Analysis of Fuzzy Information

Volume I

Mathematics and Logic

Editor
James C. Bezdek, Ph.D.
Professor and Chairman
Department of Computer Science
University of South Carolina
Columbia, South Carolina

CRC Press, Inc.
Boca Raton, Florida

Chapter 4

AN INFERENCE RULE BASED ON SUGENO MEASURE

S. T. Wierzchoń

ABSTRACT

In this paper an inference rule based on the notion of conditional Sugeno measure is proposed. Based on this rule, a counterpart of Bayes' theorem is derived under the assumption that a probability measure can be treated as the limit of a collection of Sugeno measures. This new rule allows one to find posterior probabilities from a given collection of conditional probabilities without referring to prior probabilities. Properties of this rule are discussed.

TABLE OF CONTENTS

I. INTRODUCTION

This paper concerns rules that allow one to derive so-called posterior belief functions from conditional belief functions. To be more precise, assume that X is a set of observations, Y is a set of causes underlying these observations, and both sets are finite. Suppose that we have a collection $\{Bel(\cdot|y_j),\ j = 1, \ldots, m\}$, $m = card(Y)$, of "measures" on 2^X expressing our degrees of belief that a proposition $A \subset X$ is true under the cause y_j. (We omit discussion concerning the notion of belief functions, referring the reader to Shafer's original works.[1-3] The basic notions used in Shafer's theory are explained in Section II.) Suppose, finally, we state that an observation belongs to a subset A and we ask: which cause issued this observation? The general solution to this problem was given by Smets.[4] This method — discussed in Subsection A of Section IV — is based on the assumption that each $Bel(\cdot|y_j)$ is derived from independent empirical data. Contrary to Smets' approach, we make the assumption that all the conditionals are derived from a single belief function, Bel_{XY} defined on $2^{X \times Y}$. More precisely, we assume that the conditionals as well as Bel_{XY} are Sugeno measures.[5] As we are able to fit any monotone (with respect to set inclusion) set function to the conditions of Sugeno measure, we hope that this assumption is not too restrictive.[6] The method — presented in Subsection B of Section IV — provides a very interesting conclusion in the case when conditional Sugeno measures are replaced by probability measures. We obtain a counterpart to Bayes' theorem whereby the posterior probability is computed without referring to prior probability. This is the theme of Subsection C of Section IV. Another feature of our method is that we do not use Dempster's rule of combination criticized by some authors.[7] We use instead the specific properties of Sugeno measure that are displayed in Section III; specifically, we treat a probability measure like the limit of a sequence of Sugeno measures.

II. BELIEF FUNCTIONS

The aim of this section is to explain some notions that will be used in the sequel. Readers interested in more details are referred to Shafer's works,[1-3] as well as his survey in this volume.

Suppose X is a set of results and Y is a set of causes. Suppose next that a cause, y, is transformed into a result, x, by a function, f_t, $t = 1, \ldots, s$, chosen randomly from given set F of functions that map Y onto X. Let p_t denote the probability that f_t is chosen. Suppose finally that we have observed the result, x_o. Which cause produced this result? In general it is not possible to get a precise answer to this question. However, we can try to assess the reliability of statements of the form: "The cause lies in a subset B of Y". To this end define the subsets B^* and B_* of the set F as follows:

$$B^* = \{f_t \in F \mid f_t^{-1}(x_0) \cap B \neq \emptyset\} \tag{1}$$

$$B_* = \{f_t \in F \mid f_t^{-1}(x_0) \subset B,\ f_t^{-1}(x_0) \neq \emptyset\} \tag{2}$$

This way we are able to determine numerical degrees characterizing the reliability of our suppositions, namely,

$$Pl(B) = Pr(B^*)/Pr(Y^*) \tag{3}$$

$$Bel(B) = Pr(B_*)/Pr(Y^*) \tag{4}$$

where Pr denotes the probability measure defined by the densities p_i. In general both set functions are nonadditive. It is readily seen that $Pl(\emptyset) = Bel(\emptyset) = 0$ and $Pl(Y) = Bel(Y) = 1$. Moreover, for each $B \subset Y$ there holds $Bel(B) \leq Pl(B)$ and $Bel(B) = 1 - Pl(B^c)$. (Here B_c denotes the complement of B in Y.)

The set function Bel is said to be a belief function. It measures the extent to which one finds a proposition B credible, while Pl — the so-called plausibility function — measures the extent to which this B is plausible.

To make the theory applicable, Shafer introduced a number of quantities: the basic probability function, $m: 2^Y \rightarrow [0,1]$, and the commonality function, $Q: 2^Y \rightarrow [0,1]$, being the most important.[1] The number m(B), called the m-value for brevity, measures the portion of our total finite belief that is committed exactly to the proposition B. In terms of our example,

$$m(B) = Pr\{f_i \in F \mid f_i^{-1}(x_0) = B\} \tag{5}$$

The number Q(B) measures the total portion of belief that can move freely to every point of B. Note that $m(\emptyset) = 0$, while $Q(\emptyset) = 1$. There hold the following relations:

$$Bel(B) = \sum_{\substack{C \subset Y \\ C \subset B}} m(C) \tag{6}$$

$$Q(B) = \sum_{\substack{C \subset Y \\ B \subset C}} m(C) \tag{7}$$

$$Pl(B) = \sum_{\substack{C \subset B \\ C \neq \emptyset}} (-1)^{|C|+1} Q(C) \tag{8}$$

where $|B|$ denotes the cardinality of B.

If $m(B) = 0$ for each B such that $|B| \geq 2$, then the belief function is said to be the Bayesian belief function (probability). In this case, $Bel(B) = Pl(B)$ for each $B \subset Y$. In terms of our example, this occurs if all f_i's are injective.

Let B be a fixed subset of Y and s be a number between 0 and 1. The belief function characterized by the m-values,

$$m(C) = \begin{cases} s & \text{if } C = B \\ 1 - s & \text{if } C = Y \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

is said to be a simple support function (SSF for brevity) focused on B. If $s = 0$, then we call such a belief function the vacuous belief function. It is used to represent complete ignorance, i.e., lack of evidence. On the other hand, if $s = 1$ we will say that the SSF is fully focused on the subset B (we are sure that the truth lies in B).

The most important tool in Shafer's theory is Dempster's rule of combination. Let Bel_1 and Bel_2 be two belief functions representing unrelated bodies of evidence. The orthogonal sum of these Bel's is defined by means of the corresponding m-values as follows

$$m(B) = \frac{\Sigma\{m_1(C)m_2(D) \mid C \subset Y, D \subset Y, C \cap D = B\}}{\Sigma\{m_1(E)m_2(F) \mid E \subset Y, F \subset Y, E \cap F \neq \emptyset\}} \tag{10}$$

If Bel_2 is a SSF fully focused on a subset B, then we call Bel given by Equation 10 the conditional belief function and we have

$$Bel(A \mid B) = \frac{Bel_1(A \cup B^c) - Bel_1(B^c)}{1 - Bel_1(B^c)} \tag{11}$$

In the case of the Bayesian belief function, Bel_1, Equation 11 reduces to Bayes' theorem.

III. SUGENO MEASURE

In this section we present the main results concerning a set function called Sugeno measure.

Definition:[5] A set function G: $2^X \to [0,1]$ is said to be a Sugeno measure if it satisfies the conditions

$$G(\emptyset) = 0, \; G(X) = 1 \tag{12}$$

$$G(A \cup B) = G(A) + G(B) + \lambda G(A)G(B), \; A, B \subset X, \; A \cap B = \emptyset \tag{13}$$

where the parameter $\lambda > -1$ is chosen such that the condition $G(X) = 1$ holds.

Sugeno measure is determined uniquely by the values $g_i = G(\{x_i\})$, $i = 1,2, \ldots n$; $n = |X|$. Namely,

$$G(A) = \frac{1}{\lambda}\left[\prod_{x_i \in A} (1 + \lambda g_i) - 1\right] \tag{14}$$

where λ is computed from the equation:

$$\prod_{i=1}^{n} (1 + \lambda g_i) = 1 + \lambda \tag{15}$$

The set function $H(A) = 1 - G(A^c)$ is also a Sugeno measure with the parameter:

$$\lambda_H = -\lambda_G/(1 + \lambda_G) \tag{16}$$

There holds the following relation between a Sugeno measure G and an additive set function P: $2^X \to [0,1]$ (see Reference 6):

$$G(A) = [(1 + \lambda)^{P(A)} - 1]/\lambda \tag{17}$$

$$P(A) = \frac{\ln(1 + \lambda G(A))}{\ln(1 + \lambda)} \tag{18}$$

Note that a Sugeno measure determines exactly one measure, P. We will call this P the probability measure generated by Sugeno measure. If λ tends to zero then G tends to P.

It has been shown that if $\lambda > 0$ (respectively, $\lambda < 0$), then a Sugeno measure may be treated as a belief (respectively, plausibility) function.[8] More precisely, if $\lambda > 0$, then G is a belief function with the m-values,

$$m(A) = \lambda^{|A|-1} \prod_{x_i \in A} g_i \tag{19}$$

and the commonality numbers,

$$Q(A) = |\lambda'|^{|A|-1} \prod_{x_i \subset A} g_i' \tag{20}$$

where the parameter λ' is determined by Equation 16 and

$$g_i' = g_i(1 + \lambda)/(1 + \lambda g_i) \tag{21}$$

Hence a Sugeno measure may be represented in the form

$$G(A) = \sum_{B \subset A} m(B) \tag{22}$$

if $\lambda > 0$ and in the form

$$G(A) = \sum_{\substack{B \subset A \\ B \neq \emptyset}} (-1)^{|B|+1} Q(B) \tag{23}$$

in the case of $\lambda < 0$. To distinguish between these two cases, we will write G^+ (respectively, G^-) if $\lambda > 0$ (respectively, $\lambda < 0$).

Taking into account earlier results obtained by the author, we can prove the following:[9]

Theorem 1: A Sugeno measure G is a separable support function that may be uniquely decomposed into exactly $n = |X|$ SSF's, S_i, each of which is focused on the subset

$$A_i = \{x_i\}^c,\ i = 1,2,\ldots,n \tag{24}$$

If G^+ is a Sugeno measure and $\{g_i, i = 1,2, \ldots ,n\}$ is the set of its "densities", then the m-value $m_i(A_i)$ of ith SSF S_i is defined by

$$m_i(A_i) = 1/(1 + \lambda g_i) \tag{25}$$

Conversely, if $\{S_i, i = 1,2, \ldots ,n\}$ is a sequence of SSF's focused on subsets A_i defined in Equation 24 with the m-values $m_i(A_i)$, then the orthogonal sum of S_i's is a Sugeno measure characterized by

$$g_i = \frac{1 - m_i(A_i)}{1 - a} \prod_{\substack{j=1 \\ j \neq i}}^{n} m_j(A_j) \tag{26}$$

$$Q(\{x_i\}) = (1 - m_i(A_i))/(1 - a) \tag{27}$$

$$\lambda = (1 - a)/a \tag{28}$$

$$a = \prod_{i=1}^{n} m_i(A_i) \tag{29}$$

By a separable support function, we mean a belief function that is a simple support function or that is the orthogonal sum of two or more simple support functions.[1]

Theorem 2: Let G be a Sugeno measure treated as the orthogonal sum of $n = |X|$ SSF's with the m-values given by Equation 25 and focused on the subsets A_i defined in Equation 24. Then the probability measure generated by G is defined by

$$p_i = \frac{\ln m_i(A_i)}{\sum_{i=1}^{n} \ln m_i(A_i)} \tag{30}$$

where $p_i = P(\{x_i\})$.

Theorem 3: Let in the sequence $\{S_i, i = 1,2, \ldots, n\}$ of the SSF's constituting a Sugeno measure at least one of them, say S_k, be replaced by the vacuous belief function, i.e., $m_k(X) = 1$. Then the orthogonal sum of these SSF's produces a pseudo-Sugeno measure, i.e., the measure G^- with the parameter $\lambda = -1$. In this case, $g_k = Q(\{x_k\}) = 1$.

To see what set function we obtain if all the m-values tend to unity, we need the following:

Lemma 1: Let $P: 2^X \to [0,1]$ be a probability measure determined by its densities p_i, $i = 1,2, \ldots, n$. Define a sequence $\{G_k\}$ of Sugeno measures by their "densities":

$$g_i^k = \frac{k}{1 + k} p_i \tag{31}$$

Then $P = \lim_{k \to \infty}$.

The proof of this lemma is quite similar to the proof of Theorem 9.2 in Shafer's monograph.[1] To fully characterize a Sugeno measure, we need know the value of the parameter λ. Hence the next

Lemma 2: The solution λ_k to the equation $G_k(X) = 1$ is, for sufficiently large k,

$$\lambda^k = \frac{k + 1}{k^2 b} \tag{32}$$

where

$$b = \sum_{\substack{i,j=1 \\ j>i}}^{n} p_i p_j = \left(1 - \sum_{i=1}^{n} p_i^2\right)/2 \tag{33}$$

Consider now a measure G_k determined by its "densities" (Equation 31). Applying Theorem 1, we can decompose it into n SSF's S_i^k with the m-values

$$m_i^k(A_i) = 1/\left(1 + \frac{p_i}{kb}\right) \tag{34}$$

Of course $\lim_{k \to \infty} m_i^k (A_i) = 1$. Conversely, let the m-values of a collection $\{S_i^k, i = 1,2, \ldots, n\}$ of SSF's be given by Equation 34. Then according to Equations 26 to 29, we find that $\lim_{k \to \infty} g_i^k = p_i$.

This conclusion supports the theory exhibited by Shafer in Chapter 9 of his monograph. By this theory, "a Bayesian belief function corresponds to the specification of infinite contradictory weights of evidence. Such weights of evidence cannot be combined directly, but the limiting process allows to balance each other and to produce degrees of belief".[1] To end this section, let us note that applying Equation 11 or examining Equation 17 yields:

Theorem 4: Let G be a Sugeno measure with the parameter λ. Then the conditional Sugeno measure is defined by

$$G(A \mid B) = G(A \cap B)/G(B) = [(1 + \lambda_c)^{P(A|B)} - 1]/\lambda_c \tag{35}$$

where

$$\lambda_c = \lambda G(B) = (1 + \lambda)^{P(B)} - 1 \tag{36}$$

and P is the probability measure generated by G.

IV. INFERENCES BASED ON BELIEF FUNCTIONS

Let X be a set of results and Y be a set of causes. Due to our convention, both the sets are finite with cardinalities n and m, respectively. Suppose that our knowledge about the problem is summarized by a collection of m belief functions $Bel_X(\cdot|y_j)$ defined on 2^X. The objective is to find the posterior belief function $Bel_Y(\cdot|A)$, $A \subset X$, defined on 2^Y.

A. General Case

Within the framework of Bayesian theory, the problem is solvable under the assumption that we have some prior opinion about causes expressed in terms of a probability measure P_o on 2^Y. Knowing P_o and the conditionals $P_X(\cdot|y_j)$, $j = 1,2, \ldots, m$, we are able to build up the probability measure P defined on $2^{X \times Y}$ that fulfills the two conditions:

(B1) For every $B \subset Y$, $P(X \times B) = P_o(B)$.
(B2) For every $y_j \in Y$ conditioning P on $X \times \{y_j\}$ results in the probability function $P_X(\cdot|y_j)$.

Conditioning this P on subsets $A \times Y$, we obtain the *a posteriori* probability measure $P_Y(\cdot|A)$ defined on 2^Y.

We will be faced with a much more general situation if we have conditional belief functions $Bel_X(\cdot|y_j)$ and we do not have prior knowledge about causes. The solution to this problem was proposed by Smets.[4] In his method, we build a belief function Bel_{XY} defined on $2^{X \times Y}$ that fulfills the two postulates (see also Reference 10):

(S1) For every $B \subset Y$, $B \neq Y$, $Bel_{XY} = 0$.
(S2) For every $y_j \in Y$ conditioning Bel_{XY} on $X \times \{y_j\}$ results in the belief function $Bel_X(\cdot|y_j)$.

Conditioning this Bel_{XY} on subsets of the form $A \times Y$, we obtain the posterior belief function $Bel_Y(\cdot|A)$. (Here the postulate, S1, corresponds to the assumption of the lack of knowledge about causes which is modeled by the vacuous prior function Bel_o defined on 2^Y.)

Applying this procedure we obtain (see Reference 4):

$$Pl_Y(B \mid A) = \left[1 - \prod_{y_j \in B} Bel_X(A^c|y_j)\right]/(1 - a_S)$$

$$= \frac{1}{a_S - 1}\left[\prod_{y_j \in B} (1 - (a_S - 1)\frac{Pl_X(A|y_j)}{1 - a_S}) - 1\right] \tag{37}$$

where

$$a_S = \prod_{j=1}^{m} (1 - Pl_X(A|y_j)) \tag{38}$$

It is interesting to observe that if $a_S \neq 0$, then this Pl_Y is nothing but a Sugeno measure characterized by

$$g_j^- = Pl_X(A|y_j)/|\lambda_S| \tag{39}$$

$$\lambda_S = a_S - 1 \tag{40}$$

In light of the discussion presented in Section III, we find that the resulting plausibility function is the orthogonal sum of SSF's focused on the subsets of the form $B_j = A \times \{y_j\}^c$, with the corresponding m-values (compare Equation 27)

$$m_j(B_j) = 1 - Pl_X(A|y_j) = Bel_X(A^c|y_j) \tag{41}$$

If for some j the value $Pl_X(A|y_j) = 1$, then the *a posteriori* Pl_Y becomes a pseudo-Sugeno measure. The quantity a_S in Equation 38 may be interpreted as the grade of belief that the observation is sent by a source $y_o \notin Y$.[4] The attractiveness of Smets' rule is weakened by the fact that it gives sensible results only when the cardinality of Y is fairly small.[10]

B. The Case of Sugeno Measure

Suppose now that our conditional belief functions are all Sugeno measures. To find the posterior belief function, one can proceed in one of two ways:

(a) Utilize Smets' approach.
(b) Assume that the conditionals are derived from a Sugeno measure G_{XY} defined on $2^{X \times Y}$.

In the first case, the resulting posterior is Sugeno measure $G_S^-(\cdot|A)$ characterized as follows

$$g_S^-(y_j|A) = G^-(A|y_j)/|\lambda_S| \tag{42}$$

$$\lambda_S = \prod_{j=1}^{m} (1 - G^-(A|y_j)) - 1 \tag{43}$$

Recall that $g_S^-(y_j|A)$ stands, by our convention, for $G_S^-(\{y_j\}|A)$. Here the subscript S means that the set function is derived from Smets' rule and the superscript "$-$" means that we work with plausibilities, i.e., Sugeno measures with negative parameter λ. The posterior Sugeno measure generates the probability measure P_S on 2^Y with the densities

$$p_S(y_j|A) = \frac{\ln(1 - G^-(A|y_j))}{\sum_{i=1}^{m} \ln(1 - G^-(A|y_i))} \tag{44}$$

In the second case, we use Theorem 4 that takes here the form:

$$\begin{aligned} G(A|y_j) &= G_{XY}(A \times Y \cap X \times \{y_j\})/G_{XY}(X \times \{y_j\}) \\ &= [(1 + \lambda)^{P(A \times Y \cap X \times \{y_j\})} - 1]/[(1 + \lambda)^{P(X \times \{y_j\})} - 1] \\ &= \frac{1}{\lambda_j}[(1 + \lambda)^{P(A|y_j)} - 1] \end{aligned} \tag{45}$$

Using this formula we can derive:

Theorem 5:[11] Let $\{G(\cdot|y_j), j = 1,2, \ldots, m\}$ be a collection of the conditional Sugeno measures derived from an unknown Sugeno measure G_{XY}. Then the posterior Sugeno measure $G_W(\cdot|A)$, $A \subset X$, is characterized as follows:

$$g_W(y_j|A) = \lambda_j G(A|y_j)/\lambda_W \tag{46}$$

$$\lambda_W = \prod_{j=1}^{m} (1 + \lambda_j G(A|y_j)) - 1 \tag{47}$$

Here λ_j denotes the parameter characterizing the jth conditional Sugeno measure and the subscript W is used to distinguish between Smets' and our approach. Let us note the following in connection with Theorems 4 and 5:

Fact 1: The collection $\{G(\cdot|y_j), j = 1,2, \ldots, m\}$ determines uniquely the Sugeno measure G_{XY} from which these conditionals may be derived. The measure G_{XY} possesses the "densities":

$$g(x_i, y_j) = \lambda_j g(x_i|y_j)/\lambda \tag{48}$$

where λ, the parameter characterizing this measure, is defined as

$$\lambda = \prod_{j=1}^{m} (1 + \lambda_j) - 1 \tag{49}$$

Fact 2: The probability measure generated by the posterior measure $G_W(\cdot|A)$ has the densities (compare with the Equation 44)

$$p_W(y_j|A) = P(A|y_j)r_j / \sum_{i=1}^{m} P(A|y_i)r_i \tag{50}$$

where

$$r_j = \frac{\ln(1 + \lambda_j)}{\ln(1 + \lambda)} \tag{51}$$

and λ is defined in Equation 49. The ratios r_j, $j = 1,2, \ldots, m$ play a role similar to the prior probability mass function in Bayes' theorem. If all λ_j's are sufficiently small, then

$$r_j \to \lambda_j / \sum_{i=1}^{m} \lambda_i \tag{52}$$

$$G_W(\cdot|A) \to P_W(\cdot|A) \tag{53}$$

Fact 3: The rule proposed in Theorem 5 is "symmetric" in the sense that starting from conditionals $G(\cdot|y_j)$, computing posterior measures $G_W(\cdot|A)$ and next computing, due to the rule, the measures $G''(\cdot|y_j)$, we return to the given conditionals $G(\cdot|y_j)$. This is not possible if we apply Smets' rule.

Fact 4: Comparing the values of λ_S and λ_W, we state that $\lambda_W/\lambda_S \leq 1$ which shows that $G_S(\cdot|A)$ is more "vacuous" than $G_W(\cdot|A)$ for each $A \subset X$. Hence, the sensitivity of G_W does not depend so heavily on the cardinality of the set Y.

C. The Case of Bayesian Conditionals

Comparing both rules, the most interesting results will be obtained if we assume that our knowledge is described by Bayesian conditional belief functions, i.e., probabilities. Applying Smet's rule we get a posterior plausibility that is no longer Bayesian. It is characterized as follows:

$$Pl_S(\{y_j\}|A) = P(A|y_j)/(1 - a_S) \tag{54}$$

$$a_S = \prod_{j=1}^{m} (1 - P(A|y_j)),\ A \subset X \tag{55}$$

The immediate application of our rule — as given in Theorem 5 — is not presented here. However, we can utilize the suggestion in Lemma 1 by assuming that the conditional probability $P(\cdot|y_j)$ is the limit of the sequence of Sugeno measures $G_k(\cdot|y_j)$ defined as in Equation 31. Taking into account the considerations of the last subsection and Lemma 2, we obtain:

Theorem 6: Let $\{P(\cdot|y_j),\ j = 1,2, \ldots, m\}$ be a collection of conditional probability measures. Then the posterior measure derived via Theorem 5 is the probability measure $P(\cdot|A)$ defined on 2^Y with the densities:

$$p(y_j|A) = \lim_{k\to\infty} p_W^k (y_j|A) = \lim_{k\to\infty} g_W^k (y_j|A)$$

$$= 1/\sum_{i=1}^{m} \frac{P(A|y_i)b_j}{P(A|y_j)b_i} \tag{56}$$

where p_W^k, g_W^k, and b_j are defined, respectively, by Equations 50, 46, and 33. (Here b_j is the coefficient b computed for the jth conditional.)

From this theorem, it follows that the given collection of conditionals determines (uniquely!) some prior P_o on 2^Y, namely,

$$p_j = 1/\sum_{i=1}^{m} b_j/b_i \tag{57}$$

The resulting posterior is influenced by all the values of $p(x_i|y_j)$, i.e., in its determination all available knowledge is taken into account. This is in contrast to Smets' approach, where $Pl(\{y_j\}|A) = cP(A|y_j)$ and c does not depend on j. Assume now that we have some prior P_o on 2^Y. Due to our theory, we should find the sequence G_{XY}^k of Sugeno measures defined on $2^{X\times Y}$ s.t.

$$\lim_{k\to\infty} G_{XY}^k(X \times \{y_j\}) = p_o(y_j) \tag{58}$$

$$\lim_{k\to\infty} G^k(\cdot|y_j) = P(\cdot|y_j) \tag{59}$$

However, from Lemma 2 we know that $\lim_{k\to\infty} \lambda_k = 0$ and that G^k_{XY} tends to some probability measure P_{XY} on $2^{X\times Y}$. It is clear that in order to satisfy the above two conditions, the measure P_{XY} must possess densities satisfying:

$$p_{XY}(x_i, y_j) = p_0(y_j)p(x_i|y_j) \tag{60}$$

Hence, the resulting posterior agrees with the one obtained from Bayes' theorem.

V. FINAL REMARKS

Let us further explain the result obtained in Subsection C of the last section. To do this, we introduce the following:

Lemma 3: Let P be a probability measure. Define:

$$V = \sum_{i=1}^{n} (p_i - 1/n)^2 \tag{61}$$

Then,

$$V = \frac{n-1}{n} - 2b \tag{62}$$

with b defined as in Equation 33.

The lemma gives us a powerful interpretation of the quantity b — it simply measures how dispersed the probability mass function is. Returning to our problem, it may be stated as follows: we do not know prior probabilities, and our knowledge is only represented by a set of conditionals. To refer to any prior probability, Π, let us observe that to determine the conditional probability $P(\cdot|y_j)$, we take only a part of the total probability mass function p_{XY}, namely, $\{p_{XY}(x_i,y_j), i = 1,2, \ldots, n \text{ and } j \text{ is fixed}\}$. Since,

$$p_{XY}(x_i, y_j) = p(x_i|y_j)\, \pi_j \tag{63}$$

we can define the dispersion measure of this part of P_{XY} as $b_j\pi_j^2$. Now let us find the collection of numbers π_j, $j = 1,2, \ldots, m$ such that

$$\sum_{j=1}^{m} b_j\pi_j^2 = \min \tag{64}$$

$$\sum_{j=1}^{m} \pi_j = 1 \tag{65}$$

Applying the Lagrange multiplier method, we find that the solution to this problem is given by Equation 57.

LIST OF SYMBOLS

Σ	Sigma; used as summation
Π	Sign of the product
$\subset$	Containment
$\in$	Epsilon; used as "membership in"
$\cap$	Set theoretical intersection
$\cup$	Set theoretic union
$\leqslant$	Less than or equal to
$<$	Less than
$\geqslant$	Greater than or equal to
$\neq$	Not equal to
$*$	Asterisk; used as subscript or superscript
$'$	Prime
$\rightarrow$	Approaches
∞	Infinity
$\emptyset$	Empty set
λ	Lambda
π, Π	Pi; small and big
$\| \|$	Absolute value
$\|$	Vertical dash
$\times$	Cartesian product; like in: $X \times Y$
$\cdot$	Point; like in: $Bel_X(\cdot\|y_j)$
( )	Parentheses
[ ]	Square brackets
{ }	Cubic brackets
$>$	Greater than
$\notin$	Crossed epsilon; used as "not belongs to"

REFERENCES

1. **Shafer, G.,** *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J., 1977.
2. **Shafer, G.,** Constructive probability, *Synthese*, 48, 1, 1981.
3. **Shafer, G.,** Two theories of probability, in *PSA '78*, Asquith, P. D. and Hacking, I., Eds., Philosophy of Science Association, East Lansing, Mich., 1981.
4. **Smets, Ph.,** Medical diagnosis: fuzzy sets and degrees of belief, *Fuzzy Sets Syst.*, 5, 259, 1981.
5. **Sugeno, M.,** Fuzzy measures and fuzzy integrals: a survey, in *Fuzzy Automata and Decision Processes*, Gupta, M. M., Saridis, G. N., and Gaines, B. R., Eds., North-Holland, Amsterdam, 1977, 89.
6. **Wierzchoń, S. T.,** An algorithm for identification of fuzzy measure, *Fuzzy Sets Syst.*, 9, 69, 1983.
7. **Walley, P.,** Coherent Lower and Upper Probabilities, Statistics Res. Rep., University of Warwick, Coventry, 1981.
8. **Dubois, D. and Prade, H.,** *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1977.
9. **Wierzchoń, S. T.,** On fuzzy measure and fuzzy integral, in *Fuzzy Information and Decision Processes*, Gupta, M. M. and Sanchez, E., Eds., North-Holland, Amsterdam, 1982, 79.
10. **Shafer, G.,** Belief functions and parametric models, *J. R. Stat. Soc.*, Ser. B, 44, 322, 1982.
11. **Wierzchoń, S. T.,** The use of Dempster's upper and lower probabilities in decision making, paper presented at the Polish Symposium on Interval and Fuzzy Mathematics, Technical University of Poznań, August 1983.