CONTINUOUS BELIEF FUNCTIONS FOR EVIDENTIAL REASONING

Thomas M. Strat
SRI International
333 Ravenswood Ave.
Menlo Park, CA 94025

ABSTRACT

Some recently developed expert systems have used the Shafer–Dempster theory for reasoning from multiple bodies of evidence. Many expert-system applications require belief to be specified over arbitrary ranges of scalar variables, such as time, distance or sensor measurements. The utility of the existing Shafer–Dempster theory is limited by the lack of an effective approach for dealing with beliefs about continuous variables. This paper introduces a new representation of belief for continuous variables that provides both a conceptual framework and a computationally tractable implementation within the Shafer–Dempster theory.

1. Introduction

The lack of a formal semantics for the representation and manipulation of degrees of belief has been a difficulty for expert systems. The frequent need to reason from evidence that can be inaccurate, incomplete, and incorrect has led to the recognition of evidential reasoning as an important component of expert systems [2] [3]. Evidential reasoning, based on a relatively new body of mathematics commonly called the Shafer–Dempster theory, is an extension of the more common Bayesian probability analysis. In the theory, the fundamental measure of belief is represented as an interval bounding the probability of a proposition, thus allowing the representation of ignorance as well as uncertainty. A procedure to pool multiple bodies of evidence expressed in this manner to form a consensus opinion is also provided by the theory.

Expert systems are often applied to situations involving continuous variables such as time, distance, and sensor measurements. Because the Shafer–Dempster theory is defined over discrete propositional spaces, dealing with continuous variables has been approached by partitioning the variable's range into discrete subsets of possible values. In practice however, this approach has two difficulties: conclusions are sensitive to the selected partitioning, and there is no means for specifying belief in a smoothly varying manner.

Belief as well as ignorance about a continuous variable should vary smoothly through the range of possible values. By making an appropriate restriction in the class of propositions, smoothly varying beliefs can be expressed. This restriction motivates a new, continuous representation for belief over continuous variables that is computationally practical and conceptually appealing.

The paper begins with a brief overview of the Shafer-Dempster theory. Section 3 presents a formalism for representing and manipulating evidence about a discretized scalar variable. The representation is generalized to the truly continuous case in Section 4, enabling discourse about any interval of values at any level of detail and permitting the representation of smoothly varying beliefs over those intervals. This is followed by an example which illustrates the new representation and its use. The paper concludes with a discussion of the theory's relevance and extensions.

This research was supported by the Defense Advanced Research Projects Agency under Contract No. N00039-83-K-0656 with the Naval Electronics Systems Command.

2. Review of Shafer–Dempster Theory

Suppose that there is a fixed set of mutually exclusive environmental possibilities

$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}.$$

Any proposition of interest can be represented by the subset of Θ containing exactly those environmental possibilities for which the proposition is true. The collection of all propositions (*i.e.*, the power set of Θ) constitutes the *frame of discernment.* Figure 1 shows the power set of Θ (for $n = 4$) arranged as a tree.

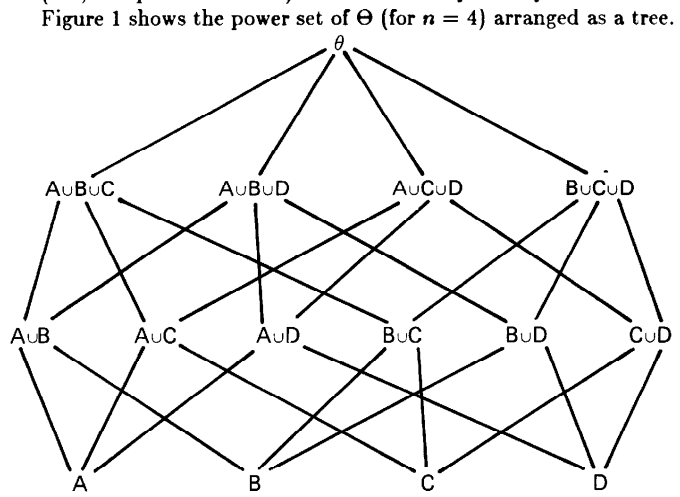


Figure 1: The Frame of Discernment: $\Theta = \{A, B, C, D\}$

The nodes of the tree are the propositions arranged such that each node logically implies its ancestors.

Bodies of evidence (*i.e.*, sets of partial beliefs) are represented by *mass distributions* that distribute a unit of belief (*i.e.*, mass) across the propositions in Θ. In other words, the mass distribution assigns a value of belief in the range $[0,1]$ to each subset of Θ, such that

$$\sum_{F_i \subseteq \Theta} M(F_i) = 1$$

$$M(\phi) = 0$$

where $M(F_i)$ is the mass attributed to proposition F_i. Viewed intuitively, mass is attributed to the most precise propositions a body of evidence supports. If a portion of mass is attributed to a proposition, it represents a minimal commitment to that proposition as well as all the propositions it implies (*i.e.*, nodes higher in the tree). At the same time, that portion of mass remains noncommital with regard to those propositions that imply it (*i.e.*, descendant nodes in the tree).

This representation allows one to specify his belief at exactly the level of detail he desires while remaining noncommital toward those propositions about which he is ignorant. Mass attributed directly to the disjunction of all propositions (*i.e.*, Θ) is neutral with respect to all propositions and represents the degree to which the evidence fails to support anything.

The support for an arbitrary proposition Q, $Spt(Q)$, is the total belief attributed by the mass distribution to propositions that imply Q (*i.e.*, the sum of the mass attributed to Q and all its descendants in the tree).

$$Spt(Q) = \sum_{F_i \subseteq Q} M(F_i)$$

The plausibility, $Pls(Q)$, is the total belief attributed to propositions that do not imply $\neg Q$.

$$\begin{aligned} Pls(Q) &= \sum_{F_i \cap Q \neq \phi} M(F_i) \\ &= 1 - \sum_{F_i \subseteq \neg Q} M(F_i) \\ &= 1 - Spt(\neg Q) \end{aligned}$$

For each proposition Q, a mass function defines an interval $[Spt(Q), Pls(Q)]$ that bounds the probability of Q. The difference $Pls(Q) - Spt(Q)$ represents the degree of ignorance; the probability of Q is known exactly if $Spt(Q) = Pls(Q)$.

Dempster's Rule of Combination pools multiple bodies of evidence represented by mass distributions. It takes arbitrarily complex mass distributions M_1 and M_2, and, as long as they are not completely contradictory, produces a third mass distribution that represents the consensus of those two disparate opinions. The rule moves belief toward propositions that are supported by both bodies of evidence and away from all others.

For all $F_i, F_j, Q \subseteq \Theta$

$$M_3(Q) = \frac{1}{1-k} \sum_{F_i \cap F_j = Q} M_1(F_i) \cdot M_2(F_j)$$

$$k = \sum_{F_i \cap F_j = \phi} M_1(F_i) \cdot M_2(F_j)$$

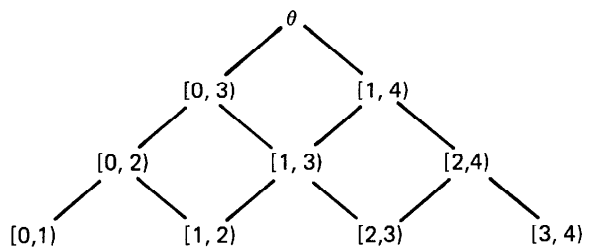


Figure 2: The Frame of Discernment of a Discretized Variable

If $k = 1$, the bodies of evidence represented by M_1 and M_2 are contradictory, and their combination is not defined. It is interesting to observe that Dempster's Rule is both commutative and associative, allowing bodies of evidence to be combined in any order and grouping. A thorough treatment of the Shafer–Dempster theory can be found in Dempster [1] and Shafer [4].

3. Discrete Analysis of a Random Variable

The standard approach to reasoning with continuous variables under the Shafer–Dempster theory has been to associate propositions with portions of the number line. Mass can then be attributed to individual propositions that correspond to arbitrary sets of points on the number line, and mass assignments from disparate sources can be combined using Dempster's Rule by computing the intersections of these sets. This approach has several undesirable properties.

Because mass must be assigned to specific propositions, computations based on such a mass function can be critically sensitive to slight variations in the proposition of interest. For example, $Spt([0,2))$[1] may differ greatly from $Spt([0,1.99))$ if there happens to be mass assigned to a proposition such as $Spt([1,2))$. This type of discontinuity is an artifact of the way the propositional space is discretized and may not be indicative of the underlying beliefs.

Secondly, the traditional approach provides no means for specifying a smoothly varying set of beliefs about the value of a continuous variable. Intuitively, one would prefer a belief function that varies gradually with both the magnitude of the proposition of interest and the level of detail of the proposition.

The following observation provides the key to overcome these difficulties: when reasoning about the value of a continuous variable, expert systems are most often interested in whether or not the value lies within some contiguous range of values. For example, a proposition of interest might be that today's temperature is between 65° and 75°. Rarely does a situation arise in which a disjoint subset would be a proposition of interest (such as "the temperature is either between 45° and 50° or between 70° and 80°"). This observation allows the frame of discernment to be

[1] Here $Spt([0,2))$ denotes the proposition that the value of the variable is in the interval $[0,2)$. We use open-ended intervals for simplicity.

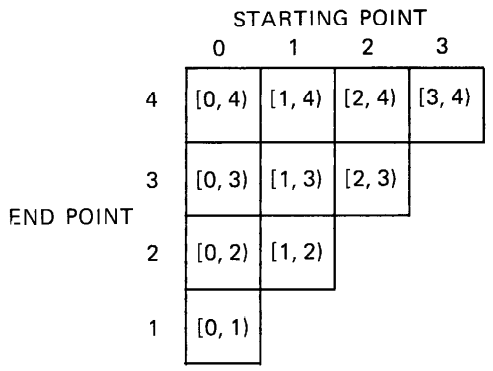
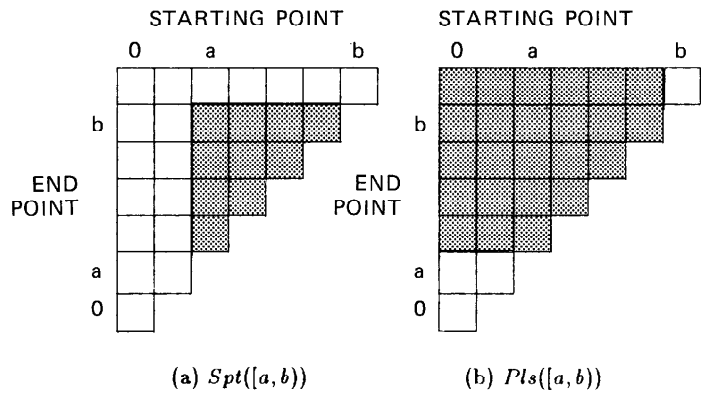


Figure 3: The Mass Function of a Discretized Variable



(a) $Spt([a,b))$ (b) $Pls([a,b))$

Figure 4: Computation of Support and Plausibility — Discrete Case

restricted to contain only contiguous intervals, yet to retain the ability to represent a wide range of interesting propositions.

The restriction provides several powerful simplifications. Imagine dividing the number line from 0 to N into N intervals of unit length. The number of propositions in this frame of discernment is reduced from 2^N (the size of the power set) to approximately $\frac{1}{2}N^2$. Figure 2 depicts the simplified tree. The computation of the intersection of pairs of propositions in Dempster's Rule is reduced to a simple intersection of contiguous intervals. Furthermore, the restricted frame of discernment is a class of subsets which is closed under the application of Dempster's Rule so that pooled evidence can always be represented in the same propositional space (*i.e.*, contiguous intervals).

The structure of the tree suggests the representation of the frame of discernment as a triangular matrix as shown in Figure 3. Here the abscissa specifies the beginning of an interval and the ordinate specifies the endpoint. The set Θ, which represents all the environmental possibilities, is the interval $[0, N)$ and is represented by the upper left-hand entry. The atomic propositions, the intervals of minimum length, are located along the diagonal. Intervals with a common starting point are located in the same column while those with a common endpoint are in the same row. It is easy to see that the matrix of Figure 3 bears a strong resemblance to the tree of Figure 2.

A mass function of a discretized variable can now be represented as a triangular matrix. To assign a mass of .1 to the interval $[2,4)$ for example, we enter .1 at the corresponding location in the matrix. Additional beliefs fill out the remainder of the matrix. As with any mass function, Shafer–Dempster theory requires that the entries in the matrix sum to one.

The computation of $Spt(Q)$ and $Pls(Q)$ can be easily understood graphically. $Spt(Q)$ is the sum of the masses of those intervals wholly contained in Q (the shaded area of Figure 4(a)), and $Pls(Q)$ is the sum of the masses of the intervals whose intersection with Q is not empty (the shaded area of Figure 4(b)). The sum of the masses in the difference of those two regions is the ignorance remaining about proposition Q. Mathematically, (using the obvious notation)[2]

$$Spt([a,b)) = \sum_{x=a}^{b-1} \sum_{y=x+1}^{b} M(x,y)$$

[2]Here we use $M(x,y)$ to represent the mass associated with the interval $[x,y)$.

$$Pls([a,b)) = \sum_{x=0}^{b-1} \sum_{y=1+\max(a,x)}^{N} M(x,y)$$

Given two mass functions represented by triangular matrices, one can obtain a third mass function that represents the pooled evidence using Dempster's Rule. The mathematics of intersecting sets is straightforward with this representation, and Dempster's Rule can be rewritten as follows:

$$\begin{aligned} M_3(a,b) = \frac{1}{1-k}\Big(&\sum_{x=0}^{a}\sum_{y=b}^{N}[M_1(x,b)\cdot M_2(a,y) + M_2(x,b)\cdot M_1(a,y)] \\ &+\sum_{x=0}^{a-1}\sum_{y=b+1}^{N}[M_1(a,b)\cdot M_2(x,y) + M_2(a,b)\cdot M_1(x,y)] \\ &- M_1(a,b)\cdot M_2(a,b)\Big) \end{aligned}$$

$$k = \sum_{p=0}^{N-2}\sum_{q=p+1}^{N-1}\sum_{r=q}^{N-1}\sum_{s=r+1}^{N}[M_1(p,q)\cdot M_2(r,s) + M_2(p,q)\cdot M_1(r,s)]$$

4. Generalization to Continuous Random Variables

The generalization from a finite number of discrete intervals to an infinite number of infinitesimal intervals is made using the standard ploys of calculus. In the limit as the width of the intervals shrinks to zero, the triangular matrix becomes a triangular region where any interval is represented by its location in Cartesian coordinates.

Let's examine some properties of the region more closely (Figure 5). The universal set Θ (the interval [0, N][3]) is located at the upper left-hand corner. Points along the hypotenuse refer to individual points along the number line. As before, points in the same vertical or horizontal line refer to intervals with identical

[3]We switch to closed intervals for the continuous case to simplify the mathematics. We are no longer concerned with an atomic set of mutually exclusive propositions.

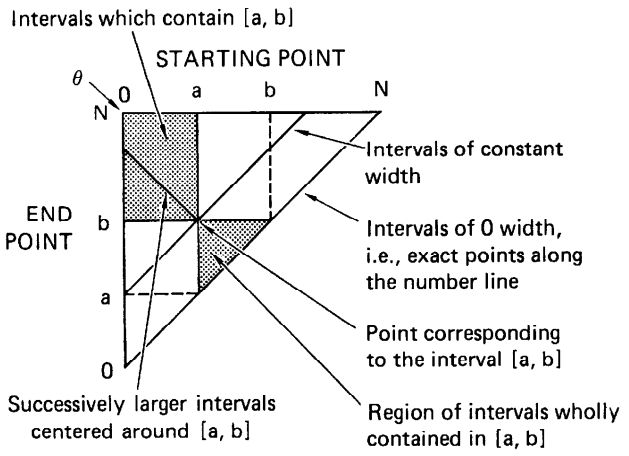


Figure 5: The Continuous Frame of Discernment

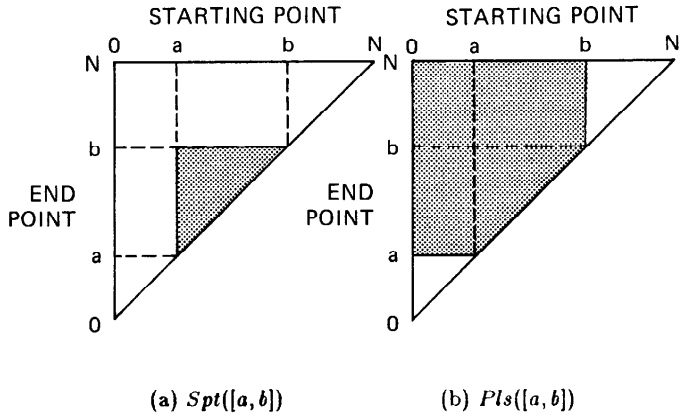


Figure 6: Computation of Support and Plausibility — Continuous Case

start or end points. Points along a northwesterly ray from some point $[a, b]$ correspond to successively larger intervals centered around $[a, b]$. Points along a northeasterly line refer to intervals of identical width, thus representing propositions with a common level of detail. The triangular region is, in a sense, the continuous analog of the tree structure of Figure 2.

A continuous mass function with all the desirable properties mentioned earlier is represented by a surface over this region. The extent to which the volume under the surface is pushed toward the northwest corner (Θ) indicates the overall degree of ignorance. Concentrating all the volume along the hypotenuse corresponds to knowing the probability density function of the variable exactly.

As an example, consider a thermometer that reads 47°. Due to the inherent inaccuracy and imprecision of any measuring device, we could represent our belief in that evidence as a continuous mass function. We would expect that the surface would have a ridge extending from the interval $[47, 47]$ toward Θ in accordance with the imprecision of the thermometer. Inaccuracy would be seen as a lateral spreading of the ridge along a northeast-southwest axis.

Analogously with the discrete case, $Spt([a, b])$ is the volume under the surface within the region shaded in Figure 6(a). Figure 6(b) shows the region containing $Pls([a, b])$. In mathematical terms,

$$Spt([a,b]) = \int_a^b \int_x^b M(x,y)\, dy\, dx$$

$$Pls([a,b]) = \int_0^b \int_{\max(a,x)}^N M(x,y)\, dy\, dx$$

The extension of Dempster's Rule to the continuous case yields the following result:

$$M_3(a,b) = \frac{1}{1-k} \int_0^a \int_b^N [M_1(x,b) \cdot M_2(a,y) + M_2(x,b) \cdot M_1(a,y) + M_1(a,b) \cdot M_2(x,y) + M_2(a,b) \cdot M_1(x,y)]\, dy\, dx$$

$$k = \int_0^N \int_p^N \int_q^N \int_r^N [M_1(p,q) \cdot M_2(r,s) + M_2(p,q) \cdot M_1(r,s)]\, ds dr dq dp$$

This can be construed as a form of convolution of the two mass functions being pooled. As in the discrete case, the resulting mass function can be represented in the same formalism.

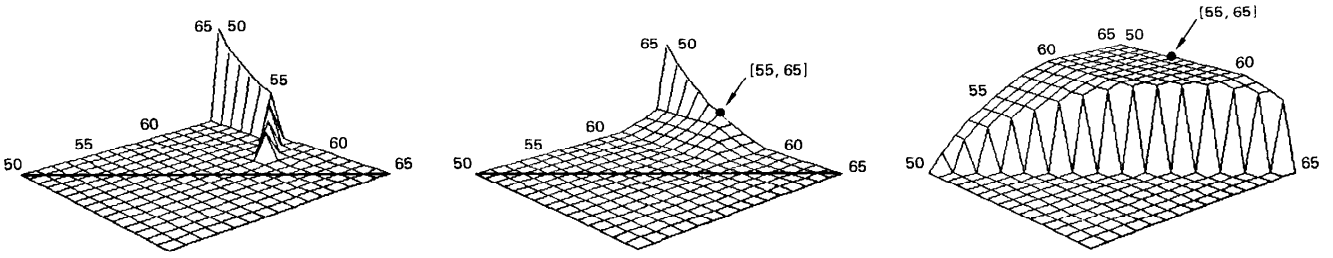
In theory, if we desire to assign mass to a precise interval $[a, b]$, we must use impulse functions of finite volume at the corresponding point. The degree to which we cannot be so precise about the interval represents the degree to which the impulse is spread out to neighboring points. If impulse functions are present, the rule of combination becomes slightly more complex since we must take care not to count certain combinations doubly. Impulse functions need only be considered when merging discrete with continuous mass functions.

5. Example

We now present a simple example to illustrate the representation and the combination of evidence:

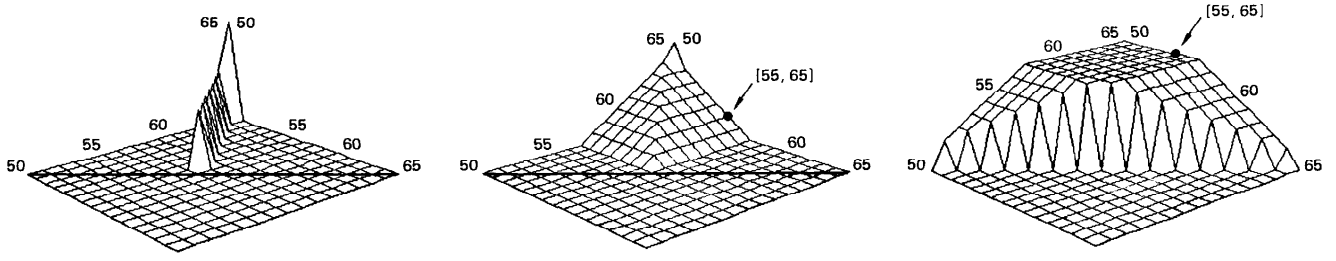
> The state highway patrol is attempting to identify speeders on Interstate 80. A patrolman on a motorcycle observes that his speedometer reads 60 mph when matching speed with a suspected speeder. Meanwhile, a parked patrolman obtains a reading on his radar gun of 57 mph for the same vehicle. Is this sufficient evidence to issue a traffic citation for speeding?

The first thing to do is to construct mass functions for both bodies of evidence. Here we will simply present intuitively reasonable functions; a formal theory for deriving mass functions from sensor measurements is the subject of a future paper. Figure 7(a) depicts the mass function for the motorcycle speedometer reading. The frame of discernment has been restricted to the range from 50 to 65 mph (*i.e.*, $\Theta = [50, 65]$) in order to focus on these values. Values outside that range are considered impossible in this example. The most precise interval that mass has been committed to is $[58, 62]$, indicating that the precision of the patrolman reading his speedometer is no better than ± 2 mph. The remainder of the mass function attributes mass to successively larger intervals centered around 60 mph (until the upper limit of 65 is reached at the bend in the ridge). This represents the unbiased ignorance associated with inaccuracy in the speedometer or with the patrolman not matching speeds properly. Note how



(a) Mass function (b) Support (c) Plausibility

Figure 7: Representation of Evidence from the Speedometer Reading



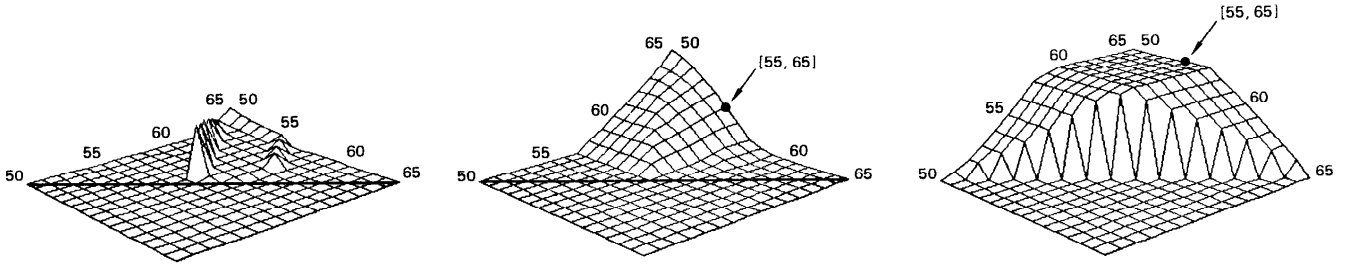
(a) Mass function (b) Support (c) Plausibility

Figure 8: Representation of Evidence from the Radar Gun

this differs from an ordinary probability distribution. The support and plausibility for each interval have been computed from the mass function and plotted in Figures 7(b) and (c). These plots clearly show how the beliefs vary smoothly as the proposition of interest is varied. Support and plausibility both increase monotonically toward one as the interval is widened. The difference between these surfaces at any point represents the ignorance remaining about the probability that the true value lies in the interval corresponding to that point. The support for the proposition "the suspect is speeding" is $Spt([55,65]) = .28$ and $Pls([55,65]) = 1.0$, indicating the probability the car was traveling greater than 55 mph is between .28 and 1.0.

Figure 8(a) shows the mass function for the evidence obtained with the radar gun. Some insight can be gained by comparing it with the speedometer mass function. The ridge, which is centered at 57 mph, is further to the left indicating a lower measured speed. There is more mass near the hypotenuse reflecting a more accurate instrument. There is a peak at Θ indicating the possibility of a gross error that provides no information about the true speed. Based on the evidence from the radar gun, this mass function provides $Spt([55,65]) = .23$ and $Pls([55,65]) = 1.0$. The support and plausibility surfaces are plotted in Figures 8(b) and (c). The values of plausibility along the hypotenuse constitute a curve showing the plausibility of any individual speed. Notice how the curve along the hypotenuse is more peaked in Figure 8(c) than in Figure 7(c), reflecting greater conviction.

Given these two mass functions, Dempster's Rule is used to compute a third mass function representing the combination of the two bodies of evidence (Figure 9). Here, the two ridges are still visible with some mass having been "spread" between the ridges. This shows support for the intermediate values that are common to both bodies of evidence. Additionally, some mass has shifted away from Θ toward the hypotenuse indicating an incremental narrowing of belief. The support and plausibility surfaces show the bounds on the probabilities of all intervals of speed. The support surface has generally risen and the plausibility surface along the hypotenuse has grown more peaked, showing that the combination of evidence has strengthened and refined our beliefs. This combination of evidence yields $Spt([55,65]) = .44$ and $Pls([55,65]) = 1.0$, meaning that there is at least a 44% chance that the car was speeding and that there is no evidence to the contrary. This may still be insufficient evidence to prove the car was speeding. The important point is that the mass function captures exactly those beliefs that are warranted by the evidence,



(a) Mass function (b) Support (c) Plausibility

Figure 9: Representation of Combined Evidence

without overcommitting or understating what is known. Additional evidence can be combined in the same fashion to yield mass functions that may or may not change our belief in propositions about the speed of the car.

6. Discussion

Restricting the frame of discernment to include only contiguous intervals along the number line provides the key to the computational and conceptual simplicity of the framework. In particular, it reduces the space of propositions from $\mathbf{O}(2^n)$ to $\mathbf{O}(n^2)$ where n is the number of atomic possibilities. In most cases, the restriction is a natural one because we would rarely expect to encounter disjoint intervals. Representing the mass function as a two-dimensional surface permits the specification of smoothly varying beliefs. A gradual shift in an interval of interest incurs a gradual change in the associated support for that interval. Similarly, a gradual widening of an interval incurs a gradual increase in support.

As an extension, one may expand the frame of discernment to include intervals that "wrap around" the endpoint N. This enlarged class of subsets would allow the representation of $M(\neg[a,b])$ and is also closed under the application of Dempster's Rule. In this case the triangular mass function becomes a full square (with a discontinuity along the diagonal) and formulas for $Spt(\cdot)$, $Pls(\cdot)$ and Dempster's Rule can be derived in an analogous fashion.

Another extension features the ability to reason over multidimensional regions. This formulation would allow for bounded areas and volumes in the frame of discernment. In the two-dimensional case, propositions of interest are restricted to be rectangles of fixed orientation. This frame of discernment is closed under Dempster's Rule and requires a four-dimensional mass function. Regions of higher dimensionality can be represented but the computational burden becomes large.

The specification of continuous mass functions is a matter for further investigation. One may envision special sensors that provide not a single value, nor a probability density function as output, but a continuous mass function by which they explicitly express their imprecision as well as their uncertainty about the measurement.

Evidential reasoning, as based on the Shafer–Dempster theory, allows belief to be represented at any level of detail and allows multiple opinions to be pooled into a consensus opinion. The ability to reason evidentially over continuous variables is crucial for expert systems that must reach decisions based on uncertain, incomplete, and inaccurate evidence about such quantities as time, distance, and sensor measurements. This paper provides a novel representation that permits a conceptually appealing implementation of Shafer–Dempster theory applied to continuous variables. It provides the means for expressing belief as a continuous function over contiguous intervals of continuously varying widths.

REFERENCES

[1] Dempster, Arthur P., "A Generalization of Bayesian Inference", *Journal of the Royal Statistical Society* **30**(Series B), 1968, pp. 205–247.

[2] Lowrance, John D., and Garvey, Thomas D., "Evidential Reasoning: A Developing Concept", Proceedings of the IEEE International Conference on Cybernetics and Society, October 1982, pp. 6–9.

[3] Lowrance, John D., and Garvey, Thomas D., "Evidential Reasoning: An Implementation for Multisensor Integration", Technical Report TN 307, Artificial Intelligence Center, SRI International, Menlo Park, California, December 1983.

[4] Shafer, Glenn A., *A Mathematical Theory of Evidence*, Princeton University Press, New Jersey, 1976.