Expert judgement models in quantitative risk assessment

Tony Rosqvist
VTT Automation
Box 1301
02044 VTT
Finland

Risto Tuominen
VTT Automation
Box 1306
33101 Tampere
Finland

Abstract

Expert judgement is a valuable source of information in risk management. Especially, risk-based decision making relies significantly on quantitative risk assessment, which requires numerical data describing the initiator event frequencies and conditional probabilities in the risk model. This data is seldom found in databases and has to be elicited from qualified experts.

In this report, we discuss some modelling approaches to expert judgement in risk modelling. A classical and a Bayesian expert model is presented and applied to real case expert judgement data. The cornerstone in the models is the log-normal distribution, which is argued to be a satisfactory choice for modelling degree-of-belief type probability distributions with respect to the unknown parameters in a risk model. Expert judgements are qualified according to bias, dispersion, and dependency, which are treated differently in the classical and Bayesian approaches. The differences are pointed out and related to the application task.

Differences in the results obtained from the different approaches, as applied to real case expert judgement data, are discussed. Also, the role of a degree-of-belief type probability in risk decision making is discussed.

1. Introduction

In risk assessment [1], the most used risk modelling approach is probably the fault- and event tree model. A fault- and event tree model entails qualitative, as well as quantitative, information that reflects the risk analyst's/risk manager's knowledge *and beliefs* related to possible consequences and frequencies of events in various risk scenarios.

Statistical data might be found to yield frequency estimates for the initiators in the fault- and event tree model, but this is rarely the case when estimating the conditional probabilities related to the various barriers in that model. Thus, expert judgement is practically always needed to complete the quantification of the risk model. This naturally raises questions about the "objectivity" of the risk assessment [2].

Every risk model is based on assumptions made by the analysts. These assumptions relate to the usability and relevance of data records of past incidents/accidents, the expertise available, the aggregation of expert judgements, the quantitative risk criteria (such as the ALARP-principle), etc. As modelling assumptions change, the outcome of the risk assessment might change. One of the main sources of uncertainty is usually related to the use of expert judgement.

In section 2, we will present a classical and a Bayesian expert model, tuned to the need of a risk analyst, with no test data for the calibration of the experts. The formulas for aggregating the judgements are derived and the means to model expert bias, dispersion, and dependency are described. In section 3, the use of the derived formulas will be demonstrated on real case expert judgements related to a part of a fault-tree model. We also want to demonstrate the effect of different assumptions made with respect to bias, dispersion and dependency as treated in the two expert modelling approaches. Sections 4 and 5 end the paper with some conclusions and a general discussion on the use of expert judgement in the context of a risk assessment.

2. Expert judgement in risk modelling

2.1 Notation

f	true frequency of an accident scenario
f	estimate of f
N	number of losses related to an accident scenario
y	true value of an unknown and unobservable frequency or probability
y	estimate of y
y_i	random variable associated with an expert judgement of y, i = 1,...,n
$\mathbf{y}$	data vector of n expert judgements
ε_i	error term related to expert i's judgement
ε	vector of error terms
Σ	co-variance matrix of expert judgements
$\tilde{y}$	random variable associated with y
$p(\tilde{y})$	analyst's prior distribution
$p(\mathbf{y}\|\tilde{y})$	likelihood of the expert judgements
$p(\tilde{y}\|\mathbf{y})$	analyst's posterior distribution
k_i	weight of expert i

2.2 Basic assumptions

In quantitative risk assessment we want to estimate the risk

$$r(N,f) = E[N]*f \qquad (1)$$

where N is the number of losses and f is the accident frequency. The accident frequency is usually the frequency of a top event of a fault tree (or an aggregated frequency obtained from top events structured by an event tree). Consequence N is conditionally independent of the occurrence rate of the top event. In fault- and event tree modelling, we discretise the outcome range of N into 4 or 5 categories (at most) and connect these to the top event of a fault tree or to the last branching stage of an event tree, where each "failure"-branch can further be modelled by a fault tree.

The single most used generic structure of a fault tree is the one illustrated in Fig. 1. It shows the logic of combining an initiator frequency with a conditional probability.

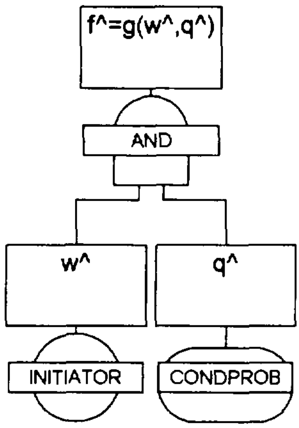

Figure 1. A generic structure of a fault tree with one initiator event and one conditional probability.

Expert judgement in quantifying the fault- and event tree model relates to the two parameters in the structure in Fig. 1: *w* (initiator frequency) and *q* (failure probability or unavailability). In contrast to the procedure of obtaining statistical point estimates for each *w* and *q* directly from data found in databases, expert judgements are modelled by probability distributions depicting a degree-of-belief probability. These distributions are combined in various ways to yield an aggregated distribution from which point estimates $\hat{w}$ and q are inferred (see Fig. 1).

The numerical ranges of the parameters w and q are [0,∞) and [0,1], respectively. The frequency *w* usually relates to some basic initiators (or precursors) such as leakage, fire, collision, etc. The conditional probability *q* usually relates to functional failures (equipment failure or human error) of safety systems and are usually numerically small. Typical ranges are 1.0E-6 to 1.0E-3 (per demand) in safety critical systems. It is, therefore, justifiable to use the *log-normal distribution* as a probability model of the above variables, as the domain of the random variable is $[0, \infty)$ in the case of frequency, and practically all probability mass will lie between $[0,1]$ in the case of conditional probability.

With the top event frequency f, a point estimate $\hat{f}$ is associated. The estimate $\hat{f}$ is basically a function of point estimates $\hat{w}_i$ and $\hat{q}_i$, inferred from the aggregated expert judgement distributions for the initiator event frequency w_i and conditional probability q_i, respectively. Denote this function by $g(\hat{w}_1,\ldots,\hat{w}_s,\hat{q}_1,\ldots,\hat{q}_q)$. The function is determined by the structure of the fault- and event tree model. The most usual point estimate for $\hat{w}$ or $\hat{q}$ is the *expectation* of the aggregated expert judgement $\tilde{w}$ or $\tilde{q}$, that is, $E[\tilde{w}]$ or $E[\tilde{q}]$. Alternatively, when using the log-normal distribution, an obvious point estimate is the *median* of the aggregated expert judgement, that is, $\tilde{w}_{.50}$ or $\tilde{q}_{.50}$.

In the following, we will focus on expert judgements related to the "true" value of the parameter w or q, and will denote this value by y. Specifically, we will associate with y a point estimate $\hat{y}$, which is either the median $\tilde{y}_{.50}$ or the expectation $E[\tilde{y}]$ obtained from the aggregated expert judgement distribution over the aggregated random variable $\tilde{y}$.

Expert judgements, related to the value y, are usually biased, dispersed and dependent. Bias is due to possible optimism or pessimism of the experts, whereas dispersion is related to the spread of the judgements over y. Expert judgements are always dependent, as they draw, more or less, on a common pool of knowledge and experience. In fact, perfect experts (clairvoyants) are completely dependent.

Bias can be accounted for by calibrating an expert according to her/his performance on test data known to the analyst. From such a test, estimates of the expert's dispersion and dependency can also be derived. Such tests, however, are not easily constructed. If such tests are not available, the above-mentioned factors can be dealt with by making subjective assumptions according to the modelling framework chosen. In the following sections we study a classical expert model and a Bayesian expert model. The models are chosen and adjusted to satisfy the needs of the risk analyst, that is, the models are based on the log-normal distribution.

2.2 A classical expert model

The basic model in the classical expert aggregation is

$$P_{\tilde{y}}(\tilde{y}) = \sum_{i=1}^{n} k_i P_{y_i}(y_i) \tag{2}$$

where the weights k_i are to be determined. Following rules have been presented for the weights [2]:

1. Assign all experts equal weight
2. Rank the experts in preference and assign weights proportional to ranks
3. Let the experts weight themselves
4. Use proper scoring rules
5. Calibrate according to test data

To grasp the relevance of the first three rules we should qualify the expert judgements according to bias and correlation. The fourth rule is associated with dishonesty and the avoidance of purposeful bias in judgements. Because dishonesty is not an issue here, as we assume that each expert is motivated to give her/his best judgement, we will not elaborate on the fourth rule further. The fifth rule relies heavily on test data, which can be difficult to obtain in practice. Thus we leave out this rule also.

Expert i is asked to give her/his best judgement about the value y. Denote this judgement by $y_{i_{.50}}$. Expert i is also requested to give her/his conservative judgement about y. The expert should feel somewhat surprised if this judgement proves to be correct. Denote this judgement by $y_{i_{.95}}$. Repeat these questions for all the experts i = 1,...,n.

Specifically, the expert judgements are interpreted as the median and 95%-percentile values of an underlying log-normal distribution depicting a degree-of-belief probability. With this information each expert's distribution is directly defined.

From (2) we can see that the aggregated distribution is a sum of log-normal distributions $P_{y_i}(y_i)$. Thus, the parameters for the aggregated distribution $P_{\tilde{y}}(\tilde{y})$ can be obtained analytically as functions of the expert's judgements $y_{i_{.50}}$ and $y_{i_{.95}}$. If we know how one expert is biased, we only need her/his judgement to estimate the value y, which we would know to coincide with the estimate $\tilde{y}_{.50}$ in average. Weighting is, however, needed if we want to estimate y by the expectation $E[\tilde{y}]$ or would like to control the variance of the estimand.

The classical expert model can be applied with two different estimands of the unknown value y. In the case where y is estimated by the median, the application entails the elicitation of the experts' median estimates $y_{i_{.50}}$, which are then used in computing the aggregated median $\tilde{y}_{.50}$. In the case where y is estimated by the expectation, the application entails the elicitation of the experts' median estimates $y_{i_{.50}}$, the 95%-percentiles $y_{i_{.95}}$, and possible (linear) correlation ρ_{ij}, which are then used in the computation of the aggregated expectation $E[\tilde{y}]$. This is also used in the case where we have some beliefs about the experts with respect to bias and dispersion as described above. The implementation of these beliefs is further described in section 3 within the context of a case study.

2.3 A Bayesian expert model

In the Bayesian framework, we define an explicit expert judgement model, where bias, variance and correlation are defined.

Assume that the experts' judgements follow the multiplicative error model [3].

$$y_i = y * \varepsilon_i \quad i = 1,\ldots,n \tag{7}$$

If we assume again that the log-transformed error terms $\ln \varepsilon_i, \quad i = 1,\ldots,n$, are normally distributed and correlated according to a joint distribution, the interpretation of the error model becomes straightforward. The factor $\ln \varepsilon_i$ reflects the expert's (systematic) bias, as well as precision. For instance, if the parameter value μ_i is greater than zero, the analyst would say that expert i is pessimistic, as the expected value of her/his judgement is greater than the "true" value. The likelihood is related to the analyst's belief $\tilde{y}$ about the true value y.

3. Comparison of expert models

Here, we want to compare the application of the classical and the Bayesian expert model to some real case data [4]. A part of a fault tree is shown in Fig. 2. Three domain experts were used to quantify the initiator event frequency and the conditional probability. The point estimate f is calculated from the aggregated point estimates [illegible] and q obtained from the judgements. Both the median and the expectation are used as point estimates.

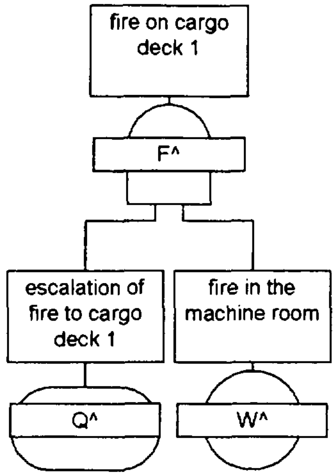

Figure 2. A part of a fault tree model of a real case. Expert judgements elicited for w and q are given in Table 1.

Table 1. Expert judgements given for a part of a fault tree.

Expert i	1	2	3
$w_{.50} / w_{.95}$	10/30	10/50	10/30
$q_{.50} / q_{.95}$	0.02/0.03	0.02/0.05	0.001/0.007

The expert models have been applied to experts who are unbiased/biased and uncorrelated/ correlated. The differences in the point estimates f, as calculated in the different situations, are assessed relative to the classical equal-weight point estimate. The following calculation cases are treated:

A. classical equal-weight (reference)
B. classical unbiased but correlated experts
C. classical biased and correlated experts
D. Bayesian unbiased but correlated experts
E. Bayesian biased and correlated experts

The median and expected value estimates $\bar{w}, \bar{q}$ and $\bar{f} = \bar{w} * q$ are shown in Table 2 for each case. Also the relative change from the equal-weight case (i.e. the reference case) is calculated. The median and expected values in all the five cases are also shown in Fig. 3.

Summary of results					
CASE	w^	q^	f^ = w^ x q^	% wrt *	% wrt **
CASE A median *	10,0000	0,0058	0,0585	0,00%	
CASE A exp.val. **	11,0778	0,0074	0,0815		0,00%
CASE B median	10,0000	0,0058	0,0585	0,00%	
CASE B exp.val.	11,5533	0,0077	0,0893		9,56%
CASE C median	10,0000	0,0033	0,0329	-43,67%	
CASE C exp.val.	11,5533	0,0077	0,0893		9,56%
CASE D median	10,0000	0,0197	0,1972	237,26%	
CASE D exp.val.	10,6192	0,0209	0,2223		172,81%
CASE E median	9,9756	0,0196	0,1953	233,97%	
CASE E exp.val.	10,5933	0,0208	0,2201		170,15%

Table 2. Summary of median and expected values estimates for the w, q and f parameters.

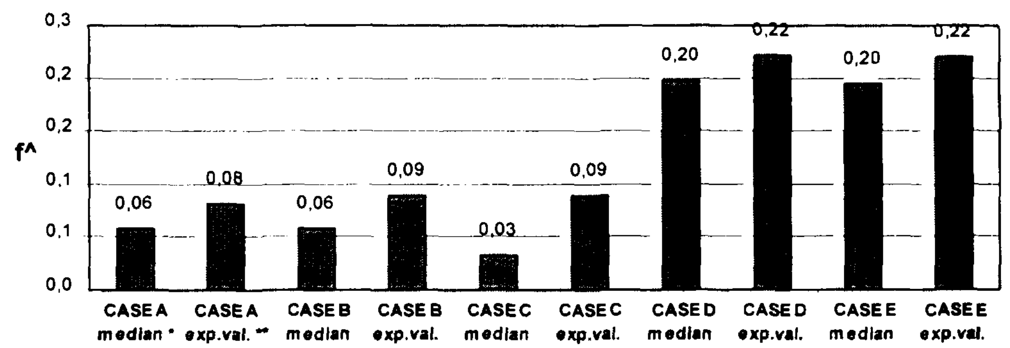

Figure 3. Median and expected value estimates of *f* calculated in the cases defined above.

4. Conclusions

From the calculations performed in section 3 the following can be concluded:

- In the classical expert model, correlation between the experts does not seem to affect the excpected value estimate of *f* significantly according to cases A and

B. Changing the between-expert correlation from the originally applied value of 0.5 to 0.9 affirms this conclusion.

- Using a log-normal distribution might give a significant difference between the median and expected value point estimates for f, as demonstrated in case C. The skewness is a function of correlation and variance, both of which are relatively large in this calculation case.
- The prior distribution affects the posterior distribution significantly, as data, i.e. expert judgements, are scarce. The difference between the the point estimates f obtained from the classical and Bayesian expert models is about one order of magnitude. Changing the prior variance to correspond to "95%-deviations" by a factor of 10 would increase the difference to two orders of magnitude. Such a prior would represent a reasonably un-informative prior.
- The significance of the prior is also shown by comparing cases D and E. Correcting the experts for assumed biases will affect the posterior point estimates for f only in the third decimal.
- Quantification of a fault tree by either a classical or a Bayesian expert model would probably result in significantly different top event point estimates (median or expected value), especially, when the number of qualified experts is small.

In summary, the aim of the study has been to show how to make the use of expert judgements more transparent and to provide better insight on the influences of the methods selected.

5. Discussion

We have described some basic expert judgement models which incorporate bias and dispersion of, and correlation between, a number of experts. These features are quantified by the analyst or decision maker as relative errors on a degree-of-belief basis. Thus, no calibration data is assumed to be available. The quantification requires some experience by the analyst with respect to the performance of the experts. Alternatively, experts could judge the bias and confidence of each other. If calibration data is available, the approach described in [5] should be considered. The crucial question is the assumed relevance of the calibration data collected: "Can the performance of experts, as assessed with respect to some data, be generalised to the problem at hand?".

(f,n)-curves derived from expert judgements along the classical or Bayesian lines are interpreted the same way. The classical approach relies on the assumption that the true parameter value is obtained as the limiting number of occurrences observed under an evolving time period (frequency value) or the relative number of occurrences observed with respect to a number of demands (conditional probability). In the Bayesian approach the posterior distribution centers over these true values as the number of observations increase. Naturally, we have to assume that the events are exchangeable, that is, we can say something about the future based on observations of the past.

In the absence of empirical data, however, we can not tune our estimates according to them. In many risk assessments the situation is precisely this, and the analyst or the decision maker is left with the expert judgements obtained. In addition, the number of experts is usually limited in practice. Which approach, the classical or the Bayesian, is then preferable? If the analyst or decision maker can be considered to be a qualified expert himself, it can be argued that the Bayesian approach is defendable, as the prior information will significantly influence the eventual estimate. If the analyst or decision maker is not an expert, and if no realistic means to define an informative prior can be found, the classical approach should be used. It is important to use informative priors in cases where empirical data is scarce.

As a consequence, every risk assessement should be clearly accompanied by a description of how the expert judgements have been used. Naturally, any risk assessment based on expert judgements can be more disputable than one based on abundant empirical data (observations). It is therefore of outmost importance that *experts consulted in risk assessments are capable of expressing their beliefs in* probabilities (in practice, medians and quantiles). Practical means to minimise the effect of bias related to known heuristics [6] in probability judgements should be developed. Furthermore, how can possible motivational biases, which are related to the experts' personal values be addressed?

References

1. Suokas J., Rouhiainen V. (eds.): Quality Management of Safety and Risk Analysis, Elsevier Science Publishers, 1993.
2. Cooke R.: Experts in Uncertainty, Oxford University Press, New York, 1991.
3. Chhibber S., Apostolakis G., Okrent D.: "A taxonomy of issues related to the use of expert judgements in probabilistic safety studies", Reliability Engineering and System Safety, 1992.
4. Rosqvist T., Tuominen R., Jalonen R., Salmikuukka J., Reunanen M.: Formal Safety Assessment - trial application of FSA to the transportation of dangerous goods on passenger/ro-ro cargo vessels, VTT Manufacturing Technology, Report VALB 353, 1998.
5. Excalibr - software package for expert data evaluation and fusion in risk and reliability assessment, Faculty of Mathematics and Informatics, Delft University of Technology, Delft, 1990.
6. Tversky A., Kahneman D.: "Judgement under uncertainty: heuristics and biases", Science, vol. 185, 1974.

Acknowledgement

The authors want to express their gratitude to M.Sc. Jukka Salmikuukka from VTT Automation for implementing the calculation formulas.