



Quantifying uncertainty by probability

This tutorial will cover methods to quantify uncertainty by probability such as Expert Knowledge Elicitation and Bayesian analysis.

We will go through hands on examples on quantifying uncertainty in assessment inputs as well as in assessment output.

All examples will be done using code from the open source program R using the participants own laptops.

It is possible to just sit in and interact without running the code.

Why is “quantification of uncertainty by probability” important

- A call to communicate uncertainty - creates a need to describe uncertainty
- Scientific assessments is not just about data analysis. It includes principles to integrate data, expert knowledge and theory/models
- In 10 years any scientist involved in producing decision support need to be able to describe and communicate uncertainty! [my belief]
- Principles to quantify uncertainty by probability may enrich your methods for research to better account for data quality, model complexity, analysis flexibility and make predictions with uncertainty

Why communicate uncertainty

- Decision making involves uncertainty
 - in facts – what will happen when we make a choice?
 - in values – what do we want when we cannot have everything?
- Unless uncertainty is known
 - a DM can place too much confidence in experts and face unexpected problems
 - or a DM can place too little confidence in experts and miss opportunities and resources to collect information has been wasted

When to communicate uncertainty

- Don't communicate with more or less detail than the DM need
- May require to communicate more things than are usual within a field, things that are assumed or ignored
- May require to communicate less about details academics like to discuss about
 - i.e. both simplifying and complicating normal scientific discourse
- ALSO – reduce to talk about uncertainty in association to decision-relevant elements
- All uncertainty must be uncovered

How to communicate

- Characterise uncertainty
- Assess uncertainty
- Convey uncertainty – create messages that afford DM the detail that their choices warrant
- Persuasive
 - when DM wants to change other DMs behaviour
 - shading or hiding uncertainty might be justified
- Non-persuasive
 - Goal to help people to make decisions that serve their own, self-defined best interests
 - Honesty is the only policy

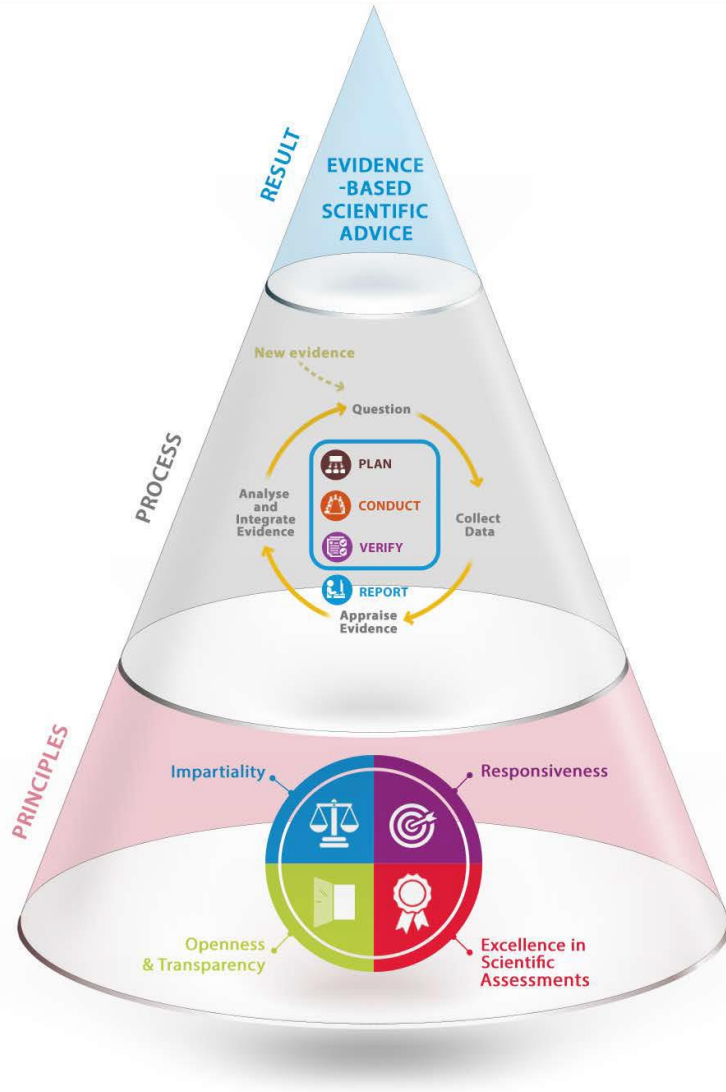
SCIENCE VERSUS SCIENTIFIC ASSESSMENTS

Science	Scientific assessment
Empirical validation	Prognosis without evaluation
Planned studies and experiments	Secondary use of information
Fully reproducible	Typically no repeatability
Publication after success	Assessment after request
Discussion of the results	Uncertainties incomplete
Peer review	Review is not standard
Open debate	Debate not possible

(Ref. Morgan & Henrion: Uncertainty, chap. 3, 1990)

Policy related research is developing standards and guidance for scientific risk assessments





- Collection, appraisal and analysis of evidence – data, literature and expert judgement – are key steps in scientific assessments.
- <https://www.efsa.europa.eu/en/methodology/evidence>



The standards for assessments changes with time

Aim	Type of decision making
Old-old aim	Risk-based decision making
Old aims	Risk-informed decision making or Evidence-based decision making
Current aim	Evidence-based and risk-informed decision making
Emerging aim	Evidence-based decision making robust to uncertainty

Different kinds of uncertainty?



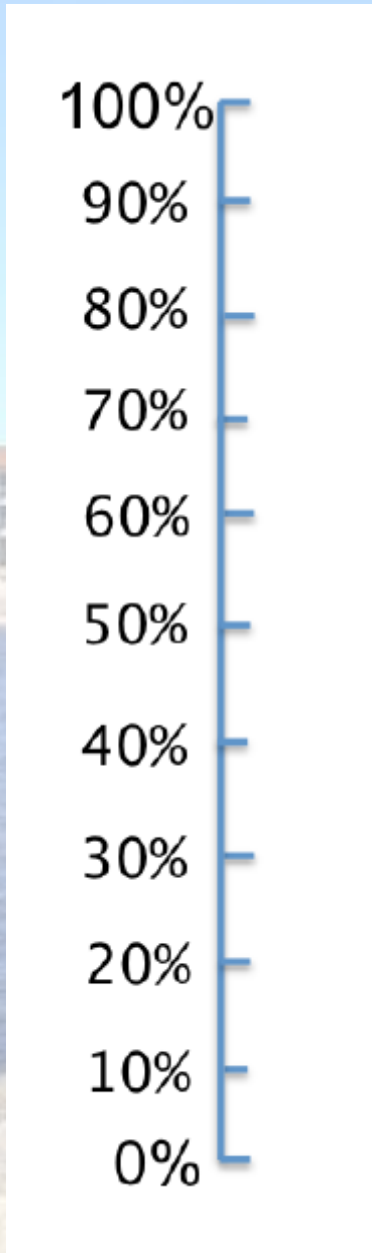
Aleatory and epistemic

Probability can be used as a measure of Relative frequency, Subjective probability and even Confidence although these describe different types of uncertainty.

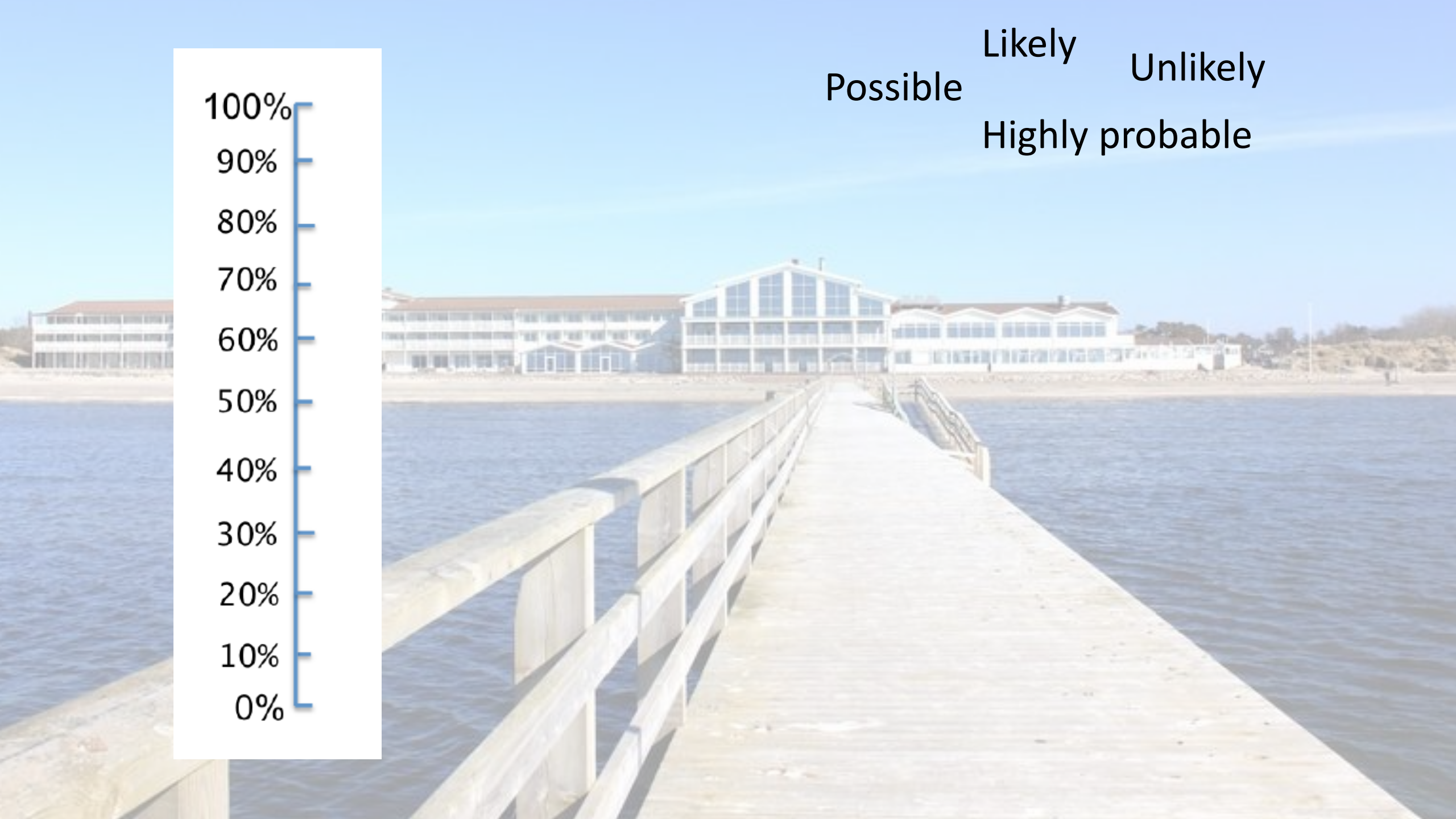


Different ways to describe uncertainty

- Qualitative expressions
- Quantitative expressions
- Combinations



Possible Likely Unlikely
Highly probable



100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

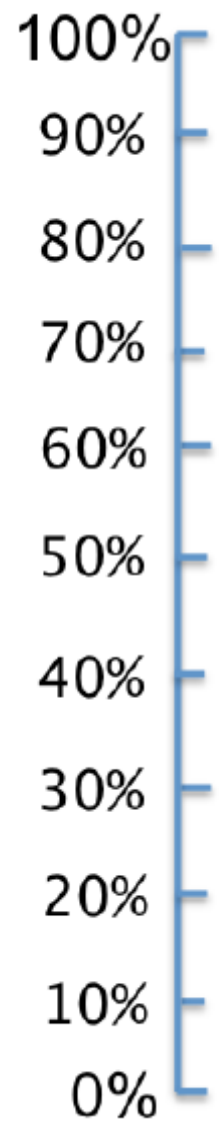
Highly probable

Likely

Possible

Unlikely





Highly probable

Likely

Possible

Unlikely

Qualitative - Quantitative

Hur stor är sannolikheten att du rekommenderar oss till en vän/kollega?

Mycket liten

Mycket stor

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Kommentar: _____

☐ Privatgäst

Epost: _____

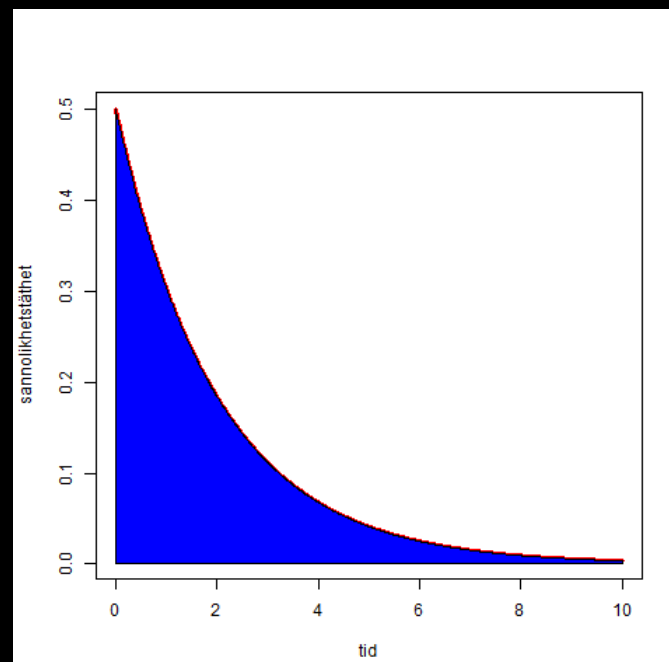
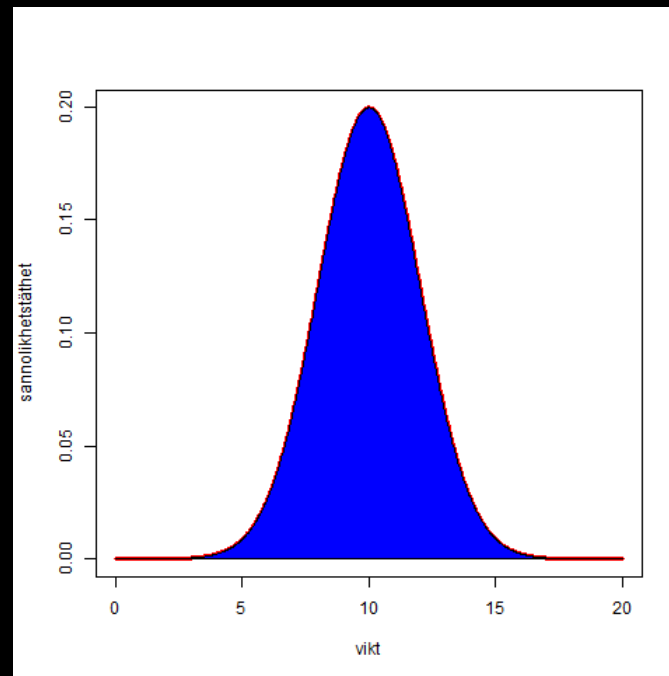
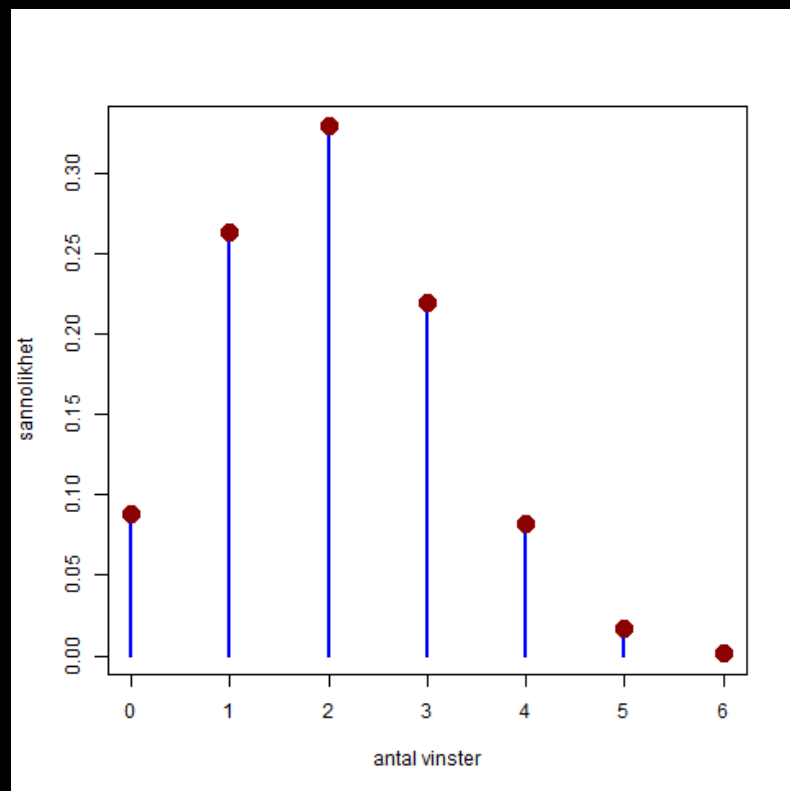
☐ Konferensgäst

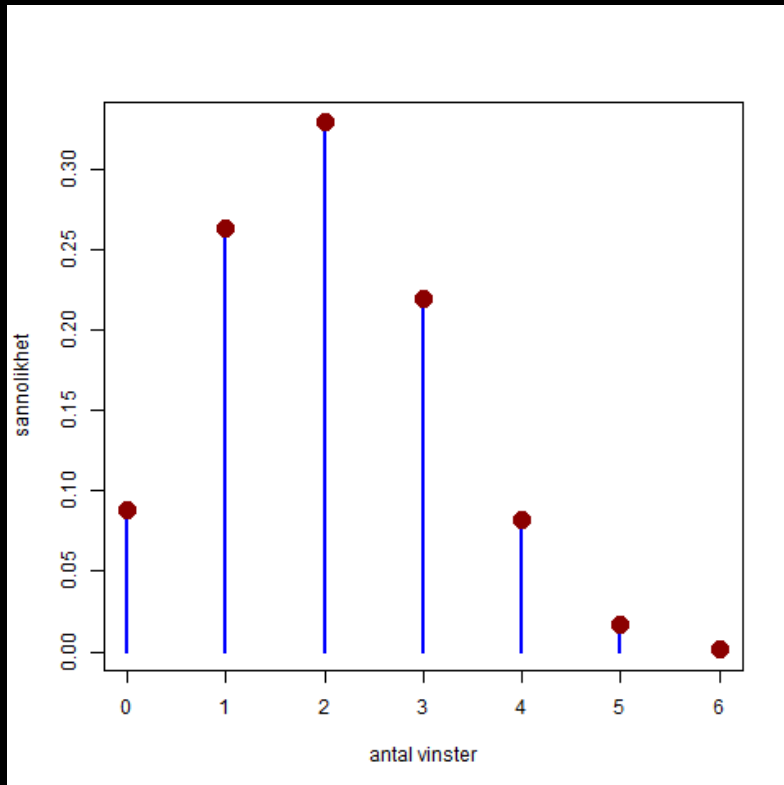
Rumsnummer: _____



0346-71 49 00 | info@strandbaden.se | www.strandbaden.se

Semi-quantitative

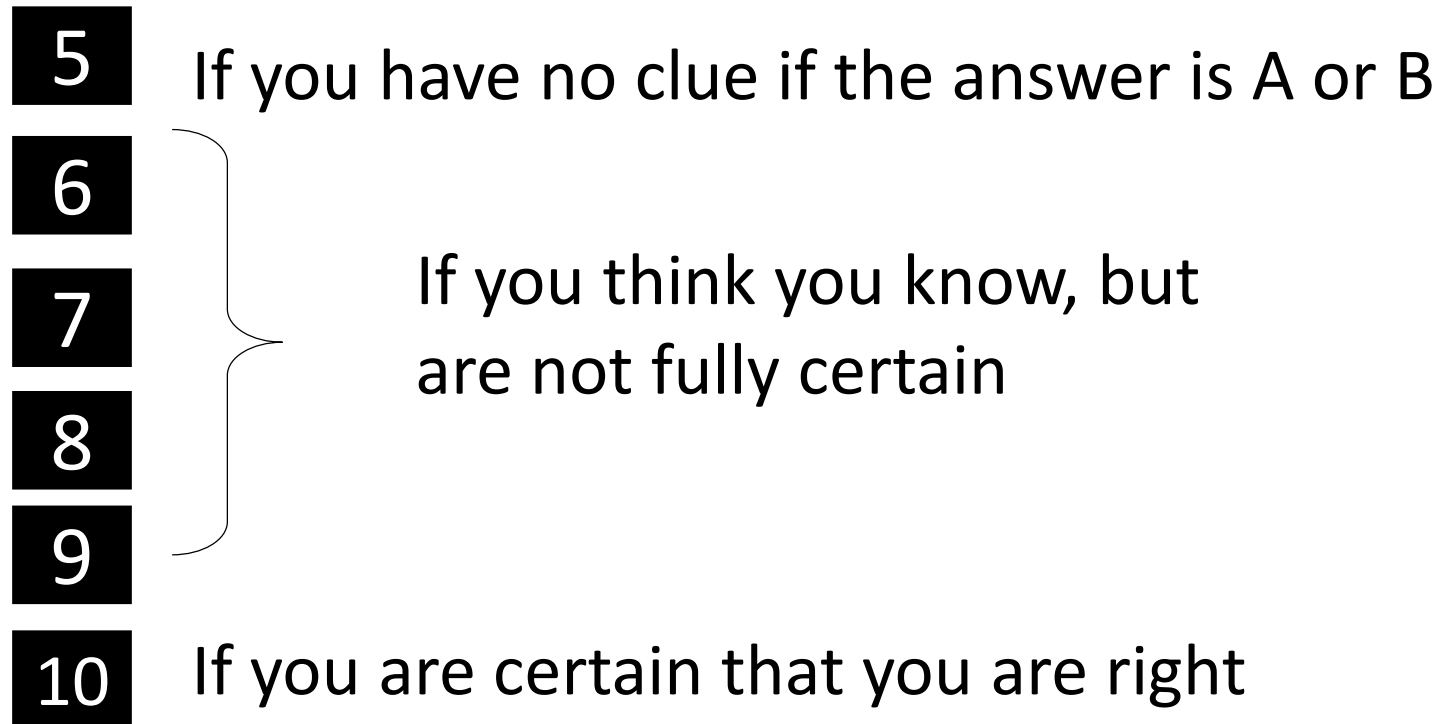




?

Quiz

- You will get questions with two possible answers: A or B
- You will answer and then state how certain you are in this answer



1. Which of these names A) Lars or B) Johan was the most common male name in Sweden during 2016?
2. Are there A) more than 5 or B) less than 5 named Ullrika with two L in Sweden?
3. Accra is the capital of A) Kazakstan or B) Ghana
4. 1 dollar is today worth A) less than 9 SEK or B) more than 9 SEK

1. Which of these names A) **Lars** or B) Johan was the most common male name in Sweden during 2016?
2. Are there A) **more than 5** or B) less than 5 named Ullrika with two L in Sweden?
3. Accra is the capital of A) Kazakhstan or B) **Ghana**
4. 1 dollar is today worth A) less than 9 SEK or B) **more than 9 SEK**

USD to SEK Chart

11 Oct 2017 00:00 UTC - 10 Oct 2018 06:28 UTC **USD/SEK**
close:**9.08583** low:**7.84339** high:**9.21320**



1. Which of these names A) **Lars** or B) Johan was the most common male name in Sweden during 2016?
2. Are there A) **more than 5** or B) less than 5 named Ullrika with two L in Sweden?
3. Accra is the capital of A) Kazakhstan or B) **Ghana**
4. One dollar is right now worth A) < 9 SEK or B) **> 9 SEK**

Your confidence	5	6	7	8	9	10
Score if you were right	0	9	16	21	24	25
Score if you were wrong	0	-11	-24	-39	-56	-75



Följer

#EFSA2018 | No need to be uncertain about communicating on **#Uncertainty** in science: David Spiegelhalter **@d_spiegel**

Översatt tweet



11:49 - 18 sep. 2018

27 Retweeter 38 gilla-markeringar



Spiegelhalter's tweet

https://twitter.com/EFSA_EU/status/1042123576713900033

There is no single 'true' uncertainty

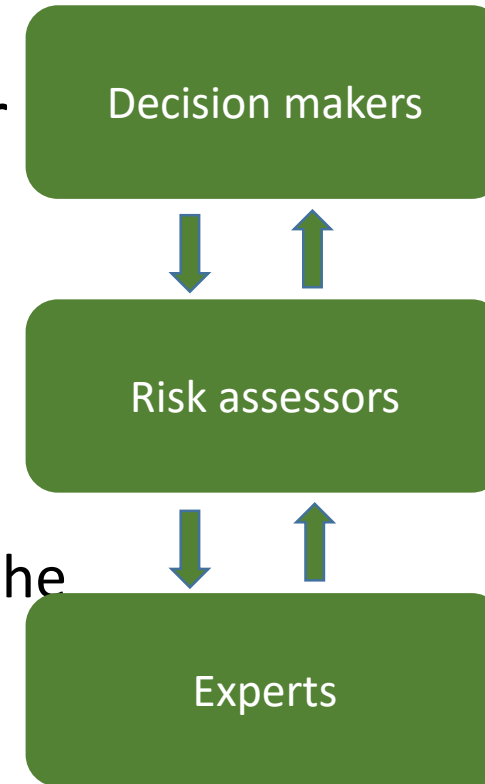
It is my, yours or our uncertainty

Reclaim uncertainty

Don't apologise for being uncertain

Who is responsible for addressing uncertainty?

- Decision-makers are responsible for resolving the impact of uncertainty on decisions – requires weighing the scientific assessment against other considerations (EFSA)
- The decision maker's asks:
 - What is the range of possible answers, and how probable are they?
 - Is further investigation needed?
 - What are the main sources of uncertainty affecting the outcome of the assessment? (helps communication with stakeholders)
- Assessors are responsible for characterizing uncertainty – requires scientific expertise (EFSA)

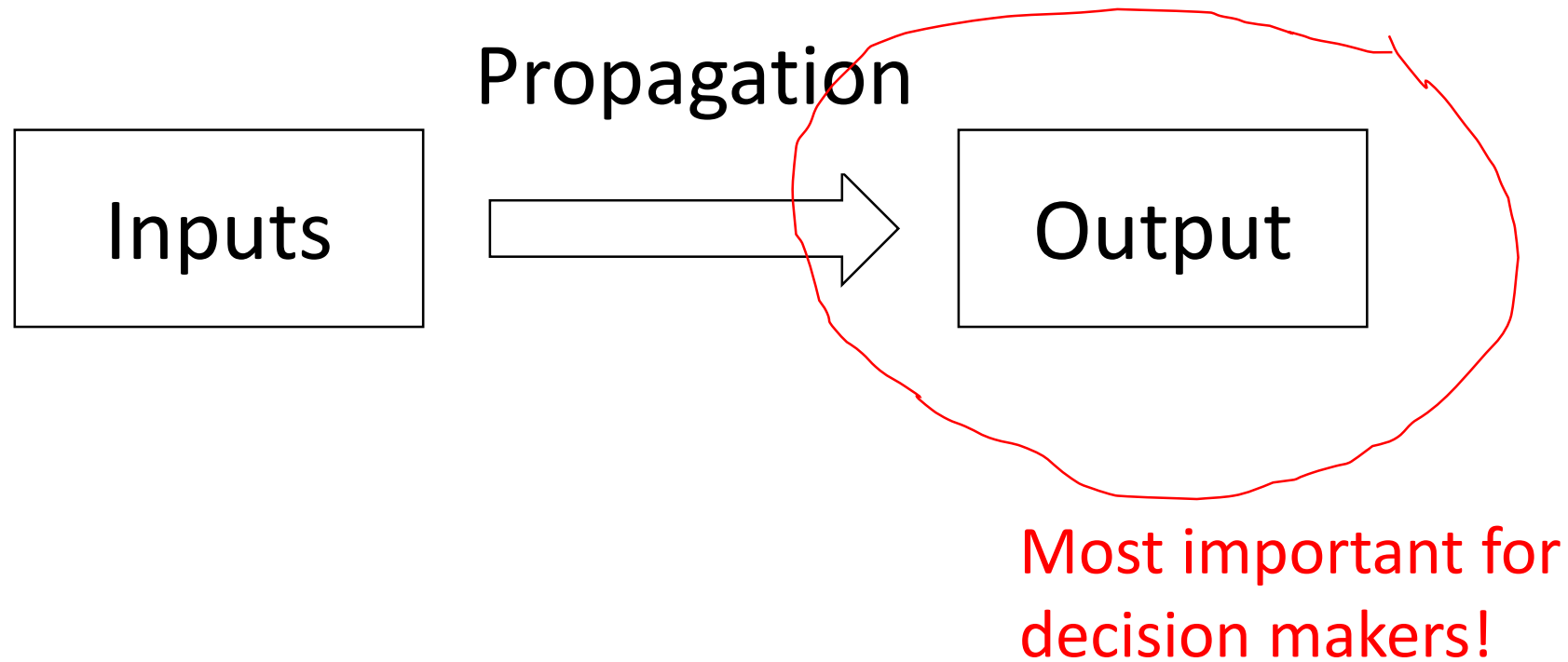


To the assessors

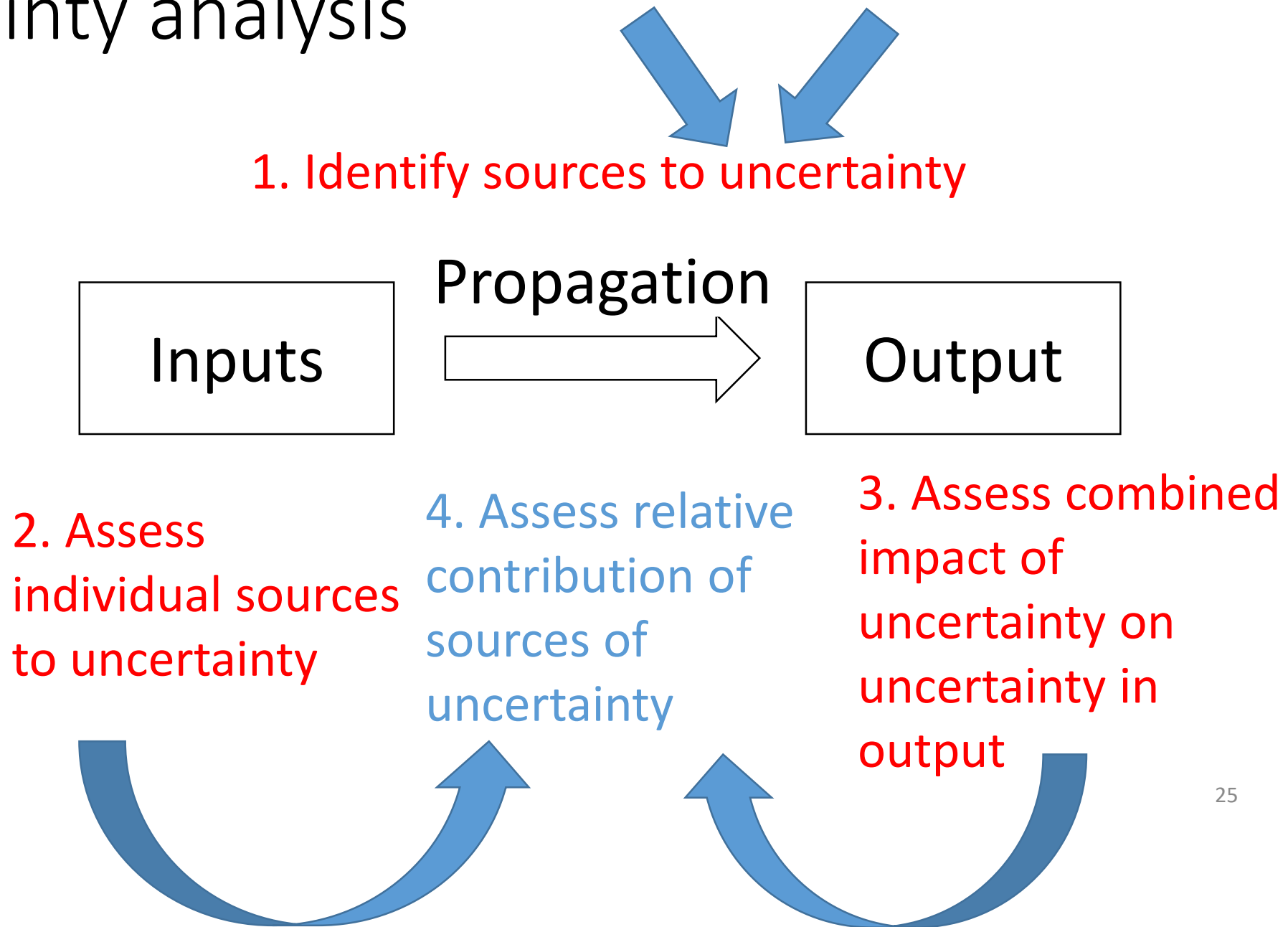
- **Combined uncertainty** should be **quantified**
 - By calculation or expert judgment
- It is **not necessary** to characterise all uncertainties individually
- It is **necessary** to characterise the combined uncertainty
 - This is what matters for decision-making
- Guidance recognises assessors may be unable to quantify some uncertainties
 - Identify and describe these

No need to quantify all sources to
uncertainty

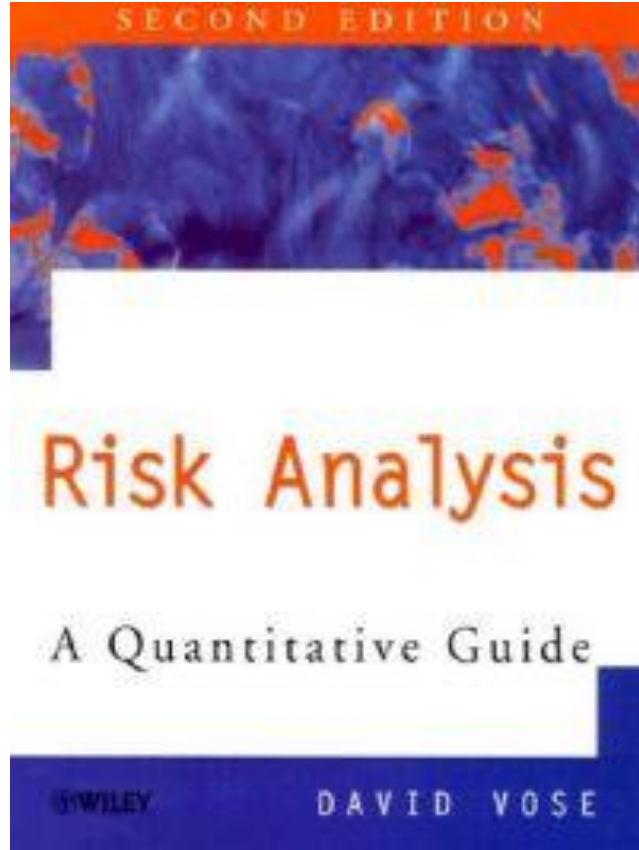
Uncertainty analysis



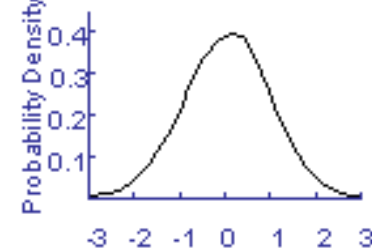
Uncertainty analysis



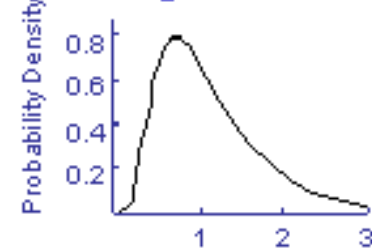
Probabilistic analysis of uncertainty



Stressing Variables

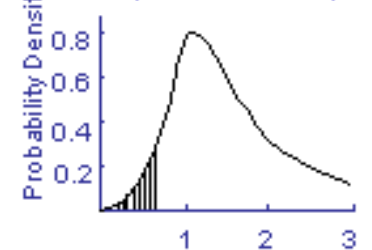


Strength Variables



Deterministic
Model

P(Failure)



Is output uncertainty aleatory or epistemic uncertainty?

Well, if the model produces an output which is the failure probability, the model express uncertainty in this probability.

If the model produce a quantity for which the system fails if it go below a certain threshold, there is only aleatory uncertainty.

Aleatory and epistemic uncertainty should be separated in the analysis.

Risk and uncertainty analysis

- Purpose: To evaluate the impact of uncertainty on objectives (ISO, EFSA)
- Different methods to identify and describe uncertainty
- Advantages with quantitative methods - objective, able to cover both inputs and outputs of an assessment
- Advantages with qualitative methods - all things shouldn't be quantified

Methods for uncertainty analysis

- Descriptive expression
- Ordinal scales
- Matrices
- Conservative assumptions
- Interval Analysis
- Probability Bounds Analysis
- Expert knowledge elicitation
- Monte Carlo simulation (1D, 2D)
- Confidence Intervals
- The Bootstrap
- Bayesian inference (i.e. full probability distributions which can give probability intervals)
- Bayesian modelling
- Sensitivity analysis (one at time, global SA)
- NUSAP (qualitative sensitivity analysis on a complex model)

Methods for uncertainty analysis

Uses probability to quantify epistemic uncertainty

- Descriptive expression
- Ordinal scales
- Matrices
- Conservative assumptions
- Interval Analysis
- Probability Bounds Analysis
- Expert knowledge elicitation
- Monte Carlo simulation (1D, 2D)
- Confidence Intervals
- The Bootstrap
- Bayesian inference (i.e. full probability distributions which can give probability intervals)
- Bayesian modelling
- Sensitivity analysis (one at time, global SA)
- NUSAP (qualitative sensitivity analysis on a complex model)

Uses bounds on probability to quantify epistemic uncertainty

Two methods to quantify uncertainty

- Expert Knowledge Elicitation
 - Extract probability distributions from experts
- Bayesian Analysis
 - Use probability distributions to describe uncertainty
 - Able to learn from data
 - Revise the probability distributions integrating expert knowledge and data

Bayesian analysis

Bayesian data analysis

Bayesian inference

Bayesian updating

Bayesian learning

Bayesian statistics

Bayesian networks

Bayesian modelling

Bayesian emulation

Expert judgement

To put it simple - any use of Bayesian probabilities for:

- Probabilistic causal modelling

or to


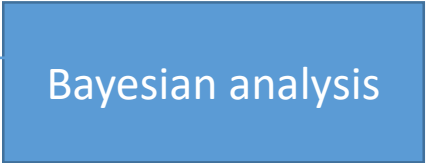
- Quantification of uncertainty

Focus of today

Inference = The act of reasoning from factual knowledge or evidence

- Predictive inference
- Probabilistic calculations
- Forward sampling
- Monte Carlo simulation
- Parametric inference
- Bayesian updating
- Backward sampling
- Markov Chain Monte Carlo simulation

I will draw these things on the board

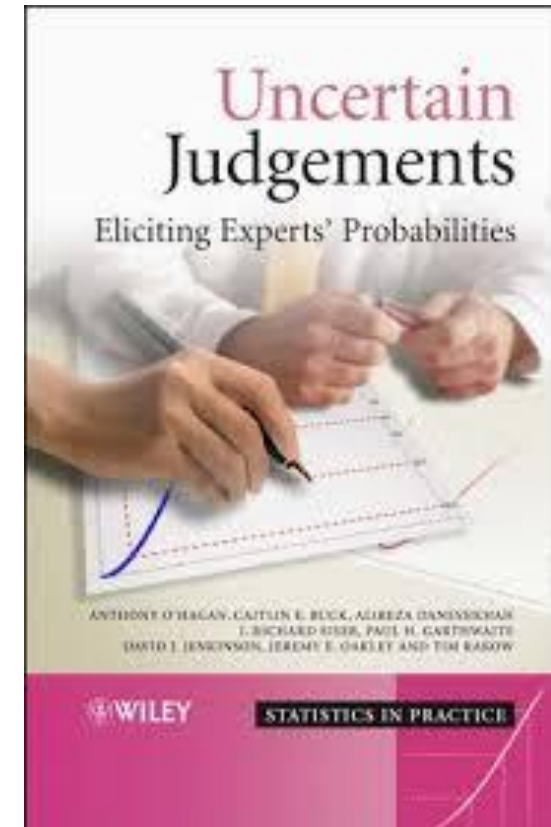
- A is the event of interest
- We want to know $P(A)$
- In general we always condition on what we know $P(A | \text{Knowledge})$
- We can formalise this into a model
- $P(A | \theta)$ We simplify the world into a set of parameters θ
- If uncertainty is important we should also consider $P(\theta)$ "prior" 
- We are then interested in $P(A | \theta)P(\theta)$
- We want to incorporate data
- In order to learn from data we need to specify a model linking data to the parameters of the model $P(\text{data} | \theta)$ "likelihood"
- How to get $P(\theta | \text{data})$? "Bayesian updating" 
- $P(\theta | \text{data})$ "parametric inference"
- $P(A | \text{data}) = P(A | \theta) P(\theta | \text{data})$ "predictive inference"

Now you are going to be experts in a session

- <https://goo.gl/forms/UFoXTLgcAOuRZGX23>
- Answer the questions as honestly as possible without looking at the answer.

Expert's Knowledge Elicitation

- Aim: to describe the Expert's Knowledge about one or more uncertain quantities in probabilistic form
- i.e. a joint probability distribution for the random variable in question
- EKE can be used to build priors distributions or prior predictive distributions





An Expert Knowledge Elicitation

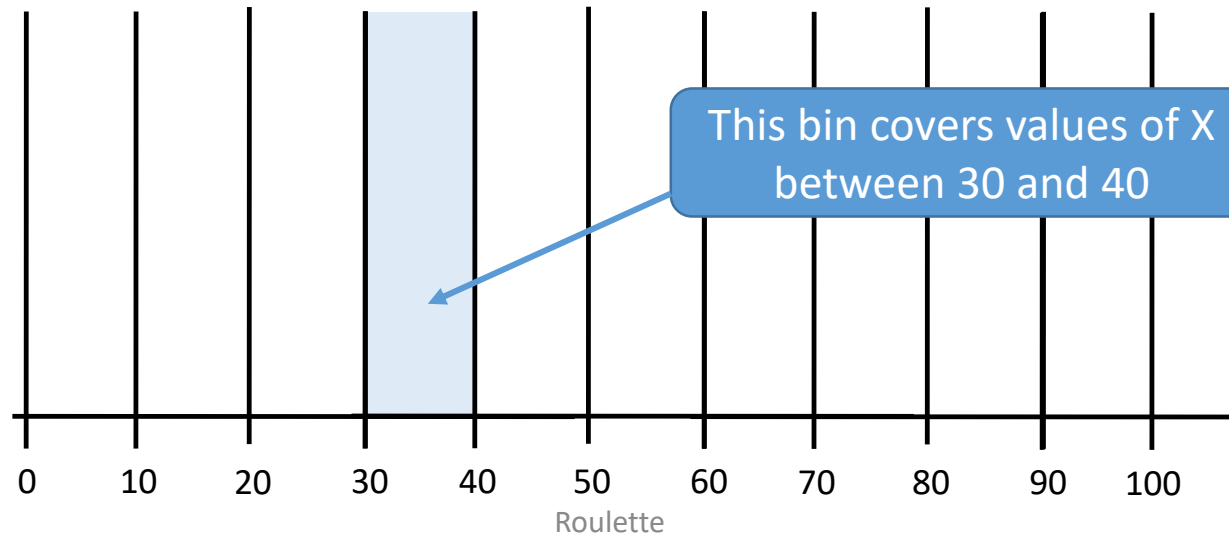
- Formulate the elicitation questions
- Ask experts about
 - Probabilities
 - Quantiles
 - Probability intervals
 - Moments or other descriptions of a probability distribution
- Fit and aggregate into a probability distribution for the uncertain quantity

Direct methods for EKE

- Simple and a bit crude
 - *Intervals* – Lower and Upper limits, then a Uniform distribution
 - *Triangular distributions* – Mode, Lower and Upper limits
- Cumulative Density Function (CDF)
 - *Quartiles* – 4 intervals, median and 25th and 75th percentiles
 - *Tertiles* – 3 intervals with equal probability
 - *Probabilities/Hybrid* – Choose probabilities and intervals
- Probability Density Function (PDF)
 - Mode/Mean, percentiles, shape,...
 - Place chips, draw it by hand...

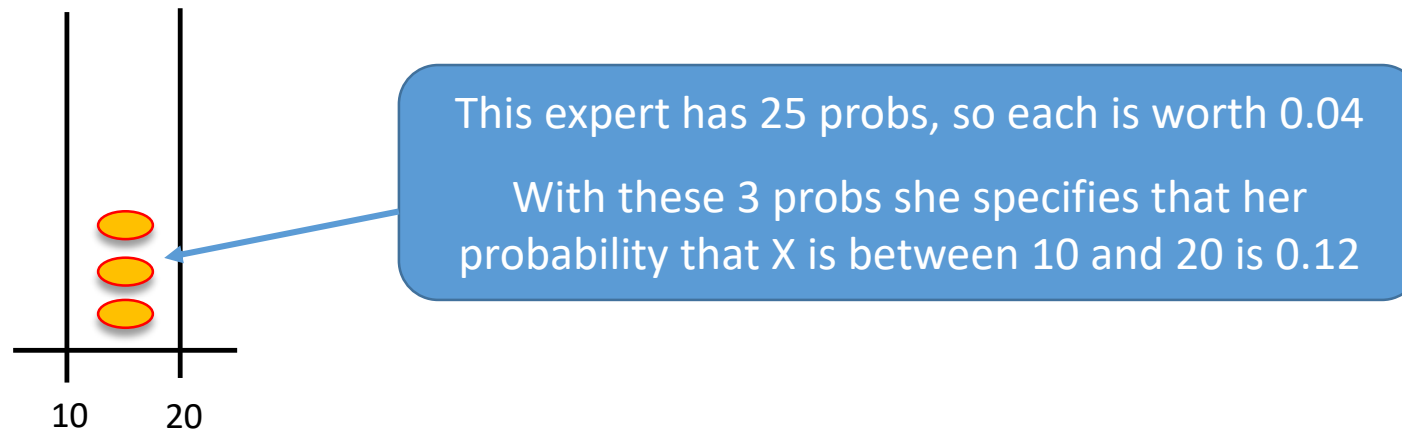
The roulette method – eliciting a pdf

- You have a grid comprising a number of columns
 - These represent ranges of possible values of X
 - Called *bins*
- You have labelled the bin boundaries
 - Showing the range of values of X covered by each bin

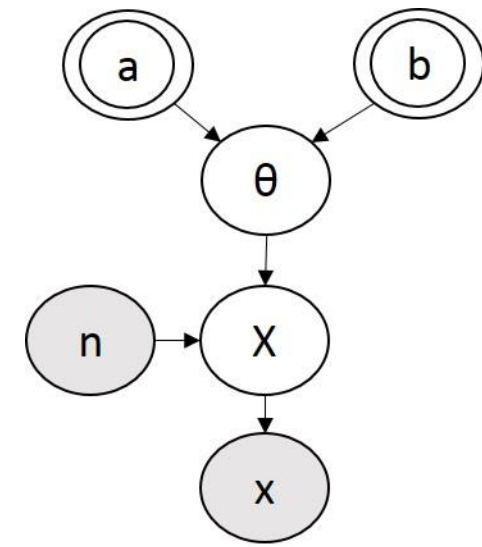


The roulette method – eliciting a pdf

- You also have a number of counters called *probs*
 - Because each one represents an amount of probability
 - For instance, if you have 20 probs, each represents a probability of 0.05 (5%)
- You are asked to place the probs in the bins on your grid, to specify your knowledge and beliefs about X



Indirect methods for EKE



- Equivalent Prior Sample (EPS)

- *What is the expected frequency of the event?*
- *What is the size a sample that you imagine to have behind this estimate?*

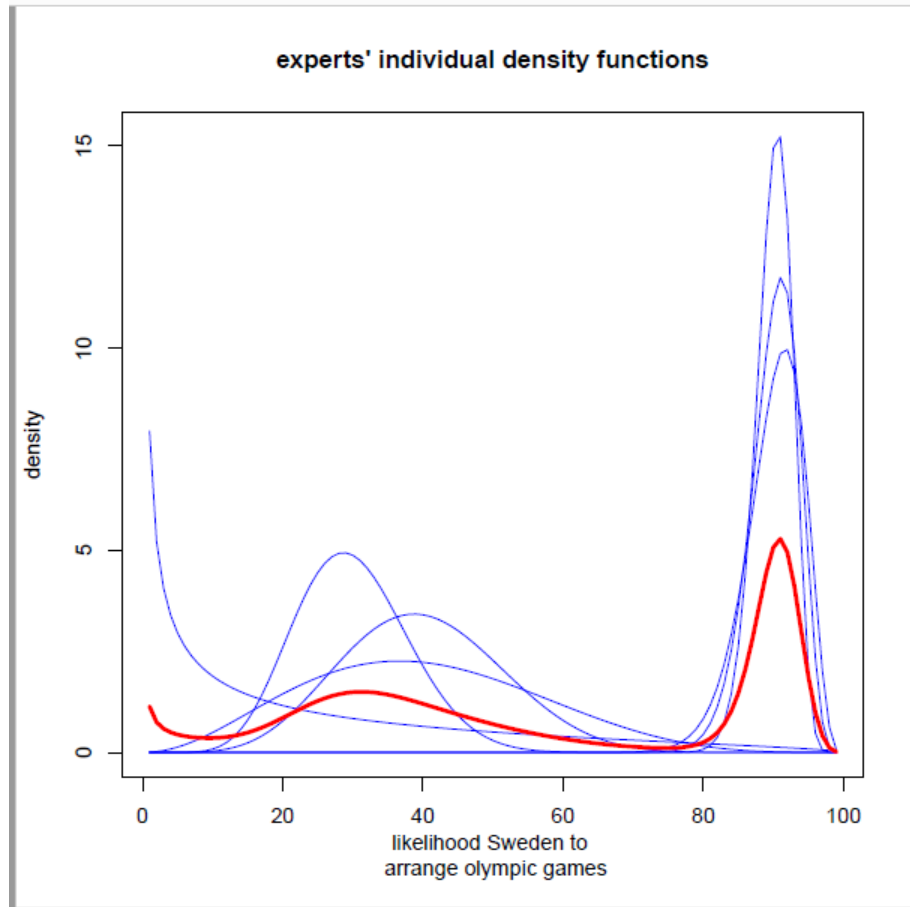
$$\frac{x}{n} = ? \quad n = ?$$

- Hypothetical Future Sample (HFS)

- *In a future sample of size 100 – in how many times has the event occurred?*

$$n = 100 \quad x = ?$$

I used an indirect method to elicit your uncertainty in Sweden arranging OS 2026



Sverige är ett steg närmare att få arrangera vinter-OS 2026.

We want the expert's uncertainty!

<i>Wrong</i>	<i>Right</i>
Your judgement of the probability that ...	Your probability that ...
Write down the median for ...	Write down your median value for ...
The probability of the quantity being below the median is 0.5	Your personal probability of the quantity being below your median value is 0.5
There should be a 50% chance that the quantity lies between the lower and upper quartiles	You should give 50% probability to the quantity lying between your lower and upper quartiles

- Never refer to a relative frequency or a proportion as a probability.

Selection of Structured EKE Software

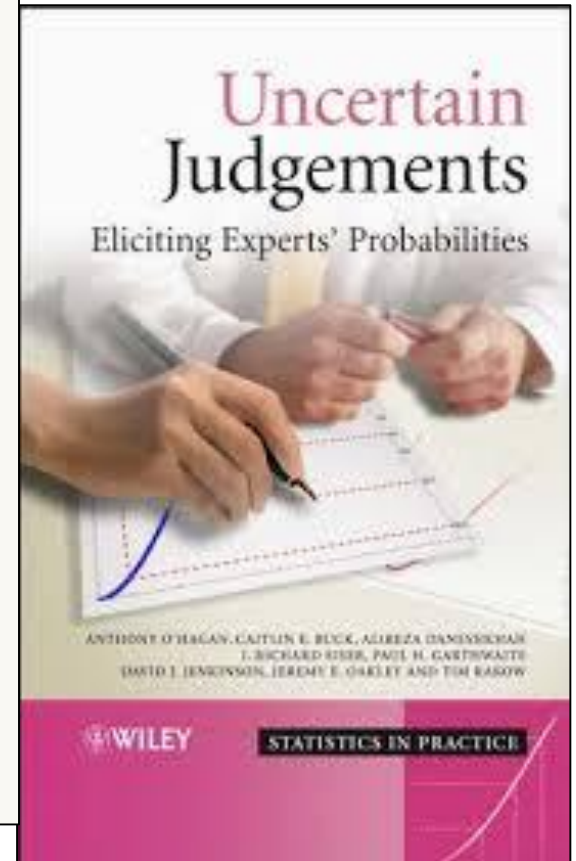
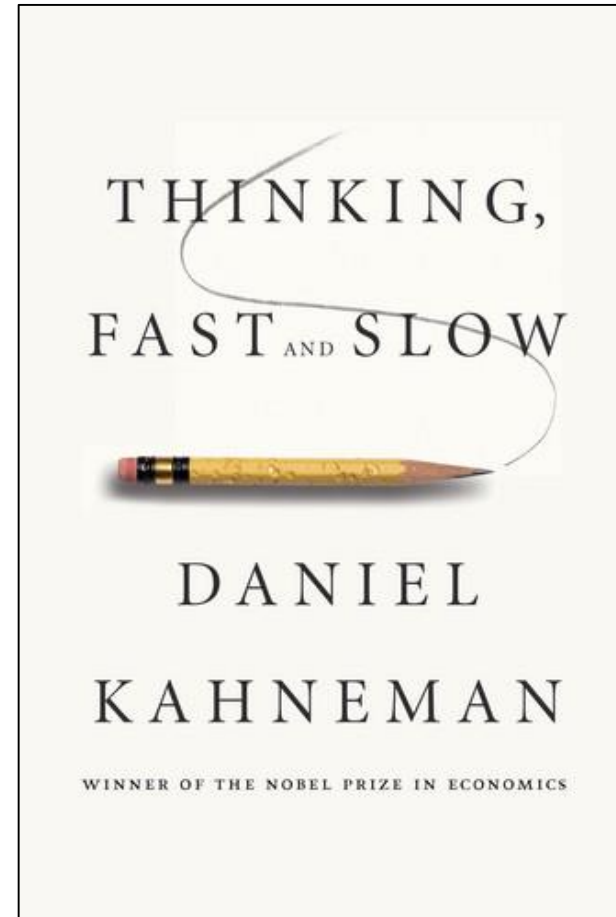
- EXCALIBUR (EXpert CALIBration): www.lighttwist.net/wp/excalibur
- ElicitN: www.downloadcollection.com/elicitn.htm
- SHELF (The SHEffield ELicitation Framework): www.tonyohagan.co.uk/shelf/
- MATCH Uncertainty Elicitation
Tool: optics.eee.nottingham.ac.uk/match/uncertainty.php#
- UncertWeb - The Elicitor: <http://elicitator.uncertweb.org/>
- Variogram elicitation: www.variogramelicitation.org
- Unicorn: www.lighttwist.net/wp/unicorn-download

Expert Knowledge Elicitation

- O' Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. E., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). Uncertain Judgements: Eliciting Expert Probabilities. Chichester: Wiley
- <https://www.efsa.europa.eu/en/efsajournal/pub/3734>
- SHELF package: **Tools to Support the Sheffield Elicitation Framework**
- <http://optics.eee.nottingham.ac.uk/match/uncertainty.php>
- David E. Morris, Jeremy E. Oakley, John A. Crowe, A web-based tool for eliciting probability distributions from experts, Environmental Modelling & Software, Volume 52, February 2014, Pages 1-4, ISSN 1364-8152, <http://dx.doi.org/10.1016/j.envsoft.2013.10.010>.

Pshycological factors and elicitation

- Anchoring and adjustment
- Availability
- Range–frequency compromise
- Representativeness and baseline neglect
- Conjunction fallacy
- The law of small numbers
- Overconfidence

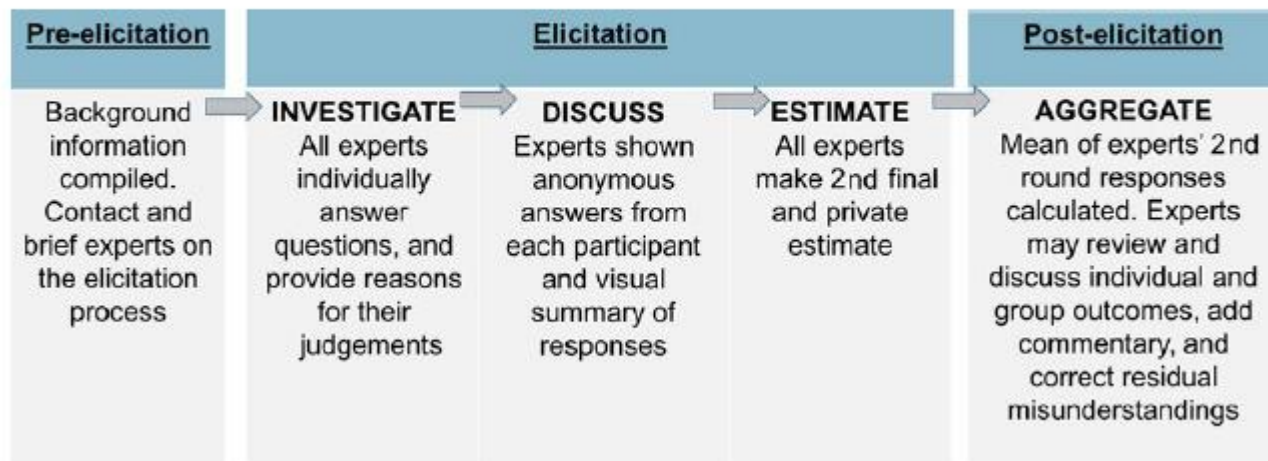


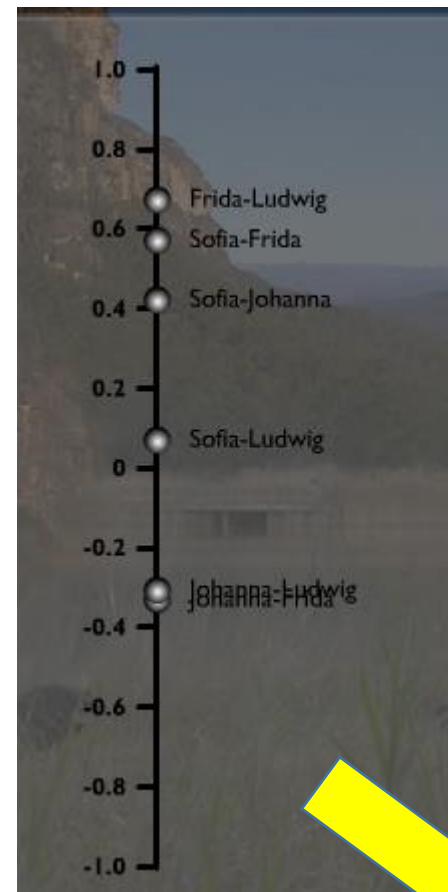
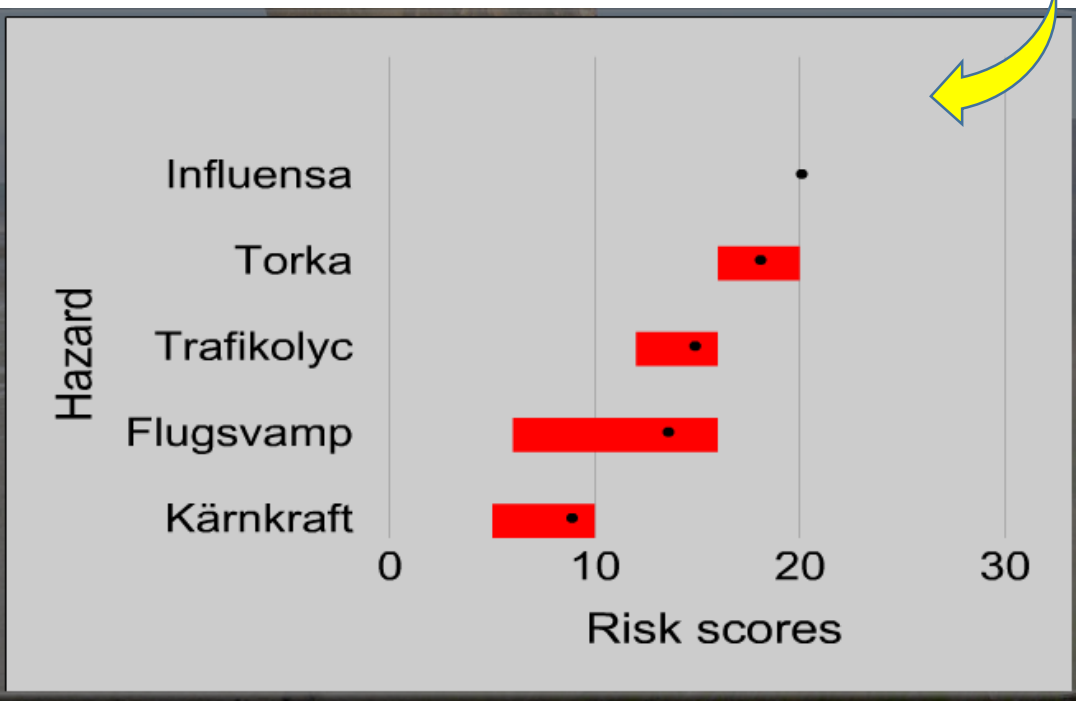
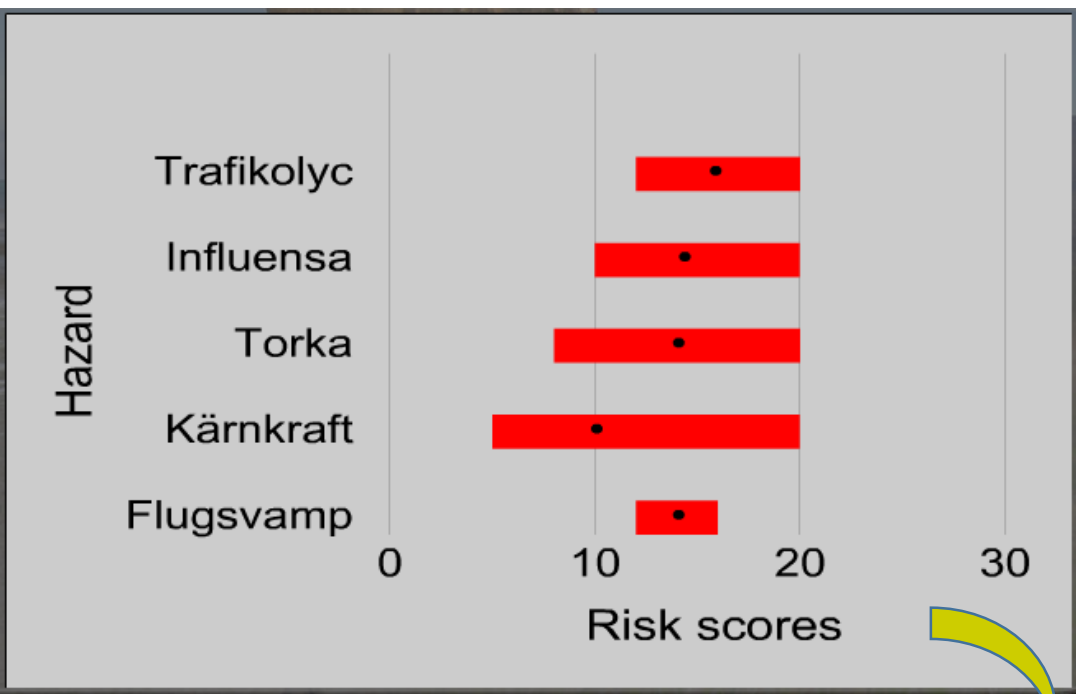
Elicitation with multiple experts

- More psychological factors add added when working with several experts.
- On the other hand, the result seems to be better (Wisdom of the crowds, multiple perspectives)
- Multiple experts needs to be aggregated somehow
- Behavioural aggregation
 - Group elicitation
 - One or several iterations, individually and in group
- Mathematical aggregation
 - Treat each expert's distribution as data and update the decision maker's belief
 - Pooled opinions – linear or logarithmic pooling
 - Calibrate experts and weight according to their performance

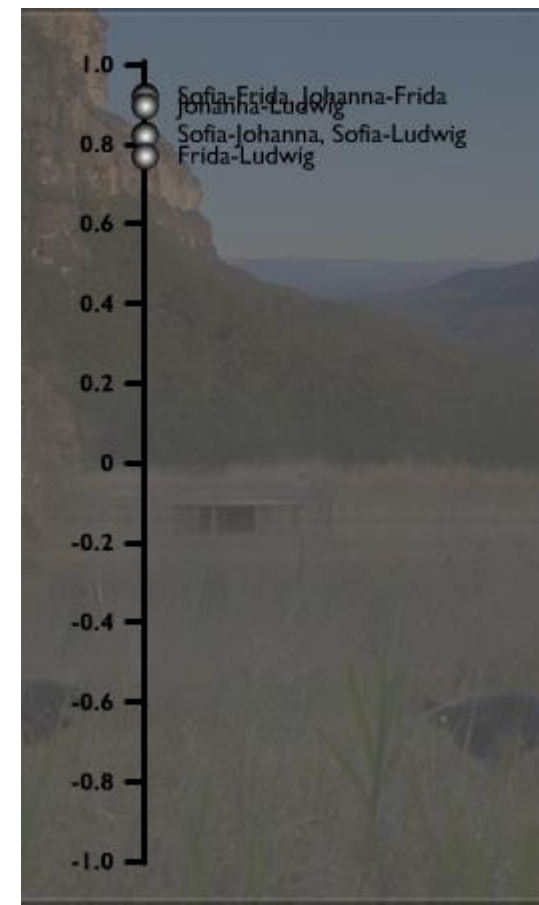
Alternative protocols for EKE

- The Sheffield protocol with group interaction of experts, consensus distributions
- The Cooke protocol with use of seed questions for the calibration of experts, no interaction
- The Delphi protocol on written expert elicitation with feedback loops, anonymous sharing of the results between iterations
- The IDEA protocol individual judgements, interaction of experts then individual judgements at the end

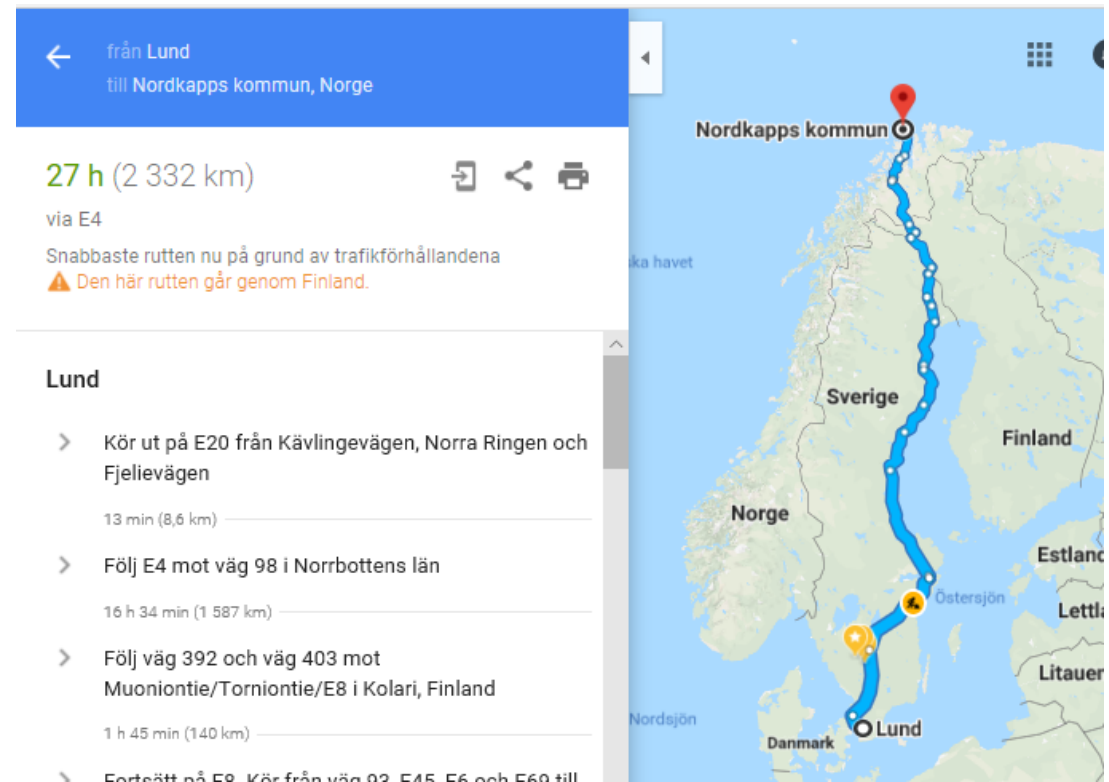
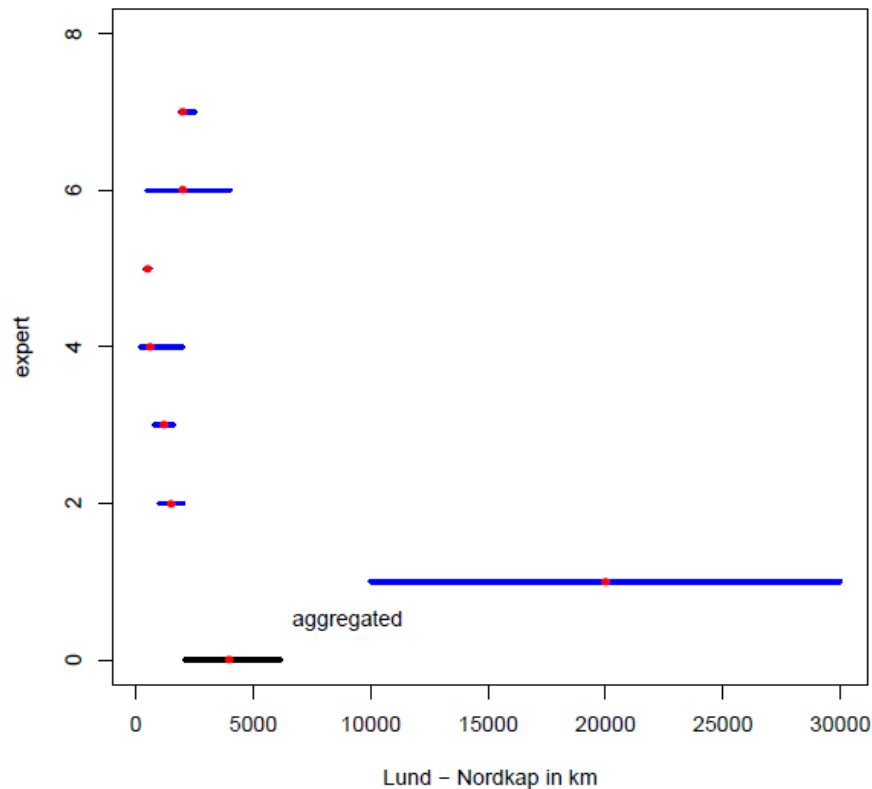




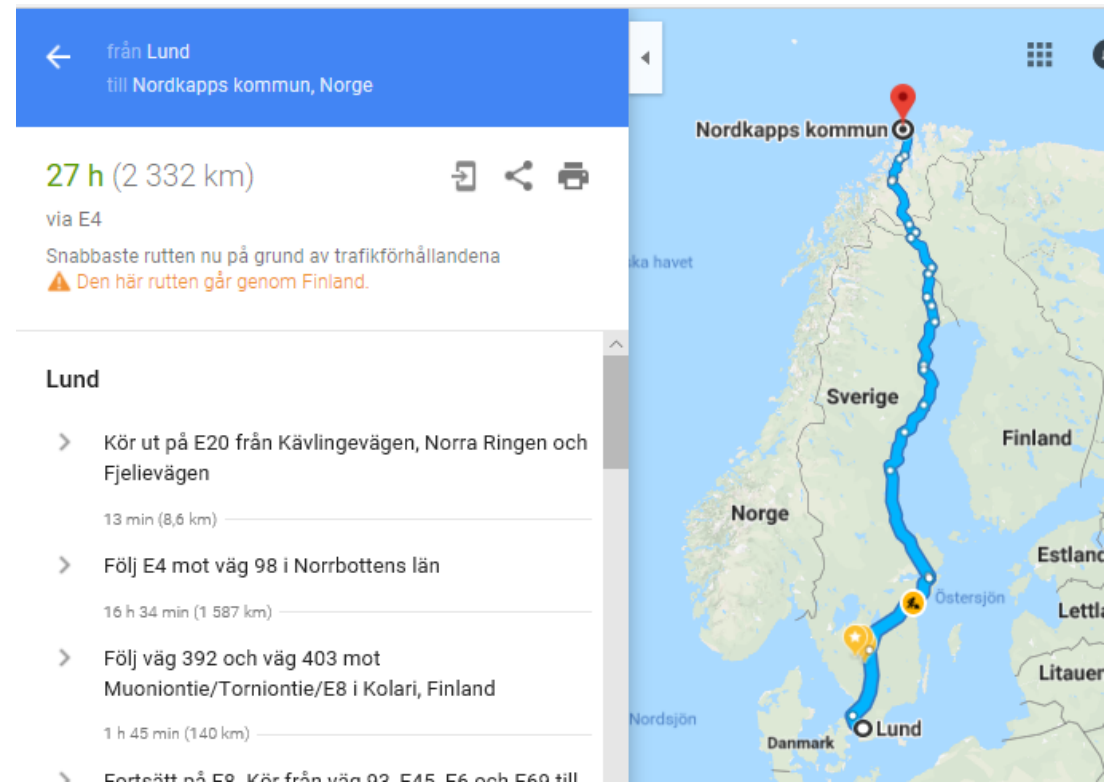
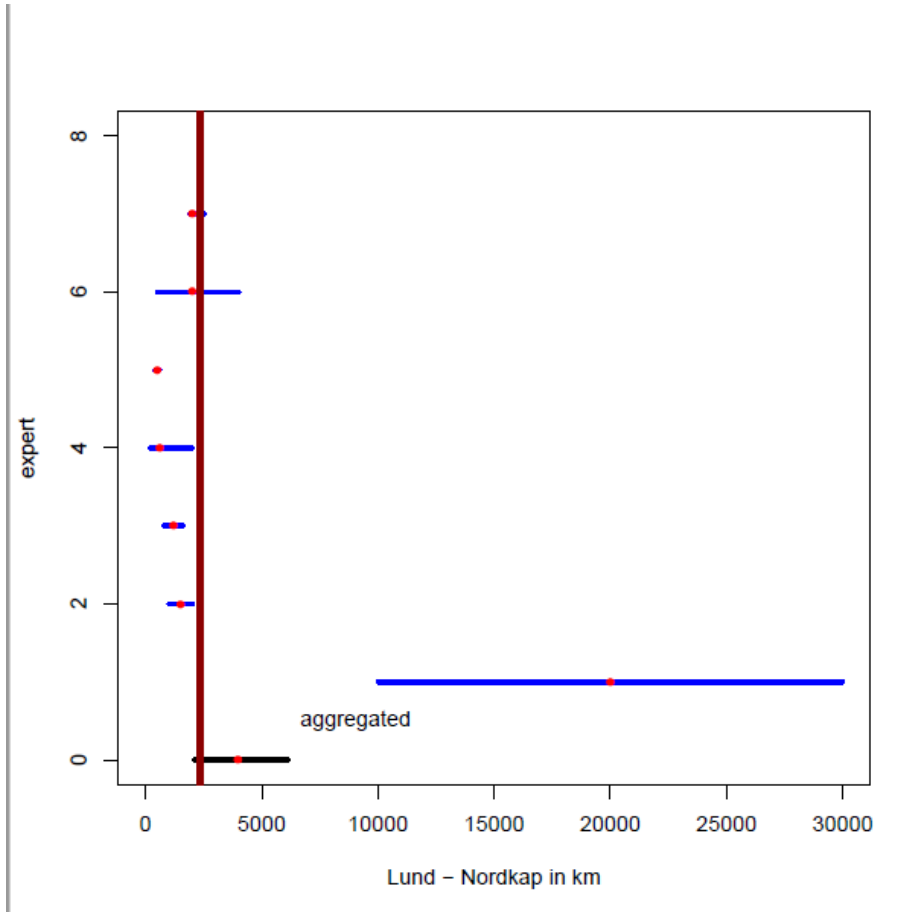
This is an exercise I do with the undergraduate students which includes one iteration



One by one, most of you could not hit the correct distance with a probability interval



....but the aggregated interval is rather precise and covers the true value.



A simple risk analysis

The code is at https://github.com/Ullrika/quantifying_uncertainty_by_probability

The name of the file is

Risk analysis with EKE and Bayesian analysis.R

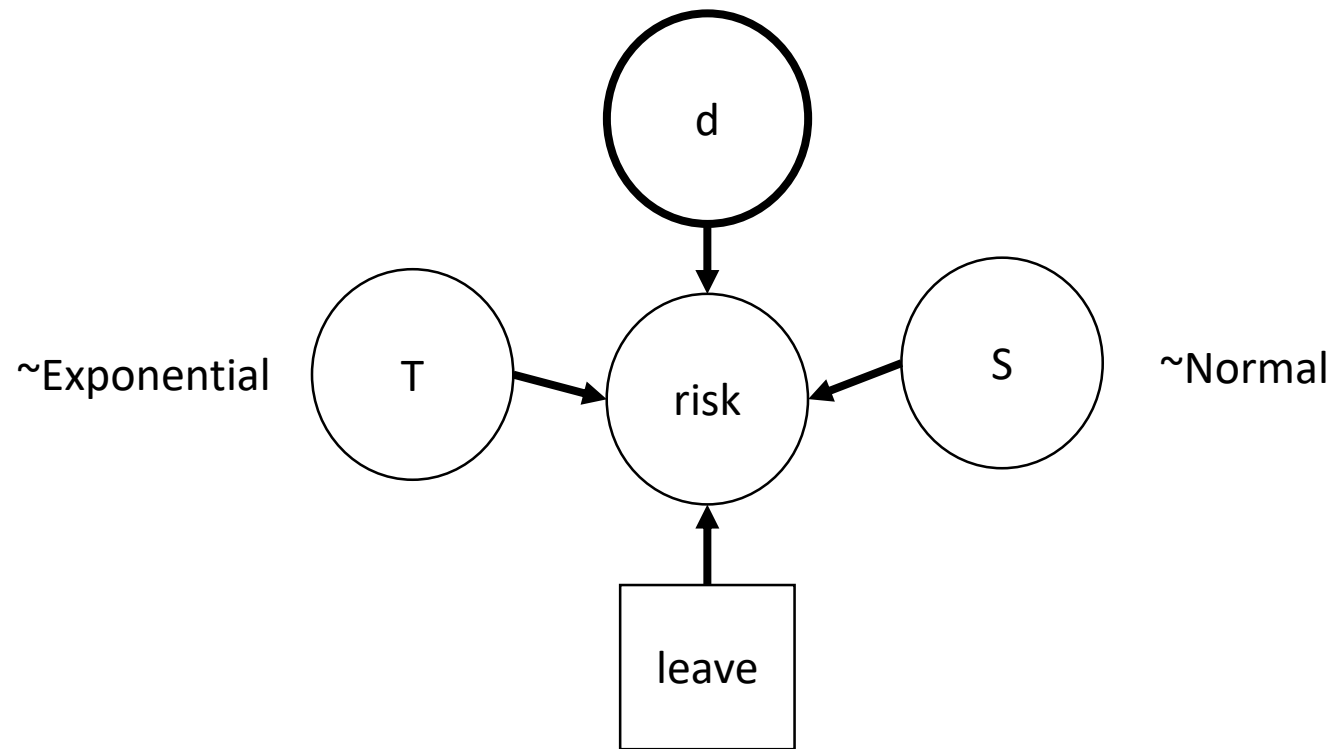
Most figures that are plotted in the code are saved in a Figures.pdf

Risk = the probability of missing the bus

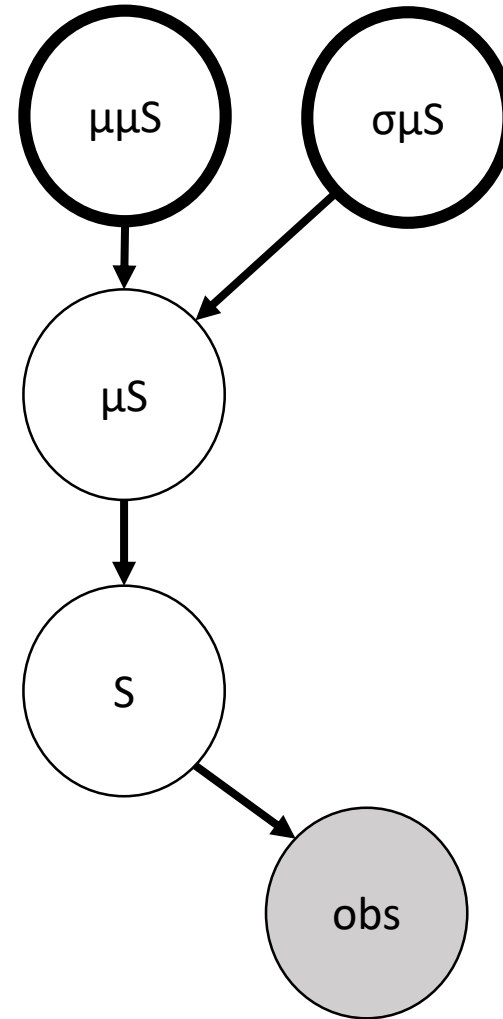
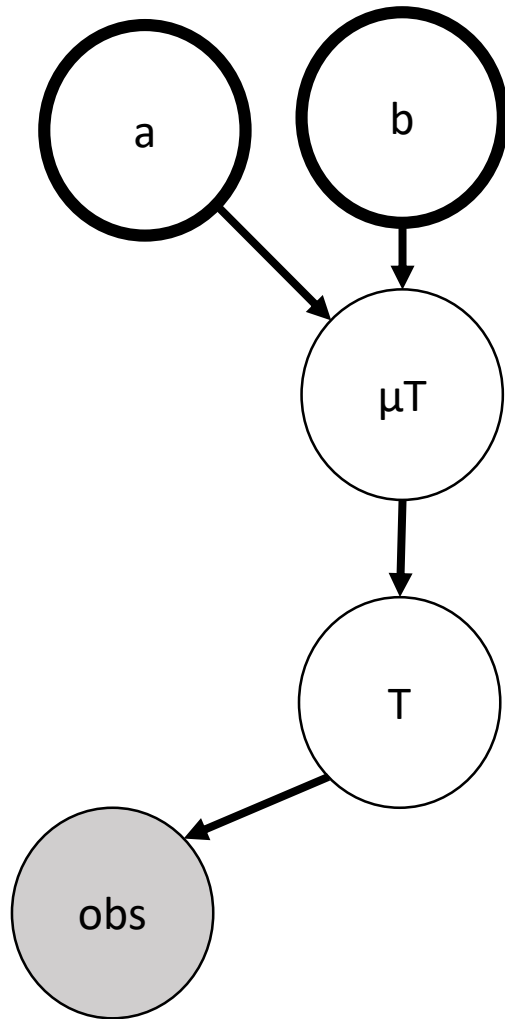
- I believe the last bus is expected to arrive in 10 minutes
- The distance to the bus stop is 540 m
- Walking speed is according to Wikipedia about 1.4 m/s
- When should I leave to make sure the risk of missing the bus is at an acceptable level, say 5%?

Bayesian analysis

This is our risk analysis model

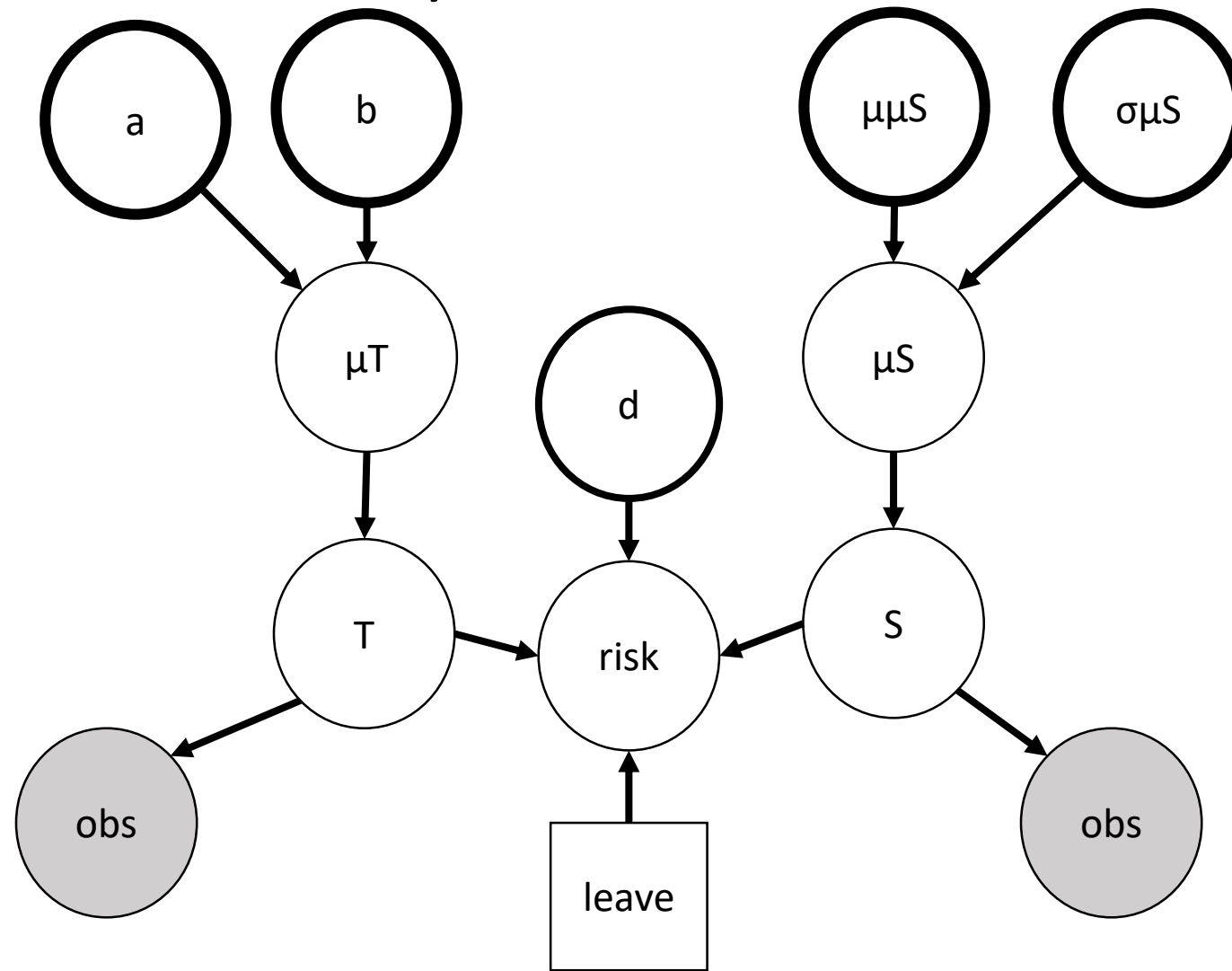


Bayesian analysis



We have data on waiting time and on walking speed. We build a probabilistic graph so we can derive the likelihood for the data given parameters.

Bayesian analysis



When we run this analysis we can run the parametric and predictive inference in the same step.

```
## You need to have JAGS on your computer
## Goto http://mcmc-jags.sourceforge.net/
## R-packages needed
## SHELF
## rjags
```

```
#####
## A simple risk model
```

```
#####
```

```
{
```

```
#####
```

```
## A risk analysis considering uncertainty in parameters
```

```
#####
```

```
{
```

```
#####
```

```
## A risk analysis in which uncertainty in parameters are quantified
```

```
## by Expert Knowledge Elicitation
```

```
#####
```

```
{
```

```
#####
```

```
## Learn about the parameters from data
```

```
#####
```

```
{
```

```
#####
```

```
## Analyse risk after learning about the parameters from data using priors
```

```
## informed by expert knowledge elicitation
```

```
#####
```

```
{
```


Recommended literature/courses/meetings

Probabilistic judgements by Anthony O'Hagan

<http://www.tonyohagan.co.uk/shelf/ecourse.html>

DataCamp Fundamentals of Bayesian Data Analysis in R by Rasmus Bååth

<https://www.datacamp.com/courses/fundamentals-of-bayesian-data-analysis-in-r>

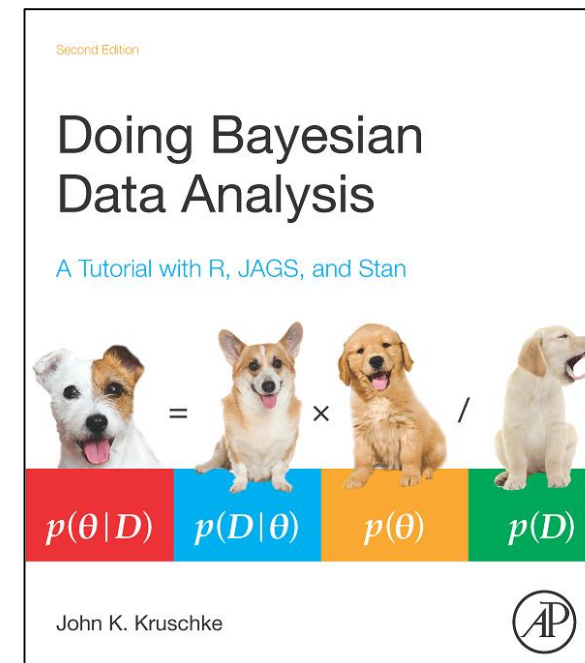
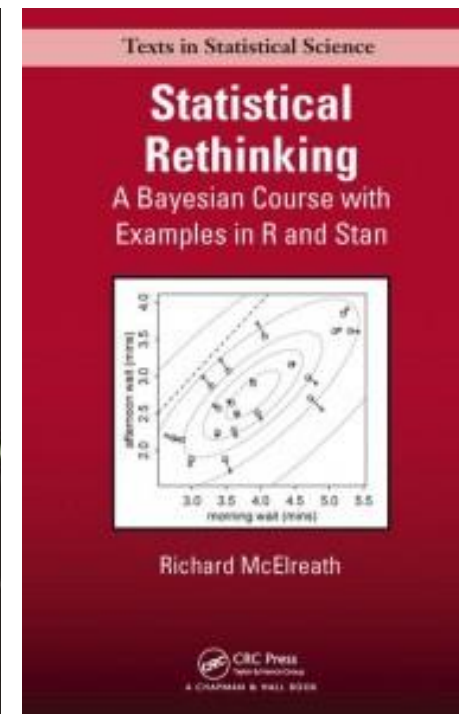
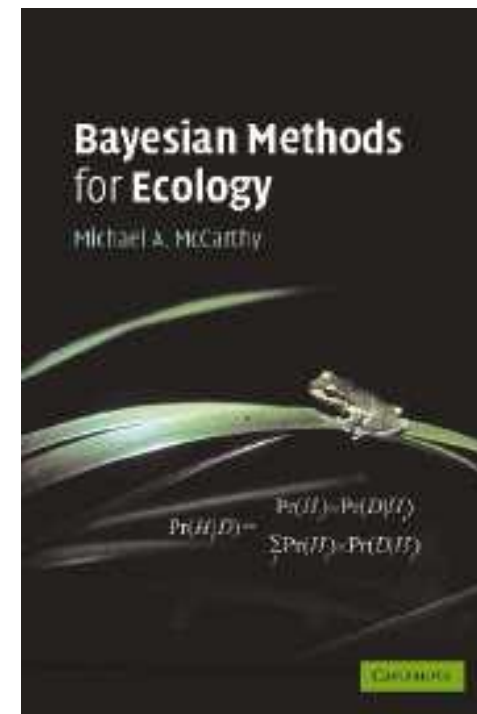
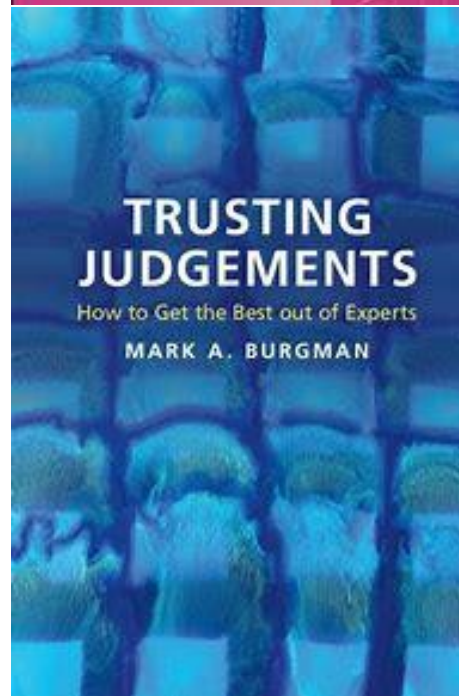
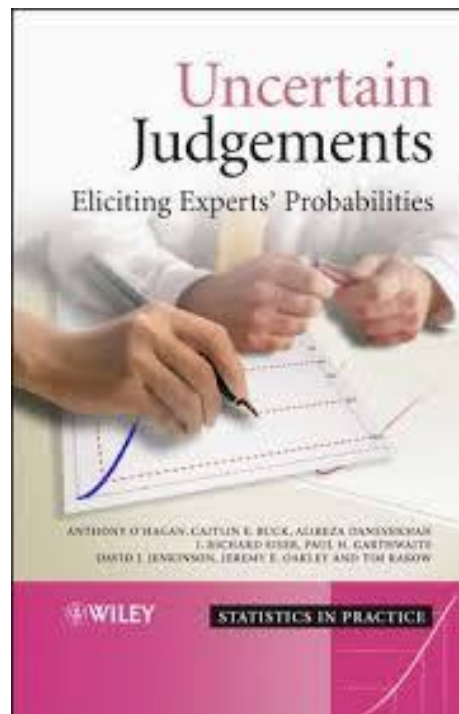
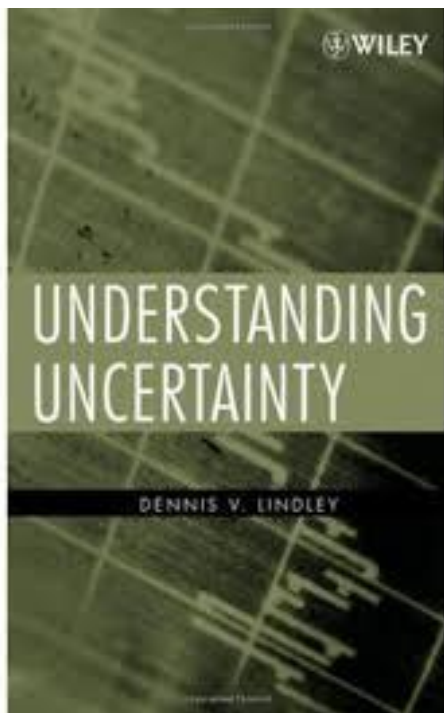
Bayesian methods course at the Department of Statistics, LU, Jonas Wallin

https://www.stat.lu.se/en/education/courses/stae02_bayesian_methods

Bayesian Analysis and Decision Theory - Graduate course, CEC, Lund, Ullrika Sahlin

Rasmus blog <http://www.sumsar.net/blog/2015/12/bayes-js-a-small-library-for-doing-mcmc-in-the-browser/>

Bayes@Lund – annual and accessible meeting in Lund (sign up to the Bayesian email list at Lund University to get early information <http://www.lucs.lu.se/bayes/>)



A protocol for summarising scientific uncertainty

1. Identify key outcomes for decision makers and how to measure them
2. Summarise variability
3. Summarise internal validity
4. Summarise external validity
5. Summarise the strength of the basic science (e.g. NUSAP)
6. Summarise uncertainty (e.g. probability distributions or credible intervals)

Recommendations in the Spiegelhalter and Riesch paper

1. Use quantitative models with aleatory and epistemic uncertainty expressed as Bayesian probability distributions.
2. Conduct sensitivity analysis to alternative model forms and assess evidential support for alternative structures, without putting probabilities on models.
3. Provide a list of known model limitations and a judgement of their qualitative or quantitative influence, possibly along the lines shown in table 6, and ensuring there has been a fully imaginative consideration of possible futures.
4. Provide a qualitative expression of confidence, or lack of it, in any analysis based on the quality of the underlying evidence, possibly expressed using an adapted GRADE scale or the IPCC guidance [46].
5. In situations of low confidence, use deliberately imprecise expressions of uncertainty about quantities, such as their orders-of-magnitude, whether they are positive or negative, or even refuse to give any judgement at all; the IPCC guidance suggests a calibrated scale for these expressions.
6. When exploring possible actions, look for robustness to error, resilience to the unforeseen, and potential for adaptivity in the face of the unexpected [10].
7. Seek transparency and ease of interrogation of any model, with clear expression of the provenance of assumptions.
8. Communicate the estimates with humility, communicate the uncertainty with confidence.
9. Fully acknowledge the role of judgement: this ‘..means engaging in policy making by fully accepting the constructive, participatory, ultimately open-ended and untamed nature of judgements under uncertainty’

Approach uncertainty scepticism (Fischoof and Davis again)

Table 3. Frequently asked questions addressing four concerns of scientists reluctant to express their uncertainty in credible-interval form

Concern 1	If I give credible intervals, people will misinterpret them, inferring greater precision than I intended.
Response	Behavioral research has found that most people (<i>i</i>) like receiving explicit quantitative expressions of uncertainty (such as credible intervals), (<i>ii</i>) can interpret them well enough to extract their main message, and (<i>iii</i>) misinterpret verbal expressions of uncertainty (e.g., “good” evidence, “rare” side effect). For audiences that receive the reports created with the protocol (Table 2), understanding should be greater if they receive credible intervals than if they have to infer them (63).
Concern 2	People cannot use probabilities.
Response	Behavioral research has found that laypeople can often provide reasonably consistent probability judgments if asked clear questions and extract needed information if provided with well-designed displays (41, 60, 74). Whether they do so well enough to satisfy their decision-making needs is an empirical question, which should be answered with evidence rather than speculation.
Concern 3	My credible intervals will be used unfairly in performance evaluations.
Response	Such judgments can protect experts from unfair evaluations, unjustly accusing them of having conveyed too much or too little confidence, especially when supported by the rationale for those judgments. The protocol provides such protection—if the experts’ management stands behind it.
Concern 4	People do not need such judgments.
Response	Decision makers must act with some degree of uncertainty. Not helping them means preferring the risk of having them guess incorrectly over the risk of expressing oneself poorly.