

行政院國家科學委員會專題研究計畫 成果報告

不明確資料之統計分析與軟計算方法

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-004-003-

執行期間：92年08月01日至93年07月31日

執行單位：國立政治大學應用數學學系

計畫主持人：吳柏林

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 9 月 6 日

# 不明確資料之統計分析與軟計算方法

計畫編號: NSC92-2213-E-004-003

吳柏林 國立政治大學應用數學系

2004/8/12

## Introduction

First, we present the evidence that other non-probabilistic uncertainty, such as belief functions ([20]) and possibility distributions ([4], [24]) can be expressed in terms of random sets. Note that this point of view has been adopted in much current research.

(a) *Belief Functions*: Belief functions are generalizations of degrees of belief in a Bayesian framework where beliefs are subjective probabilities. They are essentially perception-based information. Belief functions are acclimatized by Shafer as follows.

Let  $U$  be a finite set, say,  $U = \{u_1, u_2, \dots, u_k\}$ . A set-function  $F$  is called a belief function if  $F: 2^U \rightarrow [0,1]$  such that:

(i)  $F(\emptyset) = 0, F(U) = 1$

(ii)  $F$  is monotone of infinite order, i.e. for any  $n \geq 1$  and  $A_i \subseteq U, i = 1, 2, \dots, n$

$$F\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} F\left(\bigcap_{j \in I} A_j\right)$$

Remark: (ii) is simply the weakened form of Poincaré's equality for probability measures. Clearly  $F$  is not additive.

But then the set-function  $f: 2^U \rightarrow R$ ,  $f(A) = \sum_{B \subseteq A} (-1)^{|A-B|} F(B)$  is non negative and  $\sum_{A \subseteq \Omega} f(A) = 1$ ..

As such  $f(\cdot)$  is a Borel field probability mass (density) function on  $2^U$ . More specifically, there is a random element (or formally, a "random set")  $S: (U, A, P) \rightarrow 2^U$  such that  $P(S = A) = f(A)$ , and  $F(A) = p(S \subseteq A)$

For such  $S$ , the belief function  $F$  is nothing else than the counter-part of the distribution function of random variable. Indeed, by Möbius inversion on the finite poset  $(2^U, \subseteq)$ , we have

$$f(B) = \sum_{A \subseteq B} f(A) = p(S \subseteq B)$$

**Theorem :**  $T(K) = P_S(F_K, K \in \mathbf{K})$ .

Example. Let  $\varphi: R^d \rightarrow [0,1]$  be a possibility distribution which is semi-upper-continuous (s.u.c.), then  $S(\omega) = x \in R^d; \varphi(x) \geq \alpha(\omega)$ , where  $\alpha$  is a random variable, uniformly distributed on  $[0,1]$ , is a random closed set on  $R^d$ , whose capacity functional is

$$T(K) = \sup\{\varphi(x): x \in K\}.$$

Thus, a possibility measure generates the probability law of a random closed set.

Remark. The above capacity functional is of a very special kind, namely it is maxitive, i.e. for any  $A, B$  in  $K$ ,  $T(A \cup B) = \max(T(A), T(B))$ . It can be shown that any set-function which is maxitive is necessarily alternating of infinite order. Thus, we can look for models for random sets among such set-functions. Maxitive set-functions appear in many places, such

as *Kuratowski's measure of non-compactness*, all fractal dimensions, and as limits of probability measures in the sense of Large Deviation Principle sense. Here is an interesting example :

Let  $f \in L^\infty(\Omega, A, P)$  , i.e.  $\|f\|_\infty = \inf_{a \geq 0} : P(\omega : |f(\omega)| > a) < \infty \|f\|$ . Since  $P$  is a finite measure, it follows that  $p > 0$ ,  $f \in L^\infty(\Omega, A, P)$ . Consider the sequence of probability measures

$$P_n(A) = \int_{A \cap \Omega} |f|^n dP$$

Then,  $P_n(A)^{1/n} = (\|f_A\|_n) / \|f\|_n \rightarrow \|f_A\|_\infty / \|f\|_\infty$  as  $n \rightarrow \infty$ . But it can be checked that the set-function  $\tau(A) = \|f_A\|_\infty$  is maxitive. Thus, in a sense possibility measures are LDP limit of probability measures. This fact is pointed out in [21] where possibility measures are called *idempotent probabilities*. Possibility measures are limits of probabilities in the LDP sense. This fact not only puts the calculus of possibilities on a firm basis for deriving deduction rules, but also testifies that possibility measures are relevant for studying "rare events" which are important in decision-making.

Possibility distributions can also be viewed as *Radon-Nikodym derivatives of possibility measures* : Let the reference possibility measure  $\nu_o$  be

$$\nu_o(A) = \begin{cases} 1 & \text{if } A \neq \phi \\ 0 & \text{if } A = \phi \end{cases}$$

and consider the Choquet integral of a measurable non-negative function  $f$  with respect to a non-decreasing set-function  $\mu(c) \int_A f d\mu = \int_0^\infty [\mu(f \geq t) \cap A] dt$ , then the possibility measure  $\pi$  associated with the possibility distribution  $\varphi$  is written as  $\pi(A) = (c) \int_A \varphi d\nu_o$  so that  $\varphi = d\pi / d\nu_o$ .

This background is useful for decision analysis, for example, in designing intelligent decision-support systems to assist decision-makers to choose options for actions, we analyze a *multi-criteria decision-making problem*, in order to propose some appropriate *aggregation operators* to summary local information from profiles or observed data leading to some ranking procedure necessary for selections. This can be achieved by the aggregation operator based on Choquet integral of the profiles with respect to a possibility measure, generalizing standard criterion of using expected values in statistical decision theory. The novel aspect of this approach is that it does take into account the interactions between criteria, in the spirit of *coalition games*. We intend to pursue this research issue for applications.

#### *Statistical Analysis with coarse data.*

The analysis of perception-based information for decision-making is not an isolated area. Indeed, rooted in a realistic situation of statistics, mainly in Biostatistics, where observed data are in general of low quality (coarse data) , expressing a form of incomplete, vague data, these

coarse data such as coarse probabilities (e.g. interval valued probabilities) are treated entirely within the framework of random sets since the unobservable are nothing else than selectors of the observables which are random sets. Thus, we intend to use also results from statistics of set-valued observations for the benefit of intelligent systems designs, especially in simulations of random sets leading to assessments of possibility distributions.

The following background on *statistics with random sets* will be useful in deriving simulations for random sets which, in turn, should provide guidelines for assessments of perception-based probabilities whose values are linguistics expressed as possibility distributions.

Inference with coarse data is needed when the values of the variable of interest are not observed, but instead, they are only known to be located in some sets, such as in inference with missing data in multivariate analysis, censoring data in survival analysis, or in grouped data, in general. A simple framework is this.

Let  $X: \Omega \rightarrow U$  (finite). A coarsening of  $X$  is a non-empty random set  $S$  on  $U$  (i.e.  $S: \Omega \rightarrow 2^U$ ) such that  $P(X \in S) = 1$ .

Note that  $S$  is a coarsening of  $X$  if and only if  $\sum_A \sum_{x \in A} P(X = x, S = A) = 1$

Also, if  $S$  is a coarsening of  $X$ , then necessarily the probability law  $P_X$  has to belong to the *core* of  $S$ . Specifically, let  $f$  denote the density of  $S$ , i.e.  $f(A) = P(S=A)$ , then the distribution function  $F$  of  $S$  is  $F(A) = P(S \subseteq A) = \sum_{B \subseteq A} f(B)$

Note that by Mobius inversion,  $f(A) = \sum_{B \subseteq A} (-1)^{|A-B|} F(B)$  where  $|A|$  denotes the cardinality of

the set  $A$ . The core of  $S$ , or more precisely the *core of F*, is  $\text{core}(F) = \text{probability measures } P \text{ on } U \text{ such that } F(\cdot) \leq P(\cdot)$

Clearly, given  $X$ , there are many possible coarsening  $X$ . Among coarsening of  $X$ , there are special ones which are "nice" for statistical inference about the unknown distribution  $P_X$  of  $X$ , when  $X$  cannot be directly observed. The following popular model for coarse data will be used in our research. A coarsening  $S$  is said to be a Coarsening At Random (CAR) if  $A \in 2^U \setminus \emptyset$ ,  $P(S=A|X=x)$  is constant in  $x \in A$ .

It can be shown that the statistical meaning of this property is :

$$\lambda(A \in 2^U \setminus \emptyset) \text{ and } x \in A, P(X=x/S=A) = P(X=x/ X \in A)$$

i.e. as far as the distribution of the random variable  $X$  is concerned, it suffices to know that  $X$  falls into some know set  $A$ , in the likelihood inference, say. This popular model for coarse data has been used especially in making inference about the underlying variables of interest as well as providing simulated data. We intend to use this model of coarsening for empirical analysis of random set observations in the process of assigning possibility distributions (perception-based probabilities) to linguistic probability values.

In order to carry reasoning in decision systems, it is necessary to supply sufficient information to the knowledge base. For example, in simulations models for missions, various conditional probabilities (of rules) are needed. These are in general obtained by asking experts. Their perception-based probabilities are in general linguistic. According to Zadeh ([24]), linguistic probabilities are possibility distributions. As such, fuzzy logic operations could be used to manipulate them.

Now, from our approach to represent possibility distributions as covering functions of random sets, it is plausible that fuzzy logic can be combined with random sets to provide solid theoretical results for logical deductions. For example, when computing various logical expressions of perception-based probabilities, we first represent each such probability value as a possibility distribution which is viewed as the covering function of some random set. The natural logical connectives on random sets will suggest appropriate fuzzy logic operations, such as t-norms, t-conorms, implication operators,... , see e.g. [14], to use, as well as to prove their validity within a firm foundation of real analysis and measure theory, as in deriving probability logic in Artificial Intelligence, see e.g. [20]. We intend to pursue this important issue in this project.

### **Knowledge with uncertain conditional.**

In our previous work ([20]), see also [8], [9], [10], a second-order probability approach (in a Bayesian spirit) has been applied to situations when prior probability distributions (e.g. Dirichlet distributions) are available. These works complement Bayesian networks when the conditional independence assumption of variables involved is not plausible. However, we did encounter cases where we need to reason with assertions with "high" probability, such as reasoning with rules which have exceptions.

This is a special case of linguistic perception-based probabilities. For example, from a knowledge rule base of the form " If  $A_i$  then  $B_i$  where  $A_i$  and  $B_i$  are linguistic labels, modeled as fuzzy events or possibility distributions, and their reliability are quantified by conditional probabilities, one can assess the validity of new rule "  $C$  then  $D$ ". We intend to extend our previous work to the realm of perception-based probabilities which present much more realistic information than prior probability distributions. In this context, we intend to develop a comprehensive theory of decision based on random set connection with possibility theory, and apply it to design of cognitive decision systems.

we have shown that for interval expert estimates, membership degree corresponding to the fuzzy approach can be reformulated in statistical terms. A natural question is: vice versa, can we reconstruct all the possibilities corresponding to the statistical approach from the fuzzy values  $\mu(x_i)$ ? In other words, if we know the values  $\mu(x_i)$  for  $i$ , can then uniquely reconstruct the probabilities  $p([x^-, x^+])$  of different intervals? In short, are the fuzzy and interval-valued statistical approaches truly equivalent? The general answer to this question is "No":

**Proposition 3.** *There exist two different probability distributions on the set of all intervals*

for which  $\mu_1(x_i) = \mu_2(x_i)$ .

**Proof.** Both distributions correspond to two experts ( $n=2$ ) and a 4-element set  $X = \{x_1 < x_2 < x_3 < x_4\}$ .

In the first example, the first expert considers the interval  $[x_1, x_3]$  to be possible, and the second expert considers the interval  $[x_2, x_4]$ . Here, the probabilities of the interval  $[x_1, x_3]$  and  $[x_2, x_4]$  are equal to 0.5:  $p_1([x_1, x_3]) = p_1([x_2, x_4]) = 0.5$ , while the probabilities of all other intervals are equal to 0.

In the second example, the first expert considers the interval  $[x_1, x_4]$  (i.e., all possible values of  $x$ ) to be possible, while the second expert considers the interval  $[x_2, x_3]$ . Here the probabilities of the intervals  $[x_1, x_4]$  and  $[x_2, x_3]$  are equal to 0.5:  $p_2([x_1, x_4]) = p_2([x_2, x_3])$ , while the probabilities of all other intervals are equal to 0.

These two probability distributions are different: e.g.,  $p_1([x_1, x_3]) = 0.5 \neq 0 = p_2([x_1, x_3])$ . On the other hand, for both distributions, we have the same membership function  $\mu$ :  $\mu_1(x_1) = \mu_2(x_1) = 0.5$ ,  $\mu_1(x_2) = \mu_2(x_2) = 1$ ,  $\mu_1(x_3) = \mu_2(x_3) = 1$ , and  $\mu_1(x_4) = \mu_2(x_4) = 0.5$ . The proposition is proven.

## Conclusion

We have accomplished its objectives as follows.

- (i) Theoretical results on using random sets as a corner stone in modeling various types of uncertainty and in inference with imprecise data,
- (ii) Within such a framework, reasoning mechanisms will be rigorously developed leading to a firm decision theory for perception-based information, with applications to some cognitive decision systems, such as multi-criteria decision-support systems.
- (iii) Cases studies on *patients in the department of emergence and the tracking high speed targets*, we proposed in the project will be not only important in the academic field but also in the practical applications. It will help people to take more information into consideration when they facing the important decision problems .

## Reference

- [1] J. M. Blin, "Fuzzy relations in group decision theory", *J. of Cybernetics*, 1974, Vol.4, pp. 17-22.
- [2] Berthold, M. and Hand, D.J. (Editors) (1999) *Intelligent Data Analysis*. Springer-Verlag.
- [4] J. Goutsias, R. P. S. Mahler, and H. T. Nguyen (eds.), *Random sets: Theory and Applications*, Springer-Verlag, N.Y., 1997.
- [5] Goodman, I.R., Mahler, R. and Nguyen, H.T. (1997) *Mathematics of Data Fusion*. Kluwer Academic Press.
- [6] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, Upper Saddle River, NJ, 1995.
- [7] H. T. Nguyen, "Some mathematical tools for linguistic probabilities", *Fuzzy Sets and Systems*, 1979, Vol. 2, pp. 53-65.

- [8] B. Wu and W. Yang, “application of fuzzy statistics in the sampling survey”, In: *Development and Application for the Quantity Methods of Social Science*, Academic Sinica, Taiwan, pp. 289-316.
- [9] Goutsias, J., Mahler, R. and Nguyen, H.T. (1997) *Random Sets : Theory and Applications*. IMA Volumes in Mathematics and its Applications, no 97, Springer-Verlag.
- [7] Haddawy, P. (1994) *Representing Plans under Uncertainty*. Lecture Notes in Artificial Intelligence, No 770, Springer-Verlag.
- [10] Hall, J. and Lawry, J. (2001) *Imprecise probabilities of engineering system failure for random and fuzzy set reliability analysis*. Proceedings 2nd Symposium on Imprecise Probabilities and Their Applications , Itaca, N.Y. , 195-294.
- [11] Lehmann, D. and Magidor, M. (1992) What does a conditional knowledge entail ? *Artificial Intelligence* (55), 1-60.
- [12] Lukasiewicz, T. (2002) Probabilistic default reasoning with conditional constraints. *Annals of Mathematics and Artificial Intelligence* (34), 35-88.
- [13] Nguyen, H.T. and Bouchon-Meunier, B. (2002) *Random sets and large deviations principle as a foundation for possibility measures*.
- [14] Nguyen, H.T., Prasad, N., Walker, C. and Walker, E.A. (2002) *A First Course in Fuzzy and Neural Control*. Chapman and Hall/CRC Press, to appear Nov. 2002.
- [15] Nguyen, H.T. (2000). Some mathematical structures for computational information. *Journal of Information Sciences* (128), 67-89.
- [16] Nguyen, H, Wang, T. and Wu, B. (2003). On probabilistic methods in Fuzzy Theory. *International Journal of Intelligent Systems*. (will appear)
- [17] Terano, T. et al (2001) *New Frontiers in Artificial Intelligence*. Springer-Verlag.
- [18] Wu, B. and Sun, C. (2001). Interval-valued statistics, fuzzy logic, and their use in computational semantics. *Journal of Intelligent and Fuzzy Systems*. **11**,1-7.
- [19] Wu, B. and Tseng, N. (2002). A New Approach to Fuzzy Regression Models with Application to Business Cycle Analysis. *Fuzzy Sets and System*. 130, 33-42.
- [20] Wu, B. and Hsu, Yu-Yun. (2003). The Use of Kernel Set and Sample Memberships in the Identification of Nonlinear Time Series. *Soft Computing Journal*. (will appear).
- [21] Walley, P. (1987) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.
- [22] Zadeh, L.A. (2002) Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. *J. Statistical Planning and Inference*. (105), 233-264.