INFORMATION CONCEPTS AND AHP

by

Pedro P. Sanchez

Refik Soyer

The George Washington University

ABSTRACT

In this paper we present preliminary research on how information (or entropy) based measures can be used by a decision maker (DM) using the Analytic Hierarchy Process (AHP) to assess judgment accuracy, and to decide when to stop the process of pairwise comparisons. We introduce some information indices, and illustrate their use in the Analytic Hierarchy Process by means of some examples. Furthermore, we research the issue of how to measure the information content of redundant judgments, and investigate the relationship between information and degree of redundancy through the use of simulated judgments.

INTRODUCTION AND OVERVIEW

In a recent paper, Noble and Sanchez (1993) used the concept of entropy to assess the information content of pairwise comparison judgment matrices (PCJM's) in the Analytic Hierarchy Process (AHP). The authors studied the distribution of the entropy of the priority vectors derived from consistent pairwise comparison judgment matrices (PCJM's) using simulation, and showed that the entropy is normally distributed for these type of matrices. These simulations were based on the results of Vargas (1982) who showed that for gamma distributed judgments the priority vectors derived from perfectly consistent PCJM's are Dirichlet distributed. Noble and Sanchez (1993) noted that information content of decision maker's (DM's) judgments can provide a useful criterion for the acceptability of these judgments and the corresponding PCJM's in AHP.

In this paper we will introduce some information indices and illustrate their use in AHP for assessment of accuracy and for determining when to stop the pairwise comparisons for large dimension PCJM's.

To introduce some notation, let $\underline{p} = (p_1\ p_2\ p_3\ \dots\ p_n)$ denote the priority vector derived from a PCJM matrix of dimension n. We note that the p_i's, $\sum_{i=1}^{n} p_i = 1$, are ratio-scale numbers representing the relative preference (or importance or likelihood) of the attributes considered. We define the entropy of

priority vector $\underset{\sim}{p}$ as

$$H(\underset{\sim}{p}) = - \sum_{i=1}^{n} p_i \ln(p_i). \tag{1}$$

In information theory entropy H is defined as a measure of uncertainty about a discrete random variable X that can take finite values $\{x_1, x_2, ..., x_n\}$ such that $Pr(X=x_i) = p_i$ [see for example, Cover and Thomas (1991)]. In the context of AHP, the priority p_i can be interpreted as the probability (or the relative likelihood) that the attribute (alternative) A_i will be preferred by the DM.

Entropy of a discrete distribution with finite support is always nonnegative, and the minimum value of H is 0 if $p_i=0$ or $p_i=1$ for some i. It is well known that entropy is a concave function and its maximum is reached at the uniform distribution $p_i=1/n$ for all i. The uniform distribution is known as the maximum entropy distribution, and its entropy is given by

$$H^*(1/n, ..., 1/n) = - \sum_{i=1}^{n} (1/n) \ln(1/n) = \ln(n). \tag{2}$$

We note that $H^* \geq H$ implies that $\sum_{i=1}^{n} p_i \ln(p_i) \geq \ln(1/n)$ and the term

$$\sum_{i=1}^{n} p_i \ln(p_i) \tag{3}$$

is known as Shannon's information measure.

The concept of maximum entropy (ME) plays an important role in the comparison of the information content of probability distributions by providing benchmarks. In the context of AHP, the information content of priority vectors from different PCJM's can be compared relative to their entropy difference from the uniform priority vector associated with the initial PCJM of 1's. The entropy difference

$$(H^* - H) = \sum_{i=1}^{n} p_i \ln(p_i) - \ln(1/n) \tag{4}$$

is used by Nobles and Sanchez (1993) as the measure of information content of priority vectors. We note that entropy difference $(H^* - H)$ is scale variant and represents the information increase as a result of using a PCJM different from an initial PCJM of 1's. In other words the difference $(H^* - H)$ measures the additional amount of information added by a DM to a PCJM after an initial PCJM consisting of only 1's.

In what follows we will introduce other information measures and illustrate how they can be used in addressing important issues in AHP.

INFORMATION MEASURES FOR DISCRIMINATION

An important issue during the pairwise comparison process in AHP is to determine whether the priority vectors change significantly as a result of the pairwise comparisons provided by the DM. As noted by Harker (1985) an important issue during the pairwise comparison process is whether to stop making pairwise comparisons. Especially, in cases where the dimension of the PCJM's are large, it is desirable to develop a stopping rule. Harker (1985) has suggested some stopping rules in addressing this problem. It is possible to develop stopping rule similar to those suggested by Harker (1985) using an information measure.

A well known information index for measuring discrepancy between two discrete probability distributions $\underline{p} = (p_1 \, p_2 \, p_3 \cdots p_n)$ and $\underline{q} = (q_1 \, q_2 \, q_3 \cdots q_n)$ is the relative entropy (also referred to as cross entropy) given by

$$I(\underline{p}: \underline{q}) = \sum_{i=1}^{n} p_i \, \ln\left(\frac{p_i}{q_i}\right) \tag{5}$$

which is known as the Kullback-Leibler (KL) discrimination measure. $I(\underline{p}: \underline{q})$ is always nonnegative and is equal to zero if and only if $\underline{p} = \underline{q}$ [see Cooke (1991)].

We note that for the case when $\underline{q}$ is the uniform distribution, then $I(\underline{p}: \underline{q})$ reduces to the entropy difference $(H^* - H)$ which was used by Nobles and Sanchez (1993) as the measure of the information content of the resulting priority vectors. A new information discrimination (ID) index related to the KL measure was introduced in Soofi, Ebrahimi and Habibullah (1994) as

$$ID(\underline{p}: \underline{p}^*) = 1 - \exp[-I(\underline{p}: \underline{p}^*)] = 1 - \exp(H^* - H) , \tag{6}$$

where $\underline{p}^*$ is the uniform distribution. The ID index takes values between 0 and 1 where the values closer to 1 imply a higher information content for $\underline{p}$.

We note that the KL measure can be used to assess whether a priority vector change significantly as a result of the pairwise comparisons provided by the DM. This can be achieved by measuring the KL distance between the two priority vectors during the pairwise comparison process. If $\underline{p}_k$ denote the priority vector after the kth judgment and $\underline{p}_{k-1}$ is the priority vector after the (k-1)st judgment, then we can obtain the KL measure

$$I(\underline{p}_k: \underline{p}_{k-1}) = \sum_{i=1}^{n} p_{ik} \, \ln\left(\frac{p_{ik}}{p_{i,k-1}}\right) \tag{7}$$

and based on its magnitude determine whether to stop after the kth judgment. If the value of $I(\underline{p}_k: \underline{p}_{k-1})$ is small then it implies that the two priority vectors are close to each other and one can decide to stop the questioning process.

We will apply this idea to an example considered by Harker (1985). The example uses the AHP priorities to predict the relative distance of a set of six cities from Philadelphia. Thus, the minimum number of needed judgments is 5 (no redundancy) and the maximum number of judgments is 15 (maximum redundancy). An example of priorities for all these cases of different redundancy are given in Harker (1985). For the given example, the author, using what he refers to as the 5% stopping rule, determines to stop after the 6th judgment. In Table 1 below, we present the KL measures $I(p_k: p_{k-1})$ for k=5,6, ..., 15 which are computed using the Harker's data. We note that the KL measure associated with k=5 represents the information discrepancy between the no redundancy case and the case of no judgment (PCJM full of 1's). Thus the initial value of KL measure is high. But

JUDGMENT, k	$I(p_k: p_{k-1})$
5	0.5356
6	0.0265
7	0.0004
8	0.0147
9	0.0099
10	0.0302
11	0.0071
12	0.0020
13	0.0001
14	0.0000
15	0.0002

Table 1: Using KL measure as a stopping criterion.

after the 7th judgment the value of the KL measure is very small, which may suggest the termination of the questioning process (as in the case of Harker's rule after the sixth judgment). We note that we can also use the difference in the ID index as an equivalent criterion for the stopping problem. The (0,1) scale of the ID index may provide an easier way to interpret the information change.

ASSESSMENT OF ACCURACY USING KL MEASURE

In the previous section we illustrated the use of KL discrimination measure as a criterion to determine when to stop during the pairwise comparison process in AHP. Another important issue in AHP is to derive "accurate" priorities. This is especially important in cases where AHP is used to assess the probabilities (or relative likelihoods) of n events. In the case where the actual event is observed, one can assess the accuracy of the underlying forecasting mechanism. Evaluation of the probability forecasts is an important issue in fields such as risk analysis, where the forecasts are heavily dependent on expert judgments [see for example, Cooke (1991)].

Evaluation of the accuracy AHP priorities derived from redundant judgments have been recently addressed by Forman (1993) who used an experiment where individuals made judgments about the relative sizes of geometric shapes. The actual relative sizes of the five geometric shapes were known and these were compared to those derived by individuals using no redundancy (i.e., only the minimum number of needed judgments; in his case four were used) and the full redundancy (i.e., the maximum number of 10 judgments). There were fifty participants in the experiment. The results of the experiment are summarized in Forman (1993), who used an average quadratic score (or average mean square error) to compare the accuracy of the priorities derived from the cases of no and full redundancy. The author found that the average mean square error without redundancy was more than four times higher than the average mean square error with full redundancy.

In what follows we will use the KL discrimination measure and will analyze the experimental data obtained by Forman (1993) to investigate the relationship between accuracy and level of redundancy. If we let p denote the estimated priority vector from the experiment and q denote the actual priority vector (in this case representing the actual relative sizes of the geometric shapes) then the KL measure $I(p: q)$ defined in (5) provides a measure of accuracy. The advantage of this measure is that as the sample size increases one can show that an asymptotic result can be used for the distribution of the KL measure. As shown in Cooke (1991), if the priority vector p is estimated based on M individual priority vectors, then the quantity

$$2 \times M \times I(\underline{p} : \underline{q}) \tag{8}$$

is approximately a χ^2 random variable with (n-1) degrees of freedom, where n is the dimension of the priority vector. We note that this distribution can be used to define a statistical test in the frequentist sense.

We next apply this idea to the experimental data given by Forman (1993). In Table 2 below we present the estimated (average of M=50 priority vectors) priority vector p for the cases of no redundancy and full redundancy and the actual priority vector q. The KL measure for the full redundancy case was KL= 0.003 and using (8) we obtained Prob(KL<0.003)=0.0102 implying that it is very unlikely to observe a KL value less than 0.003. For the no redundancy case the KL measure was KL=0.019 and the probability was obtained as Prob(KL<0.019)=0.2457 which implies that KL is different than zero for the no redundancy case. Thus, our findings imply that the full redundancy case provides a more accurate priority vector.

SHAPE	ACTUAL q	NO REDUNDANCY: p	FULL REDUNDANCY: p
A	0.475	0.500	0.529
B	0.049	0.050	0.076
C	0.232	0.236	0.174
D	0.151	0.139	0.128
E	0.093	0.076	0.093

Table 2: Actual and Estimated Priorities for Geometric Shapes.

REDUNDANCY AND INFORMATION

We finally present preliminary results concerning the relationship between redundancy and the level of information content (as defined by Noble and Sanchez in (1993)) in the PCJM's. Simulations were run in which the first row of a PCJM was randomly filled up. Then redundant matrices at levels 4, 5 and 6 were created as follows. For level 4, entry (2,3) was filled up in close consistency with the first row, but entries (2,4) and (3,4) were filled up as suggested in Harker (1985); namely, the geometric mean of the intensity of all paths in the directed graph associated with the partially completed matrix which connects alternatives 2 and 4, and 3 and 4, respectively. (As noted by Harker in his paper, "... this approximation scheme in some sense mimics what the decision makers would have to perform if he or she were forced to complete a given comparison...".) For level 5, entries (2,3) and (2,4) were filled up in close consistency with the first row, but entry (3,4) as suggested in Harker (1985). Finally, for level 6, entries (2,3), (2,4) and (3,4) were filled up in close consistency with the first row. In the context of these simulations, "close consistency" mean as follows: take for example entry (2,3); to be perfectly consistent, entry (2,3) is exactly the ratio of entry (1,3) divided by entry (1,2); but instead of forcing perfect consistency, we allowed entry (2,3) to be the closest element in the set

$$\{1/9, 1/8, 1/7, 1/6, 1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

to the ratio of entry (1,3) divided by entry (1,2). The results appear to indicate that the information content of the resulting vector of priorities is rather insensitive to the level of redundancy. Each run consisted of the creation of 1000 consistent matrices (in the sense of Saaty). Level 4 resulted in a mean entropy of 0.133101, Level 5 resulted in a mean entropy of 0.133919, and Level 6 resulted in a mean entropy of 0.134010. Thus, the higher the level, the higher the entropy, but to a very minor degree. Obviously, more simulations for larger matrices is called for in a more complete analysis.

REFERENCES

Cooke, R. M. (1991). *Experts in Uncertainty*, Oxford University Press.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, John Wiley.

Forman, E. H. (1993). "Deriving Ratio Level Measures from Verbal Judgments", Technical Report, George Washington University.

Harker, P. T. (1985). "Incomplete Pairwise Comparisons in the Analytic Hierarchy Process" , Technical Report, University of Pennsylvania.

Noble, E. E. and Sanchez, P. P. (1993). "A note on the Information Content of a Consistent Pairwise Comparison Judgment Matrix of an AHP Decision Maker", *Theory and Decision*, Vol. 34, pp. 99-108.

Soofi, E. S., Ebrahimi, N. and Habibullah, M. (1994). "Information Distinguishability with Application to Analysis of Failure Data", (to appear in the *Journal of American Statistical Association*).

Vargas, L. G. (1982). "Reciprocal Matrices with Random Coefficients", *Mathematical Modeling*, Vol. 3, pp. 69-81.