LA-UR-02-1255
*Approved for public release;
distribution is unlimited.*

Title:

Information Integration Using Belief Functions

Author(s):

Alyson G. Wilson

Submitted to:

http://lib-www.lanl.gov/cgi-bin/getfile?00796754.pdf

Information Integration Using Belief Functions

Alyson G. Wilson
Statistical Sciences Group
Los Alamos National Laboratory
P.O. Box 1663, MS F600
Los Alamos, NM 87545-0600
USA
agw@lanl.gov

Abstract

This paper examines the use of belief functions (also known as *Dempster-Shafer methods*) in statistical reliability problems. Starting from the standard Bayesian model for estimating the survival probability in a binomial model, the problem is changed slightly to introduce indirect information. A Bayesian and a Dempster-Shafer approach are proposed for the new problem. The basic properties of the Dempster-Shafer method are discussed, along with connections to the theory of random sets.

1 Introduction

Consider the problem of estimating the probability that a part will survive until time t. Call the survival probability p_s, and assume that an exchangeable sample of parts, $X_1, \ldots, X_n$, is selected and tested. The standard Bayesian model for estimating p_s specifies a likelihood from $X_i \sim \text{Bernoulli}(p_s)$ and a prior distribution $\pi(p_s)$ for p_s. After observing each $X_i = x_i$, the prior distribution is updated to a posterior distribution, $\pi(p_s|x_1, \ldots, x_n)$, using Bayes rule.

Now consider a modified problem. Suppose that instead of observing X_i, whether the part i survives until time t, one observes Y_i, whether an anomaly is present in part i. Also assume the following relationships:

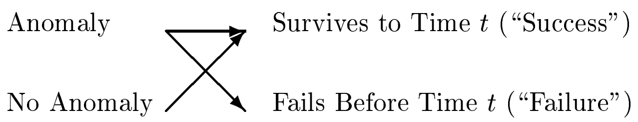

Figure 1: Anomaly to Survival Relationship.

If an anomaly is present, it is possible that the part might survive until time t or might fail before t; if no anomaly is present, the part will survive until time t. This is clearly a simplistic example, but it will serve to motivate the differences between the Bayesian and Dempster-Shafer approaches.

2 A Bayesian Approach

In the modified problem, no "direct" information about p_s is observed, so the standard Bayesian approach must be changed. Assume that an excahngeable sample of parts is selected and tested, and that data, $Y_1, \ldots, Y_n$, are observed about the presence or absence of an anomaly. Model $Y_i \sim \text{Bernoulli}(p_a)$. Since interest still centers on p_s, the probability of surviving to time t, use the Law of Total Probability to write $p_s = P(\text{survives to } t|\text{anomaly})p_a + P(\text{survives to } t|\text{no anomaly})(1 - p_a)$.

Let $p_{s|a} = P(\text{survives to } t|\text{anomaly})$. Notice from Figure 1 that $P(\text{survives to } t|\text{no anomaly}) = 1$ and thus $p_s = p_{s|a}p_a + 1 - p_a$. To complete the Bayesian modeling, a joint prior distribution must be specified on two of p_s, p_a, and $p_{s|a}$; the third has its prior distribution induced by the functional relationship.

In many situations, it might be reasonable to assumes that p_a and $p_{s|a}$ are independent. If this is the case, then observing data y_i about the presence of absence of an anomaly will not change the marginal posterior distribution of $p_{s|a}$, but will update the probability of p_a and p_s. Since the likelihood is free of $p_{s|a}$, the model is termed not *Bayesian identifiable* (Gelfand and Sahu 1999).

Herein lies both the strength and the weakness of the Bayesian approach. Specifying a prior distribution for p_a and $p_{s|a}$ provides the explicit mechanism for updating p_s when collecting "indirect" data about anomalies instead of "direct" data about survival and failure. However, the data does not provide additional information about $p_{s|a}$. The calculation of p_s is dependent on the prior specification for $p_{s|a}$, and this dependence continues even as more data is collected. In theory this is not a problem—simply specify $p_{s|a}$ so that it "correctly" represents the a priori beliefs about the probability of success given an anomaly. In practice, however, this specification may be difficult or impossible, and this leads to consideration of a Dempster-Shafer approach to the problem.

3 Basic Properties of Belief Functions

Before discussing a Dempster-Shafer approach to this specific problem, consider the following basic ideas and properties of Dempster-Shafer theory. For simplicity, this discussion focuses on finite sets; for formal treatments of infinite sets, see Kohlas and Monney (1995) and Kramosil (2001).

Start with a finite set Θ; for concreteness, let $\Theta = \{A, B\}$. Let (Θ, S_Θ, μ) be a probability space, with S_Θ a sigma field and μ a probability. Again for concreteness, let $S_\Theta = 2^\Theta = \{\emptyset, \{A\}, \{B\}, \{A, B\}\}$, the power set (set of all subsets) of Θ, and $\mu(\emptyset) = 0$, $\mu(\{A\}) = 0.3$, $\mu(\{B\}) = 0.7$, $\mu(\{A, B\}) = 1$.

Now consider another measurable space defined by a finite set Ω and a sigma field S_Ω. For concreteness, let $\Omega = \{1, 2, 3\}$ and $S_\Omega = 2^\Omega$. Let S_{2^Ω} be a sigma field associated with 2^Ω, so that a second measurable space $(2^\Omega, S_{2^\Omega})$ is defined.

Now define a map $\Gamma : \Theta \to 2^\Omega$. For concreteness, define $\Gamma(A) \to \{1, 2\}$ and $\Gamma(B) \to \{2, 3\}$. The map Γ induces a probability π on the measurable space $(2^\Omega, S_{2^\Omega})$ through the relationship $\pi(E \in S_{2^\Omega}) = \mu(\{\theta \in \Theta : \Gamma(\theta) \in E\})$. For example, $\pi(\{\{1\}, \{1, 2\}\}) = \mu(\{A\}) = 0.3$ and $\pi(\{\{2, 3\}, \{1, 2, 3\}\}) = \mu(\{B\}) = 0.7$.

Notice that this definition of the induced probability on $(2^\Omega, S_{2^\Omega})$ parallels that of a random variable, except that the map Γ is a one-to-many map from $\Theta \to 2^\Omega$ instead of a one-to-one or many-to-one map from $\Theta \to \Omega$. The map Γ is called a *set-valued random variable* or *random set*.

Even in a small problem, it may be difficult to work with the probability distribution induced on $(2^\Omega, S_{2^\Omega})$ either because of the size of S_{2^Ω} or because of the intuitive problems of working with distributions on sets of sets. Consequently, one often considers instead the function $m : 2^\Omega \to [0, 1]$ defined as $m(F \in 2^\Omega) = \mu(\{\theta \in \Theta : \Gamma(\theta) = F\})$. (Notice that in the finite case, this is the induced probability distribution π restricted to singleton subsets of the power set of 2^Ω.) It would be convenient if m defined a probability distribution on $(\Omega, 2^\Omega)$, but it does not. The function m, however, is usually called the *basic probability assignment* or *b.p.a.* on $(\Omega, 2^\Omega)$ induced by Γ. Although the b.p.a. is not a probability distribution, it does have the properties that $m(\emptyset) = 0$ and $\sum_{F \in 2^\Omega} m(F) = 1$.

In this example, the b.p.a. is

$$
\begin{array}{ll}
m(\emptyset) = 0 & m(\{1, 2\}) = 0.3 \\
m(\{1\}) = 0 & m(\{1, 3\}) = 0 \\
m(\{2\}) = 0 & m(\{2, 3\}) = 0.7 \\
m(\{3\}) = 0 & m(\{1, 2, 3\}) = 0.
\end{array}
$$

The basic probability assignment can be used to construct other functions on 2^Ω. In particular, define the following two functions from $2^\Omega \to [0, 1]$. For $F, G \in 2^\Omega$:

$$
bel(F) = \sum_{G \subseteq F} m(G)
$$

$$pl(F) = \sum_{G \cap F \neq \emptyset} m(G).$$

bel is called the *belief function* induced by Γ on 2^{Ω} and *pl* is called the *plasuability function*. *bel* and *pl* can also be written in terms of Γ and μ; in particular,

$$\begin{aligned} bel(F) &= \mu(\{\theta \in \Theta : \Gamma(\theta) \subseteq F\}) \\ pl(F) &= \mu(\{\theta \in \Theta : \Gamma(\theta) \cap F \neq \emptyset\}). \end{aligned}$$

From the concrete example, $bel(\{1,2\}) = 0.3$ and $pl(\{1,2\}) = 1$.

The belief and plausability functions have a variety of properties. In particular, belief functions are monotone capacities of order infinity and are super-additive for disjoint events. See Molchanov (1997) for a more detailed discussion.

4 A Dempster-Shafer Approach

Now consider the application of Dempster-Shafer methods to the "indirect" data reliability problem posed in Section 1. Figure 1 defines a set-valued map $\Gamma(\text{anomaly}) = \{\text{survives to t, fails before t}\}$ and $\Gamma(\text{no anomaly}) = \{\text{survives to t}\}$. This induces the following belief and plausability functions:

$$\begin{aligned} bel(\text{survive to } t) &= 1 - p_a \\ pl(\text{survive to } t) &= 1 \end{aligned}$$

and the following basic probability assignment:

$$\begin{aligned} m(\emptyset) &= 0 \\ m(\{\text{survives to t}\}) &= 1 - p_a \\ m(\{\text{fails before t}\}) &= 0 \\ m(\{\{\text{survives to t}\}, \{\text{fails before t}\}\}) &= p_a. \end{aligned}$$

For Dempster-Shafer theory to be useful in a reliability context, it must be interpretable. There are essentially two schools of interpretation for belief and plausability functions. One interpretation is as "the lower and upper bounds for some unknown probability distribution" (Kohlas and Monney 1995). Under this interpretation, no matter what distribution is selected for $p_{s|a}$, the probability of survival to t, p_s, for any given interval is bounded below by the belief function and bounded above by the plausability function. As more data is collected, the belief function is updated using Bayes rule. This interpretation lends itself to considering the possibility of assigning upper and lower bounds to the probability assigned to $p_{s|a}$ instead of simply assigning no distribution (a vacuous belief function). Assigning upper and lower bounds to the distribution of $p_{s|a}$ also leads to a belief function/plausability function pair for p_s.

The other interpretation of Dempster-Shafer theory is in terms of the "degree of support" for a hypothesis—in this case, that the part will survive until time t. A degree of support of at least $1 - p_a$ is assigned to survival, as any observations of "no anomaly" certainly support survival. However, there is never unequivocal evidence of failure—even in the presence of an anomaly, the part may survive until time t. Since there is no evidence inconsistent with survival until time t, the maximum degree of support is 1.

5 Discussion

The example considered here is representative of the class of problems where we have been considering the use of Dempster-Shafer methods. In particular, suppose that we are interested in one phenomenon, but that we can only collect data about another, related phenomenon; in other words, we have "indirect" data about the phenomenon of interest. The two phenomena are related by a set-valued map—the data that we can collect does not uniquely determine the outcome of interest. This is a natural setting to explore the applicability of Dempster-Shafer methods.

Acknowledgements

This work has benefited greatly from extensive conversations with Sallie Keller-McNulty, Nozer Singpurwalla, and Art Dempster.

References

Gelfand, A. and Sahu, S. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association 94*(445), 247–254.

Kohlas, J. and Monney, P. (1995). *A mathematical theory of hints: An approach to the Dempster-Shafer theory of evidence.* Berlin: Springer-Verlag.

Kramosil, I. (2001). *Probabilistic analysis of belief functions.* New York: Plenum.

Molchanov, I. (1997). Statistical problems for random sets. In J. Goutsias, R. Mahler, and H. Nguyen (Eds.), *Random Sets: Theory and Applications*, The IMA Volumes in Mathematics and its Applications, Volume 97 , pp. 27–46. New York: Springer.