Where do the numbers come from?

Laura Uusitalo

Aim of this talk

- To review the most important sources of probability distributions in BNs
  - Expert knowledge
  - Data
  - Models
  - Literature

- To understand their requirements, benefits and limitations

- Check out also: Aguilera et al 2011
  - https://www.researchgate.net/profile/Rosa_F_Ropero/publication/236985888_Review_Bayesian_networks_in_environmental_modelling/links/5770f95908ae6219474a3032.pdf

Drudzel & van der Gaag 2000:

2 Sources of Probabilistic Information

In most application domains, probabilistic information is available from various sources. The most common are (statistical) data, literature, and human experts. Despite the abundance of information, these sources seldom provide all numbers required for the quantitative part of a probabilistic network. As a consequence, the task of obtaining the numbers for a real-life application is hard and time consuming.

Expert knowledge: overview

- A flexible way of getting even relatively obscure distributions

- However, be extra careful with the model structure & interview setting!
  - Human brain cannot take more than 2 conditioning factors
  - See the next talk about all the ways we can be biased...

Expert knowledge: challenges

- Possibly difficult to get probability distributions out of experts!
  - used to working with real sampling or experimental data → difficult to provide estimates without
  - they may be used to classical statistical analyses: different approach to distributions!

- Uncertainty + superficial knowledge about BNs → distrust → reluctance to provide estimates?

Expert knowledge: practicalities 1/2

- Will the experts provide both the model structure and the parameters (probabilities)?

- If model structure is given
  - It must make sense to the experts!
  - They must all understand the variables in the same way!

- People have cognitive difficulty in thinking of conditional distributions with several conditioning factors (Morgan & Henrion 1990)
  - Rule of thumb: Max two parents per variable!

Expert knowledge: practicalities 2/2

- Together (consensus) or separately?
  - Pro-together: Various points of view to take into account: more balanced assessment?
  - Anti-together: Strong personalities may dominate regardless of expertise

- A mix of the two: joint discussion, separate (confidential) assessments
  - One round technique
  - Delphi method

Expert assessment: the Delphi method

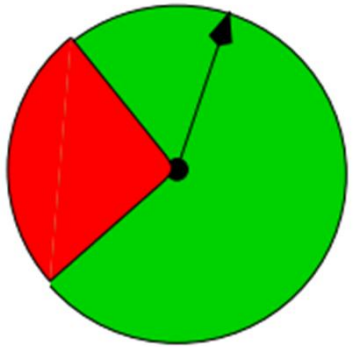
The Delphi method is a structured, consensus-seeking expert panel method:

https://www.youtube.com/watch?v=FFfKOSTftcs

Expert knowledge: probability assessment

There are multiple techniques that aim to help the expert assess the probabilities based on their knowledge:

- Probability wheel
- Probability scales (possibly with linguistic mapping)
- Betting models
- Lottery models



Expert knowledge: What I did in 2000...

Pre-smolt density capasity Expert: _____

Smoltification age	Parr density capacity	Pre-smolt density capacity									
		0-0,5	0,5-1	1-2	2-3	3-4	4-6	6-8	8-12	12-24	24-48
2	0-1										
2	1-2										
2	2-3										
2	3-4										
2	4-6										
2	6-8										
2	8-12										
2	12-16										
2	16-24										
2	24-32										
2	32-48										
2	48-64										
2	64-96										
2	96-128										
3	0-1										
3	1-2										

Data

- There are algorithms to learn conditional probabilities from data when the model structure has been fixed
  - Expectation-Maximization (EM) algorithm is the most widely used, iterative algorithm

- May be easier & more objective than expert knowledge

- However, data may be biased due to sampling, or some combinations are very poorly represented in the data

Models & literature

- More informed than raw data, less biased than experts?

- However, models rarely give exactly what we want
  - Are the parameters the same?
  - Do the models give probabilities or only expected values?
  - Are the probabilities defined in a way we want?

- Check out e.g. Uusitalo et al. 2015
  - http://www.sciencedirect.com/science/article/pii/S1364815214002813

Environmental Modelling & Software 63 (2015) 24–31

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Review

An overview of methods to evaluate uncertainty of deterministic models in decision support

Laura Uusitalo [a, *], Annukka Lehikoinen [b], Inari Helle [c], Kai Myrberg [a]

[a] Finnish Environment Institute SYKE, Marine Research Centre, Mechelininkatu 34a, FI-00251 Helsinki, Finland
[b] University of Helsinki, Department of Environmental Sciences, Fisheries and Environmental Management Group FEM, Kotka Maritime Research Centre, Heikinkatu 7, FI-48100 Kotka, Finland
[c] University of Helsinki, Department of Environmental Sciences, Fisheries and Environmental Management Group FEM, P.O. Box 65, FI-00014 University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 7 November 2013
Received in revised form
18 September 2014
Accepted 22 September 2014
Available online

Keywords:
Environmental modelling
Deterministic

ABSTRACT

There is an increasing need for environmental management advice that is wide-scoped, covering various interlinked policies, and realistic about the uncertainties related to the possible management actions. To achieve this, efficient decision support integrates the results of pre-existing models. Many environmental models are deterministic, but the uncertainty of their outcomes needs to be estimated when they are utilized for decision support. We review various methods that have been or could be applied to evaluate the uncertainty related to deterministic models' outputs. We cover expert judgement, model emulation, sensitivity analysis, temporal and spatial variability in the model outputs, the use of multiple models, and statistical approaches, and evaluate when these methods are appropriate and what must be taken into account when utilizing them. The best way to evaluate the uncertainty depends on the definitions of the

Combination of multiple sources

- Different CPTs can be populated with data from different sources
- Combination of different sources is the most common approach (Aguilera et al. 2011)

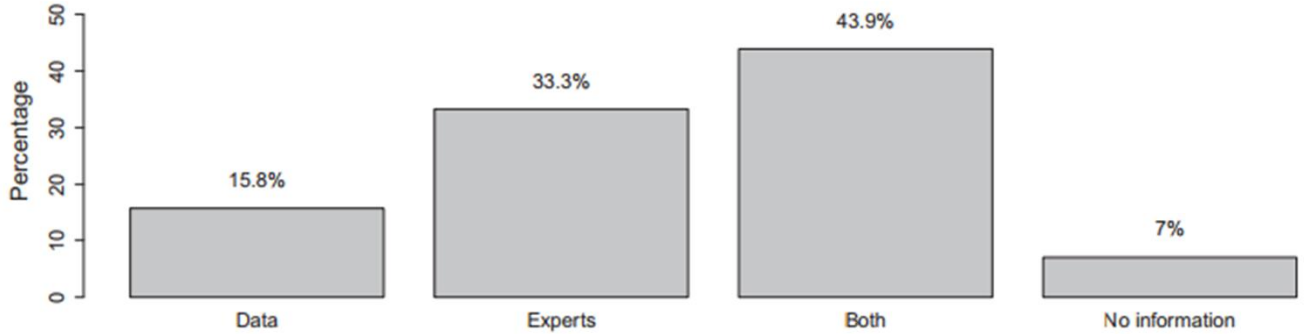


Fig. 7. Model learning in the papers reviewed. Both makes reference to papers that combine expert knowledge and data. No information means the percentage of papers that do not state how the model was learned.