NORTH-HOLLAND

A Theory of Gaussian Belief Functions

Liping Liu*

School of Business, Albany State College, Albany, Georgia

ABSTRACT

A Gaussian belief function can be intuitively described as a Gaussian distribution over a hyperplane, whose parallel subhyperplanes are the focal elements. This paper elaborates on the idea of Dempster and Shafer and formally represents a Gaussian belief function as a wide-sense inner product and a linear functional over a variable space, and as their duals over a hyperplane in a sample space. By adapting Dempster's rule to the continuous case, it derives a rule of combination and proves its equivalence to its geometric description by Dempster. It illustrates by examples how mixed knowledge involving linear equations, multivariate Gaussian distributions, and partial ignorance can be represented and combined as Gaussian belief functions.

KEYWORDS: ***expert systems, Dempster-Shafer theory, belief networks, knowledge representation, multivariate Gaussian distributions***

1. INTRODUCTION

The notion of Gaussian belief functions (GBFs) extends Dempster-Shafer theory in representing mixed knowledge, some of which is logical, some uncertain, and some vacuous. Logical knowledge is represented by a hyperplane in a sample space. Ignorance is represented by partitioning the hyperplane into parallel subhyperplanes as focal elements. Uncertain knowledge is then represented by a Gaussian distribution across the focal elements over the hyperplane. In its full generality, a GBF can be intuitively described as a Gaussian distribution across the parallel members of a partition of a hyperplane. It includes as special cases nonprobabilistic linear equations, statistical observations, multivariate Gaussian distributions over a hyperplane, and vacuous belief functions. In terms of graphical models (Kong, 1986), a general GBF can be seen as the combination of

Address correspondence to Liping Liu, Ph.D., School of Business, Albany State College, Albany, GA 31705. E-mail: LIPING@RAMS.ALSNET.PEACHNET.EDU.

Received January 1994; revised May 1995; accepted September 1995.

International Journal of Approximate Reasoning 1996; 14:95–126

0888-613X/96/$15.00

655 Avenue of the Americas, New York, NY 10010

SSDI 0888-613X(96)00115-8

its special cases, whose individual representation is often trivial. However, to represent a GBF in its full generality, advanced notions such as linear functionals and linear spaces are required. This paper elaborates on the idea of Dempster (1990b) and Shafer (1992) and formally represents a GBF as a wide-sense inner product and a linear functional over a variable space, and as their dual over a hyperplane in a sample space. When variables of interest are specified, the abstract representation can be reduced into matrices and linear and quadratic functions, which allow efficient implementation of the idea of GBFs. Using examples, the paper shows how this can be done and how statistical models and assumptions can be formally represented and combined as GBFs.

As in Dempster-Shafer theory (Shafer, 1976), knowledge represented by GBFs has two primitive operations—marginalization and combination. Marginalization corresponds to the coarsening of knowledge, and combination corresponds to the integration of knowledge. Drawing inferences consists of combining all the relevant information and marginalizing the full body of knowledge into the variables of interest. The marginalization of a GBF is most naturally described as a projection in a variable space. The combination of GBFs is geometrically described by Dempster (1990b) as intersecting hyperplanes and summing the restricted log "densities" in a sample space. To interplay these two operations, this paper adapts Dempster's rule to the continuous case and derives a rule of combination in a variable space. It shows that the resulting rule is equivalent to the geometric description in Dempster (1990b).

In terms of representation, a GBF may be equivalently represented by a Bayesian model and vice versa. However, the proposal of the notion of GBFs is dictated by the best-known properties of belief-function modeling—the representation of ignorance by vacuous belief functions, the resolution of complex representations of uncertainty into components by graphical models, and the combination of independent models by Dempster's rule. As Dempster (1990b) argued, the belief-function formalism generalizes Bayesian inference of posterior distributions while abandoning its most controversial component: improper priors. It extends, unifies, and clarifies Fisher's fiducial method of posterior reasoning while filling the void of a prior distribution in the logical structure with a vacuous belief function. Also, as its geometric description suggests, a GBF treats all the components of a statistical model (such as observations, model assumptions, and subjective beliefs) not as separate concepts, but as manifestations of a single concept. Furthermore, the specification of a graphical belief-function model is based on symmetric evidential independence assumptions that are simpler and easier to check than asymmetric Bayesian conditional-probability assumptions, whose verification and falsification are often difficult due to human beings' limited knowledge about causality.

These features of GBFs allow people to concentrate their modeling efforts on recognizing and incorporating independent components of real information.

The notion of GBFs turns out to have a wide range of real applications. Dempster (1990a, b) shows how the Kalman filter can be understood in terms of GBFs. As Dempster (1990a) shows, the state equations and the observation equations in the Kalman filter can be captured by logical belief functions. The distributional assumptions on independent random disturbances can be represented by Gaussian distributions. The values of observable variables can be represented as another set of logical belief functions. All three types of belief functions are specified locally in a belief network. The recursion involved in the filter can be regarded as a special case of the recursion involved in the computation of GBF marginals in a join tree. The full Kalman filter model results from judging all these components belief functions to be independent and combining them into a single belief function by Dempster's rule.

Because GBFs can represent statistical models, and because Dempster's rule can be used to combine knowledge from independent items of evidence, it is clear that the theory of GBFs provides a method of combining independent models. Liu (1995b) implements this idea. Specifically, information from different sources such as multiple databases is treated as independent items of evidence. The knowledge drawn from each database, such as a linear regression model or a belief network, is represented by a GBF. The models from different databases are combined in the way we combine GBFs. Combined predictions or inferences are then made, based on the combined model. Obviously, this approach is consistent with the spirit of Dempster-Shafer theory of belief functions and Dempster's rule of combination. Using concrete examples, Liu (1995b) shows how linear models, simultaneous equations, and belief networks can be combined as GBFs. He also shows how this important task can be actually performed by simple matrix operations. The proposed method has a potential application in automated learning of belief networks from multiple databases that are neither appendable nor joinable (Maier, 1983). It also generalizes the metaanalysis for integrating independent statistical findings (Hunter and Schmidt, 1990) and the Bayesian method of estimating common regression coefficients (Box and Tiao, 1973). As we will see shortly, the GBF method can combine models of different kinds that may involve different variables. In contrast, the models to be combined in the metaanalysis and the Bayesian method must be the same, and the parameters to be estimated must be common. In such a restricted case, Liu (1995b) shows that the GBF method is similar in flavor to metaanalysis and the Bayesian method. For example, for the problem of weighted means (Yates, 1939) and its generalization (Box and Tiao, 1973), both the GBF method and the Bayesian method give the same posterior distribution of common regression coefficients.

In expert systems, the number of GBFs to be combined could be very large. It is inefficient and even infeasible to combine all of them first and then make inferences. Liu (1995a) extends existing work on finite belief functions (Kong, 1986; Shafer, Shenoy, and Mellouli, 1987; Shenoy and Shafer, 1990) and proposes a local computation scheme for GBFs. The basic idea is to arrange all the GBFs into a tree-structured graph, called a join tree, and propagate knowledge by sending and absorbing messages step by step in the tree. Each step of propagation involves sending a message from a node to a neighbor. Thus, the join-tree approach consists of a series of local computations, each of which involves only a small number of variables that are near each other in the join tree. The local computation scheme has been shown to work for finite belief or probability functions (Kong, 1986; Shenoy and Shafer, 1990; Lauritzen and Spiegelhalter, 1988). Liu (1995a) shows that it also works for GBFs by proving the axioms of Shenoy and Shafer (1990), which are the conditions under which the local computation of any objects is possible.

An outline of this paper is as follows. Section 2 briefly introduces Dempster-Shafer theory of belief functions and provides some notions and terminology that are used throughout the paper. Section 3 first describes GBFs in geometric terms and then formally represent them respectively in variable spaces and sample spaces. This section uses some advanced concepts such as linear spaces and linear functionals, which are mathematically elegant but not scientifically crucial. The nontechnical reader may just read the examples and geometric descriptions to make sense of GBFs. Section 4 derives a rule for combining GBFs in variable spaces and then represents it equivalently as intersections and restricted summations in sample spaces. Section 5 concludes the paper.

2. DEMPSTER-SHAFER THEORY

The notion of belief functions can be traced to the work of Jakob Bernoulli on pooling pure evidence. In modern language, an item of pure evidence proves a claim with a certain probability but has no bearing on its negation. Probabilities in accordance with pure evidence are not additive. For example, suppose I find a scrap of newspaper predicting a blizzard tomorrow, which I regard as infallible. Also, suppose I am 75% certain that the newspaper is today's. Then, I am 75% sure of a blizzard tomorrow. However, if the newspaper is not today's, either blizzard or no blizzard could happen, since then the newspaper carriers no information on tomorrow's weather. The degree of support for no blizzard is zero and for either blizzard or no blizzard is 25%.

Bernoulli's idea of nonadditive probabilities has now been well developed by Dempster (1968), Shafer (1976), and many others, under the name of Dempster-Shafer theory of belief functions. In this theory, a piece of evidence is encoded as a probability measure. The degree of belief for a claim is interpreted as a degree of the evidential support. Degrees of belief from independent items of evidence are combined by Dempster's rule of combination. Let X be a set of discrete variables, and X^* its finite sample space.[1] Let A_X denote a subset of X^*, which is interpreted as the proposition that the true value of X is in A_X. Then the degree of evidential support for A_X is represented by $m(A_X)$. The assignment of $m(A_X)$ is in accordance with a certain item of evidence and satisfies the following axioms:

$$0 \leq m(A_X) \leq 1, \qquad m(\varnothing) = 0, \qquad \sum \{m(A_X) | A_X\} = 1. \tag{1}$$

A subset A_X is called a focal element iff $m(A_X) > 0$. Due to lack of evidence justifying a more specific allocation, a portion of our total belief allocated to a focal element A_X does not necessitate the allocation of any partial belief to its subset. For the above newspaper example, we can encode the evidence by a probability measure with p(today's) = 0.75 and p(not today's) = 0.25. Since "today's newspaper" supports the claim "blizzard" and "not today's newspaper" supports the claim "blizzard or no blizzard," the degrees of evidential support can be represented as m({blizzard}) = 0.75, m({blizzard, no blizzard}) = 0.25, and m({no blizzard}) = 0. Thus, {blizzard} and {blizzard, no blizzard} are the two focal elements. The 25% of belief for {blizzard, no blizzard} does not imply any reallocation of the belief to its subsets {blizzard} and {no blizzard}.

If all the focal elements are singletons, we call the belief function *Bayesian*. On the other hand, if the sample space is the only focal element, we call the belief function *vacuous*. One advantage of the belief-function modeling is its ability to represent ignorance and partial ignorance. In Bayesian inference, complete ignorance is often represented by a uniform prior distribution or a prior with large scale parameters, such as a Gaussian distribution with large variance. As Fisher consistently criticized (Fisher, 1959; Zabell, 1989), such priors often lack theoretical or empirical bases.

[1] In Shafer (1976), the term "frame of discernment" instead of "sample space" is used, to emphasize its epistemic nature in that a sample space is deliberately constructed according to our knowledge and opinion.

They sometimes imply vanishingly small prior probability for regions of practical interest. The belief-function formalism represents ignorance by vacuous belief functions. It clearly distinguishes lack of belief from disbelief. For example, a vacuous belief function with $m(\{\text{rainy, not rainy}\}) = 1$ will be regarded as totally different from the one with $m(\{\text{rainy}\}) = \frac{1}{2}$ and $m(\{\text{not rainy}\}) = \frac{1}{2}$.

Another advantage of the belief-function formalism is its ability to pool independent pieces of evidence by Dempster's rule. A piece of evidence is encoded as a probability measure. The pooling of two independent pieces of evidence can be encoded as the product of two probability measures. From this perspective, Dempster (1967) derived a rule for combining belief functions that represent independent pieces of evidence. Suppose there are two belief functions Bel_1 and Bel_2 respectively for sets X and Y. Their basic probability assignments are respectively $m_1(A_X)$ and $m_2(A_Y)$. Then, by Dempster's rule, the combined belief function, denoted by $\text{Bel}_1 \otimes \text{Bel}_2$, is for set $X \cup Y$ and has basic probability assignment

$$m(A_{X \cup Y}) = \alpha^{-1} \sum \{m_1(A_X) m_2(A_Y) | (A_X)^{\downarrow X \cap Y} \cap (A_Y)^{\downarrow X \cap Y} = (A_{X \cup Y})^{\downarrow X \cap Y} \}, \quad (2)$$

where α is a normalization constant given by

$$\alpha = \sum \left\{ m_1(A_X) m_2(A_Y) \middle| (A_X)^{\downarrow X \cap Y} \cap (A_Y)^{\downarrow X \cap Y} \neq \varnothing \right\},$$

and $(A_X)^{\downarrow X \cap Y}$ is the projection of A_X to the sample space of $X \cap Y$. The symbols $(A_Y)^{\downarrow X \cap Y}$ and $(A_{X \cup Y})^{\downarrow X \cap Y}$ are interpreted similarly. In general, suppose Y is a subset of X. Then

$$(A_X)^{\downarrow Y} = \{y | A_X \cap [\{y\} \times (X \setminus Y)^*] \neq \varnothing\}, \quad (3)$$

where $(X \setminus Y)^*$ is the sample space of $X \setminus Y$. Note that $(A_X)^{\downarrow X \cap Y} \cap (A_Y)^{\downarrow X \cap Y} = \varnothing$ represents that the two assertions A_X and A_Y from Bel_1 and Bel_2 are conflicting. One of them must be false, and a joint assertion is qualitatively impossible. α is the total belief committed to all the joint assertions that are qualitatively possible. If $\alpha = 0$, the two belief functions are incombinable because they have no joint assertions qualitatively possible.

Combination corresponds to the integration of knowledge. Sometimes we are interested in drawing partial knowledge from a full body of knowledge. This corresponds to the coarsening of knowledge, obtained by the marginalization of a belief function. Suppose Bel is a belief function for X with basic probability assignment $m(A_X)$, and Y is a subset of X.

Then we define $\text{Bel}^{\downarrow Y}$ as a belief function for Y with basic probability assignment $m^{\downarrow Y}$ satisfying

$$m^{\downarrow Y}(A_Y) = \sum \left\{ m(A_X) \middle| (A_X)^{\downarrow Y} = A_Y \right\}. \tag{4}$$

3. GAUSSIAN BELIEF FUNCTIONS

Variables of interest in this paper can be classified as deterministic, such as observables or controllables; random, whose distribution is Gaussian; and vacuous, on which no knowledge bears. Based on a given body of evidence, a GBF in general encodes logical and probabilistic knowledge for all the three types of variables. Logical knowledge is represented by linear equations, which are in turn represented by a hyperplane in a sample space. Probabilistic knowledge is represented by Gaussian distributions across all the members of a partition of the hyperplane into parallel subhyperplanes. Less general than an ordinary belief function, whose focal elements may have nonempty intersections, a GBF has the parallel subhyperplanes as its mutually exclusive focal elements. Let n, $n - c$, and $n - b$ denote the dimension numbers of the sample space, the hyperplane, and a focal element, respectively. In general, $c \leq b \leq n$. By appropriately setting one or two of the dimension numbers c, b, and n, a GBF can be degenerated into six nontrivial varieties, which provide building blocks for more complex GBFs. If $b = c = 0$, then the GBF is vacuous and has the sample place as its sole focal element. If $0 < c = b < n$, then the GBF is equivalent to specifying c linear equations. If $c = b = n$, the true point in the sample space is known with certainty, as might occur by direct observation. If $c = 0$ and $b = n$, then the GBF is an ordinary Gaussian probability distribution in the sample space. If $c > 0$ and $b = n$, the GBF is a Gaussian probability distribution over the hyperplane. In the latter two cases, the GBF is Bayesian, because its focal elements are singletons with zero dimension. Finally, if $0 = c < b < n$, the GBF is a proper belief function, which has a Gaussian distribution for some variables and no opinion for others.

The above geometric description of GBFs is due to Dempster (1990b). In this section, we want to represent a GBF in its full generality as a mathematical construct and a computational object. According to Dempster (1969), if all the variables of interest span a variable space—a finite-dimensional vector space whose elements are random variables—then we can consider the sample space to be its dual space, the space of all linear functionals on the variable space. Accordingly, a hyperplane in the sample space is dual to a subspace in the variable space. A wide-sense inner product in the sample space, which specifies the log "density" of a

GBF over a hyperplane, is dual to a wide-sense inner product in the variable space, which specifies the covariance of all the variables on a variable subspace. From this dual correspondence, nonprobabilistic linear equations, which are represented by a hyperplane in the sample space, can be represented by a variable subspace, in which each variable takes on a value with certainty. We will call such a variable subspace the *certainty space* of a GBF. A multivariate Gaussian distribution over a hyperplane in the sample space can be represented by a wide-sense inner product with the certainty space as its null space, which specifies the covariance between random variables. Therefore, we can fully describe a GBF in coordinate-free terms in both a variable space and a sample space.

3.1. Representation in Variable Spaces

Let **V** be a random-variable space. A GBF on **V** is a quintuplet $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$, where **C**, **B**, and **L** are nested subspaces of **V**, $\mathbf{C} \subseteq \mathbf{B} \subseteq \mathbf{L} \subseteq \mathbf{V}$, π is a wide-sense inner product of **B** with **C** as its null space, and E is a linear functional on **B**. We call **C** the *certainty space*, **B** the *belief space*, **L** the *label space*, π the *covariance*, and E the *expectation*. The expectation E and the covariance π define a Gaussian distribution for the variables in **B** by specifying their means and covariances. This Gaussian distribution is regarded as a full expression of our beliefs, based on a given body of evidence; this item of evidence justifies no beliefs about variables in **L** going beyond what is implied by the beliefs about the variables in **B**. (The evidence might justify some further beliefs about variables that are not in **L**, but these are outside the discussion so far as a belief function with space **L** is concerned.) The Gaussian distribution assigns zero variance to the variables in **C**; if X is in **C**, we are certain that it takes the value $E(X)$ with certainty. Let **F** be a subspace of **B** such that $\mathbf{B} = \mathbf{C} \oplus \mathbf{F}$. We call **F** the uncertainty space, because each variable in it has nonzero variance.

Suppose **C**, **B**, **L**, and **V** have dimensions c, b, l, and n, respectively. Then we can choose a basis $X_1, X_2, \ldots, X_n$ of **V** such that $X_1, \ldots, X_c$ is a basis of **C**, $X_1, \ldots, X_b$ is a basis of **B**, and $X_1, \ldots, X_l$ is a basis of **L**. Of course, $X_{c+1}, \ldots, X_b$ is a basis of **F**. For $i = 1, 2, \ldots, b$, let μ_i denote the mean of X_i. For $i, j = 1, 2, \ldots, b - c$, let Σ_{ij} denote the covariance between X_{c+i} and X_{c+j}. Let $\mu = (\mu_1, \ldots, \mu_b)$ and $\boldsymbol{\Sigma} = [\Sigma_{ij}]_{(b-c)\times(b-c)}$. Then E and π can be represented as follows:

$$E[(\alpha_1, \ldots, \alpha_b)] = (\alpha_1, \ldots, \alpha_b)\mu^T,$$

$$\pi[(\alpha_1, \ldots, \alpha_b), (\beta_1, \ldots, \beta_b)] = (\alpha_{c+1}, \ldots, \alpha_b)\boldsymbol{\Sigma}(\beta_{c+1}, \ldots, \beta_b)^T,$$

where $(\alpha_1, \ldots, \alpha_b)$ and $(\beta_1, \ldots, \beta_b)$ are two variables in **B**. It is easy to see that $\pi(\cdot, \cdot)$ is a wide-sense inner product on **B** with **C** as its null space: $\pi(S,T) = 0$ if S or $T \in \mathbf{C}$.

EXAMPLE 1 Let X, Y, and Z be three variables and x, y, and z be their sample values. A GBF on these variables includes a Gaussian distribution $X + X \sim N(0.5, 2)$ and a linear equation $x + y + z = 1$. Let **V** be spanned by X, Y, and Z. Let $\mathbf{L} = \mathbf{V}$. Let **C** be spanned by the variable $X + Y + Z$, and **B** by the two variables $X + Y + Z$ and $X + Y$. The linear functional E and the wide-sense inner product π on **B** are defined as follows:

$$E[\alpha_1(X + Y + Z) + \alpha_2(X + Y)] = \alpha_1 + 0.5\alpha_2,$$

$$\pi[\alpha_1(X + Y + Z) + \alpha_2(X + Y), \beta_1(X + Y + Z) + \beta_2(X + Y)]$$
$$= 2\alpha_2 \beta_2.$$

Then, by verifying its variance and mean, it is easy to see that the variable $X + Y + Z$ takes on the value 1 with certainty, and so **C** is a one-dimensional certainty space to represent the linear equation $x + y + z = 1$. Therefore, we arrive at GBF = $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$. This GBF expresses beliefs about each variable in **B** by giving its mean and variance. Suppose, for example, $Z = (X + Y + Z) - (X + Y)$. Then, $E(Z) = 1 - 0.5 = 0.5$ and $\pi(Z, Z) = 2(-1)(-1) = 2$. However, it has no opinion on variables in **L** that are not in **B**. For example, it justifies no beliefs about the variables $X, Y, X - Y$, etc.

The reason for having the variable-space representation is partly the simplicity of defining marginalization, which cannot be obtained in the sample-space representation defined shortly. In a variable space, the marginalization of a GBF is simply a projection. Suppose $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$ is a Gaussian belief function, and **M** is a subspace of **L**. Then the marginal of $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, \mathbf{E})$ on **M**, denoted by $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)^{\downarrow \mathbf{M}}$, is another GBF obtained by intersecting the certainty space **C**, belief space **B**, and label space **L** with **M** and restricting the covariance and the expectation to the new belief space:

$$(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)^{\downarrow \mathbf{M}} = (\mathbf{C} \cap \mathbf{M}, \mathbf{B} \cap \mathbf{M}, \mathbf{L} \cap \mathbf{M}, \pi|_{\mathbf{B} \cap \mathbf{M}}, E|_{\mathbf{B} \cap \mathbf{M}}). \quad (5)$$

In Example 1, if **M** is spanned by Z, then $\mathbf{C} \cap \mathbf{M} = 0$, $\mathbf{B} \cap \mathbf{M} = \mathbf{L} \cap \mathbf{M} = \{\alpha Z | \alpha \in \mathbb{R}\}$, $\pi|_{\mathbf{B} \cap \mathbf{M}}(\alpha Z, \beta Z) = 2\alpha\beta$, and $E|_{\mathbf{B} \cap \mathbf{M}}(\alpha Z) = 0.5\alpha$. On the other hand, if **M** is spanned by X, then $\mathbf{C} \cap \mathbf{M} = 0$, $\mathbf{B} \cap \mathbf{M} = 0$, $\mathbf{L} \cap \mathbf{M} = \mathbf{M}$, $\pi|_{\mathbf{B} \cap \mathbf{M}}(0) = 0$, and $E|_{\mathbf{B} \cap \mathbf{M}}(0) = 0$. The marginal is vacuous. This is intuitively reasonable, because $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$ carries no knowledge about X.

3.2. Representation in Sample Spaces

Let $\mathbf{V}^*$ denote the sample space for $\mathbf{V}$. The mathematical essentials are best conveyed by first considering $\mathbf{V}^*$ to be the dual space of $\mathbf{V}$ and each sample point to be a linear functional on $\mathbf{V}$. A Gaussian distribution on a hyperplane of $\mathbf{V}^*$ can then be represented by specifying an inner product and a linear functional on the hyperplane. An advantage with the notion of linear functionals is their independence of coordinates. A linear functional v over $\mathbf{V}$ is a real-valued function such that $v(\alpha X + \beta Y) = \alpha v(X) + \beta v(Y)$ for all variables X and Y in $\mathbf{V}$ and real numbers α and β. We can regard $v(X)$ as a sample value for X. Therefore, v specifies a sample value for each variable in $\mathbf{V}$. In particular, suppose $X_1, X_2, \ldots, X_n$ is a basis for $\mathbf{V}$. Let $v(X_i) = x_i$ for $i = 1, 2, \ldots, n$. Then v specifies a vector $(x_1, x_2, \ldots, x_n)$, which is often referred to as a sample point. Because of its linearity, v is one-to-one correspondent to $(x_1, x_2, \ldots, x_n)$. Therefore, we can treat a linear functional and a sample point interchangeably. They are different only in that the latter depends on a basis for $\mathbf{V}$ while the former does not. As a familiar example of linear functionals, the expectation E defines the mean for each variable in $\mathbf{V}$. When a basis is chosen, E is equivalent to the mean vector μ in $\mathbf{V}^*$.

Without referring to its representation in $\mathbf{V}$, a GBF can be independently represented in $\mathbf{V}^*$ by specifying a hyperplane, a partition of the hyperplane, and a wide-sense inner product and a linear functional over the hyperplane. However, to see the relationship between the two representations, we derive a dual representation in $\mathbf{V}^*$ for a given GBF $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$. To do this, we need to choose a linear functional t on $\mathbf{V}$ that agrees with E on $\mathbf{C}$—that is, $t(X) = E(X)$ for every variable X in $\mathbf{C}$. The functional t is allowed to disagree with E on variables in $\mathbf{B}$ that are not in $\mathbf{C}$. When such a t has been chosen, we say that the GBF is marked, and we call t its mark. We write $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E, t)$ for a marked GBF. Given a linear functional t, according to Dempster (1969), each subspace $\mathbf{S}$ in $\mathbf{V}$ has a dual hyperplane in $\mathbf{V}^*$ that contains t:

$$\mathbf{S}^* = \{v | v(X) = t(X) \text{ for all } X \text{ in } \mathbf{S}\}. \qquad (6)$$

Therefore, $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{L}$ have dual hyperplanes $\mathbf{C}^*$, $\mathbf{B}^*$, and $\mathbf{L}^*$, respectively. It is easy to see that these hyperplanes are nested: $\mathbf{L}^* \subseteq \mathbf{B}^* \subseteq \mathbf{C}^*$. According to the linear functional E, we can define an additional hyperplane in $\mathbf{V}^*$ as follows:

$$\mathbf{E}^* = \{v | v(X) = E(X) \text{ for all } X \text{ in } \mathbf{B}\}.$$

It follows from $E(X) = t(X)$ for all X in $\mathbf{C}$ that $\mathbf{E}^*$ is contained in $\mathbf{C}^*$ and parallel to $\mathbf{B}^*$. Since $\mathbf{C}$ is the certainty space, each variable X in it takes on a single value $E(X)$ with certainty. Because $v(X)$ is interpreted as a sample value of X and $E(X) = t(X)$ for all X in **C**, $\mathbf{C}^*$ actually specifies the location of the true sample value of X in **C**. Therefore, $\mathbf{C}^*$ is the hyperplane that represents linear equations in the GBF $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$. We are certain that the true sample point must be on $\mathbf{C}^*$, but we do not know where it is exactly on $\mathbf{C}^*$. The hyperplane $\mathbf{E}^*$ specifies its mean location. If E^* is a singleton, then the expected position of the true sample point is specific. Otherwise, $\mathbf{E}^*$ ranges from $-\infty$ to $+\infty$ along some dimensions. It means that we are completely ignorant about where the true sample point is along these dimensions. Therefore, $\mathbf{E}^*$ is actually a focal element. Any other hyperplanes, including $\mathbf{B}^*$, which are parallel to $\mathbf{E}^*$ are also focal elements. All the focal elements form a partition of the hyperplane $\mathbf{C}^*$.

The above hyperplanes are better illustrated in a coordinate system. Choose the same basis as in Section 3.1, and represent each linear functional by its corresponding sample point. Then $\mathbf{V}^* = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R}, i = 1, 2, \ldots, n\}$. Let the mark $t = (\mu_1, \ldots, \mu_c, t_{c+1}, \ldots, t_n)$. Then we have

$$\mathbf{C}^* = \{(x_1, \ldots, x_n) | x_1 = \mu_1, \ldots, x_c = \mu_c\},$$

$$\mathbf{B}^* = \{(x_1, \ldots, x_n) | x_1 = \mu_1, \ldots, x_c = \mu_c, x_{c+1} = t_{c+1}, \ldots, x_b = t_b\},$$

$$\mathbf{L}^* = \{(x_1, \ldots, x_n) | x_1 = \mu_1, \ldots, x_c = \mu_c, x_{c+1} = t_{c+1}, \ldots, x_l = t_l\},$$

$$\mathbf{E}^* = \{(x_1, \ldots, x_n) | x_1 = \mu_1, \ldots, x_c = \mu_c, x_{c+1} = \mu_{c+1}, \ldots, x_b = \mu_b\}.$$

Note that the nice look of the above hyperplanes is due to the appropriate choice of a basis. If a different basis is chosen, they may have to be expressed by linear equations.

To represent a Gaussian distribution across all the subhyperplanes on $\mathbf{C}^*$ that are parallel to $\mathbf{B}^*$, we need to define a wide-sense inner product over $\mathbf{C}^*$. Since $\mathbf{C}^*$ is not a subspace, we introduce the following operations on $\mathbf{C}^*$:

$$x \oplus y = (x - t) + (y - t) + t \qquad \text{for any } x \text{ and } y \in \mathbf{C}^*,$$

$$\alpha \otimes x = \alpha(x - t) + t \qquad \text{for any } x \text{ in } \mathbf{C}^* \text{ and any real number } \alpha.$$

It is easy to verify that $\oplus$ and $\otimes$ are closed operations on $\mathbf{C}^*$. According to Dempster (1969), the wide-sense inner product π, which is defined on **B** and takes the value 0 on **C**, has a dual operation $\pi^*(x, y)$, which is a wide-sense inner product on $\mathbf{C}^*$ with the null hyperplane $\mathbf{B}^*$ under the operations $\oplus$ and $\otimes$. That is, $\pi^*(x, y) = 0$ iff x or $y \in \mathbf{B}^*$, $\pi^*(x, y) = \pi^*(y, x)$ for any x and $y \in \mathbf{C}^*$, and

$$\pi^*[(\alpha \otimes x) \oplus (\beta \otimes y), z] = \alpha\pi^*(x, z) + \beta\pi^*(y, z)$$

for any $x, y, z \in \mathbf{C}^*$ and real numbers α and β. In coordinate terms, if $\mathbf{\Sigma}$ is the covariance matrix as in Section 3.1, then for any $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbf{C}^*$,

$$\pi^*(x, y) = (x_{c+1} - t_{c+1}, \ldots, x_b - t_b)\mathbf{\Sigma}^{-1}(y_{c+1} - t_{c-1}, \ldots, y_b - t_b)^T. \tag{7}$$

So far we have found the duals for **C**, **B**, **L**, and π. In the following we need to derive a linear functional on $\mathbf{C}^*$, which is dual to E and specifies the mean location $\mathbf{E}^*$. First we establish a one-to-one correspondence between linear functionals on $\mathbf{C}^*$ that are zero on the hyperplane $\mathbf{B}^*$ and the hyperplanes on $\mathbf{C}^*$ that are parallel to $\mathbf{B}^*$.

LEMMA 1 *In coordinate terms, $H^*(x)$ is a linear function on $\mathbf{C}^*$ that is zero on $\mathbf{B}^*$ iff there exists a $(b - c)$-dimensional vector $\boldsymbol{a}$ such that*

$$H^*(x) = \boldsymbol{a}\mathbf{\Sigma}^{-1}(x_{c+1} - t_{c+1}, \ldots, x_b - t_b)^T, \tag{8}$$

where $x = (\mu_1, \ldots, \mu_c, x_{c+1}, \ldots, x_n) \in \mathbf{C}^*$.

Proof It suffices to prove the necessity. The linearity of $H^*(x)$ implies that there exists an n-dimensional vector z such that $H^*(x) = zx^T$ for any $x = (\mu_1, \ldots, \mu_c, x_{c+1}, \ldots, x_n)$ on $\mathbf{C}^*$. We decompose z into (z_1, z_2, z_3) such that

$$H^*(x) = z_1(\mu_1, \ldots, \mu_c)^T + z_2(x_{c+1}, \ldots, x_b)^T + z_3(x_{b+1}, \ldots, x_n)^T.$$

Since $H^*(x) = 0$ for any $x \in \mathbf{B}^*$, it follows that $z_3 = 0$ and

$$z_1(\mu_1, \ldots, \mu_c)^T + z_2(t_{c+1}, \ldots, t_b)^T = 0.$$

Therefore, for any $x \in \mathbf{C}^*$,

$$H^*(x) = z_2(x_{c+1} - t_{c+1}, \ldots, x_b - t_b)^T.$$

Let $\boldsymbol{a} = z_2\mathbf{\Sigma}$. Then (8) is proved. ■

Comparing (7) and (8), we see that, for any fixed x^0 in $\mathbf{C}^*$, $\pi^*(x^0, x)$ is a linear functional on $\mathbf{C}^*$ with null hyperplane $\mathbf{B}^*$. In general, $\pi^*(x^0, x)$ is different when x^0 changes over $\mathbf{C}^*$. However, it is easy to see from (7) that $\pi^*(x^0, x)$ is invariant iff x^0 is in a hyperplane parallel to $\mathbf{B}^*$. That is, for each hyperplane $\mathbf{H}^*$ that is parallel to $\mathbf{B}^*$,

$$H^*(x) = \pi^*(x^0, x) \qquad \text{for any } x^0 \text{ in } \mathbf{H}^* \tag{9}$$

is a linear functional on $\mathbf{C}^*$ that is zero on $\mathbf{B}^*$, and the choice of x^0 does not matter. On the other hand, for each linear functional $H^*(x)$ that is zero on $\mathbf{B}^*$ and has the form (8),

$$\mathbf{H}^* = \{(\mu_1, \ldots, \mu_c, x_{c+1}, \ldots, x_n) | (x_{c+1}, \ldots, x_b) = \boldsymbol{a} + (t_{c+1}, \ldots, t_b)\} \tag{10}$$

is a hyperplane that is parallel to $\mathbf{B}^*$. Therefore, we have

LEMMA 2 *Through $\pi^*(x^0, x)$, linear functionals that are zero on $\mathbf{B}^*$ and hyperplanes that are parallel to $\mathbf{B}^*$ are in a one-to-one correspondence carried out by* (9) *and* (10).

Note that $\mathbf{E}^*$ is parallel to $\mathbf{B}^*$. As a corollary, the hyperplane $\mathbf{E}^*$ and the linear functional $E^*(x) = \pi(x^0, x)$ are one-to-one correspondent. Therefore, we can use $E^*(x)$ as the dual to E and arrive at the representation $(\mathbf{C}^*, t, \mathbf{B}^*, \mathbf{L}^*, \pi^*, E^*)$ for a marked GBF. We write t before $\mathbf{B}^*$, π^*, and E^* because all these objects depend on the choice of t. Intuitively, $(\mathbf{C}^*, t, \mathbf{B}^*, \mathbf{L}^*, \pi^*, E^*)$ expresses beliefs about which element of $\mathbf{V}^*$ is the true configuration of $\mathbf{V}$. We are certain that the true configuration is on the hyperplane $\mathbf{C}^*$ (the certainty hyperplane). Within $\mathbf{C}^*$, our belief is distributed over ellipsoidal cylinders around a smaller-dimensional hyperplane $\mathbf{E}^*$ (the expectation hyperplane) parallel to $\mathbf{B}^*$. The wide-sense inner product π^* (the concentration inner product) specifies the shape, scale, and direction of the ellipsoidal cylinders, and the linear functional E^* (the location functional) specifies $\mathbf{E}^*$ by giving its inner product with every other hyperplane parallel to $\mathbf{B}^*$ within $\mathbf{C}^*$. We call $\mathbf{B}^*$ the no-opinion-expressed space, since the GBF does not express any opinions about where the true configuration is along its coordinates. Similarly, we call $\mathbf{L}^*$ the no-opinion-allowed hyperplane, since the GBF, so long as it has the label $\mathbf{L}$, is not allowed to express any opinions about where the true configuration is along its coordinates.

EXAMPLE 2 Consider the GBF described in Example 1. If we choose X, Y, Z as a basis for $\mathbf{V}$, then each linear functional on $\mathbf{V}$ can be represented by a sample point (x, y, z) and $\mathbf{V}^* = \{(x, y, z) | x, y, z \in \mathbb{R}\}$. Let $t = (t_1, t_2, t_3)$. Since t agrees with E on $\mathbf{C}$,

$$t(X + Y + Z) = (t_1, t_2, t_3)(1, 1, 1)^T = t_1 + t_2 + t_3 = E(X + Y + Z) = 1.$$

Thus, t is a point on the hyperplane $x + y + z = 1$. Note that **C** is spanned by $X + Y + Z$, **B** is spanned by $X + Y + Z$ and $X + Y$, and **L** = **V**. Obviously, $\mathbf{L}^* = \{t\}$. The hyperplanes $\mathbf{C}^*$, $\mathbf{B}^*$, and $\mathbf{E}^*$ are as follows:

$$\mathbf{C}^* = \{(x, y, z) | z + y + z = 1\},$$

$$\mathbf{B}^* = \{(x, y, z) | x + y + z = 1, x + y = t_1 + t_2\},$$

$$\mathbf{E}^* = \{(x, y, z) | x + y + z = 1, x + y = 0.5\}.$$

Figure 1 shows these hyperplanes graphically. The GBF has no opinion about the true sample value along the dimension of solid lines in Figure 1. It has a Gaussian distribution on the hyperplane $\mathbf{C}^*$, which describes how likely it is that the true value lies on a line that is parallel to $\mathbf{E}^*$ and $\mathbf{B}^*$. Unfortunately, the distribution cannot be written explicitly in the current coordinate system. We choose $U = X + Y + Z$, $V = X + Y$, $W = X$ as another basis for **V**. Let $\mathbf{V}^* = \{(u, v, w) | u, v, w \in \mathbb{R}\}$ and $t = (1, t_2, t_3)$. Then, $\mathbf{C}^* = \{(u, v, w) | u = 1\}$, $\mathbf{B}^* = \{(u, v, w) | u = 1, v = t_2\}$, and $\mathbf{C}^* = \{(u, v, w) | u = 1, v = 0.5\}$. π^* and E^* are as follows:

$$\pi^*[(1, v', w'), (1, v, w)] = \tfrac{1}{2}(v' - t_2)(v - t_2),$$

$$E^*(1, v, w) = \tfrac{1}{2}(0.5 - t_2)(v - t_2).$$

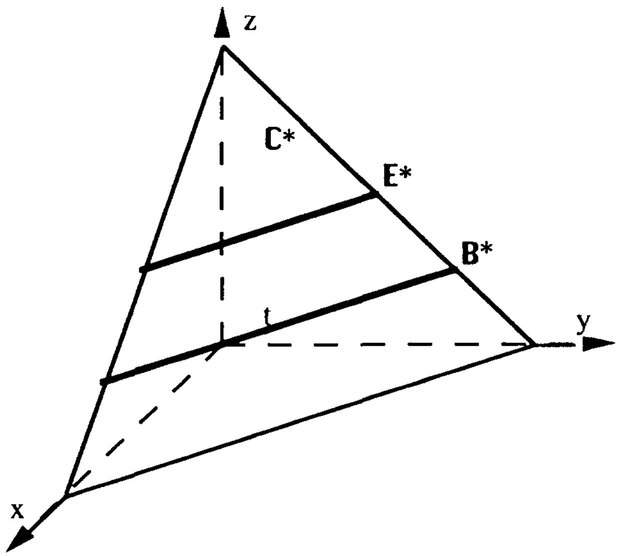

Figure 1. The graphical representation of $\mathbf{C}^*$, $\mathbf{B}^*$, and $\mathbf{E}^*$ in the (x, y, z) coordinate system.

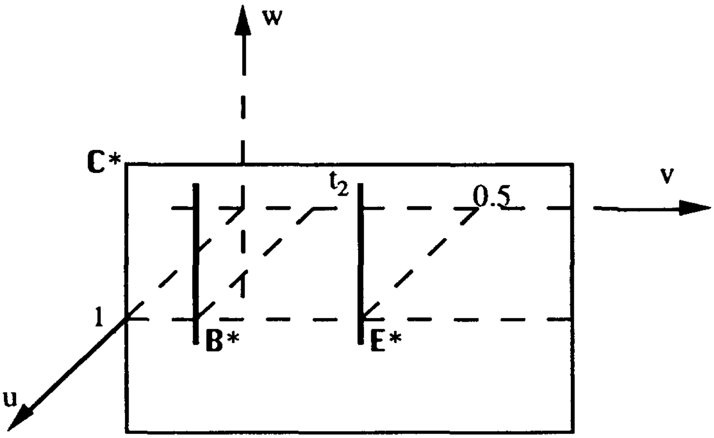

Figure 2. The graphical representation of **C***, **B***, and **E*** in the (u, v, w) coordinate system.

Based on the new coordinate system, the GBF is graphically shown in Figure 2.

4. COMBINING GAUSSIAN BELIEF FUNCTIONS

In this section we adapt Dempster's rule (2) to the case of GBFs. We achieve this progressively. We first define the combination for special cases and gradually generalize it into the full generality. As can be seen easily in Section 3.1, after we choose an appropriate basis for a variable space, each GBF consists of a Bayesian belief function for some variables and a vacuous belief function for others. Since the vacuous components do not contribute to knowledge, the combination of two GBFs is essentially the combination of their corresponding Bayesian components. Therefore, we can treat the combination of GBFs as a special case of the combination of continuous Bayesian belief functions. Following this logic, we first derive a rule for combining GBFs in a variable space. The resulting rule depends on the choice of an appropriate basis in the variable space. At this time, we are not aware of whether it can be represented in a coordinate-free way. In contrast with marginalization, combination can be most naturally described in a sample space. As Dempster (1990a, b) suggested, combination in a sample space can be phrased in coordinate-free terms as intersections of hyperplanes and additions of wide-sense inner products. In the second part of this section, we formally represent Dempster's suggested rule and show its consistency with the corresponding rule in a variable space.

4.1. Combination by Dempster's Rule

Given any continuous random vector X, a Bayesian belief function for it has singleton focal elements $\{x\}$. Its basic probability assignment can be represented by a function, say $f(x)$, which specifies the belief density committed to assertion $\{x\}$. Now suppose $f_1(x)$ and $f_2(x)$ correspond to two Bayesian belief functions for X. Let $\{x\}$ and $\{x'\}$ denote their focal elements, respectively. Since $\{x\} \cap \{x'\} = \varnothing$ if $x \neq x'$, the elements $\{x\}$ and $\{x'\}$ are consistent assertions only if $x = x'$. Discarding the belief committed to $\varnothing$, the total belief committed to all the possible joint assertions is $\int f_1(x) f_2(x)\, dx$. The total belief committed to the joint assertion $X \in (x, x + \Delta x)$ is $\int_x^{x+\Delta x} f_1(x) f_2(x)\, dx$. Therefore, the density function for the combined belief function, denoted by $f_1(x) \otimes f_2(x)$, is as follows by Dempster's rule:

$$f_1(x) \otimes f_2(x) = \alpha^{-1} f_1(x) f_2(x), \tag{11}$$

where $\alpha = \int f_1(x) f_2(x)\, dx$. Note that $f_1(x) \otimes f_2(x) \geq 0$ and $\int f_1(x) \otimes f_2(x)\, dx = 1$. Thus, $f_1(x) \otimes f_2(x)$ is indeed a probability density function.

In the special case when $f_1(x)$ and $f_2(x)$ are Gaussian, we can represent $f_1(x) \otimes f_2(x)$ explicitly. Let $d(x, \boldsymbol{\Sigma}, \mu)$ denote an n-dimensional Gaussian density function with mean μ and covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$d[x, \boldsymbol{\Sigma}, \mu] = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\tfrac{1}{2}(x - \mu) \boldsymbol{\Sigma}^{-1} (x - \mu)^T \right\}, \tag{12}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. Then, we have

LEMMA 3 *Let* $f_1(x) = d(x, \boldsymbol{\Sigma}^1, \mu^1)$ *and* $f_2(x) = d(x, \boldsymbol{\Sigma}^2, \mu^2)$. *Then* $f_1(x) \otimes f_2(x) = d(x, \sigma, a)$, *where*

$$\sigma = \left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1} \right]^{-1} \quad \textit{and}$$

$$a = \left[\mu^1 (\boldsymbol{\Sigma}^1)^{-1} + \mu^2 (\boldsymbol{\Sigma}^2)^{-1} \right] \left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1} \right]^{-1}. \tag{13}$$

Proof According to (12), we can verify that

$$f_1(x) f_2(x) = \frac{1}{(2\pi)^n |\boldsymbol{\Sigma}^1 \boldsymbol{\Sigma}^2|^{1/2}} \exp\left(-\frac{(x - a)\sigma^{-1}(x - a)^T - R}{2} \right),$$

where a and σ are expressed in (13) and

$$\begin{aligned} R &= a\sigma^{-1}a^T - \mu^1(\boldsymbol{\Sigma}^1)^{-1}(\mu^1)^T - \mu^2(\boldsymbol{\Sigma}^2)^{-1}(\mu^2)^T \\ &= \mu^1(\boldsymbol{\Sigma}^1)^{-1}\left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1}\right]^{-1}(\boldsymbol{\Sigma}^1)^{-1}(\mu^1)^T \\ &\quad + \mu^2(\boldsymbol{\Sigma}^2)^{-1}\left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1}\right]^{-1}(\boldsymbol{\Sigma}^2)^{-1}(\mu^2)^T \\ &\quad + 2\mu^1(\boldsymbol{\Sigma}^1)^{-1}\left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1}\right]^{-1}(\boldsymbol{\Sigma}^2)^{-1}(\mu^2)^T \\ &\quad - \mu^1(\boldsymbol{\Sigma}^1)^{-1}(\mu^1)^T - \mu^2(\boldsymbol{\Sigma}^2)^{-1}(\mu^2)^T. \end{aligned}$$

It follows from

$$\begin{aligned} \left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1}\right]^{-1} &= \boldsymbol{\Sigma}^1 - \boldsymbol{\Sigma}^1[\boldsymbol{\Sigma}^1 + \boldsymbol{\Sigma}^2]^{-1}\boldsymbol{\Sigma}^1 \\ &= \boldsymbol{\Sigma}^2 - \boldsymbol{\Sigma}^2[\boldsymbol{\Sigma}^1 + \boldsymbol{\Sigma}^2]^{-1}\boldsymbol{\Sigma}^2 \end{aligned}$$

that R can be simplified as

$$R = -(\mu^1 - \mu^2)[\boldsymbol{\Sigma}^1 + \boldsymbol{\Sigma}^2]^{-1}(\mu^1 - \mu^2)^T.$$

Therefore,

$$\begin{aligned} &f_1(x)f_2(x) \\ &\quad = \frac{1}{(2\pi)^n|\boldsymbol{\Sigma}^1\boldsymbol{\Sigma}^2|^{1/2}} \\ &\quad\quad \times \exp\left(-\frac{(x-a)\sigma^{-1}(x-a)^T + (\mu^1 - \mu^2)[\boldsymbol{\Sigma}^1 + \boldsymbol{\Sigma}^2]^{-1}(\mu^1 - \mu^2)^T}{2}\right) \end{aligned}$$

Note that

$$(\boldsymbol{\Sigma}^1 + \boldsymbol{\Sigma}^2)^{-1} = (\boldsymbol{\Sigma}^1)^{-1}\left[(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1}\right]^{-1}(\boldsymbol{\Sigma}^2)^{-1}.$$

Thus,

$$|\boldsymbol{\Sigma}^1 + \boldsymbol{\Sigma}^2|^{-1/2}|\boldsymbol{\Sigma}^1\boldsymbol{\Sigma}^2|^{1/2} = |(\boldsymbol{\Sigma}^1)^{-1} + (\boldsymbol{\Sigma}^2)^{-1}|^{-1/2}.$$

Therefore,

$$f_1(x)f_2(x) = \frac{1}{(2\pi)^{n/2}\left|\left[(\Sigma^1)^{-1} + (\Sigma^2)^{-1}\right]^{-1}\right|^{1/2}}$$
$$\times \exp\left\{-\frac{1}{2}(x-a)\left[(\Sigma^1)^{-1} + (\Sigma^2)^{-1}\right](x-a)^T\right\}$$
$$\times \frac{1}{(2\pi)^{n/2}|\Sigma^1 + \Sigma^2|^{1/2}}$$
$$\times \exp\left\{\tfrac{1}{2}(\mu^1 - \mu^2)(\Sigma^1 + \Sigma^2)^{-1}(\mu^1 - \mu^2)^T\right\}.$$

Therefore,

$$\int f_1(x)f_2(x)\,dx = \frac{1}{(2\pi)^{n/2}|\Sigma^1 + \Sigma^2|^{1/2}}$$
$$\times \exp\left\{\tfrac{1}{2}(\mu^1 - \mu^2)(\Sigma^1 + \Sigma^2)^{-1}(\mu^1 - \mu^2)^T\right\}.$$

and according to (13)

$$f_1(x) \otimes f_2(x) = \frac{1}{(2\pi)^{n/2}\left|\left[(\Sigma^1)^{-1} + (\Sigma^2)^{-1}\right]^{-1}\right|^{-1/2}}$$
$$\times \exp\left\{-\tfrac{1}{2}(x-a)\left[(\Sigma^1)^{-1} + (\Sigma^2)^{-1}\right](x-a)^T\right\}. \quad \blacksquare$$

Now we define the combination of two continuous Bayesian belief functions bearing on different sets of variables. Suppose $f_1(x_1, x_2)$ is the density function for random variable sets X_1 and X_2, and $f_2(x_1, x_3)$ for X_1 and X_3, where X_2 and X_3 are disjoint. Their focal elements are singletons, denoted by $\{(x_1, x_2)\}$ and $\{(x_1', x_3)\}$, respectively. By (3),

$$\{(x_1, x_2)\}^{\downarrow X_1} \cap \{(x_1', x_3)\}^{\downarrow X_1} \neq \emptyset \quad \text{iff} \quad x_1 = x_1'.$$

Discarding the belief committed to $\emptyset$, the total belief committed to all the possible joint assertions is $\alpha = \iint f_1(x_1, x_2)f_2(x_1, x_3)\,dx_1\,dx_2\,dx_3$. Therefore, the density function for X_1, X_2, and X_3 in the combined belief function, denoted by $f_1(x_1, x_2) \otimes f_2(x_1, x_3)$, is as follows by Dempster's rule:

$$f_1(x_1, x_2) \otimes f_2(x_1, x_3) = \alpha^{-1} f_1(x_1, x_2) f_2(x_1, x_3). \tag{14}$$

Since $f_1(x_1, x_2) = f_1(x_1)f_1(x_2 \mid X_1 = x_1)$ and $f_2(x_1, x_3) = f_2(x_1)f_2(x_3 \mid X_1 = x_1)$, we can verify that $\alpha = \int f_1(x_1)f_2(x_1)\,dx_1$. Then, according to (11), we have

$$f_1(x_1, x_2) \otimes f_2(x_1, x_3) = \{f_1(x_1) \otimes f_2(x_1)\}f_1(x_2 | X_1 = x_1)f_2(x_3 | X_1 = x_1). \tag{15}$$

In words, the combined density function is the product of the combination of the marginal density functions on the common variables and the conditional density functions given the value of the common variables. It is interesting to note that (15) indicates the conditional independence between X_2 and X_3 given X_1. As a basic property, marginals and conditionals of a Gaussian distribution are still Gaussian. Thus, according to Lemma 3, $f_1(x_1) \otimes f_2(x_1)$ is Gaussian, and so is $f_1(x_1, x_2) \otimes f_2(x_1, x_3)$ by (15) if $f_1(x_1, x_2)$ and $f_2(x_1, x_3)$ are both Gaussian.

LEMMA 4 *Assume $f_1(x_1, x_2)$ and $f_2(x_1, x_3)$ are Gaussian. Assume*

$$f_1(x_1) \otimes f_2(x_1) = d(x_1, \sigma_1, a_1), \tag{16}$$

$$f_1(x_2 | X_1 = x_1) = d\left[x_2, \sigma_2, a_2 + x_1(b_2)^T\right], \tag{17}$$

$$f_2(x_3 | X_1 = x_1) = d\left[x_3, \sigma_3, a_3 + x_1(b_3)^T\right]. \tag{18}$$

Then

$$f_1(x_1, x_2) \otimes f_2(x_1, x_3) = d\left[(x_1, x_2, x_3), \Omega, \left(a_1, a_2 + a_1(b_2)^T, a_3 + a_1(b_3)^T\right)\right],$$

where

$$\Omega = \begin{pmatrix} \sigma_1 & \sigma_1(b_2)^T & \sigma_1(b_3)^T \\ b_2\sigma_1 & \sigma_2 + b_2\sigma_1(b_2)^T & b_2\sigma_1(b_3)^T \\ b_3\sigma_1 & b_3\sigma_1(b_2)^T & \sigma_3 + b_3\sigma_1(b_3)^T \end{pmatrix}.$$

Proof Assume n_1, n_2, and n_3 are respectively the dimension numbers of X_1, X_2, and X_3. According to (15)–(18), we can verify that

$$f_1(x_1, x_2) \otimes f_2(x_1, x_3) = \frac{1}{(2\pi)^{(n_1+n_2+n_3)/2}(|\sigma_1| \times |\sigma_2| \times |\sigma_3|)^{1/2}} \times \exp\{-\tfrac{1}{2}g(x_1, x_2, x_3)\},$$

where

$$g(x_1, x_2, x_3) = (x_1 - a_1)(\sigma_1)^{-1}(x_1 - a_1)^T + \left[x_2 - a_2 - x_1(b_2)^T\right](\sigma_2)^{-1}\left[x_2 - a_2 - x(b_2)^T\right]^T + \left[x_3 - a_3 - x_1(b_3)^T\right](\sigma_3)^{-1}\left[x_3 - a_3 - x_1(b_3)^T\right]^T.$$

By (15), $f_1(x_1, x_2) \otimes f_2(x_1, x_3)$ is Gaussian and its marginal on X_1 is $f_1(x_1) \otimes f_2(x_1)$. Thus, $E(X_1) = a_1$. Furthermore, by (17), we have $E(X_2) = E[E(X_2|X_1)] = E[a_2 + X_1(b_2)^T] = a_2 + a_1(b_2)^T$. Similarly, we have $E(X_3) = a_3 + a_1(b_3)^T$. Therefore, there exist a three-dimensional symmetric matrix (w_{ij}) such that

$$g(x_1, x_2, x_3) = \left(x_1 - a_1, x_2 - a_2 - a_1(b_2)^T, x_3 - a_3 - a_1(b_3)^T\right)(w_{ij}) \times \left(x_1 - a_1, x_2 - a_2 - a_1(b_2)^T, x_3 - a_3 - a_1(b_3)^T\right)^T.$$

By matching the coefficients of $x_i x_j$ $(i, j = 1, 2, 3)$, we can show that

$$w_{11} = (\sigma_1)^{-1} + (b_2)^T(\sigma_2)^{-1}b_2 + (b_3)^T(\sigma_3)^{-1}b_3,$$

$$w_{12} = -(b_2)^T(\sigma_2)^{-1}, \qquad w_{13} = -(b_3)^T(\sigma_3)^{-1}, \qquad w_{22} = (\sigma_2)^{-1},$$

$$w_{23} = 0, \quad \text{and} \quad w_{33} = (\sigma_3)^{-1}.$$

Therefore,

$$(w_{ij}) = \begin{pmatrix} I & -(b_2)^T & -(b_3)^T \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} (\sigma_1)^{-1} & 0 & 0 \\ 0 & (\sigma_2)^{-1} & 0 \\ 0 & 0 & (\sigma_3)^{-1} \end{pmatrix} \times \begin{pmatrix} I & 0 & 0 \\ -b_2 & I & 0 \\ -b_3 & 0 & I \end{pmatrix},$$

$$(w_{ij})^{-1} = \begin{pmatrix} I & 0 & 0 \\ b_2 & I & 0 \\ b_3 & 0 & I \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} I & (b_2)^T & (b_3)^T \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} = \Omega,$$

and $|\Omega| = |\sigma_1| \times |\sigma_2| \times |\sigma_3|$. ■

Finally we extend Dempster's rule to the general case when two continuous belief functions contain deterministic variables, whose value is known with certainty. Without loss of generality, we assume that two belief functions Bel_1 and Bel_2 share a common deterministic vector D_1 and a common random vector X_1. The vector U is deterministic in Bel_1 but uncertain in Bel_2, and V vice versa. Bel_1 is certain about D_2 and uncertain about X_2. However, Bel_2 has no opinion about either D_2 or X_2. Similarly, Bel_2 is certain about D_3 and uncertain about X_3, but Bel_1 has no opinion about either D_3 or X_3. In summary, Bel_1 bears on deterministic vectors D_1, D_2, U and random vectors V, X_1, X_2, and Bel_2 on deterministic vectors D_1, D_3, V and random vectors U, X_1, X_3. The hypergraph representing Bel_1 and Bel_2 is shown in Figure 3. Since D_1 is a common deterministic vector, its value must be the same in both Bel_1 and Bel_2, because otherwise there are no possible joint assertions. Let $D_1 = d_1$ in both Bel_1 and Bel_2. Assume $D_2 = d_2$, $U = u$ and $D_3 = d_3$, $V = v$ with certainty in Bel_1 and Bel_2, respectively. Then focal elements for Bel_1 can be represented by $(d_1, d_2, u, v', x_1, x_2)$, and for Bel_2 by $(d_1, d_3, u', v, x_1', x_3)$. Each pair of focal elements is nonconflicting iff $u = u'$, $v = v'$, and $x_1 = x_1'$. Therefore, for the existence of possible joint assertions, U is restricted to take values u in Bel_2 and V is restricted to take values v in Bel_1. Consequently, in the combined belief function $\text{Bel}_1 \otimes \text{Bel}_2$, U and V become deterministic and

$$D_1 = d_1, \qquad D_2 = d_2, \qquad D_3 = d_3, \qquad U = u, \qquad V = v. \tag{19}$$

Let $f_1(v, x_1, x_2)$ and $f_2(u, x_1, x_3)$ be respectively the density functions for Bel_1 and Bel_2. Then the total belief committed to all the possible joint assertions is

$$\alpha = \int f_1(v, x_1, x_2) f_2(u, x_1, x_3)\, dx_1\, dx_2\, dx_3.$$

Therefore, the belief density for X_1, X_2, and X_3 in $\text{Bel}_1 \otimes \text{Bel}_2$ is

$$f_1(v, x_1, x_2) \otimes f_2(u, x_1, x_3) = \alpha^{-1} f_1(v, x_1, x_2) f_2(u, x_1, x_3). \tag{20}$$

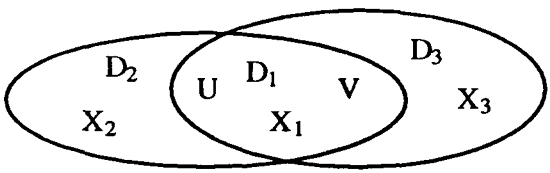

Figure 3. The belief network for Bel_1 and Bel_2.

Since $f_1(v, x_1, x_2) = f_1(v)f_1(x_1, x_2 \mid V = v)$ and $f_2(u, x_1, x_3) = f_2(u)f_2(x_1, x_3 \mid U = u)$, applying (15) and (20) leads to

$$\begin{aligned} & f_1(v, x_1, x_2) \otimes f_2(u, x_1, x_3) \\ & \quad = [f_1(x_1|V = v) \otimes f_2(x_1|U = u)] \\ & \quad\quad f_1(x_2|V = v, X_1 = x_1)f_2(x_3|U = u, X_1 = x_1). \end{aligned} \tag{21}$$

Given two GBFs $\text{Bel}_1 = (\mathbf{C}_1, \mathbf{B}_1, \mathbf{L}_1, \pi^1, E^1)$ and $\text{Bel}_2 = (\mathbf{C}_2, \mathbf{B}_2, \mathbf{L}_2, \pi^2, E^2)$, in the following we will use (21) and Lemmas 3 and 4 to obtain their combination

$$\text{Bel} = \text{Bel}_1 \otimes \text{Bel}_2 = (\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E).$$

It can be seen easily from Dempster's rule that $\mathbf{C} = \mathbf{C}_1 \oplus \mathbf{C}_2, \mathbf{B} = \mathbf{B}_1 \oplus \mathbf{B}_2$, and $\mathbf{L} = \mathbf{L}_1 \oplus \mathbf{L}_2$. However, at this time we are not able to represent E and π in a coordinate-free way. Instead we choose a convenient basis $D_1, D_2, D_3, U, V, X_1, X_2, X_3, \dots$ such that $\mathbf{C}_1$ is spanned by $\{D_1, D_2, U\}$, $\mathbf{C}_2$ by $\{D_1, D_3, V\}$, $\mathbf{B}_1$ by $\{D_1, D_2, U, V, X_1, X_2\}$, and $\mathbf{B}_2$ by $\{D_1, D_3, V, U, X_1, X_3\}$. There might be some other variables in $\mathbf{L}_1 \oplus \mathbf{L}_2$ that are not in $\mathbf{B}_1 \oplus \mathbf{B}_2$. However, specifying them is not necessary, because E and π are defined in $\mathbf{B}_1 \oplus \mathbf{B}_2$. As we know, when a basis is chosen, π^i and E^i $(i = 1, 2)$ are specified by the corresponding mean vectors and covariance matrices. The following theorem then shows how the mean vector for E and the covariance matrix for π can be represented by them.

THEOREM 1 *Given any two GBFs*

$\text{Bel}_1 = (\mathbf{C}_1, \mathbf{B}_1, \mathbf{L}_1, \pi^1, E^1)$, *where* $\mathbf{C}_1$ *is spanned by* $\{D_1, D_2, U\}$ *and* $\mathbf{B}_1$ *by* $\{D_1, D_2, U, V, X_1, X_2\}$, E^1 *is determined by the mean vector* $(d_1, d_2, u, \mu_0^1, \mu_1^1, \mu_2^1)$, *and* π^1 *is determined by the covariance matrix* $\mathbf{\Sigma}_{ij}^1$ $(i, j = 0, 1, 2)$, *and*

$\text{Bel}_2 = (\mathbf{C}_2, \mathbf{B}_2, \mathbf{L}_2, \pi^2, E^2)$, *where* $\mathbf{C}_2$ *is spanned by* $(D_1, D_3, V\}$, $\mathbf{B}_2$ *by* $\{D_1, D_3, V, U, X_1, X_3\}$, E^2 *is determined by the mean vector* $(d_1, d_3, v, \mu_0^2, \mu_1^2, \mu_3^2)$, *and* π^2 *is determined by the covariance matrix* $\mathbf{\Sigma}_{ij}^2$ $(i, j = 0, 1, 3)$,

their combination $\text{Bel}_1 \otimes \text{Bel}_2$ *is the GBF*

$$\text{Bel} = (\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$$

where $\mathbf{C} = \mathbf{C}_1 \oplus \mathbf{C}_2$, $\mathbf{B} = \mathbf{B}_1 \oplus \mathbf{B}_2$, $\mathbf{L} = \mathbf{L}_1 \oplus \mathbf{L}_2$, E *is determined by the mean vector*

$$\left(d_1, d_2, d_3, u, v, a_1, a_2 + a_1(b_2)^T, a_3 + a_1(b_3)^T\right), \tag{22}$$

and π is determined by the covariance matrix

$$\Omega = \begin{pmatrix} \sigma_1 & \sigma_1(b_2)^T & \sigma_1(b_3)^T \\ b_2\sigma_1 & \sigma_2 + b_2\sigma_1(b_2)^T & b_2\sigma_1(b_3)^T \\ b_3\sigma_1 & b_3\sigma_1(b_2)^T & \sigma_3 + b_3\sigma_1(b_3)^T \end{pmatrix}, \tag{23}$$

where

$$\sigma_1 = \left[(\Sigma^1)^{-1} + (\Sigma^2)^{-1}\right]^{-1}, \tag{24}$$

$$a_1 = \left[\mu^1(\Sigma^1)^{-1} + \mu^2(\Sigma^2)^{-1}\right]\left[(\Sigma^1)^{-1} + (\Sigma^2)^{-1}\right]^{-1}, \tag{25}$$

$$\Sigma^i = \Sigma^i_{11} - \Sigma^i_{10}(\Sigma^i_{00})^{-1}\Sigma^i_{01} \qquad (i = 1,2), \tag{26}$$

$$\mu^1 = \mu^1_1 + (v - \mu^1_0)(\Sigma^1_{00})^{-1}\Sigma^1_{01}, \tag{27}$$

$$\mu^2 = \mu^2_1 + (u - \mu^2_0)(\Sigma^2_{00})^{-1}\Sigma^2_{01}, \tag{28}$$

$$\sigma_2 = \Sigma^1_{22} - (\Sigma^1_{20}, \Sigma^1_{21})\begin{pmatrix} \Sigma^1_{00} & \Sigma^1_{01} \\ \Sigma^1_{10} & \Sigma^1_{11} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma^1_{02} \\ \Sigma^1_{12} \end{pmatrix}, \tag{29}$$

$$a_2 + x_1(b_2)^T = \mu^1_2 + (v - \mu^1_0, x_1 - \mu^1_1)\begin{pmatrix} \Sigma^1_{00} & \Sigma^1_{01} \\ \Sigma^1_{10} & \Sigma^1_{11} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma^1_{02} \\ \Sigma^1_{12} \end{pmatrix}, \tag{30}$$

$$\sigma_3 = \Sigma^2_{33} - (\Sigma^2_{30}, \Sigma^2_{31})\begin{pmatrix} \Sigma^2_{00} & \Sigma^2_{01} \\ \Sigma^2_{10} & \Sigma^2_{11} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma^2_{03} \\ \Sigma^2_{13} \end{pmatrix}, \tag{31}$$

$$a_3 + x_1(b_3)^T = \mu^2_3 + (u - \mu^2_0, x_1 - \mu^2_1)\begin{pmatrix} \Sigma^2_{00} & \Sigma^2_{01} \\ \Sigma^2_{10} & \Sigma^2_{11} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma^2_{03} \\ \Sigma^2_{13} \end{pmatrix}. \tag{32}$$

Proof From (19) it is easy to see that $\{D_1, D_2, D_3, U, V\}$ spans **C**. Thus, $\mathbf{C} = \mathbf{C}_1 \oplus \mathbf{C}_2$. According to (21), Bel has opinions about X_1, X_2, and X_3. Therefore, **B** is spanned by $\{D_1, D_2, D_3, U, V, X_1, X_2, X_3\}$. Hence we have $\mathbf{B} = \mathbf{B}_1 \oplus \mathbf{B}_2$. That $\mathbf{L} = \mathbf{L}_1 \oplus \mathbf{L}_2$ is obvious from Dempster's rule (2). It is a standard property of Gaussian distributions that $f_1(x_1 \mid V = v) = d(x_1, \Sigma^1, \mu^1)$ and $f_2(x_1 \mid U = u) = d(x_1, \Sigma^2, \mu^2)$, where Σ^i and μ^i $(i = 1,2)$ are shown in (26)–(28). By Lemma 3,

$$f_1(x_1|V = v) \otimes f_2(x_1|U = u) = d(x_1, \sigma_1, a_1).$$

By (29)–(32), it is also standard that

$$f_1(x_2|V = v, X_1 = x_1) = d\left(x_2, \sigma_2, a_2 + x_1(b_2)^T\right),$$

$$f_2(x_3|U = u, X_1 = x_1) = d\left(x_3, \sigma_3, a_3 + x_1(b_3)^T\right).$$

According to (21) and Lemma 4,

$$f_1(v, x_1, x_2) \otimes f_2(u, x_1, x_3)$$
$$= d\left[(x_1, x_2, x_3), \Omega, \left(a_1, a_2 + a_1(b_2)^T, a_3 + a_1(b_3)^T\right)\right],$$

where Ω is shown by (22). ■

In (26)–(32), $\mathbf{\Sigma}^1$ and μ^1 are respectively the conditional variance and mean of X_1 given $V = v$ in $f_1(v, x_1, x_2)$; $\mathbf{\Sigma}^2$ and μ^2 are respectively the conditional variance and mean of X_1 given $U = u$ in $f_2(u, x_1, x_3)$; a_2 and b_2 are the regression coefficients of X_2 against X_1 in $f_1(x_1, x_2 \mid V = v)$; a_3 and b_3 are the regression coefficients of X_3 against X_1 in $f_2(x_1, x_3 \mid U = u)$. In words, the combination of two GBFs is done by the following four-step procedure:

1. The certainty space of the combined GBF is the orthogonal sum of the certainty spaces of the component GBFs: A piece of evidence that supports a certainty space will be adopted as a fact in combination. If a variable is believed to take a value with certainty by one component GBF, it is believed so by the combined GBF no matter how another component GBF feels about the variable.
2. The belief space of the combined GBF is the orthogonal sum of the belief spaces of the component GBFs: The beliefs expressed by any component GBF will not be lost in the combination. If one component GBF has opinions about a certain variable, the combined GBF will adopt and somehow revise the opinions in accordance with another component GBF.
3. Suppose $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}$ are respectively the uncertainty spaces of Bel_1, Bel_2, and Bel. Given the basis of $\mathbf{C}$, compute the conditional means and variances for the basis of $\mathbf{F}_1 \cap \mathbf{F}_2$ in both distributions, the regression coefficients of the basis of $\mathbf{F}_1 - \mathbf{F}_1 \cap \mathbf{F}_2$ against the basis of $\mathbf{F}_1 \cap \mathbf{F}_2$, and the regression coefficients of the basis of $\mathbf{F}_2 - \mathbf{F}_1 \cap \mathbf{F}_2$ against the basis of $\mathbf{F}_1 \cap \mathbf{F}_2$ in the appropriate distribution.
4. Plug the results obtained in step 3 into (22)–(25), and get E and π.

Steps 1 and 2 imply that combination corresponds to knowledge integration. Steps 3 and 4 imply that the complex formulas of combination have some statistical semantics.

EXAMPLE 3 Let Bel_1 denote the GBF specified in Example 1. Let Bel_2 be another GBF bearing on random variables X, Y, and Z, which represents the following statistical models:

$$Z = 1.5X + 0.3 + \varepsilon_Z, \tag{33}$$

$$Y = 0.5Z + 0.1 + \varepsilon_Y, \tag{34}$$

where $X \sim N(0.2, 0.04)$, $\varepsilon_Z \sim N(0, 2)$, and $\varepsilon_Y \sim N(0, 1)$ are independent. The belief-network representation of the above model is shown in Figure 4. Using (12), Equations (33) and (34) can be also respectively represented by $f(z \mid X = x) = d(z, 2, 1.5x + 0.3)$ and $f(y \mid Z = z) = d(y, 1, 0.5z + 0.1)$. Noting that Y is conditionally independent of X given Z, the joint density function $f(x, y, z)$ can be obtained by multiplying $f(x)$ with $f(z \mid X = x)$ and $f(y \mid Z = z)$. We can also obtain $f(x, y, z)$ directly from (33) and (34) by computing the means and the covariances of X, Y, and Z. For example, $E[Z] = 1.5E[X] + 0.3 = 0.6$, $E[Y] = 0.5E[Z] + 0.1 = 0.4$, $\mathrm{Cov}(Y, Z) = \mathrm{Cov}(0.5Z + 0.1 + \varepsilon_Y, Z) = \mathrm{Cov}[0.5(1.5X + 0.3 + \varepsilon_Z) + 0.1 + \varepsilon_Y, 1.5X + 0.3 + \varepsilon_Z] = 1.125\ \mathrm{Var}(X) + 0.5\ \mathrm{Var}(\varepsilon_Z) = 1.045$, etc. The joint density function for Bel_2 is

$$f(x, y, z) = d\left[(x, y, z), \begin{pmatrix} 0.040 & 0.030 & 0.06 \\ 0.030 & 1.523 & 1.045 \\ 0.06 & 1.045 & 2.090 \end{pmatrix}, (0.2, 0.4, 0.6)\right].$$

Let us choose a common basis $U = X + Y + Z$, $V = X + Y$, $W = X$ for both Bel_1 and Bel_2. Then Bel_2 is represented by the distribution

$$f(u, v, w) = d\left[(u, v, w), \begin{pmatrix} 5.923 & 2.728 & 0.130 \\ 2.728 & 1.693 & 0.07 \\ 0.130 & 0.070 & 0.040 \end{pmatrix}, (1.2, 0.6, 0.2)\right].$$

According to Theorem 1, **C** is spanned by U and **B** is spanned by $\{U, V, W\}$. The common uncertain variable for both Bel_1 and Bel_2 is V, which is like X_1 in Theorem 1. Given $U = 1$, the conditional of V in Bel_1 is the distribution of V itself, $d(v, 2, 0.5)$. Thus, $\boldsymbol{\Sigma}^1 = 2$ and $\mu^1 = 0.5$. Since V is the only uncertain variable V in Bel_1, terms including $a_2 + v(b_2)^T$ and σ_2 in Theorem 1 do not exist. Given $U = 1$, the conditional distribution of V

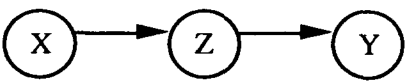

Figure 4. The belief network for Equations (33) and (34).

in Bel_2 is $d(v, 0.436, 0.508)$. Thus, $\boldsymbol{\Sigma}^2 = 0.436$ and $\mu^2 = 0.508$. Similarly, according to (31) and (32), we have

$$\sigma_3 = 0.04 - (0.13, 0.07)\begin{pmatrix} 5.923 & 2.728 \\ 2.728 & 1.693 \end{pmatrix}^{-1}\begin{pmatrix} 0.13 \\ 0.07 \end{pmatrix} = 0.037,$$

$$\begin{aligned} a_3 + v(b_3)^T &= 0.2 + (1 - 1.2, v - 0.6)\begin{pmatrix} 5.923 & 2.728 \\ 2.728 & 1.693 \end{pmatrix}^{-1}\begin{pmatrix} 0.13 \\ 0.07 \end{pmatrix} \\ &= 0.184 + 0.023v. \end{aligned}$$

Thus, $a_3 = 0.184$ and $b_3 = 0.023$. By (24) and (25) we have $\sigma_1 = 0.358$ and $a_1 = 0.507$. Plugging the above values into (22) and (23), we obtain the combined mean vector for U, V, and W as $(1, 0.507, 0.196)$, and the combined covariance matrix for V and W is

$$\begin{pmatrix} 0.358 & 0.008 \\ 0.008 & 0.037 \end{pmatrix}.$$

4.2. A Coordinate-Free Representation of Combination

The rule for combining GBFs in Section 4.1 depends on the choice of a coordinate system. In this section, we want to represent it alternatively in a coordinate-free way. The new representation turns out to be elegant and concise. It also helps us see the deep symmetry, i.e., commutativity and associativity, of combination. However, it does not imply any improvement of computational efficiencies. For the purpose of numerical computation, Liu (1995a, b) provides a third equivalent representation of the combination rule in terms of partial or full sweep operations, which essentially reduce combination of GBFs into spreadsheet manipulations.

THEOREM 2 *Suppose* Bel_1 *and* Bel_2 *are two marked GBFs represented in a sample space*:

$$\mathrm{Bel}_1 = (\mathbf{C}^{*1}, t, \mathbf{B}^{*1}, \mathbf{L}^{*1}, \pi^{*1}, E^{*1}),$$

$$\mathrm{Bel}_2 = (\mathbf{C}^{*2}, t, \mathbf{B}^{*2}, \mathbf{L}^{*2}, \pi^{*2}, E^{*2}),$$

where t is their common mark. Then

$$\begin{aligned} \mathrm{Bel}_1 \oplus \mathrm{Bel}_2 &= (\mathbf{C}^*, t, \mathbf{B}^*, \mathbf{L}^*, \pi^*, E^*) \\ &= (\mathbf{C}^{*1} \cap \mathbf{C}^{*2}, t, \mathbf{B}^{*1} \cap \mathbf{B}^{*2}, \mathbf{L}^{*1} \cap \mathbf{L}^{*2}, \pi^{*1}|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}} \\ &\quad + \pi^{*2}|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}, E^{*1}|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}} + E^{*2}|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}), \end{aligned}$$

where $\pi^{*i}|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}$ and $E^{*i}|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}$ are respectively the restrictions of π^{*i} and E^{*i} ($i = 1, 2$) to the intersection $\mathbf{C}^{*1} \cap \mathbf{C}^{*2}$.

Proof We prove Theorem 2 using Theorem 1. Thus, we assume Bel_1 and Bel_2 are the dual representations of the two GBFs $(\mathbf{C}_1, \mathbf{B}_1, \mathbf{L}_2, \pi^1, E^1)$ and $(\mathbf{C}_2, \mathbf{B}_2, \mathbf{L}_2, \pi^2, E^2)$, as defined in Theorem 1, whose combination in $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$ where $\mathbf{C} = \mathbf{C}_1 \oplus \mathbf{C}_2$, $\mathbf{B} = \mathbf{B}_1 \oplus \mathbf{B}_2$, $\mathbf{L} = \mathbf{L}_1 \oplus \mathbf{L}_2$, and E and π are specified by (22) and (23). We want to show $(\mathbf{C}^*, t, \mathbf{B}^*, \mathbf{L}^*, \pi^*, E^*)$ is dual to $(\mathbf{C}, \mathbf{B}, \mathbf{L}, \pi, E)$ with the mark t.

Given any linear functional t, suppose hyperplanes $\mathbf{S}^*$ and $\mathbf{T}^*$, defined as in (6), are dual to subspaces $\mathbf{S}$ and $\mathbf{T}$, respectively. Then, according to Dempster (1969), $\mathbf{S}^* \cap \mathbf{T}^*$ is the hyperplane dual to the subspace $\mathbf{S} \oplus \mathbf{T}$. Therefore, given t as a common mark, it is easy to see from Theorem 1 that $\mathbf{C}^* = \mathbf{C}^{*1} \cap \mathbf{C}^{*2}$, $\mathbf{B}^* = \mathbf{B}^{*1} \cap \mathbf{B}^{*2}$, and $\mathbf{L}^* = \mathbf{L}^{*1} \cap \mathbf{L}^{*2}$. Using the basis and the values of its deterministic variables specified in Theorem 1, the common mark t satisfies

$$t(D_1) = d_1, \qquad t(D_2) = d_2, \qquad t(D_3) = d_3, \qquad t(U) = u, \qquad t(V) = v.$$

Therefore, t can be written as point $(d_1, d_2, d_3, u, v, t_1, t_2, t_3, \ldots)$ in the sample space, where t_i is the value assigned to X_i by t, $i = 1, 2, 3$. Accordingly, $\mathbf{C}^*$ and $\mathbf{B}^*$ can be represented as follows:

$$\mathbf{C} = \mathbf{C}^{*1} \cap \mathbf{C}^{*2} = \{(d_1, d_2, d_3, u, v, x_1, x_2, x_3, \ldots)\},$$

$$\mathbf{B} = \mathbf{B}^{*1} \cap \mathbf{B}^{*2} = \{(d_1, d_2, d_3, u, v, t_1, t_2, t_3, \ldots)\}.$$

For any $x = (d_1, d_2, d_3, u, v, x_1, x_2, x_3, \ldots)$ and $x' = (d_1, d_2, d_3, u, v, x'_1, x'_2, x'_3, \ldots)$ in $\mathbf{C}^*$,

$$\pi^{*1}(x, x') = (v - v, x_1 - t_1, x_2 - t_2)\left(\boldsymbol{\Sigma}^1_{ij}\right)^{-1}(v - v, x'_1 - t_1, x'_2 - t_2)^T$$
$$= (x_1 - t_1, x_2 - t_2)G(x'_1 - t_1, x'_2 - t_2)^T,$$
$$\pi^{*2}(x, x') = (u - u, x_1 - t_1, x_3 - t_3)\left(\boldsymbol{\Sigma}^2_{ij}\right)^{-1}(u - u, x'_1 - t_1, x'_3 - t_3)^T$$
$$= (x_1 - t_1, x_3 - t_3)H(x'_1 - t_1, x'_3 - t_3)^T,$$

where

$$G = \left[\begin{pmatrix} \boldsymbol{\Sigma}^1_{11} & \boldsymbol{\Sigma}^1_{12} \\ \boldsymbol{\Sigma}^1_{21} & \boldsymbol{\Sigma}^1_{22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}^1_{10} \\ \boldsymbol{\Sigma}^1_{20} \end{pmatrix} (\boldsymbol{\Sigma}^1_{00})^{-1} (\boldsymbol{\Sigma}^1_{01}, \boldsymbol{\Sigma}^1_{02})\right]^{-1},$$

$$H = \left[\begin{pmatrix} \boldsymbol{\Sigma}^2_{11} & \boldsymbol{\Sigma}^2_{13} \\ \boldsymbol{\Sigma}^2_{31} & \boldsymbol{\Sigma}^2_{33} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Sigma}^2_{10} \\ \boldsymbol{\Sigma}^2_{30} \end{pmatrix} (\boldsymbol{\Sigma}^2_{00})^{-1} (\boldsymbol{\Sigma}^2_{01}, \boldsymbol{\Sigma}^2_{03})\right]^{-1}.$$

By some tedious computation we can verify that

$$G = \begin{pmatrix} (\boldsymbol{\Sigma}^1)^{-1} + (b_2)^T(\sigma_2)^{-1}b_2 & -(b_2)^T(\sigma_2)^{-1} \\ -(\sigma_2)^{-1}b_2 & (\sigma_2)^{-1} \end{pmatrix},$$

$$H = \begin{pmatrix} (\boldsymbol{\Sigma}^2)^{-1} + (b_3)^T(\sigma_3)^{-1}b_3 & -(b_3)^T(\sigma_3)^{-1} \\ -(\sigma_3)^{-1}b_3 & (\sigma_3)^{-1} \end{pmatrix},$$

where $\boldsymbol{\Sigma}^i$ $(i = 1,2)$, σ_i, a_i, and b_i $(i = 2,3)$ are listed in (26) and (29)–(32). Therefore,

$$\begin{aligned} &\pi^{*1}(x, x') + \pi^{*2}(x, x') \\ &\qquad = (x_1 - t_1, x_2 - t_2, x_3 - t_3)\Omega^{-1}(x'_1 - t_1, x'_2 - t_2, x'_3 - t_3)^T, \end{aligned}$$

where Ω is listed in (23). According to Theorem 1, $\pi^{*1} + \pi^{*2}$, when restricted to $\mathbf{C}^*$, is indeed π^*. Finally, for any point $x = (d_1, d_2, d_3, u, v, x_1, x_2, x_3, \dots)$ in $\mathbf{C}^*$,

$$\begin{aligned} E^{*1}(x) &= (\mu_0^1 - v, \mu_1^1 - t_1, \mu_2^1 - t_2)\left(\boldsymbol{\Sigma}_{ij}^1\right)^{-1}(v - v, x_1 - t_1, x_2 - t_2)^T \\ &= \left[(\mu_0^1 - v)Q + (\mu_1^1 - t_1, \mu_2^1 - t_2)G\right](x_1 - t_1, x_2 - t_2)^T, \\ E^{*2}(x) &= (\mu_0^2 - u, \mu_1^2 - t_1, \mu_3^2 - t_3)\left(\boldsymbol{\Sigma}_{ij}^2\right)^{-1}(u - u, x_1 - t_1, x_3 - t_3)^T \\ &= \left[(\mu_0^2 - u)R + (\mu_1^2 - t_1, \mu_3^2 - t_3)\right]H(x_1 - t_1, x_3 - t_3)^T, \end{aligned}$$

where G and H are as above, and Q and R are as follows:

$$Q = -(\boldsymbol{\Sigma}_{00}^1)^{-1}(\boldsymbol{\Sigma}_{01}^1, \boldsymbol{\Sigma}_{02}^1)G, \qquad R = -(\boldsymbol{\Sigma}_{00}^2)^{-1}(\boldsymbol{\Sigma}_{01}^2, \boldsymbol{\Sigma}_{03}^2)H.$$

$E^{*1}(x) + E^{*2}(x)$ is obviously a linear functional on $\mathbf{C}^*$ with null hyperplane $\mathbf{B}^*$. We can determine the location of its corresponding hyperplane that is parallel to $\mathbf{B}^*$ by (8) and (10) with $\boldsymbol{\Sigma}$ replaced by Ω in (8). By some straightforward but tedious computation, we can verify that the location is the point $(a_1, a_2 + a_1(b_2)^T, a_3 + a_1(b_3)^T)$, which, according to Theorem 1, is the mean vector for X_1, X_2, and X_3 in the combined GBF. Therefore, $E^{*1}(x) + E^{*2}(x) = E^*(x)$ when x is in $\mathbf{C}^*$. ■

Note that Theorem 2 is intuitive. As we see from Section 3.2, all the focal elements of Bel_i are the hyperplanes that are on $\mathbf{C}^{*i}$ and parallel to $\mathbf{B}^{*i}$, $i = 1,2$. Therefore, by Dempster's rule, $\mathbf{B}^{*1} \cap \mathbf{B}^{*2}$ is a typical focal element for the combined belief function. Its associated basic probability assignment is obtained by multiplying the basic probabilities assigned to

$\mathbf{B}^{*1}$ and $\mathbf{B}^{*2}$ and an appropriate normalization constant. Thus, the log "density" of the combined GBF is the sum of the component log "densities." Therefore, $\pi^* = \pi^{*1}|_{\mathbf{C}^{*1}\cap\mathbf{C}^{*2}} + \pi^{*2}|_{\mathbf{C}^{*1}\cap\mathbf{C}^{*2}}$ makes sense.

EXAMPLE 4 Let Bel_1 be the GBF specified in Example 2, and Bel_2 be dual to the Bel_2 specified in Example 3. As in Examples 2 and 3, we choose U, V, W as a basis. Since $t(U) = 1$, let the common mark $t = (1, 0.3, 2)$. Then $\mathbf{C}^{*1} = \{(1, v, w)\}$, $\mathbf{B}^{*1} = \{(1, 0.3, w)\}$, $\mathbf{L}^{*1} = \{(1, 0.3, 2)\}$, and for any two points $(1, v, w)$ and $(1, v', w')$ in $\mathbf{C}^{*1}$, we have $\pi^{*1}[(1, v', w'), (1, v, w)] = \frac{1}{2}(v' - 0.3)(v - 0.3)$ and $E^{*1}(1, v, w) = \frac{1}{2}(0.5 - 0.3)(v - 0.3)$. No variable in Bel_2 is certain or vacuous. Thus, $\mathbf{C}^{*2} = \{(u, v, w)\}$, $\mathbf{B}^{*2} = \{(1, 0.3, 2)\}$, and $\mathbf{L}^{*2} = \{(1, 0.3, 2)\}$. Let

$$\Omega = \begin{pmatrix} 5.923 & 2.728 & 0.130 \\ 2.728 & 1.693 & 0.07 \\ 0.130 & 0.070 & 0.040 \end{pmatrix}^{-1} = \begin{pmatrix} 0.658 & -1.048 & -0.305 \\ -1.048 & 2.306 & -0.629 \\ -0.305 & -0.629 & 27.09 \end{pmatrix}.$$

Then, for any two points (u, v, w) and (u', v', w') in $\mathbf{C}^{*2}$,

$$\begin{aligned} \pi^{*2}[(u', v', w), (u, v, w)] &= (u' - 1, v' - 0.3, w' - 2) \\ &\quad \times \Omega(u - 1, v - 0.3, w - 2)^T, \\ E^{*2}(u, v, w) &= (1.2 - 1, 0.6 - 0.3, 0.2 - 2) \\ &\quad \times \Omega(u - 1, v - 0.3, w - 2)^T. \end{aligned}$$

Therefore, according to Theorem 2, $\mathbf{C}^* = \mathbf{C}^{*1} \cap \mathbf{C}^{*2} = \{(1, v, w)\}$, $\mathbf{B}^* = \mathbf{B}^{*1} \cap \mathbf{B}^{*2} = \{(1, 0.3, 2)\}$, $\mathbf{L}^* = \mathbf{L}^{*1} \cap \mathbf{L}^{*2} = \{(1, 0.3, 2)\}$. For any two points $(1, v, w)$ and $(1, v', w')$ in $\mathbf{C}^*$,

$$\begin{aligned} &\pi^{*1}[(1, v', w'), (1, v, w)] + \pi^{*2}[(1, v', w'), (1, v, w)] \\ &\quad = \tfrac{1}{2}(v' - 0.3)(v - 0.3) + (1 - 1, v' + 0.3, w' - 2) \\ &\quad \times \Omega(1 - 1, v - 0.3, w - 2)^T \\ &\quad = (v' - 0.3, w' - 2)\begin{pmatrix} 2.806 & -0.629 \\ -0.629 & 27.09 \end{pmatrix}\begin{pmatrix} v - 0.3 \\ w - 2 \end{pmatrix} \\ &\quad = (v' - 0.3, w' - 2)\begin{pmatrix} 0.358 & 0.008 \\ 0.008 & 0.037 \end{pmatrix}^{-1}\begin{pmatrix} v - 0.3 \\ w - 2 \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} &E^{*1}(1, v, w) + E^{*2}(1, v, w) \\ &\quad = \tfrac{1}{2}(0.5 - 0.3)(v - 0.3) + (1.2 - 1, 0.6 - 0.3, 0.2 - 2) \\ &\quad \times \Omega(1 - 1, v - 0.3, w - 2)^T \\ &\quad = 1.714(v - 0.3) - 49.011(w - 2) \\ &\quad = (0.507 - 0.3, 0.196 - 2)\begin{pmatrix} 0.358 & 0.008 \\ 0.008 & 0.037 \end{pmatrix}^{-1}\begin{pmatrix} v - 0.3 \\ w - 2 \end{pmatrix}. \end{aligned}$$

Note that part of the above computation is to compare the results with those obtained in Example 3. If two GBFs are represented in a sample space, computing the inverse of a covariance matrix is unnecessary. Therefore, computing $\pi^{*1} + \pi^{*2}$ and $E^{*1} + E^{*2}$ only involves multiplications of matrices or additions of quadratic and linear functions. After this is done, it is also unnecessary to transform the quadratic function form of $\pi^{*1} + \pi^{*2}$ and the linear function form of $E^{*1} + E^{*2}$ into their matrix product forms.

5. CONCLUSION

This paper emphasizes how the Dempster-Shafer theory of finite belief functions is extended to the case of GBFs, which are continuous and noncondensable. We first briefly introduced the basic notions of finite belief functions. We then described GBFs in terms of this basic concepts and gave the reader a geometric picture of GBFs. The combination of GBFs is defined by the standard procedure of intersecting focal elements and multiplying the component basic probabilities, except that a basic probability assignment for a finite belief function is replaced by a density-like function in a GBF. This treatment, we believe, will give the reader who has never exposed to Dempster-Shafer theory a self-contained description of the GBF theory. It will also give a Dempster-Shafer theorist a link between finite belief function and GBFs.

A GBF can be geometrically described as a Gaussian distribution across the members of a partition of a hyperplane into parallel subhyperplanes. It includes as special cases multivariate Gaussian distributions, linear equations, and vacuous belief functions, which are nontrivial statistical models in both the classical and the Bayesian schools of thought. This paper formally represents a GBF by a wide-sense inner product and a linear functional over a variable subspace and by their duals over a hyperplane in a sample space. These abstract representations concisely show the full generality of GBFs. As illustrated by the examples in this paper as well as in Liu (1995a, b), in practical applications, many statistical and knowledge-based models turn out to be special GBFs and can be represented by quadratic and linear functions or their corresponding matrix representations. Therefore, the abstract presentation of the theory of GBFs does not hinder its effective applications and efficient implementation.

Part of the reason for having the dual representations is that marginalization can be naturally described in a variable space and combination in a sample space. As we show, the combination of two GBFs in a variable space cannot be explicitly represented by component inner products and linear functionals. The same is true for the marginalization of a GBF in a sample space. However, in applications we often have a predetermined set of variables of interest. In this case, with additional burden of conversion between covariance matrices and their inverses, both combination and marginalization can be easily done numerically in both a variable space and a sample space.

The focal elements of a GBF in general are the subhyperplanes of a hyperplane. If an appropriate basis is chosen for a variable space, this feature essentially reduces a GBF to a Bayesian belief function for some basic variables. Therefore, the combination of GBFs can be derived from that for general Bayesian belief functions, which is the adaptation of Dempster's rule. We have employed this strategy in defining the combination of GBFs in variable spaces. We could also adapt this strategy to derive the combination for non-Gaussian continuous belief functions such as t and exponent belief functions, if any.

The rule for combining GBFs in a variable space is somewhat complex. However, it acts as a basis for more efficient or more concise representations. In this paper, for example, it implies a coordinate-free representation of combination, according to which the combined GBF is obtained by intersecting the component certainty hyperplanes and summing the component inner products and linear functionals over the intersection. This alternative representation is mathematically elegant but not computationally efficient. In Liu (1995a, b), a third representation of combination is obtained using full or partial sweepings. It essentially reduces the combination of GBFs basic matrix operations, which can be done by a spreadsheet program.

ACKNOWLEDGMENT

I am indebted to Glenn Shafer, under whom it has been my privilege to study the Dempster-Shafer theory of belief functions. This paper benefited from many discussions with him and from the foundational work done in Dempster (1990a, b) and Shafer (1992). The research was supported by a research assistantship from the Harper Fund.

References

Box, G. E. P. and Tiao, G. C., *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1973.

Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist.* 38, 325–339, 1967.

Dempster, A. P., A generalization of Bayesian inference (with discussion), *J. Roy. Statist. Soc. Ser. B*, 30, 205–247, 1968.

Dempster, A. P., *Elements of Continuous Multivariate Analysis*, Addison-Wesley, Reading, MA, 1969.

Dempster, A. P., Construction and local computation aspects of network belief functions, in *Influence Diagrams, Belief Nets, and Decision Analysis*, (R. M. Oliver and J. Q. Smith, Eds.), Wiley, Chichester, 1990a.

Dempster, A. P., Normal belief functions and the Kalman filter, Research Report, Dept. of Statistics, Harvard Univ., Cambridge, MA, 1990b.

Fisher, R. A., *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh, 1959.

Hunter, J. E. and Schmidt, F. L., *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*, Sage, Newbury Park, 1990.

Kong, A., Multivariate belief functions and graphical models. Thesis, Dept. of Statistics, Harvard Univ., Cambridge, MA, 1986.

Lauritzen, S. L. and Spiegelhalter, D. J., Local computation with probabilities on graphical structures and their application to expert systems (with discussion), *J. Roy. Statist. Soc. Ser. B*, 50, 157–224, 1988.

Liu, L., Propagation of Gaussian belief functions, Research Report, School of Business, Univ. of Kansas, Lawrence, 1995a.

Liu, L., Model combination using Gaussian belief functions, Research Report, School of Business, Univ. of Kansas, Lawrence, 1995b.

Maier, D., *The Theory of Relational Databases*, Computer Science Press, Rockville, 1983.

Shafer, G., *A Mathematical Theory of Evidence*, Princeton, Princeton U.P., 1976.

Shafer, G., A note on Dempster's Gaussian belief functions, Working Paper, Univ. of Kansas, 1992.

Shafer, G., Shenoy, P. P., and Mellouli, K., Propagating belief functions in qualitative Markov trees, *Internat. J. Approx. Reason.* 3, 383–411, 1987.

Shenoy, P. P. and Shafer, G., Axioms for probability and belief-function propagation, *Uncertainty in Artificial Intelligence* 4, 169–198, 1990.

Yates, F., An apparent inconsistency arising from tests of significance based on fiducial distribution of unknown parameters, *Proc. Cambridge Philos. Soc.* 35, 579, 1939.

Zabell, S., R. A. Fisher on the history of inverse probability, *Statist. Sci.* 4, 247–263, 1989.