JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 176, 521–544 (1993)

Transformation of Mass Function and Joint Mass Function for Evidence Theory in the Continuous Domain[1]

DOUG Y. SUH

HDTV Team, Headquarters of Research and Development, Korea Academy of Industrial Technology, Duru Building, Yangjae-dong, 20-34, Seocho-bu, Seoul, Korea 137-130

AND

AUGUSTINE O. ESOGBUE

Georgia Institute of Technology, School of Industrial and Systems Engineering, Atlanta, Georgia 30332-0205

Received August 12, 1990

It has been widely accepted that expert systems must reason from multiple sources of information that is to some degree evidential—uncertain, imprecise, and occasionally inaccurate—called *evidential* information. Evidence theory (Dempster/Shafer theory) provides one of the most general frameworks for representing evidential information compared to its alternatives such as Bayesian theory or fuzzy set theory. Many expert system applications require evidence to be specified in the continuous domain—such as time, distance, or sensor measurements. However, the existing evidence theory does not provide an effective approach for dealing with evidence about continuous variables. As an extension to Strat's pioneering work, this paper provides a new combination rule, a new method for mass function[2] transformation, and a new method for rendering joint mass functions which are of great utility in evidence theory in the continuous domain. © 1993 Academic Press, Inc.

1. INTRODUCTION

When developing expert systems it is often necessary to reason from evidential information which is uncertain, imprecise, and sometimes inaccurate. Bayesian theory, Fuzzy set theory [2, 3], and Dempster/Shafer

[1] This research was supported by the Joint Services Electronics Program under Contract DAAAL-03087-K0059.

[2] A mass function represents a piece of evidence as a probability distribution function in Bayesian theory.

0022-247X/93 $5.00

theory [4] have been developed to assist one in this reasoning process. Of all these methods, Dempster/Shafer theory appears to provide the most general framework for representing and combining evidential information.

Dempster/Shafer theory (D/S theory) has been used in many expert systems recently for reasoning from multiple pieces of evidential information. Examples of systems using this theory include computer vision [5, 6], acoustic signal interpretation [7], medical diagnosis [8], etc.

Even though D/S theory is well defined for discrete sets which are both mutually exclusive and collectively exhaustive, many supporting theories have not been developed due to this theory's shorter history compared to Bayesian theory and fuzzy set theory. Moreover, many problems need to be solved when D/S theory is applied to continuous domain problems. In particular, many expert systems are required to use information from continuous variables, e.g., time, distance, and sensor measurements. Thus, there is a need to extend D/S theory to accommodate these variables.

Strat [1] and Fua [9], working separately, have introduced a continuous framework for the Dempster/Shafer theory. Strat provides the framework for accommodating continuous variables while Fua provides a method for deriving a mass function from a probability density function. Although their work provides a good starting point, many problems still remain unresolved when the theory is extended to the continuous domain. In this paper, we present some theoretical foundations, based on Strat's framework for applying the D/S theory in the continuous domain.

The major contributions of our efforts are as follows: First, Dubois' method for interpreting statistical data is extended to the continuous domain. This is a simpler alternative to Fua's method and it forms the basis for the rest of our development. Second, we describe how to derive a mass function for the function of a continuous random variable when the mass function for the variable is given. This result is more difficult to obtain than in its discrete analogue. Third, we show how to generate a joint mass function of two random variables when the mass functions for two variables are known. We note that Hughes and Maksym [7] considered this problem in the discrete domain only for a very restricted case.

This paper is organized as follows: We begin with a summary of the Dempster/Shafer theory. This is followed by a brief review of previous approaches to the application of this theory in the continuous domain. This review includes our extension of Dubois' method. In the next two sections, we present the theories developed for transformation of mass and joint mass functions followed by some examples. We conclude with a discussion of the D/S theory in the continuous domain with emphasis on its usage and extensions.

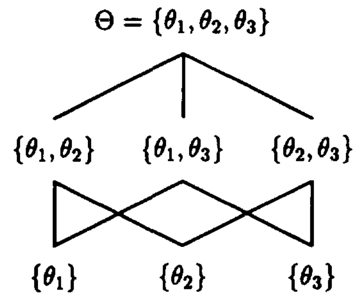

FIG. 1. Framework of Dempster/Shafer theory.

2. BACKGROUND

2.1. *Discrete Mass Function*

Consider a finite set of mutually exclusive and exhaustive propositions. The *frame of discernment* or Θ consists of the collection of all of n propositions, θ_i for $i \in [1, n]$. The frame Θ has 2^n subsets including itself and the null set ($\varnothing$). Subsets of Θ can be arranged as a tree (for $n = 3$ in Fig. 1). In the D/S theory, uncertain and imprecise evidence can be fully represented by assigning weights to all of the possible subsets of Θ.[3] Over the frame of discernment, the *mass function* is defined generally by the following:

DEFINITION 1 (A Mass Function). Let Θ be *a frame of discernment*; then a function m: $2^\Theta \to [0, 1]$ is called a *mass function* if

$$\text{(i)} \quad \sum_{A \subseteq \Theta} m(A) = 1 \qquad \text{and} \qquad \text{(ii)} \quad m(\phi) = 0. \tag{1}$$

In the D/S theory, the quantity $m(A)$ corresponds to the weight of evidence in favor of A, i.e., the probability that the evidence is exactly and completely described by A. The opinion about a proposition or a set of propositions is represented by the *probability interval* whose lower and upper bounds are the *belief*, Bel(A), and the *plausibility*, Pls(A),

[3] *Uncertain* information can be represented by assigning probability masses to two or more disjoint subsets. More probability masses are assigned to subsets of smaller sizes for representing more *precise* information.

respectively. The real probability, $P(A)$, when it exists, lies between these two values. Thus,

$$\forall A \subseteq \Theta, \qquad \mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

$$\mathrm{Pls}(A) = 1 - \mathrm{Bel}(\bar{A}) \tag{3}$$

$$\mathrm{Bel}(A) \leqslant P(A) \leqslant \mathrm{Pls}(A), \tag{4}$$

where $\bar{A}$ is the set complement of A. Each mass function, belief function, or plausibility function has $2^n - 2$ independent values. These values can be mapped in a one-to-one fashion; e.g., one and only one mass function can be derived from a belief function or a plausibility function. There is another compatible function, the *commonality function*. This function is represented by the quantity

$$Q(A) = \sum_{A \subseteq B \subseteq \Theta} m(B), \tag{5}$$

which measures the total probability mass that can move freely to the subset A.

If $m(A)$ is not zero, then A is defined to be a *focal* element. If every focal element is arranged in order so that every element except Θ is contained in the following one, then the mass function is *consonant*. If Θ is the only focal element, then the mass function is *vacuous*. This is a special case of a consonant mass function which is used to denote an initial condition in which there is no information at all.

Dempster's Combination Rule proposed in [10] is illustrated with good examples in [11]. It evolved from Bayes's rule of conditioning. Two mass functions m_1 and m_2 can be combined by an operation denoted $m_{1\oplus 2}$,

$$\begin{aligned} \forall A \subseteq \Theta, \qquad & m_{1\oplus 2}(A) = K \sum_{X \cap Y = A} m_1(X) m_2(Y), \\ \text{where} \qquad & K^{-1} = 1 - k = 1 - \sum_{X \cap Y = \phi} m_1(X) m_2(Y), \end{aligned} \tag{6}$$

where ϕ is the null set. If $k = 1$, the two mass functions are completely contradictory, and the mass function for their combination is not defined. We can combine multiple pieces of evidence by applying Eq. (6) repeatedly. As more evidence is combined, more probability mass is focused on a proposition. This means that the probability intervals get narrower and narrower and the belief in a certain proposition grows towards unity. Disadvantages of the D/S theory come from its generality. This theory requires large memory space and calculations which grow as $O(2^n)$ and $O(2^{2n})$ respectively.

Multiple pieces of evidence can be simultaneously combined in terms of commonality functions. The mass functions of M different pieces of evidence are transformed to the commonality functions by using Eq. (5) and combined by the following formulas:

$$
\begin{aligned}
&Q(A) = KQ_1(A)Q_2(A)\cdots Q_M(A) \\
&m(A) = Q(A) - \sum_{B \supset A} m(B) \qquad \text{where} \quad m(\Theta) = Q(\Theta).
\end{aligned} \tag{7}
$$

Here K is a factor used to normalize the mass function. Note that if $M = 2$, then K is the same as in Eq. (6) and the amount of required calculations grows as $O(2^n)$.

2.2. *Statistical Descriptions of a Mass Function*

In this section, we discuss the construction of a mass function from statistical data. Dubois and Prade [12] and Shafer [4] have suggested methods for transforming statistical data (i.e., the probability distribution) into their mass function. Both methods result in a consonant mass function because they assume that evidence from a source does not contradict internally.[4]

Suppose that we have a probability distribution in which the probability for a proposition θ_i is p_i for $1 \leqslant i \leqslant n$ and $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$. Then focal elements are A_i's, where $A_i = \{\theta_1, \theta_2, ..., \theta_i\}$ $(A_n = \Theta)$.

Dubois' Approach. In [12], a belief function and plausibility function can be defined by a probability distribution (the subscript $_\mathrm{d}$ denotes Dubois):

$$
\begin{aligned}
&\mathrm{Bel_d}(A) = \sum_{\theta_i \in A} \max(p_i - p_A, 0) \qquad \text{where} \quad p_A = \max\{p_i \mid \theta_i \in \bar{A}\} \\
&\mathrm{Pls_d}(\{\theta_i\}) = \sum_{k=1}^{n} \min(p_i, p_k), \qquad i = 1, ..., n.
\end{aligned} \tag{8}
$$

Shafer's approach [4] is predicated on the following two assumptions:

Assumption 1. The *plausibility* of each label is proportional to the likelihood ratio or probability and

Assumption 2. The *belief function* of a piece of evidence is consonant.

[4] No pair of disjoint sets are focal in a consonant mass function.

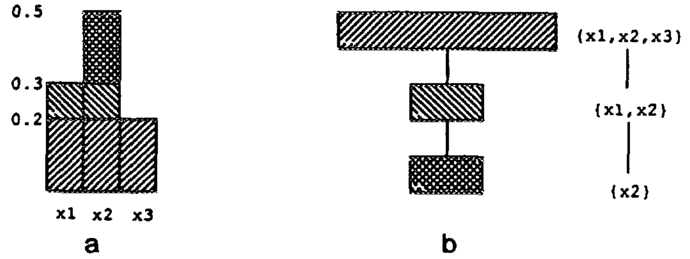

FIG. 2. (a) A likelihood distribution, (b) a mass function.

A belief function and mass function which satisfy these assumptions are defined by (the subscript $_s$ denotes Shafer)

$$\mathrm{Bel}_s(A) = 1 - \frac{\max_{\theta_i \in A} p_i}{\max_{\theta_i \in \Theta} p_i} \quad \text{and} \tag{9}$$

$$m_s(A) = \begin{cases} \mathrm{Bel}_s(A_i) - \mathrm{Bel}_s(A_{i-1}) & \text{if } A = A_i \text{ and } i = 2, ..., n \\ \mathrm{Bel}_s(A_1) & \text{if } A = A_1 = \{\theta_1\} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

We note that Dubois' and Shafer's results are not equivalent. The inequality relationship connecting their results as derived recently by Suh *et al.* [13] may be stated as

$$\mathrm{Bel}_d(A) \leqslant \mathrm{Bel}_s(A) \leqslant P(A) \leqslant \mathrm{Pls}_s(A) \leqslant \mathrm{Pls}_d(A). \tag{11}$$

Thus, Dubois' mass function is less specific according to Yager's specificity measure [14]. Another feature is its simplicity and graphical visualization. The following example shows Dubois' method for a discrete case of size 3.

EXAMPLE 1. Consider a likelihood distribution as depicted in Fig. 2a whose probabilities of elements, x_1, x_2, and x_3, are 0.3, 0.5, and 0.2. Then, since the area under $p = 0.2$ does not have any preference to any of the elements, it is assigned to the entire set $X = \{x_1, x_2, x_3\}$ in the mass function derivation. Since the probabilities of x_1 and x_2 are larger than 0.2 and not smaller than 0.3, the area between $p = 0.2$ and $p = 0.3$ is assigned to $\{x_1, x_2\}$. But only the probability of x_2 is larger than 0.3, therefore the area above $p = 0.3$ is assigned to $\{x_2\}$. Figure 2b shows the resultant mass function for this example. In Fig. 2b, the mass $m(\{x_2\})$ is represented by the area above the line $p = 0.3$ in the histogram and $m(\{x_1, x_2\})$ is the area between the line $p = 0.2$ and $p = 0.3$.[5]

[5] p is the probability assigned to each proposition.

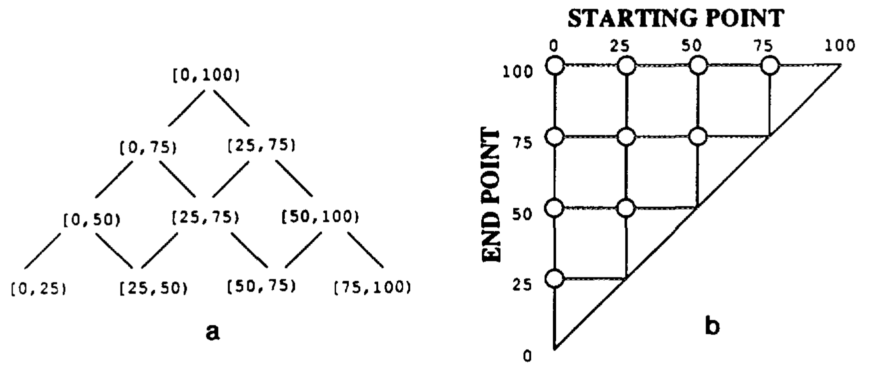

FIG. 3. Discretization of a continuous variable, shown as (a) a tree, (b) a triangle.

3. CONTINUOUS MASS FUNCTION

In some expert systems, it is necessary to deal with continuous variables such as time, distance, or other sensor measurements. Since evidence theory is defined over a finite set of discrete propositions, continuous variables have usually been handled by partitioning the variable range into discrete subsets of possible values. Since expert systems are mostly interested in whether or not a value lies within some contiguous range of values, the tree of subsets may be simplified as in Fig. 3a. This will be illustrated by means of an example presented in Strat [1]:

> 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 55° or between 70° and 80°").

Since we allow only contiguous intervals, there are $n(n+1)/2$ nonempty subsets rather than the $2^n - 1$ nonempty subsets that would be possible if this constraint were not applied.

Strat found that the approach of partitioning can be very sensitive to slight variations. For example, Bel([20, 40)) for today's temperature in °F may be greatly different from Bel([20, 39)). Strat also introduced a method for specifying a smoothly varying set of probability masses about the value of a continuous variable.

The subset tree of Fig. 3a can also be represented using a triangular matrix as shown in Fig. 3b. Here, the abscissa and ordinate specify the lower and upper bounds of an interval[6] respectively. This framework can

[6] This interval is the range of the continuous variable and not the probability interval.

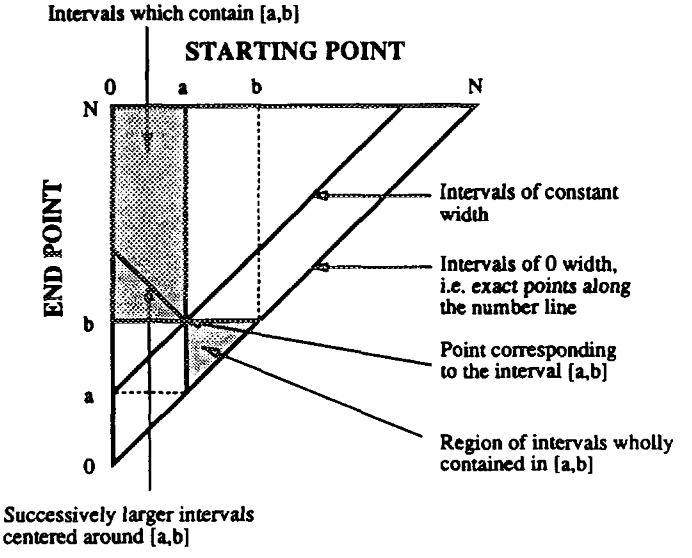

FIG. 4. The continuous frame of discernment.

be applied to the smoothly varying set of continuous variables by generalizing the triangle to the continuous domain as shown in Fig. 4. In order to assign mass to zero length intervals, e.g., [20, 20], intervals wil be closed at both ends.

Probability mass can be assigned to points, lines, and regions in the triangular frame of the continuous domain. We begin with the definition of these mass functions.

DEFINITION 2 (Point Mass Function). If probability mass is assigned to points in a continuous domain as shown in Fig. 5a, then it is defined as a point mass function. Let this be denoted by m. Each point corresponds to the interval over which the probability is uniformly distributed. Therefore, $m_x(a, b)$ represents the probability that x has a value between a and b.

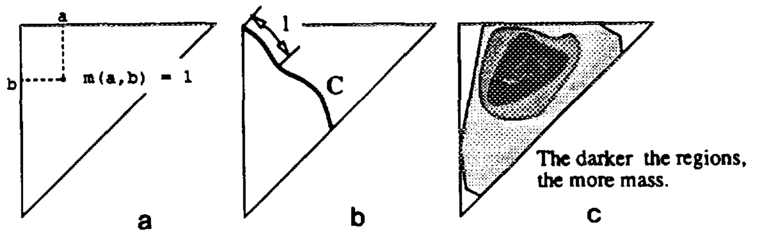

FIG. 5. (a) Point, (b) line, and (c) area mass functions.

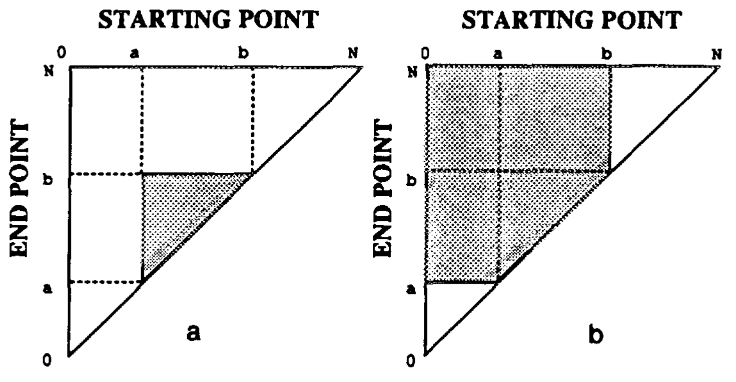

FIG. 6. Support for (a) belief and (b) plausibility of the interval $[a, b]$.

DEFINITION 3 (Line Mass Function). If probability mass is assigned to lines in the continuous domain as shown in Fig. 5b, for example, then it is called a line mass function and denoted by m'. m' denotes a probability mass assigned to unit length. Its dimension[7] is L^{-1}.

DEFINITION 4 (Area Mass Function). m'' denotes the probability mass assigned to unit area. The dimension is L^{-2}.

The point and line mass functions in Fig. 5 can also be represented using impulse function and an area mass function,

$$\begin{aligned} m(a, b) &= \int_x \int_y m''(x, y)\, \delta(x-a)\, \delta(y-b)\, dx\, dy \\ m'(x_l, y_l) &= \int_L m''(x, y)\, \delta(s-l)\, ds, \end{aligned} \tag{12}$$

where s is the arc length of L and (x_l, y_l) denotes the coordinate of the point on the line L at an arc length l from the origin. By using these relationships, all the theories or derivations defined for area mass functions can be converted to those for point and line mass functions.

Strat presents a method for calculating the *belief* and *plausibility* functions graphically and mathematically. This is done by integrating the mass function over the shaded regions shown in Fig. 6:

$$\begin{aligned} \mathrm{Bel}([a, b]) &= \int_a^b \int_x^b m''(x, y)\, dy\, dx \\ \mathrm{Pls}([a, b]) &= \int_0^b \int_{\max(a, x)}^N m''(x, y)\, dy\, dx. \end{aligned} \tag{13}$$

[7] L denotes the length unit while T and M denote the time and mass units, respectively. For example, the dimension of *velocity* is LT^{-1}.

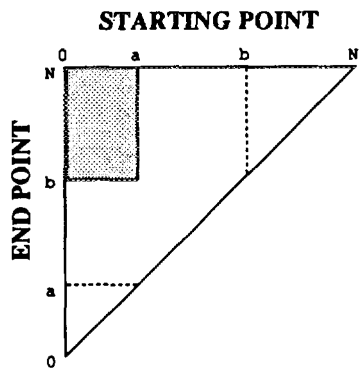

FIG. 7. Support for the commonality of the interval $[a, b]$.

Extending this approach, the continuous version of Dempster's rule is given by

$$m_1 \oplus m_2(a, b) = K \int_0^a \int_b^N (m_1''(x, b)\, m_2''(b, y) + m_1''(a, y)\, m_2''(x, b) + m_1''(a, b)\, m_2''(x, y) + m_1''(x, y)\, m_2''(a, b))\, dx\, dy \quad (14)$$

where

$$K = \frac{1}{1-k}$$

and

$$k = \int_0^N \int_p^N \int_q^N \int_r^N (m_1''(p, q)\, m_2''(r, s) + m_1''(r, s)\, m_2''(p, q))\, ds\, dr\, dq\, dp.$$

THEOREM 1. *A commonality function*[8] *can be derived from a mass function as shown in Fig. 7 and a mass function can be derived from a commonality function using the following equations*:

$$Q([a, b]) = \int_0^a \int_b^N m''(x, y)\, dy\, dx \quad (15)$$

$$m(a, b) = \left. \frac{\partial^2 Q([x, y])}{\partial x\, \partial y} \right|_{x=a,\ y=b}. \quad (16)$$

[8] Note that the commonality function is dimensionless.

Proof. This theorem can be proved by using a discrete triangular lattice and taking a limit. Let

$$a = n_a \Delta x; \qquad b = n_b \Delta x; \qquad x = n_x \Delta x; \qquad y = n_y \Delta x.$$

The mass function associated with the lattice point (n_x, n_y) is given by

$$M(n_x, n_y) = m''(x, y)(\Delta x)^2$$

$$Q([n_a, n_b]) = \sum_{n_x=0}^{n_N} \sum_{n_y=0}^{n_N} \gamma M(n_x, n_y)$$

(If $[n_a, n_b] \in [n_x, n_y]$, $\gamma = 1$, otherwise, $\gamma = 0$)

$$= \sum_{n_x=0}^{n_a} \sum_{n_y=n_b}^{n_N} M(n_x, n_y)$$

$$Q([a, b]) = \lim_{\Delta x \to 0} Q([n_a, n_b])$$

$$= \lim_{\Delta x \to 0} \sum_{n_x=0}^{n_a} \sum_{n_b=0}^{n_N} m''(x, y)(\Delta x)^2$$

$$= \int_0^a \int_b^N m''(x, y)\, dx\, dy.$$

In the discrete triangular domain,

$$M(n_a, n_b) = Q([n_a, n_b]) - Q([n_a, n_b - 1]) - Q([n_a - 1, n_b]) + Q([n_a - 1, n_b - 1])$$

$$m''(a, b) = \lim_{\Delta x \to 0} \frac{M(n_a, n_b)}{(\Delta x)^2}$$

$$= \lim_{\Delta x \to 0} \left(\left(\frac{Q([n_a, n_b]) - Q([n_a, n_b - 1])}{\Delta x} - \frac{Q([n_a - 1, n_b]) - Q([n_a - 1, n_b - 1])}{\Delta x} \right) \Big/ \Delta x \right)$$

$$= \left. \frac{\partial^2 Q([x, y])}{\partial x\, \partial y} \right|_{x=a,\, y=b}. \qquad \text{Q.E.D.}$$

M sources of evidence denoted by continuous mass functions $m_1, m_2, ..., m_M$, can be combined using Eq. (7). This gives

$$Q([a, b]) = K Q_1([a, b]) Q_2([a, b]) \cdots Q_M([a, b]),$$

where K is a normalizing factor and Q_i is derived from m_i for $i \in [1, M]$ using Eq. (15). Then the consensus mass function can be derived from $Q([a, b])$ using Eq. (16).

3.1. *Extended Dubois' Method*

In this section, Dubois' approach [12], discussed in Section 2.2, is extended to the continuous domain. Dubois' mass function is less specific according to Yager's specificity measure [14] than that of Shafer. Both methods are described in Eq. (8) and in Eqs. (9, 10), respectively.

This concept may be extended to the continuous domain. Suppose that we have a probability distribution $p_X(x)$ for a random variable X. Let the belief on an interval $[x, y]$ in Fig. 8 be represented by the shaded area; then

$$\text{For } p_X(x) = p_X(y), \qquad \text{Bel}([x, y]) = P_X(y) - P_X(x) - (y - x)\, p_X(x)$$

$$\text{where} \quad P_X(x) = \int_{-\infty}^{x} p_X(x')\, dx'. \tag{17}$$

Probability mass is assigned on a trajectory L, such that any point (x, y) on L satisfies $p_X(x) = p_X(y)$. Thus, this trajectory L is the focal line. We then have

$$\text{For } p_X(x) = p_X(y), \qquad m'(x, y)\, \Delta l = \lim_{\Delta x \to 0} (y - x)(p_X(x + \Delta x) - p_X(x))$$

$$m'(x, y) = (y - x) \frac{dp_X(x)}{dx} \cdot \frac{dx}{dl}, \tag{18}$$

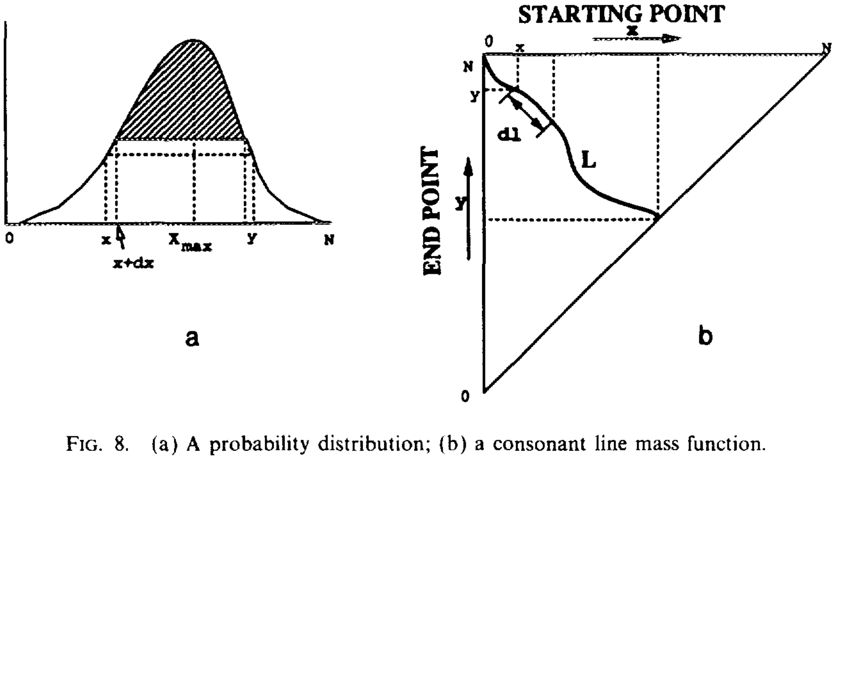

FIG. 8. (a) A probability distribution; (b) a consonant line mass function.

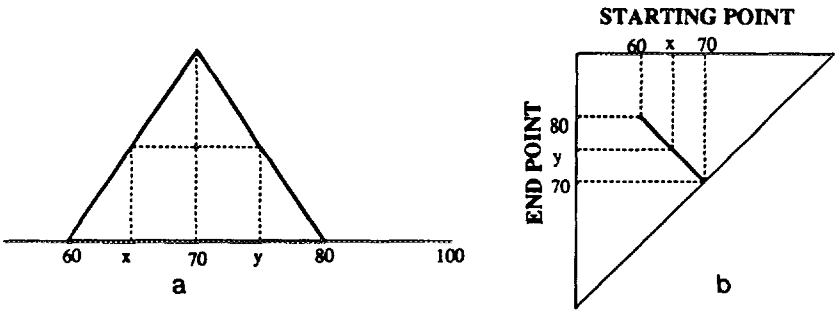

FIG. 9. (a) A convex probability distribution; (b) a line mass function.

where (x, y) denotes the coordinate of the point on the line L and the arc length between this point and the origin is l. The foregoing is illustrated by the following example.

EXAMPLE 2. Suppose over a period of time the average temperature on a given day of a year, say March 15, is computed as 70°. The probability distribution from the evidence may be shown as in Fig. 9a. The mass function can be calculated by using Eq. (18) as follows:

$$m'(x, y) = (y - x)\frac{dp_X(x)}{dx}\cdot\frac{dx}{dl}$$

$$= (140 - 2x)\cdot\frac{1}{100}\cdot\frac{1}{\sqrt{2}}.$$

The focal line is shown in Fig. 9b. This line is straight since the slopes before and after 70° are constant. The integration of $m'(x, y)$ along the line from [70, 70] to [60, 80] is unity. Thus, we have

$$\int_L m'(x, y)\,dl = \int_L m'(x, y)\,dx\,\frac{dl}{dx} = \int_{60}^{70} m'(x, y)\,dx\cdot\sqrt{2} = 1.$$

If the probability distribution function is monotonically increasing before and decreasing after x_{max} as shown in Fig. 8, i.e., *convex*, the resulting mass function is a line mass function with one focal line in the triangular continuous mass domain. If, however, the probability distribution is not convex, as shown in Fig. 10a, it may result from the combination of two different pieces of evidence and the corresponding mass function has two focal lines. This is illustrated in Fig. 10b.

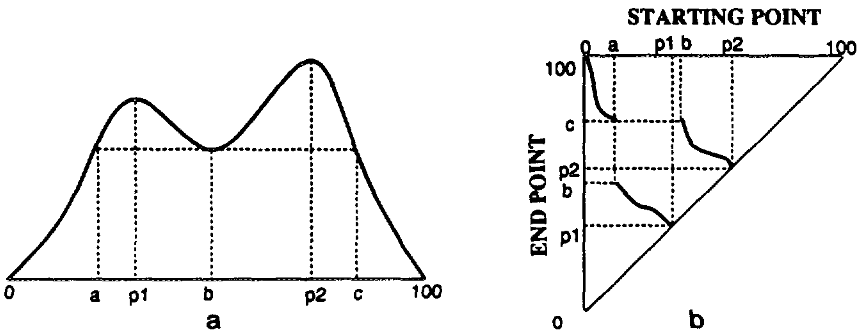

FIG. 10. (a) A non-convex probability distribution; (b) a non-convex mass function.

4. Transformation of Mass Function

In certain situations, we may wish to determine the mass function for $y = g(x)$ in terms of the mass function of the continuous random variable x. This is a transformation of the mass function from one mass domain to another. In this section, it is shown how to obtain such a transformation.

In Bayesian theory, such a transformation is easily performed by mapping one probability density function or pdf to another. Generally, if $y = g(x)$, whose real roots are $x_1, ..., x_n, ...$, and the pdf for x is $f_x(x)$, then the transformed pdf in the y-domain is

$$f_y(y) = \frac{f_x(x_1)}{|g'(x_1)|} + \cdots + \frac{f_x(x_n)}{|g'(x_n)|} + \cdots, \tag{19}$$

where $y = g(x_1) = \cdots = g(x_n) = \cdots$ and $g'(x)$ is the derivative of $g(x)$. If there exists only one solution, Eq. (19) becomes

$$f_y(y) = f_x(x) \cdot \frac{dg^{-1}(y)}{dy}, \qquad \text{where} \quad x = g^{-1}(y). \tag{20}$$

To illustrate this operation consider the following example.

EXAMPLE 3. Suppose that there is a uniform pdf for x in the interval $[0, 1]$, as shown in Fig. 11a. We wish to evaluate a pdf for $y = \sqrt[3]{x} + 1$. Consider the equation $y = g(x)$, where real root is $x = g^{-1}(y) = (y-1)^3$. Applying Eq. (20) we have

$$f_y(y) = \begin{cases} \dfrac{d[(y-1)^3]}{dy} & \text{for} \quad 1 \leqslant y \leqslant 2 \\ 0 & \text{otherwise.} \end{cases}$$

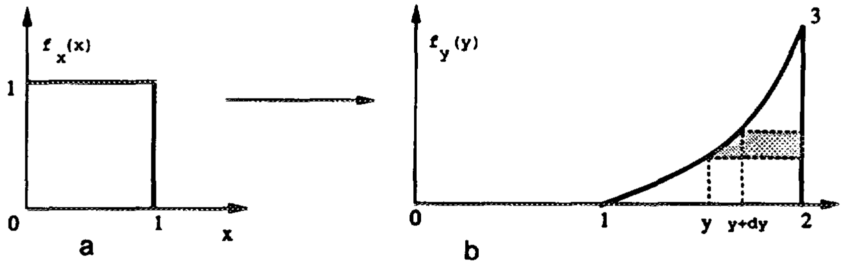

FIG. 11. pdf's for random variables (a) x, (b) y, where $y = \sqrt[3]{x} + 1$.

4.1. *Point Mass Function*

Since in the mass function probability is not assigned to a single value of x but to an interval of x, mas function transformation cannot be done by mapping in a point-to-point fashion. Probability assigned to a point denoting an interval $[a, b]$ in the continuous mass function domain corresponds to a uniform pdf in the interval (see Fig. 12). The variables x_l and x_u denote the lower and upper bounds of an interval and $[x_l, x_u]$ corresponds to a point in the continuous mass domain. The method used in Example 3 can be used to generate the transformation from the point mass function of x to the mass function $y = g(x)$.

When $g(x)$ is linear, a uniorm probability distribution is transformed to a uniform probability distribution; i.e., the point mass function of x is trivially transformed to the point mass function of y. If, however, $g(x)$ is nonlinear, then the point mass function of x is transformed to the line mass function of y, which will be constructed from $f_y(y)$. The pdf in Example 3 corresponds to the mass assigned to a point $[0, 1]$ in the m_x domain, the mass function domain of x. Then, the pdf of $y, f_y(y)$ (Fig. 11b) corresponds to a line from $[1, 2]$ to $[2, 2]$ in the m_y domain (see Fig. 13a). For $1 \leqslant y \leqslant 2$, the mass value for the interval $[y, 2]$ is the shaded area in Fig. 11b).

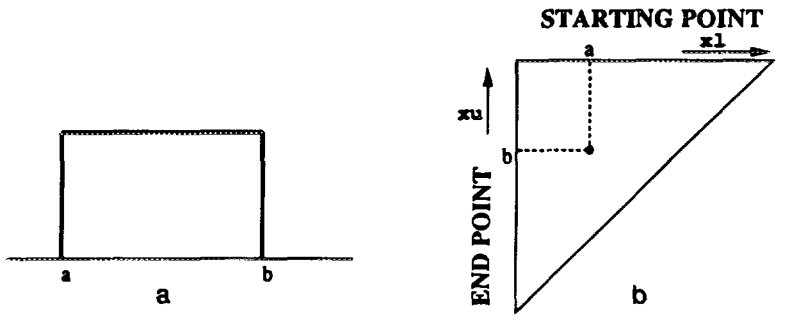

FIG. 12. (a) A uniform pdf; (b) a point mass function.

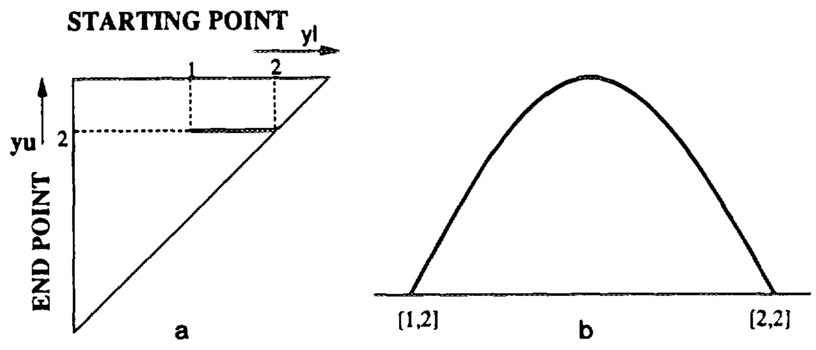

FIG. 13. (a) A line mass function; (b) a mass distribution along a focal line.

Figure 13b shows the probability mass distribution along the line from [1, 2] to [2, 2] in the m_y domain for this example:

$$m'_y(y, 2) = (2-y)\frac{df_y(y)}{dy} = 6(2-y)(y-1) \qquad \text{for} \quad 1 \leqslant y \leqslant 2.$$

The area under the parabola of Fig. 13b is unity. If $g(x)$ is defined and differentiable within the interval, then a point mass on x is transformed into a line mass function. In particular, if $g(x)$ is monotonically increasing as in Example 3, a point mass on the m_x domain is transformed to a line parallel to the y_l-axis, i.e., the line between $[g(a), g(b)]$ and $[g(b), g(b)]$, while for a monotonically decreasing $g(x)$, the line mass is parallel to the y_u-axis, the line from $[g(a), g(a)]$ to $[g(a), g(b)]$.

The line mass function can be analytically derived if $g(x)$ is a function which is monotonic within the interval $[a, b]$. For example, if $m_x(a, b) = 1$ and $g(x)$ is a monotonically increasing function in the interval $[a, b]$, then

$$m'_y(y, g(b)) = m_x(a, b)(g(b) - y)\frac{d^2 g^{-1}(y)}{dy^2} \qquad \text{for} \quad g(a) \leqslant y \leqslant g(b). \tag{21}$$

If $g(x)$ is convex, then m_y is the mass function of one line within the triangle defined by the three points of $\{[g(a), g(a)], [g(a), g(b)], [g(b), g(b)]\}$. These formulas can also be applied to transformation of line mass and area mass funtions.

4.2. *Line and Area Mass Function*

A line mass function can be viewed as an infinite number of point mass functions. It has been shown that a point mass function is transformed to a line mass function when $g(x)$ is monotonic.

If $g(x)$ is monotonic for $x \in [a, b]$, all points along the focal line in the m_x domain are transformed into lines parallel to the axis of lower bounds of y, the y_l axis, and result in an area mass function in the m_y domain. If $g(x)$ is monotonically increasing and no two points have the same upper bound in the m_x domain, then the line mass function is transformed into an area mass function that lies below the line along $[g(x_l), g(x_u)]$ in the m_y domain. In the foregoing, x_l and x_u are two random variables with $x_{min} \leqslant x_l \leqslant x_u \leqslant x_{max}$. Let the mass assigned to each unit area be denoted $m_y''(y, g(x_u))$; then from Eq. (21), we have for $g(x_l) \leqslant y \leqslant g(x_u)$

$$m_y''(y, g(x_u)) = m_x'(x_l, x_u)(g(x_u) - y)\frac{d^2 g^{-1}(y)}{dy^2}, \tag{22}$$

where m_x' is a line mass function for the random variable x.

In the same manner, the result of transforming an area mass function in the m_x domain when $g(x)$ is monotonically increasing, can be analytically expressed as an area mass function in the m_y domain by using Eq. (22). For $g(x_{min}) \leqslant y \leqslant g(x_u)$,

$$m_y''(y, g(x_u)) = \int_{x_{min}}^{g^{-1}(y)} m_x''(x_l, x_u)(g(x_u) - y)\frac{d^2 g^{-1}(y)}{dy^2}\, dx_l, \tag{23}$$

where m_x'' is the line mass function for the random variable x. We may follow this same approach to derive analytically the transformed mass function when $g(x)$ is monotonically decreasing. Even in the case when $g(x)$ is not monotonic or convex, the mass function m_y can still be calculated numerically by exploiting the ideas developed in this paper. The computational burden, however, may be expensive and the procedure, in general, may be problem dependent.

5. Joint Mass Function

In the previous section, we discussed how to transform a mass function of one random variable to that of another random variable. In this section, we show how to construct a joint mass function of two random variables, i.e., how to obtain the mass function of z from the mass functions of x and y when $z = g(x, y)$. Following our usual approach we begin with a preliminary discussion of the Bayesian procedure.

Example 4. Suppose that we are given the uniform pdf's of two independent random variables x and y, and that it is desired to evaluate the pdf of $z = x + y$ and $w = xy$.

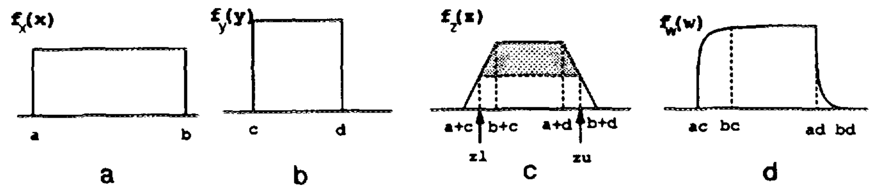

FIG. 14. Probability density functions for random variables (a) x, (b) y, (c) $z=x+y$ (for $(b-a)<(d-c)$), and (d) $w=xy$ (for $bc<ad$).

Let

$$f_X(x)=\begin{cases}\dfrac{1}{b-a} & \text{for} \quad a \leqslant x \leqslant b \\ 0 & \text{otherwise}\end{cases} \quad \text{and} \quad f_Y(y)=\begin{cases}\dfrac{1}{d-c} & \text{for} \quad c \leqslant y \leqslant d \\ 0 & \text{otherwise.}\end{cases}$$

Both $f_Z(z)$ and $f_W(w)$ can be easily derived via the following equations:

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\, dy$$

$$\text{and} \qquad f_W(w)=\int_0^{\infty} f_X\left(\frac{x}{y}\right) f_Y(y)\, \frac{1}{y}\, dy.$$

These are shown graphically in Fig. 14c and d.

5.1. *Joint Mass Function of Two Point Mass Functions*

The uniform pdf's for random variables x and y correspond to two point mass functions as in Figs. 15a and b respectively. Thus, the joint mass function of $z=x+y$ can be obtained from the joint pdf of $z=x+y$ shown in Fig. 14c. The resulting mass function is a line mass function from $(a+c, b+d)$ to $(b+c, a+d)$. Mass distribution along the focal line can be calculated by using the extended Dubois method.

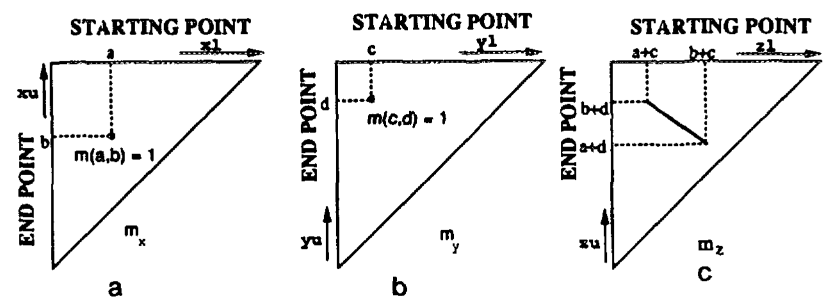

FIG. 15. (a) A point mass function of x; (b) a point mass function of y, (c) a joint mass function of $z=x+y$.

When $b-a<d-c$, the belief on the intrval (z_l, z_u) is the shaded region in Fig. 14c. For $a+c<z_l<b+c$ and $a+d<z_u<b+d$, $\mathrm{Bel}([z_l, z_u])$ can be calculated using

$$\mathrm{Bel}(z_l, z_u]) = \frac{(a+d-b-c-z_l+z_u)(z_u-a-d)}{2(d-c)(b-a)}.$$

But in order for $f_z(z)|_{z_l}$ to be the same as $f_z(z)|_{z_u}$, z_l must be a function of z_u; i.e., $z_l = b+c-(z_u-a-d)$. Since z_l and z_u are related by a linear equation, the focal line is straight and the slope in the m_z domain is 45°. That is,

$$\mathrm{Bel}([z_l, z_u]) = \frac{(z_u-b-c)(z_u-a-d)}{(d-c)(b-a)}.$$

From the belief function, the line mass fuction m'_z can be calculated. When s is defined as the distance between $(a+c,\ b+d)$ and (z_l, z_u), i.e., $s=\sqrt{2}\,(z_u-a-c)$ and $ds/dz_u=\sqrt{2}$, then

$$\begin{aligned} m'_z(z_l, z_u) &= \frac{d\,\mathrm{Bel}([z_l, z_u])}{ds} = \frac{d\,\mathrm{Bel}([z_l, z_u])}{dz_u}\cdot\frac{dz_u}{ds} \\ &= \frac{2z_u-a-d-b-c}{\sqrt{2}\,(a-b)(c-d)}, \end{aligned} \tag{24}$$

where m'_z denotes the probability assigned to unit length and integrating m'_z along the focal line yields unity. The length of the focal line is $\sqrt{2}\,|a-b|$. This subsection can be summarized by the following theorem.

THEOREM 2. *The joint mass function m_z of two point mass functions m_x and m_y is a line mass function when $z=x+y$.*

COROLLARY 2.1. *Its support is a straight line between $(a+c,\ b+d)$ and $(b+c,\ a+d)$ when $(b-a)<(d-c)$ or between $(a+c,\ b+d)$ and $(a+d,\ b+c)$ when $(b-a)\geqslant(d-c)$.*

COROLLARY 2.2. *The line mass along the support line is*

$$m'_z(z_l, z_u) = \frac{2z_u-a-d-b-c}{\sqrt{2}\,(a-b)(c-d)}.$$

5.2. *Joint Mass Function of Point and Line Mass Function*

Suppose that we have the point mass function of x ($m(a, b)=1$) and the line mass function m'_y of y. The mass function of $z=x+y$ is obtained by accumulating the point-to-point joint mass function.

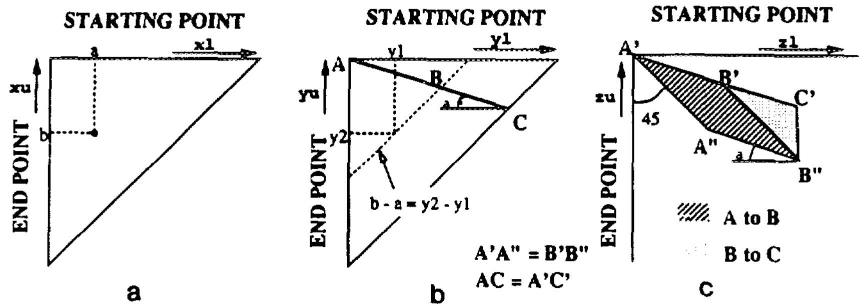

FIG. 16. (a) A point mass function of x; (b) a line mass function of y; (c) a joint mass function of $z = x + y$. (If $A = (y_{\min}, y_{\max})$, $A' = (a + y_{\min}, b + y_{\max})$.)

Each point-to-point joint mass function between m_x and the point mass function at (y_l, y_u) results in a 45° line mass function with upper end point at $(a + y_l, b + y_u)$. Since a and b are constant, the tracks of these upper end points in the m_z domain are parallel as well as those of the focal line in the m_y domain; furthermore their lengths are the same. The lower end point is determined by the length of the focal line for each point-to-point joint mass function

From A to B. If $(y_u - y_l)$ is larger than $(b - a)$ (Line $\overline{AB}$ in Fig. 16b), the length is independent of y_l and y_u, and equal to $\sqrt{2}\,(b - a)$. (See Fig. 16c A to B.) Hence we can figure out the focal area in the m_z domain by sliding a 45° tilted line with the length of $\sqrt{2}\,(b - a)$ along the focal line in the m_y domain.

From B to C. But in the region where $(y_u - y_l)$ is smaller than $(b - a)$ (Line $\overline{BC}$ in Fig. 16b), the length is dependent on y_l and y_u, and is equal to $\sqrt{2}\,(y_u - y_l)$. If the focal line in this region of the m_y domain is straight ($\overline{BC}$), then the track of the lower end points connects the lower end point for the point B and the upper end point for the point C; i.e., this track connects B'' and C' in the m_z domain. (See Fig. 16c B to C.)

5.3. *Joint Mass Function of Point and Area Mass Function*

In the previous section, we noted that a point-to-point joint mass function is a line mass function. Although the length of the line is dependent on the location of (y_l, y_u) and (b, a), the slope is always 45° for $z = x + y$. As a result, the possible focal area in the m_z domain is the shaded region in Fig. 17c. This area is uniquely defined by m_y and (a, b). Both the shape and size of $\triangle PQR$ in Fig. 17b are the same as those of $\triangle P'Q'R'$. The trapezoid $\square AQRB$ in Fig. 17b is the same as the trapezoids $\square A'Q'R'B''$ and $\square A''Q'R'B''$ in Fig. 17c. When $P = (y_{\min}, y_{\max})$ in Fig. 17b, the points in

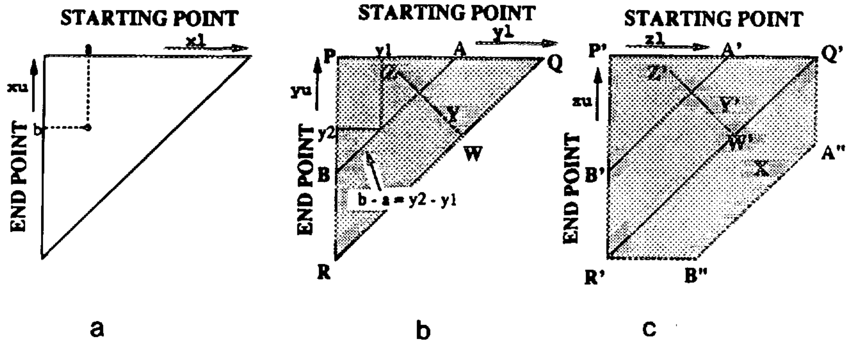

FIG. 17. A point mass function, m_x; (b) an area mass functon, m_y''; (c) a joint mass function of $z = x + y$.

Fig. 17c become $P' = (a + y_{min},\ b + y_{max})$, $Q' = (a + y_{max},\ b + y_{max})$, and $R' = (a + y_{min},\ b + y_{min})$. When $(b - a) < (y_{max} - y_{min})$, A'' is $(b + y_{min},\ b + y_{min})$ while B'' is $(a + y_{max},\ a + y_{max})$. But if $(b - a) \geqslant (y_{max} - y_{min})$, $A'' = B'' = (a + y_{max},\ b + y_{min})$.

The procedure for calculating the mass assigned to each unit area in the m_z domain is different between the regions inside $\triangle P'Q'R'$ and the regions inside the trapezoid $\square A''Q'R'B''$.

Point Y' *in* $\triangle P'Q'R'$. For the point $Y' = (z_l, z_u)$, $m_z''(z_l, z_u)$ is dependent on every point along $\overline{YZ}$ in the m_y domain where $Y = (z_l - a,\ z_u - b)$ and $Z = (z_l - b,\ z_u - a)$. In other words, $m_z''(z_l, z_u)$ can be calculated by integrating the effect of the line mass functions each of which is a joint mass function between point mass function m_x and point mass function $m_y''(y_l, y_u)\, dy_l\, dy_u$, where (y_l, y_u) lies on the line $\overline{YZ}$. Since $\overline{YZ}$ is tilted by 45°, y_u can be described as a function of y_l. Then, by using Theorem 2,

$$m_z''(z_l, z_u) = \int_{z_l - b}^{z_u - a} m_y''(y_l, y_u) \frac{2z_u - a - b - y_u - y_l}{\sqrt{2}\,(b - a)(y_u - y_l)}\, dy_l. \tag{25}$$

Point X *in* $\square A''Q'R'B''$. When $X = (z_l, z_u)$ is inside the trapezoid $\square A''Q'R'B''$, no point along $\overline{WY}$ in Fig. 17b affects $m_z''(z_l, z_u)$. A point (y_l, y_u) on $\overline{WY}$ in Fig. 17b affects mass on a straight line between $(a + y_l, b + y_u)$ and $(a + y_u,\ b + y_l)$ which does not include the point X. Equation (25) can be used to calculate $m_z''(z_l, z_u)$ by modifying the bound of y_l:

$$m_z''(z_l, z_u) = \int_{z_l - b}^{z_u + a - b} m_y''(y_l, y_u) \frac{2z_u - a - b - y_u - y_l}{\sqrt{2}\,(b - a)(y_u - y_l)}\, dy_l. \tag{26}$$

Our method can easily be applied to the problem of determining the joint mass function when both m_x and m_y are area mass functions. As a result, m_z, the joint mass function, can be calculated from any mass functions of two random variables x and y when $z = x + y$.

5.4. *Joint Mass Function for Separable Functions*

Suppose that it is desired to calculate the mass function m_z for $z = g(x, y)$ when the mass functions, m_x and m_y are given. If the function $g(x, y)$ is separable, then the joint mass function can be readily calculated by using the methods described in the previous sections.

Let $g(x, y) = g_1(x) \cdot g_2(y)$, where $p = \log(g_1(x))$ and $q = \log(g_2(x))$; then using logarithms, this separable function is transformed to

$$\log(g(x, y)) = \log(g_1(x)) + \log(g_2(y)) = p + q.$$

Note that the logarithm functions are monotonic in the region where they are defined.

The mass functions m_p and m_q are calculated by transformation of mass function. By using Eq. (25), we can obtain the joint mass function m_{p+q} that is obtained from m_p and m_q. Finally, m_z can be obtained by transforming m_{p+q} for $z = e^{p+q}$.

6. Discussion

In this paper, we have solved problems which must be resolved in order for evidence theory to be used for continuous variables. First, a new method has been developed for calculating a continuous mass function from statistical data. Second, this paper provides a method for computing the continuous mass function of a random variable that is a function of another random variable where mass function is known. Last, this paper provides a method for calculating a joint mass function from continuous mass functions of two random variables.

It has been said that evidence theory is a subset of fuzzy set theory, because the former does not have as many supporting theories as the latter. Actually, since the number of degrees of freedom of the fuzzy variable and the computation amount of the fuzzy operations are proportional to $O(n)$, where n is the number of propositions, it is easier to develop the various fuzzy operations, such as conjunction, intersection, transformation, joint fuzzy set variables, etc., either in the discrete domain or in the continuous domain. With evidence theory, one does not enjoy such a variety. For example, the conjunction of two pieces of evidence has not been well defined yet.

Moreover, when evidence theory is used in the continuous domain, there arise some problems which do not matter in the discrete domain. Since a mass function is defined by using $O(2^n)$ variables, it is not defined in the continuous domain. Strat [1] introduced a continuous framework based on the restriction that no disjoint interval is allowed. As an extension, we developed a new combination rule, a method for mass function transformation, and a method for yielding joint mass functions. These methods will help one to design expert systems which involve reasoning in the continuous domain using evidence theory.

Evidence theory can be coinsidered as a generalized Bayesian theory. For the transformation of mass functions and the generation of joint mass functions, each point on the continuous mass function domain is considered as a uniform probability distribution and these operations are basically performed point by point. The results are converted to mass functions by using the Dubois method, which is computationally less expensive than its alternative, Shafer's method.

All the theories developed in this paper provide analytic solutions for transformations and joint mass functions for the restricted cases described. They can also be implemented numerically in all cases if their domain is closed and well defined (differentiable and integrable). In the case of numerical implmentation, sampling techniques are very important for efficient calculation and accurate results.

REFERENCES

1. T. M. STRAT, Continuous belief functions for evidential reasoning, *in* "Proceedings *AAAI*-84," Austin, Texas, August 1984, pp. 308–313.
2. L. A. ZADEH, Fuzzy sets, *Inform. Control* **8** (1965), 338–353.
3. L. A. ZADEH, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* **1** (1978), 3–28.
4. G. SHAFER, "A Mathematical Theory of Evidence," Princeton Univ. Press, Princeton, NJ, 1976.
5. L. P. WESLEY, Evidential knowledge-based computer vision, *Opt. Engrg.* **25**, No. 3 (1986), 363–379.
6. D. Y. SUH, R. M. MERSEREAU, R. L. EISNER, AND R. I. PETTIGREW, Knowledge-based boundary detection applied to cardiac magnetic resonance image sequences, *in* "IEEE International Conference on Acoustics, Speech, Signal Processing," Glasgow, Vol. 3, pp. 1783–1786, May 1989.
7. R. C. HUGHES AND J. N. MAKSYM, Acoustic signal interpretation: Reasoning with non-specific and uncertain information, *Pattern Recognition* **18**, No. 6 (1985), 475–483.
8. J. YEN, GERTIS: A Dempster-Shafer approach to diagnosing hierarchical hypotheses, *Comm. ACM* **32**, No. 5 (1989), 573–585.
9. P. V. FUA, Using probability-density functions in the framework for evidential reasoning, *in* "International Conference on Information Processing and Management of Uncertainty," Paris, France, 1986, pp. 103–110.

10. A. P. DEMPSTER, A generalization of Bayesian inference, *J. Roy. Statist. Soc. Ser. B* **30** (1968), 205–247.
11. B. G. BUCHANAN AND E. H. SHORTLIFFE, "Rule-Based Expert Systems" Addison–Wesley, Reading, MA, 1984.
12. D. DUBOIS AND H. PRADE, Fuzzy sets and statistical data, *European Journal of Operational Research*, **25** (1986), 345–356.
13. D. Y. SUH AND R. M. MERSEREAU, Accuracy measure in evidence theory, submitted for publication.
14. R. R. YAGER, Entropy and specificity in a mathematical theory of evidence, *Int. J. General Systems* **9** (1983), 249–260.