ELICITATION OF EXPERT KNOWLEDGE FOR ENGINEERING RISK ANALYSIS : MYTH AND REALITY

by

Dr. Malik Ranasinghe

Abstract

This paper addresses three major issues regarding the elicitation of expert knowledge for engineering risk analysis: (1) recognition of some of the implicit assumptions and beliefs; (2) development of an approach to elicit expert knowledge as accurate, calibrated and coherent subjective probabilities; and (3) a study to explore human ability to predict future events and the validity of the implicit assumptions and beliefs in the context of the expert judgements. Directions for future work are suggested.

1.0 Introduction

Most approaches for engineering economic risk analysis use subjective probabilities to describe the uncertainty associated with input variable values [3, 6, 12, 17, 20, 23, 25]. The rationale for their use stems from the belief that actuarial or relative frequency based data are unavailable or unreliable for direct input as probability forecasts [29]. However, elicitation techniques to obtain subjective probabilities for such analyses are limited [6].

Our experience to date has shown that insufficient rigor is applied in obtaining subjective probabilities from experts for economic risk analysis. To use subjective probabilities as input to risk analyses they have to be accurate, calibrated and coherent [18]. An expert is calibrated if, for all events assigned a probability, q, the proportion that actually occur is in fact equal (or close) to q [4]. A set of subjective probabilities are coherent if they are compatible with the probability axioms [18]. A preliminary approach to elicit knowledge as accurate, calibrated and coherent subjective probabilities for economic risk analysis was suggested in ref. [21]. This approach was seen as a vital step towards standardising and computerising, to the extent possible, elicitation and verification of expert input dealing with uncertainty.

An implicit assumption regarding the elicitation of expert knowledge for engineering risk analysis is that the uncertainty of a decision variable (cost, duration, revenue, net present value and so forth) can be more accurately quantified by decomposing the decision variable to a functionally related set of input primary variables. This reflects the engineering penchant to seek more detail as a way of seeking greater precision.

Ravinder et al. [22] state that decomposition is often regarded as a useful technique for reducing the complexity of difficult judgement problems. They studied the application of decomposition to the elicitation of subjective probabilities for discrete events and concluded that if the component probabilities can be assessed with no greater precision than holistic assessment, decomposition reduced random errors associated with probability encoding and as the number of events increased, error reduction occurred only up to a point (i.e. a limit for decomposition exists).

In the context of engineering risk analysis, the main reason for decomposing decision variables to their primary variables is to develop a link between time and cost for economic analysis. It is incorrect to assume that cost is independent of time, because when a duration is either reduced (more resources) or increased (less resources) the net result is a change in the cost. The link between time and cost permits the use of net present value and internal rate of return as decision variables.

The second reason is the basis for the implicit assumption, to reduce the complexity of holistic estimation because experts in engineering economics find it easier to quantify the decomposed primary variables. The same reasoning has been used by other authors to even decompose activity duration into its primary variables [13, 16]. This raises another question; won't it be more accurate if the primary variables are further decomposed.

The main disadvantage of decomposition is the loss of the mental awareness of interdependencies between primary variables that exist when estimating from a holistic approach. While it may be possible to relate the primary variables functionally to the decision variable, it is difficult to model all the interdependencies. Secondly, even if accurate estimates for primary variables are obtained, as the decomposition is continued a model which can link them to provide a reliable estimate of the decision variable may be lacking. As Ravinder et al. [22] have shown a definite limit exists for the decomposition. Thirdly, unless decomposition improves the system significantly, it would be hard to convince an expert that decomposition is necessary for the elicitation of subjective judgements.

Dr. Malik Ranasinghe, Senior Lecturer, Department of Civil Engineering, University of Moratuwa, Sri Lanka.

While this assumption is reasonable for decision variables such as cost and revenue, which in general can be decomposed to a quantity and a rate, it becomes questionable for variables such as duration and productivity, which are sometimes estimated holistically in the elicitation of expert knowledge. The second common belief in engineering risk analysis is that when duration is estimated as a holistic value it underestimates more regularly than when it is estimated as a decomposed functional form.

This paper is structured as follows. The technique to elicit expert knowledge as accurate, calibrated and coherent subjective probabilities is briefly outlined in the next section. A study that was conducted to explore, the human ability to predict future events and the implicit assumptions and beliefs is described in the third section. We conclude by highlighting the inherent difficulties of, and the future work necessary for, developing techniques and experiments to study the elicitation of expert knowledge for engineering risk analysis.

2.0 Elicitation Technique

2.1 Definitions

The approach to elicit accurate, calibrated and coherent subjective probabilities for input variables in economic risk analysis suggested in ref. [21], utilise the interaction between an analyst and experts and consists of four stages, namely, pre-elicitation, elicitation, analysis and verification.

Analyst refers to the individual (or group of individuals) within the firm responsible for conducting economic and financial feasibility, scheduling and cost analyses. Analyst is the key person in the elicitation approach because he must elicit from the expert his belief about the uncertainty of input variables as subjective probabilities.

Expert refers to individuals both within and external to the firm who provide key input dealing with technical, economic, revenue, cost, financial, productivity and schedule information and is a person who in his area has some degree of training, experience and/or knowledge significantly greater than that of the general population [26]. These experts are drawn from the fields of economics, finance, design, construction and so forth, when the analyst believes that they possess the most relevant knowledge and information regarding the uncertainty of an input variable. In general, they are substantive experts, who in a given domain, assess events in their field of expertise [26].

2.2 Pre-elicitation

The objective of pre-elicitation is for the analyst to train the expert in the task of quantifying his belief as subjective probabilities. We adopt the three phase approach of motivating, structuring and conditioning suggested by Spetzler and Stael von Holstein [24]. During the motivating phase the analyst builds a rapport with the expert by introducing him to the elicitation task, and attempts to explore whether any motivational biases [24] are present. The analyst points out the difference between deterministic and probabilistic prediction and that there is no commitment (firm projection) inherent in a probability distribution. During the structuring phase the uncertain variables are clearly defined in terms of the structure of the problem. Also, the importance of the variables to the problem is explained to demonstrate the relevance of the elicitation process and to gain the expert's full co-operation. Such co-operation is essential for a successful elicitation [6, 14, 15, 24]. The aim of the third phase is to condition the expert to think fundamentally about his judgements and to avoid cognitive biases [24]. Winkler [27] observed that when subjects are asked for a shape of their subjective distribution many try to associate it to a normal distribution. Our experience is similar where some experts believe that percentiles should be symmetric to be consistent. The expert should be made aware of this common mistake, so that features like skewness will be considered during the elicitation.

2.3 Elicitation

With the completion of the pre-elicitation stage the expert is ready to quantify subjectively his belief about the uncertain variable. The elicitation session is based on a questionnaire suggested in ref. [21], that would elicit the 5^{th}, 25^{th}, 50^{th} (median), 75^{th} and 95^{th} percentiles of the subjective prior probability distribution for each uncertain input variable.

The questionnaire combines direct probability responses with chance responses to provide cross checking for consistency [21]. The basis for direct probability responses is the variable interval method [14]. The analyst is expected to follow the general format of the questionnaire when conducting the session. However, at his discretion the analyst can adapt the interview to suit different situations. Since the questionnaire is based primarily on the variable interval technique, it is easily standardised and automated to use for those input variables selected for interactive computer interviews.

Whenever possible a group of experts are used for the elicitation because it has been found that consensus judgements from a group to be better than the individual judgements [2, 6, 9, 11, 14, 28]. Once the initial subjective estimates are made by the expert, he is provided feedback on his assessments in the form of a discussion between the analyst and expert, and expert and expert. The expert can revise his prior judgements after the discussion. This process is based on the nominal group technique [7, 10].

Since the experts are provided feedback, their subjective estimates would be close to consensus. Also, because of the difficulty in measuring the variability in expert accuracy, consensus subjective estimates are obtained by assigning equal weights to all the experts [2, 19]. The routine aspects of feedback such as exchanging individual estimates, obtaining revised estimates and the task of combining subjective estimates can be readily automated.

2.4 Analysis

The coherence of subjective probabilities is ensured by converting subjective estimates to moments and shape characteristics of the input variable. The converted moments are its expected value and standard deviation while the shape characteristics are its skewness and kurtosis. The moments describe the location and dispersion of the uncertain input variable while shape characteristics are non-dimensional measures which describe the shape of its distribution [5]. The method suggested in ref. [23], enables the analyst to approximate the subjective estimates for an input variable to a Pearson type distribution. The high flexibility of the Pearson family which embraces a large number of unimodal distributions approximates most of the subjective estimates to Pearson type distributions (see tables for the Pearson family of distributions tabulated by Amos and Daniel [1]).

However, in some instances subjective estimates may not approximate to a Pearson distribution for the specified maximum cumulative error. In these situations the expert is made aware of the necessity for subjective probabilities to be coherent and is asked to modify the 25^{th} and 75^{th} percentile estimates. These two estimates are elicited only to approximate a Pearson type distribution. The expected value and standard deviation for the input variable are derived from the 5^{th}, 50^{th} and 95^{th} percentile values and are independent of the approximated distribution [23]. In our experience, the consensus among experts is that, if necessary, they would be willing to change within reason the 25^{th} and 75^{th} percentile estimates because those two are the least important to them.

2.5 Verification

As the final stage of the elicitation, the subjective prior probability distribution is verified to see if the expert is in total agreement with it (i.e. reflects his belief). The questionnaire performs cross checking for consistency of the percentiles as verification. In addition, the expert is informed of the expected value, standard deviation and shape characteristics of the approximated Pearson type distribution. While the skewness and the kurtois gives an indication of the shape of the distribution, a better verification method would be to provide a graphical display of the approximated density function. It is proposed to develop a computer program to display the probability density function of the approximated Pearson type to be viewed by the expert. Incorporating such a verification process to the elicitation technique would provide the analyst and expert with greater confidence that the approximated distribution represents the expert's belief.

2.6 Summary

In summary, the elicitation of subjective probabilities involves four phases - pre-elicitation, elicitation, analysis and verification. A distinct role for the computer exists for the latter three phases to expedite the process. We note the assertion by Spetzler and Stael von Holstein [24] that the role of the computer should be minimised during the elicitation process. However, because of the need to obtain assessments for a large number of input variables, use should be made both of a standard approach and the computer to increase the efficiency of the process.

3.0 The Study

3.1 General

This section describes a study that was conducted to explore the human ability to predict future events using the decomposition of an activity duration. The objective of this study was to make a small step towards exploring an issue that is largely ignored in the estimation literature and to provide the motivation for a more extensive study on the elicitation of subjective probabilities for continuous random primary variables.

The next four subsections describe hypotheses, test statistics, an experiment and the analysis to study the human ability to predict duration. In addition, the identified implicit assumptions and beliefs that exist in engineering risk analysis regarding decomposed versus holistic estimation of judgements for activity duration are explored.

3.2 Hypotheses

Nine hypotheses are suggested to study decomposed versus holistic estimation of an activity duration. The first hypothesis is about the precision of assessments for an activity duration from the two approaches. The other eight are for assessed expected values and standard deviations for an activity duration to compare holistic versus decomposed estimation.

Hypothesis 1: The precision of assessments for an activity duration from holistic or decomposed estimation are similar.

The precision of assessments from the two approaches are measured using the coefficient of variation for duration of an activity; a non-dimensional measure of variation.

$$H_0: E[V_i] = 0 ; H_1: E[V_i] \neq 0$$

where $V_i = V_{Di} - V_{Wi}$, is the difference between a pair of coefficients of variation of the assessed activity duration from decomposed (V_{Di}) and holistic (V_{Wi}) estimation.

Hypothesis 2: When experts are asked to assess the expected value for duration of an activity from the holistic approach, that assessment will be the true value.

$$H_0: E[T] = E[T_W] ; H_1: E[T] \neq E[T_W]$$

Hypothesis 3: When experts are asked to assess the standard deviation for duration of an activity from the holistic approach, that assessment will be the true value.

$$H_0: \sigma_T = \sigma_{Tw} ; H_1: \sigma_T \neq \sigma_{Tw}$$

Hypothesis 4: When experts are asked to assess the expected value for duration of an activity from the decomposed approach, that assessment will be the true value.

$H_0: E[T] = E[T_D]\,;\, H_1: E[T] \neq E[T_D]$

Hypothesis 5: When experts are asked to assess the standard deviation for duration of an activity from the decomposed approach, that assessment will be the true value.

$H_0: \sigma_T = \sigma_{TD}\,;\, H_1: \sigma_T \neq \sigma_{TD}$

Hypothesis 6: When experts are asked to assess the expected value for duration of an activity from the holistic approach, that assessment will be an underestimation of the true value.

$H_0: E[T] = E[T_W]\,;\, H_1: E[T] > E[T_W]$

Hypothesis 7: When experts are asked to assess the standard deviation of an activity from the holistic approach, that assessment will be an underestimation of the true value.

$H_0: \sigma_T = \sigma_{Tw}\,;\, H_1: \sigma_T > \sigma_{Tw}$

Hypothesis 8: When experts are asked to assess the expected value for duration of an activity from the decomposed approach, that assessment will be an underestimation of the true value.

$H_0: E[T] = E[T_D]\,;\, H_1: E[T] > E[T_D]$

Hypothesis 9: When experts are asked to assess the standard deviation for duration of an activity from the decomposed approach, that assessment will be an underestimation of the true value.

$H_0: \sigma_T = \sigma_{TD}\,;\, H_1: \sigma_T > \sigma_{TD}$

where $E[T_W]$ and $E[T_D]$ are the expected values and σ_{TW} and σ_{TD} are the standard deviations assessed for the activity duration from holistic and decomposed subjective estimation.

While a hypothesis test is done for the hypothesis, only significance tests (i.e. a hypothesis can only be rejected) are done for the next eight because there is only one sample to test all of the hypotheses. Hypothesis (1) tests whether there is a difference between the precision of assessments (i.e. coefficients of variation) from the two approaches. The next four hypotheses (2 to 5) provide the basis to compare the two approaches. By calculating the percentages of the number of times an individual hypothesis is rejected, given the group of experts and the amount of information available during the elicitation, the two approaches are compared. Hypotheses (6) to (9) are included to explore the traditional belief in engineering risk analysis that holistic approach underestimates duration more regularly than the decomposed approach (i.e. holistic approach will be rejected more times than the decomposed).

It must be stressed that the objective of this study is not to select the "better" method, but to compare the two methods available for estimating duration when experts participate in subjective elicitation. The "better" method is a consensus approach after estimating from both approaches. However, it is not practical as a subjective elicitation technique.

3.3 Test Statistics

Assumption 1: A sample of durations to complete an activity constitutes a random sample from a normal distribution with both mean and standard deviation unknown.

Since the sample of the measured durations are for the same activity it is reasonable to expect the measurements to be symmetric around the mean value. Then, test statistic for the difference between paired coefficients of variation for activity duration is [8],

$$t_{paired} = \frac{\bar{V}}{S_V/\sqrt{n}} \qquad (1)$$

where $\bar{V}$ and S_V are sample mean and standard deviation, respectively for V_i's and n is the sample size.

The rejection regions for level α tests are (see Figure 1),

Hypothesis	Rejection Region
1	$t_{paired} \geq t_{\frac{\alpha}{2},n-1}$ or $t_{paired} \leq -t_{\frac{\alpha}{2},n-1}$

Test statistic for the assessed expected value for activity duration is [8],

$$t = \frac{\bar{T} - E[T]}{S/\sqrt{n}} \qquad (2)$$

where $\bar{T}$ and S are the sample mean and standard deviation, and $E[T]$ is either the assessed $E[T_W]$ or $E[T_D]$.

The rejection regions for level α tests are (see Figures 1 and 2),

Hypothesis	Rejection Region
2 and 4	$t \geq t_{\frac{\alpha}{2},n-1}$ or $t \leq -t_{\frac{\alpha}{2},n-1}$
6 and 8	$t \geq t_{\alpha,n-1}$

Test statistic for the assessed standard deviation for activity duration is [8],

$$X^2 = \frac{(n-1)\,S^2}{\sigma_T^2} \qquad (3)$$

where S is the sample standard deviation and σ_T is the assessed σ_{TW} or σ_{TD}.

The rejection regions for level α tests are (see Figures 3 and 4),

Hypothesis	Rejection Region
3 and 5	$X^2 \geq X^2_{\frac{\alpha}{2},n-1}$ or $X^2 \leq X^2_{1-\frac{\alpha}{2},n-1}$
7 and 9	$X^2 \geq X^2_{\alpha,n-1}$

3.4 Experiment

3.4.1 The Activity

The activity to obtain a sample of durations to test the hypothesis should: permit the assessment of duration from holistic and decomposed subjective estimation; permit the measurement of actual duration; be repetitive; utilize the expertise of engineers - read, interpret and visualise engineering drawings.

While an activity such as the repetitive construction of a column, a beam or a footing is ideal, the inherent difficulties of field experiments such as: free access to a construction site; measurement of actual duration; and time constraints; makes the selection infeasible. Instead, the assembly of a LEGOLAND wheel loader (model#6658) was selected as the activity for the experiment. In addition to satisfying the requirements of an activity for the experiment, the LEGOLAND model permitted the experiment to be conducted in a laboratory setting.

3.4.2 Procedure

First, the objectives of the experiment were explained to the participants. This explanation was based on the procedure of pre- elicitation discussed in section (2.2). Thereafter, using two questionnaires the desired subjective percentile values for the activity duration were elicited from all the participants. The first questionnaire elicited the duration in minutes to assemble the complete model in accordance with the drawing (*i.e.* holistic estimation). The second elicited the duration in seconds to identify and attach one component to the model in accordance with the drawing (i.e. decomposed estimation). Finally, the actual duration to assemble the model by each participant was measured (see Table 1).

The participants were postgraduate and final year undergraduate students in civil engineering at The University of British Columbia, Vancouver, Canada, who had followed courses in engineering economics and risk analysis. The subjective elicitation based on the drawings for the LEGOLAND model and its assembly in accordance with drawings utilized their expertise as engineers.

Each participant is considered as an independent source for hypothesis testing. That is, evaluated expected values, standard deviations and coefficients of variation from the two approaches for each participant are the basis for a set of hypotheses. The measured actual durations constitute the sample to obtain statistics to test each set of hypotheses.

3.5 Analysis

The expected values, standard deviations and coefficients of variation for duration (holistic - duration to assemble the complete model; decomposed - duration to identify and attach one component to the model) are evaluated using the elicited subjective percentile values [23]. However, for hypotheses on decomposed estimation, the moments for duration to assemble the complete model have to be evaluated.

3.5.1 Moments from Decomposition

For a LEGOLAND model consisting of l components, the duration to assemble the complete model is,

$$T_D = \sum_{i=1}^{l} t \qquad (4)$$

where t is the duration to identify and attach one component to the model. It is assumed that t is identical for all the components.

The expected value for duration to assemble the complete model from decomposed estimation is,

$$E[T_D] = l\,E[t] \qquad (5)$$

where $E[t]$ is the evaluated expected value for duration to identify and attach one component to the model and assumed to be identical for all the components.

Assumption 2: The estimated coefficients of variation for duration for an activity from holistic and decomposed estimation are similar (i.e. $V_D \equiv V_W \equiv V$).

Assumption (2) is tested by hypothesis (1). Therefore, from the definition for coefficient of variation,

$$\frac{\sigma_{TW}}{E[T_W]} \equiv \frac{\sigma_{TD}}{E[T_D]} \equiv \frac{\sigma_t}{E[t]} \cong V. \qquad (6)$$

The standard deviation for duration to assemble the complete model depends on the assumption regarding the correlation between duration to assemble individual components. The variance for duration to assemble the complete model is;

$$\mu_2(T_D) = \sum_{i=1}^{l}\sum_{j=1}^{l} cov(t_i, t_j) \qquad (7)$$

From definition, $\mu_2(T_D) \geq 0$. Hence,

$$\mu_2(T_D) = l\,\sigma_t^2 + \rho l(l-1)\,\sigma_t^2 \geq 0 \qquad (8)$$

where σ_t is the evaluated standard deviation for duration to identify and attach one component to the model and ρ is the correlation coefficient between two component durations. Rewriting equation (8),

$$\mu_2(T_D) = \sigma_t^2 [l + \rho(l^2 - l)] \geq 0 \tag{9}$$

Since $\sigma_t^2 \geq 0$, for $\mu_2(T_D)$ to exist

$$[l + \rho(l^2 - l)] > 0 \tag{10}$$

Therefore,

$$\rho > \frac{-l}{l^2 - l} \tag{11}$$

Hence, lim $\rho = 0$. From definition $\rho < 1$.

At the extremes, component durations are either uncorrelated or perfect positive correlated.

When the component durations are assumed to be uncorrelated, the variance for duration to assemble the complete model from equation (8) is,

$$\mu_2(T_D) = l\sigma_t^2 \tag{12}$$

Hence, the standard deviation is,

$$\sigma_{TD} = l\sigma_t \tag{13}$$

When the component durations are assumed to be perfect positive correlated, the variance for duration to assemble the complete model from equation (8) is,

$$\mu_2(T_D) = l^2\sigma_t^2 \tag{14}$$

Hence, the standard deviation is,

$$\sigma_{TD} = l\sigma_t \tag{15}$$

where σ_t is the evaluated standard deviation for duration to identify and attach one component to the model and assumed to be identical for all the components.

Comparing equations (5), (6), (13) and (15) it is evident that relationship given by equation (15) evaluates the standard deviation for duration to assemble the complete model from decomposed estimation.

3.5.2 Experimental Results

The actual duration to assemble the LEGOLAND model and the expected values, standard deviations and coefficients of variation for duration from holistic and decomposed estimation are given in Table 1. All of the participant are given an identification #. The subjective estimates of participant # 15 were rejected.

The sample mean and standard deviation for the sample of actual measured durations to assemble the LEGOLAND model in minutes are, $\bar{T} = 15.95$ and $S = 7.99$. The sample mean and standard deviation for V_i's, the difference between paired coefficients of variation are, $\bar{V} = 0.0518$ and $S_V = 0.1714$. The test statistic t_{paired} for 27 participants from equation (1) is equal to 1.5698.

The t values are assessed expected values and X^2 values for assessed standard deviations for holistic and decomposed estimation evaluated from equations (2) and (3) are given in Table 2. Table 1, Table 2, sample means and standard deviations are obtained from a computer program called "LEGO".

3.5.3 Hypotheses Testing

All of the hypotheses are tested at confidence level $\alpha = 95\%$. Then, $t_{\frac{\alpha}{2},n-1} = 2.052$, $t_{\alpha,n-1} = 1.703$, $X^2_{\frac{\alpha}{2},n-1} = 43.194$, $X^2_{1-\frac{\alpha}{2},n-1} = 14.573$, and $X^2_{\alpha,n-1} = 40.113$ for $n = 28$.

Since t_{paired} is within the acceptance region, hypotheses (1) is accepted at 95% confidence level. Hence, the assumption that coefficients of variation for activity duration from holistic and decomposed estimation are similar is verified.

The results of the significance tests for hypotheses (2) and (9), total and percentages of the number of times an individual hypothesis is rejected at 95% confidence level are given in Table 3. 'R' indicates that the hypothesis is rejected, while 'NR' indicates that it is not rejected at 95% confidence level. The significance tests show high rejection rates and similar percentage values for both methods of estimation.

3.5.4 Discussion

The classical approaches for hypotheses testing and the generally accepted confidence levels used in this study may not be the most suitable to test human ability to predict future events because of the high variability in predictions from individual to individual. While broader confidence levels reduce the rejection rates they may not be acceptable from a statistical view point. This highlights the inherent difficulties in developing experiments to measure human ability to predict future events.

Similar rejection percentages of individual hypothesis confirm the view that neither is the "better" method. Those for hypotheses (6) to (9) contradict the traditional belief that holistic estimation underestimates duration more regularly than decomposed estimation. If decomposition is not critical to the decision problem - when only activity and project duration estimates are desired; the approach preferred by the analyst and experts can be used for subjective elicitation. However, if decomposition is important to the decision problem - [22]; economic risk analysis of engineering projects; decomposed estimation alone can be used with confidence that the precision of assessments are similar to those from holistic estimation and that it can reduce random errors [22].

4.0 Conclusions and Future Work

This paper addressed three major issues: (1) identification of some of the implicit assumptions and beliefs regarding the elicitation of expert knowledge for engineering risk analysis; (2) development of a preliminary approach to elicit expert knowledge for economic risk analysis as accurate, calibrated and coherent subjective probabilities; and (3) a preliminary study to explore human ability to predict future events and the validity of the implicit assumption and belief in the context of the expert judgements.

The elicitation approach combined the theoretical requirements of subjective probabilities with a practical process.

The process was developed by transforming proven techniques from other fields of study to the requirements of risk analysis in engineering. The role of the computer and the use of a standard approach to expedite the process was identified at every stage. The pre- elicitation stage requires a high level of person to person interaction. Hence, there is little use of the computer during pre-elicitation. Since the questionnaire to elicit the subjective percentile values is based primarily on the variable interval technique, the elicitation stage can be standardised and automated for interactive computer interviews. The routine aspects of feedback and obtaining of consensus subjective estimates can also be done by the computer. The coherence of subjective probabilities is ensured by using the Pearson family of distributions.

At present, the experience in using these approaches in field applications is limited. Future work will concentrate on building up experience from field applications and on refining and validating the elicitation approach. This is essential for the proposed technique to become a practical tool in economic risk analysis.

The study to explore human ability to predict future events consisted of a set of hypotheses, test statistics, an experiment and analysis to compare holistic versus decomposed methods for estimating duration when experts participate in subjective elicitation. While classical approaches and confidence levels used in this study may be too restrictive to test the human ability to predict future events, they provide a statistically accepted framework. The interpretation of the results as percentages of rejection rates reduced some of the restrictions. The first hypothesis verified the assumption that coefficients of variation in subjective assessments for duration from holistic and decomposed estimation were similar. The next four supported the view that neither was the "better" estimation approach to elicit subjective assessments for duration. The last four while contradicting the traditional belief in engineering about holistic estimation of duration confirmed the view regarding the "better" approach. It must be stressed that these are observations based on this study and in no way can they be generalised to the decomposition of decision variables.

The recognition, that some of the implicit assumptions and beliefs in engineering risk analysis should be explored when dealing with the human ability to predict future events and the inherent difficulties in developing experiments and methods to test such beliefs are some of the benefits of this study. It is recommended that this topic be explored further.

References

1. Amos, D.E., & Daniel, S.L. (1971), "Tables of Percentage Points of Standardised Pearson Distributions", Research Report # SC-RR - 710348, NTIS, Virginia, U.S.A.

2. Ashton, A.H., & Ashton, R.H. (1985), "Aggregating Subjective Forecasts: Some Empirical Results", Management Science, 31(12), pp.1499-1508.

3. Bjornsson, H.C. (1977), "Risk Analysis of Construction Cost Estimates", Transactions, American Association of Cost Engineers, pp.182-189.

4. Budescu, D.V., Wallsten, T.S. (1987), "Subjective Estimation of Precise and Vague Uncertainties", Judgmental Forecasting, Eds. G. Wright and P. Ayton, John Wiley & Sons, New York.

5. Bury, K.V. (1975), Statistical Models in Applied Science, John Wiley & Sons, New York.

6. Cooper, D.F., & Chapman, C.B. (1987), Risk Analysis for Large Projects, Models, Methods and Cases, John Wiley & Sons, New York.

7. Delbecq, A.L., Van de Ven, A.H., & Gustafson, D.H. (1975)., Group Techniques for Program Planning, Scott Foresman, Glenview, Illinois.

8. Devore, J.L. (1982)., Probability & Statistics for Engineering and the Sciences, Brooks / Cole Publishing Company, Monterey, California.

9. French, S. (1985), "Group Consensus Probability Distributions: A Critical Survey", Bayesian Statistics 2, Eds. J.M. Bernardo et al., Elsevier Science Publishing Co., New York, 2, pp.183-202.

10. Gustafson, D.H., Shukla, R.K., Delbecq, A., & Walster, G.W. (1973), "A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals, Interacting Groups, Delphi Groups, and Nominal Groups", Organisational Behaviour & Human Performance, 9, pp.281-291.

11. Hampton, J.M., Moore, P.G., & Thomas, H. (1973), "Subjective Probability and its Measurement", Journal of Royal Statistical Society, A, 136(Pt 1), pp.21-42.

12. Hayes, R.W., Perry, J.G., Thompson, P.A., & Wilmer, G. (1986), Risk Management in Engineering Construction, Implications for Project Managers, Thomas Telford Ltd., London.

13. Hendrickson, C., Martinelli, D., & Rehak, D. (1987)., "Hierarchical Rule-Based Activity Duration Estimation", Journal of Construction Engineering & Management, 113(2), pp.288-301.

14. Huber, G.P. (1974), Methods for Quantifying Subjective Probabilities and Multi-Attribute Utilities, Decision Science, 5(3), pp.430-458.

15. Hull, J.C. (1980), The Evaluation of Risk in Business Investment, Pergamon Press, U.K.

16. Jaafari, A. (1984)., "Criticism of CPM for Project Planning Analysis", Journal of Construction Engineering & Management, 110(2), pp.222- 233.

17. Jaafari, A. (1988), "Project Viability and Economic Risk Analysis", Journal of Management in Engineering, 4(1), pp.29-45.

18. Lindley, D.V., Tversky, A., & Brown, R.V. (1979), "On the Reconciliation of Probability Assessments", Journal of Royal Statistical Society A, 142(Pt 2), pp.146-180.

19. Lock, A. (1987), "Integrating Group Judgements in Subjective Forecasts", Judgmental Forecasting, Eds. G. Wright and P. Ayton, John Wiley & Sons, New York.

20. Perry, J.G., & Hayes, R.W. (1985), "Risk and its Management in Construction Projects", Proceedings, Institution of Civil Engineers, U.K., 78(Pt 1), pp.499-521.

21. Ranasinghe, M., & Russell, A.D. (1990), "Elicitation of Subjective Probabilities for Risk Analysis in Engineering Construction", Managing Projects, Eds. V. Ireland and T. Uher, CIB 90(4), Sydney, Australia, 376-387.

22. Ravinder, H.V., Kleinmuntz, D.W., & Dyer, J.S. (1988)., "The Reliability of Subjective Probabilities Obtained Through Decomposition", Management Science, 34(2), pp.306-312.

23. Russell, A.D., & Ranasinghe, M. (1991), "Analytical Approach for Economic Risk Quantification of Large Engineering Projects", Journal of Construction Management and Economics, U.K. (to appear).

24. Spetzler, C.S. & Stael von Holstein, C-A. (1975), "Probability Encoding in Decision Analysis", Management Science, 22(3), pp.340-358.

25. Thompson, P.A. & Wilmer, G. (1985), "CASPAR - A Program for Engineering Project Appraisal and Management", 2nd International Conference on Civil & Structural Engineering Computing, 1, pp.75-81.

26. Wallsten, T.S., & Budescu, D.V. (1983), "Encoding Subjective Probabilities: A Psychological and Psychometric Review", Management Science, 29(2), pp.151-173.

27. Winkler, R.L. (1967), "The Assessment of Prior Distributions in Bayesian Analysis", Journal of the American Statistical Association, T&M Section, 62, pp.776-800.

28. Winkler, R.L. (1968), "The Consensus of Subjective Probability Distributions", Management Science, 15(2), pp.B61-B75.

29. Wright, G. & Ayton, P. (1987), "The Psychology of Forecasting", Judgmental Forecasting, Eds. G. Wright and P. Ayton, John Wiley & Sons, New York.

Table 1: Actual and Estimated Statistics for the Activity Duration (minutes)

#	Actual Duration	Holistic Estimation			Decomposed Estimation		
		$E[T_W]$	σ_{T_W}	V_W	$E[T_D]$	σ_{T_D}	V_D
01	10.0	12.00	4.02	0.335	15.01	6.32	0.421
02	13.0	3.18	1.41	0.442	10.67	3.75	0.352
03	16.0	28.70	13.94	0.485	24.29	15.84	0.652
04	16.0	15.55	6.33	0.407	24.49	15.65	0.639
05	24.0	17.44	7.03	0.403	5.29	3.03	0.572
06	18.0	30.00	12.56	0.418	16.00	7.50	0.469
07	21.0	11.81	5.91	0.500	11.26	5.57	0.495
08	7.75	5.37	2.57	0.478	6.99	3.58	0.512
09	16.83	29.07	8.89	0.306	52.43	34.00	0.636
10	13.75	20.92	8.89	0.425	11.54	4.03	0.349
11	25.5	6.81	2.53	0.371	15.80	5.37	0.340
12	13.42	6.28	2.05	0.326	5.23	0.95	0.181
13	47.0	30.00	10.05	0.335	16.00	8.43	0.527
14	24.5	14.18	3.96	0.279	10.47	2.96	0.283
15	9.25	–	–	–	–	–	–
16	9.5	10.18	1.78	0.175	39.86	1.96	0.049
17	18.92	10.00	3.77	0.377	10.86	7.91	0.728
18	16.5	25.00	4.27	0.171	32.99	5.54	0.168
19	9.92	19.07	9.56	0.502	5.48	0.96	0.175
20	10.05	10.55	3.12	0.295	6.30	0.68	0.109
21	10.82	3.00	0.75	0.251	4.27	1.07	0.251
22	21.3	30.00	7.54	0.251	32.00	8.04	0.251
23	9.08	9.18	3.53	0.384	7.57	3.75	0.496
24	16.88	18.52	6.14	0.332	9.88	2.90	0.293
25	17.0	5.99	3.65	0.610	13.13	8.31	0.632
26	11.9	12.74	3.19	0.250	26.86	18.01	0.671
27	10.78	10.00	3.02	0.302	5.33	1.61	0.302
28	8.09	9.63	2.81	0.292	3.40	2.03	0.597

Table 2: Test Statistics for Expected Values and Standard Deviations

#	Holistic Estimation		Decomposed Estimation	
	t	χ^2	t	χ^2
01	2.6204	106.55	0.6238	43.11
02	8.4610	870.43	3.5038	122.32
03	−8.4479	8.86	−5.5249	6.87
04	0.2649	42.97	−5.6556	7.03
05	−0.9873	34.85	7.0641	187.63
06	−9.3060	10.91	−0.0299	30.58
07	2.7430	49.29	3.1116	55.55
08	7.0133	261.42	5.9386	134.12
09	−8.6931	21.78	−24.8280	1.49
10	−3.2931	21.78	2.9278	106.04
11	6.0559	269.82	0.1008	59.77
12	6.4120	410.52	7.1030	1914.45
13	−9.3060	17.05	−0.0299	24.23
14	1.1727	109.87	3.6346	196.36
16	3.8230	544.55	−15.8399	449.77
17	3.9456	121.24	3.3731	27.51
18	−5.9931	94.39	−11.2849	56.14
19	−2.0640	18.83	6.9395	1864.32
20	3.5778	177.44	6.3962	3676.43
21	8.5836	3030.90	7.7443	1498.43
22	−9.3060	30.31	−10.6311	26.64
23	4.4856	138.40	5.5587	122.15
24	−1.6996	45.67	4.0268	205.07
25	6.6033	129.07	1.8695	24.96
26	2.1301	169.33	−7.2246	5.31
27	3.9456	189.43	7.0376	665.97
28	4.1907	217.61	8.3203	418.09

Table 3: Significance Tests for Hypotheses (2) to (9) at 95% confidence level

#	Hypotheses							
	2	3	4	5	6	7	8	9
01	R	R	NR	NR	R	R	NR	R
02	R	R	R	R	R	R	R	R
03	R	R	R	R	NR	NR	NR	NR
04	NR	NR	R	R	NR	R	NR	NR
05	NR	NR	R	R	NR	NR	R	R
06	R	R	NR	NR	NR	NR	NR	NR
07	R	R	R	R	R	R	R	R
08	R	R	R	R	R	R	R	R
09	R	NR	R	R	NR	NR	NR	NR
10	R	NR	R	R	NR	NR	R	R
11	R	R	NR	R	R	R	NR	R
12	R	R	R	R	R	R	R	R
13	R	NR	NR	NR	NR	NR	NR	NR
14	NR	R	R	R	NR	R	R	R
16	R	R	R	R	R	R	NR	R
17	R	R	R	NR	R	R	R	NR
18	R	R	R	R	NR	R	NR	R
19	R	NR	R	R	NR	NR	R	R
20	R	R	R	R	R	R	R	R
21	R	R	R	R	R	R	R	R
22	R	NR	R	NR	NR	NR	NR	NR
23	R	R	R	R	R	R	R	R
24	R	R	R	R	R	R	R	R
25	R	R	NR	NR	R	R	R	NR
26	R	R	R	R	R	R	NR	NR
27	R	R	R	R	R	R	R	R
28	R	R	R	R	R	R	R	R
Total	24R	20R	22R	21R	16R	19R	16R	18R
	88.89%	74.07%	81.48%	77.77%	59.26%	70.37%	59.26%	66.66%

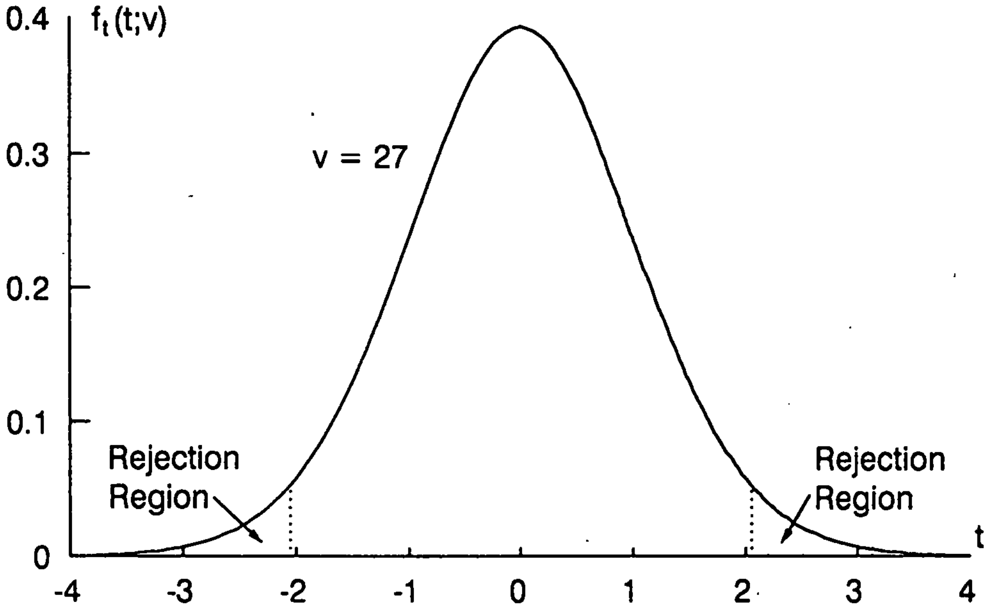

Fig.1 t Distribution for Two Tailed Test

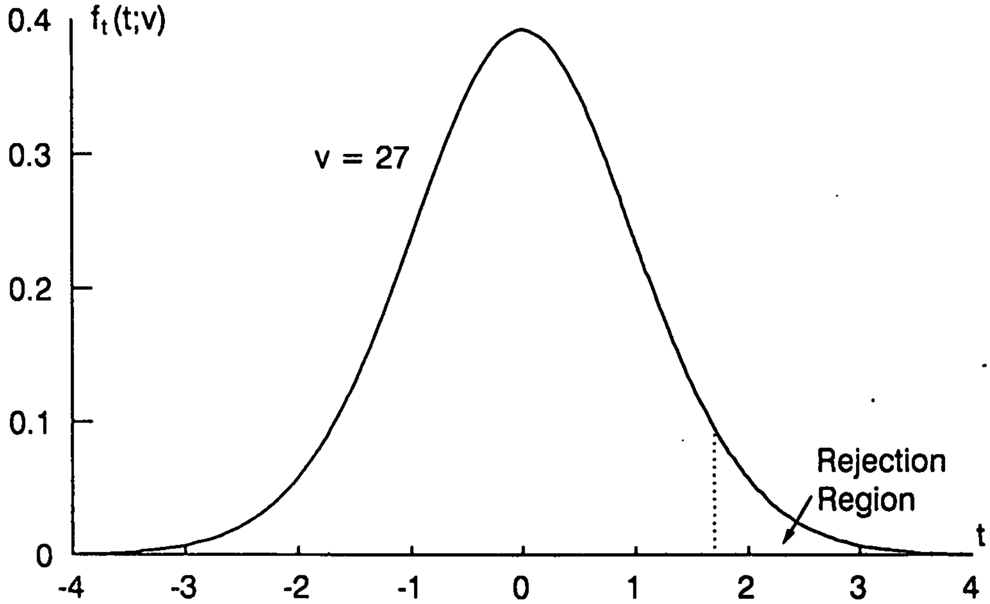

Fig.2 t Distribution for Upper Tailed Test

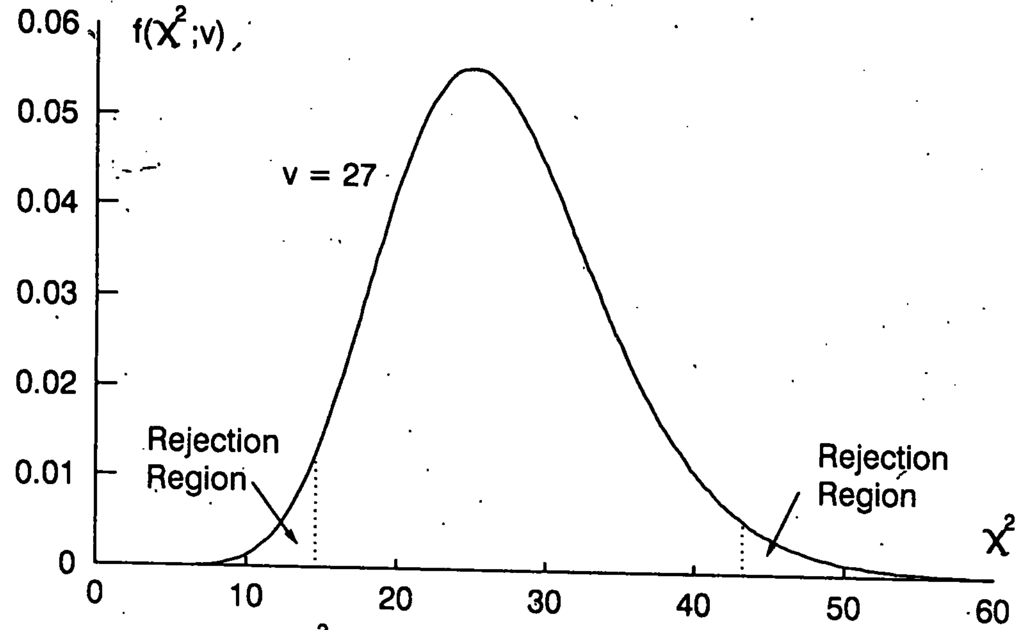

Fig.3 χ^2 Distribution for Two Tailed Test

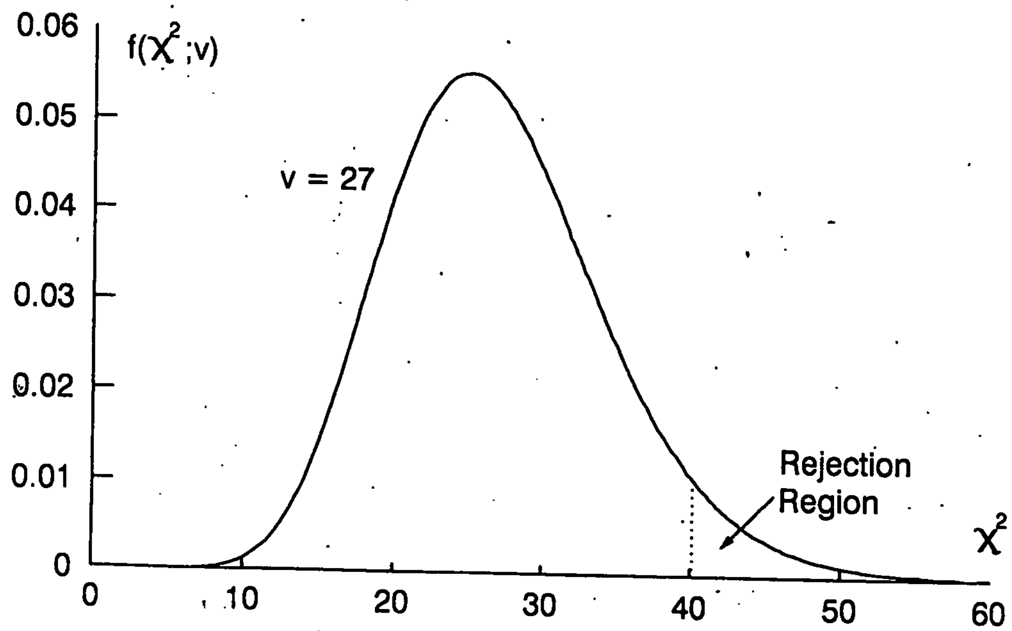

Fig.4 χ^2 Distribution for Upper Tailed Test