Discussion of Information Integration Using Belief Functions

Alyson Wilson, Ph.D.
Statistical Sciences Group, D-1
February 28, 2002

Information Integration Technology

F-22 SPO/Seek Eagle

DELPHI
Automotive Systems
Driving Tomorrow's Technology

Up Front

GONE FISSION

DR. SPOCK MEETS DR. STRANGELOVE

BALLISTIC MISSILE DEFENSE
DEPARTMENT OF DEFENSE

P&G

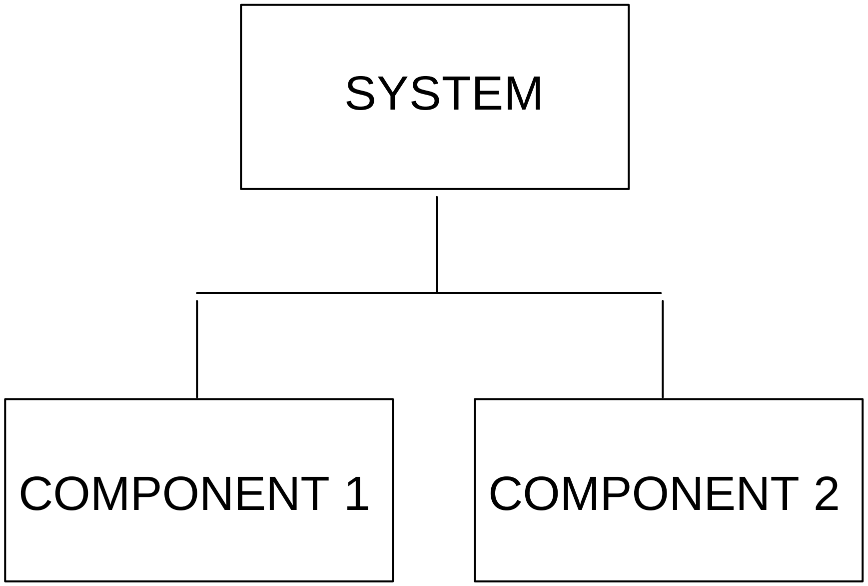
AMCOM/RDEC

BUSINESS WEEK / July 10, 2000

Why “Non-Traditional” UQ?

Historically, expert judgment elicitation has been conceived as working from what experts know (or believe or can predict) to probabilities that statisticians can use for calculations. Some types of expert knowledge may translate well into probability and some might not. So instead of trying to shoehorn everything into probabilities, a person eliciting expert judgment might recognize that the expert seems to think about the issue in a particular way and might consider using a particular type of uncertainty quantification in the analysis of the information.

Why "Non-Traditional" UQ?



Simple (Bayesian) Reliability Problem

Problem: Estimate the probability that a part will survive until time t

- A series of parts, $X_1, \ldots, X_n$ are selected and tested
- Specify a model/likelihood $X_i \sim$ Bernoulli(p_S)
- Specify a prior distribution $\pi(p_S)$
- Use Bayes Rule and update the information about p_S from $\pi(p_S)$ to $\pi(p_S | x_1, \ldots, x_n)$

"Indirect Data" Problem

Problem: Estimate the probability that a part will survive until time t

- A series of parts, $Y_1, \ldots, Y_n$ are selected and it is observed whether or not the anomaly is present

- Specify a model/likelihood $Y_i \sim \text{Bernoulli}(p_A)$

Still want to know about p_S

One Solution

Suppose we know that

Anomaly → Survive to t ("Success")

No Anomaly Fail before t ("Failure")

Using the Law of Total Probability, write

$$
\begin{aligned}
P(\text{survive to t}) &= p_S \\
&= P(\text{survive}|\text{anomaly})P(\text{anomaly}) + \\
&\quad P(\text{survive}|\text{no anomaly})P(\text{no anomaly}) \\
&= p_{S|A}p_A + (1 - p_A)
\end{aligned}
$$

Bayesian Solution Continued

- Now there are two prior parameters instead of one, but this just means I need a prior distribution $\pi(p_A, p_{S|A})$.
- I will probably assume that p_A and $p_{S|A}$ are independent.

This is the strength and weakness of the Bayesian approach. Everything works. I collect data, use Bayes rule, update. If I assume independence, the data tells me nothing about $p_{S|A}$, so I'd better be very careful about my prior specification.

How Does Dempster-Shafer Relate to Probability

Start with a probability space (X, A, μ) and a measurable space (Y, B).

Define a mapping T: X→ Y. The mapping is measurable if

$$\{x \in X : Tx \in b, b \in B\} \in A$$

And we define a measure on (Y,B) by

$$P(b) = m(\{x \in X : Tx \in b\})$$

If Y is the real line, then we call the map T a random variable.

How Does Dempster-Shafer Relate to Probability

Let 2^Y denote the power set of Y (the set of all subsets of Y) and define a sigma field C for 2^Y, so that $(2^Y, C)$ is a measurable space.

Define a measurable map $\Gamma: X \to 2^Y$. Γ induces a probability measure on $(2^Y, C)$. Since Γ is a map from a probability space to a subset of Y (instead of to a single element of Y), it is called a random set. (Some authors call the induced probability measure the random set.)

For Dempster-Shafer applications, what we usually want to talk about is $(Y, 2^Y)$, not $(2^Y, C)$.

How Does Dempster-Shafer Relate to Probability

Now some hand-waving occurs on my part for the infinite case, and even a little bit for the finite case.

For the finite case, if $C = 2^{2^Y}$, and you consider the mass assigned only to the elements of 2^Y, you do have a “measure” induced by Γ on $(Y, 2^Y)$. It is not, however, a probability measure—it is called the “basic probability assignment.”

How Does Dempster-Shafer Relate to Probability

There are basically three set functions of interest that involve Γ when you want to work with $(Y, 2^Y)$:

The basic probability assignment:

$$m(S) = m(\{x \in X : \Gamma(x) = S, S \in 2^Y\})$$

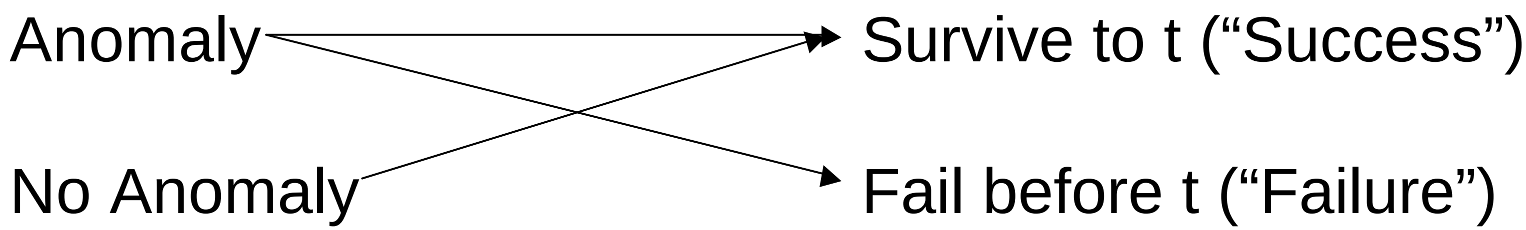
The belief function:

$$\begin{aligned} Bel(S) &= m(\{x \in X : \Gamma(x) \subseteq S, S \in 2^Y\}) \\ &= \sum_{R \subseteq S} m(R) \end{aligned}$$

The plausability function:

$$\begin{aligned} Pl(S) &= m(\{x \in X : \Gamma(x) \cap S \neq \emptyset, S \in 2^Y\}) \\ &= \sum_{R \cap S \neq \emptyset} m(R) \end{aligned}$$

Back to the Reliability Problem



This looks like our “multi-valued map” Γ from the previous discussion.

Thus, we can define belief and plausability functions:

$\mathrm{Bel}(\{\text{success}\}) = 1 - p_A$
$\mathrm{Pl}(\{\text{success}\}) = 1$

Interesting Questions

Interpretation of the belief and plausability functions:

- No matter what I choose for $p_{S|A}$, the distribution of p_S is bounded by these belief and plausability functions. (This is the "lower and upper bounds of some probability distribution" interpretation.)

- The probability of all evidence supporting success is 1 - p_A; the probability of all evidence under which success is possible is 1. (This is the "degree if support" interpretation.)

Are these the only (and/or best) interpretations?

Interesting Questions

The problem that I outlined here seems to be very similar to the kinds of problems we keep wanting to try to use Dempster-Shafer for:

- Start with parameterized distribution for something I can collect data for, but don't really want to know about
- Have a multi-valued map from this "indirect" information to the thing I really care about
- End up with a parameterized basic probability assignment/belief-plausibility function that I need to interpret

Interesting Questions

- Are there other natural statistical applications?
- How does this relate to the question of the elicitation and quantification of expert judgment and how we use it for statistical analysis?