論文 / 著書情報
Article / Book Information

題目(和文)	認識・決定過程における曖昧さ主観の解析的表現法に関する研究
Title(English)	Analytical Expression of Ambiguity and Subjectivity in Cognitive and Decision Processes
著者(和文)	廣田薫
Author(English)	KAORU HIROTA
出典(和文)	学位:工学博士, 学位授与機関:東京工業大学, 報告番号:甲第1084号, 授与年月日:1979年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:甲第1084号, Conferred date:1979/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

ANALYTICAL EXPRESSION OF AMBIGUITY AND SUBJECTIVITY IN COGNITIVE AND DECISION PROCESSES.

BY

KAORU HIROTA

JANUARY 1979

A DISSERTATION
PRESENTED TO
THE FACULTY OF THE GRADUATE SCHOOL
TOKYO INSTITUTE OF TECHNOLOGY

IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
DOCTOR OF ENGINEERING
(ELECTRICAL ENGINEERING)

ABSTRACT

Ambiguity and subjectivity have been unwieldy problems especially in human-oriented fields such as decision making and pattern recognition in a broad sense. The purpose of the present thesis is to investigate these problems from all approaches and to open a new path in the research of human-oriented problems.

The main theory is summarized as a concept of probabilistic sets, where the fundamental idea is lying in probability theory with an appropriate use of fuzzy theory. Two mutually equivalent expressions of probabilistic sets are proposed; one is called probabilistic expression and another is extended fuzzy expression. By investigating these expressions, the equality between probability and (extended) fuzzy is confirmed theoretically. The concept of Shannon's entropy is also dealt with in terms of probabilistic sets, and it is shown that entropy is an important measure of ambiguity, but that there also exist other kind of ambiguity such as the notion of vagueness. Topological structure of probabilistic sets is mentioned by several different methods. Lastly four applications are dealt with in order to clarify the description such as appraisement of recognition-performance of character readers and multiple-similarity method of OCR ASPET/71 in terms of probabilistic sets.

TABLE OF CONTENTS

CHAPTER. 1

INTRODUCTION.

Let us begin the story by discussing something rather general. Rapid and great development followed the invention of electric digital-computer in the field of information science. However, not all studies in this field are necessarily complete. On the contrary, the experiences of the old techniques, i.e. high-speed processing by using computers based on a definite algorithm, are at a standstill in a lot of human-oriented problems such as decision making and pattern recognition. Although ambiguity of objects and subjectivity of observers are closely connected with such problems, they have been neglected by most researchers since such a type of research is inefficient, laborious, tedious, expensive and probably unscuccessful. On the other hand, several researchers have been insisting that we should investigate these problems sincerely in order to realize human-oriented systems.

Realizing these points early, the present author commenced a study of this subject and recently successful results, called a theory of "probabilistic sets", have been obtained. It is the purpose of the present thesis to introduce mainly theoretical studies of them with a few applications.

As we stated above, there exist a few real problems remaining unsolved, i.e. ambiguity of objects and subjectivity of observers, in human-oriented fields such as decision-making, pattern recognition and artificial intelligence. Lately these problems have become of rather general interests, and they have been studied by many researchers from various viewpoints such as probability theory, statistics, entropy-theory, subjective probability, fuzzy theory, many-valued logic, modal-logic and quantum logic. Among these various, general studies, it should be noted that extremely many studies have been done since L.A. Zadeh proposed a concept of fuzzy sets (1965). However, there are comparatively few carefully thought-out investigations by paying attention to the inherent and special characteristics of decision making or pattern recognition. Realizing this point, the present author commenced a study of this subject and finally the study has been summarized as a theory of probabilistic sets. The stand-point is epitomized as follows: All the things we can interfere are expressed by concepts of probability, and the cases we can not intervene are unified by using fuzzy concepts. By comparing this idea with other general theories, we may conclude theoretically that the concept of probabilistic sets includes abovestated various concepts and that it will lead to a new way in human-oriented problems. To clarify the description, several fundamental applications are also given such as appraisement of recognition-performance of character readers and multiple-similarity method of OCR ASPET/71.

Contents of each chapter are summarized as follows:

Chapter 1 Introduction.

Chapter 2 Engineering approach to ambiguity and subjectivity.

The central problems of this thesis are proposed with an example of decision making model. The problems are 1) ambiguity of objects, 2) variety of ambiguity, 3) subjectivity of observers and 4) evolution of knowledge of observers. Fundamental ideas of probabilistic sets are introduced by paying attention to these unwieldy problems. However, there exist several mathematically imperfect descriptions because the purpose of this chapter is to make clear the motive of the present thesis. Mathematically detailed discussion is found later.

Chapter 3 Concepts of probabilistic sets.

Details of probabilistic set theory are dealt with from a mathematical point of view. A probabilistic set on a total space is defined by a pointwise measurable [0,1]-valued function which is called a defining function of the probabilistic set. Several operations are introduced to a family of probabilistic sets. Ordinal analysis shows that the family of probabilistic sets constitutes a complete pseudo-Boolean algebra and that it includes the notion of ordinary sets and classical fuzzy sets. Other useful notions are also mentioned such as moment analysis, expected cardinal numbers and probabilistic mappings.

Chapter 4 Extended fuzzy expression of probabilistic sets.

Chapter 3 dealt with probabilistic sets from a viewpoint of "probabilistic expression". It is also possible to introduce another mutually equivalent expression of probabilistic sets called "extended fuzzy expression". The extended fuzzy expression of a probabilistic set is given by a countably infinite set of functions called monitors. Moment analysis shows that the main information is concentrated on lower monitors such as membership function and vagueness function. We can also draw an interesting conclusion to the "fuzzy vs probability" controversy: The equality between the notion of (extended) fuzzy and that of probability is confirmed theoretically.

Chapter 5 Ambiguity based on subjective entropy.

There exists another important approach to the problem of ambiguity, i.e. the concept of Shannon's entropy. Probabilistic sets are investigated from a viewpoint of Shannon's entropy, and a notion of subjective entropy is introduced to probabilistic sets. The mutual relationships are also made clear among probability, fuzzy and entropy.

Chapter 6 Theory of subjective topology.

The purpose of the present chapter is to study fundamentals of topological structure of probabilistic sets. Subjective topology of probabilistic sets is introduced by five different methods. These five definitions of subjective topology are shown to be mutually equivalent by giving a concrete way to induce one concept from another.

Chapter 7 Applications.

Four applications are given to clarify the theoretical explanations done so far; 1) appraisement of recognition-performance of character readers, 2) multiple similarity method of OCR ASPET/71 in terms of probabilistic sets, 3) estimation of Gaussian noise-patterns, 4) Detection of directionality of picture-patterns.

Chapter 8 Conclusions.

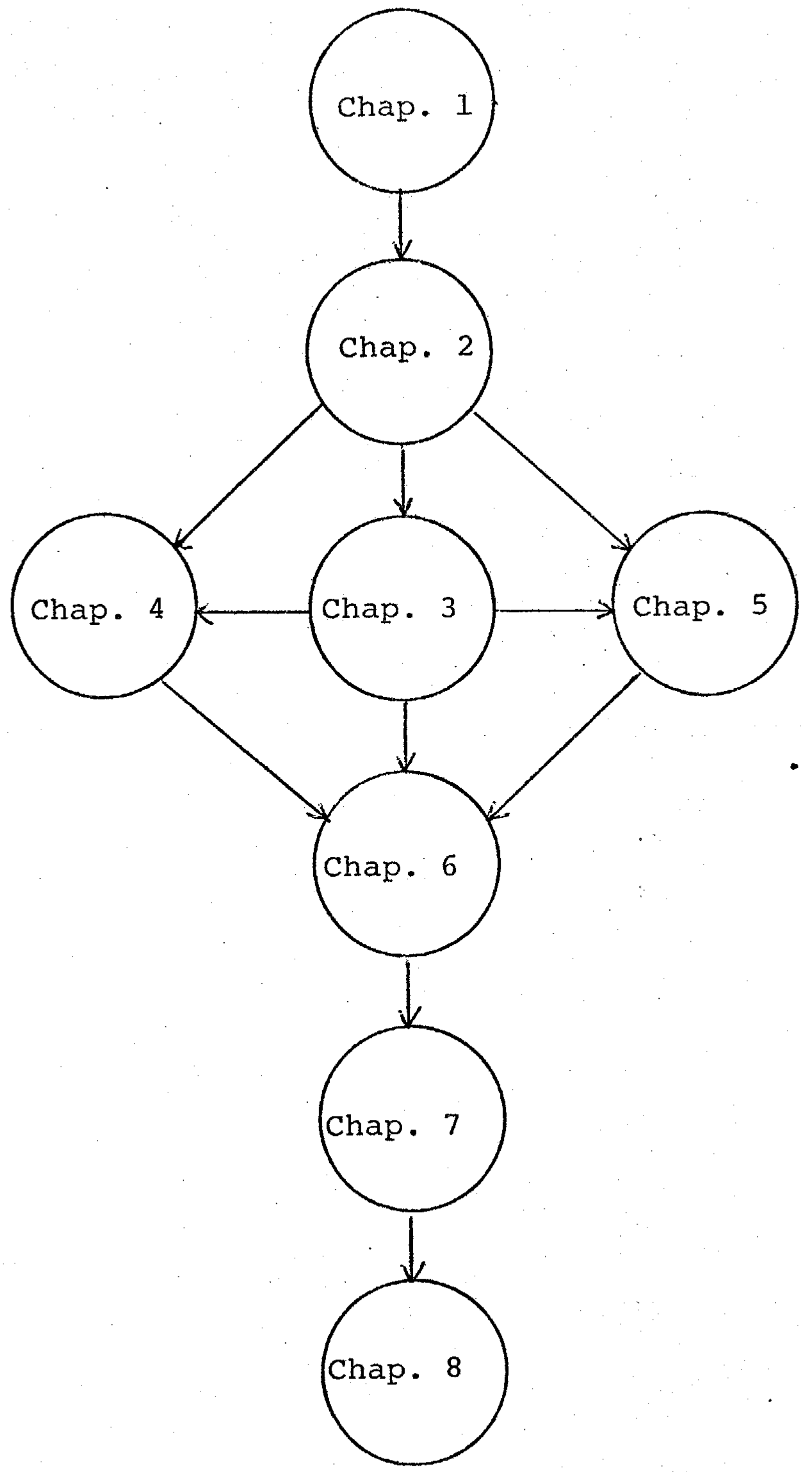


Fig. 1-1. A relation-diagram of chapters in the present thesis.

CHAPTER. 2

ENGINEERING APPROACH TO AMBIGUITY AND SUBJECTIVITY.

2-1. INTRODUCTION.

In decision making theory and in pattern recognition, there are a few real problems remaining unsolved, i.e. ambiguity of objects and subjectivity of observers. Lately these problems have become of general interests, and they have been studied by many researchers among whom the fuzzy concept by L.A.Zadeh [1] is especially excellent. However, there are few carefully thought-out investigations by paying attention to the inherent and special characteristics of decision making and pattern recognition. Realizing this point early, we have commenced a study of this subject and recently successful results have been obtained. In this chapter, we shall deal with ambiguity and subjectivity in decision making, and shall propose a method of an engineering approach to them.

The central problem of this study is proposed first. A decision making model is mentioned and an analytical expression is given on the problems of ambiguity and subjectivity in the decision making model. The standpoint is summarized as follows: All the things we can interfere are expressed by concepts of probability, and the cases we can not intervene are unified by using so-called fuzzy concepts. In order to make clear the standpoint, several comments are mentioned both from a probabilistic viewpoint and a fuzzy viewpoint. It is shown that the (classical) fuzzy concept can be derived from the notion of probability, but the converse is not true, i.e. it is impossible to derive the concept of probability only from a membership function of fuzzy sets. However, if we consider a certain kind of countable family of functions, then it becomes possible to derive the notion of probability. As a result an equality between (extended) fuzzy concepts and probability is theoretically proved. A result is also reported on a questionnaire about a decipherment of hand-written characters. It becomes clear from the result of this experiment that 1) importance of lower moments is confirmed, 2) structures of our subjectivity and ambiguity can be expressed to some extent by our studies.

It should be noted lastly that there might exist several mathematically imperfect descriptions. Since the purpose of this chapter is to make clear the motive of the present paper, the mathematically detailed discussion is found later.

2-2. A DECISION MAKING MODEL.

As stated in the introduction the central problem of this study is to investigate ambiguity and subjectivity in decision making theory. In order to make problems clear, a decision making model discussed here will be explained first.

A set of objects, denoted by

$$X = \{x\}, \tag{2-1}$$

will be called a total space. Each object x is observed by observers and each observer is supposed to make a decision according to his standards of judgement. To clarify the description, an example is given. If we regard all handwritten characters as the tôtal space, then we observers are supposed to decide which category each character belongs to. Hence, the decision making model is shown as in Fig. 2-1.

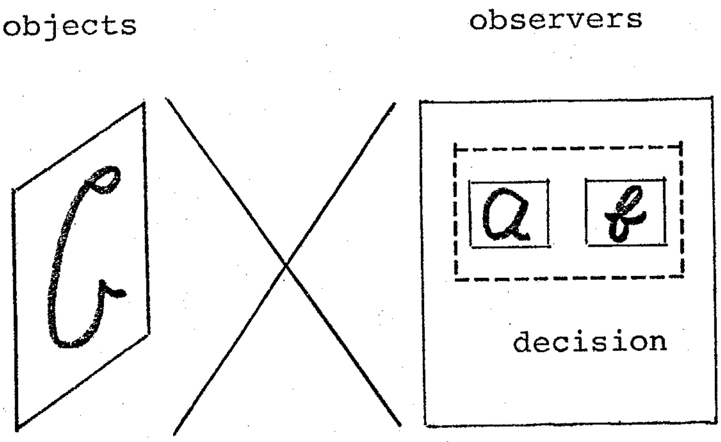


Fig. 2-1. A decision making model.

If each character is written clearly, the decision made by each observer will be a definite one. On the other hand, if it is written ambiguously, the observers may make various, different decisions. Several other problems will also arise along with it.

Let us pick up several points which are problems in decision making.

1) There exists ambiguity of property of objects.

2) The ambiguity is in general so complicated that it seems next to impossible to explain it by using classical two-valued logic alone.

3) There exists subjectivity (or personality) of observers. It also has a complicated structure, and there is a little change of the structure with observers.

4) There exists an evolution of knowledge of observers (i.e. learning and oblivion).

Although all of these points are unwieldy problems and this type of research is inefficient, laborious, tedious and maybe unsuccessful, they are all essential and unavoidable. With regard to each problem, various general studies have been made using probability theory, fuzzy set theory and so on, but more detailed investigation is required from a synthetic or practical viewpoint considering inherent and special characteristics of the decision making model. Realizing these points early, we have suggested a concept of probabilistic sets [2], and a new path has been opened in the study of decision making theory. In the following sections, we wish to introduce main ideas of our studies.

2-3. ANALYTICAL EXPRESSION OF AMBIGUITY AND SUBJECTIVITY IN A SINGLE PHENOMENON.

2-3-1. SINGLE OBJECT AND SINGLE OBSERVER.

Let us consider the simplest example, i.e. a case of single object, single observer and single phenomenon. For instance, consider the following situation in Fig. 2-1: An observer is supposed to decide whether the given character belongs to a category A or not.

In such a case {0,1}-two-valued quantization is normally chosen, i.e. if the given object satisfies (or doesn't satisfy) the phenomenon, then 1 (or 0) will be assigned by the observer (or the decision-maker). In general, however, we can make the following statement. This two-valued quantization is not always correct and a new approach is necessary. Other decisions different from 0 and 1 may be possible because there exists ambiguity. Considering ambiguity and variety of character of objects, we shall adopt [0,1]-infinite-valued quantization which is just the same as the fuzzy concept [1].

This quantization is apparently a good idea, but the problem is so complicated that all the points described in section 2-2 are not completely solved by this [0,1]-idea alone. The evaluation given by the observer might not be determined uniquely in [0,1]-interval. (If the given handwritten character is turned out to be a character A definitely, then a value 1 will be assigned without hesitation. And a value 0 will be assigned exactly in the converse situation. In ambiguous cases, however, these definite answers might be impossible. When we examine a questionnaire, it is sometimes observed that the answer given by the same testee changes from 0.75 to 0.25 after one-month-interval (see section 2-6).) Hence, it might be natural to consider the evaluation (or the decision) to be changed according to the observer's situation rather than to be a definite one. It will be easy to consider this unwieldy fact as follows: Here we would like to define the term "parameter space" $\Omega=\{\omega\}$. Each element ω of Ω corresponds to a standard of judgement of the observer, and it will be selected according to a probabilistic law $p(\omega)$,

$$\int_{\Omega} p(\omega)\cdot d\omega = 1, \quad p(\omega) \geqq 0, \tag{2-2}$$

where $p(\omega)$ follows the observer's subjectivity. Hence the observer's decision is expressed by a mapping $\mu(\omega)$ (called a defining function) from the parameter space $(\Omega, p(\omega))$ to [0,1]-interval,

$$\begin{array}{rcl} \mu(\cdot): \Omega & \longrightarrow & [0,1], \\ \omega & & \omega \\ \omega & \longmapsto & \mu(\omega) \end{array} \tag{2-3}$$

where the value $\mu(\omega)$ expresses the evaluation of the object about the phenomenon from the observer's viewpoint ω.

Since the defining function $\mu(\cdot)$ can also be considered a random variable on a probability space (i.e. the parameter space Ω), several statistics are defined such as meanvalue $E[\mu]$ and variance $V[\mu]$,

$$E[\mu] = \int_{\Omega} \mu(\omega)\cdot p(\omega)\cdot d\omega, \tag{2-4}$$

$$V[\mu] = \int_{\Omega} (\mu(\omega) - E[\mu])^2 \cdot p(\omega)\cdot d\omega. \tag{2-5}$$

The mean value $E[\mu]$ provides the first information of the observer's judgement, whereas the variance $V[\mu]$ offers the second information which indicates how definitely the decision is made by the observer. Although we must know a structure of the parameter space (Ω, $p(\omega)$) in order to calculate $E[\mu]$ and $V[\mu]$ directly, a discussion about the structure of (Ω,$p(\omega)$) is not necessary here, because the values of $E[\mu]$ and $V[\mu]$ can be estimated by several sampling values of $\mu(\omega)$. Since the parameter space (Ω,$p(\omega)$) exists in the observer's subconscious, its structure is usually unknown and it is next to impossible to estimate it. The most important point, however, is that 1) all we can get is sampling values of $\mu(\omega)$, 2) how to use them in a practical analysis.

2-3-2. SINGLE OBJECT AND PLURAL OBSERVERS.

In this section we shall deal with a parameter space Ω in a case of "single object and plural opservers". To clarify the situation, consider the following example in Fig.2-1: Many observers observe a character and all of them are supposed to infer if it is a character A.

In this case the parameter space Ω is expressed by a direct product of Θ and A,

$$\Omega = \Theta \times A, \qquad (2\text{-}6)$$

where $\Theta=\{\theta\}$ stands for a set of standards of judgement and $A=\{a\}$ stands for a set of observers. Each observer a has a set of standards of judjement θ and the evaluation given by the observer a from a viewpoint of θ is expressed by a defining function $\mu(\theta,a)$,

$$\begin{array}{rcl} \mu:\ \Omega=\Theta\times A & \longrightarrow & [0,1], \\ \omega & & \omega \\ (\theta,a) & \longmapsto & \mu(\theta,a) \end{array} \qquad (2\text{-}7)$$

instead of (2-3). A standard of judgement θ is selected by the observer a according to a probability density function (p.d.f.) $p(\theta|a)$,

$$\int_{\Theta} p(\theta|a) \cdot d\theta = 1, \quad p(\theta|a) \geqq 0. \qquad (2\text{-}8)$$

In the set of observers A, the selection rule of an observer a obeys a p.d.f. p(a),

$$\int_{A} p(a) \cdot da = 1, \quad p(a) \geqq 0. \qquad (2\text{-}9)$$

The joint distribution of θ and a is given by

$$p(\theta, a) = p(\theta|a) \cdot p(a) \quad (\geqq 0), \qquad (2\text{-}10)$$

and this function offers a p.d.f. of Ω,

$$\int_{\Omega} p(\theta, a) \cdot d\theta \cdot da = \int_{A} p(a) \cdot (\int_{\Theta} p(\theta|a) \cdot d\theta) \cdot da = 1. \qquad (2\text{-}11)$$

A standard of judgement θ is selected by a set of observers A according to

$$p(\theta) = \int_{A} p(\theta, a) \cdot da = \int_{A} p(\theta|a) \cdot p(a) \cdot da, \qquad (2\text{-}12)$$

where $p(\theta)$ is also a p.d.f. of Θ,

$$\int_{\Theta} p(\theta) \cdot d\theta = 1, \quad p(\theta) \geqq 0. \qquad (2\text{-}13)$$

Based on the Bayes' theorem, the distribution of observers a's who select a fixed standard θ obeys a conditional p.d.f. $p(a|\theta)$, where

$$p(a|\theta) = p(\theta, a) / p(\theta), \qquad (2\text{-}14)$$

$$\int_{A} p(a|\theta) da = 1, \quad p(a|\theta) \geqq 0. \qquad (2\text{-}15)$$

In the abovementioned discussion of probability measures, it must be noted that the known function is only the selection rule p(a), and that all the others are imaginary ones for discussing convenience.

2-3-3. MOMENT ANALYSIS.

In this section, we shall mention moment analysis when the parameter space is the same one given in a previous section 2-3-2.

Mean value $E[\mu(\cdot, a)]$ and variance $V[\mu(\cdot, a)]$ of each observer a are given by

$$E[\mu(\cdot,a)] = \int_\Theta \mu(\theta,a)\cdot p(\theta|a)\cdot d\theta, \tag{2-16}$$

$$V[\mu(\cdot,a)] = \int_\Theta (\mu(\theta,a)-E[\mu(\cdot,a)])^2\cdot p(\theta|a)\cdot d\theta, \tag{2-17}$$

respectively. Total mean value $E[\mu]$ and total variance $V[\mu]$ are also given by

$$E[\mu] = \int_A\int_\Theta \mu(\theta,a)\cdot p(\theta,a)\cdot d\theta da \tag{2-18}$$

$$= \int_A E[\mu(\cdot,a)]p(a)da, \tag{2-19}$$

$$V[\mu] = \int_A\int_\Theta (\mu(\theta,a)-E[\mu])^2 p(\theta,a)d\theta da \tag{2-20}$$

$$= \int_A V[\mu(\cdot,a)]p(a)da + \int_A E[\mu(\cdot,a)]^2 p(a)da - E[\mu]^2, \tag{2-21}$$

respectively. Since we are able to know both a p.d.f. p(a) and sampling values of $\mu(\theta,a)$, we can estimate $E[\mu]$ (2-19) (not (2-18)) and $V[\mu]$ (2-21) (not (2-20)) by using p(a) and estimated values of $E[\mu(\cdot,a)]$ and $V[\mu(\cdot,a)]$ from sampling values.

Of course, these concepts can be generalized. For example, n-th moment $M^n[\mu]$ and n-th moment around mean value $M_O^n[\mu]$ are given by

$$M^n[\mu] = \int_A\int_\Theta \mu(\theta,a)^n\cdot p(\theta,a)\cdot d\theta da, \tag{2-22}$$

$$M_O^n[\mu] = \int_A\int_\Theta (\mu(\theta,a)-E[\mu])^n\cdot p(\theta,a)\cdot d\theta da, \tag{2-23}$$

respectively. However, the more n increases, the less the accuracy and the concrete meaning of their estimations become.

Since the value of $\mu(\theta,a)$ exists in [0,1]-interval, we have

$$1 \geq M^m[\mu] \geq M^n[\mu] \geq 0 \quad (n \geq m). \tag{2-24}$$

Similarly, we have

$$1 \geq M_O^{2m}[\mu] \geq M_O^{2n}[\mu] \geq 0 \quad (n \geq m), \tag{2-25}$$

$$|\, M_O^{2m+1}[\mu] \,| \leq M_O^{2m}[\mu]. \tag{2-26}$$

Moreover, we can conclude that

$$\lim_{n\to\infty} M_O^n[\mu] = 0. \tag{2-27}$$

It can be said from (2-25),(2-26) and (2-27) that 1) the more n increases, the less the absolute value of moments decreases, 2) finally the value reaches the limit 0 as n tends to infinity. Hence it is confirmed theoretically that main information is concentrated on lower moments such as mean value and variance. In fact, from a practical viewpoint, since the ratio of first moment to third is almost less than 10^{-2} (cf. section 2-6), most informations are expressed only by mean value and variance.

We have already given an answer to first three problems mentioned in section 2-2. For the last problem, i.e. "evolution of knowledge", we must consider another factor, i.e. time T, and regard a parameter space as $\Omega=\Theta\times A\times T$. However, we do not discuss this complicated problem in detail here.

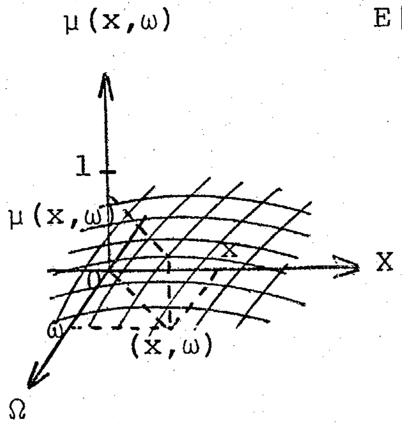
2-3-4. PLURAL OBJECTS.

In abovementioned discussion, we considered a case of single object. We shall deal with a case of plural objects here. Recall that a set of objects was called a total space and was denoted by $X=\{x\}$ (see (2-1)).

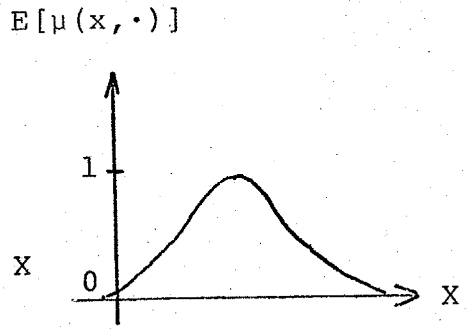
Since an evaluation is given for each object x and each parameter ω in such a case, a defining function will be expressed by

$$\begin{array}{ccc} \mu: X\times\Omega & \longrightarrow & [0,1], \\ \omega & & \omega \\ (x,\omega) & \longmapsto & \mu(x,\omega) \end{array} \qquad (2\text{-}28)$$

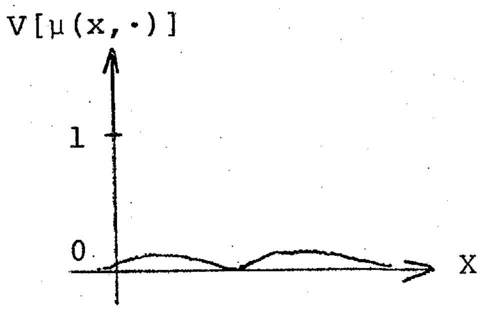
instead of (2-3),(2-7). This defining function is shown as in Fig.2-2(a), and this idea can be developed into a theory of probabilistic sets (cf. chapter 3). For each object x, abovestated moment analysis can be carried out. Mean value and variance are given as a function of object x (Fig.2-2 (b),(c)).



(a) evaluation value



(b) mean value



(c) variance

Fig.2-2. Evaluation of a single phenomenon.

Let's compare this idea with classical two-valued logic. Consider, for example, the following case,

$$X=\{x,y,z\}. \tag{2-29}$$

In such a case, a family of all subsets in two-valued logic is composed of $8(=2^3)$ elements, i.e.

$$\phi,\{x\},\{y\},\{z\},\{x,y\},\{y,z\},\{z,x\},\{x,y,z\}=X. \tag{2-30}$$

Each of these elements can be expressed by a two valued mapping χ from X to {0,1},

$$\chi: X \longrightarrow \{0,1\}. \tag{2-31}$$

(e.g. {x,y} is expressed by

$$\chi(x)=1,\ \chi(y)=1,\ \chi(z)=0. \tag{2-32}$$)

If we compare (2-31) with (2-28), it can be said that the differences between classical two-valued logic and our idea are 1) addition of a parameter space Ω (subjectivity of observers) and 2) from {0,1} to [0,1] (ambiguity and variety of objects). As a result of these differences, the number of cases we must consider increases from $8=2^3$ to infinity.

Hence, in practice, it can be concluded that if we take ambiguity and subjectivity into consideration we must carry out a systematical analysis considering topological, algebraic and ordered structure of the total space X, among which the topological structure is very important. (From this viewpoint a creative study has been made by Iijima in the field of pattern recognition [3].)

2-4. PROBABILISTIC EXPRESSION AND FUZZY CONCEPTS.

In order to make clear the standpoint and significance of our idea, we shall discuss several comments from both probabilistic viewpoints and fuzzy concepts.

L.A.Zadeh introduced a concept of fuzzy sets [1], and he tried to understand an ambiguous state positively. The fuzzy concept is expressed by a [0,1]-valued mapping called a membership function. In our idea, it may be easy to understand that the notion of membership function corresponds to a mean-value-function as in Fig.2-2 (b). It is sometimes pointed out that a concrete meaning of ambiguity is not always expressed completely only by a membership function of fuzzy sets, since the membership function assumes only one definite value in [0,1]-interval. It might be impossible to express a degree of ambiguity unless we cosidered a certain kind of ramdomness such as, for example, "mean value is 0.8 whereas there is a little uncertainty of variance 0.1".

On the other hand, there are many anti-fuzzy scientists who insist that ambiguity can be expressed by only classical two-valued logic with an appropriate use of probability. For instance, they insist that the degree of ambiguity (i.e. the value of a membership function) "0.8", for example, is a result of "true=1 with probability 0.8 and false=0 with probability 0.2". Also in our study, if we devide each standard of judgement (or equivalently each parameter) ω into further basic elements, a discussion might be possible by using only {0,1}-two-values and probability.

The most important point, however, is not a {0,1}-two-valued judgement with probability (such as "value 1 with probability 0.8

and value 0 with probability 0.2") but a [0,1]-ambiguous value itself given by observers (such as 0.8,0.75 and 0.85). The above-mentioned devision of each parameter ω is useless and next to impossible because we must intervene in the observer's subconscious (even if we could do it, only unnecessarily complicated results would be obtained). Hence, a starting point of our investigation was settled on a [0,1]-valued function μ and a probability measure p of the parameter space Ω. This way of thinking makes it possible not only not to bring in an unnecessary complexity but also to develop a systematical approach by using comprehensible quantity alone. It also seems to provide a new methodology which makes it possible to escape from a narrow-minded world of probability alone.

Basic ideas of this paper are, as were already mentioned, a defining function $\mu(x,\omega)$ and a probability measure p on Ω. An ordered pair $(\mu(x,\omega),p(\omega))$ is called a probabilistic expression of ambiguity and subjectivity. It is also possible to make another mutually equivalent expression which will be mentioned in the following.

For convenience, an object x is arbitrarily fixed and x is omitted to be written down for a while. If we consider that an important point is not who is the decision-maker nor which standard of judgement he uses, but that the evaluation-value itself such as 0.7 or 0.8, and if we do not take interest in a structure of the parameter space Ω, then we can identify the parameter space $(\Omega, p(\omega))$ with a probability space on [0,1]-interval $([0,1],f(\alpha))$ by using a defining function μ (cf. Fig. 2-3). The p.d.f. $p(\omega)$ on Ω is transformed into a p.d.f. $f(\alpha)$ on [0,1] as follows,

$$f(\alpha)d\alpha = \int_D p(\omega)d\omega, \quad \text{where } D = \{\omega | \alpha \leqq \mu(\omega) < \alpha + d\alpha\}, \tag{2-33}$$

$$\int_0^1 f(\alpha)d\alpha = \int_\Omega p(\omega)d\omega = 1, \quad f(\alpha) \geqq 0. \tag{2-34}$$

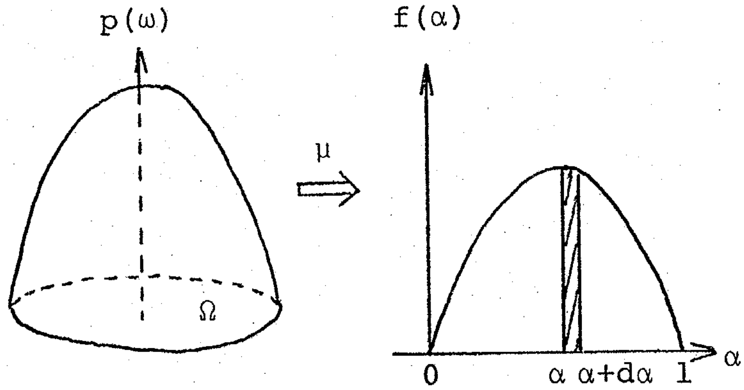


Fig.2-3. [0,1]-expression of the parameter space Ω by using a defining function μ.

From this point of view, it will be evident that all information about ($\mu(\omega)$, $p(\omega)$) is expressed completely by the p.d.f. $f(\alpha)$, and that all moments are also preserved,

$$\int_0^1 \alpha^n \cdot f(\alpha)d\alpha = \int_\Omega \mu(\omega)^n \cdot p(\omega) \cdot d\omega \overline{\overline{\Delta}} M^n[\mu]. \tag{2-35}$$

If we define a transformation from f to ϕ by

$$\phi(t) \overline{\overline{\Delta}} \int_0^1 \exp(it\alpha) \cdot f(\alpha) \cdot d\alpha, \tag{2-36}$$

then we have an inverse transformation,

$$f(\alpha) = (1/2\pi) \cdot \int_{-\infty}^{\infty} \exp(-it\alpha) \cdot \phi(t) \cdot dt. \tag{2-37}$$

Hence, the transformation from f to ϕ becomes an one to one correspondence. On the other hand, a relation

$$\left. \frac{d^n\phi(t)}{dt^n} \right|_{t=0} = i^n \cdot M^n[\mu], \tag{2-38}$$

holds (by using (2-35)). Moreover a Taylor expansion of $\phi(t)$ is always possible,

$$\phi(t) = \sum_{n=0}^{\infty} (i^n/n!) \cdot M^n[\mu] \cdot t^n. \tag{2-39}$$

If we consider the first three terms, we have an approximate expression,

$$\phi(t) = 1 + i \cdot E[\mu] \cdot t - \frac{1}{2} \cdot M^2[\mu] \cdot t^2 + o(t^2)$$

$$= \frac{1}{2}\{ (1+iE[\mu]t)^2 + (1-V[\mu]t^2) \} + o(t^2). \tag{2-40}$$

Since (2-39) holds, all moments give the same information as $\phi(t)$. Hence we have the following equivalence relations

$$\begin{aligned}(\mu(\omega),p(\omega)) &\Longrightarrow f(\alpha) \Longleftrightarrow \phi(t)\\ &\Longleftrightarrow \{E[\mu],M^2[\mu],M^3[\mu],M^4[\mu],\ldots\}\\ &\Longleftrightarrow \{E[\mu],V[\mu],M_0^3[\mu],M_0^4[\mu],\ldots\}.\end{aligned} \tag{2-41}$$

This discussion is valid for each object $x(\varepsilon X)$, so we have

$$\begin{aligned}(\mu(x,\omega),p(\omega)) &\Longrightarrow f(x,\omega) \Longleftrightarrow \phi(x,t)\\ &\Longleftrightarrow \{E[\mu(x,\cdot)],M^2[\mu(x,\cdot)],M^3[\mu(x,\cdot)],\ldots\}\\ &\Longleftrightarrow \{E[\mu(x,\cdot)],V[\mu(x,\cdot)],M_0^3[\mu(x,\cdot)],\ldots\}.\end{aligned} \tag{2-42}$$

Since each moment function is a function of x, we use the following notations,

$$m(x) \underset{\Delta}{=} E[\mu(x,\cdot)], \tag{2-43}$$

$$v(x) \underset{\Delta}{=} V[\mu(x,\cdot)], \tag{2-44}$$

$$m^n(x) \underset{\Delta}{=} M^n[\mu(x,\cdot)]. \tag{2-45}$$

Hence it can be said that the probabilistic expression $(\mu(x,\omega), p(\omega))$ is equivalent to a moment expression $\{m^n(x)\}_{n=1}^{\infty}$.

In the above discussions, we derived a countable set of functions $\{m^n(x)\}_{n=1}^{\infty}$ from a notion of probability. The converse procedure is also possible under several conditions (cf. chapter 4 theorem 4-1). But details of proofs are omitted here. In this converse situation, the function $m(x)$ is called a membership function, $v(x)$ is called a vagueness function, and in general $m^n(x)$ is called a n-th monitor. A set of monitors $\{m^n(x)\}_{n=1}^{\infty}$ is called an extended fuzzy expression, which provides the same information as the probabilistic expression $(\mu(x,\omega),p(\omega))$. In the extended fuzzy expression, it can be shown theoretically that main information is concentrated on lower monitors such as the membership function and the vagueness function. The membership function m(x) indicates the first information about the decision making, and it is the same notion as a membership function of (classical) fuzzy sets in Zadeh's sense [1]. The vagueness function v(x) provides an information of second importance, and the value indicates the disordered degree of the decision. Although the (classical) fuzzy concept, i.e. the notion of membership function alone,is not sufficient compared with the probabilistic expression, it can be said empirically that the vagueness function provides almost all insufficient informations in most applications. Hence, we can conclude theoretically that there are two mutually equivalent approaches to the problem of ambiguity and subjectivity; i.e. one is the probabilisitc expression and another is the extended fuzzy expression. Moreover, by using membership functions, vagueness functions and higher monitors successively, we can expect to obtain useful results which are different from the results given by probabilistic approaches.

2-5. EVALUATION OF PLURAL PHENOMENA.

2-5-1. VARIETY OF JUDGEMENTS - GLOBAL FEATURES AND LOCAL FEATURES -

We have been considering a case of single phenomenon. We shall also deal with a case of plural phenomena in the following. (To clarify the description let us consider the following situation, for instance, in the abovementioned example of Fig.2-1: Observers are supposed to answer whether the presented character is "either A or B".)

But, in general, our judgement turns into an intricated one compared with a $\{0,1\}$-two-valued judgement, as the considered phenomenon becomes complicated; e.g. an incompatible judgement (the presented character seems to be A from one viewpoint, and it also seems to be B from another viewpoint), a failure of transitivity (the character seems to be B rather than to be A, and it seems to be C rather than to be B, but it is more likely to be A than to be C), and so on.

In this paper, we shall clarify this complexity by introducing a notion of <u>parameter space</u> Ω. Here, each parameter ω of Ω corresponds to a standard of judgements. Hence various judgements become possible by using various parameters. There may exist abovestated contradictions among judgements by using various parameters, but the stability is retained by considering a probability density function $p(\omega)$. We shall deal with variety of judgements in the next place.

In probabilistic or statistical approaches, a criterion of least squares is commonly used. Whereas a criterion of "min-max" is regarded as of major importance in Fuzzy operations.

Let us take an example of character recognition. Here we assume that each character is represented as a two-variable function such as f(x,y) or g(x,y). One of the largest problems in character recognition is a selection of a suitable distance-measure. For example, the distance measures

$$d_1(f,g) \underset{\Delta}{=} \iint |f(x,y)-g(x,y)|dxdy, \qquad (2\text{-}46)$$

$$d_2(f,g) \underset{\Delta}{=} \{\iint (f(x,y)-g(x,y))^2 dxdy\}^{1/2}, \qquad (2\text{-}47)$$

are widely used as integral-measures (cf. multiple-similarity method based on d_2-distance by Iijima [4]), and a measure

$$d_\infty(f,g) \underset{\Delta}{=} \max_{x,y} |f(x,y)-g(x,y)|, \qquad (2\text{-}48)$$

is also used as a uniform-distance-measure. However, each of these measures has special characteristics. If there is a little difference $\varepsilon(>0)$ between the value of f(x,y) and that of g(x,y) uniformly, then the distance of f and g by d_∞-distance is very small, i.e. $d_\infty(f,g)=\varepsilon$, but the distance in terms of d_1 or d_2 becomes very large. On the other hand, if the value of f(x,y) and that of g(x, y) are the same except in a few points (in such a few exceptional points, it is assumed that there are great differences between the two), then the distance $d_\infty(f,g)$ becomes very large, but there exist no differences between f and g from a viewpoint of a distance measure of d_1 or d_2.

In general, the integral-criterion corresponds to a majority decision in which overall opinions are accepted, whereas a min-max criterion corresponds to an esteem of excellent minority opinions. A principle of the majority decision is related to a judgement by grobal features, and the esteem of an excellent minority opinion is concerned with an acceptance of various local features. In character recognition, it is necessary for us to attach importance to global features with an appropriate acceptance of excellent local features. This principle agrees with the fundamental idea of our approach based mainly on probability theory with an appropriate use of fuzzy concepts. In the following, we shall develop the study by taking this guiding principle into consideration.

2-5-2. FUNDAMENTAL OPERATIONS OF PLURAL PHENOMENA.

In (classical) {0,1}-two-valued logic, it is possible to make $16=2^4$ different kinds of binary operations, among which "AND" and "OR" operations are the most fundamental and the most important. (Consider a Venn-diagram as shown in Fig.2-4. The total space X is devided into $4=2^2$ regions by two phenomena A and B. Hence, we can obtain $16=2^4$ different figures, each of which corresponds to a binary operation of A and B. For example, the regions of 2, 3 and 4 make an "OR" operation, and the single region 3 corresponds to an "AND" operation.)

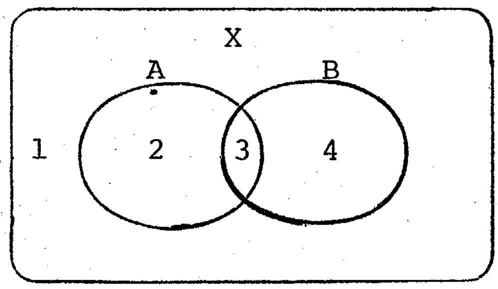


Fig.2-4. A Venn-diagram of two phenomena A and B.

The "AND" and "OR" operations play an important role in switching circuit-theory, and they also characterize the classical two-valued logic as a complete Boolean algebra. Hence, it becomes an important problem to expand these operations into the case of plural phenomena with ambiguity and subjectivity .

Let X be a total space, and Ω be a parameter space. Two phenomena A and B are expressed by mappings

$$\mu_A: X \times \Omega \longrightarrow [0,1], \quad (2\text{-}49)$$

$$\mu_B: X \times \Omega \longrightarrow [0,1], \quad (2\text{-}50)$$

respectively in the same manner as (2-28). A triplet $(\mu_A(x,\omega), \mu_B(x,\omega), p(\omega))$ is called a probabilistic expression of A and B. By considering a fact that each parameter ω of Ω is regarded as a basis of decision making, the "AND" operation $A \cap B$ is defined by

$$\mu_{A\cap B}(x,\omega) = \min\{\mu_A(x,\omega), \mu_B(x,\omega)\}, \quad (2\text{-}51)$$

and the "OR" operation $A \cup B$ is defined by

$$\mu_{A\cup B}(x,\omega) = \max\{\mu_A(x,\omega), \mu_B(x,\omega)\}. \quad (2\text{-}52)$$

In these definitions, the value of each operation is defined locally, i.e. it is defined for each parameter ω, but the result is globally stable (cf. 2-5-3 example 2-1). It must be noted that these definitions are the most suitable from a lattice theoretical viewpoint, i.e. they become fundamental operations which characterize our idea as a complete pseudo-Boolean algebra (cf. chapter 3 theorem 3-1).

2-5-3. EXTENDED FUZZY EXPRESSION OF PLURAL PHENOMENA.

Two phenomena A and B were expressed by a triplet $(\mu_A(x,\omega), \mu_B(x,\omega), p(\omega))$, called a probabilistic expression of A and B. It is the purpose of this section to introduce another mutually equivalent expression called an extended fuzzy expression of A and B. From the same standpoint as section 2-4, the information of $(\mu_A(x,\omega), \mu_B(x,\omega), p(\omega))$ is expressed completely by the following probability density function $f(x;\alpha,\beta)$ on $[0,1]^2$ for each $x \varepsilon X$,

$$f(x;\alpha,\beta)\,d\alpha\, d\beta = \int_{D_x} p(\omega)\,d\omega, \quad (2\text{-}53)$$

$$D_x = \{\omega \mid \alpha \leqq \mu_A(x,\omega) < \alpha+d\alpha,\ \beta \leqq \mu_B(x,\omega) < \beta+d\beta\}, \quad (2\text{-}54)$$

$$\int_0^1\int_0^1 f(x;\alpha,\beta)\,d\alpha d\beta = \int_\Omega p(\omega)\,d\omega = 1, \quad f(x;\alpha,\beta) \geqq 0. \tag{2-55}$$

(It should be noted that x is arbitrarily fixed in the following discussion.) Moment relations are also preserved,

$$\int_0^1\int_0^1 \alpha^n\beta^m f(x;\alpha,\beta)\,d\alpha d\beta = \int_\Omega \mu_A(x,\omega)^n \mu_B(x,\omega)^m p(\omega)\,d\omega$$
$$\overline{\overline{\Delta}}\ M^{n,m}_{A,B}(x). \tag{2-56}$$

The p.d.f. $f(x;\alpha,\beta)$ can be transformed into a function $\phi(x;s,t)$

$$\phi(x;s,t) = \int_0^1\int_0^1 \exp(i(s\alpha+t\beta))\cdot f(x;\alpha,\beta)\,d\alpha d\beta. \tag{2-57}$$

Here, this transformation has an inverse

$$f(x;\alpha,\beta) = (1/2\pi)^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-i(s\alpha+t\beta))\cdot\phi(x;s,t)\cdot dsdt. \tag{2-58}$$

The function $\phi(x;s,t)$ has a partial derivative of arbitrary order with respect to s and t, and a relation

$$\left.\frac{\partial^{n+m}\phi(x;s,t)}{\partial s^n\,\partial t^m}\right|_{s=t=0} = i^{n+m}\cdot M^{n,m}_{A,B}(x), \tag{2-59}$$

holds. Moreover, it can be expanded in a Taylor series,

$$\phi(x;s,t) = \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{r=0}^{n} i^r\binom{n}{r}M^{r,n-r}_{A,B}(x)\cdot s^r t^{n-r}. \tag{2-60}$$

Hence it is concluded that the moment matrix $\{M^{n,m}_{A,B}(x)\}_{n,m=0}^{\infty}$ has the same information as $(\mu_A(x,\omega),\mu_B(x,\omega),p(\omega))$.

Conversely, we can constitute a probability density functior on $[0,1]^2$ from a function matrix $\{m^{n,m}_{A,B}(x)\}_{n,m=0}^{\infty}$ under several conditions (cf. chapter 4 theorem 4-2). Here, the function $m^{n,m}_{A,B}(x)$ is called a (n,m)-th <u>monitor</u> of A and B, and it is regarded as the same as $M^{n,m}_{A,B}(x)$. It is also concluded that the important information is concentrated on lower monitors such as <u>membership functions</u> $m_A(x) \overline{\overline{\Delta}}\ m^{1,0}_{A,B}(x)$, $m_B(x) \overline{\overline{\Delta}}\ m^{0,1}_{A,B}(x)$, <u>vagueness functions</u> $v_A(x) \overline{\overline{\Delta}}\ m^{2,0}_{A,B}(x) - (m^{1,0}_{A,B}(x))^2$, $v_B(x) \overline{\overline{\Delta}}\ m^{0,2}_{A,B}(x) - (m^{0,1}_{A,B}(x))^2$, and a <u>co-vagueness function</u> $v_{A,B}(x) \overline{\overline{\Delta}}\ m^{1,1}_{A,B}(x) - m^{1,0}_{A,B}(x)\cdot m^{0,1}_{A,B}(x)$. An expressior by $\{m^{n,m}_{A,B}(x)\}^{\infty}_{n,m=0}$ is called an <u>extended fuzzy expression</u> of A and B.

Binary operations are also defined in terms of the extended fuzzy expression. We shall introduce only one example of union "A∪B". We can constitute a p.d.f. $f(x;\alpha,\beta)$ from $\{m^{n,m}_{A,B}(x)\}^{\infty}_{n,m=0}$. By using this p.d.f. $f(x;\alpha,\beta)$, a p.d.f. of A∪B, denoted by $f(x;\gamma)$, is obtained as follows,

$$f(x,\gamma) = \int_0^{\gamma} f(x;\gamma,\beta)\,d\beta + \int_0^{\gamma} f(x;\alpha,\gamma)\,d\alpha \,. \tag{2-61}$$

And the n-th monitor of A∪B is given by

$$m^{n}_{A\cup B}(x) = \int_0^{1} \gamma^{n} f(x,\gamma)\,d\gamma \,. \tag{2-62}$$

In applications, it is important to consider only lower monitors such as membership functions, vagueness functions and the co-vagueness function. An example is given to help the explanation done so far.

[Example 2-1]

Let the total space X be a set of real numbers, A be " numbers nearly equal to 1" on X and B be "numbers nearly equal to -1" on X. Here the probabilistic expression of A and B are given , for example, by

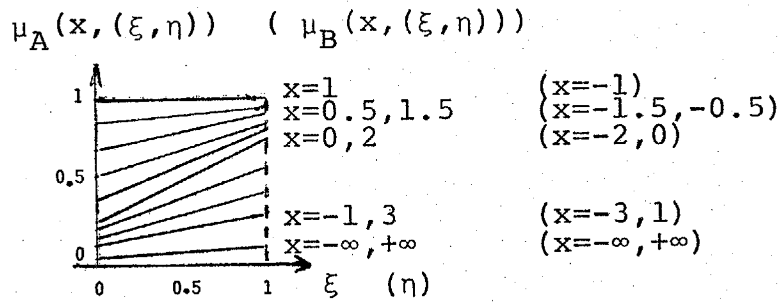
$$\omega = (\xi,\eta) \in \Omega = [0,1]^2, \tag{2-63}$$

$$p(\xi,\eta) = 1 \quad (\text{uniform distribution on } [0,1]^2), \tag{2-64}$$

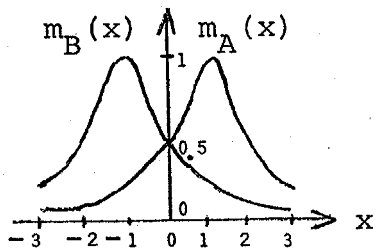
$$\mu_A(x,(\xi,\eta)) = \{\min(1,(x-1)^2)\cdot\xi + \max(1,2-(x-1)^2)/2\} / \{1+(x-1)^2\}, \tag{2-65}$$

$$\mu_B(x,(\xi,\eta)) = \{\min(1,(x+1)^2)\cdot\eta + \max(1,2-(x+1)^2)/2\} / \{1+(x+1)^2\}, \tag{2-66}$$

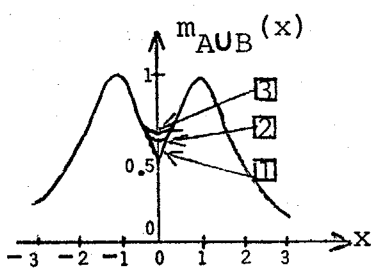
(cf. Fig. 2-5 (a)). In the extended fuzzy expression of A and B, membership functions are given by (cf. Fig. 2-5 (b)),



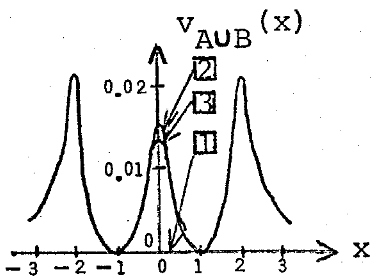
(a) defining functions of A and B. (Probabilistic Expression)



(b) membership functions of A and B. (Extended Fuzzy Expression)



(c) the membership function of $A \cup B$.



(d) the vagueness function of $A \cup B$.

Fig.2-5. A numerical example; A="numbers nearly equal to 1", B="numbers nearly equal to -1", A∪B=" numbers nearly equal to 1 or -1".

$$m_A(x) = 1/\{1+(x-1)^2\}, \quad (2\text{-}67)$$

$$m_B(x) = 1/\{1+(x+1)^2\}, \quad (2\text{-}68)$$

vagueness functions are given by,

$$v_A(x) = [\min\{m_A(x), (1-m_A(x))\}]^2/12, \quad (2\text{-}69)$$

$$v_B(x) = [\min\{m_B(x), (1-m_B(x))\}]^2/12, \quad (2\text{-}70)$$

$$v_{AB}(x) = 0, \quad (2\text{-}71)$$

and so on. The membership function of $A \cup B$, i.e. "numbers nearly equal to 1 or -1", and the vagueness function of $A \cup B$ are given as in Fig.2-5 (c) and (d) respectively; where [1] is a result by using only membership functions (2-67) and (2-68), [2] is a result by using both membership functions (2-67) (2-68) and vagueness functions (2-69) (2-70) (2-71), and [3] is a result by using all monitors (or equivalently by using the probabilistic expression (2-63) (2-64) (2-65) (2-66). To summarize our interpretation of the results, we can explain that

1) The membership function $m_{A \cup B}(x)$ in Fig.2-5 (c) provides the first approximation of probabilistic set $A \cup B$. The result [1] (which is the same result as a classical fuzzy operation) has a continuous but non-smooth (i.e. non-differentiable) point (see x=0). It will be natural to expect a smooth curve like [3] as the first approximation of "numbers nearly equal to 1 or -1". The result [2] provides a good approximation of [3] .

2) From a viewpoint of the vagueness function $v_{A \cup B}(x)$ in Fig.2-5 (d), the result [1] of classical fuzzy operation has no "vagueness" ($v_{A \cup B}(x) = 0$) but it seems unnatural. Since the vagueness function provides the second information and since it indicates a disordered degree of judgements, we would like to

expect a curve like [3] . And [2] also gives a fairly good approximation of [3] .

3) Although, the information given by the membership function alone (see [1] in Fig.2-5 (c) (d)) is not sufficient, almost all insufficient information can be expressed by the vagueness function (compare [2] with [3]).

2-6. QUESTIONNAIRE ON DECIPHERMENT OF HANDWRITTEN CHARACTERS.

As a concrete application of our idea, we put a quentionnaire into practice on decipherment of handwritten characters. We will introduce a part of the result here.

We considered 12 characters shown in Fig.2-6 and made $12^2 = 144$ cards, each of which consisted of two characters as shown in Fig.2-6.

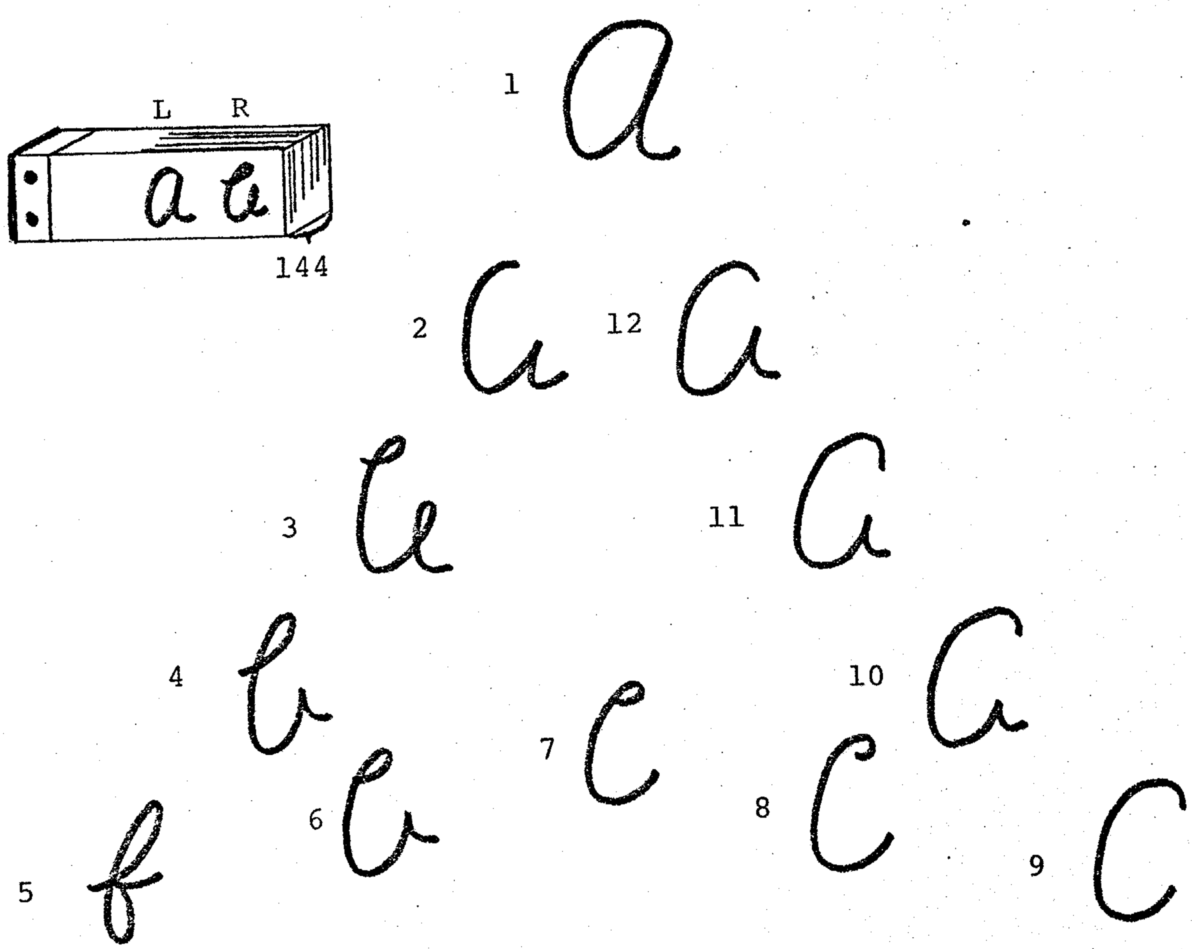


Fig. 2-6. 12 ambiguous characters and 144 cards used in a questionnaire.

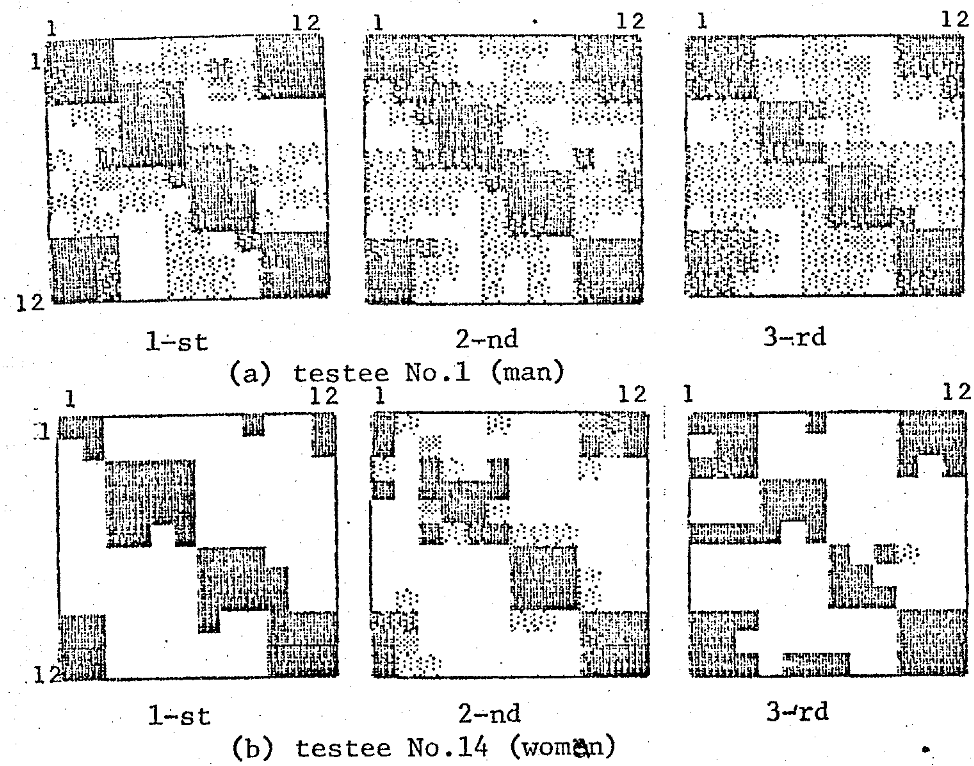


Fig.2-7. A typical result of the questionnaire.

Each card was presented to 20 testees (10 men and 10 women). Each testee was supposed to make a five-level-answer whether the presented two characters are identical ones or not. (There was about one-month-interval between two inquiries.) A typical result is shown in Fig.2-7. The result of moment analysis is also tabulated in Table 2-1 and is shown in Fig.2-8.

What is evident from this result is summarized as follows.

1) It is clearly seen from Table 2-1 (a) and (b) that both reflexivity and symmetricity are satisfied, but that transitivity is not always satisfied (i.e. we can infer nothing about a value $\mu((i,k),\omega)$ from both $\mu((i,j),\omega)$ and $\mu((j,k),\omega)$). Hence, a categorical classification of total space X (i.e. a set of all objects) is impossible based on a concept of mathematical equivalence-relation. It might be natural to identify a class of objects (e.g. a set of character "b") with our idea (i.e. a probabilistic set B) rather than with a category (or an equivalence class) in the classical sense.

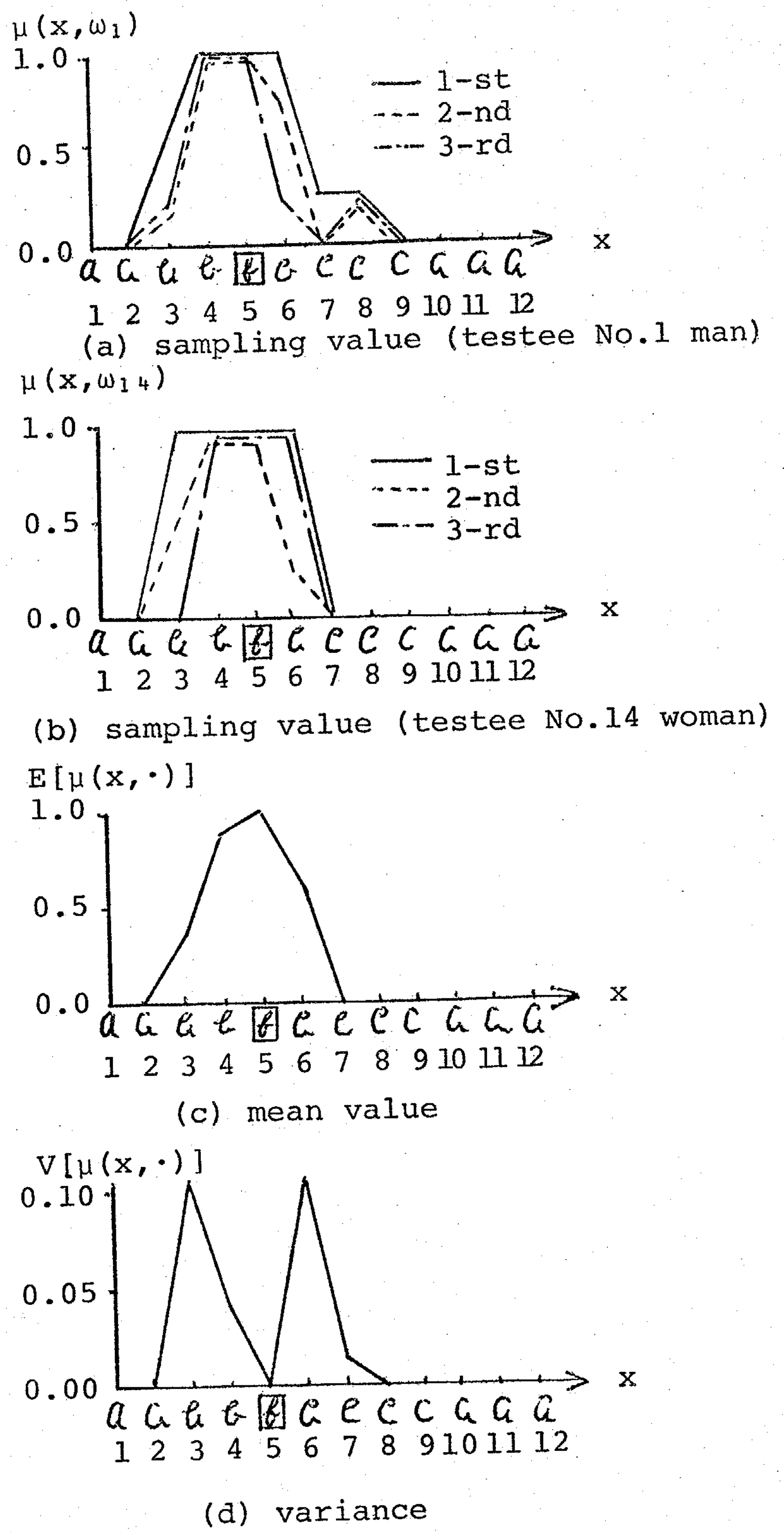


Fig.2-8. Characters which belong to category "b".

2) As is evident from Fig.2-8 (c) and (d), mean value (c) indicates an approximate evaluation-value of ambiguity (which corresponds to a concept of membership function in Zadeh's sense), whereas variance (d) clearly illustrates the degree of uncertainty (which corresponds to a notion of vagueness function of our idea). This example indicates the equality of probabilistic expression and (extended) fuzzy expression, and also assures the validity of our idea.

L \ R	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00	0.65	0.24	0.00	0.00	0.11	0.02	0.02	0.01	0.43	0.82	0.86
2	0.53	0.95	0.42	0.05	0.02	0.12	0.10	0.13	0.14	0.76	0.79	0.87
3	0.19	0.35	0.98	0.54	0.27	0.71	0.15	0.10	0.04	0.18	0.24	0.29
4	0.01	0.05	0.60	0.96	0.87	0.84	0.07	0.01	0.02	0.02	0.03	0.01
5	0.00	0.01	0.29	0.87	1.00	0.64	0.02	0.02	0.00	0.00	0.00	0.00
6	0.09	0.20	0.77	0.75	0.63	0.93	0.25	0.12	0.04	0.10	0.10	0.10
7	0.02	0.11	0.18	0.13	0.04	0.35	0.98	0.83	0.70	0.30	0.12	0.12
8	0.02	0.16	0.10	0.05	0.00	0.14	0.90	0.99	0.82	0.36	0.06	0.08
9	0.01	0.19	0.03	0.00	0.00	0.05	0.75	0.88	0.98	0.48	0.06	0.06
10	0.37	0.78	0.27	0.05	0.00	0.09	0.22	0.21	0.38	0.96	0.91	0.84
11	0.82	0.83	0.25	0.02	0.01	0.11	0.08	0.05	0.05	0.85	1.00	0.97
12	0.85	0.85	0.25	0.00	0.05	0.15	0.11	0.06	0.05	0.85	0.96	0.98

(a) total mean value (membership function).

L \ R	1	2	3	4	5	6	7	8	9	10	11	12
1	0.00	0.10	0.09	0.00	0.00	0.05	0.00	0.01	0.01	0.14	0.05	0.04
2	0.13	0.03	0.11	0.01	0.01	0.04	0.03	0.03	0.05	0.07	0.04	0.03
3	0.08	0.09	0.00	0.12	0.10	0.08	0.04	0.03	0.00	0.07	0.07	0.13
4	0.01	0.01	0.12	0.03	0.05	0.05	0.03	0.00	0.01	0.00	0.02	0.00
5	0.00	0.00	0.11	0.04	0.00	0.11	0.01	0.00	0.00	0.00	0.00	0.00
6	0.05	0.07	0.06	0.07	0.12	0.04	0.07	0.03	0.00	0.02	0.04	0.02
7	0.00	0.02	0.04	0.03	0.01	0.09	0.01	0.05	0.09	0.08	0.02	0.03
8	0.00	0.04	0.02	0.02	0.00	0.04	0.03	0.00	0.08	0.12	0.01	0.02
9	0.01	0.05	0.01	0.00	0.00	0.01	0.07	0.03	0.00	0.11	0.01	0.02
10	0.16	0.05	0.11	0.02	0.00	0.02	0.06	0.07	0.13	0.02	0.02	0.04
11	0.05	0.03	0.06	0.00	0.00	0.03	0.01	0.01	0.03	0.04	0.00	0.00
12	0.03	0.04	0.07	0.00	0.04	0.06	0.05	0.02	0.02	0.03	0.02	0.00

(b) total variance (vagueness function).

Table 2-1. Total mean value and variance (membership function and vagueness function).

A basic idea has been described on "ambiguity and subjectivity in a decision making model". The summary of the idea is as follows: All the things we can interfere are expressed by concepts of probability and the cases we can not intervene are unified by using fuzzy concepts. The significance of this idea is described both theoretically and experimentally. Equality between a concept of probability and a notion of fuzzy theory is also assured. However, theoretical details of probabilistic sets, vagueness function, monitor and so on will be mentioned at later chapters (cf. chapter 3 and chapter 4).

CHAPTER. 3

CONCEPTS OF PROBABILISTIC SETS.

3-1. INTRODUCTION

Analytical expression of ambiguity and subjectivity in pattern recognition was discussed in a previous chapter. The study was based on the following way of thinking: All the things we can interfere are expressed by concepts of probability, and the cases we can not intervene are unified by using so-called "fuzzy concepts". This chapter is concerned with an investigation of this approach from a viewpoint of probability theory and lattice theory.

Fundamental concepts of probabilistic sets are introduced first. An ordinal analysis shows that a family of probabilistic sets constitutes a pseudo-Boolean algebra. Some useful notions are also mentioned such as moment analysis, expected cardinal numbers and probabilistic mappings. By making a comparison between probabilistic sets and other general approaches, including various modified fuzzy theories, it is concluded that the notion of probabilistic sets provides a new systematical approach to the problems of ambiguity and subjectivity.

3-2. THEORY OF PROBABILITY MEASURE

To continue this thesis, it is necessary to understand several terminations and several symbols in probability theory [1], [2],[3].

Let Ω be an arbitrary space. An element ω of Ω will be called a (probability) parameter, and the space Ω will be called a parameter space.

[Def. 3-1]

A sub-family B of all subsets of Ω is called a σ-field if the following three conditions hold,

1) $\Omega \varepsilon B$, (3-1)

2) $A \varepsilon B \Longrightarrow A^C \varepsilon B$, (3-2)

3) $A_n \varepsilon B \ (n=1,2,\ldots) \Longrightarrow \bigcup_{n=1}^{\infty} A_n \varepsilon B.$ (3-3)

A pair (Ω,B) is called a measurable space.

[Prop. 3-1]

Let $(\Omega_1,B_1),(\Omega_2,B_2)$ be measurable spaces, and μ be a mapping from Ω_1 to Ω_2. Then $\mu^{-1}(B_2)$ is a σ-field of Ω_1, where

$$\mu^{-1}(B_2)= \{\{\omega|\mu(\omega)\varepsilon A\}|\ A\varepsilon B_2\}. \quad (3\text{-}4)$$

proof

1) $\Omega_1= \mu^{-1}(\Omega_2)\ \varepsilon\ \mu^{-1}(B_2)$. (3-5)

2) $\mu^{-1}(A)\ \varepsilon\ \mu^{-1}(B_2) \Longrightarrow$

$$\mu^{-1}(A)^C= \{x|\mu(x)\varepsilon A\}^C= \{x|\mu(x)\varepsilon A^C\}= \mu^{-1}(A^C)\ \varepsilon\ \mu^{-1}(B_2).$$

(3-6)

3) $\mu^{-1}(A_n)\ \varepsilon\ \mu^{-1}(B_2)\quad (n=1,2,\ldots) \Longrightarrow$

$$\bigcup_{n=1}^{\infty} \mu^{-1}(A_n)= \bigcup_{n=1}^{\infty}\{x|\ \mu(x)\varepsilon\ A_n\}=$$

$$= \{x \mid \mu(x) \in A_n \quad \text{for some } n\}$$

$$= \{x \mid \mu(x) \in \bigcup_{n=1}^{\infty} A_n\} = \mu^{-1}(\bigcup_{n=1}^{\infty} A_n) \in \mu^{-1}(B_2). \tag{3-7}$$

(Q.E.D.)

[Prop. 3-2]

Let $(\Omega_1, B_1), \Omega_2, \mu : \Omega_1 \longrightarrow \Omega_2$ be a measurable space, a parameter space, a mapping, respectively, then B_2, defined by

$$B_2 \underset{\Delta}{=} \{A_2 \mid A_2 \subset \Omega_2, \mu^{-1}(A_2) \in B_1\}, \tag{3-8}$$

is a σ-field of Ω_2.

<u>proof</u>

1) $\mu^{-1}(\Omega_2) = \Omega_1 \in B_1, \quad \therefore \Omega_2 \in B_2.$ (3-9)

2) If $A_2 \in B_2$, i.e. $\mu^{-1}(A_2) \in B_1$, then we have

$$\mu^{-1}(A_2{}^c) = \mu^{-1}(A_2)^c \in B_1, \text{ i.e. } A_2{}^c \in B_2. \tag{3-10}$$

3) If $A_n \in B_2$ $(n=1,2,\ldots)$, then we have

$$\mu^{-1}(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} \mu^{-1}(A_n) \in B_1, \text{ i.e. } \bigcup_{n=1}^{\infty} A_n \in B_2. \tag{3-11}$$

(Q.E.D.)

[Def. 3-2]

A σ-field B_2, defined in Prop 3-2, is called an <u>induced σ-field</u> from B_1 by μ.

[Def. 3-3]

Let $(\Omega_1, B_1), (\Omega_2, B_2)$ be measurable spaces. A mapping μ,

$$\mu : \Omega_1 \longrightarrow \Omega_2, \tag{3-12}$$

is said to be <u>(B_1, B_2)-measurable</u>, if $\mu^{-1}(B_2) \subset B_1$.

[Def. 3-4]

A (B_1, B_2)-measurable mapping μ is also called an (Ω_2, B_2)-valued <u>random variable</u> on (Ω_1, B_1).

Remark: We sometimes consider ([0,1], Borel sets) as (Ω_2, B_2) in this thesis. In such a case, i.e. the case that the set

Ω_2 is a subset of real numbers and the σ-field is a family of Borel sets, a random variable μ is called just a <u>measurable mapping</u>. It is well-known that a necessary and sufficient condition of μ to be a measurable mapping is that

$$\{\omega | \mu(\omega) \leqq a\} \varepsilon B_1 \quad \text{for each } a \varepsilon [0,1]. \tag{3-13}$$

[Def. 3-5]

Let (Ω, B) be a measurable space, and a correspondence P, defined by

$$P: B \longrightarrow [0,1], \tag{3-14}$$

be a set function. A set function P is said to be a <u>probability measure</u>, if the following two axioms hold,

$$1)\ P(\Omega) = 1, \tag{3-15}$$

$$2)\ A = \bigcup_{n=1}^{\infty} A_n \quad (A_n \varepsilon B,\ A_n \cap A_{n'} = \phi \quad n \neq n')$$

$$\Longrightarrow \quad P(A) = \sum_{n=1}^{\infty} P(A_n). \tag{3-16}$$

(i.e. P is completely additive.)

If P is a probability measure on (Ω, B), then a triplet (Ω, B, P) is said to be a <u>probability space</u>.

[Def. 3-6]

Two probability spaces (Ω_1, B_1, P_1) and (Ω_2, B_2, P_2) are said to be <u>isomorphic</u> if there exist measurable subsets $\Omega_1' \varepsilon B_1, \Omega_2' \varepsilon B_2$, and a mapping ψ such that

$$1)\ P_1(\Omega_1') = P_2(\Omega_2') = 1, \tag{3-17}$$

$$2)\ \psi: \Omega_1' \longrightarrow \Omega_2' \quad \text{1 to 1 onto mapping}, \tag{3-18}$$

$$3)\ \psi^{-1}(B_2 \cap \Omega_2') = B_1 \cap \Omega_1', \tag{3-19}$$

$$4)\ P_1(\psi^{-1}(A)) = P_2(A) \quad \text{for each } A \varepsilon B_2 \cap \Omega_2'. \tag{3-20}$$

The maping ψ is said to be an <u>isomorphism</u>.

If 2) and 3) in above four conditions are exchanged by 2') and 3'),

2') $\Psi : \Omega_1' \longrightarrow \Omega_2'$ onto mapping, (3-21)

3') $\psi^{-1}(B_2 \cap \Omega_2') \subset B_1 \cap \Omega_1'$, (3-22)

then (Ω_1, B_1, P_1) is said to be homomorphic to (Ω_2, B_2, P_2) by a homomorphism ψ.

3-3. THE BACKGROUND IDEA OF PROBABILISTIC SETS

Let $X=\{x\}$ be a total space, i.e. a set of all objects we want to discuss. An ordinary (sub)set A of X, as is well-known, can be defined by a characteristic function 1_A,

$$\begin{aligned} 1_A:\ & X \longrightarrow \{0,1\}. \\ & \ \ \omega \qquad\quad\ \omega \\ & x \longmapsto 1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \end{aligned} \tag{3-23}$$

Considering ambiguity, L.A.Zadeh presented a notion of fuzzy sets [4]. A fuzzy set A on X is expressed by a membership function m_A,

$$m_A: X \longrightarrow [0,1]. \tag{3-24}$$

For example, let all real numbers be the total space X. Consider " all numbers nearly equal to one" and "all numbers nearly equal to minus one". In fuzzy set theory, their membership functions are shown as in Fig.3-1.

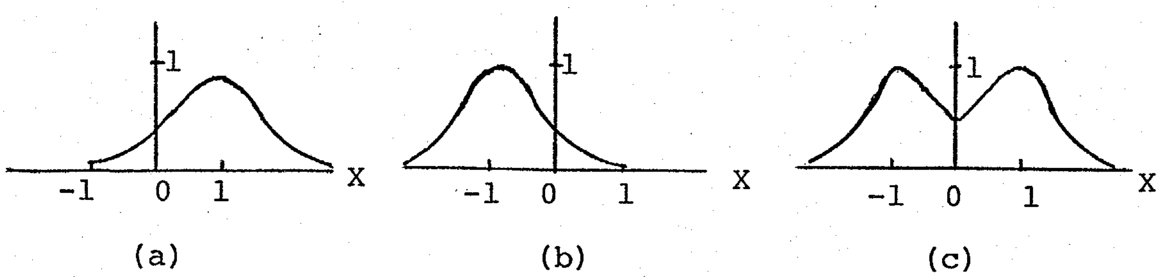


Fig.3-1. Fuzzy sets; (a) numbers near one, (b) numbers near minus one, (c) the union (numbers near one or minus one).

In this situation, however, several questions arise. One of them is that if membership functions like Fig.3-1 can be definitely determined. If the degree of ambiguity were accurately given, it would no longer be ambiguous. Although mean value or variance may be determined and a rough tendency may be given, it is impossible in general to assign definite [0,1]-values. And, to make the matters worse, the tendency varies according to observers' subjectivity, situations and so on. Hence we shall introduce a probability space (Ω,B,P), called a parameter space, whose element represents a standard of judgements. It is assumed that if a standard $\omega(\varepsilon\Omega)$ is fixed, the degree of ambiguity of considered objects (i.e. elements of the total space X) can be definitely determined. A set of all degrees of ambiguity will be called a characteristic space (Ω_c,B_c). We usually adopt ([0,1],Borel sets) as the characteristic space, because it is an infinite totally ordered set with a maximum element 1 and a minimum element 0 and because it is in harmony with characteristic functions of ordinary sets and membership functions of fuzzy sets. A probabilistic set on a total space X is defined by giving a (Ω_c,B_c)-valued random variable on (Ω,B,P) for each object $x(\varepsilon X)$, and this correspondence will be called a defining function of the probabilistic set. The corresponding probabilistic sets of Fig.3-1 are shown in Fig.3-2 (a-1),(b-1),(c-1). The parameter space (Ω,B,P) is expected to be adopted suitably according to each situation, hence in general no restrictions are added to the parameter space except that it's a probability space. For example, in the case of Fig. 3-2, the parameter space might exist in observers' subconscious and might be changed according to the situation. Hence we cannot know its exact probabilistic structure, but we can estimate it by a statistical method. One of the most important facts in proba-

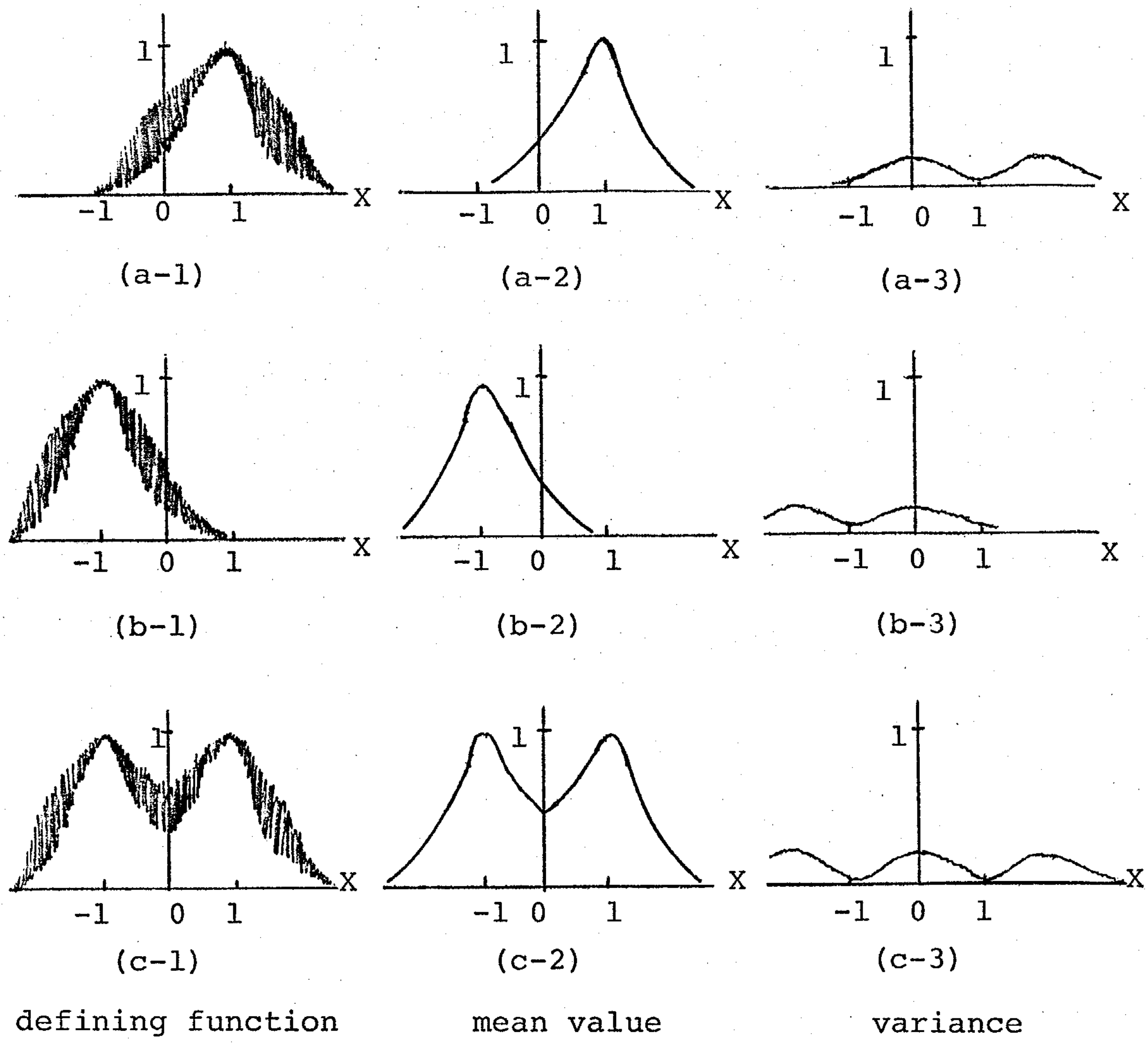


Fig.3-2. Probabilistic sets; (a) numbers near one, (b) numbers near minus one, (c) the union (numbers near one or minus one).

bilistic set theory is a possibility of moment analysis by using a probability measure P of the parameter space. For instance, in Fig.3-2, meanvalues and variances are shown in (a-2),(b-2),(c-2), and (a-3),(b-3),(c-3), respectively. The mean value indicates the first approximation of probabilistic sets and might be considered to be the same one as a membership function of corresponding fuzzy sets. In fuzzy set theory, however, a membership function of "union" generally has a continuous but non-smooth (i.e. non-differentiable) point as shown in Fig.3-1 (c) (at a point of x=0). It will be natural to expect a smooth curve like Fig.3-2 (c-2) as the first approximation of "numbers nearly equal to one or minus one". The variance provides the second information and it indicates a disordered degree of judgements. Higher moments can be considered in the same way. Moreover, it can be shown theoretically that the n-th moment around mean value tends to zero as n tends to infinity. Hence, from a practical viewpoint, it is sufficient to consider only the lower moments, i.e. mean value and variance. If we consider a probabilistic set with variance zero, it could be identified with a fuzzy set. In this sense, it can be concluded that the concepts of probabilistic sets include classical fuzzy concepts.

3-4. DEFINITIONS OF PROBABILISTIC SETS.

The abovestated discussion is a very roughly one from a mathematical viewpoint. The strict definitions are shown in this section. The mathematical foundation of this theory is probability theory.

First, we would like to define the following three terms. (The meanings were discussed in a previous section.)

[Def.3-7]

(Ω, B, P): a parameter space,

$(\Omega_c, B_c)=([0,1],\text{Borel sets})$: a characteristic space,

$M=\{\mu \mid \mu:\Omega \longrightarrow \Omega_c \ (B,B_c)\text{-measurable function}\}$:

: a family of characteristic variables.

It follows that M satisfies the following properties.

[Prop.3-3]

For arbitrary μ_i's ($\mu_i \varepsilon$ M, i=1,2,... at most countably infinite), the following properties are satisfied.

$\min(\mu_1,\mu_2)\,\varepsilon\, M,$ (3-25)

$\max(\mu_1,\mu_2)\,\varepsilon\, M,$ (3-26)

$\mu= c\,\varepsilon\, M$ where $c\varepsilon\Omega_c=[0,1]$ (μ: constant fn.), (3-27)

$|\mu_1-\mu_2|\,\varepsilon\, M,$ (3-28)

$\lambda\mu_1+(1-\lambda)\mu_2\ \varepsilon\ M$ where $0\leqq \lambda\leqq 1,$ (3-29)

$\mu_1^{\alpha}\ \varepsilon\ M$ where $\alpha\geqq 0,$ (3-30)

$\mu_1\mu_2\,\varepsilon\, M,$ (3-31)

$\inf_{i\geqq 1} \mu_i\,\varepsilon\, M,$ (3-32)

$\sup_{i\geqq 1} \mu_i\,\varepsilon\, M,$ (3-33)

$$\underline{\lim}_{i\to\infty} \mu_i = \sup_{i\geq 1} \inf_{j\geq i} \mu_j \in M, \quad (3\text{-}34)$$

$$\overline{\lim}_{i\to\infty} \mu_i = \inf_{i\geq 1} \sup_{j\geq i} \mu_j \in M. \quad (3\text{-}35)$$

<u>proof</u>

The remark described after Def.3-4 will be used in the following statements:i.e. a necessary and sufficient condition of $\mu(\omega)$ being an element of M is that the following relation holds for each $a \in [0,1]$,

$$\{\omega|\mu(\omega)\leqq a\} \in B. \quad (3\text{-}36)$$

1) $\{\omega|\min(\mu_1(\omega),\mu_2(\omega))\leqq a\} = \{\omega|\mu_1(\omega)\leqq a\}\cup\{\omega|\mu_2(\omega)\leqq a\} \in B.$ (3-37)

$$\therefore \min(\mu_1,\mu_2) \in M. \quad (3\text{-}38)$$

2) $\{\omega|\max(\mu_1(\omega),\mu_2(\omega))\leqq a\} = \{\omega|\mu_1(\omega)\leqq a\}\cap\{\omega|\mu_2(\omega)\leqq a\} \in B.$ (3-39)

$$\therefore \max(\mu_1,\mu_2) \in M. \quad (3\text{-}40)$$

3) $\{\omega|\mu(\omega)\leqq a\} = \{\omega|c\leqq a\} = \begin{cases} \phi & 0\leqq a\leq c \\ \Omega & c\leqq a\leqq 1 \end{cases} \in B,$ (3-41)

$$\therefore \mu(\omega) = c \in M. \quad (3\text{-}42)$$

4) $\{\omega||\mu_1(\omega)-\mu_2(\omega)|\leqq a\}=\{\omega| -a\leqq \mu_1(\omega)-\mu_2(\omega)\leqq a\}$

$= \{\omega|\mu_2(\omega)\leqq \mu_1(\omega)+a\}\cap\{\omega|\mu_1(\omega)\leqq\mu_2(\omega)+a\}$

$= \bigcup_{r_n\in Q} \{\omega|\mu_2(\omega)\leqq r_n\leqq \mu_1(\omega)+a\}\cap \bigcup_{r_n\in Q} \{\omega|\mu_1(\omega)\leqq r_n\leqq \mu_2(\omega)+a\}$

$= \bigcup_{r_n\in Q} (\{\omega|\mu_2(\omega)\leqq r_n\}\cap\{\omega|\mu_1(\omega)\geqq r_n-a\}) \cap$

$\cap \bigcup_{r_n\in Q} (\{\omega| \mu_1(\omega)\leqq r_n\}\cap\{\omega|\mu_2(\omega)\geqq r_n-a\}) \in B,$ (3-43)

where Q is a (countable) set of rational numbers which belong to [0,1].

$$\therefore |\mu_1-\mu_2| \in M. \quad (3\text{-}44)$$

5) If $\lambda=0$ or 1, then it is trivial. Hence, we may assume, without

loss of generality, that $0 < \lambda < 1$.

$$\{\omega \mid \lambda\mu_1(\omega)+(1-\lambda)\mu_2(\omega) \leqq a\} = \bigcup_{r_n \in Q} \{\omega \mid \lambda\mu_1(\omega) \leqq r_n \leqq a-(1-\lambda)\mu_2(\omega)\}$$

$$= \bigcup_{r_n \in Q} (\{\omega \mid \mu_1(\omega) \leqq r_n/\lambda\} \cap \{\omega \mid \mu_2(\omega) \leqq (a-r_n)/(1-\lambda)\}) \in B. \qquad (3\text{-}45)$$

$$\therefore \lambda\mu_1+(1-\lambda)\mu_2 \in M. \qquad (3\text{-}46)$$

6) $\{\omega \mid \mu_1(\omega)^{\alpha} \leqq a\} = \{\omega \mid \mu_1(\omega) \leqq a^{1/\alpha}\} \in B.$ (3-47)

$$\therefore \mu_1{}^{\alpha} \in M. \qquad (3\text{-}48)$$

7) $\{\omega \mid \mu_1(\omega)\mu_2(\omega) \leqq a\} = \{\omega \mid ((1/2)\mu_1(\omega)+(1/2)\mu_2(\omega))^2 - ((1/2)\mu_1(\omega)-(1/2)\mu_2(\omega))^2 \leqq a\} \in B.$ (3-49)

Considering (3-29)($\lambda=1/2$),(3-30),(3-28), we have (3-31).

8) $\{\omega \mid \inf \mu_i(\omega) \leqq a\} = \bigcup_{i=1}^{\infty} \{\omega \mid \mu_i(\omega) \leqq a\} \in B.$ (3-50)

$$\therefore \inf_{i \geqq 1} \mu_i \in M. \qquad (3\text{-}51)$$

9) $\{\omega \mid \sup \mu_i(\omega) \leqq a\} = \bigcap_{i=1}^{\infty} \{\omega \mid \mu_i(\omega) \leqq a\} \in B.$ (3-52)

$$\therefore \sup_{i \geqq 1} \mu_i \in M. \qquad (3\text{-}53)$$

10) Applying (3-32) and (3-33), we have (3-34).

11) Applying (3-33) and (3-32), we have (3-35).

(Q.E.D.)

The fundamental definition of probabilistic sets will be given as follows. Here a total space $X=\{x\}$, which represents a set of all objects discussed in each situation, is arbitrarily fixed.

[Def. 3-8]

A <u>probabilistic set</u> A on X is defined by a <u>defining function</u> μ_A,

$$\mu_A : X \times \Omega \longrightarrow \Omega_c, \quad (x,\omega) \longmapsto \mu_A(x,\omega) \qquad (3\text{-}54)$$

(with ω written under $X \times \Omega$ and under Ω_c)

where $\mu_A(x,\cdot)$ is the (B,B_C)-measurable function for each fixed $x(\varepsilon X)$.

For arbitrary two probabilistic sets A and B, whose defing functions are $\mu_A(x,\omega)$ and $\mu_B(x,\omega)$, respectively, A is said to be included in B ($A\subset B$) if for each $x(\varepsilon X)$ there exixts $E(\varepsilon B)$ which satisfies

$$P(E)=1, \tag{3-55}$$

$$\mu_A(x,\omega)\leqq \mu_B(x,\omega) \quad \text{for all } \omega\,\varepsilon\, E. \tag{3-56}$$

In this situation we shall sometimes use a brief notation as follows,

$$\mu_A(x,\omega)\leqq \mu_B(x,\omega) \quad \text{for all } x\varepsilon X \text{ and } P \text{ a.e. } \omega\varepsilon\Omega. \tag{3-57}$$

If both $A\subset B$ and $B\subset A$ are satisfied, A and B are said to be equivalent ($A\equiv B$). (Indeed this relation $\equiv$ satisfies an equivalence relation; i.e. reflexivity, symmetricity, and transitivity.) All equivalent probabilistic sets are considered to be the same one and are not distinguished. All probabilistic sets on X is said to be a family of probabilistic sets and is denoted by $\mathcal{P}(X)$.

Remark: An element of $\mathcal{P}(X)$ represents an equivalence class of M by the equivalence relation $\equiv$ for each $x(\varepsilon X)$.

The inclusion relation in $\mathcal{P}(X)$ satisfies reflexivity, antisymmetricity, and transitivity, hence $(\mathcal{P}(X),\subset)$ constitutes a poset (partially ordered set).

In the following, several operations in $\mathcal{P}(X)$ will be defined. A fundamental operation in $\mathcal{P}(X)$ is "union", however, it is a little complicated. Let A_γ ($\gamma\varepsilon\Gamma,\Gamma$: possibly infinite) be probabilistic sets on X whose defining functions are $\mu_{A_\gamma}(x,\omega)$ respectively. The union of $\{A_\gamma\}_{\gamma\varepsilon\Gamma}$, which is denoted by $\bigcup A_\gamma$, is defined by a defining function $\mu_{\cup A_\gamma}(x,\omega)$ which will be given by the following procedure. For the time being, consider a case that each $x(\varepsilon X)$ is arbitrarily fixed. Then $\mu_{A_\gamma}(x,\cdot)$ may be regarded as a function of $\omega\varepsilon\Omega$ (i.e. an element of M). Since $\mu_{A_\gamma}(x,\cdot)$ is a $\Omega_c=[0,1]$-valued measurable function, and since the total measure is finite (i.e. $P(\Omega)=1$), $\mu_{A_\gamma}(x,\cdot)$ is always P-integrable,

$$0\leqq \int_\Omega \mu_{A_\gamma}(x,\omega)\cdot dP(\omega)\leqq 1. \tag{3-58}$$

For arbitrarily fixed n indeces $\gamma_1,\gamma_2,\ldots,\gamma_n(\varepsilon\Gamma)$, a function max $\{\mu_{A_{\gamma_i}}(x,\cdot)\,|\,1\leqq i\leqq n\}$ is also an element of M (see [Prop.3-3] (3-26)). Hence it is also P-integrable,

$$0\leqq \int_\Omega \max\{\mu_{A_{\gamma_i}}(x,\omega)\,|\,1\leqq i\leqq n\}\cdot dP(\omega)\leqq 1. \tag{3-59}$$

The selection of $\gamma_1,\gamma_2,\ldots,\gamma_n$ from Γ being varied, the least upper bound, denoted by a(x), can be calculated,

$$a(x)= \sup\{\int_\Omega \max\{\mu_{A_{\gamma_i}}(x,\omega)\,|\,1\leqq i\leqq n\}\cdot dP(\omega)\,| \\ |\,n\varepsilon N(\text{natural numbers}),\gamma_i\varepsilon\Gamma\}, \tag{3-60}$$

$$0\leqq a(x)\leqq 1. \tag{3-61}$$

Since a(x) is a "least upper bound", there exists a countably infinite subsequence $\{\max\{\mu_{A_{\gamma_i}}(x,\omega)\,|\,1\leqq i\leqq n_j\}|n_j\varepsilon N,\gamma_i\varepsilon\Gamma\}_{j=1}^{\infty}$ such that

$$\lim_{j\to\infty} \int_\Omega \max\{\mu_{A_{\gamma_i}}(x,\omega)\,|\,1\leqq i\leqq n_j\}\cdot dP(\omega)= a(x). \tag{3-62}$$

Although an element $x(\varepsilon X)$ was arbitrarily fixed in these discussions, this procedure could be done for each $x(\varepsilon X)$. We shall define the defining function $\mu_{\cup A_\gamma}(x,\omega)$ by

$$\mu_{\cup A_\gamma}(x,\omega)= \sup\{\max\{\mu_{A_{\gamma_i}}(x,\omega)\,|\,1\leqq i\leqq n_j\}\,|\,1\leqq j<\infty\}. \tag{3-63}$$

The justification of this definition will be ensured by the following proposition 3-4.

[Prop. 3-4]

1) The union $\cup A_\gamma$ is determined uniquely by the equation (3-63), i.e. if there exists another countably infinite subsequence which satisfies (3-62), the result given by the same equation as (3-63) also belongs to the same equivalence class of M(for each $x\varepsilon X$) in the sense of Def. 3-8.

2) For all $\gamma\varepsilon\Gamma$, we have

$$A_\gamma \subset \cup A_\gamma . \tag{3-64}$$

3) If there exists A which satisfies $A_\gamma \subset A$ for all $\gamma\varepsilon\Gamma$, then we have

$$\cup A_\gamma \subset A. \tag{3-65}$$

<u>proof</u>

2) Firstly we shall give a proof of 2) by using a reductio ad absurdum. If (3-64) does not hold, then we have a $\gamma(\varepsilon\Gamma)$ and a x (εX) such that

$$\int_\Omega \max\{\mu_{\cup A_\gamma}(x,\omega),\mu_{A_\gamma}(x,\omega)\}dP(\omega) > \int_\Omega \mu_{\cup A_\gamma}(x,\omega)dP(\omega). \tag{3-66}$$

We have (by using the Lebesgue's dominated convergence theorem [5]),

the left side of (3-66)=

$$= \int_\Omega \max\{\lim_{m\to\infty} \max_{1\leq j\leq m} \max_{1\leq i\leq n_j} \mu_{A_{\gamma i}}(x,\omega), \mu_{A_\gamma}(x,\omega)\}\cdot dP(\omega)$$

$$= \int_\Omega \lim_{m\to\infty} \max\{\max_{1\leq j\leq m} \max_{1\leq i\leq n_j} \mu_{A_{\gamma i}}(x,\omega), \mu_{A_\gamma}(x,\omega)\}\cdot dP(\omega)$$

$$= \lim_{m\to\infty} \int_\Omega \max\{\max_{1\leq j\leq m} \max_{1\leq i\leq n_j} \mu_{A_{\gamma i}}(x,\omega), \mu_{A_\gamma}(x,\omega)\}\cdot dP(\omega), \tag{3-68}$$

the right side of (3-66)=

$$= \int_\Omega \lim_{m\to\infty} \max_{1\leq j\leq m} \max_{1\leq j\leq n_j} \mu_{A_{\gamma i}}(x,\omega)\cdot dP(\omega)$$

$$= \lim_{m\to\infty} \int_\Omega \max_{1\leq j\leq m} \max_{1\leq i\leq n_j} \mu_{A_{\gamma i}}(x,\omega)\cdot dP(\omega). \tag{3-69}$$

Whereas we have, by (3-60),(3-62),

$$a(x) \geqq \lim_{m\to\infty} \int_\Omega \max_{1\leqq j\leqq m} \max_{1\leqq i\leqq n_j} \mu_{A_{\gamma_i}}(x,\omega) \cdot dP(\omega)$$

$$\geqq \lim_{m\to\infty} \int_\Omega \max_{1\leqq i\leqq n_m} \mu_{A_{\gamma_i}}(x,\omega) \cdot dP(\omega) = a(x). \quad (3\text{-}70)$$

Hence we have, by (3-69),(3-70),

the right side of (3-66) = a(x). (3-71)

By (3-66),(3-68),(3-71), there exists a natural number m such that

$$\int_\Omega \max\{\max_{1\leqq j\leqq m} \max_{1\leqq i\leqq n_j} \mu_{A_{\gamma_i}}(x,\omega), \mu_{A_\gamma}(x,\omega)\} \cdot dP(\omega) > a(x). \quad (3\text{-}72)$$

This contradicts the definition of a(x) (3-60).

3) A proof of 3) will also be given by a reductio ad absurdum. If (3-65) does not hold in spite of $A_\gamma \subset A$ for all $\gamma(\varepsilon\Gamma)$, then we have,

$$\int_\Omega \mu_{\cup A_\gamma}(x,\omega)\, dP(\omega) > \int_\Omega \min\{\mu_{\cup A_\gamma}(x,\omega), \mu_A(x,\omega)\} dP(\omega) \underset{\Delta}{=} c(x). \quad (3\text{-}73)$$

Since $A_\gamma \subset \cup A_\gamma$ (3-64) and $A_\gamma \subset A$ for all $\gamma(\varepsilon\Gamma)$, we have

$$c(x) \geqq \int_\Omega \max_{1\leqq i\leqq n} \mu_{A_{\gamma_i}}(x,\omega)\, dP(\omega), \quad (3\text{-}74)$$

for all n and all $\gamma_i(\varepsilon\Gamma)$. By (3-70), we have

$$\int_\Omega \mu_{\cup A_\gamma}(x,\omega)\, dP(\omega) = a(x). \quad (3\text{-}75)$$

Hence we have

$$a(x) - (a(x) - c(x)) = c(x) \geqq \int_\Omega \max_{1\leqq i\leqq n} \mu_{A_{\gamma_i}}(x,\omega)\, dP(\omega), \quad (3\text{-}76)$$

for all n and all $\gamma_i(\varepsilon\Gamma)$. Since $a(x) - c(x) > 0$, this relation (3-76) contradicts the definition of a(x) (3-60).

1) We denote another subsequence which satisfies (3-62) by

$$\{\max_{1\leqq k\leqq n_l} \{\mu_{A_{\gamma_k'}}(x,\omega)\} \mid n_l \varepsilon N, \gamma_k' \varepsilon \Gamma\}_{l=1}^{\infty}, \quad (3\text{-}77)$$

$$\lim_{l\to\infty} \int_\Omega \max_{1\leqq k\leqq n_l} \{ \mu_{A_{\gamma_k'}}(x,\omega)\} dP(\omega) = a(x). \tag{3-78}$$

If we use the following notation,

$$\mu_{\cup A_\gamma}'(x,\omega) \overline{\underline{\Delta}} \sup_{1\leqq l\leqq\infty} \max_{1\leqq k\leqq n_l} \mu_{A_{\gamma k'}}(x,\omega), \tag{3-79}$$

then all we have to show is

$$\mu_{\cup A_\gamma}(x,\omega) = \mu_{\cup A_\gamma}'(x,\omega) \quad \text{for all } x\varepsilon X \text{ and } P \text{ a.e.} \omega\varepsilon\Omega. \tag{3-80}$$

In the following statements $x(\varepsilon X)$ is arbitrarily fixed. The same relations as 2),3) are also valid for this new subsequence, we have

$$\mu_{A_{\gamma_i}}(x,\omega) \leqq \mu_{\cup A_\gamma}'(x,\omega) \quad \text{for all } \gamma_i\varepsilon\Gamma \quad \text{and } P \text{ a.e. } \omega\varepsilon\Omega,$$

and (3-81)

$$\max_{1\leqq i\leqq n_j} \mu_{A_{\gamma_i}}(x,\omega) \leqq \mu_{\cup A_\gamma}'(x,\omega) \quad \text{for } P \text{ a.e. } \omega\varepsilon\Omega. \tag{3-82}$$

Hence we have

$$\mu_{\cup A_\gamma}(x,\omega) \leqq \mu_{\cup A_\gamma}'(x,\omega). \tag{3-83}$$

The converse relation

$$\mu_{\cup A_\gamma}'(x,\omega) \leqq \mu_{\cup A_\gamma}(x,\omega), \tag{3-84}$$

follows in much the same way.

(Q.E.D.)

Although the abovestated procedure of union is rather complicated, it can be simplified in a case that the index set Γ is at most countably infinite. For example, the union of A and B (whose defining functions are $\mu_A(x,\omega)$ and $\mu_B(x,\omega)$, respectively) may be defined by

$$\mu_{A\cup B}(x,\omega) = \max\{\mu_A(x,\omega), \mu_B(x,\omega)\}, \tag{3-85}$$

for each $x\varepsilon X$ and each $\omega\varepsilon\Omega$, and the union of $\{A_n\}_{n=1}^{\infty}$ may be defined by

$$\mu_{\cup A_n}(x,\omega)= \sup\{ \mu_{A_n}(x,\omega) \mid 1\leqq n<\infty\}, \tag{3-86}$$

for each $x\varepsilon X$ and each $\omega\varepsilon\Omega$. The complexity in a general case arises from the fact that M is not always closed by more than countably infinite operations (cf.[Prop.3-3]).

The "<u>intersection</u>" of $\{A_\gamma\}_{\gamma\varepsilon\Gamma}$, which is denoted by $\cap A_\gamma$, is a dual concept of "union" $\cup A_\gamma$, and it is defined as follows. Put

$$b(x)= \inf\{\textstyle\int_\Omega \min\{\mu_{A_{\gamma i}}(x,\omega) \mid 1\leqq i\leqq n\}\, dP(\omega) \mid n\varepsilon N, \gamma_i\varepsilon\Gamma\}, \tag{3-87}$$

$$0\leqq b(x)\leqq 1, \tag{3-88}$$

and choose a countably infinite subsequence $\{\min\{\mu_{A_{\gamma i}}(x,\omega) \mid 1\leqq i\leqq n_j\} \mid n_j\varepsilon N, \gamma_i\varepsilon\Gamma\}_{j=1}^{\infty}$ such that

$$\lim_{j\to\infty} \textstyle\int_\Omega \min\{ \mu_{A_{\gamma i}}(x,\omega) \mid 1\leqq i\leqq n_j\} dP(\omega)= b(x), \tag{3-89}$$

and define

$$\mu_{\cap A_\gamma}(x,\omega)= \inf\{\min\{ \mu_{A_{\gamma i}}(x,\omega) \mid 1\leqq i\leqq n_j\} \mid 1\leqq j<\infty\} . \tag{3-90}$$

The justification of this definition will also be ensured by the same proposition as Prop.3-4. (Change simbols $\cup,\subset$ to $\cap,\supset$, respectively in Prop.3-4.)

Some other useful concepts or operations in $P(X)$ could be defined. They will be summarized as follows. The justification of these definitions is also ensured by Prop.3-3.

[Def.3-9]

<u>total set</u> X

$$\mu_X(x,\omega)= 1 \quad \text{for all } x\varepsilon X \text{ and a.e. } \omega\varepsilon\Omega, \tag{3-91}$$

(this notation will be omitted till (3-98).)

<u>void set</u> (or <u>null set</u>) ϕ

$$\mu_\phi(x,\omega)= 0, \tag{3-92}$$

<u>complement</u> of A A^c

$$\mu_{A^c}(x,\omega) = 1 - \mu_A(x,\omega), \tag{3-93}$$

difference A-B

$$\mu_{A-B}(x,\omega) = \max\{0, \mu_A(x,\omega) - \mu_B(x,\omega)\}, \tag{3-94}$$

symmetric difference $A \Delta B$

$$\mu_{A\Delta B}(x,\omega) = |\mu_A(x,\omega) - \mu_B(x,\omega)|, \tag{3-95}$$

algebraic sum $A \oplus B$

$$\mu_{A\oplus B}(x,\omega) = \mu_A(x,\omega) + \mu_B(x,\omega) - \mu_A(x,\omega)\mu_B(x,\omega), \tag{3-96}$$

λ sum $A \underset{\lambda}{+} B$ (where $0 \leqq \lambda \leqq 1$)

$$\mu_{A\underset{\lambda}{+}B}(x,\omega) = \lambda\mu_A(x,\omega) + (1-\lambda)\mu_B(x,\omega), \tag{3-97}$$

α power A^α (where $\alpha \geqq 0$)

$$\mu_{A^\alpha}(x,\omega) = \mu_A(x,\omega)^\alpha, \tag{3-98}$$

superior limit of $\{A_n\}_{n=1}^{\infty}$

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \tag{3-99}$$

inferior limit of $\{A_n\}_{n=1}^{\infty}$

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \tag{3-100}$$

An ordered pair $(\mu_A(x,\omega), \mu_B(x,\omega))$ is said to be a direct product of A and B, and is denoted by $A \times B$.

A_y is said to be an one point probabilistic set at $y \in X$, if its defining function $\mu_{A_y}(x,\omega)$ satisfies

$$\int_\Omega \mu_{A_y}(x,\omega) \cdot dP(\omega) \begin{cases} = 0 & x \neq y \\ > 0 & x = y. \end{cases} \tag{3-101}$$

A_y is said to be a full one point probabilistic set at $y \in X$, if its defining function $\mu_{A_y}(x,\omega)$ satisfies

$$\int_\Omega \mu_{A_y}(x,\omega) \cdot dP(\omega) = \begin{cases} 0 & x \neq y \\ 1 & x = y. \end{cases} \tag{3-102}$$

3-5. SOME PROPERTIES OF PROBABILISTIC SETS.

Some properties of probabilistic sets can be characterized from a lattice theoretical viewpoint [6].

A family of probabilistic sets $(P(X),\subset)$ constitutes a poset (cf. the remark of [Def. 3-8]). For arbitrary A,B $(\varepsilon P(X))$, there exist a supremum $A\cup B$ and an infimum $A\cap B$ with respect to this partial order $\subset$(see [Prop.3-4] 2) and 3)). Hence the poset $(P(X),\subset)$ forms a lattice and the following proposition holds. (Note that a set of the following properties is a necessary and sufficient condition of being a lattice.)

[Prop.3-5]

For atbitrary $A,B,C(\varepsilon P(X))$, we have

commutativity

$A\cup B= B\cup A,$ (3-103) $A\cap B= B\cap A,$ (3-104)

associativity

$(A\cup B)\cup C= A\cup(B\cup C),$ (3-105) $(A\cap B)\cap C= A\cap(B\cap C),$ (3-106)

absorption law

$A\cup(A\cap B)= A,$ (3-107) $A\cap(A\cup B)= A.$ (3-108)

It is also possible to show that there exist pseudo complements in $P(X)$. Let A and B be arbitrarily fixed two probabilistic sets whose defining functions are $\mu_A(x,\omega)$ and $\mu_B(x,\omega)$,respectively. Consider the following equation,

$$\mu_{A'}(x,\omega)=\begin{cases} 1 & \text{if } \mu_A(x,\omega)\leqq \mu_B(x,\omega) \\ \mu_B(x,\omega) & \text{if } \mu_A(x,\omega)> \mu_B(x,\omega). \end{cases} \quad (3\text{-}109)$$

For each $x(\varepsilon X)$, $\mu_{A'}(x,\cdot)$ is a (B,B_C)-measurable function and is an element of M, since a set $\{\omega|\mu_B(x,\omega)-\mu_A(x,\omega)\geqq 0\}$ belongs to B (i.e. this set is measurable). Hence it is possible to define a probabilistic set A' by the equation (3-109). It is also clear that A' is the largest probabilistic set of those C's which satisfy $A \cap C \subset B$ ($C \in P(X)$). In this sense, A' is said to be a *pseudo complement* of A relative to B. Hence ($P(X)$, $\subset$) constitutes a *pseudo-Boolean algebra* (cf.Fig. 3-3 and [7]). (A pseudo- Boolean algebra is a relative complenented lattice with a minimum element. In this case the minimum element is ϕ.)

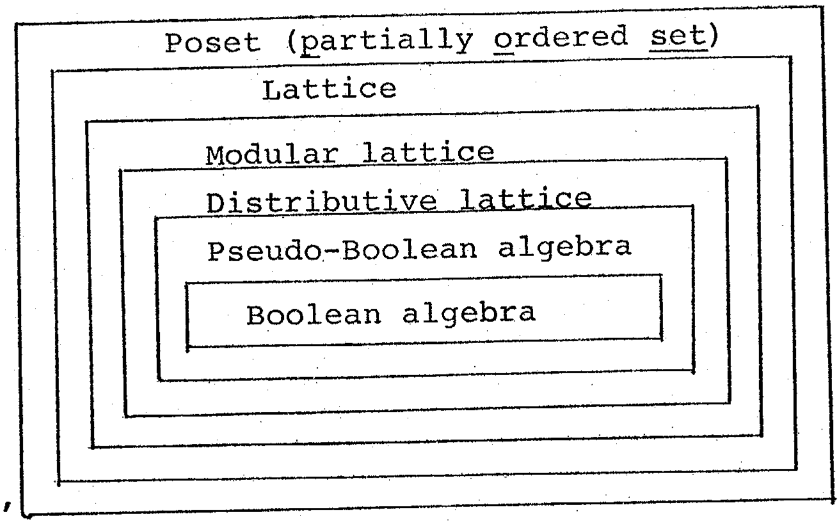


Fig.3-3. An inclusion diagram of various lattices.

Moreover, for arbitrary $\{A_\gamma\}_{\gamma \in \Gamma}$ ($\subset P(X)$) (Γ: possibly infinite), the existence of $\bigcup A_\gamma$ and $\bigcap A_\gamma$ was shown in a previous section and they played a role of a supremum and an infimum with respect to the order $\subset$. Hence it is concluded that the lattice ($P(X)$, $\subset$) is *complete*, and so we can obtain the following theorem 3-1 from a lattice theoretical viewpoint.

[Theorem 3-1]

A family of probabilistic sets ($P(X)$, $\subset$) constitutes a complete pseudo-Boolean algebra.

Remark: In ordinary set theory, a family of all subsets constitutes a complete Boolean algebra. The difference between the two is lack of complemented law (i.e. $A \cup A^c \neq X$, $A \cap A^c \neq \phi$).

In probabilistic set theory, it is essential to consider ambiguous states, so we can not get any definite informations if we know that the considered object is not in one state. In ordinary set theory, however, we can definitely conclude that the considered object is in another state from the information that it is not in one state. (Note that ordinary set theory can be considered to be a two-valued logic.) Hence the lack of complemented law is unavoidable in probabilistic set theory.

Since a notion of pseudo-Boolean algebra is included in that of distributive lattice (cf. Fig.3-3), a distributive law holds in $(\mathcal{P}(X), \subset)$. Moreover, in connection with its completeness, we can generalize commutative law, associative law, distributive law, and de-Morgan's law as follows.

[Prop.3-6]

For arbitrary subfamilies of probabilistic sets $\{A_\gamma\}_{\gamma\in\Gamma}$ and $\{B_\lambda\}_{\lambda\in\Lambda}$, we have

generalized associative law

$$\Big(\bigcup_{\gamma\in\Gamma} A_\gamma\Big) \cup \Big(\bigcup_{\lambda\in\Lambda} B_\lambda\Big) = \bigcup_{\gamma,\lambda} (A_\gamma \cup B_\lambda), \tag{3-110}$$

$$\Big(\bigcap_{\gamma\in\Gamma} A_\gamma\Big) \cap \Big(\bigcap_{\lambda\in\Lambda} B_\lambda\Big) = \bigcap_{\gamma,\lambda} (A_\gamma \cap B_\lambda), \tag{3-111}$$

generalized distributive law

$$\Big(\bigcup_{\gamma\in\Gamma} A_\gamma\Big) \cap \Big(\bigcup_{\lambda\in\Lambda} B_\lambda\Big) = \bigcup_{\gamma,\lambda} (A_\gamma \cap B_\lambda), \tag{3-112}$$

$$\Big(\bigcap_{\gamma\in\Gamma} A_\gamma\Big) \cup \Big(\bigcap_{\lambda\in\Lambda} B_\lambda\Big) = \bigcap_{\gamma,\lambda} (A_\gamma \cup B_\lambda), \tag{3-113}$$

generalized de-Morgan's law

$$\Big(\bigcup_{\gamma\in\Gamma} A_\gamma\Big)^c = \bigcap_{\gamma\in\Gamma} A_\gamma^{\,c}, \tag{3-114}$$

$$(\bigcap_{\gamma\varepsilon\Gamma} A_\gamma)^c = \bigcup_{\gamma\varepsilon\Gamma} A_\gamma{}^c. \qquad (3\text{-}115)$$

<u>proof</u>

generalized associative law (3-110)

It is sufficient to show, for each fixed $x\varepsilon X$,

$$\max\{\sup_{1\leqq j<\infty}\max_{1\leqq i\leqq n_j}\mu_{A\gamma_i}(x,\omega),\ \sup_{1\leqq l<\infty}\max_{1\leqq k\leqq n_l}\mu_{B_{\lambda k}}(x,\omega)\} =$$

$$= \sup_{1\leqq j'<\infty}\ \max_{\substack{1\leqq i'\leqq n_{j'},\\ 1\leqq \hat{i}'\leqq \hat{n}_{j'}}}\{\mu_{A\gamma_{i'}}(x,\omega),\ \mu_{B_{\lambda_{\hat{i}'}}}(x,\omega)\} \text{ for P a.e. } \omega\varepsilon\Omega. \qquad (3\text{-}116)$$

By applying Prop.3-4, we have

$$\mu_{A\gamma_i}(x,\omega) \leqq \text{the right side of (3-116) for P a.e. } \omega\varepsilon\Omega \text{ all } \gamma_i\varepsilon\Gamma, \qquad (3\text{-}117)$$

$$\therefore \max_{1\leqq i\leqq n_j}\mu_{A\gamma_i}(x,\omega) \leqq \text{the right side of (3-116)} \quad \text{for P a.e. } \omega\varepsilon\Omega, \text{ all } n_j\varepsilon N, \qquad (3\text{-}118)$$

$$\therefore \sup_{1\leqq j<\infty}\max_{1\leqq i\leqq n_j}\mu_{A\gamma_i}(x,\omega) \leqq \text{the right side of (3-116)} \quad \text{for P a.e. } \omega\varepsilon\Omega. \qquad (3\text{-}119)$$

In the same manner, we have

$$\sup_{1\leqq l<\infty}\max_{1\leqq k\leqq n_l}\mu_{B_{\lambda k}}(x,\omega) \leqq \text{the right side of (3-116)} \quad \text{for P a.e. } \omega\varepsilon\Omega. \qquad (3\text{-}120)$$

Hence we can conclude by (3-119),(3-120) that

$$\text{the left side of (3-116)} \leqq \text{the right side of (3-116)} \quad \text{for P a.e. } \omega\varepsilon\Omega. \qquad (3\text{-}121)$$

Contrary to this, since we have

$$\mu_{A\gamma_{i'}}(x,\omega) \leqq \text{the left side of (3-116)} \quad \text{for P a.e. } \omega\varepsilon\Omega, \text{ all } \gamma_{i'}\varepsilon\Gamma, \qquad (3\text{-}122)$$

$$\mu_{B_{\lambda_{\hat{i}'}}}(x,\omega) \leqq \text{the left side of (3-116)} \quad \text{for P a.e. } \omega\varepsilon\Omega, \text{ all } \lambda_{\hat{i}'}\varepsilon\Lambda, \qquad (3\text{-}123)$$

therefore

$$\max_{\substack{1 \leqq i' \leqq n_{j'} \\ 1 \leqq \hat{i}' \leqq \hat{n}_{j'}}} \{\mu_{A\gamma_{i'}}(x,\omega), \mu_{B\lambda_{\hat{i}'}}(x,\omega)\} \leqq \text{ the left side of (3-116)}$$

$$\text{for P a.e. } \omega \varepsilon \Omega, \text{ all } j' \varepsilon N. \qquad (3\text{-}124)$$

Taking the upper limit of the left side of (3-124) with respect to j', we have

$$\text{the right side of (3-116)} \leqq \text{ the left side of (3-116)}$$

$$\text{for P a.e. } \omega \varepsilon \Omega. \qquad (3\text{-}125)$$

Hence the desired result (3-116) can be obtained.

generalized associative law (3-111)

It will be easily confirmed in much the same way as (3-116). Exchange "sup" and "max" for "inf" and "min", respectively.

generalized distributive law (3-112)

All we have to show is that, under the restriction of $x \varepsilon X$ being arbitrarily fixed,

$$\min \{\sup_{1 \leqq j < \infty} \max_{1 \leqq i \leqq n_j} \mu_{A_{\gamma_i}}(x,\omega), \sup_{1 \leqq k < \infty} \max_{1 \leqq i \leqq n_k} \mu_{B_{\lambda_i}}(x,\omega)\} =$$

$$= \sup_{1 \leqq l < \infty} \max_{1 \leqq i \leqq n_l} \min\{\mu_{A_{\gamma_i}}(x,\omega), \mu_{B_{\lambda_i}}(x,\omega)\} \text{ for P a.e. } \omega \varepsilon \Omega. \qquad (3\text{-}126)$$

Since for each $\gamma_i \varepsilon \Gamma$,

$$\mu_{A_{\gamma_i}}(x,\omega) \leqq \sup_{1 \leqq j < \infty} \max_{1 \leqq i \leqq n_j} \mu_{A_{\gamma_i}}(x,\omega) \quad \text{for P a.e. } \omega \varepsilon \Omega, \qquad (3\text{-}127)$$

and since for each $\lambda_i \varepsilon \Lambda$,

$$\mu_{B_{\lambda_i}}(x,\omega) \leqq \sup_{1 \leqq k < \infty} \max_{1 \leqq i \leqq n_k} \mu_{B_{\lambda_i}}(x,\omega) \quad \text{for P a.e. } \omega \varepsilon \Omega, \qquad (3\text{-}128)$$

we can conclude that

$$\min \{\mu_{A_{\gamma_i}}(x,\omega), \mu_{B_{\lambda_i}}(x,\omega)\} \leqq \text{ the left side of (3-126)}$$

$$\text{for P a.e. } \omega \varepsilon \Omega, \qquad (3\text{-}129)$$

$$\max_{1 \leqq i \leqq n_l} \min\{\mu_{A_{\gamma_i}}(x,\omega), \mu_{B_{\lambda_i}}(x,\omega)\} \leqq \text{ the left side of (3-126)}$$

$$\text{for P a.e. } \omega \varepsilon \Omega, \qquad (3\text{-}130)$$

$\therefore$ the right side of (3-126) $\leq$ the left side of (3-126)

$$\text{for } P \text{ a.e. } \omega\varepsilon\Omega. \qquad (3\text{-}131)$$

On the other hand we have

$$\text{the left side of (3-126)} = \sup \min\{\mu_{A\gamma_i}(x,\omega), \mu_{B\lambda_i}(x,\omega)\}, \qquad (3\text{-}132)$$

where the least upper bound is taken in the region of all indeces γ_i's and λ_i's appeared in the left side of (3-126). Whereas we have

$$\min\{\mu_{A\gamma_i}(x,\omega), \mu_{B\lambda_i}(x,\omega)\} \leq \text{the right side of (3-126)}$$
$$\text{for } P \text{ a.e. } \omega\varepsilon\Omega, \qquad (3\text{-}133)$$

therefore

the left side of (3-126) $\leq$ the right side of (3-126)

$$\text{for } P \text{ a.e. } \omega\varepsilon\Omega, \qquad (3\text{-}134)$$

and the desired result (3-126) can be confirmed.

generalized distributive law (3-113)

It will be obtained in the same manner as (3-126).

generalized de-Morgan's law

In order to prove (3-114), it is sufficient to show, for each $x\varepsilon X$,

$$1 - \sup_{1\leq j<\infty} \max_{1\leq i\leq n_j} \mu_{A\gamma_i}(x,\omega) = \inf_{1\leq k<\infty} \min_{1\leq l\leq n_k} \{ 1 - \mu_{A\gamma_l}(x,\omega) \}$$
$$\text{for } P \text{ a.e. } \omega\varepsilon\Omega. \qquad (3\text{-}135)$$

We shall omit to give a proof of (3-135), since it can be obtained in much the same manner as abovementioned discussions. Another de-Morgan's law (3-115) can be obtained in the same way.

(Q.E.D.)

Some other important properties in $\mathcal{P}(X)$ will be mentioned in the following without proofs. (All the following statements will be easily confirmed by an investigation in terms of defining functions of probabilistic sets.)

[Prop.3-7]

For arbitrary $A,B,C(\varepsilon\mathcal{P}(X))$, we have

idempotent law

$$A\cup A= A, \quad (3\text{-}136) \qquad A\cap A= A, \quad (3\text{-}137)$$

involution law

$$A^{cc}= A, \quad (3\text{-}138)$$

elimination law

$$\left.\begin{array}{l} A\cup B= A\cup C \\ A\cap B= A\cap C \end{array}\right\} \longrightarrow B= C, \quad (3\text{-}139)$$

identity law

$$A\cup X= X, \quad (3\text{-}140) \qquad A\cap X= A, \quad (3\text{-}141)$$

$$A\cup \phi= A, \quad (3\text{-}142) \qquad A\cap\phi= \phi, \quad (3\text{-}143)$$

[Prop.3-8]

For arbitrary $\{A_n\}_{n=1}^{\infty}(\subset\mathcal{P}(X))$, we have

$$\underline{\lim}_{n\to\infty} A_n \subset \overline{\lim}_{n\to\infty} A_n, \quad (3\text{-}144)$$

$$\overline{\lim}_{n\to\infty} A_n^{\,c}= (\underline{\lim}_{n\to\infty} A_n)^{c}. \quad (3\text{-}145)$$

If $A_1\subset A_2\subset\ldots\subset A_n\subset\ldots$, then we have

$$\underline{\lim}_{n\to\infty} A_n= \overline{\lim}_{n\to\infty} A_n= \bigcup_{n=1}^{\infty} A_n. \quad (3\text{-}146)$$

If $A_1\supset A_2\supset\ldots\supset A_n\supset\ldots$, then we have,

$$\underline{\lim}_{n\to\infty} A_n= \overline{\lim}_{n\to\infty} A_n= \bigcap_{n=1}^{\infty} A_n. \quad (3\text{-}147)$$

If $A_{2n+1}= A$ and $A_{2n}= B$, then we have

$$\underset{n\to\infty}{\underline{\lim}} A_n = A \cap B \qquad \text{and} \qquad \underset{n\to\infty}{\overline{\lim}} A_n = A \cup B. \tag{3-148}$$

[Prop.3-9]

Each of $(P(X),\cup)$, $(P(X),\cap)$, $(P(X),\cdot)$, and $(P(X),\oplus)$ constitutes a commutative <u>monoid</u> (i.e. a commutative semigroup with a unit) and, for atbitrary $A,B,C(\varepsilon P(X))$, we have

$$A \Delta B = (A-B) \cup (B-A), \tag{3-149}$$

$$A \oplus B = (A^c B^c)^c, \tag{3-150}$$

$$A \cdot B \subset A \cap B \subset A \underset{\lambda}{+} B \subset A \cup B \subset A \oplus B. \tag{3-151}$$

Remark : In ordinary set theory, it is possible to define sixteen different kinds of binary operations. (Because the total space X can be devided in four regions for arbitrary subsets A and B, there exist $2^4=16$ combinations.) Among these sixteen binary operations, symmetric difference $A \Delta B$ has a very good property from an algebraic viewpoint, namely, it constitutes an Abelian group. In probabilistic set theory, however, $(P(X),\Delta)$ doesn't satisfy such a good property. On the contrary, it doesn't satisfy associative law.

[Prop.3-10]

Let $X_\gamma(\gamma\varepsilon\Gamma)$ be prural total spaces (possibly infinite), then we have

$$\bigcup_{\gamma\varepsilon\Gamma} P(X_\gamma) \subset P\left(\bigcup_{\gamma\varepsilon\Gamma} X_\gamma\right), \tag{3-152}$$

$$\bigcap_{\gamma\varepsilon\Gamma} P(X_\gamma) = P\left(\bigcap_{\gamma\varepsilon\Gamma} X_\gamma\right). \tag{3-153}$$

3-6. INDUCED IMAGES OF PROBABILISTIC SETS AND PROBABILISTIC MAPPINGS.

A probabilistic set A on a total space X was defined by a defining function $\mu_A(x,\omega)$. In this section, we shall consider two total spaces $X=\{x\}$, $Y=\{y\}$ and a mapping f from X to Y. We shall discuss induced images of probabilistic sets by the mapping f.

Let f be a mapping from X to Y,

$$f: X \longrightarrow Y. \quad x \longmapsto f(x) \qquad (3\text{-}154)$$

An image f(A) of an (ordinary) subset A of X by f is defined by

$$f(A)=\{f(x) \mid x \in A\} \ (\subset Y). \qquad (3\text{-}155)$$

Conversely, an inverse image $f^{-1}(B)$ of an (ordinary) subset B of Y by f is defined by,

$$f^{-1}(B)=\{x \mid f(x)=y,\ y \in B\} = \{f^{-1}(y) \mid y \in B\} \ (\subset X). \qquad (3\text{-}156)$$

In the case that A and B are probabilistic sets, the same notions can be defined as follows.

[Def.3-10]

We assume that $f:X\longrightarrow Y$ is a mapping and that $B \in \mathcal{P}(Y)$ (the defining function is $\mu_B(y,\omega)$ for $y \in Y$, $\omega \in \Omega$). An induced inverse image $f^{-1}(B)$ of the probabilistic set B by f is defined as a probabilistic set on X by the following defining function,

$$\mu_{f^{-1}(B)}(x,\omega)= \mu_B(f(x),\omega) \quad \text{for all } x\in X \text{ and P a.e. } \omega\in\Omega. \qquad (3\text{-}157)$$

Let A be a probabilistic set on X whose defining function is $\mu_A(x,\omega)$, then an induced image f(A) of A by f is defined as a probabilistic set on X by the following defining function,

$$\mu_{f(A)}(y,\omega) = \sup\{\mu_A(x,\omega) \mid x \in f^{-1}(y)\} \text{ for all } y \in Y \text{ and } P \text{ a.e. } \omega \in \Omega. \tag{3-158}$$

Remark : In the equation (3-158), if $f^{-1}(y) = \phi$ then the value of the right side is, of course, the minimum-value 0.

[Prop.3-11]

In the situation of Def 3-10, we have the following properties,

1) $B_1 \subset B_2 \implies f^{-1}(B_1) \subset f^{-1}(B_2)$, (3-159)

2) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$, (3-160)

3) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$, (3-161)

4) $f^{-1}(B^c) = f^{-1}(B)^c$, (3-162)

5) $A_1 \subset A_2 \implies f(A_1) \subset f(A_2)$, (3-163)

6) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$, (3-164)

7) $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$, (3-165)

8) $A \subset f^{-1}(f(A))$, (3-166)

9) $B \supset f(f^{-1}(B))$. (3-167)

<u>proof</u>

1) $\mu_{f^{-1}(B_1)}(x,\omega) = \mu_{B_1}(f(x),\omega) \leqq \mu_{B_2}(f(x),\omega) = \mu_{f^{-1}(B_2)}(x,\omega)$

for all $x \in X$ and P a.e. $\omega \in \Omega$. (3-168)

$\therefore f^{-1}(B_1) \subset f^{-1}(B_2)$. (3-169)

2) $\mu_{f^{-1}(B_1 \cup B_2)}(x,\omega) = \mu_{B_1 \cup B_2}(f(x),\omega) =$

$= \max\{\mu_{B_1}(f(x),\omega), \mu_{B_2}(f(x),\omega)\} =$

$= \max\{\mu_{f^{-1}(B_1)}(x,\omega), \mu_{f^{-1}(B_2)}(x,\omega)\} =$

$= \mu_{f^{-1}(B_1) \cup f^{-1}(B_2)}(x,\omega)$ for all $x \in X$ and P a.e. $\omega \in \Omega$. (3-170)

$\therefore f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$. (3-171)

3) $\mu_{f^{-1}(B_1 \cap B_2)}(x,\omega) = \mu_{B_1 \cap B_2}(f(x),\omega)$

$= \min\{\mu_{B_1}(f(x),\omega), \mu_{B_2}(f(x),\omega)\}$

$$= \min\{ \mu_{f^{-1}(B_1)}(x,\omega),\ \mu_{f^{-1}(B_2)}(x,\omega) \} =$$

$$= \mu_{f^{-1}(B_1)\cap f^{-1}(B_2)}(x,\omega) \quad \text{for all } x\varepsilon X \text{ and P a.e. } \omega\varepsilon\Omega. \quad (3\text{-}172)$$

$$\therefore f^{-1}(B_1\cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2). \quad (3\text{-}173)$$

4) $\mu_{f^{-1}(B^c)}(x,\omega) = \mu_{B^c}(f(x),\omega) = 1- \mu_B(f(x),\omega)$

$$= 1- \mu_{f^{-1}(B)}(x,\omega) = \mu_{f^{-1}(B)^c}(x,\omega) \quad \text{for all } x\varepsilon X \text{ and P a.e. } \omega\varepsilon\Omega.$$

(3-174)

$$\therefore f^{-1}(B^c) = f^{-1}(B)^c. \quad (3\text{-}175)$$

5) $\mu_{f(A_1)}(y,\omega) = \sup_{x\varepsilon f^{-1}(y)} \mu_{A_1}(x,\omega) \leqq \sup_{x\varepsilon f^{-1}(y)} \mu_{A_2}(x,\omega) =$

$$= \mu_{f(A_2)}(y,\omega) \quad \text{for all } y\varepsilon Y \text{ and P a.e. } \omega\varepsilon\Omega. \quad (3\text{-}176)$$

$$\therefore f(A_1) \subset f(A_2). \quad (3\text{-}177)$$

6) $\mu_{f(A_1\cup A_2)}(y,\omega) = \sup_{x\varepsilon f^{-1}(y)} \mu_{A_1\cup A_2}(x,\omega) =$

$$= \sup_{x\varepsilon f^{-1}(y)} \max\{\mu_{A_1}(x,\omega),\ \mu_{A_2}(x,\omega)\} =$$

$$= \max\{ \sup_{x\varepsilon f^{-1}(y)} \mu_{A_1}(x,\omega),\ \sup_{x\varepsilon f^{-1}(y)} \mu_{A_2}(x,\omega)\} =$$

$$= \max\{\mu_{f(A_1)}(y,\omega),\ \mu_{f(A_2)}(y,\omega)\} =$$

$$= \mu_{f(A_1)\cup f(A_2)}(y,\omega) \quad \text{for all } y\varepsilon Y \text{ and P a.e. } \omega\varepsilon\Omega. \quad (3\text{-}178)$$

$$\therefore f(A_1\cup A_2) = f(A_1) \cup f(A_2). \quad (3\text{-}179)$$

7) $\mu_{f(A_1\cap A_2)}(y,\omega) = \sup_{x\varepsilon f^{-1}(y)} \{\mu_{A_1\cap A_2}(x,\omega)\} =$

$$= \sup_{x\varepsilon f^{-1}(y)} \min\{ \mu_{A_1}(x,\omega), \mu_{A_2}(x,\omega)\} \leqq$$

$$\leqq \min\{\sup_{x\varepsilon f^{-1}(y)} \mu_{A_1}(x,\omega),\ \sup_{x\varepsilon f^{-1}(y)} \mu_{A_2}(x,\omega)\} =$$

$$= \min\{\mu_{f(A_1)}(y,\omega), \mu_{f(A_2)}(y,\omega)\} =$$

$$= \mu_{f(A_1)\cap f(A_2)}(y,\omega) \quad \text{for all } y\varepsilon Y \text{ and P a.e. } \omega\varepsilon\Omega. \quad (3\text{-}180)$$

$$\therefore f(A_1\cap A_2) \subset f(A_1) \cap f(A_2). \quad (3\text{-}181)$$

8) $\mu_{f^{-1}(f(A))}(x,\omega) = \mu_{f(A)}(f(x),\omega) = \sup_{z\varepsilon f^{-1}(f(x))} \mu_A(z,\omega) \geqq$

$\geqq \mu_A(x,\omega)$ for all $x\varepsilon X$ and P a.e. $\omega\varepsilon\Omega$. (3-182)

$\therefore f^{-1}(f(A)) \supset A.$ (3-183)

9) For each $y\varepsilon Y$ such that $f^{-1}(y) \neq \phi$, we have

$$\mu_{f(f^{-1}(B))}(y,\omega) = \sup_{x\varepsilon f^{-1}(y)} \mu_{f^{-1}(B)}(x,\omega) = \sup_{x\varepsilon f^{-1}(y)} \mu_B(f(x),\omega) =$$

$$= \mu_B(y,\omega) \quad \text{for P a.e. } \omega\varepsilon\Omega. \tag{3-184}$$

For each $y\varepsilon Y$ such that $f^{-1}(y)=\phi$, we have

$$\mu_{f(f^{-1}(B))}(y,\omega) = 0 \leqq \mu_B(y,\omega) \quad \text{for P a.e. } \omega\varepsilon\Omega. \tag{3-185}$$

Hence we have

$$f(f^{-1}(B)) \subset B. \tag{3-186}$$

(Q.E.D.)

A mapping f from X to Y is usually defined as a correspondence from an element $x(\varepsilon X)$ to an element $y(\varepsilon Y)$. There also exist some variations such as a set function (a correspondence from a subset $A(\subset X)$ to an element $y(\varepsilon Y)$) and a multivalued mapping (a correspondence from an element $x(\varepsilon X)$ to a subset $B(\subset Y)$). The concept of set functions and multivalued mappings play an important role in the fields of measure theory [8] and functional analysis [9], respectively. In the field of pattern recognition or learning theory, it is essential to consider an ambiguous correspondence (i.e. a probabilistic mapping) [10] [11] which will be defined as follows.

[Def.3-11]

A probabilistic mapping f from X to Y on a parameter space (Ω_m, B_m, P_m) is defined by

$$\begin{array}{rcl} f: X \times \Omega_m & \longrightarrow & Y. \\ \omega & & \omega \\ (x,\omega_m) & \longmapsto & f(x,\omega_m) \end{array} \tag{3-187}$$

Some extended concepts can be defined in connection with probabilistic mappings, such as induced images and induced inverse images of probabilistic sets by a probabilistic mapping, and some properties can also be investigated.

[Def.3-12]

Let $f: X \times \Omega_m \longrightarrow Y$ be a probabilistic mapping, and B be a probabilistic set on Y whose defining function is $\mu_B(y,\omega)$ ($y \varepsilon Y$, $\omega \varepsilon \Omega$). An induced inverse image $f^{-1}(B)$ of B by f is defined by the following defining function,

$$\mu_{f^{-1}(B)}(x,(\omega,\omega_m)) = \mu_B(f(x,\omega_m),\omega) \quad \text{for all } x \varepsilon X,$$

$$P \text{ a.e. } \omega \varepsilon \Omega,\ P_m \text{ a.e.} \omega_m \varepsilon \Omega_m. \quad (3\text{-}188)$$

If A is a probabilistic set on X whose defining function is $\mu_A(x,\omega)$ ($x \varepsilon X, \omega \varepsilon \Omega$), then an induced image f(A) of A by f is defined by the following defining function,

$$\mu_{f(A)}(y,(\omega,\omega_m)) = \sup_{\{x | f(x,\omega_m)=y\}} \mu_A(x,\omega)$$

$$\text{for all } y \varepsilon Y,\ P \text{ a.e. } \omega \varepsilon \Omega,\ P_m \text{ a.e. } \omega_m \varepsilon \Omega_m. \quad (3\text{-}189)$$

The same proposition as Prop.3-11 is also valid in this case except (3-166) and (3-167), but it is omitted here.

3-7. MOMENT ANALYSIS OF PROBABILISTIC SETS.

The parameter space (Ω, B, P) is a (probability) measure space and plays an essential role in applications of probabilistic set theory. By using the measure P of this parameter space, we can carry out moment analysis. The possibility of moment analysis is one of the most important features in probabilistic set theory and can not be found out in other general theories such as fuzzy set theory.

[Def.3-12]

Let A be a probabilistic set on X whose defining function is $\mu_A(x,\omega)$. For each fixed $x(\varepsilon X)$, mean value $E(\mu_A)(x)$, variance $V(\mu_A)(x)$, standard deviation $\sigma(\mu_A)(x)$, n-th moment $M^n(\mu_A)(x)$, n-th moment around mean value $M_o^n(\mu_A)(x)$, n-th absolute moment around mean value $\overline{M}_o^n(\mu_A)(x)$ are defined as follows.

$$E(\mu_A)(x) = \int_\Omega \mu_A(x,\omega)\cdot dP(\omega) \quad (\overline{\overline{\Delta}}\ M^1(\mu_A)(x)), \tag{3-190}$$

$$V(\mu_A)(x) = \int_\Omega (\mu_A(x,\omega) - E(\mu_A)(x))^2 dP(\omega) \quad (\overline{\overline{\Delta}}\ M_o^2(\mu_A)(x)), \tag{3-191}$$

$$\sigma(\mu_A)(x) = \sqrt{V(\mu_A)(x)}, \tag{3-192}$$

$$M^n(\mu_A)(x) = \int_\Omega \mu_A(x,\omega)^n \cdot dP(\omega) \qquad (n\varepsilon N), \tag{3-193}$$

$$M_o^n(\mu_A)(x) = \int_\Omega (\mu_A(x,\omega) - E(\mu_A)(x))^n \cdot dP(\omega), \tag{3-194}$$

$$\overline{M}_o^n(\mu_A)(x) = \int_\Omega |\mu_A(x,\omega) - E(\mu_A)(x)|^n \cdot dP(\omega). \tag{3-195}$$

The justification of abovestated definitions is ensured by Prop.3-3, and the following properties follow from these definitions.

[Prop.3-12]

In the situation of Def.3-12, we have

$$0 \leqq E(\mu_A)(x) \leqq 1 \qquad \text{for all } x\varepsilon X, \tag{3-196}$$

$$0 \leqq \ldots \leqq M^3(\mu_A)(x) \leqq M^2(\mu_A)(x) \leqq M^1(\mu_A)(x) = E(\mu_A)(x) \leqq$$

$$\leqq \{M^2(\mu_A)(x)\}^{1/2} \leqq \{M^3(\mu_A)(x)\}^{1/3} \leqq \ldots \leqq 1 \quad \text{for all } x \in X, \tag{3-197}$$

$$V(\mu_A)(x) = M^2(\mu_A)(x) - (E(\mu_A)(x))^2 \quad \text{for all } x \in X, \tag{3-198}$$

$$n \geqq m \ (\geqq 1) \Longrightarrow 0 \leqq \overline{M}_o^n(\mu_A)(x) \leqq \overline{M}_o^m(\mu_A)(x) \leqq 1 \quad \text{for all } x \in X, \tag{3-199}$$

$$\lim_{n\to\infty} M_o^n(\mu_A)(x) = \lim_{n\to\infty} \overline{M}_o^n(\mu_A)(x) = 0 \quad \text{for all } x \in X. \tag{3-200}$$

<u>proof</u>

In the following statements, $x(\varepsilon X)$ is arbitrarily fixed. Since $\mu_A(x,\cdot)$ is (B,B_c)-measurable and

$$0 \leqq \mu_A(x,\cdot) \leqq 1, \tag{3-201}$$

therefore we have

$$0 = \int_\Omega 0 \cdot dP(\omega) \leqq \int_\Omega \mu_A(x,\omega) \cdot dP(\omega) \leqq \int_\Omega 1 \cdot dP(\omega) \doteq 1. \tag{3-202}$$

If $n \geqq m$, then

$$0 \leqq \mu_A(x,\omega)^n \leqq \mu_A(x,\omega)^m \leqq 1. \tag{3-203}$$

Integrating each term, we have

$$0 \leqq M^n(\mu_A)(x) \leqq M^m(\mu_A)(x) \leqq 1. \tag{3-204}$$

If we consider the well-known Hölder's inequality [12],

$$\int_\Omega |f(\omega)g(\omega)| \cdot dP(\omega) \leqq \left(\int_\Omega |f(\omega)|^s \cdot dP(\omega)\right)^{1/s} \cdot \left(\int_\Omega |g(\omega)|^t \cdot dP(\omega)\right)^{1/t},$$

$$(1/s)+(1/t) = 1, \quad s,t > 1, \tag{3-205}$$

and if we put

$$f(\omega) \overline{\overline{\Delta}} \mu_A(x,\omega)^m, \quad g(\omega) \overline{\overline{\Delta}} 1, \quad s \overline{\overline{\Delta}} n/m, \ t \overline{\overline{\Delta}} n/(n-m), \tag{3-206}$$

then we have

$$\int_\Omega \mu_A(x,\omega)^m dP(\omega) \leqq \left(\int_\Omega \mu_A(x,\omega)^n dP(\omega)\right)^{m/n} \cdot 1, \tag{3-207}$$

therefore

$$\{M^m(\mu_A)(x)\}^{1/m} \leqq \{M^n(\mu_A)(x)\}^{1/n}. \tag{3-208}$$

By (3-204) and (3-208), we have (3-197). A direct calculation shows (3-198). We can easily verify (3-199) in much the same way as (3-197). At last we shall give a proof of (3-200) by using the Lebesgue's dominated convergence theorem [5].

$$\lim_{n\to\infty} \int_\Omega (\mu_A(x,\omega)-E(\mu_A)(x))^n \cdot dP(\omega)$$

$$= \int_\Omega \lim_{n\to\infty} (\mu_A(x,\omega)-E(\mu_A)(x))^n \cdot dP(\omega) = \int_\Omega 0 \cdot dP(\omega) = 0. \quad (3\text{-}209)$$

(Q.E.D.)

[Def.3-13]

Let A and B be probabilistic sets on X. For each fixed $x \varepsilon X$, <u>covariance</u> $C(\mu_A,\mu_B)(x)$ and <u>correlation coefficient</u> $r(\mu_A,\mu_B)(x)$ are defined by

$$C(\mu_A,\mu_B)(x) = \int_\Omega (\mu_A(x,\omega)-E(\mu_A)(x)) \cdot (\mu_B(x,\omega)-E(\mu_B)(x)) \cdot dP(\omega), \quad (3\text{-}210)$$

$$r(\mu_A,\mu_B)(x) = C(\mu_A,\mu_B)(x) / \sqrt{V(\mu_A)(x) \cdot V(\mu_B)(x)}. \quad (3\text{-}211)$$

(If $V(\mu_A)(x) \cdot V(\mu_B)(x) = 0$, $r(\mu_A,\mu_B)(x)$ is not defined.)

[Prop.3-13]

In the situation of Def.3-13, we have

$$C(\mu_A,\mu_A)(x) = V(\mu_A)(x), \quad (3\text{-}212)$$

$$C(\mu_A,\mu_B)(x) = E(\mu_A,\mu_B)(x) - E(\mu_A)(x) \cdot E(\mu_B)(x), \quad (3\text{-}213)$$

$$0 \leqq |C(\mu_A,\mu_B)(x)| \leqq V(\mu_A)(x) \cdot V(\mu_B)(x) \leqq 1, \quad (3\text{-}214)$$

$$0 \leqq |r(\mu_A,\mu_B)(x)| \leqq 1, \quad (3\text{-}215)$$

$r(\mu_A,\mu_B)(x) = \pm 1 \Longleftrightarrow$ there exist real numbers a and b such that $\mu_A(x,\omega) = a \cdot \mu_B(x,\omega)+b$ for P a.e.$\omega \varepsilon \Omega$.(3-216)

[Def.3-14]

Let $A_1, A_2, \ldots, A_n$ be probabilistic sets on X whose defining functions are $\mu_{A_1}(x,\omega)$, $\mu_{A_2}(x,\omega), \ldots, \mu_{A_n}(x,\omega)$, respectively. For arbitrary $x,y(\varepsilon X)$, <u>moment matrix</u> $M(x,y)$ and <u>variance-covar-</u>

iance matrix V(x,y) of $A_1, A_2, \ldots, A_n$ are defined by

$$M(x,y) = [m_{i,j}]_{1 \leqq i,j \leqq n}$$

$$\text{where } m_{i,j} = \int_\Omega \mu_{A_i}(x,\omega) \cdot \mu_{A_j}(X,\omega) \cdot dP(\omega), \tag{3-217}$$

$$V(x,y) = [v_{i,j}]_{1 \leqq i,j \leqq n}$$

$$\text{where } v_{i,j} = \int_\Omega (\mu_{A_i}(x,\omega) - E(\mu_{A_i})(x)) \cdot (\mu_{A_j}(x,\omega) - E(\mu_{A_j})(x)) \cdot dP(\omega). \tag{3-218}$$

[Prop.3-14]

Both M(x,x) and V(x,x) are positive-definite, symmetric matrices.

proof

We shall mention only the positiveness of M(x,x).

$$\sum_{i,j=1}^{n} a_i \bar{a}_j \int_\Omega \mu_{A_i}(x,\omega) \mu_{A_j}(x,\omega) dP(\omega) =$$

$$= \int_\Omega \sum_{i,j=1}^{n} a_i \mu_{A_i}(x,\omega) \cdot \bar{a}_j \mu_{A_j}(x,\omega) \cdot dP(\omega) =$$

$$= \int_\Omega \left| \sum_{i=1}^{n} a_i \mu_{A_i}(x,\omega) \right|^2 dP(\omega) \geqq 0, \tag{3-219}$$

for all $a_i \varepsilon C$ (complex numbers).

(Q.E.D.)

3-8. EXPECTED CARDINAL NUMBER.

In ordinary set theory, a notion of the <u>cardinal number</u> of a finite set is defined as the number of elements of the set [13]. This concept can be extended to probabilistic set theory as follows.

[Def.3-15]

Let A be a probabilistic set on X whose defining function is $\mu_A(x,\omega)$. The <u>expected support</u> of A is defined as the following (ordinary) subset of X,

$$\text{supp } A = \{x \mid x \in X,\ E(\mu_A)(x) = \int_\Omega \mu_A(x,\omega)\, dP(\omega) > 0\}, \tag{3-220}$$

and the <u>expected cardinal number</u> of A, denoted by #A, is defined by

$$\#A = \begin{cases} \sum_{x \in \text{supp}A} \left(\int_\Omega \mu_A(x,\omega)\, dP(\omega)\right) & \text{if } \#\text{supp } A \leqq \chi_0 \\ \#\ \text{supp}A & \text{if } \#\text{supp } A > \chi_0 . \end{cases} \tag{3-221}$$

where #supp A stands for the cardinal number of supp A and χ_0 stands for the cardinal number of N(all natural numbers).

[Prop.3-15]

$$\#\ \mathcal{P}(X) = (\#\ M/\equiv)^{\#X}. \tag{3-222}$$

<u>proof</u>

It is clear from the definition of probabilistic sets and $\mathcal{P}(X)$.

(Q.E.D.)

Remark : In ordinary set theory, the cardinal number of all subsets of X is equal to $2^{\#X}$. However, in probabilistic set theory, the cardinal number of $\mathcal{P}(X)$ is given by (3-222), which is much larger than $2^{\#X}$.

3-9. RELATIONS

There was no restriction about the total space X in the definition of probabilistic sets, except that it was a set. Hence, we may regard the total space as a direct product of two sets. Such a substitution makes it possible to expand a notion of (ordinary) binary relations [14].

[Def.3-16]

Let a direct product $X\times Y=\{(x,y)\mid x\in X, y\in Y\}$ be the total space. A probabilistic set R on $X\times Y$ is called a <u>relation</u> on $X\times Y$,

$$\begin{aligned} \mu_R:\ (X\times Y)\times\Omega &\longrightarrow [0,1]. \\ \omega \qquad\qquad &\qquad\ \omega \\ ((x,y),\omega) &\longmapsto \mu_R((x,y),\omega) \end{aligned} \tag{3-223}$$

A relation on $X\times X$ is called a <u>binary relation</u> on X.

Remark : To clarify the description, the concrete meaning is given: If $\mu_R((x_1,x_2),\omega)=1$ for P a.e. $\omega\in\Omega$, then there exists the relation R between x_1 and x_2. Conversely, if $\mu_R((x_1,x_2),\omega)=0$ for P a.e. $\omega\in\Omega$, then there is no connection between x_1 and x_2 from a viewpoint of the relation R. All other cases indicate that the relation R between x_1 and x_2 is ambiguous. For arbitrary relations R_1 and R_2, several composite notions can be defined in terms of operations of probabilistic sets, such as inclusion $R_1\subset R_2$, equality $R_1=R_2$, union $R_1\cup R_2$, intersection $R_1\cap R_2$, and complement $R_1^{\,c}$.

[Def.3-17]

Let R be a binary relation on X. For arbitrary $x_1,x_2\in X$, x_1 and x_2 are said to have a relation R with grade $\alpha(\in[0,1])$ if

$$\mu_R((x_1,x_2),\omega)\geqq\alpha \text{ for P a.e. } \omega\in\Omega. \tag{3-224}$$

These notions can be successfully applied to, for example, clustering analysis [15][16].

3-10. CHARACTERIZATION OF PROBABILISTIC SETS AS A NON-CLASSICAL LOGIC.

A foundation of ordinary set theory is {0,1} two-valued logic. Two valued logic can be characterized as a complete Boolean algebra, and it is called a classical logic. The concept of probabilistic sets was characterized as a complete pseudo-Boolean algebra (Theorem 3-1), hence it can be regarded as a non-classical logic. There exist so many non-classical logics such as modal logic[17], many-valued logic [18], quantum logic [19], and fuzzy logic[20]. Among these various non-classical logics, fuzzy logic is very semilar to the concept of probabilistic sets, because 1) fuzzy logic also constitutes a complete pseudo-Boolean algebra, 2) fuzzy logic has been developed as a mathematical method to deal with ambiguity and subjectivity. We shall compare the concept of probabilistic sets with other non-classical logics, especially with various fuzzy concepts.

A (classical) fuzzy set A on a total space X is defined by a [0,1]-valued mapping on X, called the membership function of A [21]. If we restrict the notion of the family of characteristic variables M (Def. 3-7) to the (proper) sub class of all constant functions, then defining functions of probabilistic sets become [0, 1]-valued functions of single variable $x(\varepsilon X)$, and they can be identified with membership functions of fuzzy sets. Hence we can conclude that the notion of probabilistic sets include classical fuzzy concepts.

There is also an investigation of probabilistic measure of fuzzy events [22]. A distinction between the total space X and the parameter space (Ω,B,P) is very important in probabilistic set theory. The concept of probabilistic sets differs intrinsically from Zadeh's way of thinking [22] in this point.

A notion of fuzzy set of type 2 [23] is introduced in order to dissolve the difficulty of settling a definite ambiguous degree. Fuzzy set of type n is also characterized by n step recursively defined ambiguity. However the number of steps (i.e. n) has no upper bound and, to make matters worse, realistic meanings decline as n increases. In probabilistic sets, the ambiguity is arranged on the parameter space and realistic meanings are made clear in connection with subjectivity and personality of observers.

A family of probabilistic sets constitutes a complete pseudo-Boolean algebra (Theorem 3-1). A pseudo-Boolean algebra is a subclass of distributive lattices (cf. Fig.3-3). Hence, from a lattice theoretical viewpoint [24], probabilistic set theory takes its position between L-fuzzy set theory [25] and Boolean algebra valued set theory [26].

In probabilistic set theory, the parameter space (Ω,B,P) plays an important role, but it has no restriction except that it's a probabilistic measure space. The most important task in applications is a suitable choice of the parameter space, especially is an establishment of probability measure P. At last, we would like to add that there is no need to recollect probabilistic randomness like casting a dice in spite of a diction "probabilistic" sets.

CHAPTER. 4

EXTENDED FUZZY EXPRESSION OF PROBABILISTIC SETS.

4-1. INTRODUCTION.

The information about a probabilistic set A is expressed completely by a pair $(\mu_A(x,\omega), P(\omega))$, where $\mu_A(x,\omega)$ is the defining function of A, and $P(\omega)$ is the probability measure of the parameter space Ω. The pair $(\mu_A(x,\omega), P(\omega))$ will be called a "probabilistic expression" of the probabilistic set A.

The main purpose of the present chapter is to introduce another mutually equivalent expression of probabilistic sets, called an "extended fuzzy expression". The extended fuzzy expression of A is given by a countably infinite set of functions (called "monitors") $\{m_A^n(x)\}_{n=1}^{\infty}$. The moment analysis shows that the main information is concentrated on lower monitors such as $m_A^1(x)$ (called a "membership function" of A) and $v_A(x) \triangleq m_A^2(x) - (m_A^1(x))^2$ (called a "vagueness function" of A). It is sufficient practically to consider both the membership function and the vagueness function. From the possibility of two mutually equivalent expressions of probabilistic sets, we can draw an interesting conclusion to the "fuzzy vs probability" controversy: The classical fuzzy concept, i.e. the notion of membership functions alone introduced by L.A.Zadeh ('65), is not sufficient. However, if other concepts such as vagueness functions are taken into consideration in addition to the notion of membership functions, then the modified fuzzy concepts, called extended fuzzy expression, provide the same information as the notion of probability. Hence the equality between the notion of "(modified) fuzzy" and that of "probability" is confirmed theoretically. Moreover, by using membership functions, vagueness functions and higher monitiors successively, we can expect to obtain useful results which are different from the results given by probabilistic approaches.

We shall also mention about the extended fuzzy expression of plural probabilistic sets and the fundamental operations of probabilistic sets in terms of the extended fuzzy expression.

4-2. EXTENDED FUZZY EXPRESSION OF A SINGLE PROBABILISTIC SET.

A probabilistic set A on the total space X was defined by a defining function $\mu_A(x,\omega)$ from $X\times\Omega$ to $\Omega_c=[0,1]$ (Def.3-8); where $\mu_A(x,\cdot)$ was the (B,B_c)-measurable function for each fixed $x\varepsilon X$, and the probability space (Ω,B,P) was called the parameter space. Hence the probabilistic set A can be expressed completely by both $\mu_A(x,\omega)$ and $P(\omega)$ (exactly speaking, by both $\mu_A(x,\omega)$ and (Ω,B,P)). A pair $(\mu_A(x,\omega),P(\omega))$ will be called a probabilistic expression of the probabilistic set A. It is also possible to express the probabilistic set A by another expression, called an extended fuzzy expression which will be given in the following.

Consider the following induced probability measure $\Phi_A(x,\cdot)$ on (Ω_C,B_C) (=([0,1],Borel sets)) for each fixed $x(\varepsilon X)$,

$$\Phi_A(x,E) \underset{\Delta}{=} P(\{\omega|\mu_A(x,\omega)\varepsilon E\}) \quad \text{for all } E\varepsilon B_c. \tag{4-1}$$

If we do not consider the structure of Ω (since the important point is not each parameter $\omega\varepsilon\Omega$, but the evaluation value $\mu_A(x,\omega)$ in Ω_c), then we may consider that $\Phi_A(x,\cdot)$ has the same information as $(\mu_A(x,\omega),P(\omega))$. We can consider the following transformation from $\Phi_A(x,\cdot)$ to $\phi_A(x,t)$ (where $t\varepsilon R$(real numbers)) for each fixed $x\varepsilon X$,

$$\phi_A(x,t) \underset{\Delta}{=} \int_0^1 \exp(it\alpha)\cdot d\Phi_A(x,\alpha). \tag{4-2}$$

This transformation has an inverse (cf. Levy-Haviland inversion formula in probability theory [1]). Hence it can be concluded that $\phi_A(x,t)$ has the same information as $\Phi_A(x,\cdot)$. Moreover, the following propositions hold.

[Prop. 4-1]

$$\left.\frac{\partial^n \phi_A(x,t)}{\partial t^n}\right|_{t=0} = i^n \cdot M^n[\mu_A(x,\cdot)]. \tag{4-3}$$

<u>proof</u>

First, we shall give the proof in the case of n=1.

$$\lim_{h\to 0} \frac{\phi_A(x,t+h)-\phi_A(x,t)}{h}$$

$$= \lim_{h\to 0} \int_0^1 \frac{\exp(i(t+h)\alpha)-\exp(it\alpha)}{h}\, d\Phi_A(x,\alpha)$$

$$= \lim_{h\to 0} \int_0^1 \frac{\exp(ih\alpha)-1}{h} \exp(it\alpha)\ d\Phi_A(x,\alpha)$$

$$= \int_0^1 \lim_{h\to 0} \frac{\exp(ih\alpha)-1}{h} \exp(it\alpha)\ d\Phi_A(x,\alpha)$$

(cf. Lebesgue's dominated convergence theorem [2])

$$= \int_0^1 i\alpha \cdot \exp(it\alpha) \cdot d\Phi_A(x,\alpha). \tag{4-4}$$

$$\therefore \frac{\partial \phi_A(x,t)}{\partial t} = i\int_0^1 \alpha \cdot \exp(it\alpha) \cdot d\Phi_A(x,\alpha). \tag{4-5}$$

$$\lim_{t\to 0} \frac{\partial \phi_A(x,t)}{\partial t} = i\int_0^1 \lim_{t\to 0} \alpha \cdot \exp(it\alpha) \cdot d\Phi_A(x,\alpha).$$

(cf. Lebesgue's theorem [2])

$$= i \int_0^1 \alpha\ d\Phi_A(x,\alpha)$$

$$= i \int_\Omega \mu_A(x,\omega) \cdot dP(\omega)$$

$$= i \cdot E[\mu_A(x,\cdot)]. \tag{4-6}$$

$$\left.\frac{\partial \phi_A(x,t)}{\partial t}\right|_{t=0} = i \cdot E[\mu_A(x,\cdot)]. \tag{4-7}$$

The desired result (4-3) will be obtained in almost the same manner.

(Q.E.D.)

[Prop.4-2]

$$\phi_A(x,t) = \sum_{n=0}^{\infty} \frac{i^n}{n!} M^n[\mu_A(x,\cdot)]\, t^n. \tag{4-8}$$

proof

From a previous proposition 4-1 and Taylor's formula [3], we have

$$\phi_A(x,t) = \sum_{n=0}^{N} \frac{i^n}{n!} M^n[\mu_A(x,\cdot)]t^n + R_N(x,t), \tag{4-9}$$

where $R_N(x,t)$ is the remainder,

$$R_N(x,t) = \frac{1}{(N+1)!} \left. \frac{\partial^{(N+1)}\phi_A(x,s)}{\partial s^{(N+1)}} \right|_{s=\theta t} t^{(N+1)} \quad (0<\theta<1), \tag{4-10}$$

$$= \frac{(it)^{(N+1)}}{(N+1)!} \int_0^1 \alpha^{(N+1)} \cdot \exp(i\theta t\alpha) \cdot d\Phi_A(x,\alpha). \tag{4-11}$$

It can be shown that $R_N(x,t)$ uniformly tends to zero as N tends to infinity:

$$\begin{aligned} |R_N(x,t)| &\leqq \frac{|t|^{(N+1)}}{(N+1)!} \int_0^1 \alpha^{(N+1)} \cdot d\Phi_A(x,\alpha) \\ &\leqq \frac{|t|^{(N+1)}}{(N+1)!} \int_0^1 d\Phi_A(x,\alpha) \\ &= \frac{|t|^{(N+1)}}{(N+1)!} \\ &\leqq \frac{C^{(N+1)}}{(N+1)!} \qquad \text{for all } t\varepsilon(-C,C) \\ &\xrightarrow{N\to\infty} 0 \qquad \text{uniformly.} \end{aligned} \tag{4-12}$$

Hence we can conclude that $\phi_A(x,t)$ is an analytic function of t for each $x\varepsilon X$ and that (4-8) is valid.

(Q.E.D.)

[Prop.4-3]

Since each n-th moment $M^n[\mu_A(x,\cdot)]$ is a function of $x\varepsilon X$, we shall use the following notations,

$$m_A{}^n(x) \overline{\overline{\Delta}} M^n[\mu_A(x,\cdot)] \tag{4-13}$$

$$\phi_A(x,t) = \sum_{n=0}^{\infty} (i^n/n!) \cdot m_A{}^n(x) \cdot t^n. \tag{4-14}$$

Then we have

1) $n \geqq m \Longrightarrow 1 = m_A{}^0(x) \geqq m_A{}^m(x) \geqq m_A{}^n(x) \geqq 0.$ (4-15)

2) For each fixed $x \in X$, $\phi_A(x,t)$ (4-14) is a <u>positive definite function</u> of $t \in R$ in Bochner's sense [4], i.e. for each $n \in N$, $t_1, t_2, \ldots, t_n \in R$, $z_1, z_2, \ldots, z_n \in C$ (complex numbers), we have

$$\sum_{i,j=1}^{n} z_i \phi_A(x, t_i - t_j) \overline{z_j} = (\ldots z_i \ldots) \begin{pmatrix} & \vdots & \\ \ldots & \phi_A(x, t_i - t_j) & \ldots \\ & \vdots & \end{pmatrix} \overline{\begin{pmatrix} \vdots \\ z_j \\ \vdots \end{pmatrix}} \geq 0. \quad (4\text{-}16)$$

3) If we consider the following function $\psi_A(x,t)$,

$$\psi_A(x,t) \underset{\Delta}{=} \phi_A(x,-it), \quad (4\text{-}17)$$

then $\psi_A(x,t)$ is a monotone non-decreasing function of t for each fixed $x \in X$. ($\psi_A(x,t)$ is called a moment generating function [5].)

<u>proof</u>

1) It was already shown (cf. Prop.3-12 (3-197)).

2) Since we have

$$\phi_A(x,t) = \int_0^1 \exp(it\alpha) \cdot d\Phi_A(x,\alpha), \quad (4\text{-}18)$$

a direct calculation shows the desired result (4-16), i.e.

$$\sum_{i,j} z_i \phi_A(x, t_i - t_j) \overline{z_j} =$$

$$= \int_0^1 \sum_{i,j} z_i \exp(i(t_i - t_j)\alpha) \overline{z_j}\, d\Phi_A(x,\alpha)$$

$$= \int_0^1 \left| \sum_i z_i \cdot \exp(it_i\alpha) \right|^2 d\Phi_A(x,\alpha) \geq 0. \quad (4\text{-}19)$$

3) It is clear because we have

$$\psi_A(x,t) = \int_0^1 \exp(t\alpha) \cdot d\Phi_A(x,\alpha). \quad (4\text{-}20)$$

(Q.E.D.)

[Prop.4-4]

Let $\{m_A^n(x)\}_{n=1}^{\infty}$ be a countable set of functions which satisfy three conditions 1) 2) 3) mentioned in Prop.4-3. Then we can construct uniquely a probability measure $\Phi_A(x,\alpha)$ on (Ω_c, B_c) (=([0,1], Borel sets)).

proof

1) For the time being, $x(\varepsilon X)$ is arbitrarily fixed. Since, by a condition 2), $\phi_A(x,t)$ is continuous at $t=0$ and is a positive definite function of t for each fixed $x\varepsilon X$, there exists a measure $\Phi_A(x,\alpha)$ on R (not necessarily on [0,1]) and we have

$$\phi_A(x,t) = \int_{-\infty}^{\infty} \exp(it\alpha)\cdot d\Phi_A(x,\alpha), \quad (4\text{-}21)$$

(by using Bochner's theorem in Fourier transformation theory [6]). Moreover $\Phi_A(x,\alpha)$ becomes a probability measure, because we have

$$\phi_A(x,0) = m_A{}^0(x) = 1 = \int_{-\infty}^{\infty} d\Phi_A(x,\alpha), \quad (4\text{-}22)$$

by using (4-14), (4-15), (4-21).

2) The uniqueness of the probability measure $\Phi_A(x,\alpha)$ will be shown in the following. It is sufficient to show that from the premise of

$$\phi_A(x,t) = \int_{-\infty}^{\infty} \exp(it\alpha)\cdot d\Phi_A(x,\alpha)$$

$$= \int_{-\infty}^{\infty} \exp(it\alpha)\cdot d\Phi_A(x,\alpha), \quad (4\text{-}23)$$

we can conclude $\Phi_A(x,\alpha) = \Phi_A{}'(x,\alpha)$. Here we can define a probability measure $\Psi(\alpha)$ by

$$\Psi(\alpha) = \frac{1}{2}\,(\Phi_A(x,\alpha) + \Phi_A{}'(x,\alpha)). \quad (4\text{-}24)$$

Let D be a set of all atoms with respect to the probability measure Ψ,

$$D = \{a_n \mid \Psi(\{a_n\}) > 0,\ n=1,2,\ldots\}. \quad (4\text{-}25)$$

(It must be noted that the set D is at most countably infinite.) Hence we have

$$\Phi_A(x,\{a\})= \Phi_A'(x,\{a\})= 0, \qquad (4\text{-}24)$$

for each element a that doesn't belong to D. (Note that $\Psi(\{a\})= 0$ for such an element a.) Let's consider an interval $I=(a,b]$. Then we have in general,

$$\Phi_A(x,I^i)\leqq \Phi_A(x,I)\leqq \Phi_A(x,I^a), \qquad (4\text{-}25)$$

$$\Phi_A'(x,I^i)\leqq \Phi_A'(x,I)\leqq \Phi_A'(x,I^a). \qquad (4\text{-}26)$$

If we choose the end-points a,b from D^c, then we have by (4-24),(4-23) and Levy-Haviland inversion formula [1],

$$\Phi_A(x,I)= \Phi_A(x,I^i)= \Phi_A(x,I^a)= \Phi_A'(x,I^i)= \Phi_A'(x,I^a)= \Phi_A'(x,I) \qquad (4\text{-}27)$$

Let $I_0=(-\infty,a_0]$ be an arbitrarily fixed interval. We can choose a countably infinite interval-sequence $\{I_n=(a_n,b_n]\}_{n=1}^{\infty}$ (where $a_n,b_n\varepsilon D^c$) such that

$$I_n \nearrow I_0, \qquad (4\text{-}28)$$

because the set D^c is dense in R(all real numbers). Since we have by (4-27)

$$\Phi_A(x,I_n)= \Phi_A'(x,I_n) \quad \text{for all } n=1,2,... \qquad (4\text{-}29)$$

and since probability measures are monotone, we have

$$\Phi_A(x,I_0)= \Phi_A'(x,I_0). \qquad (4\text{-}30)$$

Hence we can obtain

$$\Phi_A(x,\alpha)= \Phi_A'(x,\alpha). \qquad (4\text{-}31)$$

3) Finally we shall give a proof of $\Phi_A(x,\alpha)$ being a probability measure on $[0,1]= \Omega_c$. Here all we have to show is

$$\Phi_A(x,[0,1])= 1. \qquad (4\text{-}32)$$

For an arbitrary number a(>1), consider the following set U_a,

$$U_a = (-\infty,-a) \cup (a,+\infty). \tag{4-33}$$

Considering the condition (4-15), we have

$$1 \geqq f_n(x) \searrow \geqq 0, \tag{4-34}$$

where

$$f_n(x) \overline{\overline{\Delta}} \int_{-\infty}^{\infty} \alpha^{2n} d\Phi_A(x,\alpha). \tag{4-35}$$

Since

$$f_n(x) \geqq \int_{U_a} \alpha^{2n} \cdot d\Phi_A(x,\alpha) \geqq a^{2n} \cdot \Phi_A(x,U_a), \tag{4-36}$$

we have

$$\Phi_A(x,U_a) = 0 \qquad \text{for all } a>1. \tag{4-37}$$

(If $\Phi_A(x,U_a) > 0$, then (4-36) and (4-34) cause a contradiction.) Considering that $\Phi_A(x,\alpha)$ is monotone and that $a(>1)$ can be arbitrarily chosen, we have

$$\Phi_A(x,U_1) = 0. \tag{4-38}$$

Hence we have

$$\Phi_A(x,[-1,1]) = 1. \tag{4-39}$$

Let a be an arbitrary, negative number. We shall prove that

$$\Phi_A(x,(-\infty,a)) = 0. \tag{4-40}$$

In fact, since we have

$$\begin{aligned} \Psi_A(x,t) &= \int_{-1}^{1} \exp(t\alpha) d\Phi_A(x,\alpha) \\ &\geqq \int_{-1}^{a} \exp(t\alpha) \cdot d\Phi_A(x,\alpha) \\ &\geqq \exp(ta)\ \Phi_A(x,(-\infty,a)) \qquad \text{for } t< 0, \end{aligned} \tag{4-41}$$

we can conclude (4-40). (If $\Phi_A(x,(-\infty,a)) > 0$, then the right side of (4-41) (hence $\Psi_A(x,t)$) tends to $+\infty$ as t tends to $-\infty$. This contradicts the third assumption, i.e. $\Psi_A(x,t)$ is a monotone non-decreasing function of t.) Since $\Phi_A(x,\alpha)$ is monotone, we have

$\Phi_A(x,(-\infty,0)) = 0.$ (4-42)

Considering (4-39) and (4-42), we can obtain the desired result (4-32).

(Q.E.D.)

The noteworthy point from above propositions is summarized as the following theorem 4-1.

[Theorem 4-1]

The probabilistic expression $(\mu_A(x,\omega),P(\omega))$ of the probabilistic set A has another equivalent expression $\{m_A{}^n(x)\}_{n=0}^{\infty}$ where

1) $n \geqq m \Rightarrow 1 = m_A{}^0(x) \geqq m_A{}^m(x) \geqq m_A{}^n(x) \geqq 0,$ (4-43)

2) For each fixed $x \in X$,

$$\phi_A(x,t) \underset{\Delta}{=} \sum_{n=0}^{\infty} (i^n/n!) \cdot m_A{}^n(x) \cdot t^n \tag{4-44}$$

is a positive definite function of $t \in R$.

3) For each fixed $x \in X$,

$$\psi_A(x,t) \underset{\Delta}{=} \phi_A(x,-it) = \sum_{n=0}^{\infty} (1/n!) \cdot m_A{}^n(x) \cdot t^n \tag{4-45}$$

is a monotone non-decreasing function of t.

[Def.4-1]

A set of countably infinite functions $\{m_A{}^n(x)\}_{n=0}^{\infty}$ with three conditions 1) 2) 3) in theorem 4-1 is called an <u>extended fuzzy expression</u> of the probabilistic set A. The function $m_A{}^n(x)$ is called a n-th <u>monitor</u> of A; especially $m_A{}^1(x)$ is called a <u>membership function</u> of A, and a function $v_A(x) \underset{\Delta}{=} m_A{}^2(x) - (m_A{}^1(x))^2$ is called a <u>vagueness function</u> of A.

Many discussions have been done on fuzzy concepts since L. A. Zadeh presented his paper [7]. Some anti-fuzzy scientists have been persisting that the fuzziness is not a new concept and that so-called fuzzy concepts can be derived by probability and two-valued logic. Whereas other fuzzy scientists have been insisting that the fuzziness is one thing and the (probabilistic) randomness is another. From the facts described above, we may conclude that,

1) There are two possible approaches to the problem of ambiguity and subjectivity, i.e. a probabilistic expression and an extended fuzzy expression.
2) The two approaches are mutually equivalent from a theoretical viewpoint. The most important point, however, is that there may exist several differences between the two approaches in applications. In probabilistic expression, the basic idea is to give a probability distribution, whereas, in extended fuzzy expression, it is to provide a membership function, a vagueness function and so on. Hence it may be possible to obtain different kinds of useful results according to each approach.
3) It was shown that although the notion of membership function (by L.A.Zadeh) provided an important information, it was not sufficient. We must consider a vagueness function and other higher monitors in order to get a good approximation. However, it was also shown that the important information was concentrated on lower monitors such as a membership function and a vagueness function.

4-3. EXTENDED FUZZY EXPRESSION OF PLURAL PROBABILISTIC SETS.

At first we shall deal with a case of two probabilistic sets. Let A and B be two probabilistic sets on X whose defining functions are $\mu_A(x,\omega)$ and $\mu_B(x,\omega)$ respectively. A triplet $(\mu_A(x,\omega),\mu_B(x,\omega),P(\omega))$ is called a probabilistic expression of A and B. We shall derive another expression, called an extended fuzzy expression of A and B, in the following.

We can introduce a probability measure $\Phi_{A,B}(x,\cdot)$ on $(\Omega_C\times\Omega_C, B_C\times B_C)$ $(=([0,1]^2,\text{Borel sets}))$ by

$$\Phi_{A,B}(x,E\times E') \underset{\Delta}{=} P(\{\omega|\mu_A(x,\omega)\varepsilon E\}\cap\{\omega|\mu_B(x,\omega)\varepsilon E'\}) \quad \text{for all } E,E'\varepsilon B_C. \tag{4-46}$$

This probability measure $\Phi_{A,B}(x,\cdot)$ provides the same information as the probabilistic expression. (The reason has already mentioned in the case of a single probabilistic set.) It is easy to see that moment relations are preserved,

$$\int_0^1\int_0^1 \alpha^n\beta^m\cdot d\Phi_{A,B}(x,(\alpha,\beta)) = \int_\Omega \mu_A(x,\omega)^n\cdot\mu_B(x,\omega)^m\cdot dP(\omega) \underset{\Delta}{=} E[\mu_A(x,\cdot)^n\mu_B(x,\cdot)^m]. \tag{4-47}$$

We can consider a transformation from $\Phi_{A,B}(x,\cdot)$ to $\phi_{A,B}(x,(s,t))$ by

$$\phi_{A,B}(x,(s,t)) \underset{\Delta}{=} \int_0^1\int_0^1 \exp(i(s\alpha+t\beta))\cdot d\Phi_{A,B}(x,(\alpha,\beta)). \tag{4-48}$$

[Prop.4-5]

The transformation (4-48) has an inverse, i.e.

$$\int_0^1\int_0^1 \chi(\alpha;a_1,a_2)\cdot\chi(\beta;b_1,b_2)\cdot d\Phi_{A,B}(x,(\alpha,\beta))$$

$$= \lim_{T\to+\infty} \left(\frac{1}{2\pi}\right)^2 \int_{-T}^{T}\int_{-T}^{T} \frac{e^{-ia_2s}-e^{-ia_1s}}{-is}\cdot\frac{e^{-ib_2t}-e^{-ib_1t}}{-it}\,\phi_{A,B}(x,(s,t))\cdot dsdt, \tag{4-49}$$

where

$$0 \leqq a_1 < a_2 \leqq 1, \quad 0 \leqq b_1 < b_2 \leqq 1$$

$$\chi(\alpha; a_1, a_2) = \begin{cases} 1 & a_1 < \alpha < a_2 \\ 1/2 & \alpha = a_1 \text{ or } a_2 \\ 0 & \text{otherwise.} \end{cases} \qquad (4\text{-}50)$$

<u>proof</u>

It will be sufficient to show a one dimensional case, i.e.

$$\int_0^1 \chi(\alpha; a_1, a_2) \cdot d\Phi(\alpha) = \lim_{T \to +\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ia_2 s} - e^{-ia_1 s}}{-is} \phi(s) ds. \quad (4\text{-}51)$$

(Here we use simplified notations.)

By applying the Fubini's theorem [8], we have

$$F(T) \underset{\Delta}{=} \int_{-T}^{T} \frac{e^{-ia_2 s} - e^{-ia_1 s}}{-is} \phi(s) ds$$

$$= \int_{-T}^{T} \left(\int_{a_1}^{a_2} e^{-ixs} dx\right) \left(\int_0^1 e^{i\alpha s} d\Phi(\alpha)\right) ds$$

$$= \int_0^1 d\Phi(\alpha) \int_{a_1}^{a_2} dx \int_{-T}^{T} e^{is(\alpha - x)} ds$$

$$= 2\int_0^1 d\Phi(\alpha) \int_{a_1}^{a_2} \frac{\sin T(x-\alpha)}{x-\alpha} dx$$

$$= 2\int_0^1 d\Phi(\alpha) \int_{T(a_1-\alpha)}^{T(a_2-\alpha)} \frac{\sin u}{u} du$$

$$= 2\int_0^1 \{G(T(a_2-\alpha)) - G(T(a_1-\alpha))\} \, d\Phi(\alpha), \qquad (4\text{-}52)$$

where $\quad G(x) = \int_0^x \frac{\sin u}{u} du. \qquad (4\text{-}53)$

Since $G(x)$ is continuous and since $\lim_{x \to \infty} G(x) = (\pi/2)$, $\lim_{x \to -\infty} G(x) = -(\pi/2)$, $G(x)$ is bounded and the right hand side of (4-52) is also bounded. Hence we can apply the Lebesgue's dominated convergence theorem [2], and we have

$$\lim_{T \to +\infty} F(T) = 2\int_0^1 \lim_{T \to \infty} \{G(T(a_2-\alpha)) - G(T(a_1-\alpha))\} \cdot d\Phi(\alpha)$$

$$= 2\pi \int_0^1 \chi(\alpha; a_1, a_2) \cdot d\Phi(\alpha). \qquad (4\text{-}54)$$

(Q.E.D.)

We shall give several comments. Let $F_{A,B}(x,(\alpha,\beta))$, defined by

$$F_{A,B}(x,(\alpha,\beta)) \underset{\Delta}{=} \Phi_{A,B}(x,(-\infty,\alpha]\times(-\infty,\beta]), \tag{4-55}$$

be a cumulative distribution function. Then we have

1) $F_{A,B}(x,(\alpha,\beta)) \nearrow$ with respect to α and β, (4-56)

2) $\lim_{\alpha,\beta\to-\infty} F_{A,B}(x,(\alpha,\beta)) = 0$, $\lim_{\alpha,\beta\to+\infty} F_{A,B}(x,(\alpha,\beta)) = 1$, (4-57)

3) $\lim_{h,k\to+0} F_{A,B}(x,(\alpha+h,\beta+k)) = F_{A,B}(x,(\alpha,\beta))$ (right continuous).(4-58)

Conversely, it is well-known [9] that a probability measure $\Phi_{A,B}(x,\cdot)$ can be uniquely constructed by $F_{A,B}(x,(\alpha,\beta))$ which satisfies (4-56),(4-57),(4-58). Therefore there exists an one to one correspondence between $\Phi_{A,B}(x,\cdot)$ and $F_{A,B}(x,(\alpha,\beta))$. If $\Phi_{A,B}(x,\cdot)$ is absolutely continuous and can be expressed by

$$\Phi_{A,B}(x,E) = \iint_E f_{A,B}(x,(\alpha,\beta))\cdot d\alpha d\beta \quad \text{for all } E\in B_C\times B_C, \tag{4-59}$$

where $f_{A,B}(x,(\alpha,\beta))$ is a non-negative, Baire function of α and β with an integral value 1, then $f_{A,B}(x,(\alpha,\beta))$ is said to be a <u>probability density function</u> (p.d.f.) of $\Phi_{A,B}(x,\cdot)$. If there exists the p.d.f. $f_{A,B}(x,(\alpha,\beta))$ of $\Phi_{A,B}(x,\cdot)$, then we have

$$F_{A,B}(x,(\alpha,\beta)) = \int_{-\infty}^{\alpha}\int_{-\infty}^{\beta} f_{A,B}(x,(\alpha,\beta))\cdot d\alpha d\beta. \tag{4-60}$$

[Prop. 4-6]

If $\phi_{A,B}(x,(s,t))$ belongs to $L^1(R^2)$ as a function of (s,t), then $\Phi_{A,B}(x,\cdot)$ has a continuous p.d.f. $f_{A,B}(x,(\alpha,\beta))$ and the following relation holds,

$$f_{A,B}(x,(\alpha,\beta)) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(\alpha s+\beta t)}\cdot\phi_{A,B}(x,(s,t))\cdot dsdt. \tag{4-61}$$

<u>proof</u>

We have in general

$$\Phi_{A,B}(x,(a_1,a_2)\times(b_1,b_2)) \leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\chi(\alpha;a_1,a_2)\cdot\chi(\beta;b_1,b_2)\cdot d\Phi_{A,B}(x,(\alpha,\beta))$$

$$\leqq \Phi_{A,B}(x,[a_1,a_2]\times[b_1,b_2]). \tag{4-62}$$

The sign of equality in (4-62) is valid, if $F_{A,B}(x,(\alpha,\beta))$ is continuous on the boundary of $\{(\alpha,\beta)\,|\,a_1\leqq \alpha\leqq a_2,\ b_1\leqq \beta\leqq b_2\}$. If we take a limit $h,k\to+0$, where positive numbers h and k are chosen such that $F_{A,B}(x,(\alpha,\beta))$ is continuous on the boundary of $\{(\alpha,\beta)\,|\,\alpha_0-h\leqq \alpha\leqq \alpha_0+h,\ \beta_0-k\leqq \beta\leqq \beta_0+k\}$ (this procedure is possible because of (4-56),(4-57),(4-58)), then we have

$$f_{A,B}(x,(\alpha_0,\beta_0))$$

$$= \lim_{h,k\to+0} \frac{F_{A,B}(x,(\alpha_0+h,\beta_0+k))-F_{A,B}(\alpha_0-h,\beta_0+k))-F_{A,B}(x,(\alpha_0+h,\beta_0-k))+F_{A,B}(\alpha_0-h,\beta_0-k))}{(2h)(2k)}$$

$$= \lim_{h,k\to+0} \frac{1}{4hk}\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{-i(\alpha_0+h)s}-e^{-i(\alpha_0-h)s}}{-is}\cdot \frac{e^{-i(\beta_0+k)t}-e^{-i(\beta_0-k)t}}{-it}\cdot\phi_{A,B}(x,(s,t))\cdot dsdt$$

$$= \lim_{h,k\to+0}\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i\alpha_0 s}\cdot\frac{\sin sh}{sh}\cdot e^{-i\beta_0 t}\cdot\frac{\sin tk}{tk}\cdot\phi_{A,B}(x,(s,t))\cdot dsdt$$

$$= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\lim_{h,k\to+0} e^{-i(\alpha_0 s+\beta_0 t)}\cdot\frac{\sin sh}{sh}\,\frac{\sin tk}{tk}\,\phi_{A,B}(x,(s,t))\cdot dsdt$$

$$= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(\alpha_0 s+\beta_0 t)}\phi_{A,B}(x,(s,t))\, dsdt. \tag{4-63}$$

(Here we used the Lebesgue's dominated convergence theorem [2] and the fact that $\phi_{A,B}(x,(s,t))$ was integrable.)

The continuity of $f_{A,B}(x,(\alpha,\beta))$ (i.e. $f_{A,B}(x,(\alpha+h,\beta+k))\to f_{A,B}(x,(\alpha,\beta))$ $(h,k\to0)$) is also confirmed by using the Lebesgue's dominated convergence theorem.

(Q.E.D.)

[Prop.4-7]

There exists a partial derivative of $\phi_{A,B}(x,(s,t))$ of arbitrary order with respect to s and t, and we have

$$\left.\frac{\partial^{n+m}\phi_{A,B}(x,(s,t))}{\partial s^n \partial t^m}\right|_{s=t=0} = i^{n+m}\, E[\mu_A(x,\cdot)^n \mu_B(x,\cdot)^m]. \quad (4\text{-}64)$$

proof

$$\frac{\partial\phi_{A,B}(x,(s,t))}{\partial s} = \lim_{h\to 0} \frac{\phi_{A,B}(x,(s+h,t)) - \phi_{A,B}(x,(s,t))}{h}$$

$$= \lim_{h\to 0} \iint \frac{e^{i((s+h)\alpha+t\beta)} - e^{i(s\alpha+t\beta)}}{h}\, d\Phi_{A,B}(x,(\alpha,\beta))$$

$$= \lim_{h\to 0} \iint \frac{e^{ih\alpha}-1}{h}\, e^{i(s\alpha+t\beta)} d\Phi_{A,B}(x,(\alpha,\beta))$$

$$= \iint \lim_{h\to 0} \frac{e^{ih\alpha}-1}{h}\, e^{i(s\alpha+t\beta)}\, d\Phi_{A,B}(x,(\alpha,\beta))$$

(cf. Lebesgue's dominated covergence theorem [2])

$$= i \iint \alpha\, e^{i(s\alpha+t\beta)} d\Phi_{A,B}(x,(\alpha,\beta)). \quad (4\text{-}65)$$

$$\lim_{s,t\to 0} \frac{\partial\phi_{A,B}(x,(s,t))}{\partial s} = i \iint \lim_{s,t\to 0} \alpha\, e^{i(s\alpha+t\beta)} d\Phi_{A,B}(x,(\alpha,\beta))$$

$$= i\iint \alpha\, d\Phi_{A,B}(x,(\alpha,\beta)) = i\cdot E[\mu_A(x,\cdot)]. \quad (4\text{-}66)$$

We can obtain (4-64) almost in the same manner.

(Q.E.D.)

[Prop.4-8]

$$\phi_{A,B}(x,(s,t)) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{r=0}^{n} i^n \binom{n}{r} E[\mu_A(x,\cdot)^r \mu_B(x,\cdot)^{n-r}] s^r t^{n-r}. \quad (4\text{-}67)$$

proof

By applying the Taylor's theorem of a function of two variables [10], we have

$$\phi_{A,B}(x,(s,t)) = \sum_{n=0}^{N} \frac{1}{n!} \sum_{r=0}^{n} i^n \binom{n}{r} E[\mu_A(x,\cdot)^r \mu_B(x,\cdot)^{n-r}] s^r t^{n-r} + R^N_{A,B}(x,(s,t)), \quad (4\text{-}68)$$

where the remainder $R^N_{A,B}(x,(s,t))$ is given by

$$R^N_{A,B}(x,(s,t)) = \frac{1}{(N+1)!} \sum_{r=0}^{N+1} \binom{N+1}{r} \left.\frac{\partial^{N+1}\phi_{A,B}(x,(s,t))}{\partial s^r\, \partial t^{N+1-r}}\right|_{\substack{s=\theta s\\ t=\theta t}} s^r t^{N+1-r}$$

$$0< \theta<1$$

$$= \frac{1}{(N+1)!} \sum_{r=0}^{N+1} \binom{N+1}{r} i^{N+1} \iint \alpha^r \beta^{N+1-r} e^{i\theta(s\alpha+t\beta)} d\Phi_{A,B}(x,(\alpha,\beta))\cdot \cdot s^r t^{N+1-r}. \quad (4\text{-}69)$$

$$|R^N_{A,B}(x,(s,t))| \leqq \frac{1}{(N+1)!} \sum_{r=0}^{N+1} \binom{N+1}{r} \iint |\alpha|^r |\beta|^{N+1-r} \cdot d\Phi_{A,B}(x,(\alpha,\beta))\, |s|^r |t|^{N+1-r}$$

$$= \frac{1}{(N+1)!} \sum_{r=0}^{N+1} \binom{N+1}{r} E[\mu_A(x,\cdot)^r \, \mu_B(x,\cdot)^{N+1-r}]\, |s|^r |t|^{N+1-r}$$

$$\leqq \frac{1}{(N+1)!} \sum_{r=0}^{N+1} \binom{N+1}{r} |s|^r |t|^{N+1-r}$$

$$= \frac{(2C)^{N+1}}{(N+1)!} \quad \text{for all } (s,t) \varepsilon \{(s,t) \mid s^2+t^2 \leqq C^2\}$$

$$\xrightarrow{N\to+\infty} 0 \quad \text{uniformly.} \tag{4-70}$$

Therefore, we have (4-67).

(Q.E.D.)

[Prop. 4-9]

We shall use the following notations in the same manner as the case of a single probabilistic set,

$$m^{n,m}_{A,B}(x) \underset{\Delta}{=} E[\mu_A{}^n(x,\cdot)\mu_B{}^m(x,\cdot)], \tag{4-71}$$

$$\phi_{A,B}(x,(s,t)) \underset{\Delta}{=} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{r=0}^{n} i^n \cdot \binom{n}{r} \cdot m^{r,n-r}_{A,B}(x) \cdot s^r \cdot t^{n-r}. \tag{4-72}$$

Then we have,

1) $1 = m^{0,0}_{A,B}(x) \geqq m^{n,m}_{A,B}(x) \geqq m^{n',m'}_{A,B}(x) \geqq 0 \quad (n' \geqq n,\ m' \geqq m),$ (4-73)

2) $\psi_{A,B}(x,(s,t)) \underset{\Delta}{=} \phi_{A,B}(x,(-is,-it)),$ (4-74)

is a monotone non-decreasing function of (s,t) for each $x(\varepsilon X)$.

3) For each fixed x, $\phi_{A,B}(x,(s,t))$ is a positive definite function of (s,t); i.e. for $n \varepsilon N$, $(s_1,t_1),\ldots,(s_n,t_n) \varepsilon R$, $z_1,\ldots,z_n$ εC, we have

$$\sum_{i,j=1}^{n} z_i \cdot \phi_{A,B}(x,(s_i-s_j,t_i-t_j)) \cdot \overline{z_j}$$

$$= (\ldots,z_i,\ldots) \begin{pmatrix} & \vdots & \\ \ldots, & \phi_{A,B}(x,(s_i-s_j,t_i-t_j)) & \ldots \\ & \vdots & \end{pmatrix} \overline{\begin{pmatrix} \vdots \\ z_j \\ \vdots \end{pmatrix}} \geqq 0. \tag{4-75}$$

<u>proof</u>

1) ,2) It is clear from the following equations,

$$m^{n,m}_{A,B}(x)= \int_\Omega \mu_A(x,\omega)^n \mu_B(x,\omega)^m dP(\omega), \tag{4-76}$$

$$\psi_{A,B}(x,(s,t))= \int_0^1\int_0^1 \exp(s\alpha+t\beta)\,d\Phi_{A,B}(x,(\alpha,\beta)). \tag{4-77}$$

3) Since

$$\phi_{A,B}(x(s,t))= \int_0^1\int_0^1 \exp(i(s\alpha+t\beta))\,d\Phi_{A,B}(x,(\alpha,\beta)), \tag{4-78}$$

we have

$$\sum_{i,j} z_i\phi_{A,B}(x,(s_i-s_j,t_i-t_j))\overline{z_j}$$

$$= \iint \sum_{i,j} z_i\exp(i((s_i-s_j)\alpha+(t_i-t_j)\beta)\cdot\overline{z_j}\; d\Phi_{A,B}(x,(\alpha,\beta))$$

$$= \iint \sum_{i,j} z_i\cdot\exp(i(s_i\alpha+t_i\beta))\cdot\exp(-i(s_j\alpha+t_j\beta))\cdot\overline{z_j}\; d\Phi_{A,B}(x,(\alpha,\beta))$$

$$= \iint \left|\sum_i z_i\cdot\exp(i(s_i\alpha+t_i\beta))\right|^2 d\Phi_{A,B}(x,(\alpha,\beta)) \geqq 0. \tag{4-79}$$

(Q.E.D.)

[Prop.4-10]

Let $\{m^{n,m}_{A,B}(x)\}^{\infty}_{n,m=0}$ be a countably infinite function matrix with three properties 1),2),3) described in Prop.4-11. Then we can constitute uniquely a probability measure $\Phi_{A,B}(x,\cdot)$ on $(\Omega_c\times\Omega_c, B_c\times B_c)$ $(=([0,1]^2,\text{Borel sets}))$.

<u>proof</u>

1) Since $\phi_{A,B}(x,(s,t))$ defined by (4-72) is continuous at $(s,t)=(0,0)$ for each x and is a positive definite function, we can define a measure $\Phi_{A,B}(x,\cdot)$ on R^2 (cf. the Bochner's theorem [6]) and we have

$$\phi_{A,B}(x,(s,t))= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(s\alpha+t\beta)}\,d\Phi_{A,B}(x,(\alpha,\beta)). \tag{4-80}$$

Moreover, the measure $\Phi_{A,B}(x,\cdot)$ becomes a probability measure on R^2, since we have

$$\phi_{A,B}(x,(0,0))= m^{0,0}_{A,B}(x)= 1= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\Phi_{A,B}(x,(\alpha,\beta)). \tag{4-81}$$

2) We shall prove the uniqueness of $\Phi_{A,B}(x,\cdot)$. It is sufficient

to show $\Phi_{A,B}(x,\cdot) = \Phi'_{A,B}(x,\cdot)$ from the assumption of

$$\phi_{A,B}(x,(s,t)) = \iint \exp(i(s\alpha+t\beta))\cdot d\Phi_{A,B}(x,(\alpha,\beta))$$
$$= \iint \exp(i(s\alpha+t\beta))\cdot d\Phi'_{A,B}(x,(\alpha,\beta)). \qquad (4\text{-}82)$$

Consider the following probability measure $\Psi(\cdot)$ on R^2,

$$\Psi(\cdot) = \frac{1}{2}\{\Phi_{A,B}(x,\cdot)+\Phi'_{A,B}(x,\cdot)\}, \qquad (4\text{-}83)$$

and the following hyperplanes on R^2 (i.e. lines)

$$H_1(a) = \{(a,\beta) \mid \beta\varepsilon(-\infty,\infty)\}, \qquad (4\text{-}84)$$

$$H_2(a) = \{(\alpha,a) \mid \alpha\varepsilon(-\infty,\infty)\}. \qquad (4\text{-}85)$$

Since a set of a's such that

$$\Psi(H_i(a)) > 0 \quad i=1 \text{ or } 2, \qquad (4\text{-}86)$$

is at most countable, we shall write it by

$$D \underset{\Delta}{=} \{a_{i,n} \mid i=1,2,\ n=1,2,\ldots.\}. \qquad (4\text{-}87)$$

Then we have

$$\Phi_{A,B}(x,H_i(a)) = \Phi'_{A,B}(x,H_i(a)) = 0 \quad i=1,2,\ a\varepsilon D^c. \qquad (4\text{-}88)$$

If we choose an interval I,

$$I = \{(\alpha,\beta) \mid a_1 < \alpha \leqq a_2,\ b_1 < \beta \leqq b_2\}, \qquad (4\text{-}89)$$

such that $a_1, a_2, b_1, b_2 \varepsilon D^c$, then we have, by (4-82) and the Levy-Haviland inversion formula [1],

$$\Phi_{A,B}(x,I) = \Phi_{A,B}(x,I^a) = \Phi_{A,B}(x,I^i) = \Phi'_{A,B}(x,I^a) = \Phi'_{A,B}(x,I^i)$$
$$= \Phi'_{A,B}(x,I). \qquad (4\text{-}90)$$

Let an interval I_0 be

$$I_0 \underset{\Delta}{=} (-\infty,\alpha_0]\times(-\infty,\beta_0], \qquad (4\text{-}91)$$

where $(\alpha_0,\beta_0)(\varepsilon R^2)$ is arbitrarily fixed. Then we have

$$\Phi_{A,B}(x,I_0) = \Phi'_{A,B}(x,I_0). \qquad (4\text{-}92)$$

(Since D^c is dense, we can select a countable sequence of intervals $\{I_n\}_{n=1}^{\infty}$ such that

$$I_n = (\alpha_n^1,\alpha_n^2]\times(\beta_n^1,\beta_n^2] \quad \text{where } \alpha_n^1,\alpha_n^2,\beta_n^1,\beta_n^2 \varepsilon D^c, \qquad (4\text{-}93)$$

$$I_n \nearrow I_0. \qquad (4\text{-}94)$$

Here,

$$\Phi_{A,B}(x,I_n) = \Phi'_{A,B}(x,I_n) \quad n=1,2,\ldots\ldots , \tag{4-95}$$

is valid by (4-90). Considering the monotone-property of probability measures, we have (4-92).)

Therefore, we have $\Phi_{A,B}(x,\cdot) = \Phi'_{A,B}(x,\cdot)$. (Note the abovementioned remark that there exists an one to one correspondence between the cumulative distribution function and the probability measure.)

3) Finally, we shall prove that $\Phi_{A,B}(x,\cdot)$ is a probability measure on $[0,1]^2$, i.e.

$$\Phi_{A,B}(x,[0,1]\times[0,1]) = 1. \tag{4-96}$$

Let a be an arbitrary, positive number which is greater than 1, and put

$$U_a \triangleq (-\infty,-a)\times(-\infty,\infty) \cup (a,\infty)\times(-\infty,\infty). \tag{4-97}$$

Since the following equations hold,

$$\begin{aligned} m_{A,B}^{2n,0}(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\alpha^{2n}\cdot d\Phi_{A,B}(x,(\alpha,\beta)) \\ &\geqq \iint_{U_a} \alpha^{2n}\cdot d\Phi_{A,B}(x,(\alpha,\beta)) \\ &\geqq a^{2n}\cdot\Phi_{A,B}(x,U_a), \end{aligned} \tag{4-98}$$

$$1 \geqq m_{A,B}^{2n,0}(x) \searrow \geqq 0, \tag{4-99}$$

we have

$$\Phi_{A,B}(x,U_a) = 0 \quad \text{for all } a>1. \tag{4-100}$$

Considering the fact that $\Phi_{A,B}(x,\cdot)$ is monotone, we have

$$\Phi_{A,B}(x,U_1) = 0. \tag{4-101}$$

We can also verify that

$$\Phi_{A,B}(x,V_1) = 0, \tag{4-102}$$

where

$$V_1 = (-\infty,\infty)\times(-\infty,-1) \cup (-\infty,\infty)\times(1,\infty), \tag{4-103}$$

in almost the same manner as (4-101). Hence, from (4-101) and (4-102), we have

$$\Phi_{A,B}(x,[-1,1]\times[-1,1])= 1, \tag{4-104}$$

i.e. $\Phi_{A,B}(x,\cdot)$ becomes a probability measure on $[-1,1]^2$.

In the next place, put

$$U_a= \{(\alpha,\beta) | \ \alpha< a, \ \beta\varepsilon \ (-\infty,\infty)\} \quad \text{where } a\varepsilon[-1,0). \tag{4-105}$$

If we substitute t=0 in (4-74) and (4-80), then we have

$$\psi_{A,B}(x,(s,0))= \phi_{A,B}(x,(-is,0))$$

$$= \int_{-1}^{1}\int_{-1}^{1} \exp(s\alpha)\cdot d\Phi_{A,B}(x,(\alpha,\beta))$$

$$\geqq \int_{-1}^{a}\int_{-1}^{1} \exp(s\alpha)\cdot d\Phi_{A,B}(x,(\alpha,\beta))$$

$$\geqq \exp(s\alpha)\cdot\Phi_{A,B}(x,U_a) \quad \text{for } s< 0. \tag{4-106}$$

If $\Phi_{A,B}(x,U_a)> 0$, then the right hand side of (4-106) tends to $+\infty$ as s tends to $-\infty$. This contradicts the monotone-assumption of $\psi_{A,B}(x,(s,t))$. Hence we have

$$\Phi_{A,B}(x,U_a)= 0 \quad \text{for all } a\varepsilon[-1,0). \tag{4-107}$$

Since the measure $\Phi_{A,B}(x,\cdot)$ is monotone, we have

$$\Phi_{A,B}(x,U_0)= 0. \tag{4-108}$$

We can also conclude that

$$\Phi_{A,B}(x,V_0)= 0, \tag{4-109}$$

where

$$V_0= \{(\alpha,\beta) | \ \alpha \ \varepsilon(-\infty,\infty), \ \beta< 0\} . \tag{4-110}$$

Considering (4-108) and (4-109), we can obtain the desired result

$$\Phi_{A,B}(x,[0,1]\times[0,1])= 1, \tag{4-111}$$

i.e. $\Phi_{A,B}(x,\cdot)$ is a probability measure on $[0,1]^2$.

(Q.E.D.)

With these propositions, the following theorem is demonstrated.

[Theorem 4-2]

The probabilistic expression $(\mu_A(x,\omega), \mu_B(x,\omega), P(\omega))$ of two probabilistic sets A and B has another equivalent expression $\{m_{A,B}^{n,m}(x)\}_{n,m=0}^{\infty}$ where

1) $1 = m_{A,B}^{0,0}(x) \geqq m_{A,B}^{n,m}(x) \geqq m_{A,B}^{n',m'}(x) \geqq 0 \quad (n' \geqq n,\ m' \geqq m),$ (4-112)

2) For each fixed $x \in X$,

$$\phi_{A,B}(x,(s,t)) \overline{\overline{\Delta}} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{r=0}^{n} i^n \cdot \binom{n}{r} \cdot m_{A,B}^{r,n-r}(x) \cdot s^r \cdot t^{n-r}, \tag{4-113}$$

is a positive definite function of (s,t).

3) For each fixed $x \in X$,

$$\psi_{A,B}(x,(s,t)) \overline{\overline{\Delta}} \phi_{A,B}(x,(-is,-it)), \tag{4-114}$$

is a monotone non-decreasing function of (s,t).

[Def.4-2]

A countably infinite function-matrix $\{m_{A,B}^{n,m}(x)\}_{n,m=o}^{\infty}$ with three conditions 1)2)3) in theorem 4-2 is called an extended fuzzy expression of the probabilistic sets A and B. The function $m_{A,B}^{n,m}(x)$ is called a (n,m)-th monitor of A and B; especially, $m_A(x) \overline{\overline{\Delta}} m_{A,B}^{1,0}(x)$ is called a membership function of A, $m_B(x) \overline{\overline{\Delta}} m_{A,B}^{0,1}(x)$ is called a membership function of B, $v_A(x) \overline{\overline{\Delta}} m_{A,B}^{2,0}(x) - (m_{A,B}^{1,0}(x))^2$ is called a vagueness function of A, $v_B(x) \overline{\overline{\Delta}} m_{A,B}^{0,2}(x) - (m_{A,B}^{0,1}(x))^2$ is called a vagueness function of B, $v_{A,B}(x) \overline{\overline{\Delta}} m_{A,B}^{1,1}(x) - m_{A,B}^{1,0}(x) \cdot m_{A,B}^{0,1}(x)$ is called a co-vagueness function of A and B.

In the extended fuzzy expression of A and B, the important information is concentrated on lower monitors. In most applications, it will be sufficient practically to give the 1-st and the

2-nd monitors, i.e.

$$\begin{pmatrix} 1 & m_A(x) & v_A(x) \\ m_B(x) & v_{A,B}(x) & \times \\ v_B(x) & \times & \times \end{pmatrix}. \tag{4-111}$$

We have been discussing about the extended fuzzy expression of two probabilistic sets. It will be easily verified to expand this notion into the case of more than two (in general, n) probabilistic sets. Moreover it is possible to develop this extended fuzzy expression to the case of infinite (possibly non-countably infinite) probabilistic sets: Let $\{A_\gamma\}_{\gamma\varepsilon\Gamma}$ be a family of probabilistic sets (possibly infinite), whose probabilistic expression is given by $(\{\mu_{A_\gamma}(x,\omega)\}_{\gamma\varepsilon\Gamma},P(\omega))$. Then its extended fuzzy expression is given as follows. Let $I=\{\gamma_1,\gamma_2,\ldots,\gamma_m\}$ be an arbitrary, finite subset of Γ. An extended fuzzy expression of $\{A_\gamma\}_{\gamma\varepsilon I}$ is given by a set of monitors $\{m^{n_1,n_2,\ldots,n_m}_{A_{\gamma_1},A_{\gamma_2},\ldots,A_{\gamma_m}}(x)\}^{\infty}_{n_1,n_2,\ldots,n_m=0}$.

Consider a class of such sets of monitors for all finite subsets I's of Γ. Then the class is called an extended fuzzy expression of $\{A_\gamma\}_{\gamma\varepsilon\Gamma}$.

4.4 OPERATIONS OF PROBABILISTIC SETS IN TERMS OF EXTENDED FUZZY EXPRESSION.

We defined several operations of probabilistic sets such as union (3-63), intersection (3-90) and λ-sum [Def.3-9]. However, all of these operations were defined in terms of defining functions. It is also possible to define these operations in terms of monitors.

Let $\{m_{A,B}^{n,m}(x)\}_{n,m=0}^{\infty}$ be a set of monitors of two probabilistic sets A and B. Then, for each $x \varepsilon X$, we can constitute uniquely a probability measure $\Phi_{A,B}(x,\cdot)$ on $[0,1]$ by Prop 4-10. By using this probability measure, we can define monitors of various binary operations of A and B as follows:

1) the union of A and B ($A \cup B$)

For each fixed $x \varepsilon X$, we can constitute a probability measure $\Phi_{A \cup B}(x,\cdot)$ on ([0,1], Borel sets) by

$$\Phi_{A \cup B}(x,(a,b)) \underset{\Delta}{=} \Phi_{A,B}(x,E), \tag{4-112}$$

$$\text{where } E = \{(\alpha,\beta) \mid a < \max(\alpha,\beta) < b,\ \alpha,\beta \ \varepsilon [0,1]\}. \tag{4-113}$$

Then the n-th monitor is given by

$$m_{A \cup B}{}^{n}(x) = \int_0^1 \alpha^n \cdot d\Phi_{A \cup B}(x,\alpha). \tag{4-114}$$

2) the intersection of A and B ($A \cap B$)

$$\Phi_{A \cap B}(x,(a,b)) \underset{\Delta}{=} \Phi_{A,B}(x,E), \tag{4-115}$$

$$\text{where } E = \{(\alpha,\beta) \mid a < \min(\alpha,\beta) < b,\ \alpha,\beta \ \varepsilon [0,1]\}. \tag{4-116}$$

$$m_{A \cap B}{}^{n}(x) = \int_0^1 \alpha^n \cdot d\Phi_{A \cap B}(x,\alpha). \tag{4-117}$$

3) the λ-sum of A and B ($A \underset{\lambda}{+} B$)

$$\Phi_{A \underset{\lambda}{+} B}(x,(a,b)) \underset{\Delta}{=} \Phi_{A,B}(x,E), \tag{4-118}$$

$$\text{where } E = \{(\alpha,\beta) \mid a < \lambda\alpha + (1-\lambda)\beta < b,\ \alpha,\beta \ \varepsilon [0,1]\}. \tag{4-119}$$

$$m_{A \underset{\lambda}{+} B}{}^{n}(x) = \int_0^1 \alpha^n \cdot d\Phi_{A \underset{\lambda}{+} B}(x,\alpha). \tag{4-120}$$

4) the algebraic sum of A and B $(A \oplus B)$

$$\Phi_{A\oplus B}(x,(a,b)) \underset{\Delta}{=} \Phi_{A,B}(x,E), \tag{4-121}$$

where $E = \{(\alpha,\beta) \mid a < \alpha+\beta-\alpha\beta < b,\ \alpha,\beta \varepsilon [0,1]\}$. (4-122)

$$m_{A\oplus B}{}^{n}(x) = \int_{0}^{1} \alpha^{n} \cdot d\Phi_{A\oplus B}(x,\alpha). \tag{4-123}$$

5) the algebraic product of A and B $(A \cdot B)$

$$\Phi_{A\cdot B}(x,(a,b)) \underset{\Delta}{=} \Phi_{A,B}(x,E), \tag{4-124}$$

where $E = \{(\alpha,\beta) \mid a < \alpha\beta < b,\ \alpha,\beta \varepsilon [0,1]\}$. (4-125)

$$m_{A\cdot B}{}^{n}(x) = \int_{0}^{1} \alpha^{n} \cdot d\Phi_{A\cdot B}(x,\alpha). \tag{4-126}$$

Other operations can be defined in almost the same manner.

Youth lives on hope, old age on remembrance.

CHAPTER. 5

AMBIGUITY BASED ON SUBJECTIVE ENTROPY.

5-1. INTRODUCTION.

The problems of ambiguity and subjectivity were visualized in chapter 2, and a new approach, called probabilistic set theory, was also proposed. There are two mutually equivalent expressions of probabilistic sets, i.e. probabilistic expression and extended fuzzy expression.

However, there exists another general approach to ambiguity, i.e. the concept of Shannon's entropy. In this chapter, probabilistic sets are investigated from a viewpoint of Shannon's entropy. It is summarized as a notion of subjective entropy, and the mutual relationship of probability, fuzzy theory and entropy is also investigated.

The important points are summarized as follows:

1) the subjective entropy can be expressed by using both membership function of fuzzy sets and probability,
2) the notion of subjective entropy is superior to classical fuzzy theory,
3) the concept of subjective entropy and that of vagueness function (of extended fuzzy expression) are independent each other, hence the notion of subjective entropy is properly included in that of probabilistic sets.

In the first place, we shall deal with subjective entropy of a single probabilistic set, and after that a case of plural probabilistic sets is mentioned. Lastly, a result is reported on a questionnaire about a decipherment of handwritten characters.

5-2. THE SIMPLEST STRUCTURE OF PROBABILISTIC SETS.

Let X={x} be a total space, i.e. a set of objects we want to discuss. A probabilistic set A on X is defined by a defining function μ_A,

$$\begin{array}{rcl} \mu_A: X\times\Omega & \longrightarrow & [0,1], \\ \omega & & \omega \\ (x,\omega) & \longmapsto & \mu_A(x,\omega) \end{array} \qquad (5\text{-}1)$$

where (Ω,B,P) is a probability space, called a parameter space (cf. Def. 3-8). The pair $(\mu_A(x,\omega),P(\omega))$ is called a probabilistic expression of A. There also exists another mutually equivalent definition of probabilistic sets, i.e. an extended fuzzy expression. It consists of a countable family of monitors $\{m_A{}^n(x)\}_{n=1}^{\infty}$ (cf. chap. 4).

To clarify the description, an example is given. Let $X=\{x_1, x_2, x_3\}$ be a set of handwritten characters as shown in Fig.5-1, and A be a probabilistic set of "characters which look like an alphabetical character A". Here, the parameter space Ω is a set of eight testees $\omega_1,\omega_2,\ldots,\omega_8$, and a uniform probability measure is assumed;

$$\Omega=\{\omega_1,\omega_2,\ldots,\omega_8\}, \qquad (5\text{-}2)$$

$$P(\{\omega_i\})=1/8. \qquad (5\text{-}3)$$

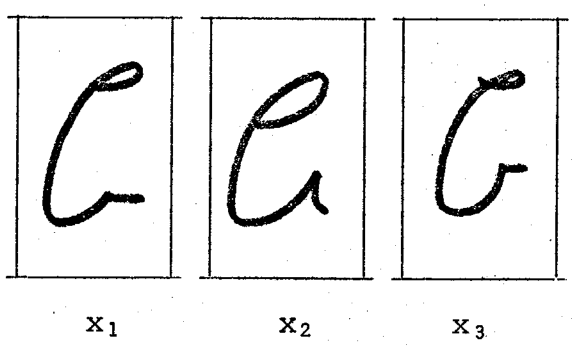


Fig.5-1. Ambiguous characters. (an example of a total space X.)

Consider the case that each character x_i is presented to all testees, and that each testee is supposed to make a five level answer whether the presented character looks like A or not. The defining function $\mu_A(x_i,\omega_j)$ is characterized as ω_j's answer to the question whether x_i is A. (The value range from 0(="no") to 1(="yes") with 0.25 in between.) The result is shown, for example, as in Table 5-1. From a viewpoint of the extended fuzzy expression of A, lower monitors are also shown in Table 5-1; i.e. the membership function $m_A(x_i)$ and the vagueness function $v_A(x_i)$. (Note that the membership function and the vagueness function correspond to a mean-value and a variance in probabilistic expression, respectively.)

x_i	$\mu_A(x_i,\omega_j)$								$m_A(x_i)$	$v_A(x_i)$	$H(x_i,A)$
	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8			
x_1	0.	0.5	0.5	0.25	0.5	0.75	0.5	1.	0.5	0.08	2.70
x_2	1.	0.	1.	0.	1.	1.	0.	0.	0.5	0.25	1.
x_3	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.	1.

Table 5-1. A numerical example of a probabilistic set A.

Since we have,

$$m_A(x_1) = m_A(x_2) = m_A(x_3) = 0.5, \tag{5-4}$$

there are no differences among three characters from a viewpoint of the membershipfunction. Whereas the vagueness function $v_A(x_i)$ has a maximum 0.25 at x_2. But the evaluation value $\mu_A(x_2,\omega_j)$ is either 0 (by four testees) or 1 (by also four testees), so the structure of x_2 from a viewpoint of the probabilistic set A is simple compared with that of x_1. (Note that $\mu_A(x_1,\omega_j)$ attains five different values from 0 to 1 with 0.25 in between.) It will be natural to consider that x_1 has a very complicated structure, since five different standards of judgements are taken part in the answers given by testees. But we cannot guess this interesting fact directly by using a membership function nor a vagueness function. Hence we would like to introduce a new measure which indicated a degree of complexity of probabilistic sets.

With regards to this point, it is necessary for us to make clear the notion of "complexity" (or conversely "simplicity") of probabilistic sets. We have already defined a binary operation of λ-sum $A \underset{\lambda}{+} B$ of two probabilistic sets A and B by

$$\mu_{A \underset{\lambda}{+} B}(x,\omega) = \lambda\mu_A(x,\omega) + (1-\lambda)\mu_B(x,\omega) \quad (0 \leq \lambda \leq 1), \tag{5-5}$$

(cf. (3-97)). The operation λ-sum was defined by introducing information of A and B in the ratio $\lambda:(1-\lambda)$. If we reconsider this operation and we assume that an arbitrarily given probabilistic set is decomposed into two probabilistic sets as in the right side of (5-5), then the two probabilistic sets generated by the decomposition may be considered to have a simpler structure than the original, given probabilistic set. We would like to define, as might be suspected, the simplest structure of probabilistic sets as the structure of probabilistic sets that could not be decomposed into more simplified one. We shall define this fact as a "pure state" as follows.

[Def. 5-1]

An object $x(\varepsilon X)$ is said to be in a <u>pure state</u> with respect to a probabilistic set A (on X), if there exists a (0,1)-positive number λ and the following statement is satisfied; From the assumption of

$$\mu_A(x,\omega) = \lambda\mu_B(x,\omega) + (1-\lambda)\mu_C(x,\omega) \quad \text{for P a.e. } \omega\varepsilon\Omega, \tag{5-6}$$

we can conclude that,

$$\mu_A(x,\omega) = \mu_B(x,\omega) = \mu_C(x,\omega) \quad \text{for P a.e. } \omega\varepsilon\Omega. \tag{5-7}$$

The pure state is exactly characterized as a limiting state that can not be decomposed into more detailed one. Another representation in the following theorem allows the facts to be read easily.

[Theorem 5-1]

A necessary and sufficient condition of $x(\varepsilon X)$ being in a pure state with respect to A is that the value of the defining function is either 0 or 1, i.e.

$$\mu_A(x,\omega) \;\varepsilon\{0,1\} \text{ for P a.e. } \omega\varepsilon\Omega. \tag{5-8}$$

proof

1) First we show (5-8) from (5-6) and (5-7) by a contrapositive. If (5-8) is invalid, then we have

$$\int_\Omega \mu_A(x,\omega)\,dP(\omega) = m \quad \text{where} \quad 0< m< 1, \tag{5-9}$$

$$\Omega_0 = \{\omega \,|\, \mu_A(x,\omega) \leqq m\}, \quad P(\Omega_0) > 0. \tag{5-10}$$

Consider the case of $m \geqq (1/2)$ and put

$$\mu_B(x,\omega) = 1_{\Omega_0{}^c}(\omega)\cdot\mu_A(x,\omega) + (1/m)\cdot 1_{\Omega_0}(\omega)\cdot\mu_A(x,\omega), \tag{5-11}$$

$$\mu_C(x,\omega) = 1_{\Omega_0{}^c}(\omega)\cdot\mu_A(x,\omega) + (2-(1/m))\cdot 1_{\Omega_0}(\omega)\cdot\mu_A(x,\omega), \tag{5-12}$$

where $1_S(\omega)$ is a characteristic function of an ordinary set S of Ω, i.e.

$$1_S(\omega) = \begin{cases} 1 & \omega \;\varepsilon\; S \\ 0 & \omega \;\varepsilon\; S^c. \end{cases} \tag{5-13}$$

Then we have

$$\mu_B(x,\omega) \neq \mu_C(x,\omega) \qquad \text{on} \quad \Omega_0, \tag{5-14}$$

$$\mu_A(x,\omega) = (1/2)\cdot\mu_B(x,\omega) + (1-(1/2))\cdot\mu_C(x,\omega). \tag{5-15}$$

Hence we can conclude that x is not in a pure state with respect to A. It will also be easily verified in the case of m<(1/2).

2) In the next place, we derive (5-7) under assumptions of (5-8) and (5-6). Put

$$\Omega \underset{\Delta}{=} \Omega_0 \cup \Omega_1 = \{\omega \mid \mu_A(x,\omega)=0\} \cup \{\omega \mid \mu_A(x,\omega)=1\}, \quad (5\text{-}16)$$

$$\mu_A(x,\omega) = \lambda\mu_B(x,\omega)+(1-\lambda)\mu_C(x,\omega) \quad (0< \lambda< 1). \quad (5\text{-}17)$$

Then we have, for each element ω of Ω_0,

$$0= \lambda\mu_B(x,\omega) + (1-\lambda)\mu_C(x,\omega), \quad (5\text{-}18)$$

$$\therefore\ 0= \mu_A(x,\omega)= \mu_B(x,\omega)= \mu_C(x,\omega). \quad (5\text{-}19)$$

For each element ω of Ω_1, we have also

$$1= \mu_A(x,\omega)= \mu_B(x,\omega)= \mu_C(x,\omega), \quad (5\text{-}20)$$

in almost the same manner.

(Q.E.D.)

Let us refer to the first example of Table 5-1. It will be easily confirmed that the character x_2 is in a pure state with respect to A. In the case of x_1, the answer $\mu_A(x_1,\cdot)$ varies in value from 0 to 1 with 0.25 in between. Here the value $\mu_A(x_1,\omega_4)= 0.25$, for example, may be expressed by both {0,1}-two values and a notion of probability, if we make a decomposition of (5-6) by meddling in ω_4's state of mind. But the discussion of meddling in a state of mind is practically impossible and it does not mean anything. The important point is not a state of mind but an evaluation value of $\mu_A(x_i,\omega_j)$ itself. This way of thinking has developed into the theory of probabilistic sets.

5-3. SUBJECTIVE ENTROPY OF PROBABILISTIC SETS IN FINITE STATES.

In this section, we introduce a measure which expresses the complexity of probabilistic sets based on a concept of entropy by C.E.Shannon [1].

Before going into the main argument, the necessary fundamentals will be explained first. Let p be an incidence probability of a certain phenomenon. Then the entropy of the phenomenon, as is well known, is defined by Shannon as follows,

$$h(p) = -\{p \cdot \log_2 p + (1-p) \cdot \log_2 (1-p)\}. \tag{5-21}$$

Fig. 5-2 shows the Shannon's entropy function h(p). What is evident from Fig.5-2 is that

1) h(p) is a smooth, convex function of p.
2) h(p) has a minimum value 0 at p=0 or p=1, i.e. if the phenomenon is completely known then the entropy is zero.
3) h(p) has a maximum 1 at p=0.5, i.e. the entropy attains a maximum 1 (bit) if the possibility of the phenomenon is fifty-fifty.

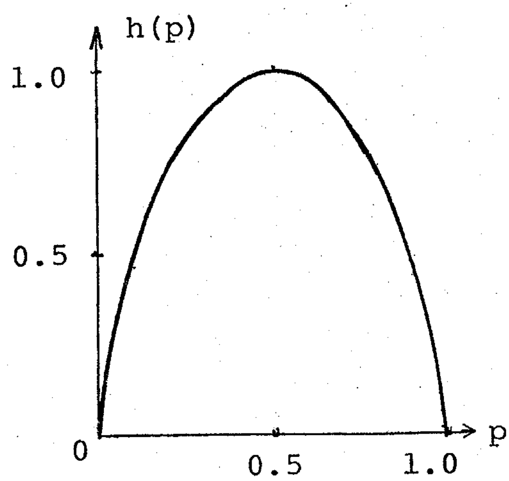


Fig.5-2. Shannon's entropy function

With these preliminaries, a theory of subjective entropy can be developed. In the beginning, we consider the simplest case, i.e. we assume that an object $x \in X$ is in a pure state with respect to a probabilistic set A on X. Then all informations are expressed by

$$\Omega_1(x) \underset{\Delta}{=} \{\omega \mid \mu_A(x,\omega)=1\}, \tag{5-22}$$

$$\Omega_0(x) \underset{\Delta}{=} \{\omega \mid \mu_A(x,\omega)=0\}, \tag{5-23}$$

$$\Omega = \Omega_1(x) \cup \Omega_0(x), \tag{5-24}$$

$$p(x) \underset{\Delta}{=} P(\Omega_1(x)), \quad 1-p(x) = P(\Omega_0(x)). \tag{5-25}$$

Hence, the ambiguity in Shannon's sense is given by $h(p(x))$. Since the partition (5-24) of the parameter space Ω depends upon observers' subjectivity, we call the value $h(p(x))$ a <u>subjective entropy</u> of x with respect to A, and denote it by $H(x,A)$, i.e.

$$H(x,A) = -\{p(x)\cdot\log_2 p(x) + (1-p(x))\cdot\log_2(1-p(x))\}. \tag{5-26}$$

In general, however, there exist ambiguous evaluations other than 0 and 1. (For example, see x_1 in Table 5-1.) We interpret these various, ambiguous evaluations as follows; Each of them is obtained as a result of partially averaging operations of fundamental parameters, where fundamental parameters are assumed to output evaluations of either 1(= "yes") or 0(= "no"). It will be necessary to explain this in detail. Suppose that there exist n different evaluations,

$$0 \leqq \alpha_1 < \alpha_2 < \ldots < \alpha_n \leqq 1, \tag{5-27}$$

and that each evaluation α_i occures with frequency $p_i(x)$. (In the example of Table 5-1, $\alpha_1 = 0$, $\alpha_2 = 0.25$, ... , $\alpha_5 = 1$.) This result is considered to be given by the following 2n-partition of the parameter space Ω (cf. Fig.5-3),

$$\Omega = \Omega_0(x) \cup \Omega_1(x), \tag{5-28}$$

$$\Omega_0(x) = \bigcup_{i=1}^{n} \Omega_o^i(x), \quad \Omega_1(x) = \bigcup_{i=1}^{n} \Omega_1^i(x). \tag{5-29}$$

For each i, a sub-unification of Ω, denoted by

$$\Omega^i(x) = \Omega_o^i(x) \cup \Omega_1^i(x), \tag{5-30}$$

is performed, and the evaluation

$$\alpha_i = P(\Omega_1^i(x)) / P(\Omega^i(x)), \tag{5-31}$$

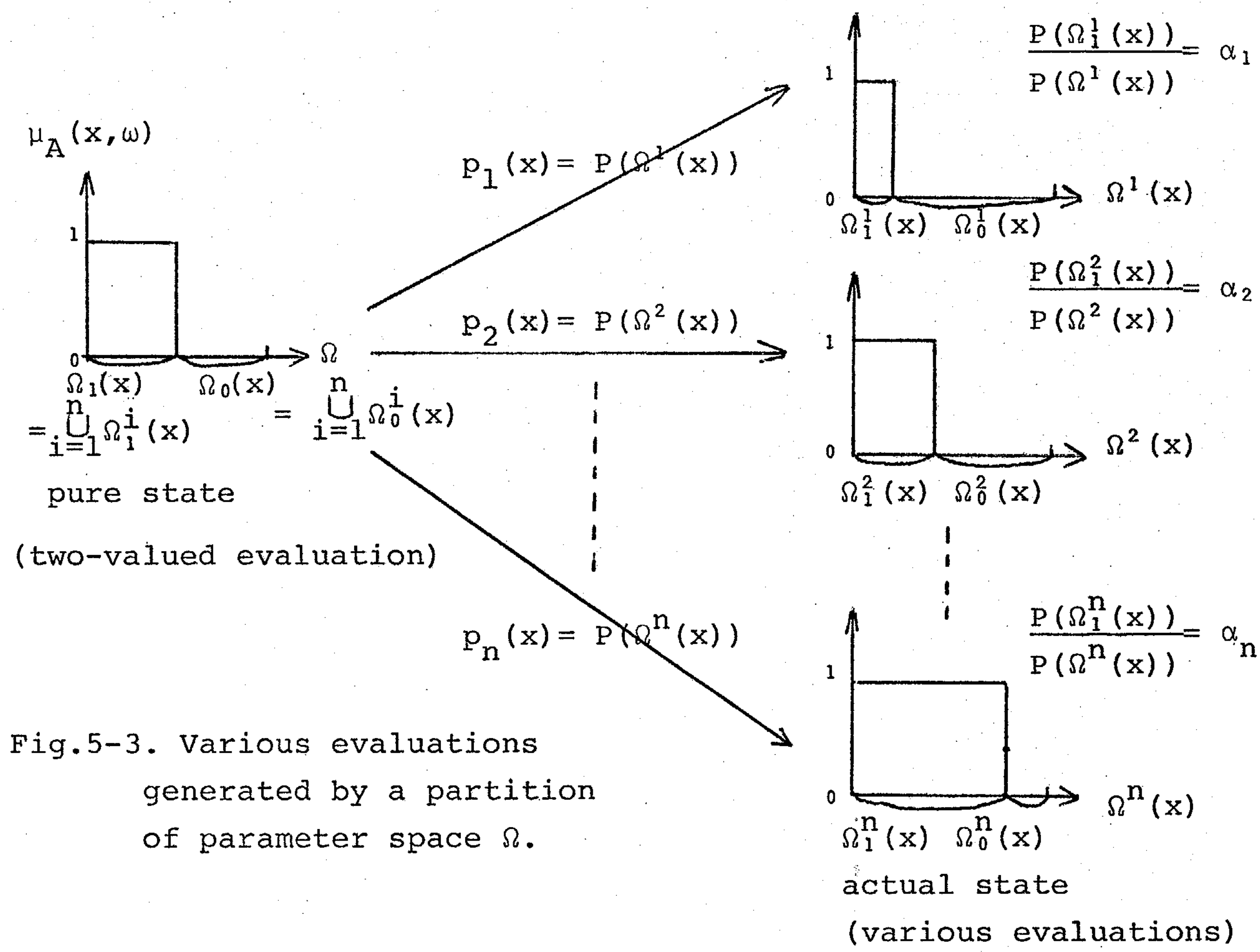


Fig.5-3. Various evaluations generated by a partition of parameter space Ω.

is obtained with frequency

$$p_i(x)= P(\Omega^i(x)), \tag{5-32}$$

as a result of an averaging operation in $\Omega^i(x)$. Here, a noteworthy point is that the value of membership function $m_A(x)$ is preserved in the averaging operation, i.e.

$$m_A(x)= P(\Omega_1(x)) \tag{5-33}$$

$$= \sum_{i=1}^{n} P(\Omega_1^i(x))= \sum_{i=1}^{n} \alpha_i \cdot p_i(x), \tag{5-34}$$

whereas the value of vagueness function $v_A(x)$ is not preserved but in general decreases.

Since n different kinds of criteria are required in order to make n different evaluations (5-27), the structure of the probabilistic set becomes complicated compared with the structure in a pure state. If we consider that the complexity depends on the 2n-partition (5-28),(5-29) of Ω, then the complexity is expressed by the entropy of the 2n-partition of Ω. We call the value a <u>subjective entropy</u> of x with respect to A, and denote it by H(x,A),

$$H(x,A) = -\sum_{i=1}^{n}\{P(\Omega_0^i(x))\cdot\log_2 P(\Omega_0^i(x)) + P(\Omega_1^i(x))\cdot\log_2 P(\Omega_1^i(x))\}. \tag{5-35}$$

Although we can not calculate this quantity directly (since the 2n-partition (5-28),(5-29) depends on observers subjectivity and the exact structure is in general unknown), we can estimate it by using the following equations,

$$P(\Omega_1^i(x)) = \alpha_i\cdot p_i(x), \quad P(\Omega_0^i(x)) = (1-\alpha_i)\cdot p_i(x). \tag{5-36}$$

Note that both α_i's and $p_i(x)$'s are known. The result may be summarized as the following theorem.

[Theorem 5-2]

Under an assumption of (5-27) (where each evaluation α_i occurs with frequency $p_i(x)$), the subjective entropy H(x,A) of $x(\varepsilon X)$ with respect to A is given by

$$H(x,A) = -\sum_{i=1}^{n}\{\alpha_i\cdot p_i(x)\cdot\log_2\alpha_i p_i(x) + (1-\alpha_i)\cdot p_i(x)\cdot\log_2(1-\alpha_i)p_i(x)\} \tag{5-37}$$

$$= -\sum_{i=1}^{n} p_i(x)\cdot\log_2 p_i(x) + \sum_{i=1}^{n} p_i(x)\cdot h(\alpha_i). \tag{5-38}$$

Several examples will be given to help the explanation done so far. Consider the simplest case first, i.e. n=1 in (5-27). In such a case, the evaluation is α_1-constant and $p_1(x) = 1$. Hence we have

$$H(x,A) = -\{\alpha_1\cdot\log_2\alpha_1 + (1-\alpha_1)\cdot\log_2(1-\alpha_1)\}$$

$$= h(\alpha_1) = h(P(\Omega_1(x))), \tag{5-39}$$

and this value is the same as that in a pure state (5-26). A point to which special attention should be paid is that the vagueness function has a minimum 0 since all values of the evaluation are uniformly equalized to α_1, where as it has a maximum in a pure state. Hence we can make the following statement, i.e. the concept of subjective entropy and that of vagueness function are mutually independent. We call this state a stationary state, since the evaluation is kept unchanged. In Table 5-1, for example, x_3 is in the stationary state with respect to A.

In the next place, consider the second simplest case, i.e. n=2 and

$$\Omega_1^1(x) = \phi, \qquad \Omega_0^1(x) = \Omega_0(x), \qquad \alpha_1 = 0, \tag{5-40}$$

$$\Omega_1^2(x) = \Omega_1(x), \quad \Omega_0^2(x) = \phi, \qquad \alpha_2 = 1. \tag{5-41}$$

Then we have

$$H(x,A) = -\{p(\Omega_0(x))\cdot\log_2 P(\Omega_0(x)) + P(\Omega_1(x))\cdot\log_2 P(\Omega_1(x))\}$$

$$= h(P(\Omega_1(x))), \tag{5-42}$$

and this is the same as Shannon's entropy itself (5-21). Hence, we can conclude that the notion of subjective entropy is an expansion of that of Shannon's entropy which is defined by both {0,1}-two values and probability.

If the number of states n becomes large and the structure becomes complicated, then the value of H(x,A) also increases. For example, in Table 5-1, the subjective entropy of x_2 (in pure state) and that of x_3 (in stationary state) are the same

$$H(x_2,A) = H(x_3,A) = 1, \tag{5-43}$$

whereas that of x_1 (n=5) is larger than 1, i.e.

$$H(x_1, A) = 2.70. \tag{5-45}$$

The outline of the discussion is as follows; A notion of subjective entropy $H(x,A)$ is introduced as an expansion of Shannon's entropy. It is a measure which indicates the degree of complexity of probabilistic sets, and is an independent measure of the vagueness function.

5-4. AN AXIOMATIC DEFINITION OF SUBJECTIVE ENTROPY.

Although a pure state plays an important role in the definition of subjective entropy, it is a fictious, extreme state. In practice there are deviations from the pure state due to various reasones, in general [0,1]-ambiguous evaluations are observed. Hence it is worth investigating the subjective entropy based on general ambiguous states not on pure states.

Let an object $x(\varepsilon X)$ and a probabilistic set A (on X) be arbitrarily fixed for convenience. Denote $p_i(x)$ by just p_i. Then the value of H(x,A) is determined by both α_i's and p_i's (i=1,..., n). Hence, we can regard H(x,A) as a function of these 2n-variables, and we denote it by $f_n(\alpha_1,p_1;\ldots;\alpha_n,p_n)$, i.e.

$$f_n:[0,1]^n \times \Delta^n \longrightarrow [0,\infty),$$
$$\begin{array}{ccc} \omega & & \omega \end{array}$$
$$(\alpha_1,..,\alpha_n,p_1,..,p_n) \longmapsto f_n(\alpha_1,p_1;..;\alpha_n,p_n)$$
$$= -\{\sum_{i=1}^{n} p_i\alpha_i\cdot\log_2 p_i\alpha_i + p_i(1-\alpha_i)\cdot\log_2 p_i(1-\alpha_i)\},$$

where (5-46)

$$\Delta^n = \{(p_1,p_2,\ldots,p_n) \mid p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1\}. \quad (5\text{-}47)$$

We shall investigate fundamental properties of a set of functions $\{f_n\}_{n=1}^{\infty}$. Since

$$f_1(\alpha_1,1) = -\{\alpha_1\cdot\log_2\alpha_1 + (1-\alpha_1)\cdot\log_2(1-\alpha_1)\} = h(\alpha_1), \quad (5\text{-}48)$$

holds in the case of n=1, we obtain the following property.

Ax.1) $f_1(\alpha,1)$ is a continuous function of α, and we have

$$f_1(1/2,1) = 1. \quad (5\text{-}49)$$

It is very natural that f_1 is a continuous measure of the variable α. The condition (5-49) offers a normarizarion of this measure, i.e. the unit of the measure is chosen so that a uniformly ambiguous state ($\alpha_1 = 0.5$ constant) takes a value 1. Since the value of f_n doesn't change by the change of the indices i's of α_i's and p_i's , we have;

Ax.2) For an arbitrary permutation $(i_1, i_2, \ldots, i_n)$ of $(1,2,\ldots,n)$, a relation

$$f_n(\alpha_1, p_1; \ldots; \alpha_n, p_n) = f_n(\alpha_{i_1}, p_{i_1}; \ldots; \alpha_{i_n}, p_{i_n}), \qquad (5\text{-}50)$$

holds.

This property also seems to be a natural requirement, since the complexity of probabilistic sets should not depend upon the numerical order of the states α_i's. If a state α_n is replaced by one of extremes $\{0,1\}$, then we obtain the following from (5-46).

Ax.3) $f_n(\alpha_1, p_1; \ldots; \alpha_{n-1}, p_{n-1}; \alpha_n, p_n)$

$$= f_n(\alpha_1, p_1; \ldots; \alpha_{n-1}, p_{n-1}; 1, p_n) + p_n \cdot f_1(\alpha_n, 1)$$

$$= f_n(\alpha_1, p_1; \ldots; \alpha_{n-1}, p_{n-1}; 0, p_n) + p_n \cdot f_1(\alpha_n, 1). \qquad (5\text{-}51)$$

From this equation, we can infer that the value of states α_i's has symmetry with respect to a value 0.5, i.e. a value 0 (in general α) is on a level with a value 1 (in general $1-\alpha$) from a viewpoint of subjective entropy. In addition to this property, we have the following recurrence fromula of f_{n+1} and f_n from (5-46),

Ax.4) $f_{n+1}(\alpha_1, p_1; \ldots; \alpha_{n-1}, p_{n-1}; \alpha_n, p_n; \alpha_{n+1}, p_{n+1})$

$$= f_n(\alpha_1, p_1; \ldots; \alpha_{n-1}, p_{n-1}; 1, p_n + p_{n+1}) +$$

$$+ (p_n + p_{n+1}) \cdot f_2(\alpha_n, p_n/(p_n + p_{n+1}); \alpha_{n+1}, p_{n+1}/(p_n + p_{n+1})).$$

(5-52)

We have derived four properties Ax.1)∿Ax.4) from the definition of subjective entropy (5-46). Moreover, we can verify the validity of a converse procedure.

[Theorem 5-3]

The function (5-46) satisfies the abovementioned four properties Ax.1)∿Ax.4). Conversely, the function which satisfies Ax.1)∿Ax.4) must have the form of (5-46).

proof

All we have to show is a converse part, i.e. we shall derive (5-46) from Ax.1)∿Ax.4). If we put

$$p_1, p_2 \geqq 0, \quad p_3 > 0, \quad p_1+p_2+p_3 = 1, \ \alpha_1 = \alpha_2 = \alpha_3 = 1, \tag{5-53}$$

and apply Ax.4), then we have

$$f_3(1,p_1;1,p_2;1,p_3) = f_2(1,p_1;1,1-p_1) + \\ + (1-p_1)\cdot f_2(1,p_2/(1-p_1);1,p_3/(1-p_1)), \tag{5-54}$$

$$f_3(1,p_2;1,p_1;1,p_3) = f_2(1,p_2;1,1-p_2) + \\ + (1-p_2)\cdot f_2(1,p_1/(1-p_2);1,p_3/(1-p_2)). \tag{5-55}$$

Note that left sides of (5-54) and (5-55) are equal (from Ax.2)). If we adopt the following notation,

$$f_2(1,p;1,(1-p)) \underset{\Delta}{=} g(p), \tag{5-56}$$

for convenience, then we have from (5-54) and (5-55),

$$g(p_1) + (1-p_1)\cdot g(p_2/(1-p_1)) \\ = g(p_2) + (1-p_2)\cdot g(p_1/(1-p_2)). \tag{5-57}$$

Substitute $p_1=0$ in (5-57) and note the fact that p_2 can be arbitrarily changed. Then we obtain

$$g(0) = 0. \tag{5-58}$$

Integrate both members of (5-57) with respect to p_2 from 0 to $1-p_1$ (this integration is possible by Ax.1)), then we have

$$(1-p_1)\cdot g(p_1) + (1-p_1)\int_0^{1-p_1} g(p_2/(1-p_1))\cdot dp_2 \\ = \int_0^{1-p_1} g(p_2)dp_2 + \int_0^{1-p_1}(1-p_2)\cdot g(p_1/(1-p_2))\cdot dp_2, \tag{5-59}$$

therefore

$$(1-p_1)\cdot g(p_1) + (1-p_1)^2\cdot\int_0^1 g(t)dt =$$
$$= \int_0^{1-p} g(p_2)dp_2 + p_1{}^2\cdot\int_{p_1}^1 g(t)/t^3\cdot dt. \qquad (5\text{-}60)$$

All terms in (5-60) except the first one in the left side are differentiable with respect to p_1. Hence the first term, especially $g(p_1)$, also becomes differentiable, and we obtain,

$$(1-p_1)\cdot g'(p_1) - g(p_1) - 2(1-p_1)\cdot\int_0^1 g(t)dt =$$
$$= -g(1-p_1) + 2p_1\int_{p_1}^1 g(t)/t^3\cdot dt - g(p_1)/p_1. \qquad (5\text{-}61)$$
$$\therefore (1-p_1)\cdot g'(p_1) = 2(1-p_1)\cdot\int_0^1 g(t)dt +$$
$$+ 2p_1\int_{p_1}^1 g(t)/t^3\cdot dt - g(p_1)/p_1. \qquad (5\text{-}62)$$

The right side of (5-62) is differentiable with respect to p_1, so $g'(p_1)$ in the left side has also a derivative $g''(p_1)$, and we obtain

$$-g'(p_1) + (1-p_1)\cdot g''(p_1) =$$
$$= -2\int_0^1 g(t)dt + 2\int_{p_1}^1 g(t)/t^3\cdot dt - 2g(p_1)/p_1{}^2 -$$
$$- (g'(p_1)p_1 - g(p_1))/p_1{}^2. \qquad (5\text{-}63)$$

From (5-62) and (5-63), we have

$$g''(p_1) = (-2/p_1(1-p_1))\cdot\int_0^1 g(t)dt, \qquad (5\text{-}64)$$
$$\therefore g(p_1) = -2\int_0^1 g(t)dt\cdot[p_1\log p_1 + (1-p_1)\log(1-p_1)] + c_1p_1 + c_2.$$
$$(5\text{-}65)$$

Since $g(p_1) = g(1-p_1)$, we obtain $c_1 = 0$, and we also have $c_2 = 0$ from (5-58). Hence, we obtain

$$g(p) = f_2(1,p;1,(1-p))$$
$$= -2\int_0^1 g(t)dt\cdot[p\cdot\log p + (1-p)\cdot\log(1-p)]. \qquad (5\text{-}66)$$

This is equal to $f_1(p,1)$ (by Ax.1),Ax.2),Ax.3)). If we apply the normalization-condition of Ax.1), we have

$$-2\int_0^1 g(t)dt = -1/\log 2. \qquad (5\text{-}67)$$

Hence we can obtain

$$f_1(p,1) = f_2(1,p;1,(1-p))$$
$$= -[p \log_2 p + (1-p)\cdot\log_2(1-p)]. \qquad (5\text{-}68)$$

From the facts described above, and by using a mathematical induction, we can conclude that

$$f_n(\alpha_1,p_1;\alpha_2,p_2;\ldots;\alpha_n,p_n) =$$
$$= -\sum_{i=1}^{n} (p_i\alpha_i\cdot\log_2 p_i\alpha_i + p_i(1-\alpha_i)\cdot\log_2 p_i(1-\alpha_i)), \qquad (5\text{-}69)$$

as desired. (In the case of n=2, apply Ax.3),Ax.2),Ax.3) and (5-68) successively. And apply Ax.4) in the induction from n= k to n= k+1.)

(Q.E.D.)

It is well-known that the notion of Shannon's entropy can be derived from several axioms. For example, the Faddeev's axiom is widely known [2]. We can regard abovementioned four properties Ax.1)∿ Ax.4) as an extension of the Faddeev's axiom from a case of {0,1}-two values into that of $\{\alpha_1,\ldots,\alpha_n\}$-many values. Hence, an axiomatic definition of subjective entropy is established by theorem 5-3 in the case of finitely many-valued probabilistic sets.

5-5. SUBJECTIVE ENTROPY IN INFINITE STATES.

In section 5-3, we defined a subjective entropy in finite states. Here, we shall also study it in infinite states. (Consider the case, for example, that the value of $\mu_A(x,\omega)$ ranges the whole interval of [0,1].)

Let Δ be an arbitrary ε-partition of [0,1], where

$$\Delta:\ 0 = \alpha_0 < \alpha_1 < \dots < \alpha_{n-1} < \alpha_n = 1, \tag{5-70}$$

$$\delta_\Delta = \max\{\alpha_{i+1} - \alpha_i \mid 0 \leq i \leq n-1\} \leq \varepsilon. \tag{5-71}$$

We can induce the following partition of the parameter space Ω by this ε-partition,

$$\Omega = \Omega_1(x) \cup \dots \cup \Omega_n(x), \tag{5-72}$$

$$\Omega_i(x) \cap \Omega_j(x) = \phi \qquad (i \neq j), \tag{5-73}$$

$$\begin{aligned} \Omega_i(x) &= \{\omega \mid \alpha_{i-1} \leqq \mu_A(x,\omega) < \alpha_i\} \quad (i = 1,2,\dots,n-1), \\ \Omega_n(x) &= \{\omega \mid \alpha_{n-1} \leqq \mu_A(x,\omega) \leqq \alpha_n = 1\}. \end{aligned} \tag{5-74}$$

Choose n points $\{\xi_i\}_{i=0}^{n-1}$ in [0,1] such that they are arranged in [0,1]-interval alternatively with $\{\alpha_i\}_{i=0}^{n}$, and define

$$\Delta_\xi:\ 0 = \alpha_0 \leqq \xi_0 \leqq \alpha_1 \leqq \xi_1 \leqq \dots \leqq \alpha_{n-1} \leqq \xi_{n-1} \leqq \alpha_n = 1. \tag{5-75}$$

Put

$$p_i(x) \underset{\Delta}{=} P(\Omega_i(x)), \tag{5-76}$$

and consider the following quantity,

$$-\sum_{i=0}^{n-1} \{p_i(x)\xi_i \cdot \log_2 p_i(x)\xi_i + p_i(x)(1-\xi_i) \cdot \log_2 p_i(x)(1-\xi_i)\}, \tag{5-77}$$

on refferring to (5-37). This quantity will be changed according to an arrangement of α_i's and ξ_i's. We denote its infimum by $H_\varepsilon(x,A)$,

$$H_\varepsilon(x,A) \underset{\Delta}{=} \inf\{-\Sigma\,(p_i(x)\xi_i \cdot \log_2 p_i(x)\xi_i + p_i(x)(1-\xi_i) \cdot$$

$\cdot \log_2 p_i(x)(1-\xi_i)) | \Delta_\xi, \ \Delta$ is an ε-partition$\}$,

(5-78)

and call it an ε-subjective entropy of x with respect to A. An approximation of subjective entropy may be given by $H_\varepsilon(x,A)$ in the sense that two evaluations in [0,1]-interval must be identical if the difference between the two is at most ε.

Under these considerations, we define subjective entropy $H(x,A)$ of x with respect to A as a limit of $H_\varepsilon(x,A)$, i.e.

$$H(x,A) = \lim_{\varepsilon \to 0} H_\varepsilon(x,A). \tag{5-79}$$

The existence of this limit in $[0,+\infty]$-interval will be proved in a later section, i.e. the proof is included in that of theorem 5-5, (it will be confirmed by using (5-100) and (5-101)). If the value of $\mu_A(x,\omega)$ ranges the whole interval of [0,1], $H(x,A)$ may be $+\infty$. In such a case, we attach importance to $H_\varepsilon(x,A)$ for a suitably chosen $\varepsilon(>0)$. (Note that $H_\varepsilon(x,A)$ always takes a finitely positive value.) There is no restriction on the probabilistic set A in the general definition (5-79) of subjective entropy. Hence, the subjective entropy can be defined for all probabilistic sets by (5-79). It will be easily verified that subjective entropy by the general definition (5-79) provides the same result as that in finite states (5-37), if the range of $\mu_A(x,\omega)$ is $\{\alpha_1,\alpha_2,\ldots,\alpha_n\}$. (It mainly depends on the continuity of Shannon's entropy function $h(\cdot)$.) The validity of the general definition (5-79) is also confirmed in this point.

5-6. A MAXIMIZATION AND MINIMIZATION PROBLEM OF SUBJECTIVE ENTROPY.

We shall investigate the complexity and the simplicity of probabilistic sets by solving a maximization and minimization problem of subjective entropy.

As we have already mentioned in a previous section, the maximum value of $H(x,A)$ can reach any large number when x is in an infinite state with respect to A, and the structure of probabilistic sets also becomes infinitely complicated. Therefore, we sometimes consider ε-subjective entropy instead of subjective entropy. In such a case, however, an approximation by finite states becomes very important.

For arbitrary two probabilistic set A and B, we denote the mean-value of a symmetric difference $A \Delta B$ (cf. (3-95)) by $d_x(A,B)$,

$$d_x(A,B) \underset{\Delta}{=} E[\mu_{A\Delta B}(x,\cdot)]$$

$$= \int_\Omega |\mu_A(x,\omega)-\mu_B(x,\omega)|dP(\omega). \tag{5-80}$$

Then $d_x(\cdot,\cdot)$ provides a pseudo-metric in $P(X)$, since it satisfies non-negativeness, symmetricity and transitivity. (Note the fact that $P(X)$ constitutes a strongly closed, convex set on $L^1(\Omega)$.) We can approximate $\mu_A(x,\omega)$ by finite states as closely as possible in the sense of $d_x(\cdot,\cdot)$, i.e. for arbitrary $\varepsilon>0$, there exist positive numbers $\alpha_1,\alpha_2,\ldots,\alpha_n$ and a partition of the parameter space Ω,

$$\Omega = \Omega_1(x) \cup \ldots \cup \Omega_n(x), \quad \Omega_i(x) \cap \Omega_j(x) = \phi \quad (i \neq j), \tag{5-81}$$

such that

$$\int_\Omega |\mu_A(x,\omega) - \sum_{i=1}^{n} \alpha_i \cdot 1_{\Omega_i(x)}(\omega)| \cdot dP(\omega) \leq \varepsilon \ , \tag{5-82}$$

where

$$1_{\Omega_i(x)}(\omega) = \begin{cases} 1 & \omega \in \Omega_i(x) \\ 0 & \omega \notin \Omega_i(x). \end{cases} \tag{5-83}$$

In fact, if we put

$$n = [1/\varepsilon] + 1,\ \alpha_i = (i-1)/n,\ \Omega_i(x) = \{\omega \mid \alpha_i \leqq \mu_A(x,\omega) < \alpha_{i+1}\}, \tag{5-84}$$

then we have

$$0 \leqq \mu_A(x,\omega) - \sum_{i=1}^{n} \alpha_i \cdot 1_{\Omega_i(x)}(\omega) \leqq \varepsilon, \tag{5-85}$$

and obtain (5-82).

Based on the abovementioned arguments, we consider a maximization problem of $H(x,A)$ under an assumption that the value of $\mu_A(x,\omega)$ is restricted to

$$0 \leqq \alpha_1 < \alpha_2 < \dots < \alpha_n \leqq 1, \tag{5-86}$$

where $\mu_A(x,\omega) = \alpha_i$ occurs with probability $p_i(x)$. For convenience, denote $p_i(x)$ by just p_i, and $H(x,A)$ by $f(p_1, p_2, \dots, p_n)$, i.e.

$$f(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} \{p_i\alpha_i \cdot \log_2 p_i\alpha_i + p_i(1-\alpha_i) \cdot \log_2 p_i(1-\alpha_i)\}. \tag{5-87}$$

We consider the following maximization problem; Maximize $f(p_1, \dots, p_n)$ subject to

$$\sum_{i=1}^{n} p_i(x) = 1, \quad p_i \geqq 0. \tag{5-88}$$

We shall use the Lagrange's method of indeterminate coefficients.

$$f(p_1, \dots, p_n) = -\sum_{i=1}^{n} \{p_i\alpha_i \cdot \log_2 p_i\alpha_i + p_i(1-\alpha_i) \cdot \log_2 p_i(1-\alpha_i)\} + \lambda\left(\sum_{i=1}^{n} p_i - 1\right). \tag{5-89}$$

$$\partial f/\partial p_i = -\{\log_2 e - \log_2 p_i + h(\alpha_i)\} + \lambda = 0, \tag{5-90}$$

$$\therefore\ p_i = 2^{\{\lambda - \log_2 e + h(\alpha_i)\}}. \tag{5-91}$$

We obtain the maximum of $f(p_1, \dots, p_n)$ by determining λ from (5-91) and (5-88). The result is summarized in the following theorem.

[Theorem 5-4]

If the value of $\mu_A(x,\omega)$ is restricted to (5-86), the subjective entropy H(x,A) has a maximum $H(x,A)_{max}$,

$$H(x,A)_{max} = \log_2 \sum_{i=1}^{n} 2^{h(\alpha_i)}, \tag{5-92}$$

when

$$p_i(x) = 2^{h(\alpha_i)} / \sum_{i=1}^{n} 2^{h(\alpha_i)}. \tag{5-93}$$

In the next place, we debate a minimization problem of H(x,A). We can infer from the definition of subjective entropy and the example of table 5-1 that H(x,A) attains a minimum in a pure state or a stationary state. In fact, we have the following theorem.

[Theorem 5-5]

Under the condition of

$$E[\mu_A(x,\cdot)] = \int_\Omega \mu_A(x,\omega)\,dP(\omega) \equiv m, \tag{5-94}$$

H(x,A) has a minimum $H(x,A)_{min}$,

$$H(x,A)_{min} = -\{m\cdot\log_2 m + (1-m)\cdot\log_2(1-m)\} = h(m), \tag{5-95}$$

when x is in a pure state or in a stationary state with respect to A.

<u>proof</u>

Since we have already mentioned that H(x,A) is equal to h(m) in a pure state or in a stationary state (cf. the last part of section 5-3), we shall prove that

$$H(x,A) \geqq h(m) - \varepsilon, \tag{5-96}$$

for an arbitrary positive number ε and an arbitrary probabilistic set A which satisfies $H(x,A) < +\infty$.

Consider a partition Δ_ξ in (5-75) and denote the quantity of (5-77) by $f(\Delta_\xi)$, i.e.

$$f(\Delta_\xi) = -\sum_{i=0}^{n-1} p_i(x)\cdot\log_2 p_i(x) + \sum_{i=0}^{n-1} p_i(x)\cdot h(\xi_i). \qquad (5\text{-}97)$$

Since $h(\cdot)$ in the second term is a continuous function on $[0,1]$-interval, there exist a minimum and a maximum of $f(\Delta_\xi)$ by changing $\{\xi_i\}_{i=0}^{n-1}$ (here, Δ is fixed). We denote $\{\xi_i\}_{i=0}^{n-1}$ which offers the minimum and the maximum by $\{\xi_i^{min}\}_{i=0}^{n-1}$ and $\{\xi_i^{max}\}_{i=0}^{n-1}$ respectively. (Of course,

$$\alpha_i \leqq \xi_i^{max}, \xi_i^{min} \leqq \alpha_{i+1} \text{ for all } i. \qquad (5\text{-}98)\ )$$

Then we obtain

$$\begin{aligned} 0 &\leqq \max\{f(\Delta_\xi) \mid \xi\} - \min\{f(\Delta_\xi) \mid \xi\} \\ &= \sum_{i=1}^{n-1} p_i(x)\cdot(h(\xi_i^{max}) - h(\xi_i^{min})\}. \end{aligned} \qquad (5\text{-}99)$$

The right side of (5-99) tends to 0 as δ_Δ(5-71) tends to 0, i.e.

$$\begin{aligned} 0 &\leqq \max\{f(\Delta_\xi) \mid \xi\} - \min\{f(\Delta_\xi) \mid \xi\} \\ &\longrightarrow 0 \qquad (\delta_\Delta \longrightarrow 0). \end{aligned} \qquad (5\text{-}100)$$

(Note that $h(\cdot)$ is uniformly continuous.)

Next, we prove that

$$\min\{f(\Delta_\xi) \mid \xi\} \leqq \min\{f(\tilde{\Delta}_\xi) \mid \xi\}, \qquad (5\text{-}101)$$

where $\tilde{\Delta}$ is a refinement of Δ ($\Delta \leqq \tilde{\Delta}$). If we take the function-form of $f(\Delta_\xi)$ into the consideration, it is sufficient to show

$$\begin{aligned} &\min[-\sum_{i=1}^{n} \{p_i(x)\eta_i\cdot\log_2 p_i(x)\eta_i + p_i(x)(1-\eta_i)\cdot \\ &\qquad \cdot\log_2 p_i(x)(1-\eta_i)\} \mid \beta_i \leqq \eta_i \leqq \beta_{i+1}] - \\ &- \min[-\sum_{i=1}^{n} p_i(x)\eta\cdot\log_2 \sum_{i=1}^{n} p_i(x)\eta + \sum_{i=1}^{n} p_i(x)(1-\eta)\cdot \\ &\qquad \cdot\log \sum_{i=1}^{n} p_i(x)(1-\eta)\} \mid \beta_1 \leqq \eta \leqq \beta_{n+1}] \geqq 0, \end{aligned} \qquad (5\text{-}102)$$

where

$$\begin{aligned} &(0 \leqq)\ \beta_1 < \beta_2 < \ldots < \beta_{n+1}\ (\leqq 1), \\ &\qquad \beta_i \leqq \eta_i \leqq \beta_{i+1},\ \beta_1 \leqq \eta \leqq \beta_{n+1}. \end{aligned} \qquad (5\text{-}103)$$

The inequality (5-102) will be easily obtained by a direct calculation.

If we put

$$f(\Delta) \underset{\overline{\Delta}}{=} - \sum_{i=0}^{n-1} p_i(x)\cdot \log_2 p_i(x) + \sum_{i=0}^{n-1} p_i(x)\cdot h(\alpha_i), \tag{5-104}$$

then we have ,from (5-100),

$$0 \leqq f(\Delta) - \min\{f(\Delta_\xi)|\xi\} \leqq \varepsilon/2, \tag{5-105}$$

for the ε of (5-96) and a sufficiently small partition Δ. On the other hand, since (5-101) holds and since

$$H(x,A) \geqq \min\{f(\Delta_\xi)|\xi\}, \tag{5-106}$$

holds (cf. the definition of H(x,A) (5-79)), we obtain

$$H(x,A) \geqq f(\Delta) - (\varepsilon/2). \tag{5-107}$$

A direct calculation shows

$$f(\Delta) - h(\sum_{i=0}^{n-1} p_i(x)\alpha_i) \geqq 0. \tag{5-108}$$

Hence we have, from (5-107) and (5-108),

$$H(x,A) \geqq h(\sum_{i=0}^{n-1} p_i(x)\alpha_i) - (\varepsilon/2). \tag{5-109}$$

If the partition Δ is sufficiently small, then we can make the following quantity,

$$m - \sum_{i=0}^{n-1} \alpha_i\cdot p_i(x) \quad (\geqq 0), \tag{5-110}$$

sufficiently small. (Apply an approximation by a simple function to the condition (5-94).) Hence, we may assume that

$$|h(m) - h(\sum_{i=0}^{n-1} \alpha_i\cdot p_i(x))| \leqq \varepsilon/2 , \tag{5-111}$$

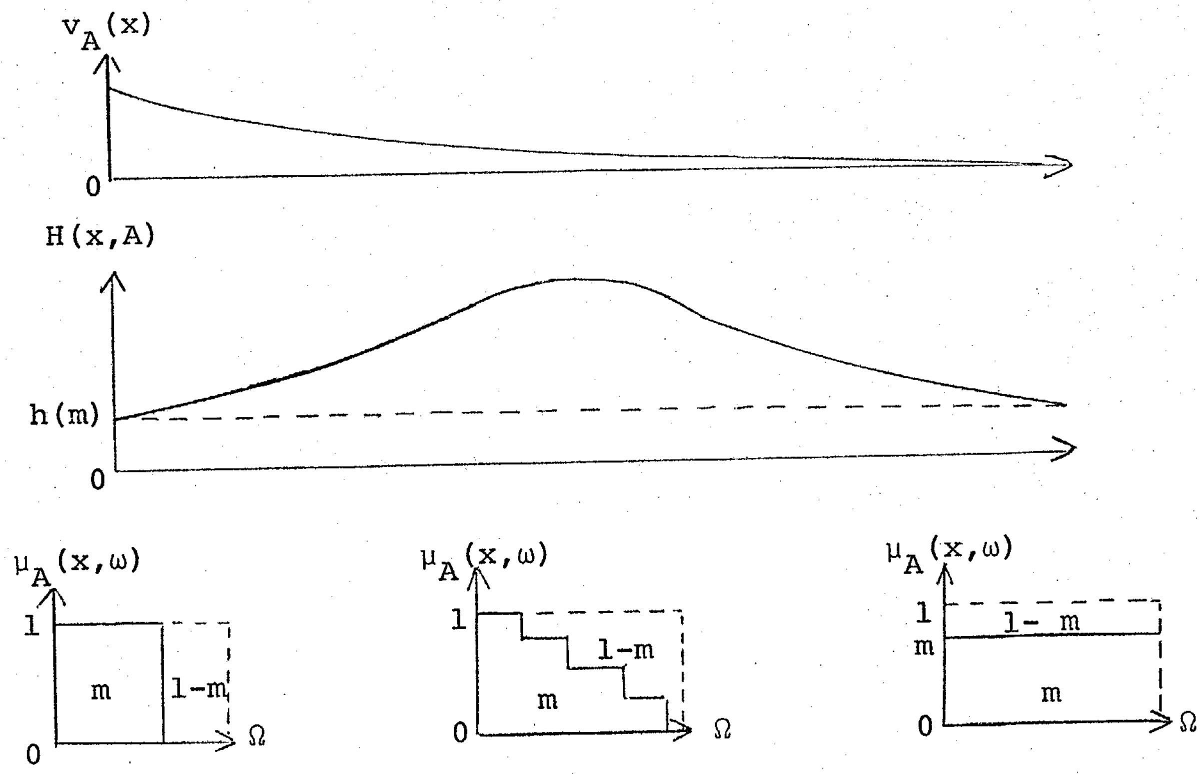
for a sufficiently small partition Δ. (Note the uniformly continuity of h(·).) If we substitute (5-111) to (5-109), we can obtain the desired result (5-96), i.e.

$$\begin{aligned} H(x,A) &\geqq (h(m) - (\varepsilon/2)) - (\varepsilon/2) \\ &= h(m) - \varepsilon. \end{aligned} \tag{5-112}$$

Since the positive number ε is arbitrarily chosen, we can conclude that

$$H(x,A) \geqq h(m). \tag{5-113}$$

(Q.E.D.)



pure state ({0,1}-two-valued logic + probability)

actual state

stationary state (Zadeh's (classical) fuzzy concept)

Fig. 5-4. Fuzzy vs probability from viewpoints of a subjective entropy $H(x,A)$ and a vagueness function $v_A(x)$.

The noteworthy points from the above theorem are summarized as follows (cf. Fig.5-4): As we have already mentioned, a stationary state corresponds to a (classical) fuzzy concept by Zadeh, whereas a pure state corresponds to a notion of probability and {0,1}-two-valued logic. From a viewpoint of subjective entropy $H(x,A)$, both states are the same and they attain a minimum $h(m)$ under the assumption of $E[\mu_A(x,\cdot)] = m$. On the other hand, from a viewpoint of a vagueness function $v_A(x)$, the stationary state attains a minimum 0, whereas the pure state attains a maximum.

Hence it becomes clear that the notion of (subjective) entropy is an important measure of ambiguity, but that there exists another kind of ambiguity such as the notion of vagueness function and so on.

5-7. SUBJECTIVE ENTROPY OF PLURAL PHENOMENA.

It is the main purpose of this section to investigate subjective entropy of plural phenomena.

Let $(\mu_A(x,\omega), \mu_B(x,\omega), P(\omega))$ be a probabilistic expression of two probabilistic sets A and B on X (cf. section 4-3). We have characterized a pure state as the simplest structure of a single probabilistic set (cf. Def.5-1,Theorem 5-1). It is also possible to consider a pure state in plural probabilistic sets. If an object $x(\varepsilon X)$ belongs to a pure state with respect to both A and B, then the value of $\mu_A(x,\omega)$ and $\mu_B(x,\omega)$ has $2^2=4$ combimations and we have the following partition of the parameter space Ω in almost the same manner as theorem 5-1 (cf. Fig.5-5)..

$$\Omega = \Omega_{11}(x) \cup \Omega_{10}(x) \cup \Omega_{01}(x) \cup \Omega_{00}(x), \tag{5-114}$$

where

$$\Omega_{ij}(x) = \{\omega \mid \mu_A(x,\omega) = i,\ \mu_B(x,\omega) = j\} \quad (i,j = 0 \text{ or } 1), \tag{5-115}$$

$$p_{ij}(x) \underset{\Delta}{=} P(\Omega_{ij}(x)). \tag{5-116}$$

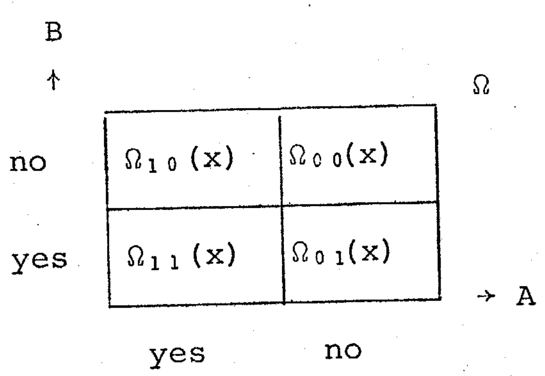


Fig.5-5. A partition of Ω in the case that x belongs to a pure state with respect to both A and B.

We may consider the actual, various evaluations as being obtained by an averaging operation of the abovestated pure state. Let the number of states be l (with respect to A) and m (with respect to B),

$$0 \leqq \alpha_1 < \alpha_2 < \dots < \alpha_l \leqq 1, \tag{5-117}$$

$$0 \leqq \beta_1 < \beta_2 < \dots < \beta_m \leqq 1. \tag{5-118}$$

Then the parameter space Ω is divided into $l \cdot m$ regions as follows,

$$\Omega = \bigcup_{i=1}^{l} \bigcup_{j=1}^{m} \Omega^{ij}(x), \tag{5-119}$$

where

$$\Omega^{ij}(x) = \{\omega \mid \mu_A(x,\omega) = \alpha_i,\ \mu_B(x,\omega) = \beta_j\}, \tag{5-120}$$

$$p(\alpha_i, \beta_j\, ; x) \underset{\Delta}{=} P(\Omega^{ij}(x)). \tag{5-121}$$

Each $\Omega^{ij}(x)$ consists of the following four parts in the same manner as (5-114) and (5-115),

$$\Omega^{ij}(x) = \Omega^{ij}_{11}(x) \bigcup \Omega^{ij}_{10}(x) \bigcup \Omega^{ij}_{01}(x) \bigcup \Omega^{ij}_{00}(x), \tag{5-122}$$

$$\begin{aligned} &\Omega^{ij}_{11}(x) \subset \Omega_{11}(x), \quad \Omega^{ij}_{10}(x) \subset \Omega_{10}(x), \\ &\Omega^{ij}_{01}(x) \subset \Omega_{01}(x), \quad \Omega^{ij}_{00}(x) \subset \Omega_{00}(x). \end{aligned} \tag{5-123}$$

Hence, we have

$$\alpha_i = \{P(\Omega^{ij}_{11}(x)) + P(\Omega^{ij}_{10}(x))\} \,/\, P(\Omega^{ij}(x)), \tag{5-124}$$

$$\beta_j = \{P(\Omega^{ij}_{11}(x)) + P(\Omega^{ij}_{01}(x))\} \,/\, P(\Omega^{ij}(x)). \tag{5-125}$$

For convenience, we introduce the following notations (cf. Fig. 5-6),

$$p(\alpha_i; x) = \sum_{j=1}^{m} p(\alpha_i, \beta_j; x) \quad (i=1,\ldots,l), \tag{5-126}$$

$$p(\beta_j; x) = \sum_{i=1}^{l} p(\alpha_i, \beta_j; x) \quad (j=1,\ldots,m), \tag{5-127}$$

$$\sum_{i=1}^{l} \sum_{j=1}^{m} p(\alpha_i, \beta_j; x) = \sum_{i=1}^{l} p(\alpha_i; x) = \sum_{j=1}^{m} p(\beta_j; x) = 1. \tag{5-128}$$

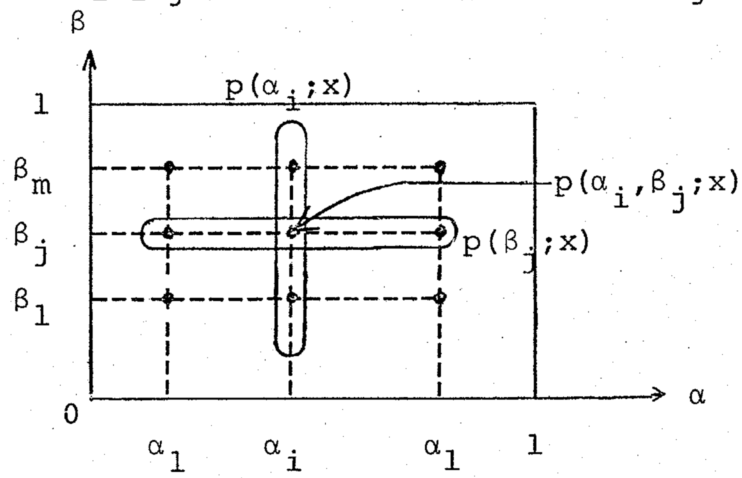


Fig.5-6. A relation between evaluation values and their probability distribution.

If we regard the various evaluations (5-117)(5-118) as a result obtained by 4lm-partition of Ω (5-119) (5-122), then its complexity is measured by a notion of Shannon's entropy, and we denote it by H(x;A,B),

$$H(x;A,B) = -\sum_{i=1}^{l}\sum_{j=1}^{m}\{P(\Omega_{11}^{ij}(x))\cdot\log_2 P(\Omega_{11}^{ij}(x)) + $$
$$+ P(\Omega_{10}^{ij}(x))\cdot\log_2 P(\Omega_{10}^{ij}(x)) + P(\Omega_{01}^{ij}(x))\cdot\log_2 P(\Omega_{01}^{ij}(x))$$
$$+ P(\Omega_{00}^{ij}(x))\cdot\log_2 P(\Omega_{00}^{ij}(x))\}. \quad (5\text{-}129)$$

We call H(x;A,B) a <u>subjective entropy</u> of x with respect to A and B. It should be noted that we could not calculate H(x;A,B) (5-129) directly, because the 4lm-partition of Ω (5-119)(5-122) existed in observers' subconscious and in general its structure was unknown. But we can estimate it by using evaluation values α_i's, β_j's and their probability, both of which are known.

[Theorem 5-6]

$$H(x;A,B) = -\sum_{i=1}^{l}\sum_{j=1}^{m}\{\alpha_i\beta_j p(\alpha_i,\beta_j;x)\cdot\log_2\alpha_i\beta_j p(\alpha_i,\beta_j;x) + $$
$$+ \alpha_i(1-\beta_j)p(\alpha_i,\beta_j;x)\cdot\log_2\alpha_i(1-\beta_j)p(\alpha_i,\beta_j;x) + $$
$$+ (1-\alpha_i)\beta_j p(\alpha_i,\beta_j;x)\cdot\log_2(1-\alpha_i)\beta_j p(\alpha_i,\beta_j;x) + $$
$$+ (1-\alpha_i)(1-\beta_j)p(\alpha_i,\beta_j;x)\cdot\log_2(1-\alpha_i)(1-\beta_j)p(\alpha_i,\beta_j;x)\} \quad (5\text{-}130)$$

$$= -\sum_{i=1}^{l}\sum_{j=1}^{m}p(\alpha_i,\beta_j;x)\cdot\log_2 p(\alpha_i,\beta_j;x) + $$
$$+ \sum_{i=1}^{l}p(\alpha_i;x)\cdot h(\alpha_i) + \sum_{j=1}^{m}p(\beta_j;x)\cdot h(\beta_j). \quad (5\text{-}131)$$

<u>proof</u>

Firstly, it should be noted that the parameter space (Ω, B,P) was a probability space. We can consider a conditional probability measure by $\Omega^{ij}(x)$ (5-122), and regard $\Omega^{ij}(x)$ as a sub-probability space, where its probability measure is defined as the following $P_{\Omega^{ij}(x)}(\cdot)$,

$$P_{\Omega^{ij}(x)}(\cdot) = P(\cdot \cap \Omega^{ij}(x)) / P(\Omega^{ij}(x)). \quad (5\text{-}132)$$

Since the partition of $\Omega^{ij}(x)$ (5-122) is a partition of fundamental parameters, we may consider it an independent partition with respect to $P_{\Omega^{ij}(x)}(\cdot)$. Then we obtain, for example,

$$P_{\Omega^{ij}(x)}(\Omega_{11}^{ij}(x)) = P_{\Omega^{ij}(x)}(\Omega_{11}^{ij}(x) \cup \Omega_{10}^{ij}(x)) \cdot$$
$$\cdot P_{\Omega^{ij}(x)}(\Omega_{11}^{ij}(x) \cup \Omega_{01}^{ij}(x)). \qquad (5\text{-}133)$$

Since we have, from (5-132),(5-124) and (5-125),

$$P(\Omega_{11}^{ij}(x)) / P(\Omega^{ij}(x)) = \alpha_i \beta_j, \qquad (5\text{-}134)$$

we can conclude

$$P(\Omega_{11}^{ij}(x)) = \alpha_i \cdot \beta_j \cdot p(\alpha_i, \beta_j; x). \qquad (5\text{-}135)$$

In much the same way as (5-135), we obtain

$$P(\Omega_{10}^{ij}(x)) = \alpha_i (1-\beta_j) p(\alpha_i, \beta_j; x), \qquad (5\text{-}136)$$

$$P(\Omega_{01}^{ij}(x)) = (1-\alpha_i) \beta_j p(\alpha_i, \beta_j; x), \qquad (5\text{-}137)$$

$$P(\Omega_{00}^{ij}(x)) = (1-\alpha_i)(1-\beta_j) p(\alpha_i, \beta_j; x). \qquad (5\text{-}138)$$

Rearranging (5-129) by substituting (5-135)∿ (5-138) into, we have (5-131) and (5-132).

(Q.E.D.)

The validity of the definition of H(x;A,B) is confirmed in much the same way as the case of a single probabilistic set (cf. the last part of section 5-3). And an argument of H(x;A,B) in infinite states ($l \to \infty$, $m \to \infty$) is also possible. However, all of them are omitted here.

5-8. MUTUAL SUBJECTIVE ENTROPY.

We investigate the interaction of two evaluations from a viewpoint of probability theory and entropy theory. Let $p(\alpha_i,\beta_j;x)$ be a probability of x being evaluated α_i from a viewpoint of A and β_j from that of B (cf. (5-120),(5-121)). If a relation,

$$p(\alpha_i,\beta_j;x)= p(\alpha_i;x)\cdot p(\beta_j;x), \tag{5-139}$$

holds, then two evaluations are given independently and there is no interaction between the two. Taking the characteristic of our sensorium into consideration, we express the degree of interaction of two phenomena on the following log scale,

$$\log_2 p(\alpha_i,\beta_j;x)/(p(\alpha_i;x)\cdot p(\beta_j;x)). \tag{5-140}$$

(If this value equals zero, then the evaluations are given independently.) We denote the average value of this quantity by I(x;A,B),

$$I(x;A,B)= \sum_{i=1}^{l}\sum_{j=1}^{m} p(\alpha_i,\beta_j;x)\cdot\log_2 p(\alpha_i,\beta_j;x)/ /(p(\alpha_i;x)\cdot p(\beta_j;x)), \tag{5-141}$$

and call it a <u>mutual subjective entropy</u> of x with respect to A and B.

The symmetricity is easily confirmed,

$$I(x;A,B)= I(x;B,A). \tag{5-142}$$

Moreover, if we apply the well-known inequality of probability vectors [3], i.e.

$$\Sigma p_i= \Sigma q_i= 1, \quad p_i,q_i\geqq 0$$

$$\Longrightarrow \Sigma\, q_i\log_2(1/p_i)\geqq \Sigma q_i\log_2(1/q_i) \tag{5-143}$$

(the equality holds iff $p_i= q_i$ for all i),

then we have

$$I(x;A,B)\geqq 0, \tag{5-144}$$

(the equality is valid iff (5-139) holds for all i and j). Hence, we can conclude that as the value of I(x;A,B) increases, the degree of interaction also increases further.

We have already introduced four quantities, i.e. H(x;A), H(x;B), H(x;A,B) and I(x;A,B). Fig.5-7 shows a mutual relationship of them.

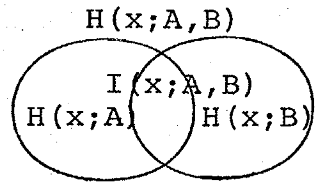


Fig.5-7. A Venn diagram of various subjective entropy of two phenomena.

A direct calculation shows

$$H(x;A) + H(x;B) = H(x;A,B) + I(x;A,B). \quad (5\text{-}145)$$

What is evident from (5-145) is that

1) the sum of two subjective entropy of a single phenomenon (i.e. the left side) is greater than the subjective entropy of two phenomena H(x;A,B).

2) the difference is given by I(x;A,B) and it indicates the degree of interaction of two phenomena.

5-9. ANALYSIS OF INTERACTION BASED ON VARIOUS SUBJECTIVE ENTROPY.

As we have shown in Fig.5-7, we defined four various notions of subjective entropy. As is evident from Fig.5-7, other four different combinations are possible, since the total area is divided into three parts and there exist $2^3=8$ combinations.

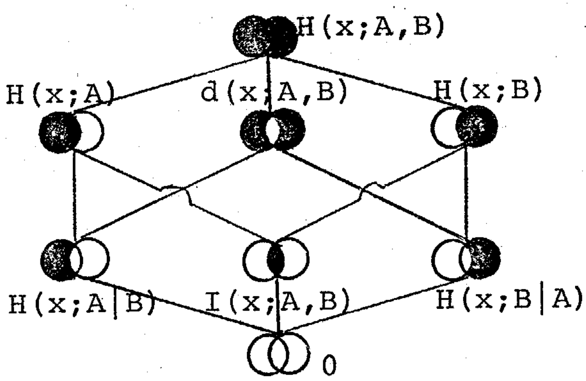


Fig.5-7. A Hasse diagram of 8 different subjective entropy.

Fig.5-7 shows all of these 8 notions.

We shall give a definition about them in the following.

[Def. 5-2]

<u>conditional subjective entropy</u>

$$H(x;A|B) = H(x;A,B) - H(x;B) \tag{5-146}$$

$$= \sum_{j=1}^{m} p(\beta_j;x)\{ - \sum_{i=1}^{l} p(\alpha_i|\beta_j;x)\cdot\log_2 p(\alpha_i|\beta_j;x) + $$
$$+ \sum_{i=1}^{l} p(\alpha_i|\beta_j;x)\cdot h(\alpha_i) \quad (\geqq 0), \tag{5-147}$$

where

$$p(\alpha_i|\beta_j;x) \underset{\Delta}{=} p(\alpha_i,\beta_j;x)/\,p(\beta_j;x). \tag{5-148}$$

<u>dispersion</u>

$$d(x;A,B) = H(x;A|B) + H(x;B|A) \tag{5-149}$$

$$= 2\cdot H(x;A,B) - H(x;A) - H(x;B) \tag{5-150}$$

$$= - \sum_{i=1}^{l} \sum_{j=1}^{m} p(\alpha_i,\beta_j;x)\cdot\log_2 p(\alpha_i|\beta_j;x)p(\beta_j|\alpha_i;x) + $$
$$+ \sum_{i=1}^{l} p(\alpha_i;x)\cdot h(\alpha_i) + \sum_{j=1}^{m} p(\beta_j;x)\cdot h(\beta_j) \quad (\geqq 0). \tag{5-151}$$

Several mathematical properties are summarized in the following proposition.

[Prop.5-1]

1) $H(x;A|A)= \sum_{i=1}^{l} p(\alpha_i;x)\cdot h(\alpha_i),$ (5-152)

2) $d(x;A,A)= 2\cdot H(x;A|A) = 2 \sum_{i=1}^{l} p(\alpha_i;x)\cdot h(\alpha_i),$ (5-153)

3) $I(x;A,B)= H(x;A) - H(x;A|B)= H(x;B) - H(x;B|A),$ (5-154)

4) $H(x;A) + H(x;B) \geqq H(x;A,B),$ (5-155)

5) $H(x;A,B)\geqq H(x;A),$ (5-156)

6) $H(x;A)\geqq H(x;A|B),$ (5-157)

7) $I(x;A,A)= H(x;A) - H(x;A|A)$

$= - \sum_{i=1}^{l} p(\alpha_i;x)\cdot \log_2 p(\alpha_i;x),$ (5-158)

8) $d(x;A,B)= d(x;B,A),$ (5-159)

9) $d(x;A,B) + d(x;B,C)\geqq d(x;A,C),$ (5-160)

10) $d(x;A,B)= \frac{1}{2}\{d(x;A,A) + d(A;B,B)\}.$ (5-160)

proof

They are easily verified from each definition. Hence, all of them are omitted.

(Q.E.D.)

The important points from the above proposition are summarized as follows;

1) Although all of these notions are an expansion of Shannon's theory, there are several differences between the two. For example, $H(x;A|A)$ and $d(x;A,A)$ are not necessaily zero. Since we consider ambiguous evaluations other than 0 and 1, the value itself has ambiguity and abovementioned effects arise.

2) In Shannon's theory, a corresponding notion to dispersion provides a distance measure [4], i.e. it satisfies positivity, symmetricity and trigonometric law. However, dispersion $d(x;\cdot,\cdot)$ is not a distance measure, since it doesn't satisfies positivity (cf. (5-153)). The value $d(x;A,A)$, which is not always zero, provides the dispersiveness of x from a viewpoint of A.

An example will be given to clarify the explanation done so far. Let an object x be an ambiguous character as shown in Fig.5-8. Consider the following three probabilistic sets : A= "characters which look like a character a", B= "characters which look like a character b", R= " characters which look like neither a nor b (or rejected characters)". Let a parameter space Ω be the same one mentioned in (5-2) (5-3). Then defining functions $\mu_A(x,\omega_i)$, $\mu_B(x,\omega_i)$, $\mu_R(x,\omega_i)$ are given, for example, as tabulated in Table 5-2. We have defined various operations in $P(X)$ such as union and intersection. Several results are also tabulated in Table 5-2.

x =

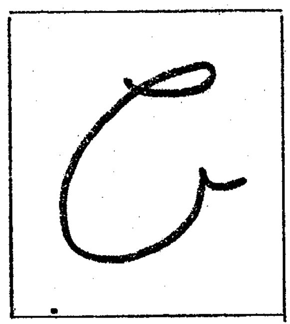


Fig.5-8. An ambiguous character.

As seen from this table, the example is made up very skillfully, and we can not find any differences among three probabilistic sets A,B and R from viewpoints of probabilistic expression, extended fuzzy expression and subjective entropy. However, a different circumstance comes up, if we take mutual interaction into consideration. Various subjective entropy of plural probabilistic sets are shown in Table 5-3 and Fig.5-9. It can be seen from Table 5-3 and Fig.5-9 that

1) the interaction between A and R is the same as that between B and R,

2) the interaction between A and B is different from abovestated ones.

S \ i	defining function $\mu_S(x,\omega_i)$ 1	2	3	4	5	6	7	8	membership function $m_S(x)$	vagueness function $v_S(x)$	subjective entropy $H(x,S)$
A∩B∩R	0.	0.5	0.25	0.25	0.5	0.25	0.	0.	0.21875	0.02014	2.11551
A∩B	0.	0.5	0.5	0.25	0.5	0.25	0.5	0.	0.3125	0.04297	2.2028
A∩R	0.	0.5	0.25	0.25	0.5	0.5	0.	0.5	0.3125	0.04297	2.2028
B∩R	0.5	0.5	0.25	0.5	0.5	0.25	0.	0.	0.3125	0.04297	2.2028
A	0.	0.5	0.5	0.25	0.5	0.75	0.5	1.	0.5	0.07813	2.7028
B	1.	0.5	0.5	0.75	0.5	0.25	0.5	0.	0.5	0.07813	2.7028
R	0.5	0.75	0.25	0.5	1.	0.5	0.	0.5	0.5	0.07813	2.7028
A∪B	1.	0.5	0.5	0.75	0.5	0.75	0.5	1.	0.6875	0.04297	2.2028
A∪R	0.5	0.75	0.5	0.5	1.	0.75	0.5	1.	0.6875	0.04297	2.2028
B∪R	1.	0.75	0.5	0.75	1.	0.5	0.5	0.5	0.6875	0.04297	2.2028
A∪B∪R	1.	0.75	0.5	0.75	1.	0.75	0.5	1.	0.78125	0.02014	2.11551

Table 5-2. Another numerical example of probabilistic sets.

· \ :	A	B	R
A	3.4056	3.4056	4.4056
B	3.4056	3.4056	4.4056
R	4.4056	4.4056	3.4056

(a) $H(x;\cdot,:)$

· \ :	A	B	R
A	2.	2.	1.
B	2.	2.	1.
R	1.	1.	2.

(b) $I(x;\cdot,:)$

· \ :	A	B	R
A	0.7028	0.7028	1.7028
B	0.7028	0.7028	1.7028
R	1.7028	1.7028	0.7028

(c) $H(x;\cdot|:)$

· \ :	A	B	R
A	1.4056	1.4056	3.4056
B	1.4056	1.4056	3.4056
R	3.4056	3.4056	1.4056

(d) $d(x;\cdot,:)$

Table 5-3. Various subjective entropy of the example given in Table 5-2.

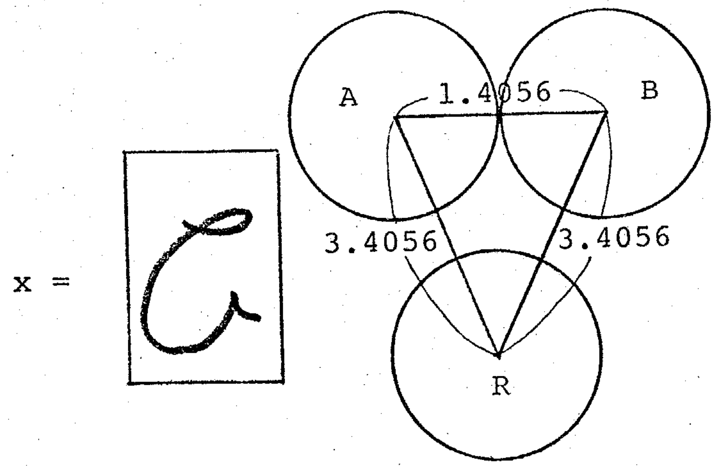


Fig. 5-9. Dispersion of the object x
in Fig.5-8.

5-10. A RESULT OF A QUESTIONNAIRE.

We shall reconsider the questionnaire mentioned in section 2-6 from a viewpoint of subjective entropy. Several results are shown in Table 5-4, Fig.5-10, Fig.5-11 and Fig.5-12.

L \ R	1	2	3	4	5	6	7	8	9	10	11	12
1	0.000	2.640	2.275	0.136	0.000	1.329	0.550	0.258	0.122	2.637	1.566	1.601
2	2.706	0.615	2.766	0.930	0.547	1.866	1.532	1.209	1.717	2.308	2.107	1.508
3	2.020	2.659	0.607	2.741	2.410	2.454	1.881	1.436	0.285	1.978	2.363	2.285
4	0.122	0.968	2.632	0.346	1.562	1.264	1.131	0.402	0.258	0.465	0.528	0.327
5	0.000	0.276	2.448	1.525	0.000	2.768	0.569	0.550	0.000	0.000	0.000	0.000
6	1.085	2.082	2.168	2.120	2.502	0.610	2.273	1.651	0.785	1.456	1.412	1.508
7	0.465	1.492	2.057	1.678	0.819	2.598	0.122	1.738	2.380	2.573	1.590	1.651
8	0.681	1.904	1.509	0.921	0.238	1.286	1.349	0.139	1.213	2.639	1.014	1.318
9	0.122	2.092	0.750	0.000	0.136	0.920	2.358	1.427	0.136	2.864	1.106	1.099
10	2.332	2.189	2.319	0.806	0.136	1.366	2.258	2.150	2.471	0.494	1.325	1.641
11	2.035	1.876	2.367	0.465	0.326	1.534	1.318	0.383	0.802	1.749	0.000	0.550
12	1.709	1.762	2.410	0.238	0.421	1.761	1.326	1.051	0.836	1.803	0.528	0.327

Table 5-4. Subjective entropy of the questionnaire.

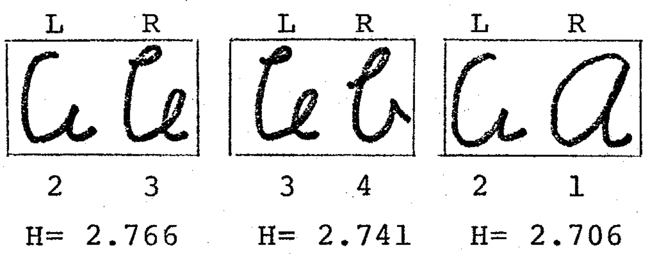


Fig.5-10. Ambiguous characters whose subjective entropy is greater than 2.7. (a maximum= 2.91 by Th.5-4)

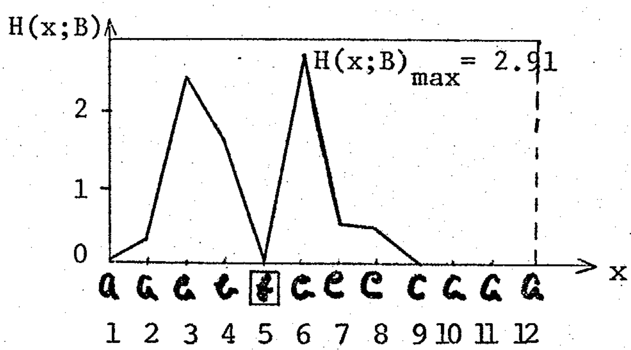


Fig.5-11. Subjective entropy H(x;B).

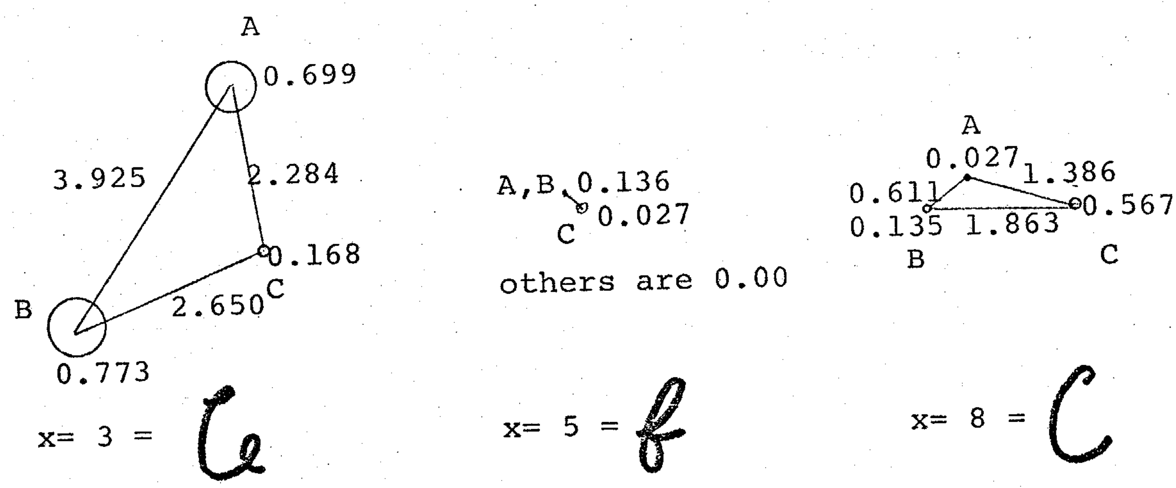


Fig.5-12. Dispersion d(x;·,·).

CHAPTER. 6

THEORY OF SUBJECTIVE TOPOLOGY.

6-1. INTRODUCTION.

We have been investigating a concept of probabilistic sets, which is an extension of ordinary set theory and classical fuzzy set theory.

In ordinary set theory, various studies have been developed based on it such as theory of general topology [1], theory of algebra [2] and measure theory [3]. As applications of fuzzy set theory, a lot of analogous studies have also been done [4] [5] [6] [7]. Among these various applications, it is very important to consider topological structure especially in the field of pattern recognition. The purpose of the present chapter is to study fundamentals of topological structure of probabilistic sets. It is summarized as a concept of subjective topology.

Subjective topology of probabilistic sets is introduced by five different methods; neighbourhood system of one point probabilistic sets, system of open sets, system of closed sets, closure operator, and interior operator. These five different definitions of subjective topology are shown to be mutually equivalent by giving a concrete way to induce one concept from another. Two other things are also mentioned; classification of probabilistic sets from a viewpoint of subjective topology, and ordinal structure of subjective topology.

6-2. INTRODUCTION TO SUBJECTIVE TOPOLOGY.

We introduce a notion of subjective topology in $\mathcal{P}(X)$ (a family of probabilistic sets on a total space X) by five different methods, and investigate their mutual relationships.

[Def.6-1] (neighbourhood system)

Suppose a sub-family of probabilistic sets $\mathcal{V}(A_x)$ $(\subset \mathcal{P}(X))$ is defined for each one point probabilistic set A_x (cf.(3-101)), where $\mathcal{V}(A_x)$ satisfies,

(NB.1) $X \varepsilon \mathcal{V}(A_x)$, (6-1)

(NB.2) $U \varepsilon \mathcal{V}(A_x) \Longrightarrow A_x \subset U$, (6-2)

(NB.3) $U \varepsilon \mathcal{V}(A_x) \Longrightarrow {}^{\forall}A_x' \subset A_x,\quad U \varepsilon\ \mathcal{V}(A_x')$, (6-3)

(NB.4) $U, V \varepsilon \mathcal{V}(A_x) \Longrightarrow U \cap V\ \ \varepsilon\ \ \mathcal{V}(A_x)$, (6-4)

(NB.5) $U \varepsilon \mathcal{V}(A_x),\ U \subset V \Longrightarrow V\ \varepsilon\ \mathcal{V}(A_x)$, (6-5)

(NB.6) $U \varepsilon \mathcal{V}(A_x) \Longrightarrow {}^{\exists}V(\subset U) \varepsilon\ \mathcal{V}(A_x)$ such that

$${}^{\forall}A_y \subset V,\quad {}^{\exists}A_y' \subset A_y,\ U \varepsilon \mathcal{V}(A_y'). \qquad (6\text{-}6)$$

Each element of $\mathcal{V}(A_x)$ is called a neighbourhood of A_x, and a family $\{\mathcal{V}(A_x)\}_{A_x \subset X}$ is called a neighbourhood system. A space X is called a subjective topological space (or briefly, topological space) if a neighbourhood system $\{\mathcal{V}(A_x)\}_{A_x \subset X}$ is given.

[Def.6-2] (system of open sets)

Suppose a sub-family $\mathcal{O}(\subset \mathcal{P}(X))$ is given, where $\mathcal{O}$ satisfies

(O.1) $\phi,\ X \varepsilon\ \mathcal{O}$, (6-7)

(O.2) $O_1, O_2 \varepsilon\ \mathcal{O} \Rightarrow O_1 \cap O_2 \varepsilon\ \mathcal{O}$, (6-8)

(O.3) $O_\gamma \varepsilon\ \mathcal{O}\ (\gamma \varepsilon \Gamma) \Longrightarrow \bigcup_{\gamma \varepsilon \Gamma} O_\gamma \varepsilon\ \mathcal{O}$. (6-9)

Then $\mathcal{O}$ is called a system of opensets and each element of $\mathcal{O}$ is called an open set. A pair $(X,\mathcal{O})$ is called a subjective topological space.

[Def. 6-3] (system of closed sets)

Suppose a sub-family $\mathcal{C}(\subset P(X))$ is given, where $\mathcal{C}$ satisfies

(C.1) $X, \phi \ \varepsilon\ \mathcal{C}$, (6-10)

(C.2) $C_1, C_2 \ \varepsilon\ \mathcal{C} \Longrightarrow C_1 \cup C_2 \ \varepsilon\ \mathcal{C}$, (6-11)

(C.3) $C_\gamma \ \varepsilon\ \mathcal{C}\ (\gamma \varepsilon \Gamma) \Longrightarrow \bigcap_{\gamma \varepsilon \Gamma} C_\gamma \ \varepsilon\ \mathcal{C}$. (6-12)

Then $\mathcal{C}$ is called a <u>system of closed sets</u> and each element of $\mathcal{C}$ is called a <u>closed set</u>. A pair $(X, \mathcal{C})$ is called a <u>subjective topological space</u>.

[Def. 6-4] (closure operator)

Suppose the following correspondence $\cdot^a$ is given on $P(X)$,

$$\cdot^a : \begin{array}{ccc} P(X) & \longrightarrow & P(X), \\ \omega & & \omega \\ A & \longmapsto & A^a \end{array} \qquad (6\text{-}13)$$

(CO.1) $\phi^a = \phi$, (6-14)

(CO.2) $A \subset A^a$, (6-15)

(CO.3) $(A \cup B)^a = A^a \cup B^a$, (6-16)

(CO.4) $A^{aa} = A^a$. (6-17)

Then X is called a subjective topological space based on a <u>closure operator</u> $\cdot^a$. A probabilistic set A^a is called a <u>closure</u> (abgeshlossene Hülle, in German) of A.

[Def. 6-5] (interior operator)

Suppose the following correspondence $\cdot^i$ is given on $P(X)$,

$$\cdot^i : \begin{array}{ccc} P(X) & \longrightarrow & P(X), \\ \omega & & \omega \\ A & \longmapsto & A^i \end{array} \qquad (6\text{-}18)$$

(IO.1) $X^i = X$, (6-19)

(IO.2) $A^i \subset A$, (6-20)

(IO.3) $(A \cap B)^i = A^i \cap B^i$, (6-21)

(IO.4) $A^{ii} = A^i$. (6-22)

Then X is called a subjective topological space based on an interior operator $\cdot^{i}$. A probabilistic set A^{i} is called an interior of A.

We defined subjective topology by five different notions. However, we can infer that there exists duality between open sets and closed sets, and between the closure operator and the interior operator. Moreover, we can verify that each of these five notions can be derived from others, i.e. five notions are mutually equivalent.

[Prop.6-1]

Suppose that the subjective topology on X is defined by a neighbourhood system as in def.6-1. If we put

$$\mathcal{O} \underset{\Delta}{=} \{O \mid O \in \mathcal{P}(X),\ A_x \subset O \Rightarrow {}^{\exists}A_x' \subset A_x \text{ such that } O \in \mathcal{V}(A_x')\}, \tag{6-23}$$

then $\mathcal{O}$ satisfies the axiom of open sets in def.6-2.

proof

(O.1) Since $\phi \in \mathcal{P}(X)$ and since there exists no one point probabilistic set A_x which is included in ϕ, we have $\phi \in \mathcal{O}$. Since $X \in \mathcal{P}(X)$ and $X \in \mathcal{V}(A_x)$ for all A_x by (NB.1), we have $X \in \mathcal{O}$.

(O.2) Suppose $O_1, O_2 \in \mathcal{O}$. Then $O_1 \cap O_2 \in \mathcal{P}(X)$. For all $A_x \subset O_1 \cap O_2$, we have an $A_x'(\subset A_x)$ such that $O_1 \in \mathcal{V}(A_x')$ (note $A_x \subset O_1$). Since $A_x' \subset A_x \subset O_1 \cap O_2 \subset O_2$, there exists an $A_x''(\subset A_x')$ such that $O_2 \in \mathcal{V}(A_x'')$. By (N.B.3), we obtain $O_1 \in \mathcal{V}(A_x'')$. Hence we also obtain $O_1 \cap O_2 \in \mathcal{V}(A_x'')$ by (N.B.4). And we conclude $O_1 \cap O_2 \in \mathcal{O}$.

(O.3) Suppose $O_\gamma \in \mathcal{O}(\gamma \in \Gamma)$. Then $\bigcup_{\gamma \in \Gamma} O_\gamma \in \mathcal{P}(X)$. For all $A_x \subset \bigcup_{\gamma \in \Gamma} O_\gamma$, there exists a $\gamma \in \Gamma$ such that $A_x' \underset{\Delta}{=} A_x \cap O_\gamma \neq \phi$. (If $A_x \cap O_\gamma = \phi$

for all γ, then $\bigcup_{\gamma\varepsilon\Gamma}(A_x \cap O_\gamma) = A_x \cap \bigcup_{\gamma\varepsilon\Gamma} O_\gamma = \phi$ and we have a contradiction.) Hence we obtain $A_x' \subset O_\gamma \varepsilon \mathcal{O}$, and there exists an $A_x'' \subset A_x'$ such that $O_\gamma \varepsilon \mathcal{V}(A_x'')$. By considering (NB.5) and $O_\gamma \subset \bigcup_{\gamma\varepsilon\Gamma} O_\gamma$, we obtain $\bigcup_{\gamma\varepsilon\Gamma} O_\gamma \varepsilon \mathcal{V}(A_x'')$. We conclude $\bigcup_{\gamma\varepsilon\Gamma} O_\gamma \ \varepsilon \ \mathcal{O}$.

(Q.E.D.)

[Prop.6-2]

Suppose that a subjective topology is defined by a system of open sets as in def.6-2. If we put

$$\mathcal{V}(A_x) \underset{\Delta}{=} \{U \mid \exists O \varepsilon \mathcal{O} \ \text{ s.t. } A_x \subset O \subset U \ \varepsilon \ P(X)\}, \qquad (6\text{-}24)$$

for each one point probabilistic set A_x, then $\mathcal{V}(A_x)$ satisfies the axiom of neighbourhood system mentioned in def.6-1.

<u>proof</u>

(NB.1) Since $X \ \varepsilon \ \mathcal{O}$(by (O.1)), we have $A_x \subset X(\varepsilon \ \mathcal{O}) \subset X$. Hence,

$X \ \varepsilon \ \mathcal{V}(A_x)$.

(NB.2) If $U \varepsilon \mathcal{V}(A_x)$, then there exists $O \varepsilon \mathcal{O}$ such that $A_x \subset O \subset U \ \varepsilon \ P(X)$.

Hence, $A_x \subset U$.

(NB.3) Suppose $U \ \varepsilon \ \mathcal{V}(A_x)$ and $A_x' \subset A_x$, then there exists $O \ \varepsilon \ \mathcal{O}$ such that $A_x'(\subset A_x) \subset O \subset U \ \varepsilon \ P(X)$. Therefore we have $U \ \varepsilon \ \mathcal{V}(A_x')$.

(NB.4) Suppose $U, V \ \varepsilon \ \mathcal{V}(A_x)$, then there exist $O_1, O_2 \varepsilon \mathcal{O}$ such that $A_x \subset O_1 \subset U \ \varepsilon \ P(X)$ and $A_x \subset O_2 \subset V \ \varepsilon \ P(X)$. Whereas $O_1 \cap O_2 \varepsilon \mathcal{O}$ (by (O.2)) and $A_x \subset O_1 \cap O_2 \subset U \cap V \ \varepsilon P(X)$. Hence we have $U \cap V \ \varepsilon \ \mathcal{V}(A_x)$.

(NB.5) If $U \ \varepsilon \ \mathcal{V}(A_x)$ and $U \subset V$, then we have $O \ \varepsilon \ \mathcal{O}$ such that $A_x \subset O \subset U$. Therefore $A_x \subset O \subset V$. And we conclude $V \ \varepsilon \ \mathcal{V}(A_x)$.

(NB.6) Suppose $U \ \varepsilon \ \mathcal{V}(A_x)$, then we have $O \ \varepsilon \ \mathcal{O}$ such that $A_x \subset O \subset U \ \varepsilon \ P(X)$. If we put $V \underset{\Delta}{=} O$, then we obtain $A_x \subset O = V \ \varepsilon \ P(X)$, and therefore $V \ \varepsilon \ \mathcal{V}(A_x)$. On the other hand, we have $A_y \subset V = O \subset U \ \varepsilon \ P(X)$ for each $A_y \subset V$. Hence we conclude $U \ \varepsilon \ \mathcal{V}(A_y)$.

(Q.E.D.)

[Prop.6-3]

Suppose that a subjective topology is defined by a system of open sets as in def.6-2, and define

$$\mathcal{C} \underset{\Delta}{=} \{C \mid C \in P(X),\ C^c \in \mathcal{O}\}. \tag{6-25}$$

Then $\mathcal{C}$ satisfies the axiom of closed sets in def.6-3.

proof

(C.1) We have $\phi, X \in \mathcal{O}$ by (O.1). Hence, $\phi = X^c$, $X = \phi^c \in \mathcal{O}$, i.e. $X, \phi \in \mathcal{C}$.

(C.2) Suppose $C_1, C_2 \in \mathcal{C}$, i.e. $C_1{}^c, C_2{}^c \in \mathcal{O}$. Then $C_1{}^c \cap C_2{}^c \in \mathcal{O}$ by (O.2). On the other hand $C_1{}^c \cap C_2{}^c = (C_1 \cup C_2)^c$. Hence, $(C_1 \cup C_2)^c \in \mathcal{O}$ and we conclude $C_1 \cup C_2 \in \mathcal{C}$.

(C.3) Suppose $C_\gamma \in \mathcal{C}$ $(\gamma \in \Gamma)$, i.e. $C_\gamma{}^c \in \mathcal{O}$. From (O.3), we have $\bigcup_{\gamma \in \Gamma} C_\gamma{}^c \in \mathcal{O}$. Whereas $\bigcup_{\gamma \in \Gamma} C_\gamma{}^c = (\bigcap_{\gamma \in \Gamma} C_\gamma)^c$. Hence we have $(\bigcap_{\gamma \in \Gamma} C_\gamma)^c \in \mathcal{O}$ and we obtain $\bigcap_{\gamma \in \Gamma} C_\gamma \in \mathcal{C}$.

(Q.E.D.)

[Prop.6-4]

Suppose that a subjective topology is defined by a system of closed sets as in def.6-3. Put

$$\mathcal{O} \underset{\Delta}{=} \{O \mid O \in P(X),\ O^c \in \mathcal{C}\}. \tag{6-26}$$

Then $\mathcal{O}$ satisfies the axiom of open sets defined in def.6-2.

proof

It will be easily verified in almost the same manner as Prop.6-3.

(Q.E.D.)

[Prop.6-5]

Suppose that a subjective topology is introduced by the closure operator in def.6-4. If we define

$$\mathcal{C} \underset{\Delta}{=} \{C \mid C \in P(X),\ C^a = C\}, \tag{6-27}$$

then $\mathcal{C}$ satisfies the axiom of closed sets in def.6-3.

proof

(C.1) From (CO.2) we have $X \subset X^a$. Since X is a maximum element in $P(X)$, we obtain $X^a \subset X$. Hence we have $X = X^a$, and $X \varepsilon \mathcal{C}$.

(C.2) Suppose $C_1, C_2 \varepsilon \mathcal{C}$ i.e. $C_1{}^a = C_1$, $C_2{}^a = C_2$. By applying (CO.3), we obtain $(C_1 \cup C_2)^a = C_1{}^a \cup C_2{}^a = C_1 \cup C_2 \varepsilon P(X)$, and $C_1 \cup C_2 \varepsilon \mathcal{C}$.

(C.3) Suppose $C_\gamma \varepsilon \mathcal{C} (\gamma \varepsilon \Gamma)$, i.e. $C_\gamma{}^a = C_\gamma$ for all $\gamma \varepsilon \Gamma$. In general, we have $A^a \subset B^a$ from an assumption of $A \subset B$. (Since $B = B \cup A$, we have, by (CO.3), $B^a = (B \cup A)^a = B^a \cup A^a \supset A^a$.) Since $\bigcap_{\gamma \varepsilon \Gamma} C_\gamma \subset C_\gamma$, we have $(\bigcap_{\gamma \varepsilon \Gamma} C_\gamma)^a \subset C_\gamma{}^a = C_\gamma$ for all γ. Hence, we obtain $(\bigcap_{\gamma \varepsilon \Gamma} C_\gamma)^a \subset \bigcap_{\gamma \varepsilon \Gamma} C_\gamma$. On the other hand we have $\bigcap_{\gamma \varepsilon \Gamma} C_\gamma \subset (\bigcap_{\gamma \varepsilon \Gamma} C_\gamma)^a$ by (CO.2). Therefore $(\bigcap_{\gamma \varepsilon \Gamma} C_\gamma)^a = \bigcap_{\gamma \varepsilon \Gamma} C_\gamma \varepsilon P(X)$, and we conclude $\bigcap_{\gamma \varepsilon \Gamma} C_\gamma \varepsilon \mathcal{C}$.

(Q.E.D.)

[Prop.6-6]

Suppose that a subjective topology is defined by a system of closed sets as in def.6-3. Define a mapping $\cdot^a$ by

$$\begin{aligned} \cdot^a : P(X) &\longrightarrow P(X). \\ A &\longmapsto A^a \underset{\Delta}{=} \bigcap \{C \mid A \subset C \varepsilon \mathcal{C}\} \end{aligned} \qquad (6\text{-}28)$$

(A^a is defined as the smallest closed set that contains A.)

Then $\cdot^a$ satisfies the axiom of a closure operator in def.6-4.

proof

(CO.1)

Since we have $\phi \varepsilon \mathcal{C}$ by (C.1), we obtain $\phi^a = \phi$.

(CO.2)

By a definition of (6-28), we have $A \subset A^a$ for all $A \varepsilon P(X)$.

(CO.3)

$(A \cap B)^a = \bigcap \{C \mid (A \cup B) \subset C \varepsilon \mathcal{C}\} = \bigcap \{C \mid A \subset C, B \subset C, C \varepsilon \mathcal{C}\}$

$$\supset \begin{cases} \bigcap\{C \mid A \subset C \in \mathcal{C}\} = A^a \\ \bigcap\{C \mid B \subset C \in \mathcal{C}\} = B^a. \end{cases}$$

Hence we have $(A \cup B)^a \supset A^a \cup B^a$. On the other hand we have $A^a \in \mathcal{C}$ for each $A \in P(X)$ by the definition (6-28) and (C.3). By using (CO.2) which has already proved, we have $A \subset A^a \subset A^a \cup B^a$ and $B \subset B^a \subset A^a \cup B^a$. Hence $A \cup B \subset A^a \cup B^a \in \mathcal{C}$. Reconsidering (6-28), we obtain $(A \cup B)^a \subset A^a \cup B^a$. We conclude $(A \cup B)^a = A^a \cup B^a$.

(CO.4) From (CO.2), we have $A \subset A^a \subset A^{aa}$ for each $A \in P(X)$. Hence $A^a \subset A^{aa}$. Since $A^a \in \mathcal{C}$, we have $(A^a)^a \subset A^a$ (by the definition (6-28)). Therefore we conclude $A^{aa} = A^a$.

(Q.E.D.)

[Prop.6-7]

If a subjective topology is introduced by an interior operator $\cdot^i$ as in def.6-5, and if we put

$$\mathcal{O} \underset{\Delta}{=} \{O \mid O \in P(X),\ O^i = O\}, \tag{6-29}$$

then $\mathcal{O}$ satisfies the axiom of open sets in def.6-2.

<u>proof</u>

It will be easily proved in much the same way as prop.6-5. In the proof of (O.3), we use the following property,

$$A \subset B \Rightarrow A^i \subset B^i.$$

(Q.E.D.)

[Prop.6-8]

Define a mapping $\cdot^i$ by

$$\begin{array}{rccl} \cdot^i : & P(X) & \longrightarrow & P(X); \\ & \omega & & \omega \\ & A & \longmapsto & A^i \underset{\Delta}{=} \bigcup\{O \mid A \supset O \in \mathcal{O}\} \end{array} \tag{6-30}$$

under an assumption that a subjective topology is introduced by the system of open sets in def.2-2. Then $\cdot^i$ satisfies the axiom of an interior operator. (In (6-30), A^i is defined as the largest open set that is included in A.)

proof

In much the same way as prop.6-6.

(Q.E.D.)

[Prop.6-9]

Suppose that a subjective topology is introduced by a closure operator $\cdot^a$ as in def.6-4. If we define a mapping $\cdot^i$ by

$$\begin{array}{cccc} \cdot^i: & P(X) & \longrightarrow & P(X), \\ & \omega & & \omega \\ & A & \longmapsto & A^i \underset{\Delta}{=} A^{cac} \end{array} \qquad (6\text{-}31)$$

then $\cdot^i$ satisfies the axiom of an interior operator in def.2-5.

proof

(IO.1)

$X^i = X^{cac} = \phi^{ac} = \phi^c = X.$

(IO.2)

For each $A \in P(X)$, $A^c \in P(X)$ and $A^c \subset A^{ca}$ by (CO.2). Hence, we have $A = A^{cc} \supset A^{cac} = A^i$.

(IO.3)

By applying the de Morgan's law (cf. Prop.3-6) and (CO.3), we have $(A \cap B)^i = (A \cap B)^{cac} = (A^c \cup B^c)^{ac} = (A^{ca} \cup B^{ca})^c = A^{cac} \cap B^{cac} = A^i \cap B^i$. Hence we have $(A \cap B)^i = A^i \cap B^i$.

(IO.4)

For each $A \in P(X)$, we have, by (CO.4), $A^{ii} = A^{caccac} = A^{caac} = A^{cac} = A^i$.

(Q.E.D.)

[Prop.6-10]

Define a mapping $\cdot^a$ by

$$\begin{array}{cccc} \cdot^a: & P(X) & \longrightarrow & P(X), \\ & \omega & & \omega \\ & A & \longmapsto & A^a \underset{\Delta}{=} A^{cic} \end{array} \qquad (6\text{-}32)$$

under an assumption that a subjective topology is defined by an interior operator in def.6-5. Then $\cdot^a$ satisfies an axiom of the closure operator in def.6-4.

proof

In almost the same manner as Prop.6-9.

(Q.E.D.)

Abovementioned propositions are summarized in the following theorem.

[Theorem 6-1]

There exist five mutually equivalent definitions of subjective topology; neighbourhood system (def.6-1), system of open sets (def.6-2), system of closed sets (def.6-3), closure operator (def.6-4), interior operator (def.6-5).

proof

The relation is shown in Fig.6-1.

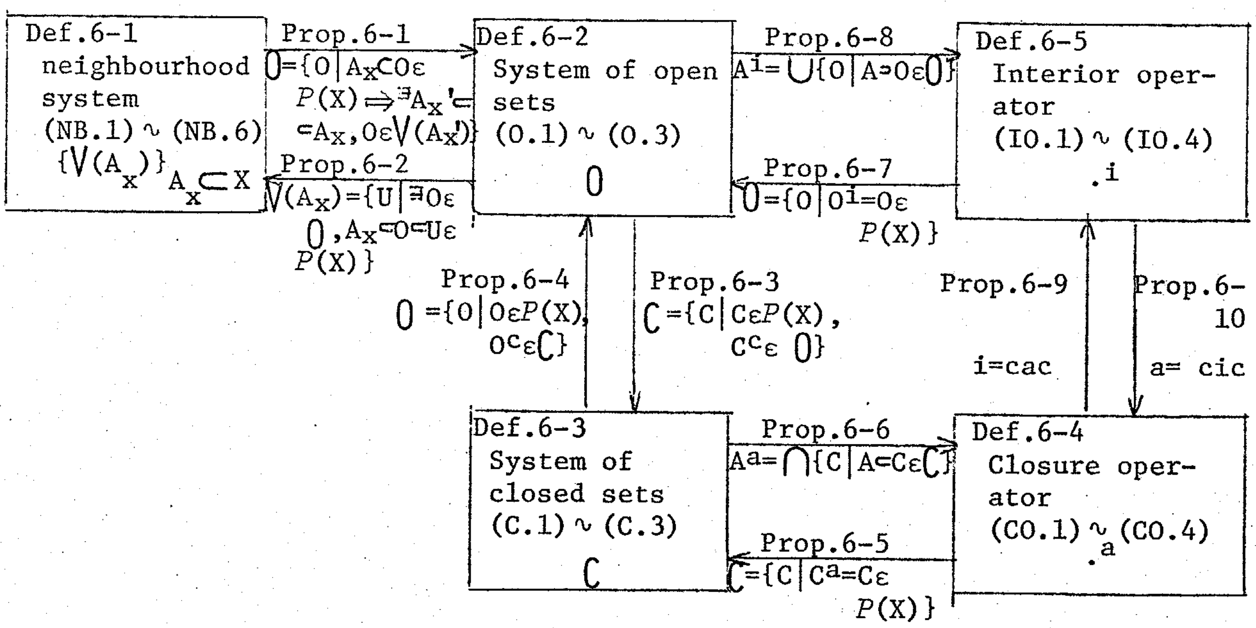


Fig.6-1. The relation among various definitions of subjective topology.

(Q.E.D.)

If a subjective topology is introduced in X by one of the five methods, we call the total space X a <u>subjective topological space</u>, and denote it by (X,τ) where τ indicates the subjective topology.

One thing which we would like to mention at this stage is that we may introduce a subjective topology by other methods which are different from abovestated five definitions. However, it is not necessary to explain from a viewpoint of engineering aspect.

6-3. CLASSIFICATION OF PROBABILISTIC SETS BASED ON SUBJECTIVE TOPOLOGY.

Let (X,τ) be a subjective topological space. Then we can consider three monomial operators on $\mathcal{P}(X)$, i.e. closure operator a, interior operator i, and complement operator c. Based on these operators, we shall classify probabilistic sets in the following.

[Prop.6-11]

Three monomial operators

$$\begin{aligned} a,i,c:\ & \mathcal{P}(X,\tau) \longrightarrow \mathcal{P}(X,\tau), \\ & \quad \omega \qquad\qquad\qquad \omega \\ & \quad A \longmapsto A^a, A^i, A^c \end{aligned} \tag{6-33}$$

satisfy the following properties.

1) $a \circ a = a$, (6-34)

2) $i \circ i = i$, (6-35)

3) $c \circ c = i_d$ (identity operator), (6-36)

4) $c \circ a \circ c = i$, (6-37)

5) $c \circ i \circ c = a$ (6-38)

6) $a \circ i \circ a \circ i = a \circ i$, (6-39)

7) $i \circ a \circ i \circ a = i \circ a$, (6-40)

8) $a \circ c = c \circ i$, (6-41)

9) $i \circ c = c \circ a$, (6-42)

<u>proof</u>

1) ∿ 5) clear.

6) For each $A \in \mathcal{P}(X,\tau)$, we have $A^{ai} \subset A^a$. Hence, we obtain $A^{aia} \subset A^{aa} = A^a$ and $A^{aiai} \subset A^{ai}$. On the other hand, we have $A^{ai} \subset A^{aia}$, and $A^{ai} = A^{aii} \subset A^{aiai}$. Therefore we conclude $A^{aiai} = A^{ai}$.

7) For each $A \varepsilon P(X,\tau)$, we have $A^{i} \subset A^{ia}$, $A^{ii} = A^{i} \subset A^{iai}$ and $A^{ia} \subset$ $\subset A^{iaia}$. Whereas, $A^{iai} \subset A^{ia}$, $A^{iaia} \subset A^{iaa} = A^{ia}$ and $A^{iaia} \subset A^{ia}$. Hence, we obtain $A^{iaia} = A^{ia}$.

8) For each $A \varepsilon P(X,\tau)$, we have $A^{c} \varepsilon P(X,\tau)$ and $(A^{c})^{cac} = (A^{c})^{i}$ by 4). Hence, $A^{ac} = A^{ci}$.

9) For each $A \varepsilon P(X,\tau)$, we have $A^{c} \varepsilon P(X,\tau)$. We obtain $(A^{c})^{cic} = (A^{c})^{a}$ by 5), and $A^{ic} = A^{ca}$.

(Q.E.D.)

[Theorem 6-2]

For each $A \varepsilon P(X,\tau)$, we can generate 14 probabilistic sets by operations of a, i, and c. And they are shown as in Fig.6-2.

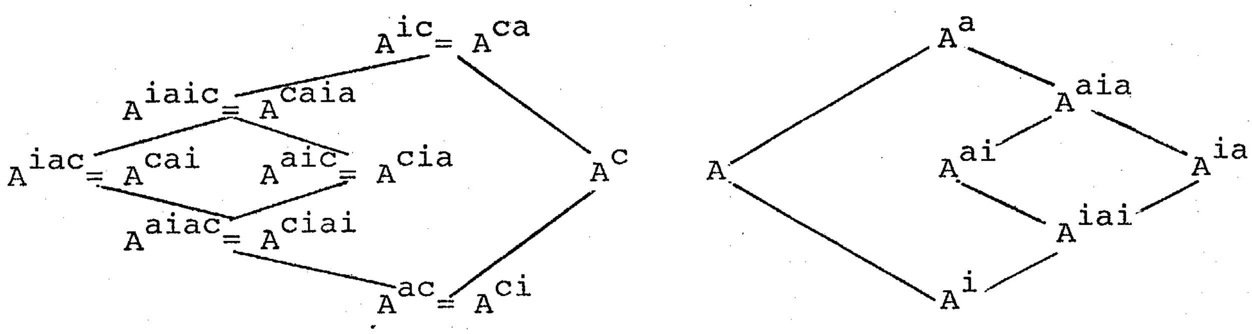
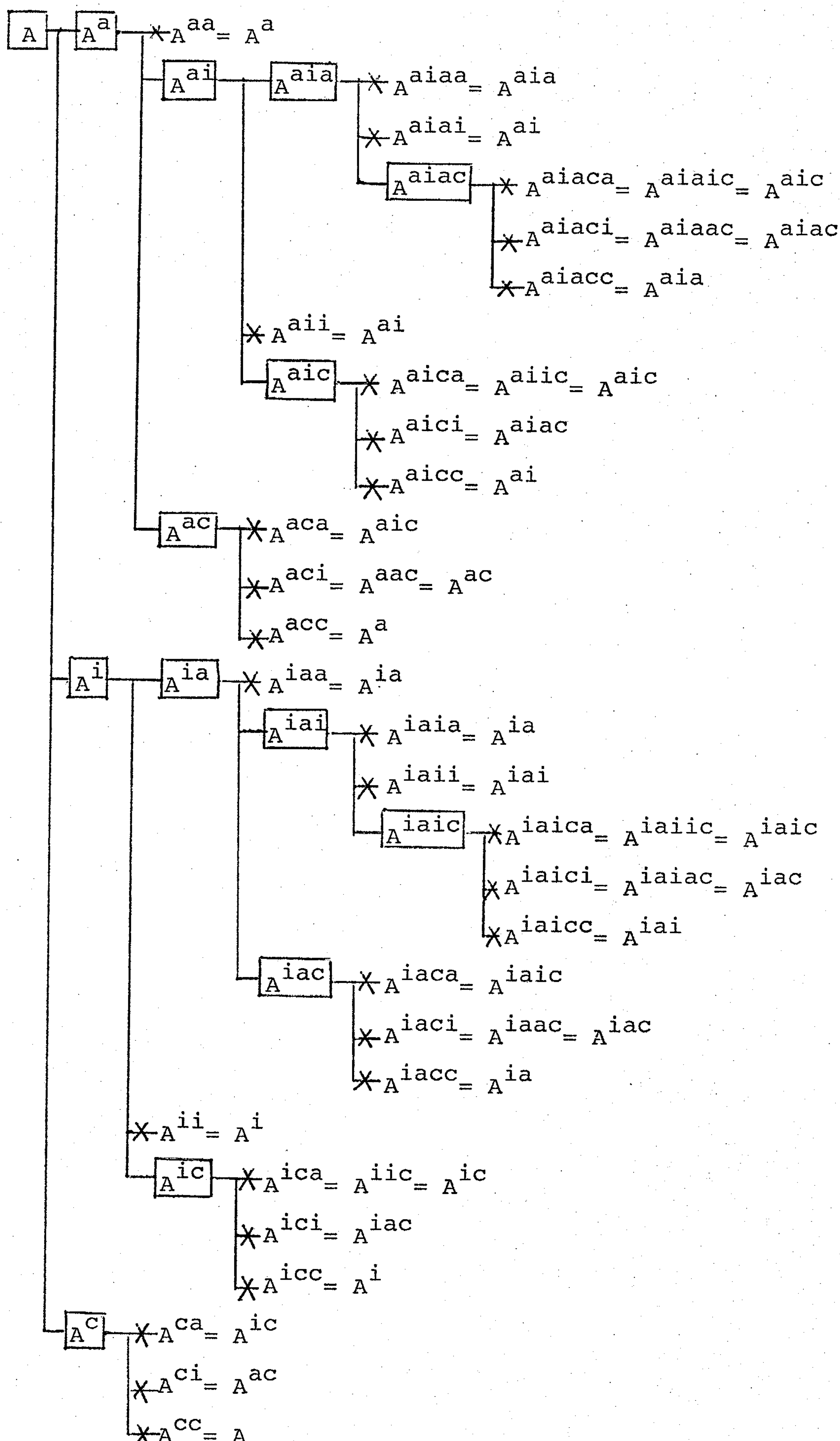


Fig.6-2. Hasse diagram of probabilistic sets induced by operators a,i and c.

<u>proof</u>

It will be verified by the following derivation tree that there is a possibility of being generated 14 probabilistic sets from A by operators a,i and c.

All we have to show is to give an example of a probabilistic set A which induces 14 different probabilistic sets in Fig.6-2. Let a total space X be a set of real numbers, and a parameter space (Ω,B,P) be the Wiener's probability spce,


A
A^{a}
$A^{aa} = A^{a}$
A^{ai}
A^{aia}
$A^{aiaa} = A^{aia}$
$A^{aiai} = A^{ai}$
A^{aiac}
$A^{aiaca} = A^{aiaic} = A^{aic}$
$A^{aiaci} = A^{aiaac} = A^{aiac}$
$A^{aiacc} = A^{aia}$
$A^{aii} = A^{ai}$
A^{aic}
$A^{aica} = A^{aiic} = A^{aic}$
$A^{aici} = A^{aiac}$
$A^{aicc} = A^{ai}$
A^{ac}
$A^{aca} = A^{aic}$
$A^{aci} = A^{aac} = A^{ac}$
$A^{acc} = A^{a}$
A^{i}
A^{ia}
$A^{iaa} = A^{ia}$
A^{iai}
$A^{iaia} = A^{ia}$
$A^{iaii} = A^{iai}$
A^{iaic}
$A^{iaica} = A^{iaiic} = A^{iaic}$
$A^{iaici} = A^{iaiac} = A^{iac}$
$A^{iaicc} = A^{iai}$
A^{iac}
$A^{iaca} = A^{iaic}$
$A^{iaci} = A^{iaac} = A^{iac}$
$A^{iacc} = A^{ia}$
$A^{ii} = A^{i}$
A^{ic}
$A^{ica} = A^{iic} = A^{ic}$
$A^{ici} = A^{iac}$
$A^{icc} = A^{i}$
A^{c}
$A^{ca} = A^{ic}$
$A^{ci} = A^{ac}$
$A^{cc} = A$

$$(\Omega,B,P) = ((0,1],\text{Borel sets, Lebesgue measure}). \quad (6\text{-}43)$$

Let the subjective topology on X be a trivial extension of the ordinary topology of real numbers. Consider a probabilistic set A whose defining function is given by,

$$\mu_A(x,\omega) = f(x)\cdot\omega \quad x \in X=R,\ \omega\in\Omega=(0,1]. \quad (6\text{-}44)$$

where

$$f(x) = \begin{cases} 1 & x \in (-2,-1)\cup(-1,0)\cup\{(0,1)\cap Q\}\cup\{2\} \\ 0 & \text{otherwise}. \end{cases} \quad (6\text{-}45)$$

Then we can induce 14 different probabilistic sets from A as shown in Fig.6-3.

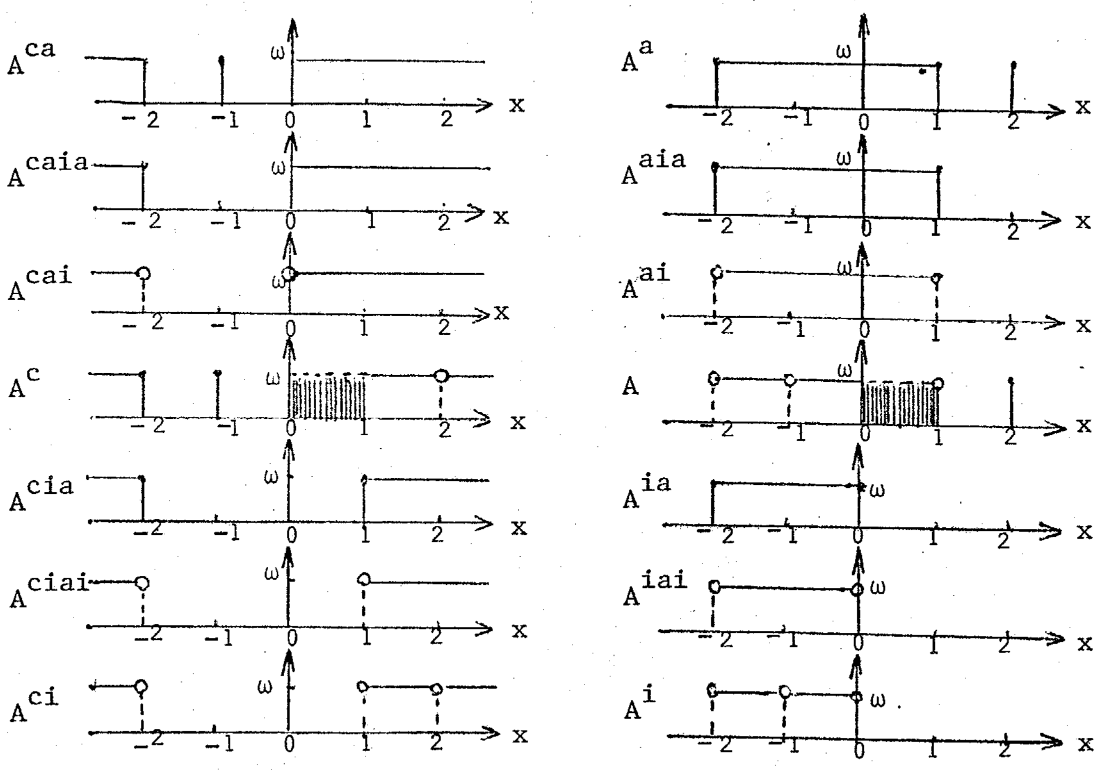


Fig.6-3. Defining functions of 14 probabilistic sets induced by A.

(Q.E.D.)

Based on this theorem, we propose several notions.

[Def.6-6] (regular open, regular closed)

Let (X,τ) be a subjective topological space. A probabilistic set $A(\varepsilon \mathcal{P}(X,\tau))$ is called a <u>regular open set</u> if

$$A= A^{ai}, \tag{6-46}$$

and a probabilistic set $A(\varepsilon \mathcal{P}(X,\tau))$ is called a <u>regular closed set</u> if

$$A= A^{ia}. \tag{6-47}$$

A set of regular open sets and a set of regular closed sets are denoted by

$$\mathcal{O}_{reg}= \{A \mid A \varepsilon \mathcal{P}(X,\tau),\ A= A^{ai}\}, \tag{6-48}$$

$$\mathcal{C}_{reg}= \{A \mid A \varepsilon \mathcal{P}(X,\tau),\ A= A^{ia}\}, \tag{6-49}$$

respectively.

Remark:

1) $A \varepsilon \mathcal{O}_{reg} \underset{n.\&s.}{\Longleftrightarrow} A= A^{i}= A^{iai}= A^{ai},$ (6-50)

2) $A \varepsilon \mathcal{C}_{reg} \underset{n.\&s.}{\Longleftrightarrow} A= A^{a}= A^{aia}= A^{ia},$ (6-51)

3) $\mathcal{O}_{reg} \subset \mathcal{O} \subset \mathcal{P}(X,\tau),\quad \mathcal{C}_{reg} \subset \mathcal{C} \subset \mathcal{P}(X,\tau).$ (6-52)

[Def.6-7] (pre-open, pre-closed)

Let (X,τ) be a subjective topological space. A probabilistic set $A(\varepsilon \mathcal{P}(X,\tau))$ is called a <u>pre-open set</u> if

$$A= A^{ia}, \tag{6-53}$$

and a probabilistic set $A(\varepsilon \mathcal{P}(X,\tau))$ is called a <u>pre-closed set</u> if

$$A= A^{ai}. \tag{6-54}$$

A set of pre-open sets and a set of pre-closed sets are denoted by

$$\mathcal{O}_{pre} = \{A \mid A \in P(X,\tau),\ A^a = A^{ia}\}, \qquad (6\text{-}55)$$

$$\mathcal{C}_{pre} = \{A \mid A \in P(X,\tau),\ A^i = A^{ai}\}, \qquad (6\text{-}56)$$

respectively.

Remark:

1) $A \in \mathcal{O}_{pre} \underset{n.\&s.}{\Longleftrightarrow} A^a = A^{aia} = A^{ia}$, (6-57)

2) $A \in \mathcal{C}_{pre} \underset{n.\&s.}{\Longleftrightarrow} A^i = A^{iai} = A^{ai}$, (6-58)

3) $\mathcal{O}_{reg} \subset \mathcal{O} \subset \mathcal{O}_{pre} \subset P(X,\tau)$, (6-59)

$\mathcal{C}_{reg} \subset \mathcal{C} \subset \mathcal{C}_{reg} \subset P(X,\tau)$. (6-60)

[Def.6-8]

Let $A, B \in P(X,\tau)$ and $A \subset B$. A is said to be <u>dense</u> in B if $A^a \supset B$.

$A(\in P(X,\tau))$ is said to be <u>nowhere dense</u> if $A^{ai} = \phi$.

[Prop.6-12]

1) $\phi,\ X \in \mathcal{O}_{reg}$, (6-61)

2) $O_1, O_2 \in \mathcal{O}_{reg} \Longrightarrow O_1 \cap O_2 \in \mathcal{O}_{reg}$, (6-62)

3) $A \in \mathcal{O}_{reg} \Longrightarrow A^c \in \mathcal{C}_{reg}$, (6-63)

4) $X,\ \phi \in \mathcal{C}_{reg}$, (6-64)

5) $C_1, C_2 \in \mathcal{C}_{reg} \Longrightarrow C_1 \cup C_2 \in \mathcal{C}_{reg}$, (6-65)

6) $A \in \mathcal{C}_{reg} \Longrightarrow A^c \in \mathcal{O}_{reg}$. (6-66)

<u>proof</u>

1) $\phi^{ai} = \phi^i = \phi,\quad X^{ai} = X^i = X$.

2) It is sufficient to show $(O_1 \cap O_2)^{ai} = O_1 \cap O_2$ under an assumption of $O_j^{ai} = O_j$ $(j = 1,2)$.

$O_1 \cap O_2 \subset O_1^a \cap O_2^a$. $\therefore (O_1 \cap O_2)^a \subset (O_1^a \cap O_2^a)^a = O_1^a \cap O_2^a$.

$(O_1 \cap O_2)^{ai} \subset (O_1^a \cap O_2^a)^i = O_1^{ai} \cap O_2^{ai} = O_1 \cap O_2$.

$(O_1 \cap O_2)^a \supset O_1 \cap O_2$ $\quad\therefore (O_1 \cap O_2)^{ai} \supset (O_1 \cap O_2)^i = O_1^i \cap O_2^i = O_1 \cap O_2$ $\quad\therefore (O_1 \cap O_2)^{ai} \supset O_1 \cap O_2$. Hence we have $(O_1 \cap O_2)^{ai} = O_1 \cap O_2$.

3) Since $A^{ai} = A$, we have $(A^c)^{ia} = A^{aca} = A^{aic} = A^c$.

4) 5) 6) They are verified almost in the same manner as 1) 2) 3).

(Q.E.D.)

Remark:

$\mathcal{O}_{reg}$ and $\mathcal{C}_{reg}$ are not always closed by $\cup$ and $\cap$, respectively.

[Prop.6-13]

1) $\phi,\ X \in \mathcal{O}_{pre}$, (6-67)

2) $O_1, O_2 \in \mathcal{O}_{pre} \Longrightarrow O_1 \cup O_2 \in \mathcal{O}_{pre}$, (6-68)

3) $A \in \mathcal{O}_{pre} \Longrightarrow A^c \in \mathcal{C}_{pre}$, (6-69)

4) $\phi,\ X \in \mathcal{C}_{pre}$, (6-70)

5) $C_1, C_2 \in \mathcal{C}_{pre} \Longrightarrow C_1 \cap C_2 \in \mathcal{C}_{pre}$, (6-71)

6) $A \in \mathcal{C}_{pre} \Longrightarrow A^c \in \mathcal{O}_{pre}$. (6-72)

proof

1) $\phi = \phi^a = \phi^{ia}$, $X = X^a = X^{ia}$.

2) It is sufficient to show $(O_1 \cup O_2)^a = (O_1 \cup O_2)^{ia}$ under an assumption of $O_j{}^a = O_j{}^{ia}$ $(j = 1,2)$.

$$O_1 \cup O_2 \supset (O_1 \cup O_2)^i \qquad \therefore (O_1 \cup O_2)^a \supset (O_1 \cup O_2)^{ia}$$

$$O_1 \cup O_2 \supset O_1{}^i \cup O_2{}^i, \quad (O_1 \cup O_2)^i \supset (O_1{}^i \cup O_2{}^i)^i = O_1{}^i \cup O_2{}^i.$$

$$(O_1 \cup O_2)^{ia} \supset (O_1{}^i \cup O_2{}^i)^a = O_1{}^{ia} \cup O_2{}^{ia} = O_1{}^a \cup O_2{}^a = (O_1 \cup O_2)^a.$$

$\therefore (O_1 \cup O_2)^{ia} \supset (O_1 \cup O_2)^a$. Hence we have $(O_1 \cup O_2)^a = (O_1 \cup O_2)^{ia}$.

3) Since $A^a = A^{ia}$, we have $(A^c)^{ai} = A^{cai} = A^{ici} = A^{iac} = A^{ac} = A^{ci}$, i.e. $(A^c)^{ai} = (A^c)^i$.

4) 5) 6) They are verified almost in the same way as 1)2)3).

(Q.E.D.)

Remark:

$\mathcal{O}_{pre}$ and $\mathcal{C}_{pre}$ are not always closed by $\cap$ and $\cup$, respectively.

6-4. ORDINAL STRUCTURE OF SUBJECTIVE TOPOLOGY.

We defined several definitions of subjective topology. Subjective topology of X is not always uniquely determined. On the contrary, we can consider various subjective topologies on the same total space X. We investigate their relationships from a viewpoint of ordinal structures.

[Example 6-1]

Let X be a total space. We can always define the following two subjective topologies on X. One is called a trivial topology τ_O, where τ_O is defined by the following sustem of open sets,

$$\mathcal{O}= \{\phi, X\}. \tag{6-73}$$

Another is called a discrete topology τ_I, where τ_I is defined by

$$\mathcal{O}= P(X). \tag{6-74}$$

Here, τ_O is the weakest topology and τ_I is the strongenst topology in the following sense.

[Prop.6-14]

Suppose that two subjective topologies are defined on X and denote them by (X,τ_1) and (X,τ_2), where τ_1 and τ_2 are defined by

topology	n.b.d. system	open sets	closed sets
τ_1	$\{V(A_x)\}_{A_x \subset X}$	$\mathcal{O}_1$	$\mathcal{C}_1$
τ_2	$\{V(A_x)\}_{A_x \subset X}$	$\mathcal{O}_2$	$\mathcal{C}_2$

closure operator	interior operator
a_1	i_1
a_2	i_2 .

Then we have the following mutually equivalent five propositions.

1) $\mathcal{V}_1(A_x) \subset \mathcal{V}_2(A_x)$ for all $A_x \subset X$, (6-75)

2) $\mathcal{O}_1 \subset \mathcal{O}_2$, (6-76)

3) $\mathcal{C}_1 \subset \mathcal{C}_2$, (6-77)

4) $A^{a_1} \supset A^{a_2}$ for all $A \varepsilon P(X)$, (6-78)

5) $A^{i_1} \subset A^{i_2}$ for all $A \varepsilon P(X)$. (6-79)

proof

1) → 2)

For each $A_x \subset O_1$, where $O_1 \varepsilon \mathcal{O}_1$ is arbitrarily fixed, we have $O_1 \varepsilon \mathcal{V}_1(A_x)$ (by prop.6-2). Hence we have $O_1 \varepsilon \mathcal{V}_2(A_x)$ by 1) and $O_1 \varepsilon \mathcal{O}_2$ by prop.6-1. Therefore we conclude $\mathcal{O}_1 \subset \mathcal{O}_2$.

2) → 1)

Let $U \varepsilon \mathcal{V}_1(A_x)$ be arbitrarily fixed. Then there exists $O \varepsilon \mathcal{O}_1$ such that $A_x \subset O \subset U$ by prop.6-2. The assumption 2) shows $O \varepsilon \mathcal{O}_2$. Hence $U \varepsilon \mathcal{V}_2(A_x)$ and we can conclude $\mathcal{V}_1(A_x) \subset \mathcal{V}_2(A_x)$. Others are easily verified in almost the same manner as a case of ordinary theory of general topology [8].

(Q.E.D.)

[Def.6-9]

Let τ_1 and τ_2 be two subjective topologies on X. The subjective topology τ_2 is said to be stronger than τ_1, and τ_1 is said to be weaker than τ_2, if one of 1)∿ 5) in prop.6-14 (hence all of them) holds. We denote it by

$$\tau_1 \leqq \tau_2. \tag{6-80}$$

A set of all subjective topologies on X is denoted by

$$\mathcal{T}(X) = \{\tau \mid \tau \text{ is a subjective topology on X }\}. \tag{6-81}$$

Remark:

It will be easily confirmed that $(\mathcal{T}(X), \geqq)$ constitutes a poset, i.e. for arbitrary τ_1, τ_2, $\tau_3 \in \mathcal{T}(X)$, we have

1) $\tau_1 \leqq \tau_1$, (6-82)

2) $\tau_1 \leqq \tau_2$ and $\tau_2 \leqq \tau_1 \rightarrow \tau_1 = \tau_1$, (6-83)

3) $\tau_1 \leqq \tau_2$ and $\tau_2 \leqq \tau_3 \rightarrow \tau_1 \leqq \tau_3$. (6-84)

Moreover, it can be shown that $(\mathcal{T}(X), \leqq)$ constitutes a complete lattice [9].

From a mathematical viewpoint, other useful, higher notions can be introduced such as "continuity", "compactness" and "separation axiom". However, they are omitted here, since the purpose of the present chapter is to introduce fundamentals of subjective topology.

CHAPTER. 7

APPLICATIONS.

7-1. INTRODUCTION.

We have been investigating the problems of "ambiguity and subjectivity" from mainly a theoretical point of view. In this chapter, we visualize the problem from a different angle, i.e. from a viewpoint of applications. Examples in the following will be given to help the explanation done so far. However, because we do not have enough space to describe every thing about our studies, we will introduce the following four examples.

Section 7-2 deals with an appraisement of recognition-performance of character readers. As an example of practical optical character readers, we refer to ASPET/71, especially its guiding principle "multiple similarity method" in terms of the concept of probabilistic sets (in section 7-3). In section 7-4, estimation of independent Gaussian noise patterns is investigated. A new statistic called B.V.Q. (bounded variation quantity) is also proposed for the purpose of estimation of standard deviation. Lastly, a method is mentioned on detecting the directionality of picture patterns. A result is also reported about texture analysis of electron-microscopic photographs of metals by using the B.V.Q. based on probabilistic set theory.

7-2. APPRAISEMENT OF RECOGNITION-PERFORMANCE OF CHARACTER READERS.

7-2-1. FUNDAMENTAL IDEAS OF APPRAISEMENT.

Many investigations have been made about the method of character recognition. Basic and applied researches have been done, and recently comparatively many character readers were put into practice [1]. However, their recognition-performance is not necessarily appraised correctly. Reports of this type of research have apparently not been published to date.

It is sometimes seen in demonstrations that participants are requested to fill out test-sheets. The written characters in test-sheets are put into the character reader. Then participants are informed whether the characters are recognized just the same as what they wrote or not. And each of them evaluates the recognition-performance of the character reader based on the output-result.

However, this type of evaluation is not always pertinent, since character readers should respond to, if possible, generally many people (not to specially selected individuals). The important point is not which way the given character is written by each individual, but how it is recognized by many people in the society. Therefore further investigation is necessary on analysis of ambiguity of handwritten characters. Based on the investigation, we shall deal with appraisement of recognition-performance of character readers.

Let a total space X,

$$X = \{x\}, \qquad (7\text{-}1)$$

be a set of all handwritten characters, and let a parameter space

$$\Omega = \{\omega\}, \tag{7-2}$$

be a set of members of the society. Here, it is assumed that the evaluation of each member ω is accepted with a weight $p(\omega)$,

$$\int_{\Omega} p(\omega)\cdot d\omega = 1, \quad p(\omega) \geqq 0. \tag{7-3}$$

It is also assumed that there exist concepts of characters in the society (e.g. the alphabet A,B,...,Z, "reject" and so on), where each concept is called a <u>category</u> and they are denoted by

$$C_0, C_1, C_2, \ldots, C_n. \tag{7-4}$$

Then an evaluation may be given for each member ω to the question whether the character x belongs to a category C_i or not. The evaluation is denoted by $\mu_i(x,\omega)$, where $\mu_i(x,\omega)$ takes its value in [0,1]-interval. The value 0 corresponds to "no", and contrary to this the value 1 corresponds to "yes". If the value $\mu_i(x,\omega)$ is not 1 or 0, then the evaluation becomes an ambiguous one. However, it should be noted that such an ambiguous evaluation was more important in appraisement of character readers.

Each category C_i is expressed as a probabilistic set on X whose defining function is

$$\begin{aligned} \mu_i:\ & X\times\Omega \longrightarrow [0,1]. \\ & \quad\ \omega \qquad\quad\ \ \omega \\ & (x,\omega) \longmapsto \mu_i(x,\omega) \end{aligned} \tag{7-5}$$

A pair $(\mu_i(x,\omega), p(\omega))$ is called a probabilistic expression of C_i (cf. section 4-2). There exists another mutually equivalent expression called an extended fuzzy expression (cf.chapter 4). The extended fuzzy expression of C_i consists of a countable family of monitors. It is shown that important information is concentrated on lower monitors such as a membership function m_i

$$m_i : X \longrightarrow [0,1], \quad x \longmapsto m_i(x) \qquad (7\text{-}6)$$

(with $x \in X$ and $m_i(x) \in [0,1]$)

and a vagueness function v_i

$$v_i : X \longrightarrow [0,1]. \quad x \longmapsto v_i(x) \qquad (7\text{-}7)$$

(with $x \in X$ and $v_i(x) \in [0,1]$)

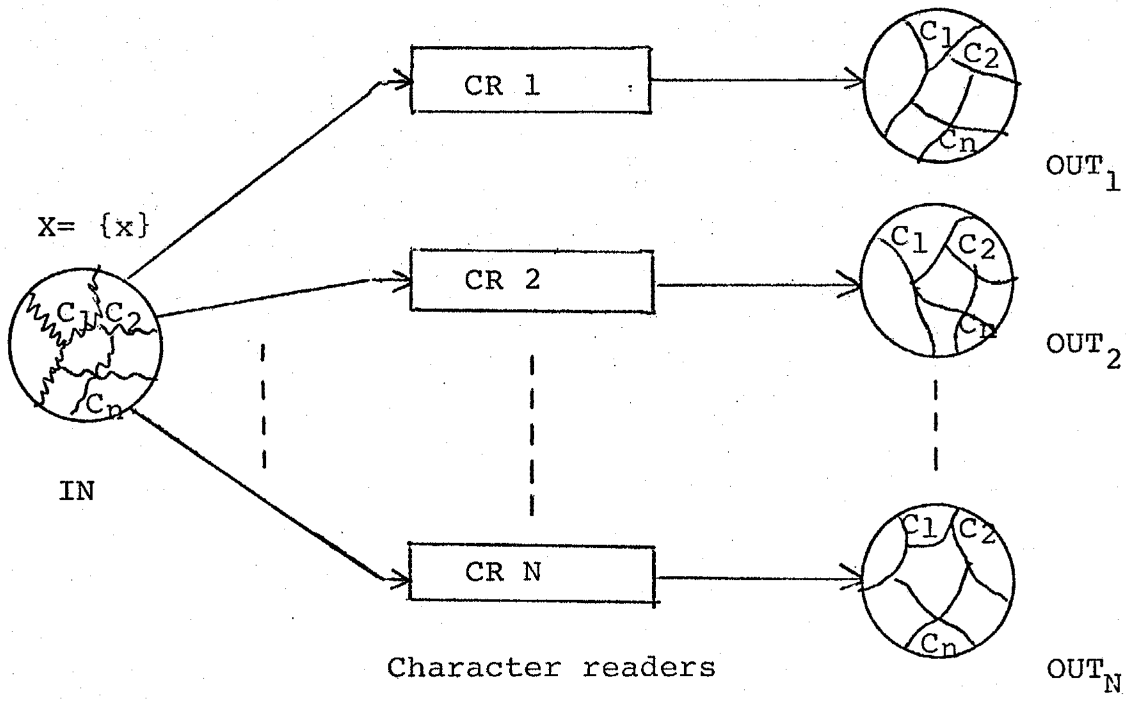
The value $m_i(x)$ indicates the average evaluation of the society to the question whether the character x belongs to the category C_i or not. Whereas the value $v_i(x)$ indicates a disordered degree (or "vagueness") of the average evaluation. There also exist the following relationships between the probabilistic expression and the extended fuzzy expression,

$$m_i(x) = \int_\Omega \mu_i(x,\omega) \cdot p(\omega) \cdot d\omega, \qquad (7\text{-}8)$$

$$v_i(x) = \int_\Omega (\mu_i(x,\omega) - m_i(x))^2 p(\omega) \cdot d\omega. \qquad (7\text{-}9)$$

Hence we can estimate the value of membership function and vagueness function by a statistical method (cf.section 2-6). Strictly speaking, it is necessary to know the value of monitors whose degree is more than 2 for the purpose of a complete expression of C_i. But it is impossible in general to estimate them accurately. It is empirically sufficient to use both membership function and vagueness function.

A set of input characters $X=\{x\}$ is classified by the concept of category C_i which is empirically established by the society. However it will be natural to expect that each character x (εX) does not always correspond to one definite category. Especially characters which belong to the border of categories will be evaluated in various ways, i.e. their vagueness function has a non-zero, positive value. The function of each character reader is considered to give an "{0,1}-definite" output for each input character x, even if the character is an ambiguous one. Since we can expect, in general, N different outputs for N different character readers, we would lide to know from the outputs which character reader was the best one for our society (Fig.7-1).



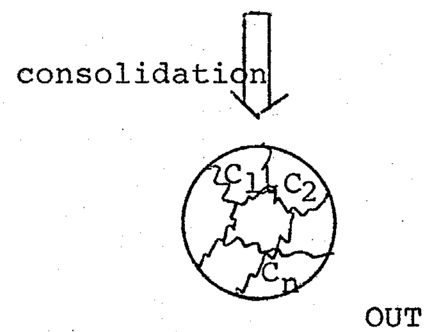


Fig.7-1. A function-diagram of N character readers.

It will be natural to consider that if a character reader outputs the same result as that of categorical classification in X, then it is the best one. Of course it is impossible to expect such a character reader, because each output of character readers is a {0,1}-definite one, whereas a set of input characters X contains a lot of [0,1]-ambiguous ones. However, we should deal with the "goodness" of character readers by comparing inputs "IN" with outputs "OUT"; where IN is established by the society and OUT is given by character readers. Therefore the point is to introduce a reasonable criterion d(IN,OUT) which indicates the difference between IN and OUT.

Let us reconsider the ambiguity of written characters. For each member (or individual) $\omega \varepsilon \Omega$, the evaluation may be a {0,1}-definite one even if the presented character is written ambiguously. But such an ambiguous character will be evaluated in various ways according to members of the society. As a result of various evaluations, a [0,1]-ambiguous value of membership functions and a non-zero "vagueness" arise. The same situation can be seen in a set of character readers. For each character reader, the answer may be a {0,1}-definite one but the answer will be changed according to character readers. If we consolidate the outputs given by character readers, then we have the same kind of ambiguous evaluations as ones given by members of the society. Therefore we can compare two quantities, i.e. IN and OUT, both of which have the same kind of ambiguous structure.

However, what we would like to know is not the value of a criterion d(IN,OUT), but the value of the difference between IN and the output OUT_j of j-th character reader. It is convenient if the difference between IN and OUT_j is given in the same form as d(IN,OUT). In order to derive a function-form of the criterion d(IN,OUT), we must take the consolidation-operation into consideration. We denote the operation by E, for example (cf.Fig.7-1) OUT is written by

$$OUT = E\{OUT_1, OUT_2, \ldots, OUT_N\}. \tag{7-10}$$

Based on the abovestated arguments, we require the following relation of invariance between the consolidation-operation E and the difference measure $d(\cdot,\cdot)$,

$$(d(IN,OUT)=)\ d(IN,E\{OUT_1,OUT_2,\ldots,OUT_N\})=$$

$$= E\{d(IN,OUT_1),d(IN,OUT_2),\ldots,d(IN,OUT_N)\}. \quad (7\text{-}11)$$

It will also be convenient if the measure $d(\cdot,\cdot)$ consists of lower monitors such as membership functions and vagueness functions. We shall introduce a measure $d(\cdot,\cdot)$ which satisfies abovestated requirements in the following section.

7-2-2. A CRITERION OF RECOGNITION-PERFORMANCE.

Before going into the main argument, we make preparation several fundamentals about relationships between the consolidation-operation and lower monitors. Let $\mu(x,\omega)$, $m(x)$ and $v(x)$ be a defining function, a membership function and a vagueness function respectively. Then we have

$$m(x) = \int_{\Omega} \mu(x,\omega)\, p(\omega) \cdot d\omega, \qquad (7\text{-}12)$$

$$v(x) = \int_{\Omega} (\mu(x,\omega) - m(x))^2\, p(\omega) \cdot d\omega, \qquad (7\text{-}13)$$

(cf. (7-8) and (7-9)).

In the next place, consider the case of two parameter spaces Ω_1 and Ω_2. Suppose that the membership function and the vagueness function are given by $\{m_1(x), v_1(x)\}$ in Ω_1 and $\{m_2(x), v_2(x)\}$ in Ω_2, and that two parameter spaces are consolidated with weight p_1 and p_2,

$$p_1 + p_2 = 1, \quad p_1, p_2 \geq 0. \qquad (7\text{-}14)$$

We obtain the membership function $m(x)$ and the vagueness function $v(x)$ in the cosolidated parameter space $\Omega = \Omega_1 \cup \Omega_2$ as follows;

$$m(x) = \int_{\Omega_1 \cup \Omega_2} \mu(x,\omega)\, p(\omega) \cdot d\omega$$
$$= p_1 \cdot m_1(x) + p_2 \cdot m_2(x), \qquad (7\text{-}15)$$

$$v(x) = \int_{\Omega_1 \cup \Omega_2} (\mu(x,\omega) - (p_1 m_1(x) + p_2 m_2(x)))^2\, p(\omega) \cdot d\omega$$
$$= p_1 \cdot v_1(x) + p_2 \cdot v_2(x) + p_1 p_2 (m_1(x) - m_2(x))^2. \qquad (7\text{-}16)$$

It should be noted that $m(x)$ was given by a weighted sum but that $v(x)$ was larger than that.

Consider the case of n parameter spaces,

$$\Omega = \Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_n, \qquad (7\text{-}17)$$

$$1 = p_1 + p_2 + \ldots + p_n, \quad p_i \geq 0. \qquad (7\text{-}18)$$

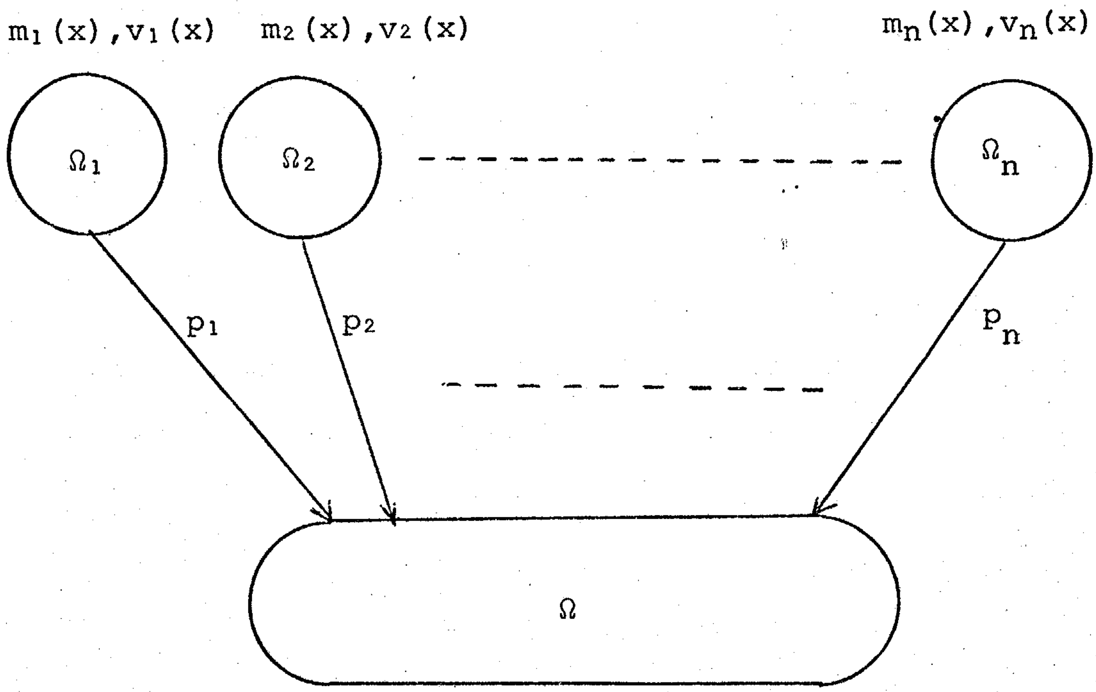
It is clear that the membership function is given by a weighted sum,

$$m(x) = \sum_{i=1}^{n} p_i \cdot m_i(x). \qquad (7\text{-}19)$$

Then the vagueness function is calculated as follows;

$$\begin{aligned} v(x) &= \int_\Omega (\mu(x,\omega) - \sum_i p_i m_i(x))^2 p(\omega)\,d\omega \\ &= \int_\Omega (\sum_i p_i(\mu(x,\omega) - m_i(x)))^2 p(\omega)\,d\omega \\ &= \sum_{i,j} p_i p_j \{\int_\Omega \mu(x,\omega)^2 p(\omega)\,d\omega - (m_i(x)+m_j(x)) \int_\Omega \mu(x,\omega) p(\omega)\,d\omega \\ &\qquad + m_i(x) m_j(x)\} \\ &= \sum_{i=1}^{n} p_i v_i(x) + \frac{1}{2} \sum_{i,j=1}^{n} p_i p_j (m_i(x) - m_j(x))^2. \qquad (7\text{-}20) \end{aligned}$$

The result is shown in Fig.7-2.



$$\begin{cases} m(x) = \sum_i p_i m_i(x) \\ v(x) = \sum_i p_i v_i(x) + \frac{1}{2} \sum_{i,j} p_i p_j (m_i(x) - m_j(x))^2 \end{cases}$$

Fig.7-2. Consolidation of n parameter spaces.

Let us reconsider the main problem. As we have already mentioned, the fundamental idea is a comparison of IN with OUT (cf.

section 7-2-1), where IN and OUT is expressed by $\mu_{IN}(x,\omega)$, $m_{IN}(x)$, $v_{IN}(x)$ and $\mu_{OUT}(x,\omega)$, $m_{OUT}(x)$, $v_{OUT}(x)$, respectively. The difference between IN and OUT is represented by an operation "symmetric defference" INΔOUT (cf.(3-95)),

$$\mu_{IN\Delta OUT}(x,\omega)= |\mu_{IN}(x,\omega) - \mu_{OUT}(x,\omega)|. \tag{7-21}$$

In order to emphasize the difference, we take its 2-power (cf. (3-98)) $(IN\Delta OUT)^2$,

$$\mu_{(IN\Delta OUT)^2}(x,\omega)= (\mu_{IN}(x,\omega) - \mu_{OUT}(x,\omega))^2, \tag{7-22}$$

Then its membership function becomes,

$$m_{(IN\Delta OUT)^2}(x)= \int_{\Omega}(\mu_{IN}(x,\omega) - \mu_{OUT}(x,\omega))^2 p(\omega)d\omega \tag{7-23}$$

$$= v_{IN}(x) + v_{OUT}(x) + (m_{IN}(x)- m_{OUT}(x))^2, \tag{7-24}$$

where it is assumed that the parameter spaces Ω_{IN} and Ω_{OUT} are independent, i.e.

$$\Omega= \Omega_{IN} \times \Omega_{OUT}, \tag{7-25}$$

$$p(\omega)= p_{IN}(\omega)\cdot p_{OUT}(\omega). \tag{7-26}$$

This membership function satisfies the abovementioned invariance (7-11): Suppose that there exist N character readers whose lower monitors are expressed by $\{m_j(x),v_j(x)\}_{j=1}^{N}$, and that the weight is given by $\{p_j\}_{j=1}^{N}$. Then we have, by (7-19) and (7-20),

$$m_{OUT}(x)= \sum_{j=1}^{N} p_j m_j(x), \tag{7-27}$$

$$v_{OUT}(x)= \sum_{j=1}^{N} p_j v_j(x) + \frac{1}{2} \sum_{i,j=1}^{N} p_i p_j (m_i(x)-m_j(x))^2. \tag{7-28}$$

On the other hand, since we have

$$m_{(IN\Delta j)^2}(x)= v_{IN}(x) + v_j(x) + (m_{IN}(x) + m_j(x))^2, \tag{7-29}$$

in the same way as (7-24), we obtain

$$\sum_{j=1}^{N} p_j \cdot m_{(IN\Delta j)^2}(x)=$$

$$= v_{IN}(x) + \sum_{j=1}^{N} p_j v_j(x) + m_{IN}(x)^2 - 2 \cdot m_{IN}(x) \sum_{j=1}^{N} p_j m_j(x) + \sum_{j=1}^{N} p_j m_j(x)^2$$

$$= v_{IN}(x) + v_{OUT}(x) + (m_{IN}(x) - m_{OUT}(x))^2. \qquad (7\text{-}30)$$

The right side of (7-30) provides the same quantity as (7-24), and the assertion is confirmed.

It is clear from (7-23) that this quantity provides the square of L^2-distance measure. This is also a convenient property. We adopt this quantity (7-29) as a basis of the criterion and denote it by D(IN,j)(x), i.e.

$$D(IN,j)(x) = v_{IN}(x) + v_j(x) + (m_{IN}(x) - m_j(x))^2 \qquad j=1,2,\ldots,N. \ (7\text{-}31)$$

It should be noted that in most cases the output of character readers is a definite one, i.e. $v_j(x)=0$ and $m_j(x)$ is either 1 or 0, so we obtain the following inequality,

$$0 \leqq v_{IN}(x) \leqq D(IN,j)(x) \leqq 1. \qquad (7\text{-}32)$$

Of course it is desirable that D(IN,j)(x) takes a small value, but its lower bound is given by $v_{IN}(x)$ ("the vagueness of the character x"). This lower bound $v_{IN}(x)$ doesn't depend on character readers, but on the society. Therefore it will be natural for the purpose of appraisement of character readers to take away the vagueness from D(IN,j)(x), and we designate it by d(IN,j)(x), i.e.

$$d(IN,j)(x) = (m_{IN}(x) - m_j(x))^2. \qquad (7\text{-}33)$$

It should be noted that d(IN,j)(x) also satisfied the invariance-relation (7-11). This quantity is a criterion for each character x. The criterion for X={x} is given by its average,

$$d(IN,j) = \int_X d(IN,j)(x)\cdot p(x)\cdot dx$$

$$= \int_X (m_{IN}(x) - m_j(x))^2 p(x)\cdot dx, \qquad (7\text{-}34)$$

where each character x is supposed to appear with frequency p(x),

$$\int_X p(x)dx = 1, \quad p(x) \geqq 0. \qquad (7\text{-}35)$$

Based on the concept of probabilistic sets, we showed bound (7-32) of recognition-performance and proposed a criterion (7-34) of appraisement. However, further experiments like ones mentioned in section 2-6 and 5-10 will be necessary in order to establish the technical system. It will also be possible and be interesting to discuss the criterion from a viewpoint of subjective entropy. We would like to leave this for one of further studies.

7-3. MULTIPLE SIMILARITY METHOD OF OCR "ASPET/71".

In applications of probabilistic set theory, one of the most important points is to choose a suitable parameter space according to each situation. No restrictions are imposed on the parameter space (Ω, B, P) except that it is a probability space, but we have been mentioning only the case of Ω being a set of people. We shall give other examples in the present and the following sections.

As we have mentioned in a previous section, character recognition is a categorizing process of each of unknown input character patterns into one of known finite number of character categories. Various practical methods of realizing this process has been devised. Among them, pattern matching method is one of the most commonly used techniques in which the similarity of input pattern is tested with reference patterns of each category. Multiple similarity method has been developed as an advanced, practical technique of the pattern matching method by T.Iijima, who is an advisor of the present author [2]. A practical optical character reader ASPET/71 (analogue spatial processor developed by Electro-technical laboratory and Toshiba) capable of reading printed characters with very poor print quality has been successfully designed and put into practice based on mainly the theory of the multiple similarity method [3]. The multiple similarity method is also possible to interpret by using the concept of probabilistic sets. We shall deal with it in the following.

Let each character pattern be expressed by a light energy distribution function of an element of $L^2(R^2)$. Let an input

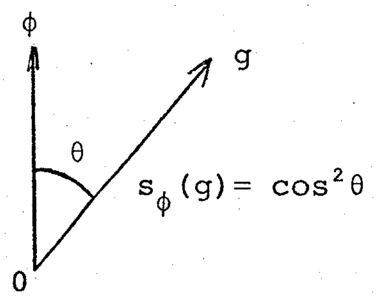


Fig.7-3. An illustration of simple similarity in function space $L^2(R^2)$.

pattern be g and a reference pattern be ϕ. Then simple similarity is defined by the following $s_\phi(g)$,

$$s_\phi(g) \underset{\Delta}{=} (\phi, g)^2 / \|\phi\|^2 \|g\|^2 = \cos^2\theta, \qquad (7\text{-}36)$$

where

$$0 \leqq s_\phi(g) \leqq 1, \qquad (7\text{-}37)$$

(cf.Fig.7-3). Based on the concept of simple similarity, multiple similarity s(g) is defined as follows,

$$s(g) = \sum_{n=0}^{N} p_n \cdot s_{\phi_n}(g), \qquad (7\text{-}38)$$

$$p_n = \lambda_n / \sum_{m=0}^{N} \lambda_m \quad (\geqq 0) \quad \left(\sum_{n=0}^{N} p_n = 1 \right), \qquad (7\text{-}39)$$

where λ_n is n-th eigen value which corresponds to the mode function ϕ_n [3].

As is evident from (7-38) and (7-39), the multiple similarity method can be interpreted in terms of probabilistic sets; We can regard each mode function ϕ_n and the ratio p_n as a parameter and its probability,respectively,

$$\Omega = \{\omega_n \underset{\Delta}{=} \phi_n\}_{n=0}^{N}, \qquad (7\text{-}40)$$

$$p(\omega_n) = \lambda_n / \sum_{m=0}^{N} \lambda_m. \qquad (7\text{-}41)$$

Then $s_{\phi_n}(g)$ can be interpreted as a defining function of a probabilistic set,

$$\mu(g, \omega_n) \underset{\Delta}{=} s_{\phi_n}(g). \qquad (7\text{-}42)$$

The multiple similarity s(g) becomes the "expectation of the defining function" or the membership function of the probabilistic set,

$$s(g) = \int_\Omega \mu(g, \omega)\, dP(\omega) = m(g). \qquad (7\text{-}43)$$

Fig.7-4 allow the points to be read easily.

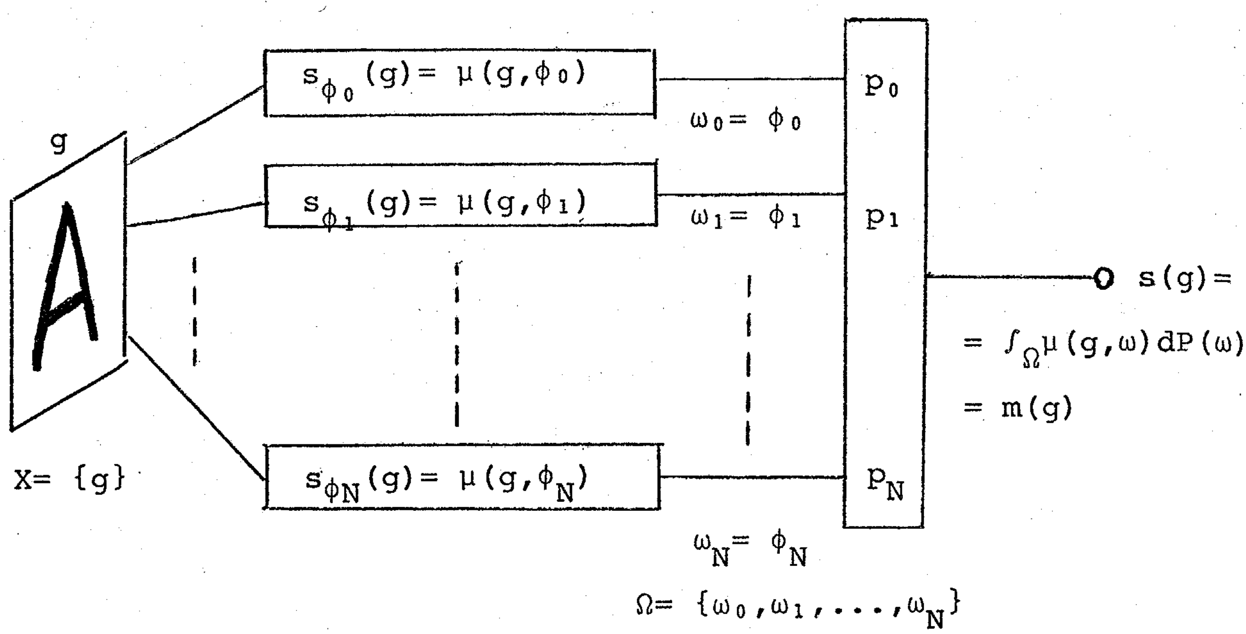


Fig.7-4. Multiple similarity method from a viewpoint of probabilistic sets.

In the case of OCR "ASPET/71", three parameters are chosen, i.e. N=2, and ϕ_0, ϕ_1 and ϕ_2 are made by $\phi, \partial\phi/\partial x$ and $\partial\phi/\partial y$ by taking the translation of characters in sheets into consideration.

7-4. ESTIMATION OF GAUSSIAN NOISE-PATTERNS.

Let's consider a discrete time series,

$$\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots, \tag{7-44}$$

where

$$X_i \sim N(m,\sigma^2) \quad \text{for all } i= 0,\pm 1,\ldots , \tag{7-45}$$

and each of them is independent. ($N(m,\sigma^2)$ stands for a normal distribution with mean m and standard deviation σ.) This pattern, as is well-known, is called an independent Gaussian process, and often appears in noise analysis. This pattern is characterized completely by only two parameters, i.e. the mean value m and the standard deviation σ,

$$\Omega= \{\omega_1= m,\ \omega_2= \sigma\}. \tag{7-46}$$

These noise patterns are expressed as probabilistic sets, and the problem is how to estimate m and σ from observed data.

In the abovementioned situation (7-44) and (7-45), consider the following two statistics,

$$E(n,i)= \frac{1}{n}\sum_{j=0}^{n-1} X_{i-j}, \tag{7-47}$$

$$Var(n,i)= \frac{1}{n-1}\sum_{j=0}^{n-1} (X_{i-j}- E(n,i))^2, \tag{7-48}$$

where i stands for the present time and n stands for the number of observed data. Then, as is well-known, the expectation and the error-variance of these two statistics are given as follows;

$$E[E(n,i)]= m, \tag{7-49}$$

$$V[E(n,i)]= \sigma^2/n, \tag{7-50}$$

$$E[Var(n,i)]= \sigma^2, \tag{7-51}$$

$$V[Var(n,i)]= 2\sigma^4/(n-1). \tag{7-52}$$

Hence the mean value m and the standard deviation σ can be estimated by (7-47) and (7-48), and the estimation error tends to zero as n tends to infinity. However, the convergence-ratio is a little bit different, i.e. in the case of mean value m the order of the convergence-ratio is $1/n$, whereas it is $1/\sqrt{n}$ in the case of standard deviation. But it can be improved by using a concept of B.V.Q. (bounded variation quantity), which was proposed by the present auther [4]. And we shall propose a new statistic $V(n,i)$ in the following.

By using the statistic based on the B.V.Q., the standard deviation σ can be estimated with $1/n$ as the order of the convergence-ratio. In the same situation as (7-44) and (7-45), the bounded variation quantity is given by

$$V(n,i)=\frac{\sqrt{\pi}}{2n}\sum_{j=0}^{n-1}|X_{i-j}-X_{i-(j+1)}|. \tag{7-53}$$

With regard to the expectation and the error-variance of this statistic, we have the following proposition.

[Prop.7-1]

$$E[V(n,i)]=\sigma, \tag{7-54}$$

$$V[V(n,i)]=(\pi-2)\sigma^2/(2n). \tag{7-55}$$

proof

Since $X_{i-j}\sim N(m,\sigma^2)$ for all j, and since X_{i-j} and $X_{i-(j+1)}$ are independent, we obtain

$$X_{i-j}-X_{i-(j+1)}\sim N(0,(\sqrt{2}\sigma)^2). \tag{7-56}$$

If we calculate moments, we will have

$$E[|X_{i-j}-X_{i-(j+1)}|]=2\sigma/\sqrt{\pi}, \tag{7-57}$$

$$E[|X_{i-j}-X_{i-(j+1)}|^2]=2\sigma^2. \tag{7-58}$$

Hence, we obtain

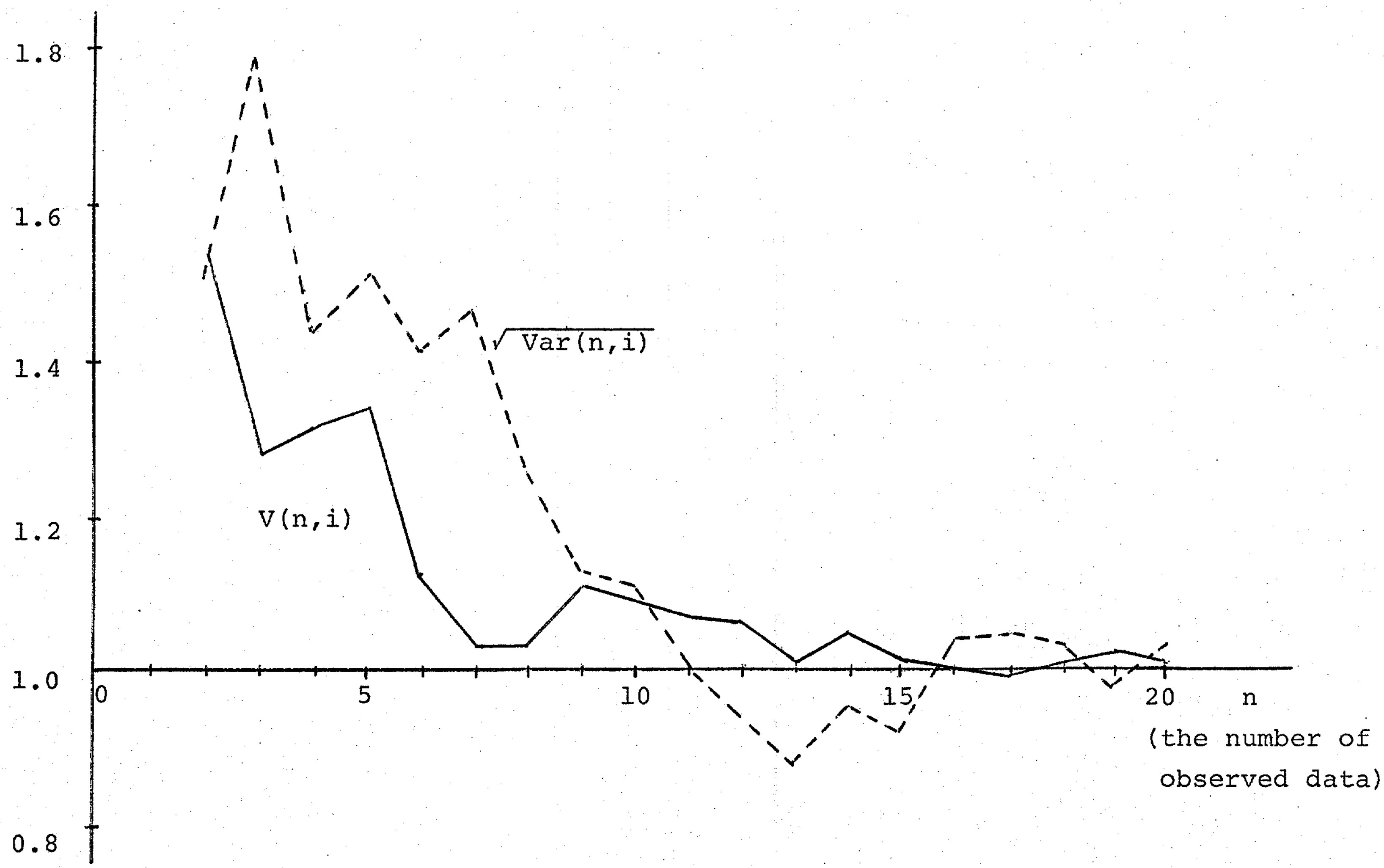


Fig.7-5. The estimation of standard deviation in N(0,1).

$$E[V(n,i)] = \frac{\sqrt{\pi}}{2n} \sum_{j=0}^{n-1} E[|X_{i-j} - X_{i-(j+1)}|]$$

$$= \frac{\sqrt{\pi}}{2n}\, n\, \frac{2\sigma}{\sqrt{\pi}} = \sigma, \tag{7-59}$$

$$V[V(n,i)] = E[V(n,i)^2] - E[V(n,i)]^2$$

$$= (\pi/4n^2)\{n2\sigma^2 + 2(n(n-1)/2)(2\sigma/\sqrt{\pi})^2\} - \sigma^2$$

$$= (\pi-2)\sigma^2/2n. \tag{7-60}$$

(Q.E.D.)

Comparing (7-51)(7-52) with (7-54)(7-55), we can summarize he advantage of bounded variation estimation as follows.

) The order of convergence-ratio is improved from $1/\sqrt{n}$ to $1/n$.

) The coefficient is improved from $\sqrt{2\sigma^4} = 1.414\sigma^2$ to $(\pi-2)\sigma^2/2 = 0.571\sigma^2$.

) In the calculation of V(n,i), the main operation is a difference (see (7-53)), so the reduction of bits is possible. (In (7-48), because the main operation is a square-multiplication of X_{i-j}, the necessary bits are doubled.) Moreover, the computation-speed of difference is, in general, faster than that of multiplication. Hence the bounded variation estimation is very convenient, especially in the case of real-time-processing with a microcomputer.

An example of a simulation experiment is shown in Fig.7-5. The abovestated advantages 1) and 2) can be seen in this example.

Based on abovestated result defining functions $\mu(\{X_i\},\omega_1)$ and $\mu(\{X_i\},\omega_2)$ are estimated from (7-47) and (7-53), respectively. Since the complexity of calculation and the convergence speed are almost the same in both cases, we can estimate them by a real-time-parallel-micro-processing system (cf. Photo.7-1). A result

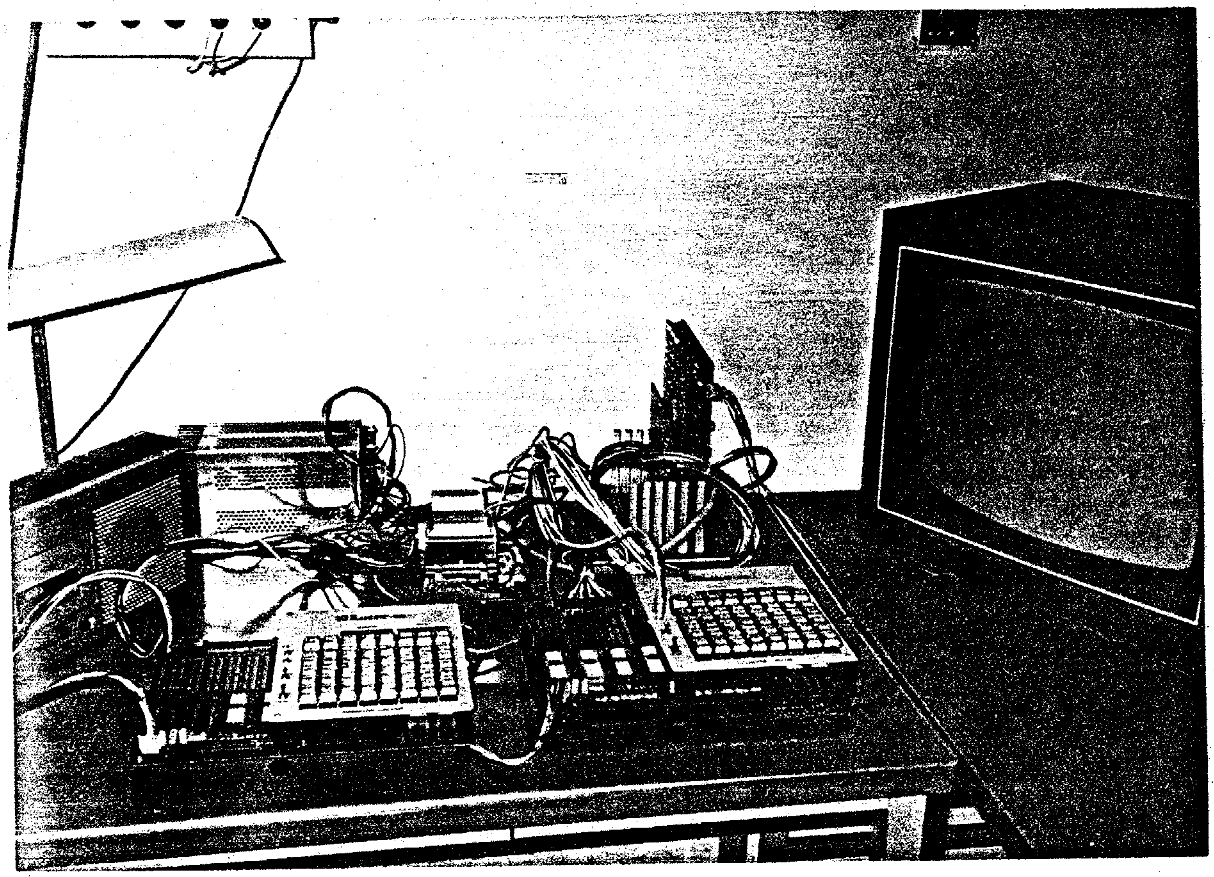
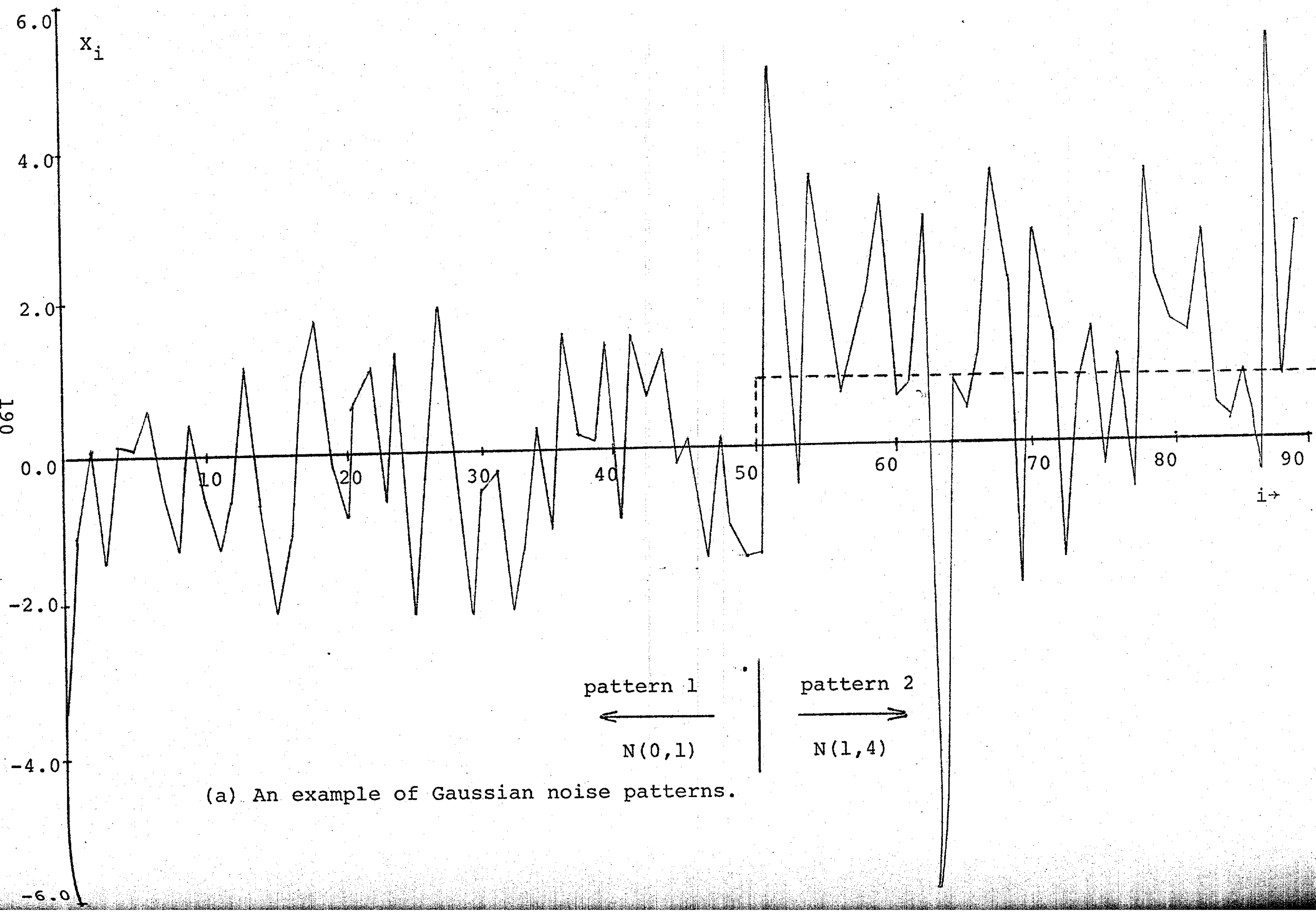
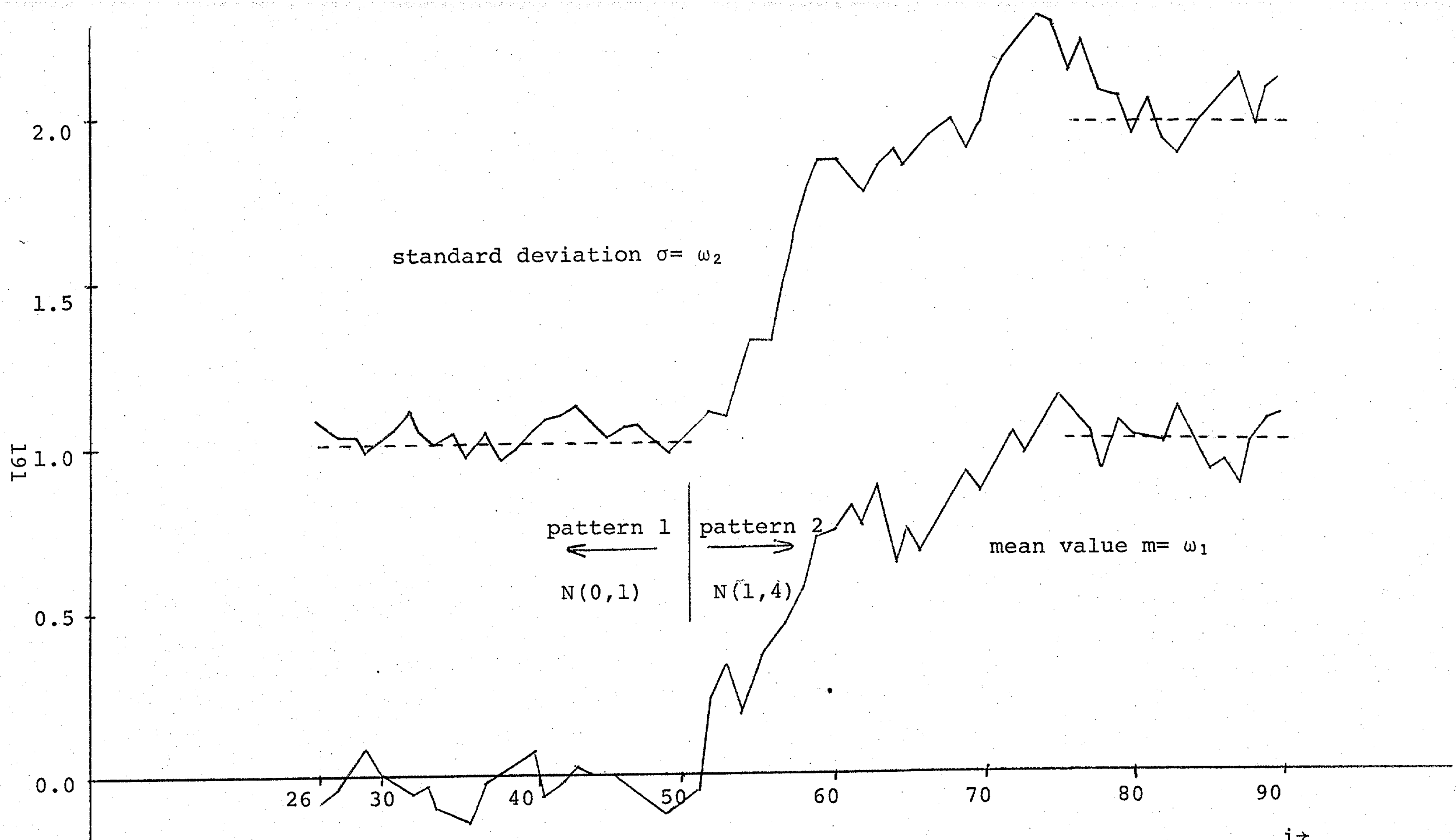


Photo. 7-1. A parallel-micro-processing system.

of a simulation experiment is also shown in Fig.7-6 (a) and (b).



(a) An example of Gaussian noise patterns.



(b) The estimation of two parameters of patterns in (a).

Fig.7-6. A result of simulation experiment.

7-5. TEXTURE ANALYSIS BASED ON DIRECTIONALITY.

Texture analysis is a classifying process of each picture-pattern into several regions each of which has the same "texture". The texture of picture-patterns is characterized by several features such as coarseness, contrast and directionality. Among them, directionality can be easily detected by using BVQ and the concept of probabilistic sets. The outline will be described below.

Firstly, we shall mention BVQ in the case of single variable function. Let f(x) be a real-valued function on [a,b], and Δ be an arbitrary finite partition of [a,b],

$$\Delta:\ a = x_0 < x_1 < x_2 < \dots < x_n = b. \qquad (7\text{-}61)$$

We define the variation of f with respect to Δ as follows,

$$V(f,\Delta) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|. \qquad (7\text{-}62)$$

If the partition Δ' is a refinement of $\Delta(\Delta' > \Delta)$, then we have

$$0 \leqq V(f,\Delta) \leqq V(f,\Delta') < +\infty. \qquad (7\text{-}63)$$

Considering this property, we define the variation of f as follows,

$$V(f) = \sup \{V(f,\Delta) \mid \Delta\text{:finite partition of } [a,b]\}. \qquad (7\text{-}64)$$

This quantity V(f) is nonnegative but may be infinite for some functions. A function f is said to be a <u>bounded variation function</u> if

$$V(f) < +\infty\ , \qquad (7\text{-}65)$$

and the value V(f) is called the <u>bounded variation quantity</u> BVQ of f. Another representation of V(f) is also possible, i.e.

$$V(f) = \int_{-\infty}^{\infty} |f'(x)|\,dx, \qquad (7\text{-}66)$$

where the derivative is taken in the sense of distribution [4].

In the next place, consider the case of picture-patterns. Let each picture-pattern be expressed by a real-valued function of two-variables $f(x,y)$. Let us consider a straight line of direction $\theta\,(0 \leqq \theta < \pi)$,

$$ax+by=c, \tag{7-67}$$

on a x-y plane as shown in Fig.7-7. If we define a coordinate transformation from (x,y) to (ξ,η) by,

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}, \tag{7-68}$$

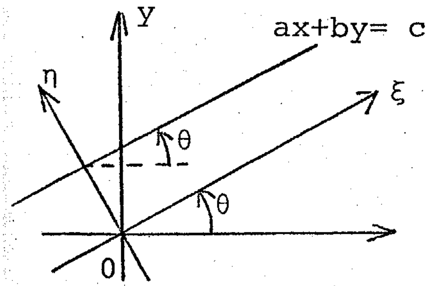


Fig.7-7. A straight line ax+by= c.

and if we restrict a domain of $f(x,y)$ to the straight line $ax+by=c$, then the function

$$f(x,y)=f(\xi\cdot\cos\theta-\eta\cdot\sin\theta,\ \xi\cdot\sin\theta+\eta\cdot\cos\theta) \tag{7-69}$$

will be regarded as a function of one variable ξ. Under an assumption that f is a bounded variation function of ξ, the variation $V(f,ax+by=c)$ of f on the straight line $ax+by=c$ will be given by

$$V(f,ax+by=c)=\int_{-\infty}^{\infty}|\partial f/\partial\xi|\cdot d\xi, \tag{7-70}$$

(cf. (7-66)). Since this quantity $V(f,ax+by=c)$ is a function of η, an integration

$$V(f,\theta)=\int_{-\infty}^{\infty}V(f,ax+by=c)\,d\eta \tag{7-71}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\partial f/\partial\xi|\,d\xi d\eta, \tag{7-72}$$

may be calculated. This quantity will be called the variation of f in the θ-direction. The <u>bounded variation quantity</u> (B.V.Q.) $V(f)$ of f will be characterized as the mean value of $V(f,\theta)$

$$V(f)=\frac{1}{\pi}\int_0^{\pi}V(f,\theta)\,d\theta$$

$$=\frac{1}{\pi}\int_0^{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\partial f/\partial\xi|\,d\xi d\eta d\theta, \tag{7-73}$$

and it will also be written symbolically

$$V(f)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{(1/\pi)\int_0^{\pi}|df(x,y)/d\theta|\,d\theta\}dxdy, \tag{7-74}$$

where

$$df(x,y)/d\theta = \cos\theta(\partial f(x,y)/\partial x) + \sin\theta(\partial f(x,y)/\partial y). \quad (7\text{-}75)$$

The BVQ V(f) is defined by (7-73) based on a concept of distribution, but (7-73) is also valid in an ordinary sense if the function f belongs to C^1-class. Moreover, in this case, the equation (7-74) is also valid and provides the same quantity as (7-73).

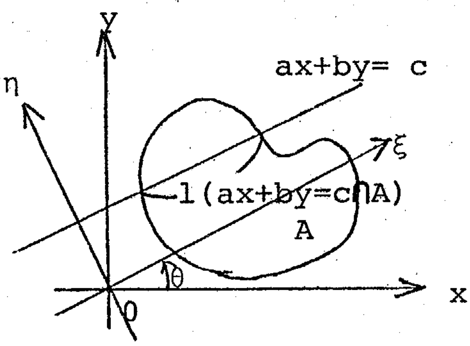


Fig.7-8. Observed area A.

The BVQ V(f) can be used as a comparatively excellent feature for the purpose of detecting the directionality of pictures. Let f be a picture-pattern and A be an obseved area (an open bounded domain, cf. Fig.7-8). The directionality of f in the area A can be detected as follows; Let the x-axis be $\theta = 0$, and let the counterclockwise direction be the positive direction of $\theta (0 \leqq \theta < \pi)$. A quantity $V(\theta)$ is defined for each direction θ, by using the concept of BVQ. Consider a straight line ax+by= c of direction θ. And calculate the average variation of f in A along ax+by=c,

$$l(ax+by=c \cap A)^{-1} \int_{ax+by=c\cap A} |\partial f/\partial \xi| d\xi. \quad (7\text{-}76)$$

This quantity will be changed when ax+by=c is moved parallel. We denote $V(\theta)$ for the average value of them in A. The quantity $V(\theta)$ provides the average variation of f in the θ-direction, and it has the following property. If the picture f has a directionality in the θ_0-direction (i.e. if the value of f doesn't change in the θ_0-direction), then we have,

$$\mathrm{Min}\,\{V(\theta)\,|\,\theta\} = V(\theta_0) = 0, \quad (7\text{-}77)$$

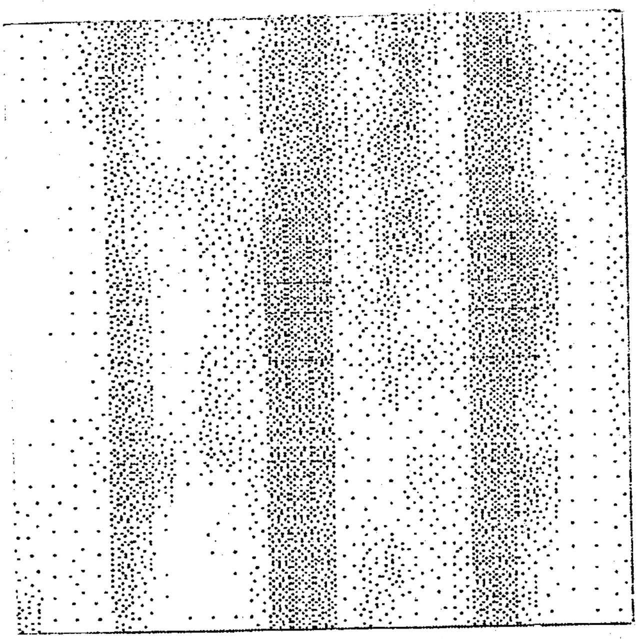
$$\mathrm{Max}\,\{V(\theta)\,|\,\theta\} = V((\theta_0 + (\pi/2))_{\mathrm{mod}\ \pi}). \quad (7\text{-}78)$$

Hence we can find the directionality of f by the fluctuation of the value $V(\theta)$. We can also guess the contrast and the coarseness of f in the θ-direction from the magnitude of the value $V(\theta)$. Moreover, it should be noted that the computation of $V(\theta)$ is very easy compared with methods such as the power spectrum method of Fourier-transformation [5]. This property of $V(\theta)$ will be understood by the example shown in Fig.7-9. Each picture has 32×32 pixels and each pixel has 256 gray-levels. $V(\theta_i)$ is calculated in 8 directions ($i= 1 \sim 8$).

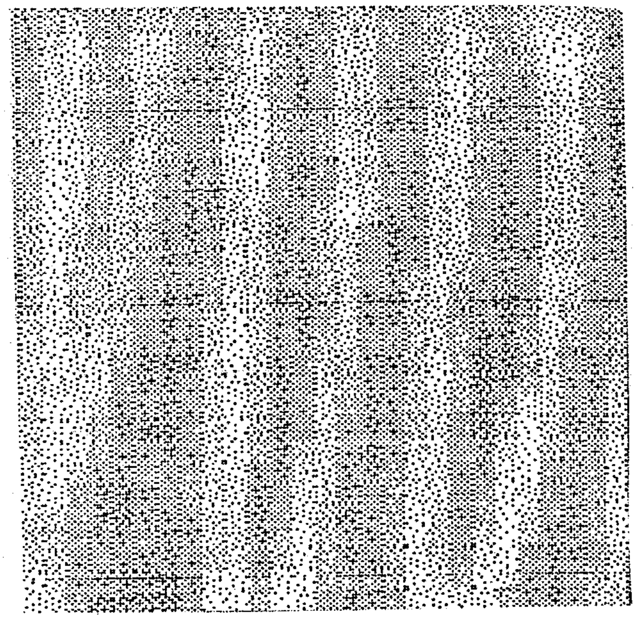
Of course, the BVQ itself is not a perfect feature of picture-patterns, and it is rather a local feature. In order to compensate this insufficiency, the integral quantity is useful. Textures of picture-patterns can be expressed as probabilistic sets to some extent by using avovementioned features for parameters. A simulation experiment of texture analysis is performed on electron-microscopic photographs of metals (Fig.7-10). Each picture is divided into 8×8 blocks and each block is constructed by 32×32 pixels and 256 gray-levels. In each block, the BVQ $\{V(\theta_i)\}_{i=1}^{8}$, and the average density (the integral quantity) are calculated. Each of these quantities corresponds to a parameter of the parameter space Ω,

$$\begin{aligned} &\Omega = \{\omega_1, \omega_2, \ldots, \omega_8, \omega_9\} \\ &\omega_i \propto V(\theta_i), \quad p(\omega_i) = 1/12, \quad \text{for } i = 1 \sim 8, \qquad (7\text{-}79) \\ &\omega_9 \propto \text{the average density}, \quad p(\omega_9) = 1/3. \end{aligned}$$

Each block is considered to be a probabilistic set, and 64 blocks (64 probabilistic sets) are classified into several regions (cf. Fig.7-10). It should be noted that since all operations in the calculation were very simple, the processing time was rather short (about 1 minute for each picture using a mini-computer and a FORTRAN program).



(a)



(b)

i	$V(\theta_i)$ of (a)	$V(\theta_i)$ of (b)
1	24793	30597
2	24673	29933
3	24438	29687
4	16229	20500
5	8590	11843
6	16454	22986
7	24487	29687
8	24524	32469

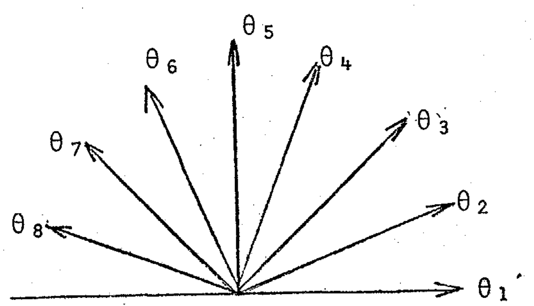
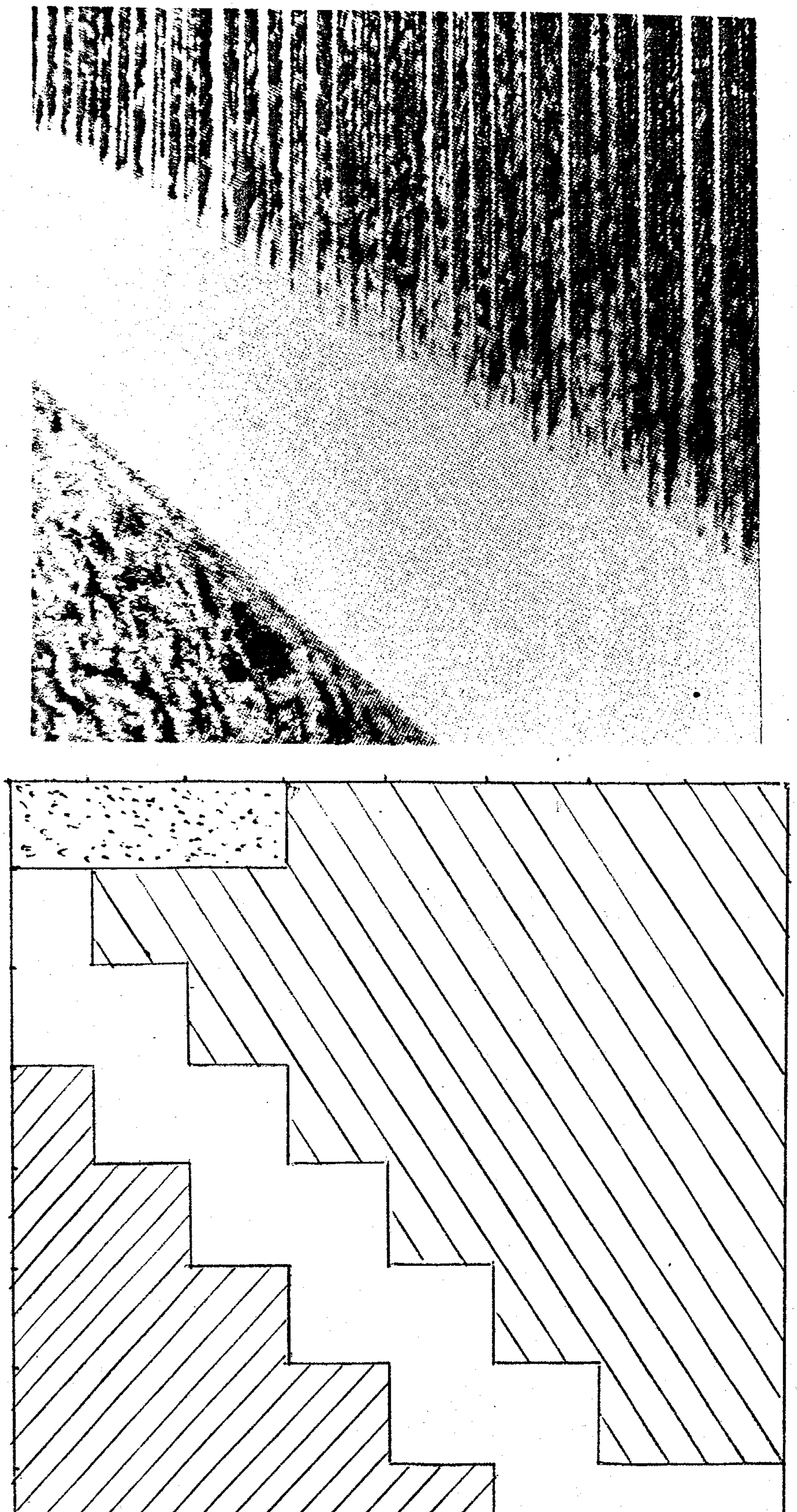
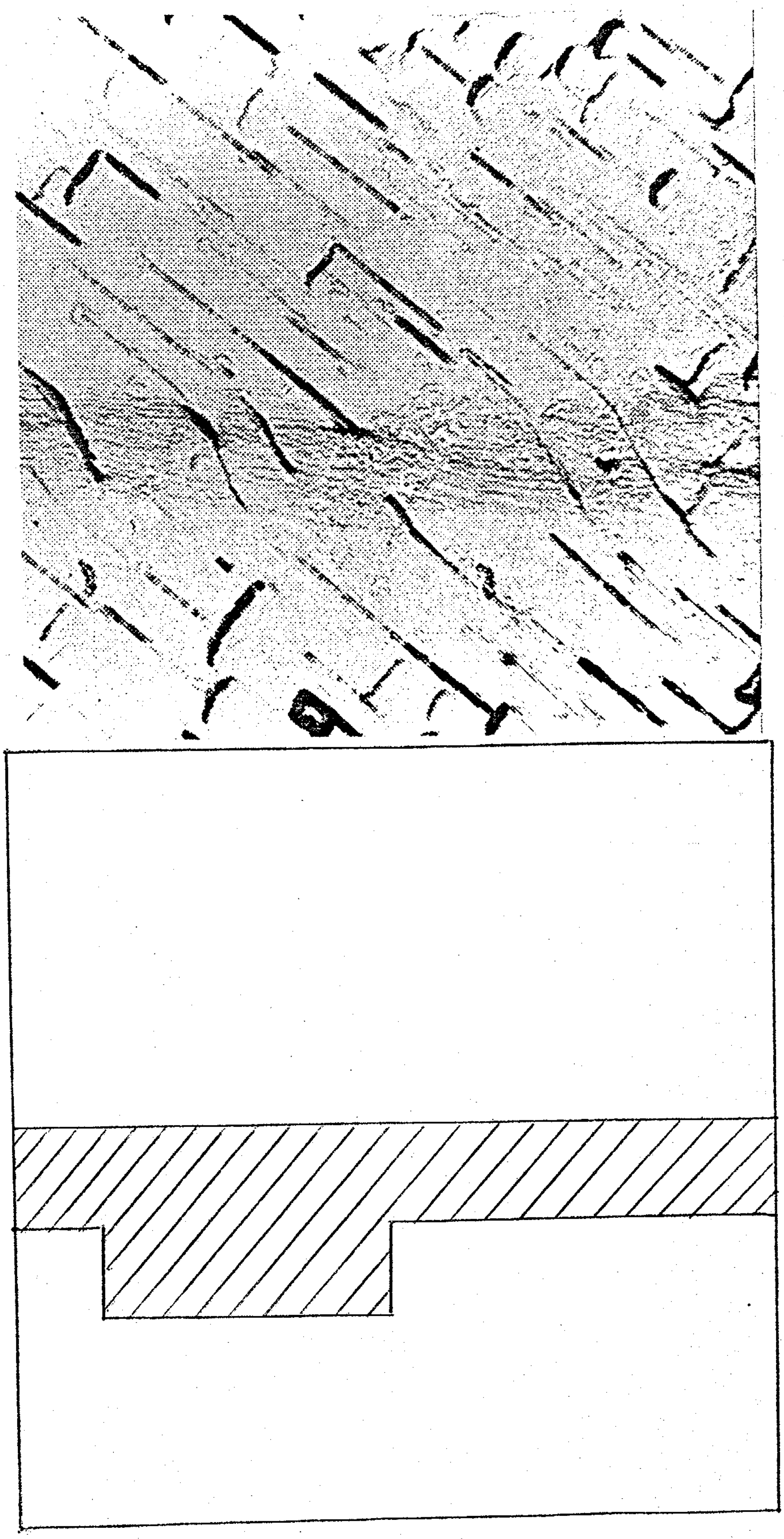


Fig.7-9. Ditection of directionality of picture-patterns.



(a)



(b)

Fig.7-10. Texture analysis of electron-microscopic photographs of metals.

CHAPTER. 8

CONCLUSIONS.

We have been investigating ambiguity and subjectivity in cognitive and decision processes mainly from a theoretical viewpoint. The outline of the results will be described below.

Firstly we attempted to obtain a new interpretation about the problem of ambiguity and subjectivity in human-oriented fields. And it was summarized as a concept of probabilistic sets. The fundamental idea was as follows: All the things we can interfere are expressed by concepts of probability, and the cases we can not intervene are unified by using fuzzy concepts. The theory was characterized as a complete pseudo-Boolean algebra from a lattice theoretical point of view. We also proposed two mutually equivalent expressions of probabilistic sets, i.e. probabilistic expression and extended fuzzy expression. As a result of these expressions, we could draw an interesting conclusion to the "fuzzy vs probability" controversy, i.e. the equality between the two was confirmed. From a viewpoint of information theory, the concept of Shannon's entropy was investigated in terms of probabilistic sets. It was shown that entropy was an important measure of ambiguity, but that there also existed other kind of ambiguity such as the notion of vagueness. Topological structure of probabilistic sets was also dealt with by several different methods. To clarify the description, four examples are given such as appraisement of recognition-performance of character readers.

Various unwieldy problems which arise despite one's pleasure have been visualized by our studies. Investigations from all approaches were being carried out and a new path has opend in the research of human-oriented problems such as decision making and pattern recognition in a broad sense. However this type of research has just been started and the careful examination of the results will lead to a new way in research. The following points are left as future problems: As we have already mentioned, OCR ASPET/71 is a good example of practical applications. Other concrete applications are now in the improvement stage, where multiple-parallel processing will play an important role by using plural micro-computers. But some time is necessary until these plans are put into practical use. There is no alternative but to wait for future research and development. The present auther will be very glad if his study is any help to the people concerned.

ACKNOWLEDGEMENT

The present author would like to acknowledge the continuing guidance and encouragement of Prof. Taizo IIJIMA (T.I.T.).

He is very grateful to Prof. H.ENOMOTO, Associate Prof. M. SHIMURA, Associate Prof. M.SUGENO and Associate Prof. H.OGAWA (T.I.T.) who supervised this thesis.

He is also very grateful to Prof. H.UMEGAKI, Prof. T.TERANO (T.I.T.), Prof. K.HORIUCHI (Waseda Univ.), Prof. M.M.GUPTA (Univ. of Saskatchewan), Prof. L.A.Zadeh (Univ. of California), Dr. A. IGARASHI (Electro-technical laboratory) and Associate Prof. H. ETOH (Tsukuba Univ.) for their continuing encouragement.

He wishes to thank Associate Prof. Y.ISOMICHI (Hiroshima Univ.), Dr. M.OZAWA (T.I.T.), Prof. H.NGUYEN (Univ. of Massachusetts), Mr. W.PEDRYCZ (Univ. of Gliwice) and Mr. I.TOKUHIRO (Waseda Univ.) who made useful discussions and helped the present study.

He also wishes to thank Dr. A.NUSHI (Tokyo Univ.), Ms YOKOHAMA (Tsuda Women's Univ.), Ms SAKAI (Denka Co.Ltd.) who contributed to the questionnaire, and Mr. ARAKI (Waseda Univ.), Mr. KUROSAWA (Toshiba Co.Ltd.), Mr. NITTA, Mr. NAKAJIMA (T.I.T.) and other members of Prof. IIJIMA's Lab. who contributed to the experiments.

It is a pleasure to acknowledge the hospitality and encouragement of the members of the informal, theoretical group of pattern recognition; Assosiate Prof. S.KONDO (Tokai Univ.), Dr. S.KAWARAI (Anritsu Co.Ltd.), Dr. K.KAMOI (Asahi-kogaku Co.Ltd.), Mr. OHISHI, Mr. KITANO (Waseda Univ.).

This research was supported financially in part by the YAMADA Foundation. (The title is "Concepts of Probabilistic Sets and their Applications to Pattern Recognition".)

Finally, a personal note of gratitude goes to my parents and Ms M. SHIRATORI without whose encouragement and forbearance this thesis would not have been accomplished.

BIBLIOGRAPHY

CHAPTER 2

[1] L.A.Zadeh:"Fuzzy Sets" Inf. and Control, vol.8, pp.338-353, (1965)

[2] K.Hirota:"Kakuritsu-Shugoron to Sono Oyo-rei (Probabilistic Sets and its Applications)" The Behaviormetric Society of Japan 3-rd Conf., pp.24-27, (1975) (in Japanese)

[3] T.Iijima:"Pattern Recognition" Corona-sha Publ.,Tokyo,(1973)

[4] T.Iijima and et.al.:"A Theory of Character Recognition by Pattern Matching Method" Proc. of 1-st IJCPR, pp.50-56,(1973)

CHAPTER 3

[1] K.Itoh:"Kakuritsu Ron (Probability theory)" Iwanami Press, Tokyo,(1972) (in Japanese)

[2] J.L Doob."Stochastic Processes" John Wiley & Sons,Inc., New York, (1953)

[3] A.N.Kolmogoroff:"Grundbergriffe der Wahrscheinlichkeitsrechnung" Springer Verlag, Berlin (1933) (in German)

[4] L.A.Zadeh:"Fuzzy Sets" Inf. & Control, vol.8, pp.338-353 (1965)

[5] P.R.Halmos:"Measure Theory" Van Nostrand, p.110, New York (1950)

[6] G.Birkhoff:"Lattice Theory", AMS.Colloq. Publ., New York (1969)

[7] T.Iwamura:"Soku Ron (Lattice Theory)" Kyoritsu Press., Tokyo, (1966) (in Japanese)

[8] P.R.Halmos:"Measure Theory" Van Nostrand, p.110, New York, (1950)

[9] K.Yoshida:"Functional Analysis" Springer Verlag, Berlin, (1970)

[10] Y.Isomichi:"The Stochastic Theory of Learning Systems" Electro-technical Laboratoty Technical Report, No.750, p.31, (1975) (in Japanese)

[11] K.Hirota:"Probabilistic Set Theory", IECEJ PRL76-36, pp.1-10 (1976) (in Japanese)

[12] G.F.Simmons:"Introduction to Topology and Modern Analysis", McGraw-Hill, New York, p.218, (1963)

[13] P.R.Halmos:"Naive Set Theory" Van Nostrand, New York, (1950)

[14] K.Shiga:"Set, Topology,Measure", Hirokawa Press, Tokyo, (1971) (in Japanese)

[15] R.T.Yeh & S.T.Bang:"Fuzzy relations, Fuzzy Graphs, and their Applications to Clustering Analysis" Fuzzy Sets and their Applications to Cognitive and Decision Processes, Academic Press, pp.125-149, (1975)

[16] E.H.Ruspini:"A Theory of Fuzzy Clustering" IEEE Conf. on Decision and Control, 77CH 1269-0SC, pp.1378-1383 (1977)

[17] Feys, Robert:"Modal Logics", Gauthier-Villars (1965)

[18] Rischer, Nicholas:"Many-valued Logic" McGraw Hill (1969)

[19] S.Watababe:"Knowing and Guessing" John Wiley ans Sons,(1969)

[20] T.Kitagawa:"Fuzziness in Informative Logics" Fuzzy Sets and their Applications to Cognitive and Decision Processes, Academic Press, pp.97-124 (1975)

[21] L.A.Zadeh:"Fuzzy Sets" Inf and Control, vol.8, pp.338-353 (1965)

[22] L.A.Zadeh:"Probabilistic Measure of Fuzzy Events",J. of Math. Anal. and Appl., 23, pp.421-427, (1968)

[23] M.Mizumoto & K.Tanaka:"Some Properties of Fuzzy sets of Type 2" Inf. and Control, vol.31, pp.312-340, (1976)

[24] G.Birkhoff:"Lattice Theory" AMS Colloq. Publ., New York (1969)

[25] J.A.Goguen:"L-Fuzzy Sets", J.of Math. Anal. and Appl., 18 pp.145-174, (1967)

[26] K.Nanba:"Shugo Ron (Set Theory)" Science-sha Publ., Tokyo (1975) (in Japanese)

CHAPTER 4

[1] Haviland:"On the Inversion Formula for Fourier-Stieltjes Transforms in more than One Dimension", Amr.J.Math., 57 (1935)

[2] P.R.Halmos:"Measure Theory" Van Nostrand, p.110, New York (1950)

[3] T.Takagi:"Kaiseki Gairon" Iwanami Press, Tokyo (1950) (in Japanese)

[4] K.Yoshida:"Theory of Hilbert Space" Kyoritsu Press., Tokyo p.89, (1953) (in Japanese)

[5] S.Kakeshita:"Probability Theory" Science-sha Publ., Tokyo, p.89, (1973) (in Japanese)

[6] K.Yoshida:"Functional Analysis" Springer Verlag, Berlin, p.346, (1971)

[7] L.A.Zadeh:"Fuzzy Sets" Inf. and Control, vol.8, pp.338-353, (1965)

[8] P.R.Halmos:"Measure Theory" Van Nostrand, New York, p.148, (1950)

[9] K.Itoh:"Probability Theory" Iwanami Press, Tokyo, p.33, (1972) (in Japanese)

[10] Y.Komatu:"Kaiseki Gairon I" Hirokawa Publ., Tokyo, p.265, (1962) (in Japanese)

CHAPTER 5

[1] C.E.Shannon & W.Weaver:"The Mathematical Theory of Communications" Univ. of Illinois Press, Urbana (1949)

[2] A.D.Fadeev:"On the notion of Entropy of Finite Probability Space" Uspehi Math.Nauk, 11, pp.227-231,(1956) (in Russian)

[3] H.Mine:"Introduction to Information Theory", Asakura Press, Tokyo, p.51, (1964) (in Japanese)

[4] H.Totoki:"Introduction to Ergodic Theory" Kyoritsu Press, Tokyo, p.85, (1966) (in Japanese)

CHAPTER 6

[1] J.L.Kelly:"General Topology" Van Nostrand, New York (1955)

[2] Van der Waerden:"Moderne Algebra" Springer Verlag, Berlin (1937) (in German)

[3] P.R.Halmos:"Measure Theory" Van Nostrand, New York (1950)

[4] C.L.Chang:"Fuzzy Topological Spaces" J.of Math. Anal. and Appl. 24, pp.182-190, (1968)

[5] C.K.Wong:"Covering Properties of Fuzzy Topological Spaces" J. of Math. Anal. and Appl., 43, pp.697-704, (1973)

[6] R.Lowen:"Initial and Final Fuzzy Topologies and the Fuzzy Tychonoff Theorem" J. of Math. Anal. and Appl.,58,pp.11-21, (1977)

[7] M.Sugeno:"Fuzzy Measure and Fuzzy Integral" Society of Instrment and Control Engineers, Vol.8, No.2, pp.94-102, (1972) (in Japanese)

[8] J.L.Kelly:"General Topology" Van Nostrand, New York (1955)

[9] K.Hirota:"Introduction to Subjective Topology on Probabilistic Sets" IECEJ PRL77-41, pp.1-10, (1977) (in Japanese)

CHAPTER 7

[1] M.Nishimaki et.al.:"Special Issue on Optical Character Recognition" J. of IECEJ, vol.61, No.2, (1978) (in Japanese)

[2] T.Iijima et.al.:"A Theory of Character Recognition by Pattern Matching Method" Proc. of 1-st IJCPR, pp.50-56,(1973)

[3] T.Iijima & K.Mori:"A New Generation of OCR-ASPET/71" J. of Nikkei Electronics, No.30, pp.66-80, May (1972),(in Japanese)

[4] K.Hirota & T.Iijima:"The Bounded Variation Quantity (BVQ) and its Application to Feature Extractions" Proc. of 4-th IJCPR, pp.457-461, (1978)

[5] R.Bajsy et.al.:"Computer Description of Real Outdoor Scenes" Proc. of 2-nd IJCPR, pp.174-179, (1974)

APPENDIX I

Summary of an invited lecture given by the present author at Takenaka-hall, Ginza, Tokyo in Sept. 14-th 1978.

パターン認識の研究と独創性

広田　薫

私がまだ小学生であった頃、すなわち現在から遡ること10数年前のことです。私の社会科の先生が、欧米先進諸国の経済や産業がいかに進んでいるかを説明し、アジアの多くの後進国の中における日本の状況などを、口角あわをとばしてしゃべってくださったことを、ときどき、つい昨日のことであったかのように思い出します。しかし、それはもう二度と帰らぬ、私にとってとても楽しくかつなつかしい思い出の一つです。

さて、ここ数年、高度成長をなしとげた日本は、経済大国として、欧米先進諸国を追いぬき、追い越しすぎて勇み足をふんだりする一方で、開発途上国という丁寧な言葉を用いて、それらの国々を援助指導するという、先進国としての自覚を持つようになりました。そこで、ここでは、私の仕事である学問や研究、特に情報科学という分野に焦点をあわせ、日本の立場を考察し、問題を明確にしたうえで、私が現在研究をしているパターン認識という分野を紹介してみたいと思います。

現在の日本の研究の世界的にみたレベルは、分野にもよるでしょうが、かなり高いものと考えてよいでしょう。一例として、数学の関数解析や数値計算で有名な東大の藤田宏教授によれば、「昨年マレーシアで開かれた東南アジアの数学関係の国際会議で、第一線の研究者がやっている研究が、本郷の学生の卒論や高々修論程度であった。」ということであり、「日本からも、若手の数学者が指導に行くべきだ。」とのことです。それでは、現状に満足していて良いのでしょうか？

もちろん、上を望めば限がありませんから、これで充分と満足するわけにはいきません。より質の高い学問や研究が望まれます。私は、研究において、最も大切なことは、独創性ということだと思います。いわゆる、無から有を生じさせる能力が、どれだけその研究者にあるかということです。このことは、電子回路という言葉の名付け親であり、国大協でも活躍された私の大先輩、川上正光先生が、東工大学長時代に、特に強調されていたことでもありました。この観点から、日本における情報科学研究をみた場合、まだまだ、合格点をあげるわけにはいかないように思えます。学会誌や論文誌にみられる研究は、例えば、「米国のある研究機関で、ごく最近新らしい方向への研究が始まり、我々のところでも、その路線で研究を進めた結果、彼等が予想していたよりはずっと良好な結果を得ることができた。」という感じのものが極めて多いと感じているのは私だけでしょうか？つまり、小さな有を得て、大きな有に育てあげるという努力が多いということです。日本人は、生来真面目であり、良く勉強をします。例の受験勉強方式で、鍛えあげられたこともあって、海外の文献もできるだけ多く読もうとします。その結果その文献の中に問題点をみつけ出し、自分ならこうするということで、研究を開始し、悪戦苦闘の末に良い結果を得て発表をするということになるのでしょう。そのために、新らしく研究を始めようという者にとっては、日本人のかなり完成した研究よりも、海外の未完成だが興味ある研究のほうが魅力があり、ややもすると、日本人の研究を軽視し、欧米人の研究をありがたがるということになりがちです。このような態度が続くかぎり、本当の意味での先進国にはなれません。

無から有を生じさせ、新らしい研究の芽生えを、日本人自らの手で作り出すためには、独創性というものをもっともっと大事にし、こわさないということが肝要です。例えば、人文社会系統の日本人研究者の方々のお話を聞いたり、論文を読んでいつも感じるのは、引用があまりに多いとこうことです。ときには、誰がどんなことを言ったかを、どれだけ多く知っているかということが、その研究者の能力を示す唯一の指標であると、考えているのではないかと思いたくなるほどです。無から有を生じさせて、かつ価値のある独創的な研究というものは、本来引用文献など必要のないものであるはずです。ニュートンが、リンゴの落下から万有引力の概念を得たように、あるいは、アルキメデスが、溢れそうになった湯船に入って、比重の概念を得たように、ヒントは、身のまわりの自然界や人間の行動現象にあるといえます。他人の論文からヒントを得ようという精神では、第一級の研究は望めません。

しかし、このような態度で望んだら、すぐに第一級の研究の見返りが得られるかというと、決してそうではありません。仮に、1万回バッターボックスに立ったとしても、1本のヒットが打てれば良いほうであり、ホームランなどはまず望めません。それどころか、もし1回で

も、ボールがバットをかすってくれれば、喜ぶべきであり、１万回の連続三振の可能性がかなり濃厚と思っていたほうが良いでしょう。真の研究とは、それほど難かしくかつ厳しいものです。さらに、バッターボックスに立ったとして、バットにボールが当たったとしても、ヒットになるか否かが判明するのに、かなり時間がかかります。良い当たりをしたと思っても、世の中には、非常に巧妙なプレイをする野手がいるのです。ファーストまであと一歩というところまで待っていて、その瞬間にアウトにするという、いやらしいものもおります。

また、その場合には、正しい判断を下す能力のある日本人審判が、是非とも必要です。ノーベル賞物理学者の江崎玲於奈博士が、ソニー時代に、例のエサキダイオードの基本特性を引っ提げて、東工大で開催された日本物理学会で発表したら、実験の誤りと一笑されながら、改良して米国のフィジカルレビュー誌に発表したら、トランジスターの父ショックレーの絶賛を受け、世界的に有名となったという例は、日本人研究者として残念に思います。横文字の文献だけでなく、日本人の研究でも良いものを見抜き尊重するという態度を持たなくてはいけません。私の指導教官飯島泰蔵教授は、パターン認識関係で独創的手法ですぐれた研究を多く残していますが、論文は日本語で書き、欧米の研究者に自分の論文を読ませるために、日本語の勉強させるくらいの気概を持っています。自分では、第二級の研究しかしていないのに、国際交流が必要だと称して、下手な英語をふりまわして、海外を飛びまわる研究者は、自分の行動を一考する必要があるでしょう。斯く言う私自身も、国際会議と称して、米国まで出かけていったほうであり、自分の日本語の論文を外人に読ませるまでには、まだ徹しきれておりませんから、大いに反省しなければなりません。

以上述べたことから、日本が研究面でも先進国であるためには、無から有を生ずる質の良い研究を、もっともっと心掛け、海外の研究者に日本の研究動向を注視されるようになることが必要といえるでしょう。そのためには、効率の良い利益の多いことばかりでなく、研究にも冗長性を取り入れ、あえて無駄なことをする勇気を持つこと、およびそのような態度を育む土壌を作ることが不可欠であるといえます。

さて、これまでは、評論家の立場で偉そうなことを言ってきましたが、研究者は、単に問題点をみつけ出して批判をし、こうあるべきだと主張するだけではいけません。ましてや、自分の主張と同じ路線で世論が動きだし、その結果が不幸にも行き詰まりを生じ、どうしようもなくなった場合に、なるべく表に出ないようにして、人々の熱がさめるのを待ち、体面を保つというような、責任の無い行動は取るわけにはいきません。自らの主張を、自らの手で実行し、実現を第一に考えるべきです。仮に失敗した場合は、全責任を負うという覚悟で、深い洞察力を伴った主張を、すべきです。

さて、以上のような考え方のもとで、私自身が研究を進めているパターン認識というものを、わかりやすく説明したいと思います。パターン認識と称するものを研究する者の数は、現在では国内でも数100名、海外も考えれば極めて多くにのぼるものと思われます。情報科学の一分野として、研究が始まったのは、電子計算機が出現してから10数年後の60年代初期ですから、歴史は20年弱であり、最近ようやくパターン認識というものを、一つの学問としてとらえようという段階まで来たというのが現状です。研究関連分野は、単に工学だけでなく、数学や物理などの理学、心理学や哲学などの人文社会学、それに生理学や医学など、極めて広範囲にわたります。私自身も、普段はかなり難かしい事柄を扱っているわけですが、ここでは、基本的に重要な事柄のみ集約し、原点に帰って考え直してみることにします。

前置きが長くなりましたが、パターン認識とは何でしょうか。まず、パターンとしては、例えば、文字や音声を考えればよいでしょう。そのようなパターンを見たり聞いたりして、それが何という文字かあるいは音かを識別することがパターン認識といえます。単に認識といえば、哲学では、それだけで一つの分野をなしているほどに意味深長なものですが、パターン認識といった場合は、人間が関与している事が重要です。先日、哲学をやっている東大の村上陽一郎助教授に、パターン認識を、どう定義するかと尋ねたところ、人間の行なう認識であると答えてくださいましたが、ある程度はあたっていると思います。けれども、これではまだ認識という言葉をそのまま用いており、妥当な定義というわけにはいきません。

そこで、私自身は、パターン認識というものを、次の

ように定義しています。すなわち「扱うべき対象（これをパターンと呼ぶ。）と、それらを観測するもの（これを認識主体という。）が存在し、主体がその対象を観測することにより、自己の内部にその情報モデルを構築すること。」を、パターン認識と定義します。つまり、パターン認識では、単にパターンを観測するだけではなく、その結果として、自分の内部にそれに関する何等かの情報を再構成し、自分なりの解釈を与えることが重要なことになります。従って、客観的には同一の対象であっても、主観が入ることにより全く別のものとみられることもありうるわけです。従って、何の意識もなしに、ただ目前にある風景が目に入ってくるという状態では、パターン認識を行なっていないことになります。この事情は、ハワイ大学の渡辺慧教授も、その著書の「認識とパタン」（岩波新書36）の中で、「見る」ことと、「見なす」ことのちがいとして、うまく表現しています。

パターン認識は、以上述べてきたように、人間の有する非常に優れた機能の一つであるわけです。従って、次に来る考えとして、その機能、（できれば人間のレベルよりもっと進んだ形で）人工的に実現するにはどうしたらよいか、ということが当然生じてくるはずです。いわゆるパターン認識機械の構成問題であり、現在極めて多くの研究者がこの問題に取り組んでいます。初期のころには、高度に発達した電子計算機を用いれば、比較的簡単に実現できるだろうと考えられていましたが、実際にやってみると意外に難かしいことがわかりました。現在、どの程度の機械が実用化されているかを、紹介しましょう。文字の読取装置に関しては、郵便番号読取り機がよく知られていますが、対象は０から９までの10数字です。印字の場合は、もう少し字数が多くて、数字とアルファベットそれにいくつかの特殊記号の識別装置がありますし、最近では、当用漢字の読取り装置もできています。しかし、手書き文字の場合は、まだまだ実用段階にはいっておりません。一方、音声に関しては、新幹線の座席予約などの、限定語に関する基本的研究はなされておりますが、やはり、私達人間が他人の声を聞いて識別をするような高度な機械はできておりません。

私達が、日常、いとも簡単に文字を判読し、音を聞きわけているのに、あの電子計算機を用いてさえパターン認識機能が実現できぬ理由はどこにあるのでしょうか。それは、計算機の得意とする「明確なアルゴリズムとして捉えたものを、高速で実行する。」という能力と、人間のパターン認識能力が、根本的に異った性格のものだからといえましょう。すなわち、パターン認識の場面では、計算機で実行するための大前提である明確なアルゴリズムを組むことができないのです。人間は、一つの与えられた文字を見て、この特徴があってかつ、あの特徴がないから、アという文字であるというように、一つ一つの特徴を丹念に調べていって、その結果はじめて識別を行なうというような、プログラムに書ける行動を通じて認識をしているわけではありません。全ての特徴を、一つ一つ逐次的に調べていったら、その量は天文学的な量になってしまい、とても実行できません。ただ、その極めて多くの特徴量の中から、適当なものをうまく取捨選択して、識別をしています。しかし、その取捨選択の過程は、決して明確なものではなく、時には、いわゆる「第六感」と称するものまで動員して実行しています。この、第六感と称する能力は、計算機には殆んどありません。従って、もともと性格が全く異なる計算機を、だましながら用いて、何とか人間のパターン認識機能を実現しようとする現在の研究方法では無理があり、人間の有する能力よりかなり低い所にひかれた一定水準を超すことは難かしいと考えます。もう少し、原点に戻った発想の転換が必要でしょう。このような理由に基ずき、私自身は現在、従来の研究方針とは、かなり異った立場から、研究を進めております。現在の電子計算機と性格の異なる、別の情報処理機械を構成するには、まだ極めて多くの問題点を解決する必要がありますが、その基礎理論の基本的考え方だけ、言及したいと思います。

まず、人間のパターン認識の場面において、従来は扱いにくいという理由でなるべく避けて通ってきたが、私自身は本質的に重要と考えているいくつかの問題点を列挙し、私の考え方を紹介しようと思います。

電子計算機の基本は、yes と no の二値論理だといわれています。通常 yes は１に、no は０に対応させ、計算機の内部では、１と０の２種類の数の組合せだけで動作をします。これを、パターン認識の場面にあてはめて考えてみますと、あるパターンが、考えている性質を有するときが１であり、逆に有していないときが０ということになります。しかし、一般的には性質の有無は、０

と１二値ですむほど単純なものではありません。むしろ、我々は、性質の有無がはっきりしている場合よりも、大体そうらしいとか、違うかもしれないとか、あるいは全くわからないというような、曖昧な判断をすることのほうが多いといえます。従来は、この曖昧な場合でも、無理やり０か１にしたりして、曖昧状態を避けて通ってきました。しかし、私はこの曖昧さこそ、人間の持つ一つの本質的機能と考え、積極的に取りあげるべきだと主張します。しかし、この曖昧さを認めることになると、従来の数学では許容されないことになり、理論の基礎がぐらつくという大変な事が生じます。

従来の研究の曖昧さに関する消極的態度の原因の一つは、このへんにあるものと私は考えます。しかし、私は、数学は与えられるものではなく、むしろ我々に都合の良いように造り出し、その造り出したものを、あたかも道具のように自由に使いこなすものと考えています。(しかし、現代数学は高度に抽象化し、難かしくなりました。理論研究者の中には、数学を与えられたものと思い、逆に数学によって、道具のようにふりまわされている者がいるように感じているのは私だけでしょうか？）この観点から、パターン認識というものを征服するのに便利なように、使いやすい道具としての数学を造り出すことを、私は自分自身の研究の一部にしております。

さて、曖昧さという重要性を指摘しましたが、一口に曖昧さと言っても、その構造は多様性に富んだかなり複雑なものです。例えば、パターン認識の定義のところで述べたように、同一の対象であっても、観測者によって別のものとみなされることもあります。単なる白い布というパターンが、白い布と認識されたり、幽霊と認識されたりするという場面を考えればよいでしょう。この他にも、多くの例が考えられますが、とにかくパターン認識の場面では、主観というものが極めが重要な要因となってきます。しかし、自然科学が、客観の学問といわれてきたように、主観（或いは個性）というものを積極的に取りあげることには、極めて大きな反発があります。けれども、私自身は、主観を取りあげないパターン認識学はありえないと考えています。従って、主観や個性を積極的に反映した理論の研究も、私の研究テーマの一つとして選ばれております。

ところで、認識機械を実現するためには、主観や個性の構造解析をする必要があります。主観や個性は、過去の経験などにより培われた非常に複雑な多様構造を持つものと考えられますが、常に学習や忘却という時間的変化（これを、私は知識の進展と呼んでいます。）がみられるということも重要な事柄の一つです。

この他にも、問題点は多く考えられますが、既成の手法のみでは扱いきれない曖昧さや主観というものを解明していくことこそ、パターン認識研究の進むべき方向と考えております。数式を全く用いないで、考え方の基本のみを述べたため、抽象的でわかりにくくなったかもしれません。あるいは、逆に、あたりまえの事しか言っていないと感じられたかもしれません。もし、あたりまえの事と感じていただければ、私自身、大変うれしく思います。というのは、私は、学問というものは、やたらに難かしそうなカモフラージュを全て取り去り、裸にすれば、全ての人があたりまえの事として認めるシンプルなものであるべきと考えるからです。ごく一部の研究者集団にしか理解されぬ学問は、その研究者自身が本質を理解していないか、裸にしたら何も残らなくなるため、あえてそのままにしてありがたがっているものと考えます。

現在、私は、パターン認識研究チームの打者として、あえて皆さんの目の前でバッターボックスに立ちました。はたして、１万回に１回のヒットが打てるでしょうか？その結果が判明するのは、まだまだ先の事です。しかし、私の打席には、とある財団から数100万円の研究助成も懸かっています。何として、打たねばなりません。「有言実行」私の好きな言葉の一つです。

（昭和53年９月14日竹門会ホールにて）

APPENDIX II

PUBLICATIONS LIST OF THE PRESENT AUTHOR

Master thesis

Kakuritsu Shugo Iso Kukanron to sono Oyorei ni kansur Kiso-kenkyu (Fundamental studies of Probabilistic, topological spaces and their applications) (in Japanese)
Tokyo Institute of Technology, 1976

Doctoral dissertation

Analytical Expression of Ambiguity and Subjectivity in Cognitive and Decision Processes
Tokyo Institute of Technology, 1979

Papers (in English)

1) Concepts of Probabilistic Sets
   IEEE Int. Conf. on Decision and Control, New-Orleans USA, 77CH 1269-0SC, pp.1361-1366, 1977
2) Probabilistic sets - Expansion of Fuzzy Concepts based on Probability Theory -
   Summary of papers on general Fuzzy problems No.3, Working group on Fuzzy Systems in Tokyo, pp.135-153, 1977
3) A decision making Model - A new approach based on the concepts of probabilistic sets - (IIJIMA)
   IEEE Int. Conf. on Cybernetics and Society, Tokyo, 78 CH 1306-0SMC, pp.1348-1353, 1978
4) The Bounded Variation Quantity (B.V.Q.) and its application to Feature Extractions (IIJIMA)
   The 4-th Int. Joint Conf. on Pattern Recognition, Kyoto pp.456-461, 1978
   Pattern Recognition (the Journal of the Pattern Recognition Society) (invited and in preparation)
5) Extended Fuzzy Expression of Probabilistic Sets - Analytical Expression of Ambiguity and Subjectivity in Pattern Recognition -
   Seminar Report on Applied Functional Analysis, The Math. Society of Japan, (to appear)

6) Concepts of Probabilistic Sets
   Int. Journal of Fuzzy Sets and Systems (submitted)

Book (in English)

1) Extended Fuzzy Expression of Probabilistic Sets
   A part of a book entitled "Advances in Fuzzy Set Theory and Applications", Edited by M.M.Gupta et al., North-Holland Publ. Comp. (in prepatation)

Papers (in Japanese)

1) Logical basis in probabilistic set theory - Probabilistic expression of ambiguity and subjectivity -
   IECEJ papers (D) Feb. 1979 (to appear)
2) Analysis of single probabilistic set based on the concept of subjective entropy -Theoretical approach to ambiguity and subjectivity using Shannon's entropy -
   IECEJ papers (D) Feb. 1979 (to appear)

Papers (in Japanese & in Oral Publication)

The Institute of Electronics and Communication Engineers of Japan

Dept. of Pattern Recognition and Learning

1) A decision making model using the probabilistic set theory
   IECEJ PRL76-21, pp.19-28, July 1976 (Hiroshima Univ.)
2) Probabilistic set theory
   IECEJ PRL76-36, pp.1-10, Oct. 1976 (Tohoku Univ.)
3) Probabilistic set theory (II)
   IECEJ PRL76-63, pp.19-28, Jan. 1977 (Osaka Univ.)
4) The learning theory considering subjectivity ambiguity and evolution of knowledge
   IECEJ PRL77-10, pp.41-50, May 1977 (Nagoya Univ.)
5) Introduction to subjective topology on probabilistic sets
   IECEJ PRL77-20, pp.1-10, July 1977 (Kanazawa Univ.)
6) Picture analysis by bounded variation method
   IECEJ PRL77-41, pp.25-34, Oct. 1977 (Kyushu Univ.)
7) Analytical expression of ambiguity and subjectivity in pattern recognition
   IECEJ PRL78-25, pp.23-32, July 1978 (Hokkaido Univ.)
8) Analysis of single probabilistic set based on the concept

of subjective entropy - Theoretical approach to ambiguity and subjectivity using Shannon's entropy -
IECEJ PRL78-43, pp.1-10, Oct. 1978 (Tohoku Univ.)
9) Logical basis in probabilistic set theory - Probabilistic expression of ambiguity and subjectivity -
IECEJ PRL78-44, pp.11-20, Oct. 1978 (Tohoku Univ.)

IECEJ Zenkoku-taikai
1) Kakuritsu Shugo-ron to Tegakimoji Shikibetu Chosa heno Oyo
S.51, Vol.5, No.1169, March 1976 (T.I.T.)
2) Kakuritsu Shugo-ron
S.52, Vol.1, No.27, March 1977 (Waseda Univ.)
3) Pattern Ninshiki ni okeru Shukan-Joho-ryo
S.52 Joho-bumon, No.3, Aug. 1977 (Tohoku Univ.)
4) Yukai Hendo-ryo ni motozuku Gazo no Hoko-sei Kenshutu-ho oyobi Texture Kaiseki heno Oyo
S.53, Vol.5, No.1012, March 1978 (T.I.T.)

The Mathematical Society of Japan
1) Kakuritsu Shugoron to Shukan Iso
1976 Shuki Sogo Bunka Kai, Oyo-Sugaku, Oct.1976, (T.I.T.)
2) Pattern Ninshiki ni okeru Shukan Iso to Shukan Joho-ryo
1977 Shunki Sogo Bunka Kai, Oyo-Sugaku, Apl. 1977, (Kyoto Univ.)
3) Kakuritsu Shugo no Kakucho Fuzzy Hyogen
Seninar on Applied Functional Analysis, July 1978, (Tsukuba Univ.)

Monbu-sho Kagaku Kenkyu-hi Kenkyu
1) Kakuritsu Shugoron
S.51 Monbusho Kakenhi Hokoku, "Fuzzy system and Artificial Intelligence" pp.193-213

The Operations Research Society of Japan
1) Fuzzy Iso to Bunri Kori
1975 Shunki Kenkyu Happyo Kai, pp.114-115, Apl. 1975 (Tokai Univ.)
2) Kakuritsu Shugoron to Ishi Kettei Model
1977 Shunki Kenkyu Happyo Kai, pp.99-100, March 1977 (Waseda Univ.)

Society of Instrument and Control Engineers

1) Kakuritsu Shugoron to sono Oyo Rei II
   18-th SICE, pp.91-92, Nov. 1975, (Kokuritsu Kyoiku Kaikan)

The Behaviour-Metric Society of Japan

1) Kakuritsu Shugoron to Sono Oyo Rei
   3 kai Sokai, pp.24-27, Sept.1975, (Aoyama Gakuen Univ.)
2) Tegaki Moji Shikibetsu Chosa heno Kakuritsu Shugoron no Oyo
   4 kai Sokai, pp.124-127, Sept.1976, (Tokyo Univ.)
3) Shukan Johoryo ni yoru Kodo Gensho no Keiryo
   5 kai Sokai, pp.5-6, Sept. 1977, (Okayama Univ.)

APPENDIX III

PERSONAL HISTORY OF THE PRESENT AUTHOR

Name	HIROTA Kaoru
Nationality	Japanese
Date of Birth	January 6, 1950
Place of Birth	Niigata Prefecture, Japan
Education	1965-1970 Dept. of Electricity, Nagaoka Technical College
	1970-1974 Dept. of Electronics, Faculty of Engineering, Tokyo Institute of Technology
	1974-1976 Dept. of Electrical Engineering, Graduate School of Engineering and Science (Master Course), Tokyo Institute of Technology
	1977-present Dept. of Electrical Engineering, Graduate School of Engineering and Science (Doctor Course), Tokyo Institute of Technology
Address of TIT	Prof. IIJIMA's Lab., Dept. of Computer Science, O-okayama, 2-12-1 Meguro-ku, Tokyo 152, Japan
Home Address	1-2-6 Nakanuki-cho, Nagaoka-city, Niigata Prefecture 940, Japan
Professional Society	The Institute of Electronics and Communication Engineers of Japan
	The Mathematical Society of Japan
	Society of Instrument and Control Engineers
	Information Processing Society of Japan
	The Operations Research Society of Japan
Activity	Reviewer, International Journal of Fuzzy Sets and Systems
	Member, Working Group of Fuzzy Systems, Tokyo AVIRG (Audio Visual Information Research Group)

SUBJECT INDEX