Institute for Energy

Training Material for Formal Expert Judgement

Authors: *K. Simola, A. Mengolini, R. Bolado-Lavin, L. Gandossi*

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

EUR 21770 EN

Mission of the Institute for Energy

The Institute for Energy provides scientific and technical support for the conception, development, implementation and monitoring of community policies related to energy.
Special emphasis is given to the security of energy supply and to sustainable and safe energy production.

European Commission
Directorate-General Joint Research Centre (DG JRC)
Institute for Energy
Petten
The Netherlands

Contact:
Anna Mengolini Tel.: +31 (0) 224 56 5253
E-mail: anna.mengolini@jrc.nl

http://ie.jrc.cec.eu.int/

http://www.jrc.cec.eu.int/

Luxembourg: Office for Official Publications of the European Communities, 2005

EUR 21770 EN

Printed in the Netherlands,(DG JRC, Institute for Energy , PR & Communication)

Cover: R. Houghton, JRC IE, PR & Communication
(No commercial use. Credit "Audiovisual Library European Commission".)

TRAINING MATERIAL FOR FORMAL EXPERT JUDGEMENT

K. Simola, A. Mengolini, L. Gandossi & R. Bolado-Lavin

July 2005

TABLE OF CONTENTS

FOREWORD

This document has been produced at JRC, Institute for Energy, within the frame of the institutional action SAFELIFE - Safety of Ageing Components in Nuclear Power Plants. SAFELIFE provides an integrated approach to R&D activities on critical issues for plant life management on ageing nuclear power installations.

This document provides training material to be used in connection to formal expert judgement, especially in the field of structural integrity. This training material was initially developed for a JRC-IE case study on the use of formal expert judgement in the field of structural integrity. Additional training material focused on important issues identified during the case study will be developed to support the specific needs and characteristics of structural integrity problems. The training material together with the document summarising formal expert judgement [1] should be seen as an information package useful for technical experts involved in formal expert judgement exercises.

1 INTRODUCTION

The process of formal expert judgement usually consists of the following steps:

1) Identification and selection of issues about which the expert judgements should be made.
2) Identification and selection of experts.
3) Training of experts and definition of variables to be elicited.
4) Individual work of experts.
5) Elicitation.
6) Analysis and aggregation of results and, in case of disagreement, attempt to resolve differences.
7) Documentation of results, including expert reasoning in support of their judgement.

Experts may be asked for judgements in different forms, such as *single point estimates, ranking alternatives with paired comparisons, discrete event probabilities*, or as *distributions of continuous or discrete uncertain quantities*. In our case we focus on the last type of expert judgement (continuous uncertain quantities), i.e. elicitation of probability distributions.

Ideally the experts should have a solid background in probability theory and statistics. However, this is often hard to achieve, especially if probability and statistics are not used daily in their work. Even if the experts are familiar with most of the concepts, they may lack the knowledge of subjective interpretation of probability and may not be aware e.g. of the cognitive biases related to judgements. Thus a training session is an important part of the expert elicitation process.

In order to familiarise experts with the process of providing subjective assessments and understanding subjective probability related issues, a <u>training session</u> should be given. There should be a clear definition of the issues on which experts have to make judgements and, as a help, decomposition can be used in case of complex issues. The training of experts is discussed in more detail in Chapter5.

The expert training covers the following issues:

- Familiarising the experts with the expert judgement process and motivating them to provide formal judgements.
- Giving training on concepts and laws of probability, and on expressing judgements formally.
- Informing the experts about possible biases in expert judgement and the application of debiasing techniques.
- Exercises.

2 TRAINING MATERIAL

The training material presented in Appendices 1-3 consists of three parts:

1. An introduction to expert judgements and training on concepts and laws of probability (Appendix 1)
2. Training on heuristics and biases in expert judgements (Appendix 2)
3. Exercises (Appendix 3)

2.1 Introduction to expert judgement and basic concepts of probabilities

Providing formal expert judgements is usually unfamiliar to experts. Further they may worry that their judgements may be misused or misinterpreted. Thus it is very important to familiarise experts with the process. The need and purpose of expert judgements should be made clear, and it should be stressed that there is not only one right answer. The formal expert judgement is rather a tool to summarise the current information, and it identifies where sufficient knowledge exists and where more research is needed.

Since expressing judgements as probabilities is seldom part of daily life of experts, it is useful to explain basic concepts and main properties of probabilities during the training. Use of expert opinions to produce probability distributions to express the uncertainties is based on the concept of subjective probability. Thus it is very important to explain the various concepts or interpretations (e.g. classical, frequentistic and subjective) of probability.

Experts should be trained to some extent to explicitly express their judgements, and this can be helped with practical examples. Most expert judgements can be aided by decomposing the problem (disaggregation), and examples of decomposition can be helpful. Problem decomposition is widely used in scientific studies to simplify a complex problem into components that are more manageable and more easily solved. These less complex assessments are then recombined into a probability distribution for the quantity of interest. Examples of modes of decomposition are event trees, fault trees and functional decompositions. Decomposition may also use physical models of the phenomena. In such case the physical relationship between the quantity of interest and several constituents is expressed through a mathematical function.

Appendix 1 contains the training slides on formal expert judgements and concepts and laws of probability.

2.2 Training on biases and debiasing techniques

Training should also be provided on the heuristics and on the biases they lead to. Training on biases may help individuals to make better probability assessments.

Their knowledge is therefore essential in the elicitation process to avoid systematic errors.

According to the subjectivistic school of probability, the probability of an event is a measure of a person's degree of belief that the event will occur. In most cases experts must synthesize or construct probability values and distributions when an analyst asks for them. In this process of estimating probabilities or determine degree of belief, experts use "rules of thumb", the so-called ***heuristics***.

Heuristics are easy and intuitive ways to deal with uncertainties, but since they are at best only approximate procedures, they can lead to predictable "errors". By "error" we mean a violation of the axioms of probability or an estimate that is not in accord with the expert's beliefs and that the expert would like to correct if the matter was brought to his/her attention. 'Errors' could also be systematic underestimation or over estimation of quantities. These "errors" in the context of expert elicitation are called ***biases***. Because of the existing biases the question of how to minimize biases and systematic errors in elicitation is essential.

Awareness of heuristics and biases may help individuals to make better probability assessments, and thus they should be introduced and discussed during the training for expert judgements.

Appendix 2 contains the training slides on heuristics and biases.

2.3 Exercises

It is common in the expert elicitation process to have exercises for expressing uncertainty with probabilities. These exercises do not have necessarily to be related to the area of expertise of the experts. In the exercises, experts will be asked to give estimates for a set of *seed variables*. Seed variables are variables whose values are known by the normative expert(s), but not by the substantive experts. They are used as feedback to expert, and they can help them to estimate their subjective sense of uncertainty.

In our exercise, experts were asked to provide 5%, 50% and 95% quantiles for the distributions of the seed variables, and afterwards the true values were shown to the experts. In principle, events that are assigned a given probability should occur with a relative frequency equal to that probability. For example, if we have a set of 20 seed variables, for a well-calibrated expert, approximately one out of 20 true values should fall below the estimated 5% quantiles, and one over the 95% quantiles. In 10 cases the true values should be larger than the expert's median and in 10 cases smaller. Comparing the true and the estimated values, the experts can identify whether they tend to be e.g. overconfident, or give systematically too high/low values. If 3 or 4 values fall outside the 90 % bands, it can be interpreted as sampling fluctuations, but if e.g. 10 out of 20 true values are outside the bands, there is reason to suspect that the expert chooses the uncertainty bands too narrowly.

Appendix 3 contains an introduction for the exercises. The questions used in the case study have been excluded. Questions on general knowledge whose values are known can be chosen for the exercises.

3 CONCLUSIONS

This report documents the training material for expert judgement applications in the field of structural integrity. Due to this specific application field, some emphasis is given on lifetime distributions. Otherwise, the training material is of very generic nature and can be used for any application of formal expert judgement.

Based on the experience obtained from the JRC-IE internal case study, we recommend further development of following issues in the training:

- An example of a deterministic and probabilistic structural analysis to highlight the identification and treatment of uncertain parameters.
- Guidance on propagation and management of uncertainties.

REFERENCES

[1] Simola, K., Mengolini, A. & Bolado-Lavin, R. 2005. *Formal expert judgement. An overview*. EUR 21772 EN.

APPENDIX 1

Expert Judgement Training

Introduction to Expert Judgement
Concepts and laws of probability

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

INTRODUCTION TO EXPERT JUDGEMENT

CONCEPTS AND LAWS OF PROBABILITY

Expert Judgement Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Contents

- About expert judgements
- Expert elicitation approaches, NUREG-1150
- Concepts of probability
- Probability distributions
- Expert elicitation
- Heuristics and biases
- Exercises

- Presentation and discussion of the case

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Need for expert judgements

Expert judgements are useful in evaluating uncertainties when

- there is not enough data or good models available

or

- their use is too expensive or time consuming

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

About expert judgements (EJs)

People make judgements daily:

"I think it's going to rain today"
"I do not believe that republicans will win the presidential election"

Advantages of a formal EJ process:

- Explicit
- Systematic
- Uncertainties are described with probabilities
- Attention is paid to possible biases in judgements

=> ***more precise and reliable judgements***

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Parties involved in an EJ process

Decision maker
- uses the outcome of the EJ process

Normative expert(s)
- Experts in decision analysis, probability an statistics, cognitive psychology
- Lead the EJ process

Technical experts (a.k.a. domain / substantive experts)
- Know well the issue
- Analyse the problem, evaluate the values of parameters and their uncertainties

Generalists
- Wide background knowledge in the area of interes
- Clear idea about the targets of the whole project

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Criteria for selection of experts

- Reputation in the field of interest
- Experimental experience in the field of interest
- Number and quality of publications in the field of interest
- Familiarity with uncertainty concepts
- Diversity in background
- Awards
- Balance of views
- Interest in the project
- Availability for the project

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Expert elicitation approaches

History of structured use of expert opinions:
After WW II development of Delphi method & scenario analysis (RAND Corporation)

Most well known methodology in nuclear industry:
"NUREG-1150" Severe accident risks: an assessment for five U.S. NPP (1990)

Benchmark exercise on Expert judgement Techniques (EU Concerted action 1996-1998)
comparison of several approaches

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Phases of NUREG-1150 approach (1/3)

1 Identification and selection of the case
- Goal of elicitation, variables to be quantified

2 Identification and selection of experts

Technical experts
- Represent highest knowledge in the field
- Independent
- Diversity in background

Normative experts
- Knowledge of probability theory, statistics, formal expert elicitation

Expert Judgment Training

Phases of NUREG-1150 approach (2/3)

3 Discussion on the issue

- Refining the definition of variables to be quantified

4 Training of experts

- Concepts of probability
- Heuristics and biases in EJ
- Elicitation, examples

5 Individual work of the experts

- Literature reviews, analyses, simulations

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Phases of NUREG-1150 approach (3/3)

6 Expert elicitation

- Interviews of experts: 1) experts present the rationale behind their assessment in non-quantitative manner 2) Individual interviews for obtaining quantitative assessments

- Documentation and validation of reasoning

7 Aggregation

- Aggregation of experts' estimated
- Review of disagreements
  misjudgements / truly conflicting opinions

8 Documentation and communication

Approach to be used in this case

Same basic steps and principles as in NUREG-1150

Lighter approach

- Less demanding requirements for experts
- Reduced resources for analyses

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Concepts of probability
&
Probability distributions

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Concepts of probability

Classical probability

Logical probability

Frequency interpretations

Propensity interpretations

Subjective probability

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Classical probability

- *The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the* ***measure of this probability****, which is thus* ***simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible****. (Laplace 1814)*

- When are events of the same kind?
- What about probabilities in infinite spaces?
  E.g. Bertrand's paradox

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Frequentistic interpretation of probability

- *The probability of an attribute A in a finite reference classB is the relative frequency of actual occurrences of A within B*
- Called also statistical / empirical / objective probability
- Unrepeatable events?
- Excludes many natural uses of probability:
  "probably I will miss my bus today"
  "it is highly improbable that there is life on Mars"

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Subjective interpretation of probability

- ***Probability is degree of belief***
- Called also Bayesian interpretation of probability
- Probability expresses the observer's uncertainty of the outcome
- Dependent on the observer
- Dependent on the information available
  - Subjective probability is updated with new information

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Subjective interpretation of probability

- A coin has been tossed but we have not yet seen the result. What is the probability of tails?
- What is the probability that the population of Paris was larger than the population of London in 1930?
- The events are no more random
- If correct answers are unknown, the estimates are uncertain

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Axioms and laws of probability

I Probability of an impossible event is 0
Probability of a certain event is 1
The probability for an outcome of event A, P(A), is a number between $0 \leq P(A) \leq 1$

II The probability of the complement of A, “not A”:
P(not A) = 1 – P(A)

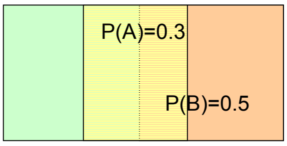
III Intersection of two events
$P(A\&B)=P(A|B)P(B)= P(B|A)P(A)$

P(A|B) is the conditional probability:
probability of A given that B has occurred

Expert Judgment Training

Intersection of two events - example

$P(A\&B)=P(A|B)P(B)= P(B|A)P(A)$



$P(A|B)=1/4$
$P(B)=1/2$
$P(A\&B)=1/4 \times ½ = 1/8$

Axioms and laws of probability

- Probability of the union of two events: $P(A+B)=P(A)+P(B)-P(AB)$
- If the events are mutually exclusive, $P(A+B)=P(A)+P(B)$

In a certain manufacturing process the probability of type I defect is 2%, the probability of a type II defect is 4% and the probability of having both types of defects is 1%.
What is the probability that a randomly chosen product has a defect of any kind (I, II or both)?

P (A or B) = 0,02 + 0,04 – 0,01 = 0,05

- Decomposition of an event:
- $P(A) = P(A|B)P(B) + P(A|\text{not } B)P(\text{not } B)$

Axioms and laws of probability

Bayes' rule

P(A|B) = P(A)xP(B|A)/P(B) [or P(A|B) = P(A&B)/P(B)]

Important e.g. for illustrating a bias called ***base rate fallacy***:

Someone has the symptoms of a disease which takes two forms, both fatal, requiring two different medicines.

Only one medicine can be taken and medicine A does not work for form B of the disease and medicine B does not work for form A of the disease.

Form A of the disease occurs 10% of the time in the population whilst form B occurs 90% of the time. After taking an 80% reliable A/B test it says that this person has form A of the disease. Which medicine to take?

P(A) = 0.1 P(B) = 0.9

$P(T_A|B) = P(T_B|A) = 0.2$ $P(T_A|A) = P(T_B|B) = 0.8$

$P(T_A) = P(T_A|A)P(A) + P(T_A|B)P(B) = 0.8x0.1 + 0.2x0.9 = 0.26$

Probability of having the form A, given the test result T_A:

$P(A|\ T_A) = P(A)xP(T_A\ |A)/P(T_A) = 0.1x0.8\ /\ 0.26 =$ 0.31

while $P(B|\ T_A) =$ 0.69

A person is likely to take the treatment for form A of the disease despite a 20% chance that he could have form B and only 10% of people in the population have form A.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Probability distributions

- **Discrete distributions**
  - e.g. Bernoulli distribution
  - Binomial distribution
  - Poisson distribution

- **Continuous distributions**
  - e.g. Uniform distribution
  - Normal distribution
  - Beta distribution
  - Lifetime distributions: defined in the interval (0,∞)
    - Exponential, Weibull, Gamma, Lognormal,...

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Binomial distribution

Two outcomes of events A & notA: P(A) = p (P(notA)=1-p)

Probability of obtaining exactly *k* times A in *n* independent trials:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad X \sim Bin(n,p)$$

Useful for systems with two possible outcomes of events (failure/no failure) in cases with known, finite number of independent trials, and their ordering does not effect the outcome

Coin tossing
Probability of detection in ultrasonic testing

More than two possible outcomes => multinomial distribution

Expert Judgment Training

Poisson distribution

Poisson distribution treats systems in which randomly occurring phenomena cause irreversible transitions from one state to another

Examples:

- radioactive decay
- number of road accidents within a year in certain crossing
- randomly occurring shocks causing degradation in a structure

$$P(X=k)=\frac{\theta^k}{k!}e^{-\theta}$$

Definitions related to probability distributions

- **Probability density function *f(x)***
  probability that outcome *x* for an experiment occurs within *dx* about *x*

- **Cumulative probability function *F(x)***

$F(x)=P(X\le x)=\int_{x_{min}}^{x} f(u)du$ probability that X occurs between (x_{min}, x)

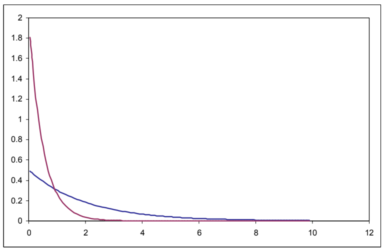
Central concepts in reliability and failure analyses

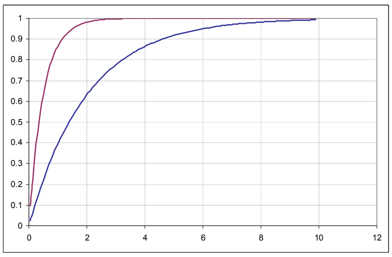
- **Reliability function *R(t)* = 1-*F(t)***
  R(t) = 1-*F(t)* i.e. probability that the item will not fail between (0,t)

- **Hazard rate (conditional failure rate) λ*(t)***
  λ*(t)* = f(t)/R(t) i.e. probability that the item will fail within *dt* given that it has survived to time *t*
  Ageing items have an increasing hazard rate

Exponential distribution

$$f(t) = \lambda e^{-\lambda t}$$





Lifetime distribution of random failures. Constant hazard rate (given the device has survived time T, the probability of failure within Δt is independent of T)

Very commonly used in reliability estimates for active components, such as pumps, valves,...

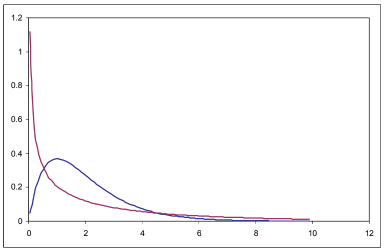
Cannot be used if an ageing phenomenon is present

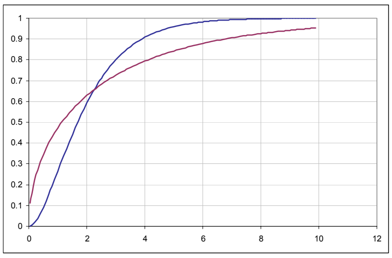
EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Gamma distribution

$$f(t) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} t^{\alpha-1} e^{-t/\beta}$$





Suitable distribution to describe failure time of a system subjected to repetitive random shocks that occur according to a Poisson distribution

Failure probability depends upon how many shocks the device has suffered, i.e. its age.

Gamma function should be used if the mean rate of wear is constant but the rate of wear is subject to random variations

Note: if α=1, we obtain the exponential distribution. α >1 indicates ageing

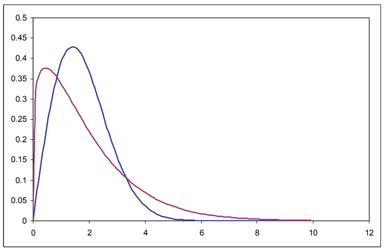
Expert Judgment Training

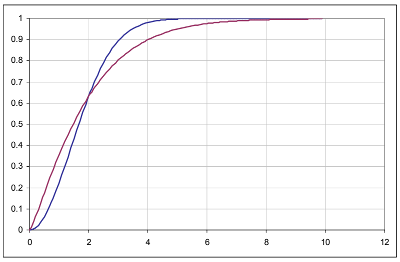
EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Weibull distribution

$$f(t) = \frac{\alpha}{\beta^{\alpha}} t^{\alpha-1} e^{-(x/\beta)^{\alpha}}$$





Appropriate for fitting data for which the conditional probability of failure satisfies a power law as a function of time.

Failure distribution of a device consisting of large number of identical components failing independently according to gamma distribution, and all have to function for device not to fail.

Note: if the shape parameter α=1, we obtain the exponential distribution.

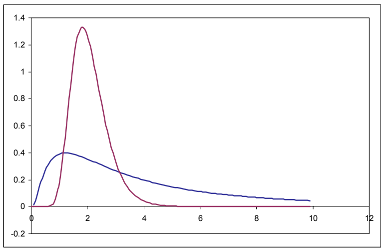
α >1 indicates ageing

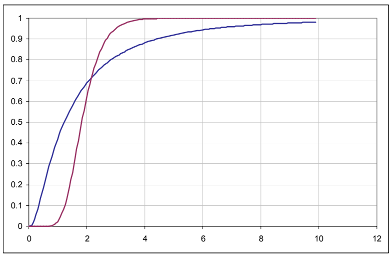
EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Lognormal distribution

$$f(t) = \frac{1}{\sqrt{2\pi}\alpha t} \exp\left[-\frac{[\ln(t/\beta)]^2}{2\alpha^2} \right]$$





Logarithm of failure time t is normally distributed

Suitable distribution when considerable uncertainty in failure parameters

Is typically used if data for rarely occurring events is not extensive, and component failure rates may vary by factors.

Skewness to higher times incorporates the general behaviour of the data for unlikely phenomena: occurrence of infrequent but large values

Expert Judgment Training

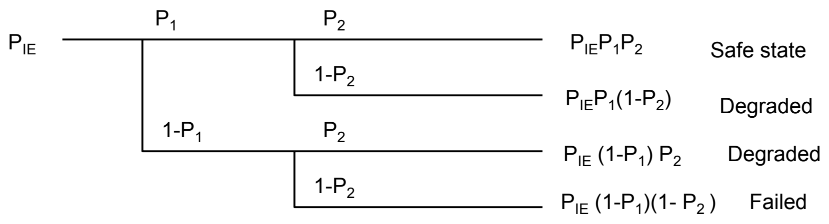
Decomposition of variables

Often the original problem can be decomposed in smaller parts

- more manageable, easier to evaluate
- combined using the rules of probability calculus

=> Better estimate of the variable

Example: event tree



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Elicitation

Median value:

$X>f_{50}$ is as possible as $X<f_{50}$

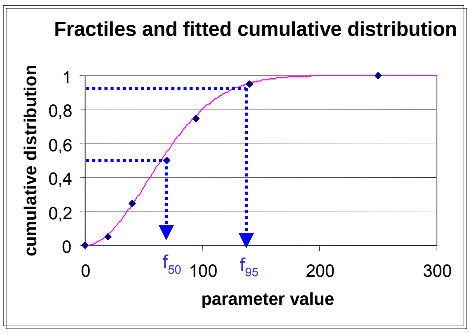
Max and min values for X

Fractiles

f_5, f_{25}, f_{75}, f_{95} ...

E.g. f_5: $P(X \leq f_5) = 0.05$

Expert thinks that the probability that the parameter value is smaller or equal to f_5 is the same as that he would draw the right card among twenty cards

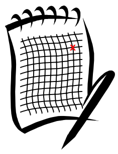


EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

Illustration of probabilities

Balls in an urn: what is the chance of picking up a red ball...



Paper with e.g. a 10 x 10 grid: chance of picking up a certain square = 1/100

Or 10 x 10 cm of millimeter paper: chance of picking up a certain mm^2 = 1/10000 (10^{-4})

Consider 1 km long road.
A probability 10^{-6} corresponds to the chance of selecting a certain 1 mm of the road...

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

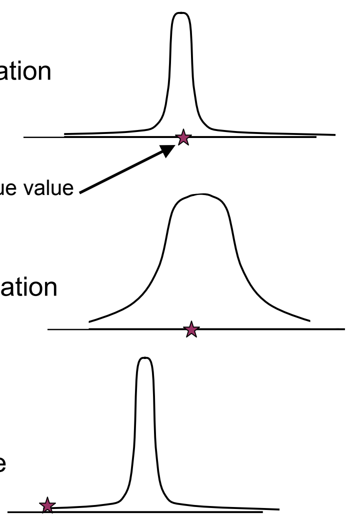
Joint Research Centre

Performance measures

Good calibration and high information

Good calibration but lower information

Poor calibration, over-confidence



APPENDIX 2

Expert Judgement Training

Heuristics and biases in Expert Judgement

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

HEURISTIC AND BIASES IN EXPERT JUDGEMENT

Expert Judgement Training

Source: Kahneman,D. , Slovic P. and Trevesky, A. (1982). *Judgement under uncertainty: Heuristics and biases.*

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

HEURISTICS AND BIASES

Many decisions are based on beliefs concerning the *likelihood of uncertain events.*

These beliefs are usually expressed in statements such as "*I think that…", "chances are ….,* "*it is likely that…*"…

What determines such beliefs? How do people *assess the probability of an uncertain event* or the *value of an uncertain quantity*?

People rely on a limited number of *heuristic principles*

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

HEURISTICS AND BIASES

Heuristics are "*rules of thumb*", easy and intuitive ways to deal with uncertainties

Heuristics are often useful, but may often lead to *severe and systematic errors*

Awareness of heuristics and biases may help individuals to make better probability assessments

Biases: *Cognitive, Structural, Motivational, Background, Affective*

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

BIASES

Cognitive biases	the way we process information, the way we reason linked to rational and experiential level
Structural biases	the way individual the elicitation process is structured
Motivational biases	the way we distort our judgment due to our beliefs and ideology
Background biases	our experience
Affective biases	the way we feel, emotional level
Statistical biases	Lack of capability to deal with statistical concepts

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

HEURISTICS

- Representativeness
- Availability
- Affect heuristic
- Adjustment and anchoring
- Overconfidence
- Control

Joint Research Centre

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

REPRESENTATIVENESS

An Example of Representativeness

John is an athletic young guy who owns a sport car and has a charming blond girlfriend.

Is John a nurse or a professional football player?

We often judge that if X represents well group A,
the probability that X belongs to A is high

Joint Research Centre

Expert Judgment Training

REPRESENTATIVENESS : BIASES

We can further distinguish:

- **Biases due to misconception of chance**
- **Biases due to insensitivity to prior probability**
- **Biases due to insensitivity to sample size**

REPRESENTATIVENESS : BIASES

Misconception of chance (1)

We expect that a sequence of events generated by a random process will represent the essential characteristics of that process even when the sequence is short.

For example:

1) HTHTTH

2) HHHTTT

3) HHHHTH

People believe 1 more likely than 2 and 3

Why???

People tend to expect in the small behaviour that behaviour that they know exists in the large: "*Belief in the law of small numbers*""

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

REPRESENTATIVENESS : BIASES

Insensitivity to prior probability (2)

We often pay too much attention to specific details
while not paying enough attention to base rates

though

Base rates of outcomes should be a major
factor in estimating their frequency

For example

Personality tests administered to 30 engineers and 70 lawyers.

On the basis of this information, thumbnail descriptions of the 30 lawyers and 70 engineers were written.

For each description the probability that the person is an engineer was asked.

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

REPRESENTATIVENESS : BIASES

Insensitivity to prior probability (2)

Example (continue)

Jack is 45 year old. He is married and has four children. He is generally conservative, careful and ambitious. He shows no interest in political and social issues and spends most of his free time on his hobbies which include home carpentry, sailing and mathematical puzzles. **Is he an engineer or a lawyer??**

Judged with high probability to be an engineer

Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues. **Is he an engineer or a lawyer??**

Judged to be at 50% an engineer, forgetting prior probability 30:70

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

REPRESENTATIVENESS : BIASES

Insensitivity to prior probability (2)

Example (continue)

Responses were based on how much the described person was judged to be a lawyer or an engineer

Reversing the proportions - 70 engineers and 30 lawyers - have no effect on estimating the person's profession, given the description.

Why??

Giving worthless evidence causes the subject to ignore the prior probabilities; when no specific evidence is given, prior probabilities are properly utilized

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

REPRESENTATIVENESS : BIASES

Insensitivity to sample size (3)

The size of a sample withdrawn from a population should affect the likelihood of obtaining certain results in it

For example: Large hospitals: 45 babies/day

Small hospital: 15 babies/day

It is known that 50 % of babies are boys

For 1 year, the nº of days with more than 60% boys were recorded

Question: which hospital do you think recorded more days?

Answer : in 95 interviews only 21 gave the right answer

Why??

We tend to ignore sample size and to use only superficial similarity measures

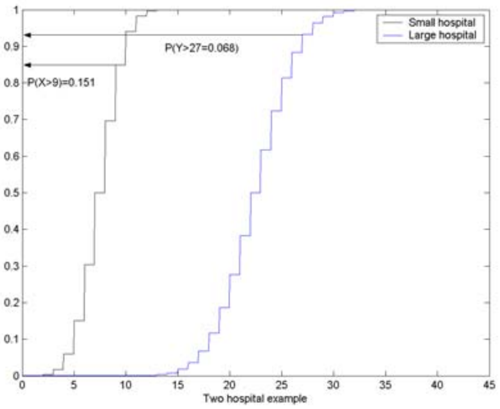
Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

REPRESENTATIVENESS : BIASES

- Larger absolute variations expected in larger samples/populations
- Larger relative variations expected in smaller samples/populations



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

AVAILABILITY

An Example of Availability

Which is the probability to encounter patrolmen on the way to work?

We think about how often we have encountered patrolmen during the daily drive to work in the last ten years

Sometimes people assess the probability of an event by the ease with which instances or occurrences can be brought to mind.

Availability can be a useful heuristic, when the frequency of the available situation is likely to be well correlated to actual frequency

however

Availability is also influenced by other factors, and therefore reliance on such heuristic can lead to biases.

AVAILABILITY: BIASES

We can further distinguish:

- **Biases due to the retrievability of instances**
- **Biases due to ease of imaginability**
- **Perception of risk**

AVAILABILITY: BIASES

Retrievability of instances (1)

When the size of a class is judged by the availability of its instances, a class whose instances are easily retrieved will appear more numerous than a class of equal frequency but whose instances are less easily retrieved.

For example:

In a study, groups of subjects heard a list of well-known personalities of both sexes, all containing the same amount of men and women. They were asked to judged whether the list contained more men or women.

When the list contained names of men that were relatively more famous than the women, the subjects judged that such list contained more men than women (and vice versa).

Why??

Subjects erroneously judge that sex (class) that have the more famous personalities was the more numerous.

AVAILABILITY: BIASES

Ease of imaginability (2)

Example

Suppose you need to assess the frequency of a class whose instances are not stored directly in memory, but that can be generated (imagined) according to some given rule.

For instance, you are asked to consider a group of 10 people who form committee of k members, with $2 \leq k \leq 8$.

•Is it possible to form more committees of 2 people or of 5 people?

•Is it possible to form more committees of 2 people or of 8 people?

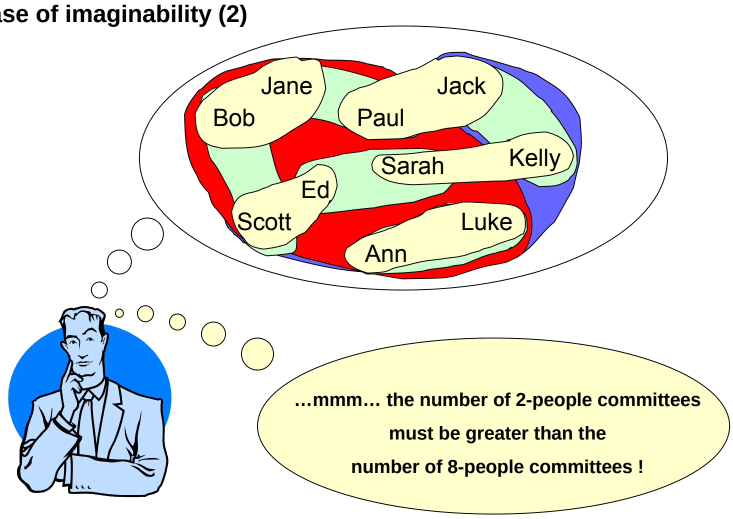
The exact solution is given by the binomial coefficient: $\binom{10}{k}$

This number is maximum for k=5 (252). Moreover, as any choice of k members also define a choice of (10-k) members, the number of k-members committee is equal to the number of (10-k)-members committees!

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

AVAILABILITY: BIASES

Ease of imaginability (2)



AVAILABILITY: BIASES

Ease of imaginability (2)

This was proved in a study.

Subjects estimated a number of committees that was decreasing with increasing committee size k.

For instance, the median estimate for 2-members committees was 70, whereas the estimate for 8-members committees was 20 !

The exact solution is 45 in both cases !!

AVAILABILITY: BIASES

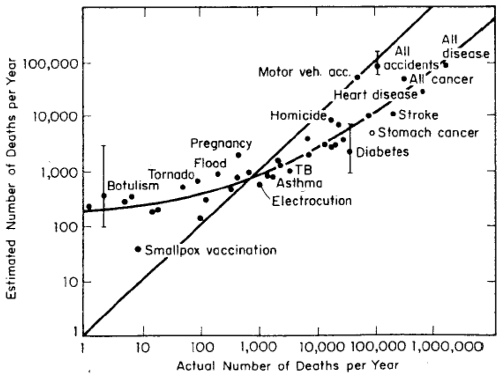
Perception of risk (3)

People asked to estimate the probability of death from various causes

Availability heuristic determine biases in the responses

Overestimation of less frequent/more “publicized” causes; ex: botulism

Underestimation of more frequent/less publicized causes; ex: stroke, cancer



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

AFFECT HEURISTIC

Thought is made of images

Each image gets marked with "***positive***" or "***negative***" mark linked to somatic or body states.

When a negative mark is associated with a future outcome: an alarm is activated

When a positive mark is associated with a future outcome: there is a signal of incentive

The experiential system relies on images

and associations linked by experience to emotion and affect

Biases that have been attributed to ***availability*** heuristic may be due at least in part to ***affect***

Availability may work not only through ease to recall, but also because remembered images come marked with affect.

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

ADJUSTMENT and ANCHORING

In many situations, a natural starting point, or anchor, is selected as first approximation to a value of the quantity to be estimated

The value is then adjusted to reflect supplementary information.

The initial value may be suggested by the formulation of the problem, or may be the result of a partial computation.

In either case, adjustment are typically insufficient and the estimates are biased towards the initial values

Expert Judgment Training

ADJUSTMENT & ANCHORING: BIASES

We can further distinguish:

- **Insufficient adjusting**
- **Biases in the evaluation of conjunctive and disjunctive events**

ADJUSTMENT & ANCHORING: BIASES

Insufficient adjusting (1)

Example 1:

In a study, groups of subjects were asked to estimate the percentage Q of African countries in the United Nations.

A “wheel of fortune” was spun to produce a quantity A, with 1<A<100.

The subjects were led to believe that the result was a random number between 0 and 100, but in fact the wheel yielded either 10 or 65.

After obtaining A, subjects were asked if Q was greater or less than A.

They were then asked to adjust their response up or down from A.

Result

The median percentage Q of African countries in the United Nations was 25 when A was 10 and was 45 when A was 65.

Joint Research Centre

ADJUSTMENT & ANCHORING: BIASES

Insufficient adjusting (1)

Example 2:

In another study, 2 groups of high school pupils were asked to estimate, within 5 seconds, a numerical expression written on the blackboard.

One group estimated the product: **8x7x6x5x4x3x2x1**

The other group estimated the product: **1x2x3x4x5x6x7x8**

Because *adjustments* are typically *insufficient*, the experimenter predicted underestimation in both cases.

Further, because the result of the first steps of the multiplication are higher in the first case than in the second, the first expression was judged larger than the second *(anchoring of result to incomplete computation)*.

Median estimate for ascending sequence = 512

Median estimate for descending sequence = 2250

Correct result: 40320 !

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

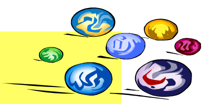
ADJUSTMENT & ANCHORING: BIASES

Evaluation of conjunctive and disjunctive events (2)

Example 1 :

Suppose now you are given the following three events. On which one would you bet 10 € as being the most likely?

1) Drawing a red marble from an urn containing 50 red and 50 blue marbles;
2) Drawing a red marble seven times in a row (with replacement) from an urn containing 90 red and 10 blue marbles;
3) Drawing a red marble at least once in seven successive tries (with replacement) from an urn containing 10 red and 90 blue marbles.



Events of type 1) are **simple** events.

Events of type 2) are called **conjunctive** events

Events of type 2) are called **disjunctive** events

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

ADJUSTMENT & ANCHORING: BIASES

Evaluation of conjunctive and disjunctive events (2)

Example 1:

It was shown in a study that:

When comparing a simple event and a conjunctive event, a significant majority of subjects preferred to bet on (i.e. to consider more likely) the conjunctive event.

At the same time, when comparing a simple event and a disjunctive event, a significant majority of subjects preferred to bet on the simple event.

That is, in this study for a majority of subjects:

P(conjunctive event) > P(simple event) > P(disjunctive event)

In reality:

0.48 < 0.5 < 0.52

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

ADJUSTMENT & ANCHORING: BIASES

Evaluation of conjunctive and disjunctive events (2)

Example 1:

The explanation given to this bias is given in terms of anchoring:

The stated probability of the simple event (**success at any one stage**) provides a natural starting point for the estimation of the probabilities of both conjunctive and disjunctive events.

Since adjustment from the starting point is typically insufficient, the final estimates remain too close to the probability of each elementary event in both cases.

Note the the overall probability of a conjunctive event is **lower** than the probability of each elementary event, while the probability of a disjunctive event is **higher** than the probability of each single elementary event.

As a consequence of anchoring, the overall probability will be overestimated in conjunctive problems and underestimated in disjunctive problems

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

OVERCONFIDENCE AND CALIBRATION

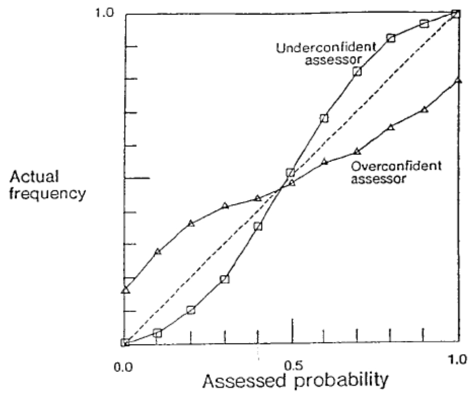
Not directly related with an estimation heuristic, however in some cases anchoring can aggravate it. Overconfidence can express itself as poor calibration.

Calibration curve: assessed probabilities are plotted against actual frequency

Well calibrate judge: *curve near the diagonal*

Under confident judge: assessed probabilities nearer to 0.5 than they should be

Overconfident judge: assessed probabilities too near certainty (0 or 1)



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

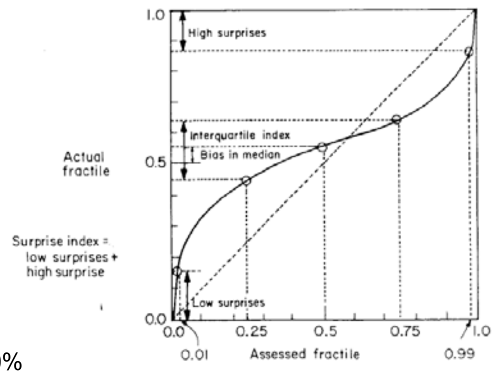
Joint Research Centre

OVERCONFIDENCE AND CALIBRATION

Interquartile range: interval between 25% and 75% quantiles. 50% of true values should fall in this range

Surprise index: percentage of true values that fall below the lowest or above the highest quantiles

Perfectly calibrated experts:
Interquartile index of about 50%
Surprise index of about 2%



EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

LACK OF CAPABILITY TO DEAL WITH STATISTICAL CONCEPTS: BIASES

- Difficulties to distinguish means and medians
  - many experts provide medians when asked for means (it is easier to compute a median than a mean: equal probability vs. integral or sum of a series)
- Difficulties to evaluate measures of spread (standard deviations).
  - Sometimes this problem is linked to the problem of overconfidence
- Difficulties to use Bayes Theorem as the main tool to update information
- Excessive tendency to fit opinions to the normal model
  - Extensive use of this model in probability courses
  - Uncertainty reducing mechanism (symmetry).

Expert Judgment Training

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Joint Research Centre

CONTROL

In certain case we tend to act as if we can influence
the situation over which in reality we have no control at all.

For example two opponents have to cut a deck of cards
subjects are asked to bet on one of the opponents

One opponent was instructed to be shy and insecure
The other was instructed to be confident and self-possessed

Subjects thought to have a better chance of winning against
the insecure opponent, betting a higher amount on money on him

Why???

We think we can control the situation

Expert Judgment Training

APPENDIX 3

Expert Judgement Training

Introduction to Exercises

ASSESSMENT OF UNCERTAIN QUANTITIES

In the following, you will be asked to assess and give your estimates regarding twenty different quantities. These range from familiar to less familiar ones.

For each question, you are asked to give:

- The median value;
- The 5% percentile;
- The 95% percentile;

The **median value** should be what you think is the best estimate you can come up with. In other words, it can be seen as the number you judge to be as likely to be above the true (unknown) value as it is to be below it.

The **5% percentile** is the number that you judge only in 5% of the cases could be exceeded downwards by the true (unknown) value.

The **95% percentile** is the number that you judge in 5% of the cases could be exceeded upwards by the true (unknown) value.

Let us consider an example.

I am asked to assess the following quantity: the number of airplanes that departed from Schiphol Airport yesterday.

I know that Schiphol is pretty busy, having flown from there many times in the past. Last time I was there my flight was late, they did not let me in the Business Suite and I was utterly bored, so I counted the numbers of flights departing between 3 p.m. and 4 p.m. and found out that these were roughly 50.

So now I use this knowledge and I make the assumption that this rate would probably occur continuously between 7.00 a.m. and 10 p.m. That makes 15 hours and 750 flights. So I round it up to 800 flights to consider also the flights that depart in the middle of the night.

I am thus quite happy to say that my median estimate is 800. Now I ask myself: what would the 5% percentile? Well, after all, my estimate was based on a single sample taken on a different day of the week. Maybe there are more flights per hour in general, maybe my assumption to keep the rate constant over 15 hours is quite off the mark. I am pretty sure (and willing to take a bet with odd 5 to 95) that there must have been at least 700 departures from Schiphol yesterday. Also, I am pretty (95%) sure that the number of flights cannot exceed 1000 (I think I have heard once that Frankfurt, which I am pretty sure is bigger than Schiphol, had 1100 departures on the peak day of the season last year).

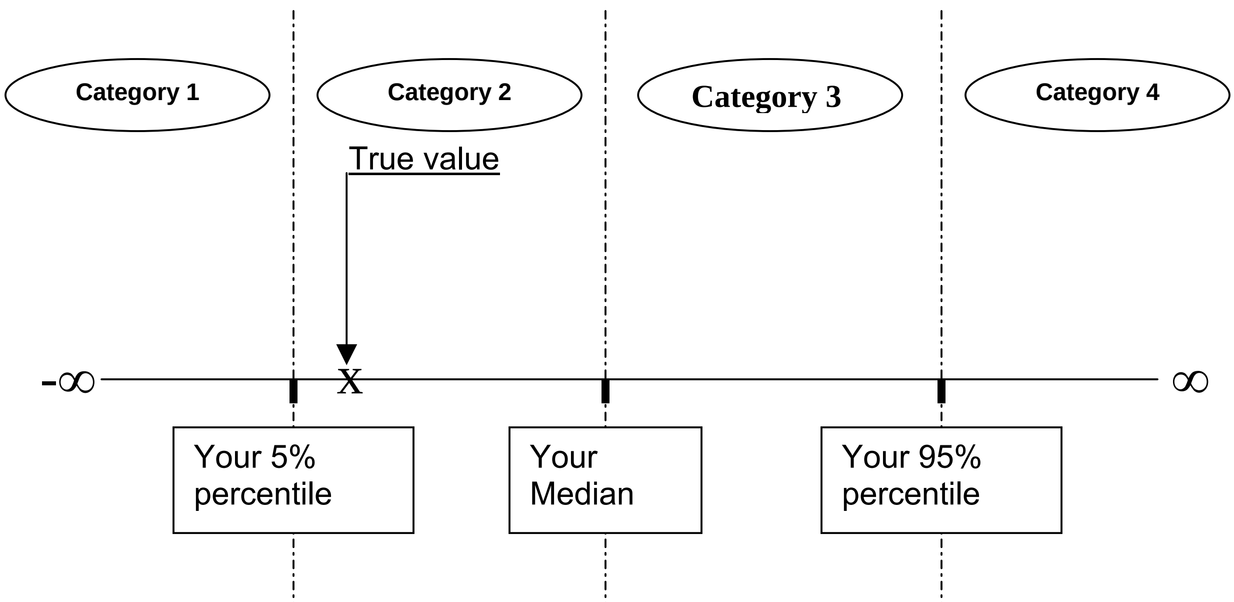
So:

5% percentile = 700
median = 800
95% percentile = 1000

Please note that the distance between the 5% percentile and the median is different from the distance between the median and the 95% percentile. This can be done as a distribution can be skewed to the left or to the right. This is entirely up to your judgement.

At the end of the exercise, we will give the correct answers. Please note that you will be self-evaluating your effort. We will not ask you to make your estimates public. So do not worry too much if some of your estimates are wildly off the mark. This is bound to happen some times. Just remember our training, if you are very uncertain about some quantity, just spread out your 5 and 95% percentiles to take this into account.

At the end, we will only ask you where each true value falls with respect to your percentiles. Namely, we will ask you in which category (defined by your estimates) the true value falls.



European Commission

EUR 21770 EN – DG JRC – Institute for Energy

Training Material for
Formal Expert Judgement

Authors:

K. Simola
A. Mengolini
R. Bolado
L. Gandossi

Abstract

This document provides training material for expert judgement applications in the field of structural integrity. Some emphasis is given on lifetime distributions. However, the training material is of a very generic nature and can be used for any application of formal expert judgement.