence to be sums of random variables of which we may, by Monte Carlo simulation, sample the apparent distribution of possible characteristic values. From these empirical distributions, we may assess the apparent likelihood of obtaining a positive weighted majority margin for each outranking situation. And depending on the seriousness of the decision issue, we may hence recommend to accept only those outranking statements that show a sufficiently high likelihood of 90% or 95%, for instance. We could also, in the limit accept only those statements which appear to be certainly supported by a weighted majority of criterial significance.

The paper is structured as follows. A first section is concerned with how to model the uncertainty we face for assessing precise numerical criteria significance weights. The second section illustrates how the likelihood of outranking situations may be estimated. And, the last section introduces the concept of confidence level of the valued outranking digraph.

# 2 Modelling uncertain criteria significances

We have already extensively discussed some time ago (see [3]) the operational difficulty to numerically assess with sufficient precision the actual significance that underlies each criterion in a multiple criteria decision aid problem. Even, when considering that all criteria are equisignificant, it is not clear how precisely (how many decimals ?) such a numerical equality should be taken into account when computing the outranking characteristic values. In case of unequal significance of the criteria, it is possible to explore the stability of the Condorcet digraph with respect to the ordinal criteria significance structure (see [5] and [7]). One may also use indirect preferential observations for assessing via linear programming computations apparent significance ranges for each criterion (see [4]).

Here we propose instead to consider the significance weights of a family $F$ of $n$ criteria to be independent random variables $W_i$, distributing the potential significance weights of each criterion $i = 1, ..., n$ around the mean value $E(W_i)$ with variance $Var(W_i)$.

We consider four different models for taking into account the uncertainty with which we know the numerical significance weights: uniform, triangular, and two models of $beta(a, b)$ laws, one widespread and one very concentrated.

1. A continuous uniform law on the range 0 to $2 * E(W_i)$.

Thus $W_i \sim \mathcal{U}(0, 2E(W_i))$ and $Var(W_i) = \frac{1}{3}E(W_i)^2$;

2. A Beta law with parameters $a = 2$ and $b = 2$.
   Thus, $W_i \sim \mathcal{B}eta(2, 2) \times 2E(W_i)$ and $Var(W_i) = \frac{1}{5}E(W_i)^2$.

3. A triangular law on the same range with mode $E(W_i)$.
   Thus $W_i \sim \mathcal{T}r(0, 2E(W_i), E(W_i))$ with $Var(W_i) = \frac{1}{6}E(W_i)^2$;

4. A narrower Beta law with parameters $a = 4$ and $b = 4$.
   Thus $W_i \sim \mathcal{B}eta(4, 4) \times 2E(W_i)$, $Var(W_i) = \frac{1}{9}E(W_i)^2$

It is worthwhile noticing that the four uncertainty models models admit the same expected value, $E(W_i)$, however, with a standard deviation which goes decreasing as shown in Fig. 1.

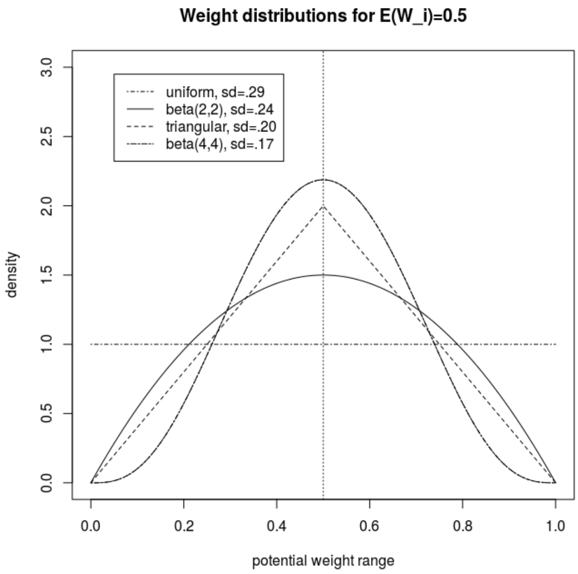


Figure 1: Four models of uncertain significance weights

# 3 Likelihood of an outranking situation

Now, $r(x \succcurlyeq y) = \sum_i W_i \times r(x \succcurlyeq_i y)$ with $r(x \succcurlyeq_i y) \in \{-1, 0, 1\}$ becomes a simple sum of positive or negative independent random variables with known means and variances. We know from the Central Limit Theorem (CLT) that such a sum of random variables tends, with $n$ getting large, to a Gaussian distribution $Y$ with $E(Y) = \sum_i E(W_i) \times r(x \succcurlyeq y)$ and $Var(Y) = \sum_i Var(W_i) \times |r(x \succcurlyeq y)|$.

**Example 3.1.** Let us consider two decision alternatives $x$ and $y$ being evaluated on a family of 7 equi-significant criteria, such that four out of the seven criteria positively support that $x$ outranks $y$, and three criteria support that $x$ does not outrank $y$. In this case, $r(x \succcurlyeq y) = 4w - 3w = w$ where $W_i = w$ for $i = 1, ..., 7$ and the outranking situation is positively validated. Suppose now that the significance weights $W_i$ appear only more or less equivalent and let us model this numerical uncertainty with independent triangular laws: $W_i \sim \mathcal{T}r(0, 2w, w)$ for $i = 1, ...7$. The expected credibility of the outranking situation, $E(r(x \succcurlyeq y)) = 4w - 3w = w$, will remain the same, however with a variance of $7 \times \frac{3}{18}w^2$. If we take a unit weight $w = 1$, we hence obtain a standard deviation of 1.08. Applying the CLT we notice that, under the given hypotheses, the likelihood of obtaining a negative majority margin will be about 17%. A Monte Carlo simulation with 10000 runs empirically confirms the effective convergence to a Gaussian: $r(x \succcurlyeq y) \rightsquigarrow \mathcal{N}(1.03, 1, 089)$ (see Figure 2), with an empirical proportion of negative majority margins $r(x \succcurlyeq y) \leq 0.0$ of indeed about 17% .

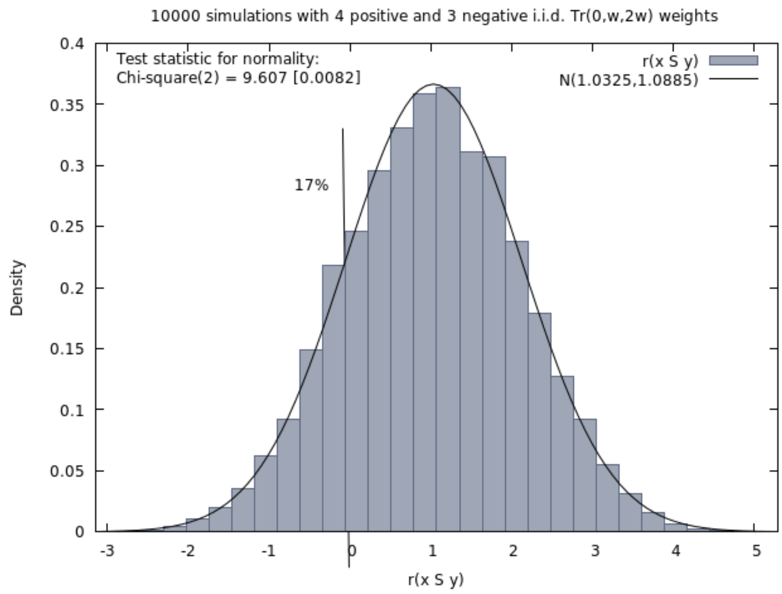


Figure 2: Distribution of outranking credibility $r(x \succcurlyeq y)$

**Example 3.2.** The second example concerns two decision alternatives $x$ and $y$ that are evaluated on a family of 7 criteria, denoted $g_i$ of unequal significance weights $w_i$ for $i = 1, ..., 7$ (see Tab. 1). The performances on the seven criteria are measured on a rational scale from 0 (worst) to 100 points (best). Let us suppose that both decision alternatives are evaluated as shown in Tab. 1. A performance difference of 10 points or less is considered insignificant, whereas a difference of 20 points and more is considered to be significant. The overall bipolar outranking

Table 1: Pairwise comparison of two decision alternatives

| $g_i$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_6$ | $g_7$ |
|---|---|---|---|---|---|---|---|
| $w_i$ | 7 | 8 | 3 | 10 | 1 | 9 | 7 |
| $x$ | 14.09 | 71.41 | 87.92 | 38.73 | 26.50 | 93.00 | 37.15 |
| $y$ | 64.03 | 87.51 | 67.04 | 82.24 | 80.84 | 80.82 | 10.64 |
| $x - y$ | -49.94 | -16.10 | +20.88 | -43.51 | -54.34 | +12.18 | 26.51 |
| $r(x \succcurlyeq_i y)$ | $-1$ | 0 | $+1$ | $-1$ | $-1$ | $+1$ | $+1$ |

credibility $r(x \succcurlyeq y)$ (see [6]) is given as follows:

$$r(x \succcurlyeq y) \;=\; \sum_{i=1}^{7} r(x \succcurlyeq_i y) \times w_i \;=\; -7 + 0 + 3 - 10 - 1 + 9 + 7 \;=\; +1\;;. \quad (1)$$

The outranking situation "$(x \succcurlyeq y)$" is thus positively validated (see Eq. 1). However, in case the given criteria significance weights (see Tab. 1) are not known with certainty, how confident can we be about the actual positiveness of $r(x \succcurlyeq y)$? If we suppose now that the random significance weights $W_i$ are in fact independently following a continuous triangular law on the respective ranges 0 to $2w_i$, the CLT approximation will make $r(x \succcurlyeq y)$ tend to a Gaussian distribution with mean equal to +1 and standard deviation equal to 6.94. The likelihood of $r(x \succcurlyeq y) \geqslant 0.0$ equals thus approximately $1.0 - P(\frac{z-1}{6.94} \leqslant 0.0) \;=\; 1.0 - 0.43 \;\approx\; 55.7\%$, a result we can again empirically verify with a Monte Carlo sampling of 10000 runs (see Fig. 3). Under the given modelling of the uncertainty in the setting of the criteria significance weights, the credibility of the outranking situation between alternatives $x$ and $y$ is neither convincingly positive, nor negative. Neither the relational situation may hence be validated, nor invalidated.

By requiring a certain level of likelihood for all pairwise outranking situations' credibility, we may thus enforce the actual confidence we may have in the bipolarly-valued outranking digraph. If, for instance, we would require that an outranking statement, to be validated, must admit a positive credibility with a high likelihood, 90% or more for instance, the first and, even more, the second outranking situation discussed previously, will not be validated, nor invalidated with enough confidence. We will set their characteristic values hence equal to the indeterminate value 0.0.

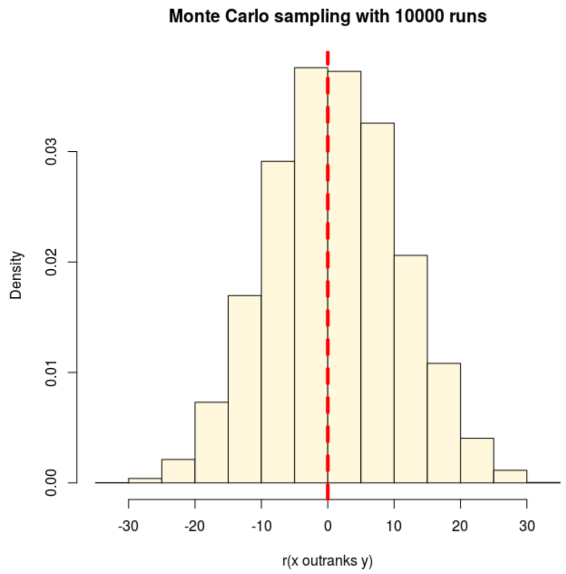


Figure 3: Distribution of outranking credibility $r(x \succcurlyeq y)$

# 4 Confidence level of the outranking digraph

Most of the MCDA decision aiding problematiques like best choice, ranking or sorting recommendations exploiting a valued outranking digraph rely in fact on the majority cut of the valued outranking digraph [2, 3, 8, 9, 6, 7]. The previous example gives the hint how we may appreciate the very confidence we may have in a given majority when the criteria significance weights are not precisely given. Knowing a priori the distribution of the significance weight of each criterion will indeed be sufficient in practice for computing, with the so given means and variances, the CLT based likelihood of the fact that a bipolar outranking characteristics $r(x \succcurlyeq y)$ is positively validating, respectively negatively invalidating, the outranking situation "$(s \succcurlyeq y)$".

**Example 4.1.** We may illustrate this approach with a small random performance tableau (see Tab. 2) showing the performance evaluations of seven decision alternatives on the same family of performance criteria we used for Example 3.2. To operate with a full outranking model, we furthermore consider that a large performance difference of 80 points and more will trigger a veto situation (see [6]). [ht] When using deterministic criteria significance weights, we obtain the following bipolarly valued outranking relation (see Tab. 3): We recover the weakly positive credibility $(r(a01 \succcurlyeq a02) = +1/45)$ of the outranking situation between alternative 'a01' and alternative 'a02' discussed in Example 3.2. Notice by the way

Table 2: Random performance tableau

| $g_i$ | $w_i$ | a01 | a02 | a03 | a04 | a05 | a06 | a07 |
|---|---|---|---|---|---|---|---|---|
| $g_1$ | 7.00 | 14.09 | 64.03 | 73.43 | 36.47 | 30.61 | 85.90 | 97.83 |
| $g_2$ | 8.00 | 71.41 | 87.51 | 61.90 | 84.67 | 60.41 | 54.46 | 45.77 |
| $g_3$ | 3.00 | 87.92 | 67.04 | 25.17 | 34.23 | 87.30 | 43.05 | 30.35 |
| $g_4$ | 10.00 | 38.73 | 82.24 | 94.06 | 86.05 | 34.05 | 97.23 | 72.21 |
| $g_5$ | 1.00 | 26.50 | 80.84 | 71.94 | 21.63 | 56.38 | 88.13 | 15.03 |
| $g_6$ | 9.00 | 93.00 | 80.82 | 23.21 | 57.24 | 81.39 | 16.57 | 93.03 |
| $g_7$ | 7.00 | 37.15 | 10.64 | 64.79 | 98.94 | 69.95 | 24.66 | 13.57 |

Table 3: Credibility of the deterministic outranking relation (range = $[-45, +45]$)

| $r(x \succcurlyeq y)$ | a01 | a02 | a03 | a04 | a05 | a06 | a07 |
|---|---|---|---|---|---|---|---|
| a01 | - | $\mathbf{+1}$ | $\mathbf{-5}$ | $-11$ | $+22$ | $+9$ | 0 |
| a02 | $+16$ | - | $+21$ | 0 | $+25$ | $+14$ | $+22$ |
| a03 | $+21$ | $+5$ | - | $-3$ | $+21$ | $+34$ | $+13$ |
| a04 | $+21$ | $+45$ | $+29$ | - | $+19$ | $+19$ | $+45$ |
| a05 | $+28$ | $-7$ | $+10$ | $-5$ | - | $+9$ | $+2$ |
| a06 | $+6$ | $+5$ | $+31$ | $-3$ | $+7$ | - | $+20$ |
| a07 | $+45$ | $+11$ | $+1$ | $+0$ | $+15$ | $+13$ | - |

the slightly negative credibility ($-5/45$) of the outranking situation between alternative ‘a01’ and ‘a03’ . How confident are these deterministic statements, if the significance weights are not precisely given. If we assume now that the potential criteria significances $w_i$ are distributed following independent triangular laws $\mathcal{T}(0, 2w_i, w_i)$ for $i = 1, ..., 7$, we obtain the following CLT likelihoods (see Tab. 4): If we, furthermore, require for each credibility degree ($r(x \succcurlyeq y)$ a likelihood of 0.90 and more for convincingly validating, respectively invalidating, the corresponding outranking statement, we obtain the following result (see Tab. 5): We notice here that the outranking situations between ‘a01’ and ‘a02’, respectively ‘a03’, with credibility likelihoods lower than 90%, are all put to doubt. In total 15 pairwise outranking statements out of the potential $7 \times 6$ statements are thus considered not confident enough. Their credibility $r(x \succcurlyeq y)$ is put to the indeterminate value 0. It is worthwhile noticing that all outranking statements with a credibility less than $\pm 7/45$ (a qualified majority of 57.8%) are thus put to doubt. However, the outranking situation between ‘a01’ and ‘a06’, obtaining a deterministic credibility

Table 4: CLT likelihood of $r(x \succcurlyeq y) > 0$ or $< 0$

| p-value | a01 | a02 | a03 | a04 | a05 | a06 | a07 |
|---|---|---|---|---|---|---|---|
| a01 | - | +0.56 | +0.74 | +0.94 | +1.00 | +0.88 | +0.50 |
| a02 | +0.99 | - | +1.00 | +0.50 | +1.00 | +0.99 | +1.00 |
| a03 | +1.00 | +0.74 | - | +0.65 | +1.00 | +1.00 | +0.95 |
| a04 | +1.00 | +1.00 | +1.00 | - | +0.99 | +1.00 | +1.00 |
| a05 | +1.00 | +0.82 | +0.90 | +0.74 | - | +0.88 | +0.62 |
| a06 | +0.83 | +0.74 | +1.00 | +0.65 | +0.82 | - | +1.00 |
| a07 | +1.00 | +0.95 | +0.56 | +0.50 | +0.98 | +0.97 | - |

Table 5: 90% confident outranking relation

| $r(x \succcurlyeq y)$ | a01 | a02 | a03 | a04 | a05 | a06 | a07 |
|---|---|---|---|---|---|---|---|
| a01 | - | 0 | 0 | −11 | +22 | 0 | 0 |
| a02 | +16 | - | +21 | 0 | +25 | +14 | +22 |
| a03 | +21 | 0 | - | 0 | +21 | +34 | +13 |
| a04 | +21 | 0 | +29 | - | +19 | +19 | +45 |
| a05 | +28 | 0 | 0 | 0 | - | 0 | 0 |
| a06 | 0 | 0 | +31 | 0 | +7 | - | +20 |
| a07 | 0 | +11 | 0 | 0 | +15 | +13 | - |

of +9 (a majority of 60%), but with only a likelihood of 88%, appears as well not confident enough.

The quality of the CLT convergence will, in general, depend, first, on the number of effective criteria, i.e. non abstaining ones, involved in each pairwise comparison and, secondly, on the more or less differences in shape of the individual significance weight distributions. Therefore, with a tiny performance tableau, less than 25 decision actions and less than 10 criteria, we may estimate the likelihood of all pairwise outranking situations with a Monte Carlo simulation consisting of a given number of independent runs. Indeed, the present computational power available, even on modest personal computers, allow us to sufficiently sample a given outranking digraph construction.

**Example 4.2.** For instance, if we sample 1000 MC simulations of the previous outranking relation (see Tab. 3), by keeping the same uncertainty modelling of the criteria significances with random weights distributed like $\mathcal{T}(0, 2w_i, w_i)$, we obtain very similar empirical likelihoods (see Tab. 6). We may thus verify again the very

Table 6: Empirical likelihoods of $r(x \succcurlyeq y) > 0$ or $< 0$ with a sample of 1000 runs

| p-value | a01 | a02 | a03 | a04 | a05 | a06 | a07 |
|---|---|---|---|---|---|---|---|
| a01 | - | +0.56 | +0.73 | +0.95 | +1.00 | +0.87 | +0.5 |
| a02 | +0.99 | - | +1.00 | +0.5 | +1.00 | +0.99 | +1.00 |
| a03 | +1.00 | +0.73 | - | +0.64 | +1.00 | +1.00 | +0.95 |
| a04 | +1.00 | +0.72 | +1.00 | - | +0.99 | +1.00 | +1.00 |
| a05 | +1.00 | +0.81 | +0.90 | +0.74 | - | +0.88 | +0.62 |
| a06 | +0.83 | +0.73 | +1.00 | +0.65 | +0.82 | - | +1.00 |
| a07 | +1.00 | +0.95 | +0.57 | +0.50 | +0.99 | +0.97 | - |

accurate convergence (in the order of $\pm 1\%$) of the CLT likelihoods, a convergence we already observed in Example 3.2, even with a small number of criteria.

# 5 Conclusion

When modelling preferences following the outranking approach, the sign of the majority margins do sharply distribute validation and invalidation of pairwise outranking situations. How can we be confident in the resulting outranking digraph, when we acknowledge the usual imprecise knowledge of criteria significance weights and a small majority margin? To answer this question, we propose to model the significance weights as random variables following more less widespread distributions around an average weight value that corresponds to the given deterministic weight. As the bipolarly valued random credibility of an outranking statement results from a simple sum of positive or negative independent and similarly distributed random variables, we may apply the CLT for computing likelihoods that a given majority margin is indeed positive, respectively negative. To test the effective convergence of the CLT likelihoods, we apply Monte Carlo simulations of outranking digraph constructions. Our computational results confirm a satisfactory convergence even for a random performance tableau with only seven criteria.

# References

[1] Roy B (1985) Méthodologie Multicritère d'Aide à la Décision. Economica Paris 1

[2] Bisdorff, R. (2000). Logical foundation of fuzzy preferential systems with application to the electre decision aid methods. Computers and Operations Research 27:673-687 2, 6

[3] Bisdorff, R. (2002). Logical Foundation of Multicriteria Preference Aggregation. In: Bouyssou D et al (eds) Essay in Aiding Decisions with Multiple Criteria. Kluwer Academic Publishers 379-403 2, 6

[4] Dias, L. C. (2002). Exploring the Consequences of Imprecise Information in Choice Problems Using ELECTRE. In Bouyssou, D. et al. (eds) Aiding Decisions with Multiple Criteria (Springer ORMS) 44: 175-193 2

[5] Bisdorff, R., P. Meyer and Th. Veneziano (2009). Inverse analysis from a Condorcet robustness denotation of valued outranking relations. In F. Rossi and A. Tsoukiás (Eds.), Algorithmic Decision Theory. Springer-Verlag Berlin Heidelberg, LNAI 5783, pp. 180-191 2

[6] Bisdorff, R. (2013) On polarizing outranking relations with large performance differences. Journal of Multi-Criteria Decision Analysis 2, 5, 6

[7] Bisdorff, R., P. Meyer and Th. Veneziano (2014). Elicitation of criteria weights maximising the stability of pairwise outranking statements. Journal of Multi-Criteria Decision Analysis (Wiley) 21: 113-124 2, 6

[8] Bisdorff, R., M. Pirlot and M. Roubens (2006). Choices and kernels from bipolar valued digraphs. European Journal of Operational Research, 175: 155-170. 6

[9] Bisdorff, R., P. Meyer and M. Roubens (2008). RUBIS: a bipolar-valued outranking method for the choice problem. 4OR, A Quarterly Journal of Operations Research, Springer-Verlag, 6 (2): 143-165 6