Precise versus Bounded Probability

ULLRIKA SAHLIN, LUND UNIVERSITY

THINKOVER AT LIVERPOOL RISK INSTITUTE ONLINE 2020

Outline

- Introduction to probability as a measure of epistemic uncertainty
- Discussion
- Bounded probability
- Discussion
- More detailed discussion about the challenge problems for those interested

Declaration of interest

Uncertainty is an uncomfortable position.
But certainty is an absurd one.

Voltaire

Epistemic uncertainty in assessments

- Uncertainty
  - Uncertainty is any limitations in knowledge – epistemic
  - Someone is uncertain – personal
  - This uncertainty can change when new knowledge becomes available – changeable
- Scientific assessment
  - A scientific procedure to produce an answer to a question asked by a decision maker – the DM should be willing to accept the answer as their own
  - The assessors (Un)Certainty about the answer is of importance – could decrease or increase a DM's confidence in the answer
  - Honest communication of uncertainty is the only option

The terms parameter and variable are used for many different things – to facilitate a structured discussion about uncertainty the following terminology is chosen

Parameters and variables in assessments

- An assessment model consists of assessment variables in a structure that express scientific theory/hypothesis, mechanistic understanding, causal relationships and random variability (aleatory uncertainty).
- Parameters are numbers with a fixed value within an assessment model – uncertainty about parameters are epistemic, it is useful to use parameters that has a meaning
- Example: $X \sim N(\mu, \sigma^2)$

Assessment variable

Parameters within the assessment model

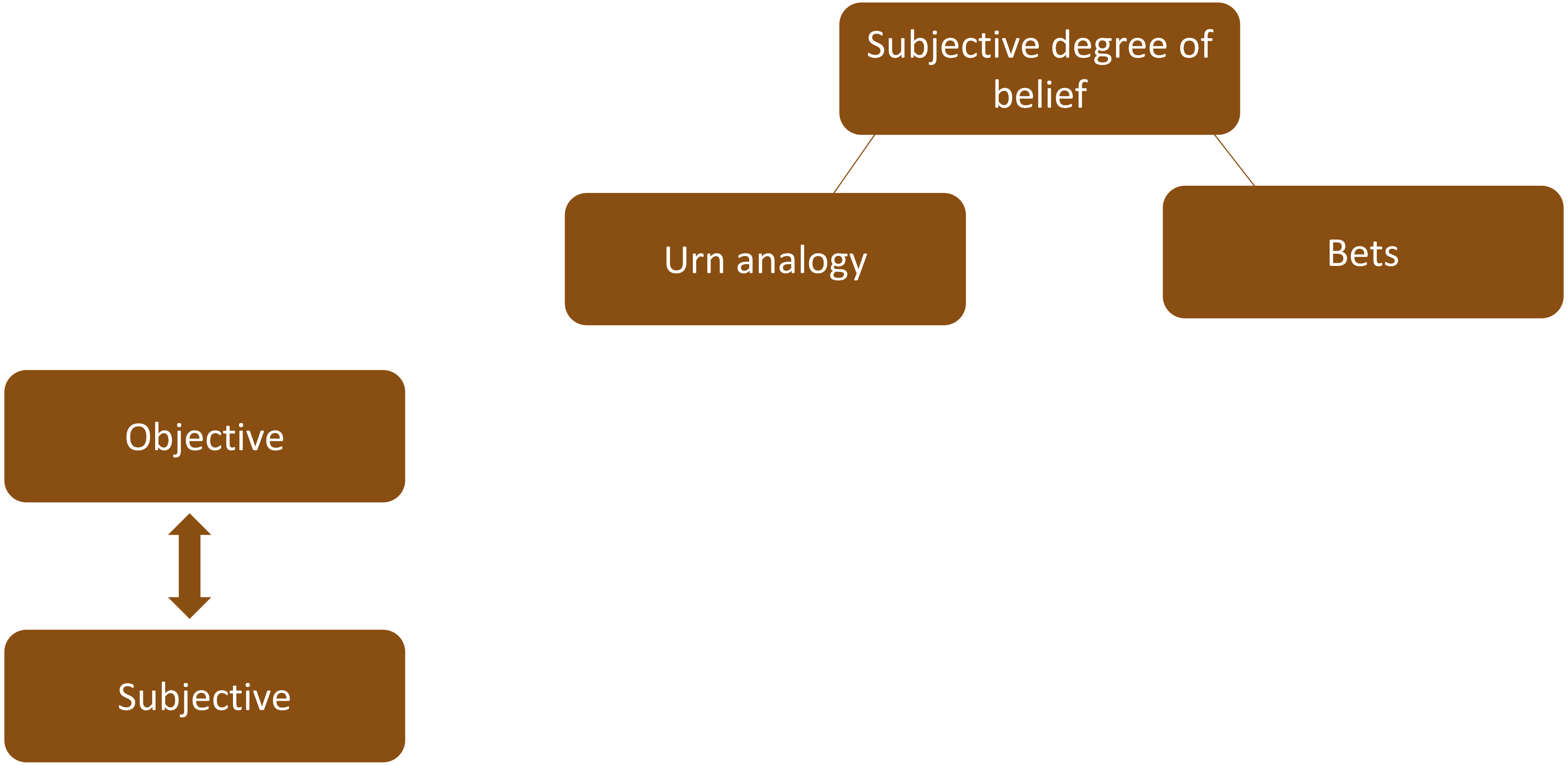
There are different interpretations of probability as a measure of epistemic uncertainty

Probability as a measure of uncertainty

- Probability as a concept appear in the Middle Ages
- Pascal and Fermat (1600) – classical urn probability
- Laplace (1749-1827) – inductive reasoning based on probability, probability calculus, CLT, OLS
- Keynes (1883-1946) – objective or logic probabilities
- Ramsey (1903-1939) – subjective probability based on gambles. Probability is personal and can be used for epistemic uncertainty. Repeated experiments or iterative learning are important for updating subjective probabilities and they should converge to a common conclusion in light of evidence
- Savage (1917-1971) – normative decision theory based on subjective probability. Subjective probability is not the only measure of epistemic uncertainty
- Kolmogorov (1903-1987) – probability as a mathematical measure. Axioms. No specific interpretation
- Raiffa (1924-2016) – Bayesian decision theory. Subjective probability follow Kolmogorov's axioms. Introduced conjugate priors
- De Finetti (1906-1985) – subjective probability measured through games. Bayesian statistical theory e.g. exchangeability and conjugate models
- Jaynes (1922-1998) – Bayesian inference as extended logic
- …

There are different interpretations of probability as a measure of epistemic uncertainty

Probability as a measure of uncertainty



Probability as a measure of uncertainty

The "assessment perspective"

- Model: P(variables|parameters) = P(X|μ,σ)
- Measure of uncertainty: P(parameters) = P(μ,σ)

epistemic or aleatory!

epistemic

$$\mu \sim N(\mu_0, \sigma_0)$$

$$\sigma \sim \ldots$$

$$X \sim N(\mu, \sigma^2)$$

Probability as a measure of uncertainty

The ”assessment perspective”

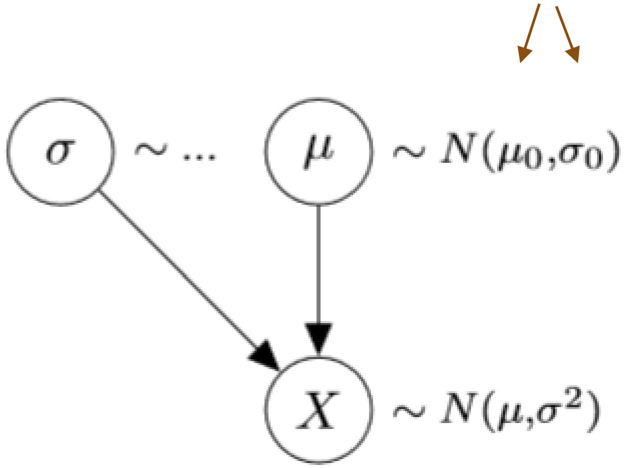
- Model: P(variables|parameters) = P(X|μ,σ) **epistemic or aleatory!**
- Measure of uncertainty: P(parameters) = P(μ,σ) **epistemic**

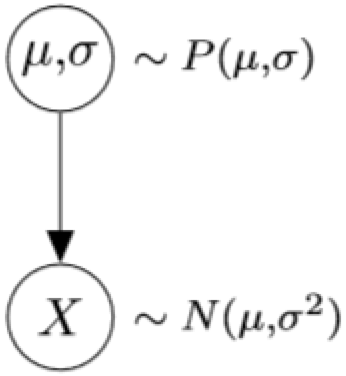
$$\mu \sim N(\mu_0, \sigma_0)$$

$$\sigma \sim \ldots$$

$$X \sim N(\mu, \sigma^2)$$

hyperparameters





Probability as a measure of uncertainty

The "assessment perspective"

- Model: P(variables|parameters) = $P(X|\mu,\sigma)$

epistemic or aleatory!

- Measure of uncertainty: P(parameters) = $P(\mu,\sigma)$

epistemic

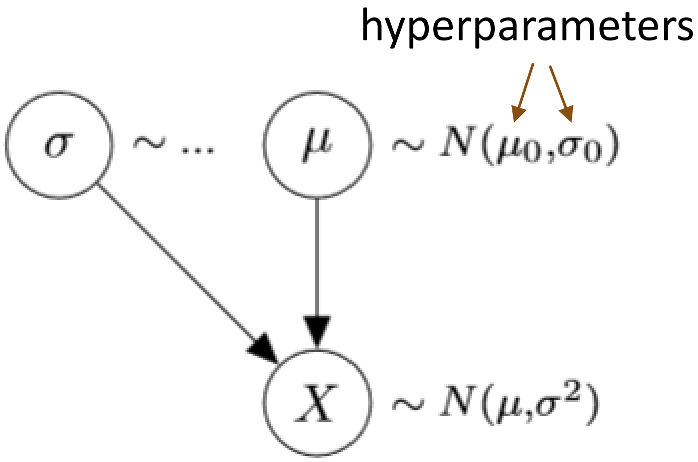
- We can summarise uncertainty about
  - parameters e.g. $\mu<1$
  - a function of parameters e.g. $h(\mu,\sigma)<1$, where $h(\mu,\sigma):=P(X<1|\mu,\sigma)$

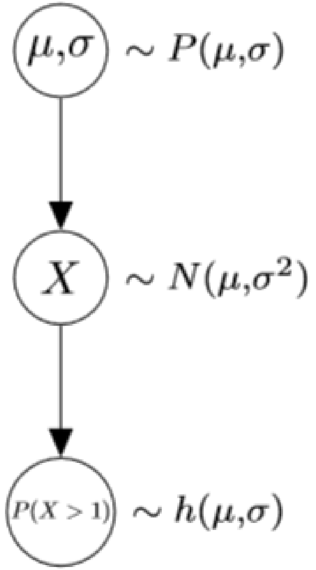
h can be a model of e.g. risk or utility

$$\mu \sim N(\mu_0, \sigma_0)$$

$$\sigma \sim \ldots$$

$$X \sim N(\mu, \sigma^2)$$





The ”assessment perspective” on Bayesian inference

Some comments:

- P(assessment variables|parameters) are representing aleatory uncertainty and can be interpreted as relative frequencies
- Subjective probability on relative frequency approach
- Note:
  - In Bayesian inference both variables and parameters are mathematically treated as random variables – however, they represent different types of uncertainty
  - It is the assessors task to make sure these are not mixed when propagating uncertainty in the assessment
- Can or must this perspective apply to P(data|parameters), i.e. the likelihood as well?

How do we get the probabilities?

Data

Expert knowledge

P

Probability as a measure of uncertainty

- P(parameters) should represent the assessors uncertainty
- The assessor may consult one or several experts
- Structured process for Expert Knowledge Elicitation
- Behavioural or mathematical aggregation into one distribution

Bayesian inference

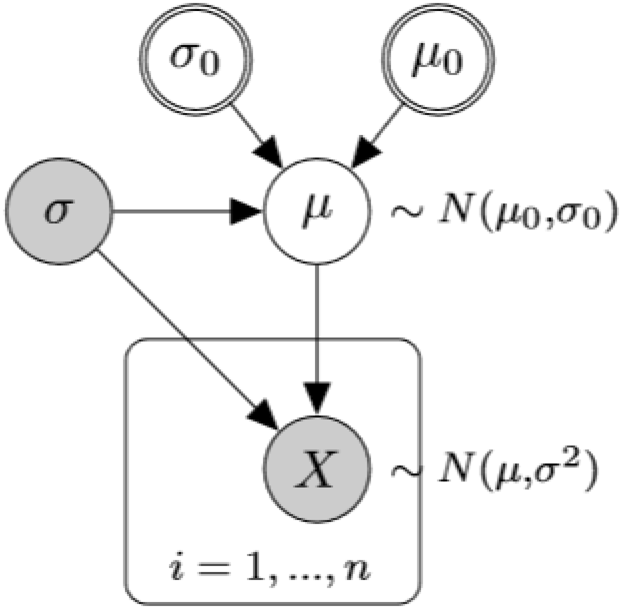
- Statistical theory to learn from data where epistemic uncertainty is quantified by probability and it is possible to specify a probabilistic model for data
- Probabilistic model for data P(data|parameters) – the likelihood
- Probability distribution for uncertainty about the parameters
  - Prior distribution – before data is seen P(parameters)
  - Posterior distribution – after data is seen P(parameters|data)
- Bayesian inference integrates expert knowledge (the prior) and data
- Sequential updating as data becomes available
- Propagation of uncertainty by probability calculus
- Bayesian updating is done analytically, e.g. using conjugate models, or approximately, e.g. by MCMC sampling or ABC

Bayes rule

Bayesian posterior

$X \sim N(\mu, \sigma^2)$

- Assume σ is known
- Data are n independent observations of X
- Specify a model that associates data to the parameters we want to make inference on
- Tasks:
  - 1) conclusion about a parameter e.g. μ>1 — Derive the posterior!
  - 2) conclusion about a quantity of interest e.g. h:=P(X>1| μ,σ) — Derive the uncertainty in the quantity of interest!



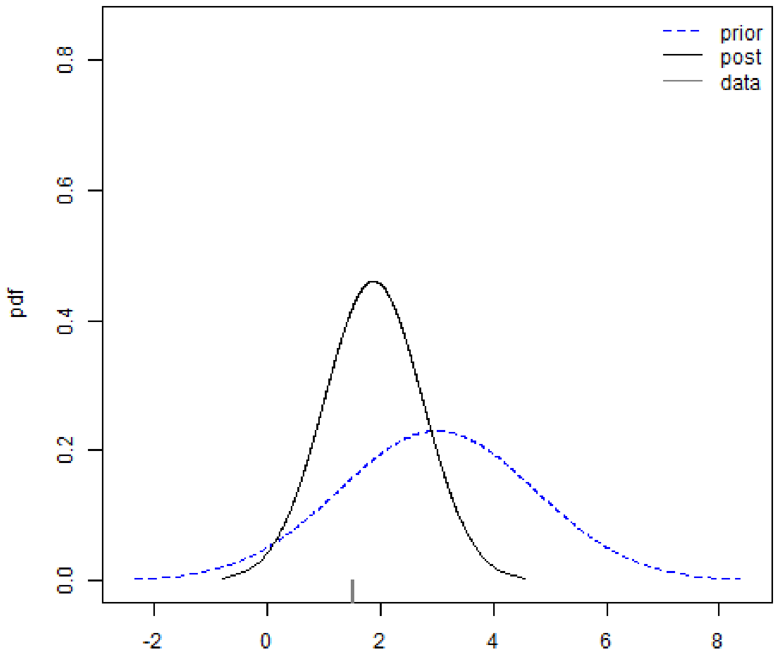
$$\mu \sim N(\mu_0, \sigma_0)$$ **Prior**

1) $$\mu|(x_1, ..., x_n), \sigma \sim N(\mu_1, \sigma_1^2)$$ **Posterior**

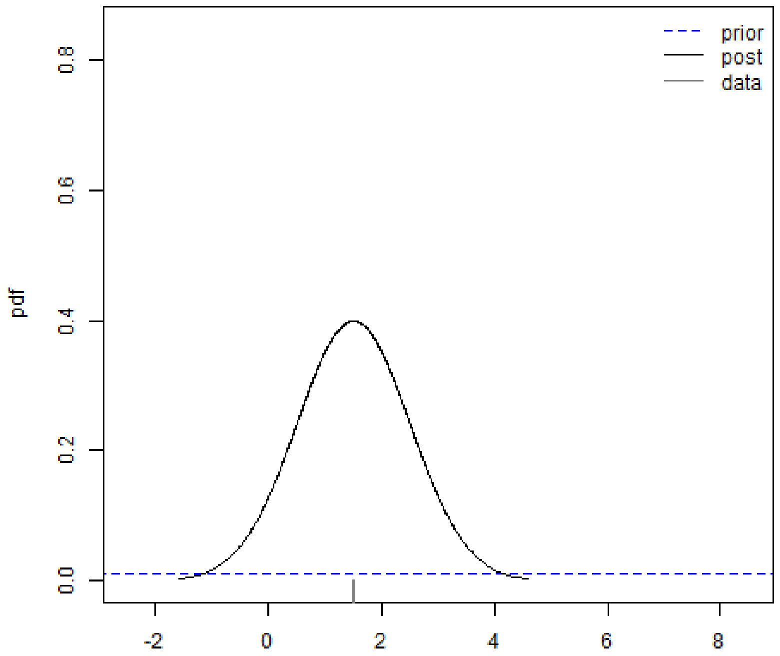
$$\mu_1 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}\right) \qquad \sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

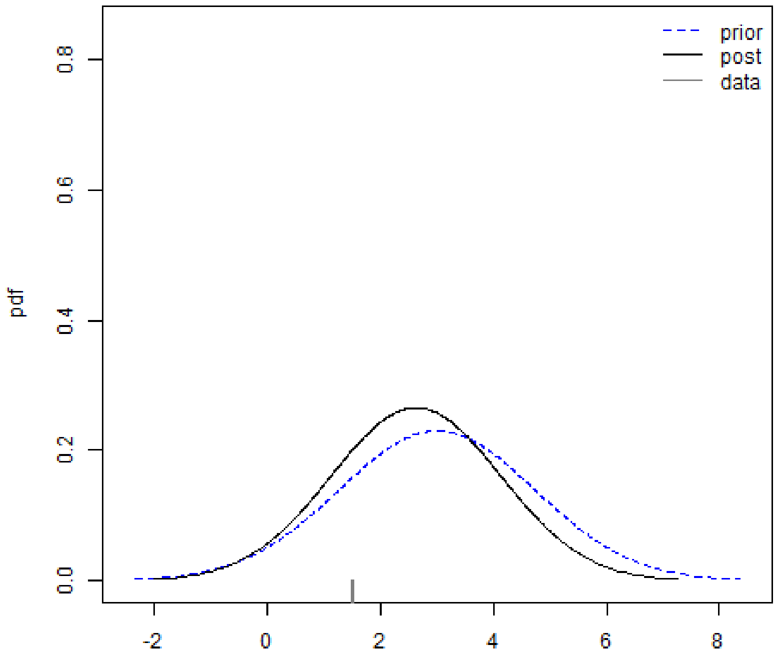
Conjugate model for normal with known variance

2) $$h|(x_1, ..., x_n), \sigma = \int_1^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu_1}{\sigma}\right)^2} dt$$ **Uncertain quantity of interest**

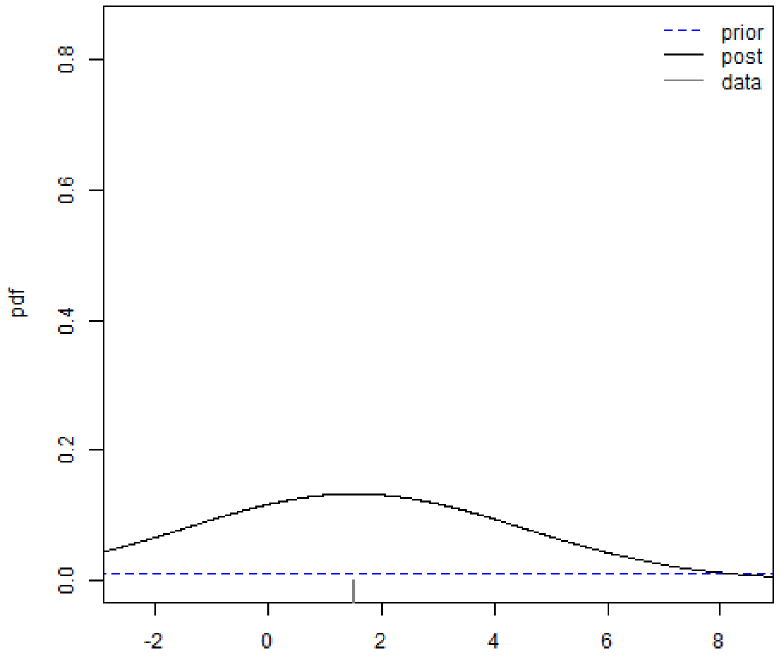


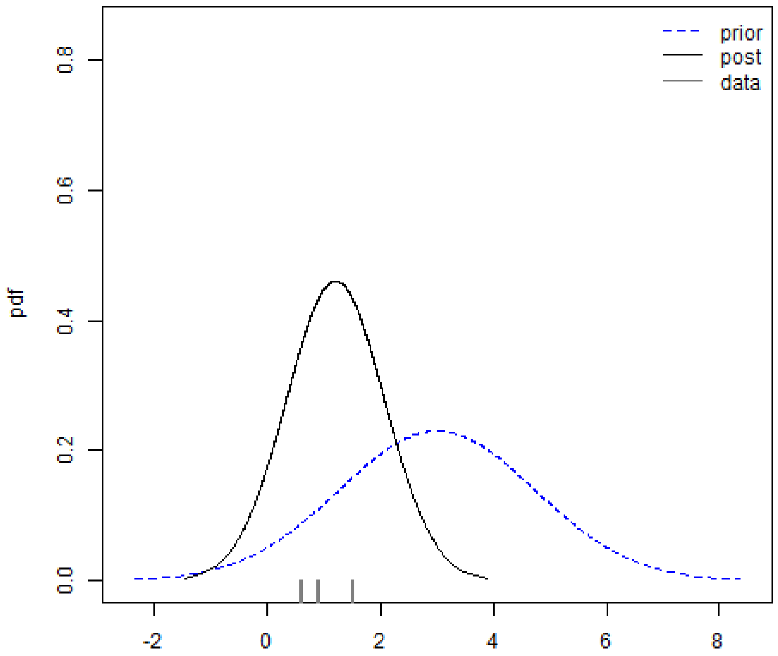
Small error σ



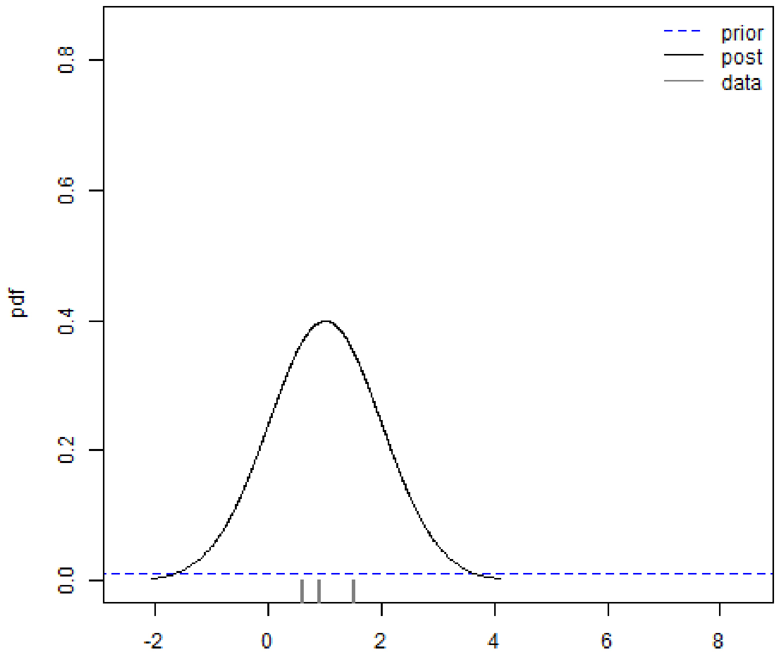


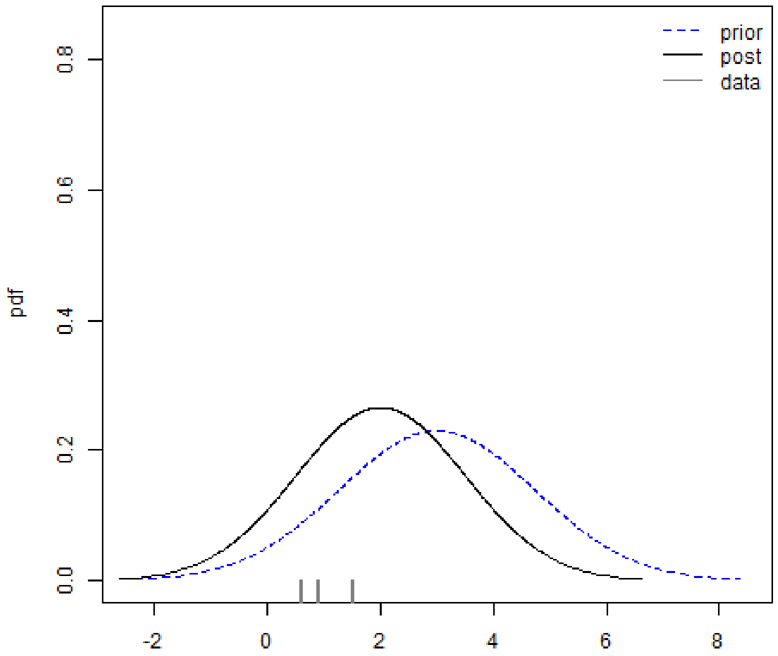
Large error σ



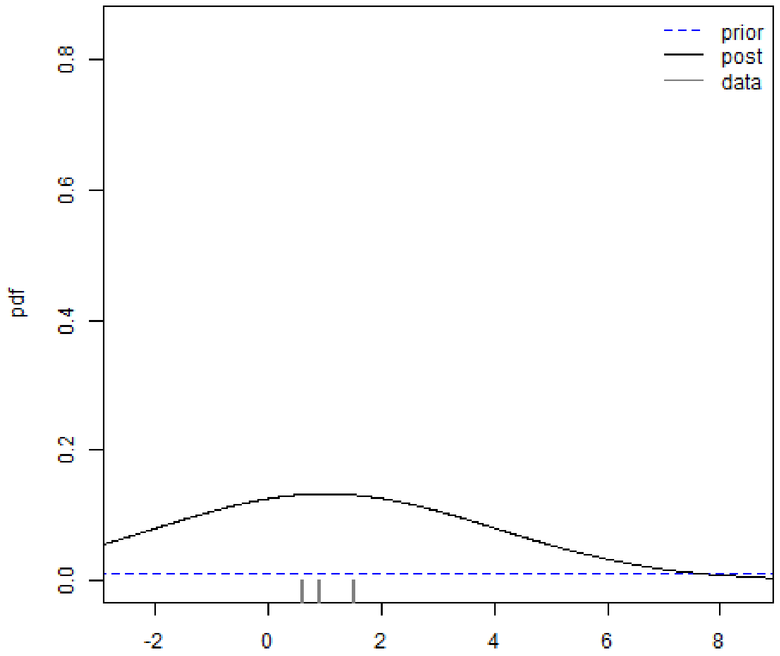


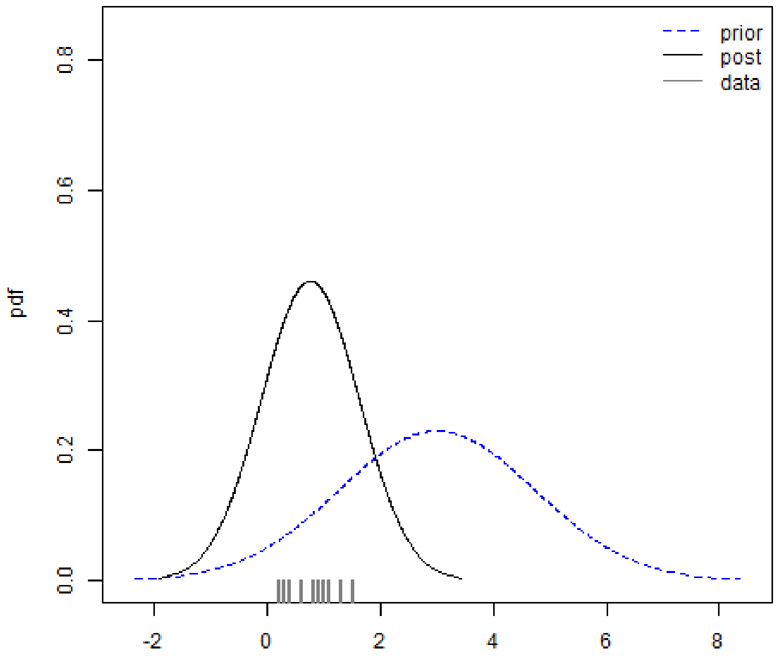
Small error σ



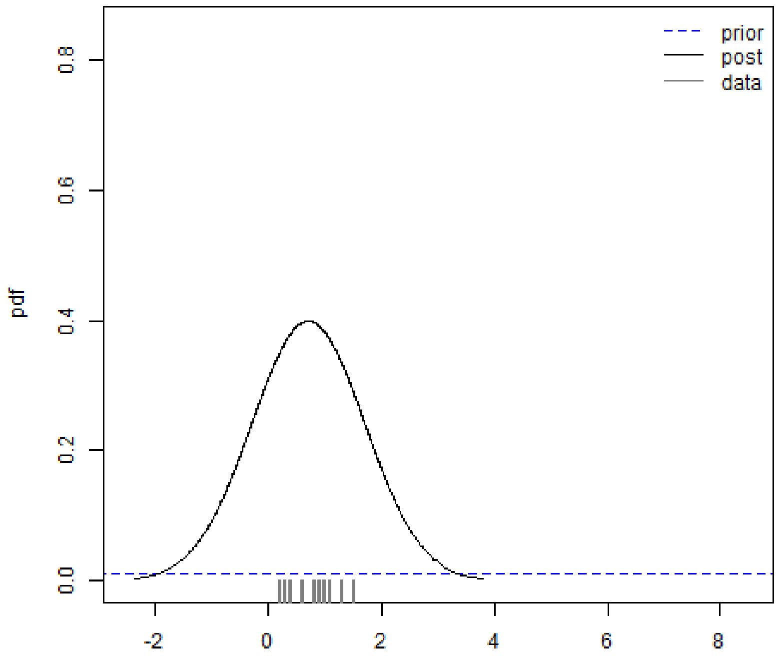


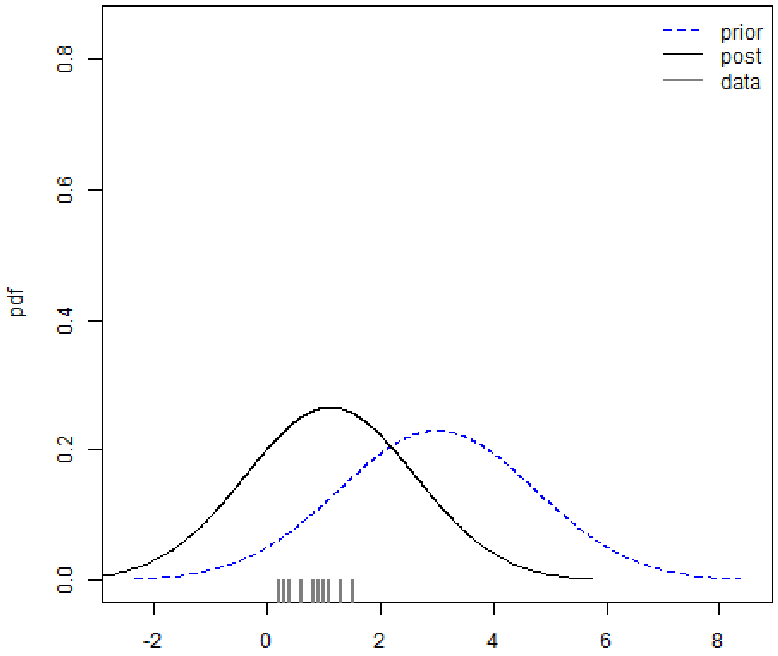
Large error σ



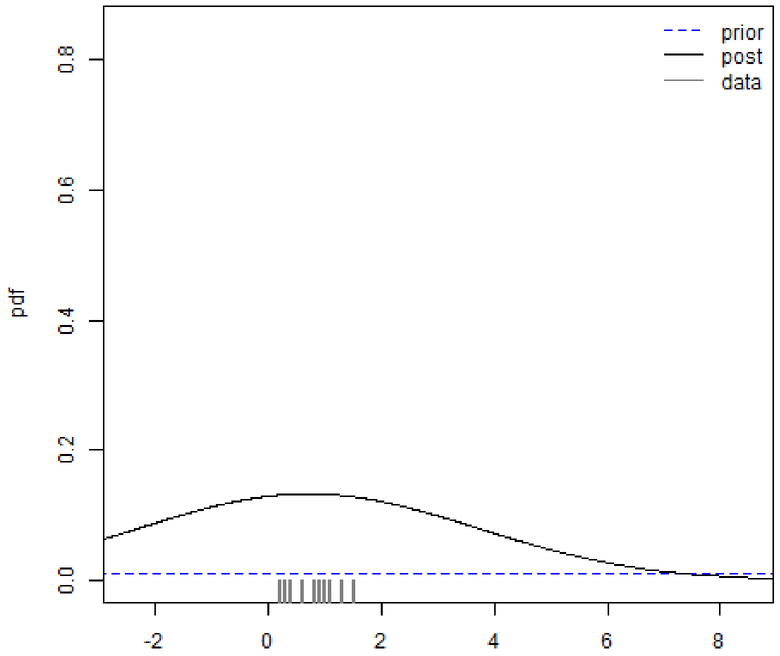


Small error σ



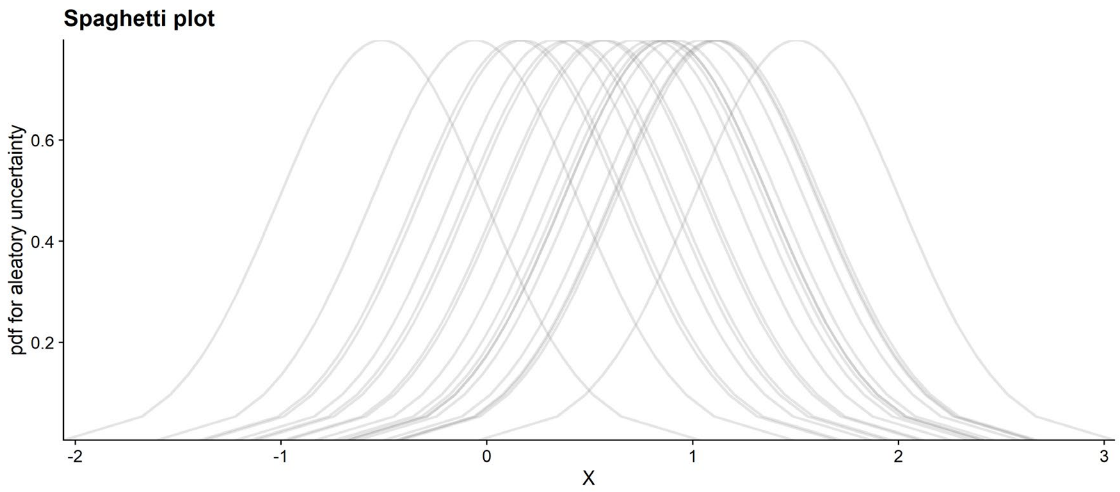


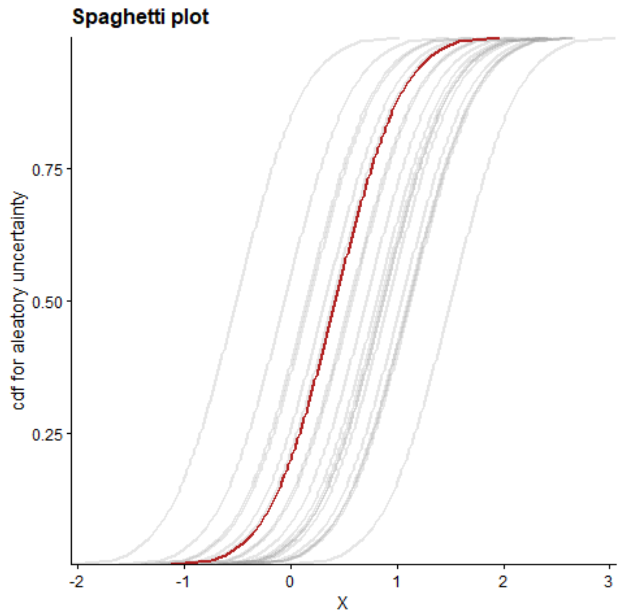
Large error σ



Uncertainty in a quantity of interest

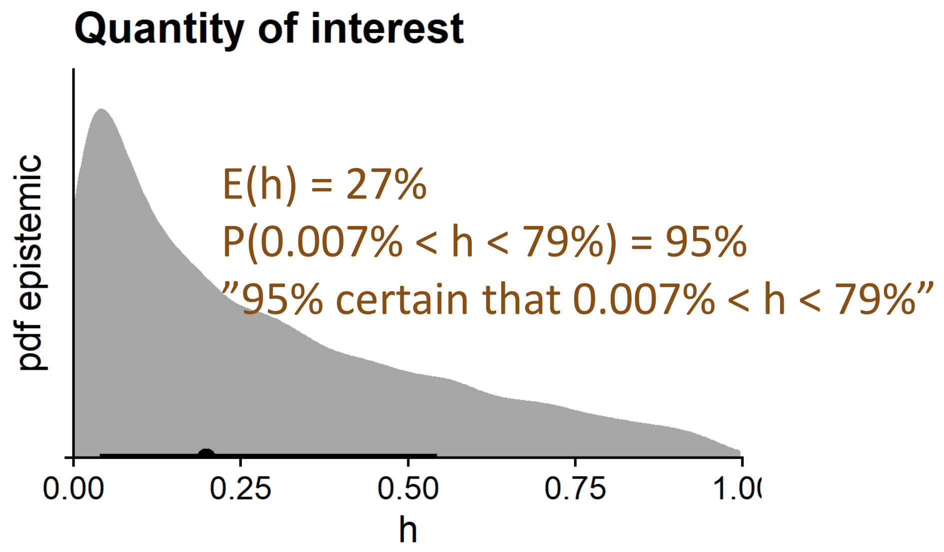
- Express uncertainty using precise probability implies to make statements about $h := P(X > 1 | \mu,\sigma)$, e.g. $P(h < 5\%)$ which could mean that something is "safe"
- Predictive distribution $P(X|data)$ – under the "assessment perspective" this is a mixture of aleatory and epistemic uncertainty – should not be used
- Propagate uncertainty to h! – this is what we should do instead
- 2D distributions/spaghetti plots
  - visualisation of X separating aleatory and epistemic uncertainty

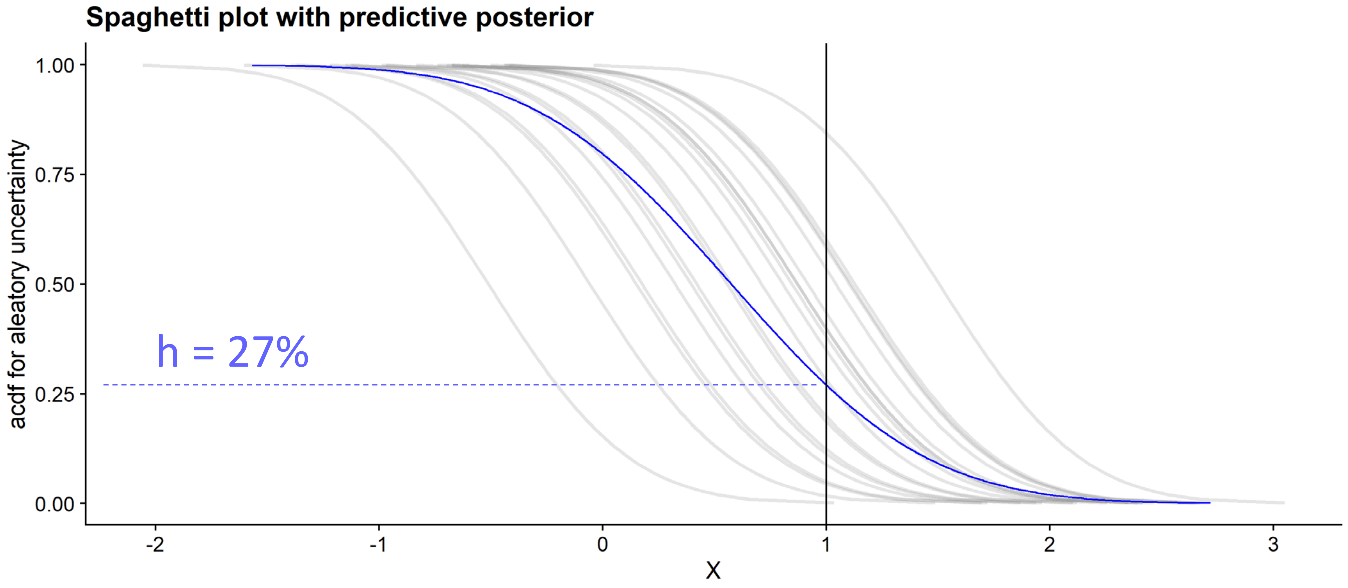




Uncertainty in a quantity of interest

- Uncertainty about h where h:=P(X > 1 | μ,σ)
- Predictive distribution P(X|data) – is a mixture of aleatory and epistemic uncertainty
- Propagate uncertainty to h





Uncertainty in a quantity of interest

- If the quantity of interest is not a parameter the following approach prevent us from mixing aleatory and epistemic uncertainty:
- Bayesian calibration of parameters in an assessment model (backward sampling)
  - P(data|parameters) can be seen as epistemic since data ”are history”
- Probabilistic uncertainty analysis propagating epistemic uncertainty P(parameters|data) to the quantity of interest (forward sampling)
  - P(variables|parameters) is aleatory
  - P(quantity of interest) is epistemic

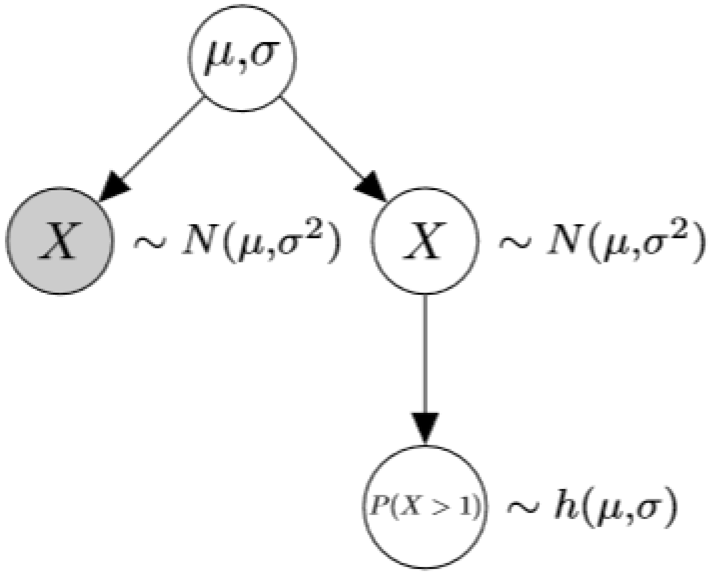
This is what I call

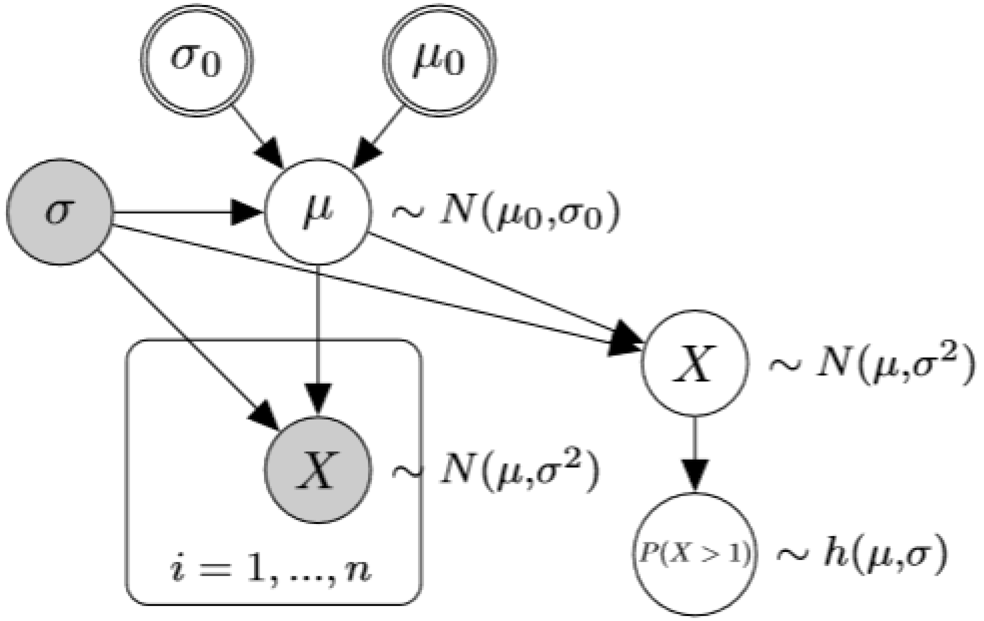
Bayesian Evidence Synthesis

Bayesian evidence synthesis

- Bayesian inference to calibrate the assessment model and propagate uncertainty to the quantity of interst (which could be a parameter or a function of parameters, h)

Same as to the right but with less details



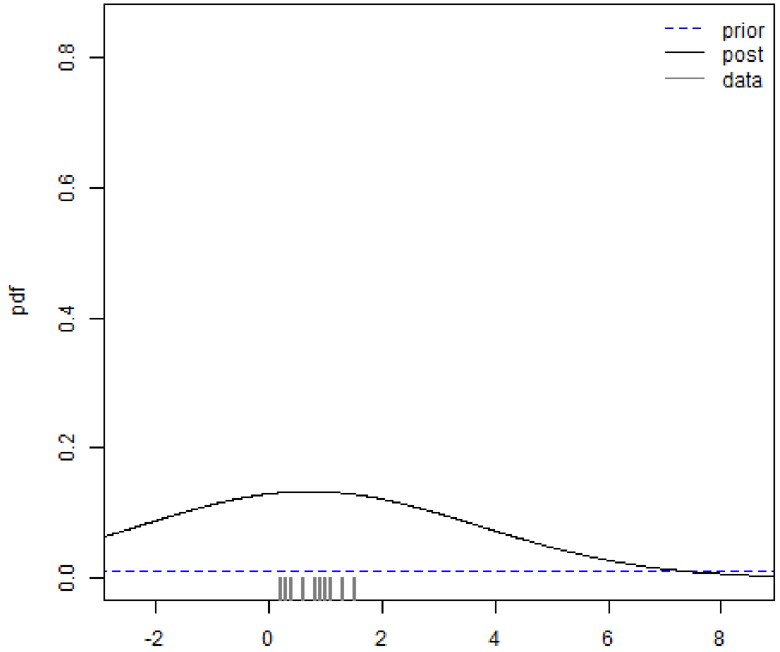


In what way is probability a good or bad way to quantify epistemic uncertainty?

Is bounded probability a useful alternative or complement to precise probability?

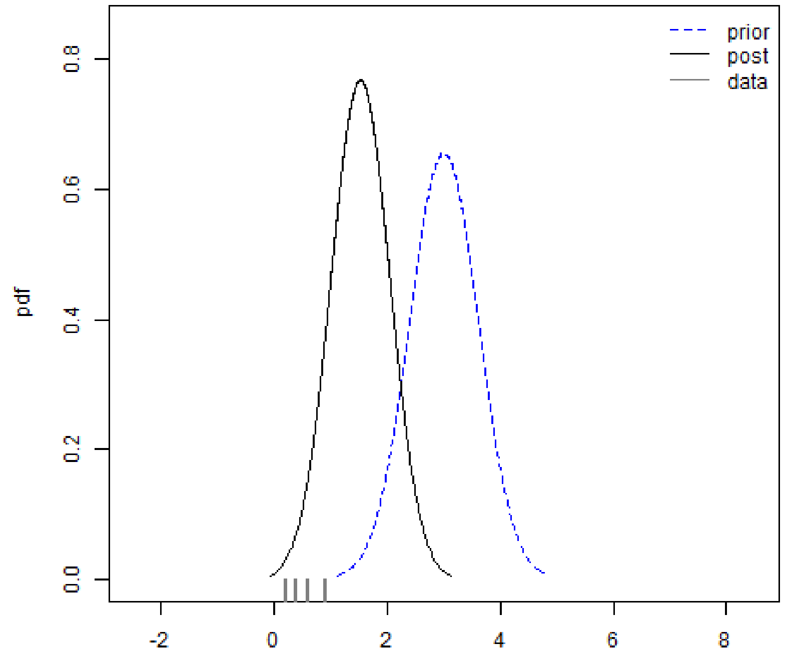
Caveats about precise probability

- Weak expert knowledge combined with weak data
  - Non-ideal situation for an assessment
  - Low confidence in uncertainty about the quantity of interest not seen in the analysis
  - Avoid using flat priors as default



Caveats about precise probability

- Strong expert knowledge in conflict with data (prior-data conflicts)
  - Inference tends to end up in the middle of prior and data
  - Low confidence in uncertainty about the quantity of interest not seen in the analysis
  - Indicates low confidence but is difficult to detect



Caveats about precise probability

- Experts may struggle in making precise probability statements

What if we are uncertain about our subjective probability?

- Could we express uncertainty in hyperparameters with probability as well?
- Examples of comments on using higher order probabilites:
- Hume (1739). It is meaningless to assign a degree of uncertainty to an uncertainty. When to stop?
- Savage's (a.k.a. Woodbury's) argument: If there is second or higher order uncertainty, these should be used as weights of the first order uncertainty. The weighted average is enough to express the relevant epistemic uncertainty
- Nils-Eric Sahlin: A weighted average of first order uncertainty may loose important information about uncertainty. Subjective probability as basis for action and second order probabilites as epistemic probabilities (degree of confidence)

Alternatives to precise probability

- Interval on parameters – aleatory p-box
  - Intervals come from experts
  - How to combine multiple parameters not clear
  - Quick to propagate and may be sufficient

$$X \sim N((\underline{\mu}, \overline{\mu}), \sigma^2)$$

- Bounded probability for epistemic uncertainty – epistemic "p-box"
  - Updating possible by Bayesian inference under a set of priors
  - Can be combined with precise probability
  - Uncertainty in parameters or quantity of interest turns into an optimisation problem

$$\mu \sim N(\mu_0, \sigma_0) \text{ where } (\mu_0, \sigma_0) \in \mathbf{M}$$

This is what I call

Robust Bayesian inference

Types of bounded probability

Examples

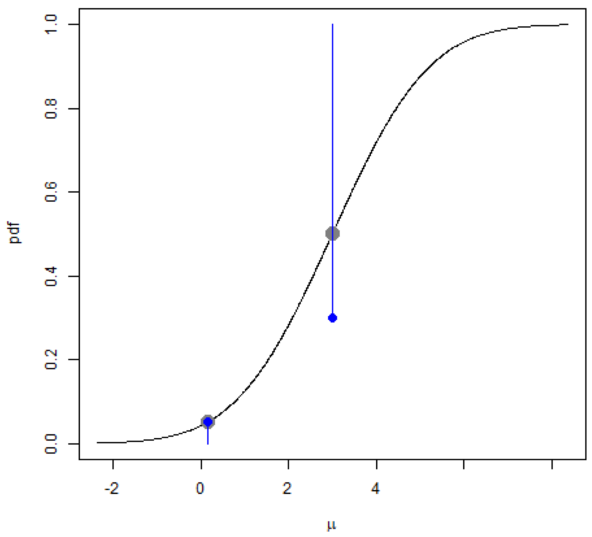
- Interval probability (Keynes, Kyburg, Weichselberger)
- Imprecise probability (Walley)
- Bounds emerging from convex sets of probability distributions (Levi)
- Subjective probability combined with possibility theory (Gärdenfors and Sahlin)
- …

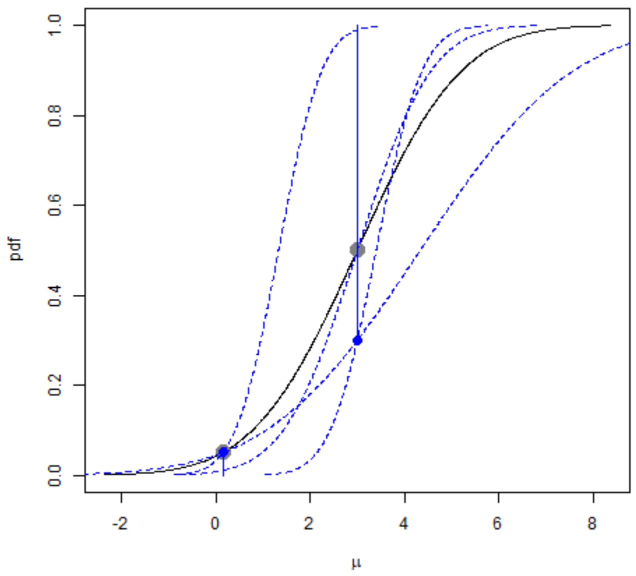
Robust Bayesian inference

- Updating and propagation using Iterative Importance Sampling
- Works well with conjugate models
  - Specify the set of priors from the hyperparameters
  - Search for bounds on uncertainty about the quantity of interest derived from the posteriors
  - No need to derive bounds on the set of posteriors and then propagate
- Methods to combine IIS with MCMC sampling are in development

Eliciting bounded probabilities

- Precise probability:
  - μ < 3 with probability equal to 50%
  - μ > 0.15 with probability 95%
- Bounded probability:
  - μ < 3 with probability greater or equal to 50%
  - μ > 0.15 with probability greater or equal to 95%
  - Define conditions for the set of priors that agree with the bounded probabilites





Confidence theory

- Confidence structures on single parameters with certain coverage properties based on data
- Limited to observables with data
- Do not use expert information
- How to derive confidence structures on combinations of parameters?
- How to derive confidence structure of quantity of interest that is a function of parameters?

References on precise probability

- Kadane – Principles of Uncertainty
- Spiegelhalter – The Art of Statistics – learning from data
- Lindley – Understanding Uncertainty
- McElreath – Statistical Rethinking

LUNDS UNIVERSITET

Ullrika.Sahlin@cec.lu.se

Challenge problems

A handful of random samples were collected to estimate the quantity *X* in units of µg per liter: 12.1, 6.45, 73, 24.6, 15.2, 44.3, 19.0. This variable is believed to be roughly lognormal.

Problem 1.0 What can we say about the chance that *X* exceeds 100?

Problem 1.1 An expert panel provided quantiles representing their uncertainty about the median of the quantity *X* as being 8, 10 and 20 corresponding to the 25%, 50% and 75% quantiles. What can we say now about the chance that X exceeds 100?

Problem 1.2 The chemist reported that the detection limit was 20 µg per liter, so samples of *X* below this value might actually be zeros, and could be as high as the detection limit. What can we say now about the chance that *X* exceeds 100?

Problem 1.3 The samples of *X* were not collected randomly, but rather from a 'hotspot', which implies they may be overestimates, not representative of the true distribution. What can we say now about the chance that *X* exceeds 100?

Problem 1.4 Random sample data is available for a separate but commensurate variable *Y*. The sample values are 2.1, 55, 68, 12, 26, 33, 29, 36, 54, 1.0, 28, 22. What is the chance that $X+Y > 100$?

Problem 1.5 The last six *Y*-values were associated with the samples observed for *X*. What can we say now about the chance that *X* exceeds 100?

Challenge problems

If the chance that a ‘bad thing’ happens in any one trial is p, then the frequency that a bad thing happens in at least one out of N future trials is $Q = 1 - (1 - p)^N$. The event of interest is whether at least one person is infected out of those arriving during a single day, but we are not sure about the frequency p and the number of trials N on any given day. We want to express our uncertainty that Q is greater than 0.05.

The available data about p is that the bad thing happened only once in 153 previous trials. We believe the number of future trials N will be the result of a Poisson process, which in the past has had counts of 12, 0, 21, 14, 6.

Problem 2.0 What is the estimate of Q, and what is our uncertainty that Q is greater than 0.05?

Problem 2.1 There is doubt about whether or not the bad thing actually occurred in the 153 trials. What can be said about Q?

Problem 2.2 There were more sample values previously observed for the Poisson process, but the counts were binned so they are only known as interval ranges. The number N was in the range [0,4] six times, in [5,9] four times, in [10,14] eight times, in [15,19] three times, and once each in [20,24] and [25,29]. (The binned observations were collected before the counts {12, 0, 21, 14, 6} and can be pooled with them.) What can be said about Q?

Problem 2.3 We suspect that the frequency p depends on the number of trials N such that p increases with N. What can be said about Q?