EXPERT AGGREGATION WITH DEPENDENCE

M. J. Kallen[1], R.M. Cooke[2]

[1]Department of Mathematics, Delft University of Technology, Delft, The Netherlands

[2]Department of Mathematics, Delft University of Technology, Delft, The Netherlands

Appeared in *Probabilistic Safety Assessment and Management* E.J. Bonano, A.L. Camp, M.J. Majors, R.A. Thompson (eds),Elsevier, 2002; 1287-1294.

ABSTRACT

The measure for expert dependence proposed by Jouini and Clemen (clemen) is implemented for expert judgement data gathered at the T.U. Delft. Experts show less dependence than might have been supposed, though more sensitive measures might reveal more. Clemen's copula for aggregation is implemented and performance is compared with performance-based combinations for two illustrative cases.

KEYWORDS

Expert judgement, dependence, copula, Frank's copula

INTRODUCTION

To date, there have been 15 clusters of expert judgment studies performed by the TU Delft. These include 28 different expert panels, which gave assessments over a variety of topics. In all cases experts assessed items for which the true values were known, in addition to variables of interest. A brief listing of these studies given in table 1 extends the work in [Goossens, 1998]. In this paper we study dependence in the manner suggested by Jouini and Clemen [Clemen, 1996]. For two illustrative cases we implement the copula method for combining expert assessments with dependence and compare performance of the copula based combination with other combinations. The purpose of the p-value will explained in a later section.

To illustrate the calculations in the following sections, we will use the EUNRCDIS case as an example. This is the EU-USNRC dispersion module performed by the TU Delft and SANDIA. The publication of this study can be found in [Harper, 1995].

Case	Name	#experts	#vbls/#seed	P-value
1a	DSM-1	10	14/8	0.492
1b	DSM-2	8	39/12	0.59
2a	ESTEC-1	4	48/13	0.786
2b	ESTEC-2	7	58/26	≈ 0
2c	ESTEC-3	6	22/12	0.454
3	AOT	9	38/38	≈ 0
4	GROUND	7	38/10	0.132
5a	TUDDISP	11	58/36	≈ 0
5b	TNODISP	7	58/36	≈ 0
5c	TUDDEPOS	4	56/24	0.034
6a	ACNEXPTS	7	43/10	0.324
6b	NH3EXPTS	6	31/10	0.456
6c	SO3EXPTS	4	28/7	0.614
7	WATERPOL	11	21/11	0.124
8a	EUNRCDIS	8	101/23	≈ 0
8b	EUNRCDD	8	70/14	0.034
8c	EUNRCA_S	7	8/8	0.172
8d	EUNRCWD	7	50/19	≈ 0
8e	EUNRCEAR	9	15/15	0.014
8f	EUNRCSOI	4	31/31	≈ 0
9a	GAS95 (env. panel)	15	48/28	n/a
9b	GAS95 (corr. panel)	12	58/11	n/a
10	MVBLBARR	8	52/14	0.176
11	REALESTR	5	45/31	0.692
12	RIVRCHNL	6	14/8	0.504
13	MONT1	11	13/8	0.516
14	THRMBLD	6	48/48	≈ 0
15	DIKRING	17	87/47	≈ 0

TABLE 1: List of completed TUDelft expert judgment studies.

DEPENDENCE BETWEEN EXPERTS

Following [Clemen, 1996] we can consider only the expert's median assessments and consider the events that these medians are above or below the true values. Each expert receives a 1 if his assessment is above the true value and a -1 otherwise. We therefore only use the assessments for which the realizations are known (sometimes called seed variables), and compute the product moment correlation between the 1's and -1's. In some cases it may happen that all the median assessments of one expert fall on one side of the true value, in which case the expert has no variance. The correlation matrix for the EUNRCDIS case is given in table 2. All the experts are positively dependent with the average correlation being about 0.55. We therefore have a case with the desired features for the Clemen model which we will apply later on.

	E1	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**
E1	1	0.4792	0.6283	0.4792	0.5231	0.6937	0.3144	0.3874
E2	0.4792	1	0.3973	0.4773	0.5649	0.6908	0.5505	0.5800
E3	0.6283	0.3973	1	0.7628	0.6485	0.7073	0.2791	0.2333
E4	0.4792	0.4773	0.7628	1	0.7405	0.6908	0.5505	0.3973
E5	0.5231	0.5649	0.6485	0.7405	1	0.7542	0.6009	0.6485
E6	0.6937	0.6908	0.7073	0.6908	0.7542	1	0.5677	0.5089
E7	0.3144	0.5505	0.2791	0.5505	0.6009	0.5677	1	0.5004
E8	0.3874	0.5800	0.2333	0.3973	0.6485	0.5089	0.5004	1

TABLE 2: Correlation matrix for the experts of the EUNRCDIS case.

STATISTICAL SIGNIFICANCE

In table 3 are all the 1's and -1's for each item i and each expert j in the EUNRCDIS case. If we consider these assessments as realizations of random variables X_j, then the probability of expert j's assessment being a 1 is given by $P(X_j = 1)$. The values of these probabilities are on the bottom line of table 3. Obviously $P(X_j = -1) = 1 - P(X_j = 1)$.

The q_i in the right column is the weight for each 'word', which we define as the combination of 1's and -1's for each item. For item 1 ($i = 1$) this word is given by the sequence $w_{1,\cdot} = \{-1, 1, -1, 1, 1, -1, 1, 1, 1\}$ and $q_1 = 1$, which means that there is only one such sequence for the items. We can normalize q_i by dividing this number with the total amount of items, so that we obtain the probability of a certain word appearing in the assessments. If we call this vector of probabilities Q, then

$$Q_i = P(w_{i,\cdot}) \approx \frac{q_i}{n},$$

where n is the number of items. As can be seen in table 3, if a particular word appears more than once, only one word receives a weight whereas the other identical words receive zero weight. Since we can estimate $P(X_j = 1)$ and $P(X_j = -1)$ for each expert, we can also estimate the probability of each $w_{i,\cdot}$ being assessed by the experts under the hypothesis of independence. Take W as the vector of these probabilities and assume the expert assessments are independent, then

$$W_i = P(X_1 = w_{i,1}, \ldots, X_m = w_{i,m}) = P(X_1 = w_{i,1}) \cdots P(X_m = w_{i,m})$$

for m experts. W is the theoretical distribution (i.e. that the experts are independent) and Q is the distribution of the data.

	$j=1$	2	3	4	5	6	7	8	q_i
$i=1$	-1	1	-1	1	1	-1	1	1	1
2	-1	1	-1	-1	-1	-1	-1	1	1
3	-1	1	-1	1	1	1	1	1	1
4	-1	1	-1	-1	-1	-1	-1	-1	2
5	-1	-1	-1	-1	-1	-1	-1	-1	8
6	-1	-1	-1	-1	-1	-1	-1	-1	0
7	1	1	1	1	1	1	1	1	2
8	-1	-1	-1	-1	1	-1	-1	1	1
9	1	1	1	1	1	1	-1	1	1
10	1	1	1	1	1	1	1	1	0
11	-1	1	1	1	1	1	-1	1	1
12	1	1	1	1	1	1	-1	-1	1
13	-1	1	1	1	1	1	1	-1	1
14	-1	-1	1	1	1	-1	-1	-1	1
15	-1	-1	-1	-1	-1	-1	-1	-1	0
16	-1	-1	-1	-1	-1	-1	-1	-1	0
17	-1	-1	1	1	-1	-1	-1	-1	1
18	-1	-1	-1	-1	-1	-1	-1	-1	0
19	-1	-1	-1	-1	-1	-1	-1	-1	0
20	-1	-1	-1	1	-1	-1	-1	-1	1
21	-1	-1	-1	-1	-1	-1	-1	-1	0
22	-1	-1	-1	-1	-1	-1	-1	-1	0
23	-1	1	-1	-1	-1	-1	-1	-1	0
$P(X_j = 1)$	0.174	0.478	0.348	0.478	0.435	0.304	0.217	0.348	

TABLE 3: Assessments for item i by each expert j.

The relative information between the two probability vectors Q and W is given by

$$I(Q, W) = \sum_{i=1}^{n} Q_i \ln\left(\frac{Q_i}{W_i}\right).$$

This measure is always non-negative and equals zero if and only if $Q = W$. For the EUNRCDIS case, $I(Q,W) = 2.67$.

We test the hypothesis that 'the experts are independent' against 'the experts are not independent'. We do this by simulating the distribution of $I(Q,W)$ under the assumption that the events $X_j < 1$ are independent. We reject this assumption if this actual value $I(Q < W)$ is in the top 5% of all the simulated values. We have performed 500 simulations like this for the EUNRCDIS data and all the simulated values for the relative information were smaller than the original value, therefore $P(I(Q,W) \leq 2.67) = 1$ and thus the p-value is approximately zero. We have also done this for all the other cases and the p-value for each case is given in table 1. The assumption of independence is rejected at the 5% level in about one half of the cases, which may be a weaker indication of dependence than many would have expected. Inspecting the table 1 also reveals that there seems to be a negative correlation between the number of seed variables and the p-value. The low p-values coincide with high numbers of seed variables.

CLEMEN'S COPULA METHOD

We implement the expert opinion aggregation model as suggested by [Clemen, 1996]. A Bayesian decision maker who is interested in modelling dependence among the experts by combining their assessments in a copula representing the amount of dependence (i.e. correlation). [Clemen, 1996] recommends Frank's copula. With this copula and the univariate distributions of the experts' assessments, the decision maker can construct a likelihood function for these opinions. If we take H_i and h_i as expert i's cumulative distribution function and density respectively, then the likelihood function is given by

$$f_m = c_m[1 - H_1(\theta), \ldots, 1 - H_m(\theta)]h_1(\theta) \cdots h_m(\theta), \tag{1}$$

where m is the number of experts and

$$c_m(u_1, \ldots, u_m) = \frac{\partial^m}{\partial u_1 \cdots \partial u_m} C_m(u_1, \ldots, u_m). \tag{2}$$

Here C_m is the copula and with Frank's copula:

$$C_m(u_1, \ldots, u_m) = \log_\alpha \left[1 + \frac{(\alpha^{u_1} - 1) \cdots (\alpha^{u_m} - 1)}{(\alpha - 1)^{n-1}}\right], \quad 0 < \alpha < 1.$$

The dependence is captured through α, whose value can be obtained from table I in the article. For this we need Kendall's τ [Kendall, 1961], which is a measure of dependence. In the case of our binomial random variable X_j this τ is equal to the correlation between the experts' assessments.

To implement Clemen's model, we solve for the likelihood function (1) numerically. It is easy to check that for m experts, (2) becomes

$$c_m(u_1, \ldots, u_m) = \ln(\alpha)^{m-1} \sum_{j=1}^{m} \beta_j \frac{\alpha^{u_1} \cdots \alpha^{u_m} (\alpha^{u_1} - 1)^{j-1} \cdots (\alpha^{u_m} - 1)^{j-1}}{[(\alpha - 1)^{m-1} + (\alpha^{u_1} - 1) \cdots (\alpha^{u_m} - 1)]^j}, \tag{3}$$

where β is a vector of coëfficiënts which depends on m. This vector must be determined numerically.

From the average correlation of 0.55, we find that $\alpha = 0.0011957$. The results for item 10 and item 21 are plotted in figure 1 and they look reasonable. Especially for item 21, where the experts'

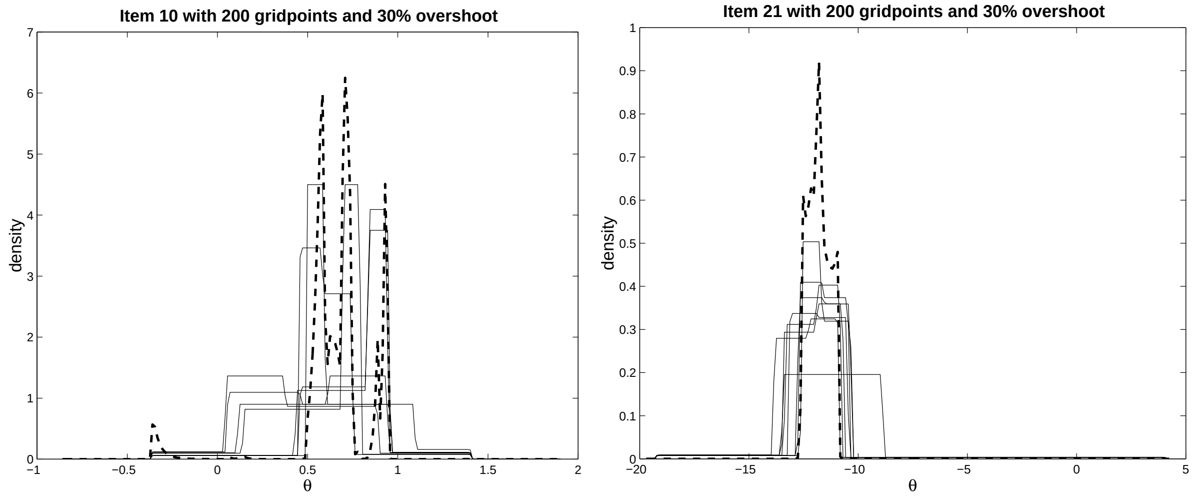

Figure 1: Posterior density (dashed line) for item 10 (left) and 21 (right) from the EUNRCDIS data, using Clemen's method based on the experts' densities (full lines).

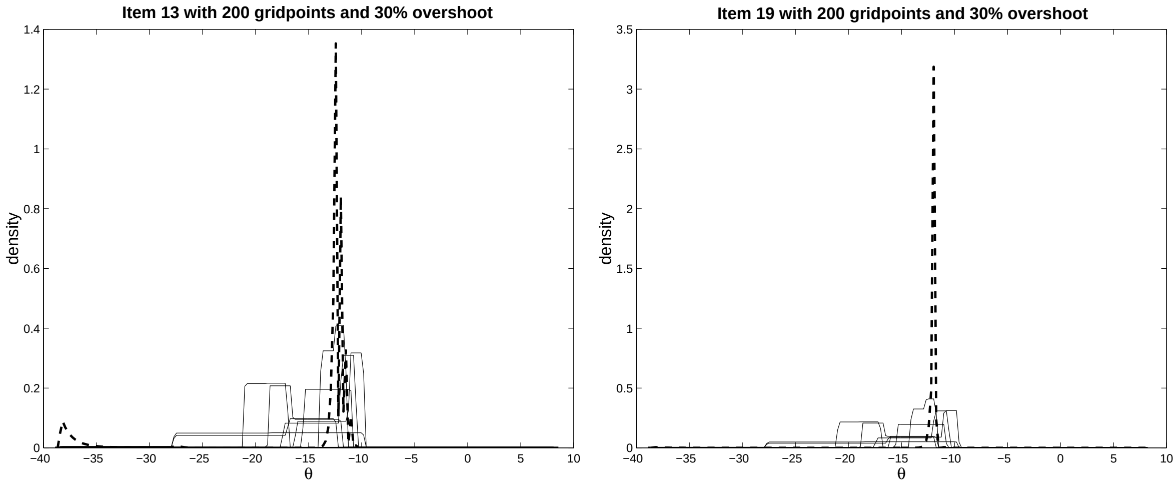

Figure 2: Same as figure 1 for item 13 and 19 of the EUNRCDIS data.

assessments are very close together, the posterior density follows the data very well. In figure 2 we have plotted items 13 and 19. The assessment data for these two items is almost identical. The model also produces similar results, although the mass is slightly more concentrated for item 19.

As Frank's copula can be highly peaked at the extremes, numerical stability is an issue, especially when α is small. The likelihood function given by equation (1) behaves very much like the geometric mean: very sensitive to the lower values and insensitive to the higher values. Figure 3 shows the same results for item 10 and 13 as in figure 1 and 2, but this time on a logarithmic scale. This shows how incredibly small the values of the result actually are. If the PDF from one expert is zero at a certain value, then the result will be forced to zero. This problem is also addressed in [Clemen, 1996], where the decision maker is advised to be very careful with the tails of the assessments.

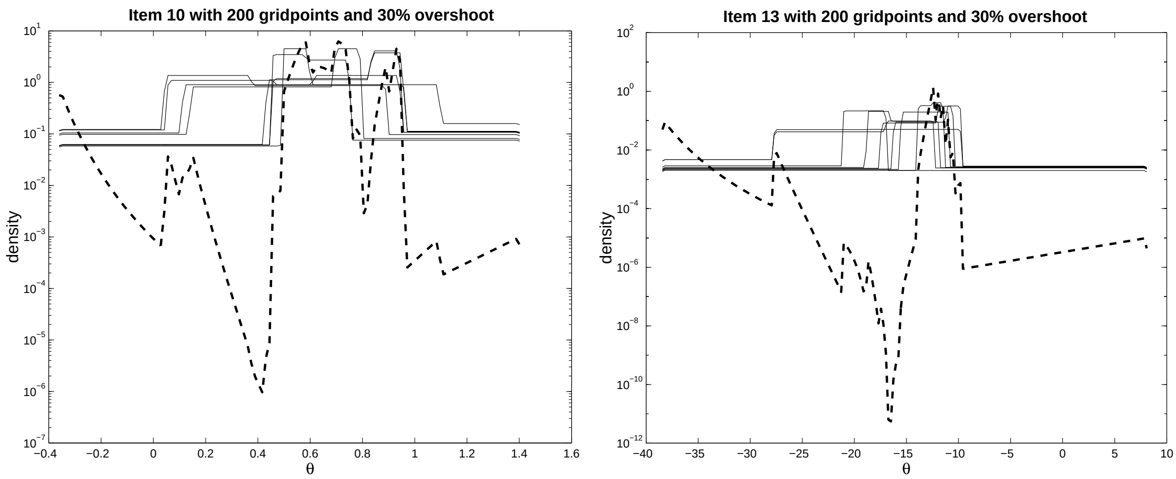

Figure 3: Results for item 10 (left) and item 13 (right) for the EUNRCDIS data plotted on a logarithmic scale.

PERFORMANCE OF THE COPULA METHOD

Performance is measured by calibration and information [Cooke, 1991]. The results for two illustrative cases shown below were obtained using a 30% overshoot on the intrinsic range.

Table 4 shows the calibation and information scores for the global weight combination, the item weight combination, the equal weight combination, the best expert and the Clemen copula combination. The Clemen combination shares the lowest calibration with two other experts, but has a slightly higher information score.

	Combined DM's:			Experts:	
	global	equal	item	best	Clemen
Calibration	0.36	0.15	0.9	0.13	0.0001
Information	1.24	0.894	1.116	1.276	2.534
Combination	0.4443	0.1341	1.005	0.1659	2.534×10^{-3}

TABLE 4: Expert Calibration results for the EUNRCDIS case.

As a second example we use the data from the REALESTR case [Qing, 2002]. The case involves 5 investment managers who each give assessments on the rent indices of office space for the major cities in the Netherlands in the future. The case involves 45 items and there are 16 seed variables. A year after this study, the new prime rent indices became available and the experts' assessments where checked again by adding these values to the original 16 seed variables. For this example we will only use the 16 original seed variables.

The average correlation between the experts' assessments is 0.32, which gives an $\alpha = 0.054042$. The results are given in table 5. The global and item weight decision makers are the same as the best expert. The Clemen expert is the third best expert out of the six in total, which is better than the previous case.

	Combined DM's:			Experts:	
	global	equal	item	best	Clemen
Calibration	0.33	0.12	0.33	0.33	0.02
Information	0.8572	0.2068	0.8572	0.8572	1.384
Combination	0.2829	0.0248	0.2829	0.2829	0.0277

TABLE 5: Expert Calibration results for the REALESTR case.

CONCLUSIONS

We reject the hypothesis of dependence between the experts in about half of the cases. We also have found that there is no significant relation between dependence and calibration. We do see an apparent relation between dependence and the number of seed variables. This relationship should be studied further with more sensitive dependence measures.

The copula method for aggregating expert opinions does not yield good results for the EUNR-CDIS case. The copula method in [Clemen, 1996] also requires the experts to be pairwise equally dependent in the sense of their correlation. In most practical cases the decision maker does not know anything about the variable(s) being assessed and about the level of dependence between the experts. We have taken the average correlation between the experts' assessment as the measure for pairwise dependence between the experts. Better performance can be expected from the multi-variate normal copula, as suggested in [Clemen and Reilly, 1999]. In the two cases presented here the calibration and information performance of Clemen's model is mixed. It is worthwhile trying other copula.

References

[Bedford, 2001] Bedford, T., Cooke, R.M., *Mathematical tools for probabilistic risk analysis.* Cambridge University Press, U.K., 2001.

[Clemen, 1996] Jouini, Mohamed N. and Clemen, Robert T., *Copula models for aggregating expert opinions* in Operations Research, Vol.44, No.3, 444-457, 1996.

[Clemen and Reilly, 1999] Clemen, R.T. and Reilly, T. *Correlations and copulas for decision and risk analysis* in Management Science, vol. 45, no. 2, 208-223.

[Cooke, 1991] Cooke, R.M., *Experts in Uncertainty: opinion and subjective probability in science.* Oxford University Press, Oxford, 1991.

[Goossens, 1998] Goossens L.H.J. et al, *Evaluation of weighting schemes for expert judgment studies* in Mosleh, A. and Bari, R. (eds). Probabilistic Safety Assessment and Management, Vol.3, 2113-2118. Springer, New York 1998.

[Harper, 1995] Harper, F.T. et al, *A joint USNRC/CEC consequence uncertainty study: summary of objectives, approach, application, and results for the dispersion and deposition uncertainty assessment.* Prepared for the U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6244, EUR 15855, Washington, U.S.A. and Brussels-Luxembourg, 1995. (Volumes I,II,III)

[Kendall, 1961] Kendall, M.G., Stuart, A., *The advanced theory of statistics: inference and relationship*, Vol. 2. Charles Griffin & Co. Ltd., London, 1961.

[Qing, 2002] Qing, X., *Risk analysis for real estate investment*, PhD thesis, Dept. of Architecture, TU Delft, 2002.