Expert Probabilities in Bayesian Networks for Early Warning

Linda C. van der Gaag

Decision-support Systems Research Group
Department of Information and Computing Sciences,
Utrecht University, The Netherlands

Topic of the talk

The talk discusses the use of expert probabilities in stochastic models for decision support:

- decision-support systems are automated systems that assist people in taking decisions in complex situations;
- modern decision-support systems range from simple decision trees to sophisticated stochastic models;
- while commonly constructed from data, some models are built on expert knowledge.

The talk reports experiences with developing a decision-support system for early warning of Classical Swine Fever in pigs.

EPIZONE: Early warning of CSF

In view of the following considerations:

- Classical Swine Fever (CSF) is an infectious disease in pigs with a potential for rapid spread;
- the disease has a low prevalence and often remains undetected for a long time;
- an outbreak of the disease has major socio-economic consequences;

we built an early-warning system for Classical Swine Fever, for use by pig veterinary practitioners throughout the European Union.

The CSF system and its input

Our early-warning system for CSF takes input data:

- from a group of sick pigs, a veterinarian selects up to five individual animals for further inspection;
- for each selected pig, the system asks for a variety of clinical signs:

Ataxia	Fever	Nasal secretion
Conjunctivitis	Huddling	Petechia
Cyanosis	Lack of appetite	Respiratory problems
Diarrhoea	Lethargy	...

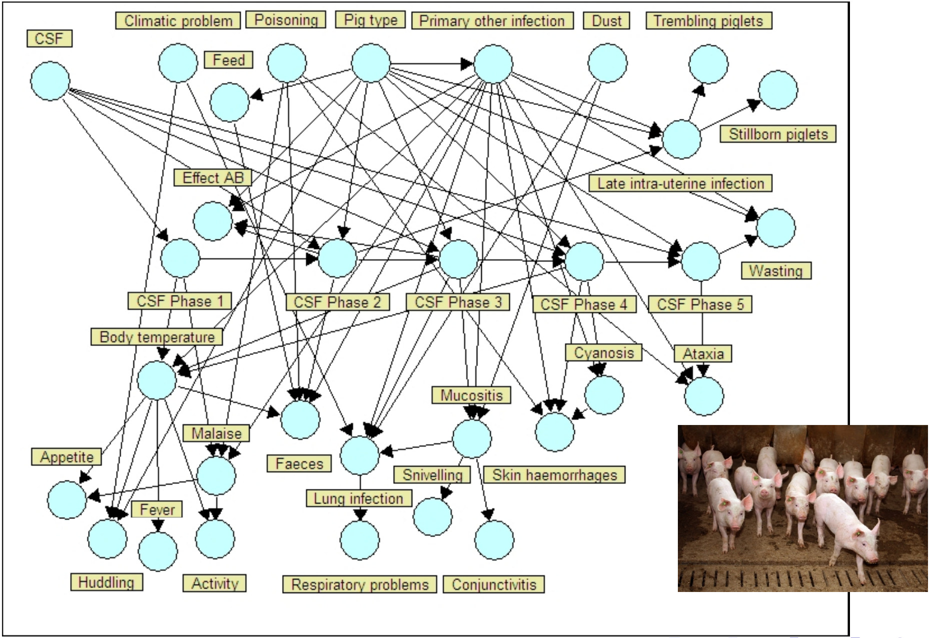
- in addition, the system asks for group summary information:
  - the time of onset of clinical signs;
  - the spread of clinical signs;
  - mortality;
  - ...

The CSF system and its output

Our early-warning system for CSF generates output:

- the system computes the probability of the clinical signs being caused by the CSF virus:
  - for this group of animals, the probability of CSF being present equals 0.07, which is some 4000 times the probability for an arbitrary group of diseased animals.
- based on this probability, the system recommends further actions to be taken:
  - re-visit the farm in a couple of days;
  - send in samples to rule out CSF;
  - call the veterinary authorities;
  - . . .

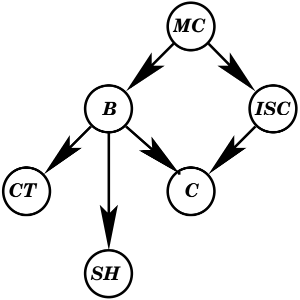
Behind the screens of the CSF system



Bayesian networks in general

A Bayesian network encodes a joint probability distribution in

- a graphical structure, representing the stochastic variables and their (in)dependency relationships;
- an associated set of parameter probabilities.



$$
\begin{aligned}
p(b \mid mc) &= 0.20 \\
p(b \mid \overline{b}) &= 0.05 \\[6pt]
p(isc \mid mc) &= 0.80 \\
p(isc \mid \overline{mc}) &= 0.20 \\[6pt]
p(ct \mid b) &= 0.95 \\
p(ct \mid \overline{b}) &= 0.10
\end{aligned}
\qquad
\begin{aligned}
p(mc) &= 0.20 \\[6pt]
p(c \mid b, isc) &= 0.80 \\
p(c \mid \overline{b}, isc) &= 0.80 \\
p(c \mid b, \overline{isc}) &= 0.80 \\
p(c \mid \overline{b}, \overline{isc}) &= 0.05 \\[6pt]
p(sh \mid b) &= 0.80 \\
p(sh \mid \overline{b}) &= 0.60
\end{aligned}
$$

Algorithms are available for computing probabilities of interest.

The Bayesian network of the CSF system

The CSF early-warning system embeds a Bayesian network with:

- 32 stochastic variables:
  - clinical signs, internal effects of the viraemia, risk factors, alternative explanations, ...
- 63 links between the variables:
  - an elevated body temperature can cause an animal to seek warmth by huddling, ...
- some 1500 (conditional) probabilities:

	Malaise *Body temperature*	no normal	no elevated	yes normal	yes elevated
Appetite	normal	0.995	0.75	0.15	0.10
	decreased	0.005	0.25	0.85	0.90

The construction of the CSF network

The Bayesian network of CSF was constructed by hand:

- literature;
- expert knowledge:
  - in-depth interviews with two pig researchers and with a group of Dutch swine practitioners with CSF experience;
  - case reviews with experts in the six partner countries within the EPIZONE project;
  - written questionnaires with (practising) experts from all countries involved;
- a small collection of data.

The expert meetings

To gather detailed knowledge for our model, we interviewed multiple experts in plenary meetings per country:

The expert meetings

To gather detailed knowledge for our model, we interviewed multiple experts in plenary meetings per country:

We studied within-expert and between-experts properties of the probability assessments obtained.

Probability assessments from multiple experts

The following probabilities were requested from all experts, in the displayed order:

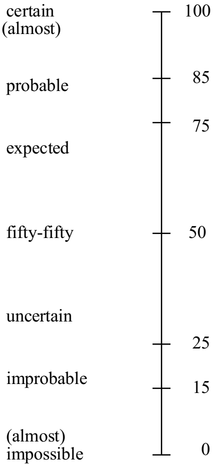
Probability
$p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$
$p_2 = \Pr(\text{conjunct} \mid \text{csf, resp})$
$p_3 = \Pr(\text{conjunct} \mid \text{csf, intest})$
$p_4 = \Pr(\text{conjunct} \mid \text{csf, resp+intest})$
$p_5 = \Pr(\text{sniv} \mid \text{muco})$
$p_6 = \Pr(\text{sniv} \mid \text{no-muco})$

The requested probabilities are mathematically independent, yet not unrelated from a domain perspective.

The elicitation method used

For the assessment task, we used a tailored elicitation method:

Consider a pig *without an infection of the mucous* in the upper respiratory tract. How likely is it that this pig is *snivelling* ?



The method was demonstrated through a plenary instruction.

The assessments obtained

For the probability $p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$, the following assessments were obtained:

Country	*Assessments*							*Range*	*Mean*
$\mathcal{A}$	0.60	0.75	0.75	0.75	0.80			[0.60, 0.80]	0.73
$\mathcal{B}$	0.30	0.40	0.50	0.71	0.75	0.85		[0.30, 0.85]	0.59
$\mathcal{C}$	0.15	0.15	0.20	0.25	0.30			[0.15, 0.30]	0.21
$\mathcal{D}$	0.40	0.50	0.75	0.90	0.95			[0.40, 0.95]	0.70
$\mathcal{E}$	0.70	0.75	0.79					[0.70, 0.79]	0.75
$\mathcal{F}$	0.15	0.34	0.50	0.64	0.75	0.75	0.79	[0.15, 0.79]	0.56
NL	0.29								

Analysis of variance showed that:

- the null hypothesis of equal country means was rejected at a significance level of 0.05;
- upon post-hoc testing, the country mean of $\mathcal{C}$ was found to differ significantly from those of $\mathcal{A}$ and $\mathcal{E}$.

The assessments obtained

The analyses per probability showed very little consensus per country and across countries:

- the results may have been influenced by the varying levels and expertise of the assessors:
  - people reason about probabilities by mentally considering and counting possibilities;
- the results may have been influenced by uncontrolled factors:
  - language barriers, attitudes of the experts, atmosphere in the group, remarks out loud;
- the results may have arisen from actual differences in pig husbandry between the countries . . .

Expected qualitative orderings

We consider again:

Probability
$p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$
$p_2 = \Pr(\text{conjunct} \mid \text{csf, resp})$
$p_3 = \Pr(\text{conjunct} \mid \text{csf, intest})$
$p_4 = \Pr(\text{conjunct} \mid \text{csf, resp+intest})$
$p_5 = \Pr(\text{sniv} \mid \text{muco})$
$p_6 = \Pr(\text{sniv} \mid \text{no-muco})$

- the requested probabilities are mathematically independent;
- the probabilities are not unrelated: based upon domain knowledge, the following orderings are expected:
  - $p_1 \preceq p_3 \preceq p_2 \preceq p_4$
  - $p_6 \preceq p_5$

The qualitative orderings obtained

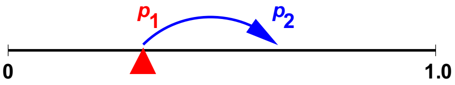
We investigated the qualitative orderings, per assessor, of:

Probability
$p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$
$p_2 = \Pr(\text{conjunct} \mid \text{csf, resp})$
$p_3 = \Pr(\text{conjunct} \mid \text{csf, intest})$
$p_4 = \Pr(\text{conjunct} \mid \text{csf, resp+intest})$
$p_5 = \Pr(\text{sniv} \mid \text{muco})$
$p_6 = \Pr(\text{sniv} \mid \text{no-muco})$

- the ordering $p_6 \preceq p_5$ was found with 97% of the assessors;
- the ordering $p_1 \preceq p_3 \preceq p_2 \preceq p_4$ was found with 62% of the assessors . . .

The qualitative orderings obtained

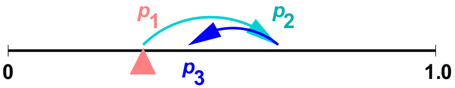
Using anchoring and adjusting as a heuristic, people choose a relevant probability for an anchor to tie their assessment to by adjustment:



- for the ordering $p_1 \preceq p_3 \preceq p_2 \preceq p_4$, some 65% of the violations were caused by an adjustment
  - in the expected direction,
  - yet not by the expected (relative) amount.
- the expected pairwise orderings were found with 86% of the assessors.

The qualitative orderings obtained

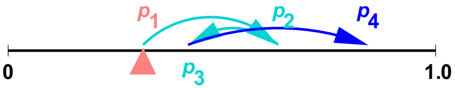
Using anchoring and adjusting as a heuristic, people choose a relevant probability for an anchor to tie their assessment to by adjustment:



- for the ordering $p_1 \preceq p_3 \preceq p_2 \preceq p_4$, some 65% of the violations were caused by an adjustment
  - in the expected direction,
  - yet not by the expected (relative) amount.
- the expected pairwise orderings were found with 86% of the assessors.

The qualitative orderings obtained

Using anchoring and adjusting as a heuristic, people choose a relevant probability for an anchor to tie their assessment to by adjustment:



- for the ordering $p_1 \preceq p_3 \preceq p_2 \preceq p_4$, some 65% of the violations were caused by an adjustment
  - in the expected direction,
  - yet not by the expected (relative) amount.
- the expected pairwise orderings were found with 86% of the assessors.

Violation of expected orderings

We consider again:

Probability
$p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$
$p_2 = \Pr(\text{conjunct} \mid \text{csf, resp})$
$p_3 = \Pr(\text{conjunct} \mid \text{csf, intest})$
$p_4 = \Pr(\text{conjunct} \mid \text{csf, resp+intest})$

If the assessments for the parameters violate the expected ordering

$$p_1 \preceq p_3 \leq p_2 \preceq p_4$$

the results computed from the network may be counterintuitive.

Violation of expected orderings

We consider again:

Probability	
$p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$	$= 0.58$
$p_2 = \Pr(\text{conjunct} \mid \text{csf, resp})$	$= 0.67$
$p_3 = \Pr(\text{conjunct} \mid \text{csf, intest})$	$= 0.56$
$p_4 = \Pr(\text{conjunct} \mid \text{csf, resp+intest})$	$= 0.72$

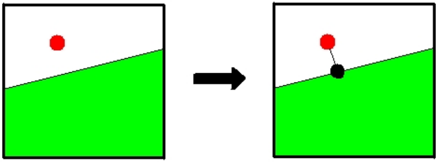
If the assessments for the parameters violate the expected ordering

$$p_1 \preceq p_3 \leq p_2 \preceq p_4$$

the results computed from the network may be counterintuitive.

Exploiting expected orderings

Expected orderings can be used as hard constraints on probability assessments:



Isotonic regression is a general technique for correcting ordering violations:

- the resulting probability assessments are guaranteed to show the expected qualitative orderings;
- the resulting assessments have minimum-distance properties.

Isotonic regression (special case)

Isotonic regression is a statistical technique for estimating parameters $Z = \{z_1, \ldots, z_n\}$, $n \geq 2$, constrained by a qualitative ordering $\preceq$:

- a real-valued function g on Z is isotonic with respect to $\preceq$ if for any $z, z' \in Z$:

$$z \preceq z' \quad \longrightarrow \quad g(z) \leq g(z')$$

- given a real-valued function f on Z, isotonic regression computes the function g^* which minimizes

$$\sum_{i=1}^{n} (f(z_i) - g(z_i))^2$$

over all isotonic functions g on Z.

Applying isotonic regression

For our totally ordered probabilities, the pool adjacent violators (PAV) algorithm constructs the isotonic regression function g^*:

- iterate over the expert assessments $f(p_i)$ in increasing order:
  - if the subsequent assessments $f(p_j), f(p_i)$ violate the ordering, replace them by their average $g(p_j) = g(p_i)$;
  - continue pooling assessments in decreasing order until the ordering is satisfied;

after which all ordering violations have been corrected.

An example:

Probability
$p_1 = 0.75$
$p_3 = 0.85$
$p_2 = 0.75$
$p_4 = 1.00$

Applying isotonic regression

For our totally ordered probabilities, the pool adjacent violators (PAV) algorithm constructs the isotonic regression function g^*:

- iterate over the expert assessments $f(p_i)$ in increasing order:
  - if the subsequent assessments $f(p_j), f(p_i)$ violate the ordering, replace them by their average $g(p_j) = g(p_i)$;
  - continue pooling assessments in decreasing order until the ordering is satisfied;

after which all ordering violations have been corrected.

An example:

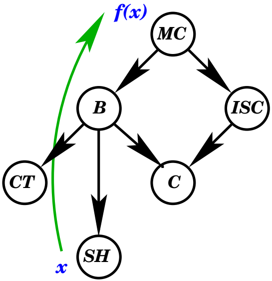
Probability		
$p_1 = 0.75$		
$p_3 = 0.85$	$\rightarrow$	0.80
$p_2 = 0.75$	$\rightarrow$	0.80
$p_4 = 1.00$		

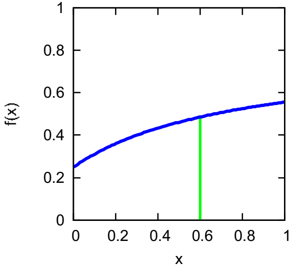
Applying isotonic regression

After applying isotonic regression per assessor, the following parameter probabilities result for the CSF network:

Probability			
$p_1 = \Pr(\text{conjunct} \mid \text{csf, no-other})$	$= 0.58$	$\rightarrow$	0.54
$p_2 = \Pr(\text{conjunct} \mid \text{csf, resp})$	$= 0.67$	$\rightarrow$	0.68
$p_3 = \Pr(\text{conjunct} \mid \text{csf, intest})$	$= 0.56$	$\rightarrow$	0.56
$p_4 = \Pr(\text{conjunct} \mid \text{csf, resp+intest})$	$= 0.72$	$\rightarrow$	0.73

Sensitivity analysis is a general technique for studying the effects of parameter inaccuracies on the output of a mathematical model:





For a Bayesian network,

- a specific parameter probability x is varied;
- a specific output probability of interest is expressed as a function $f(x)$ of x.

The general form of a sensitivity function

For a Bayesian network, the effects of parameter variation are described by a sensitivity function $f(x)$:

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d}$$

where

- $x \in [0, 1]$ is the parameter probability under study;
- $f(x) \in [0, 1]$ denotes a (prior or posterior) output probability;
- a, b, c, d are constants built from the network's other parameters.

Efficient algorithms are available for computing the constants in $f(x)$.

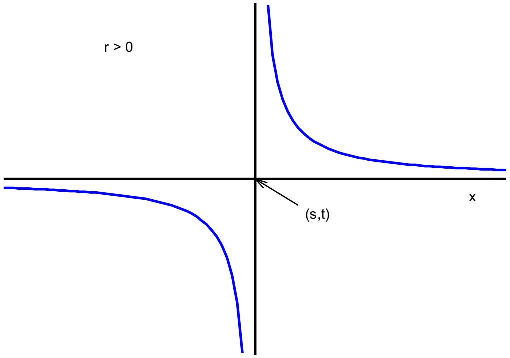
The general form of a sensitivity function

A sensitivity function takes the form of a part of a hyperbola branch:

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d} = \frac{r}{x - s} + t$$

$$\text{with} \quad s = -\frac{d}{c}, \quad t = \frac{a}{c}$$

$$r = \frac{b \cdot c - a \cdot d}{c^2}$$



- the hyperbola has the two asymptotes $x = s$ and $f(x) = t$;
- the vertex of the hyperbola is the point $(x, f(x))$ where $|f'(x)| = 1$.

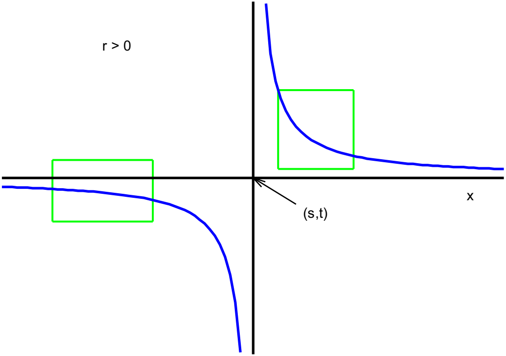
The general form of a sensitivity function

A sensitivity function takes the form of a part of a hyperbola branch:

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d} = \frac{r}{x - s} + t$$

with $\quad s = -\frac{d}{c}, \quad t = \frac{a}{c}$

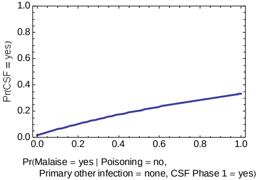
$$r = \frac{b \cdot c - a \cdot d}{c^2}$$



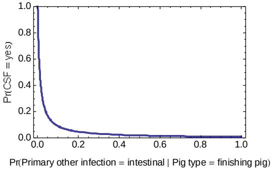
- the hyperbola has the two asymptotes $x = s$ and $f(x) = t$;
- the vertex of the hyperbola is the point $(x, f(x))$ where $|f'(x)| = 1$.

Some example sensitivity functions

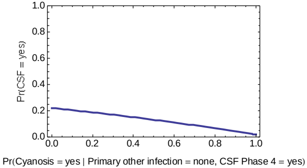
Some sensitivity functions from the CSF network are:



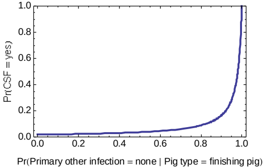
$s = -53, t = 1$



$s = -0.01, t = 0$



$s = 4.82, t = 1$



$s = 1.02, t = 0$

Robustness information from sensitivity functions

Robustness pertains to the stability of a network's output in terms of the assessments for its parameter probabilities:

- the output is robust if varying the network's assessments reveals little effect on the output; otherwise, it is not robust.

A sensitivity function conveys robustness information for a single output probability, through

- the value of its first derivative at the original assessment for the parameter under study;
- the location of its vertex relative to this original assessment.

The sensitivity value

Consider a sensitivity function $f(x)$ and its first derivative $f'(x)$:

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d} \qquad\qquad f'(x) = \frac{a \cdot d - b \cdot c}{(c \cdot x + d)^2}$$

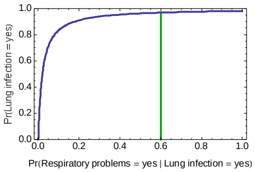
The value $|f'(x_0)|$ for the original assessment x_0 for the parameter x, is the sensitivity value for x_0:

- if $|f'(x_0)| > 0$, then the output probability is sensitive to deviations of x from x_0;
- the larger the sensitivity value, the stronger the effect of deviations from x_0 can be.

Some example sensitivity values

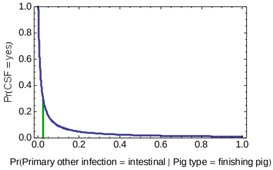
Sensitivity analysis of the CSF network revealed:

- many parameters with small sensitivity values:



the sensitivity value at $x_0 = 0.6$ is 0.052

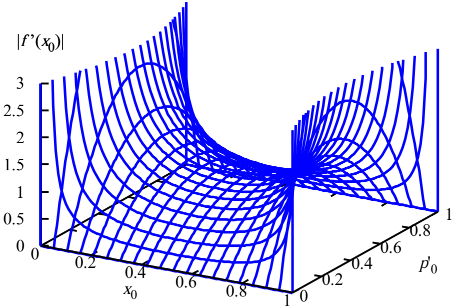
- some parameters with quite large sensitivity values:



the sensitivity value at $x_0 = 0.025$ is 8.163

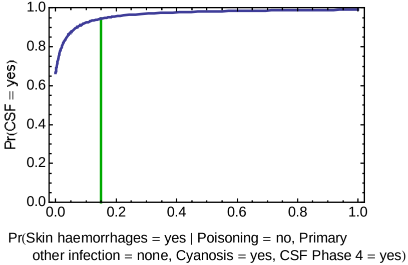
Bounds on the sensitivity value

The sensitivity value is highly dependent of the original assessment x_0 and the associated original output probability p_0:



Large sensitivity values are found only for the more extreme x_0.

The usefulness of the sensitivity value



The sensitivity value at $x_0 = 0.15$ equals 0.31, which suggests little effect on the output probability:

- deviations from x_0 to larger values indeed have little effect;
- deviations from x_0 to smaller values, however, can have a considerable effect !

Extra information from vertex proximity

From a sensitivity function $f(x) = \dfrac{a \cdot x + b}{c \cdot x + d}$, the vertex $(x_s, f(x_s))$ has

$$x_s = \frac{-d \pm \sqrt{|a \cdot d - b \cdot c|}}{c}$$

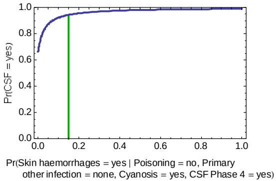
Now,

- if the assessment x_0 for x is close to x_s, then the sensitivity value at x_0 is not a good approximation of the effect of parameter variation;
- the further x_0 lies from x_s, the better the sensitivity value describes the effect of deviations from x_0.

Some example locations of the vertex

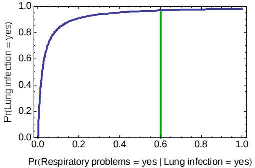
Sensitivity analysis of the CSF network revealed:

- various parameters whose assessment x_0 lies close to x_s:



with $x_0 = 0.15$, the x-coordinate of the vertex is 0.0185

- various parameters whose assessment x_0 lies away from x_s:



with $x_0 = 0.6$, the x-coordinate of the vertex is 0.0338

Concluding observations

For some fields of application, Bayesian networks have to rely on expert probabilities:

- expert probabilities are inaccurate and include biases;
- expert probabilities show very little consensus numerically.

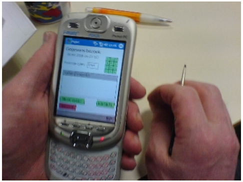
Several techniques are available for studying and reducing the effects of inaccurate expert probabilities:

- isotonic regression enforces more robust qualitative ordering information to hold;
- sensitivity analysis shows which parameter probabilities require further elaboration.

Wrapping up . . .

To study the performance of the CSF network, data were gathered from individual animals using a standardised protocol:

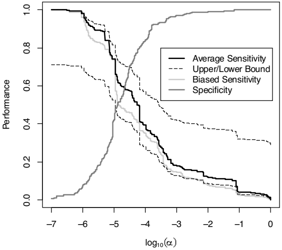
- data from experimental infections:
  - experimental infection studies in Denmark, Germany and the Netherlands;
  - data were recorded every two or three days;
  - for each animal, some 15 clinical signs were scored;
- field data:



  - practitioners from the Netherlands, Belgium, Denmark, Germany and Italy;
  - up to five animals per pen;
  - for each animal, 15 clinical signs were scored and the most likely diagnosis was recorded.

Wrapping up ...

Initial evaluation results for individual animals are promising:



cut-off value	*specificity*	*sensitivity*
0.00001	0.42	0.74
0.00005	0.77	0.52
0.0001	0.84	0.39
0.0005	0.95	0.23
0.001	0.97	0.18
0.005	0.99	0.15

where sensitivity values have been corrected for differences in population and environment conditions.