

A COMPARATIVE ASSESSMENT OF DEMPSTER-SHAFER AND BAYESIAN  
BELIEF IN CIVIL ENGINEERING APPLICATIONS

by

WUBEN LUO

B.Sc. Tsinghua University, P.R. China, 1985

A THESIS SUBMITTED PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES  
Department of Civil Engineering

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 1988

© Wuben Luo, 1988

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Civil Engineering

The University of British Columbia  
Vancouver, Canada

Date June 15<sup>th</sup>, 1988

### Abstract

The Bayesian theory has long been the predominate method in dealing with uncertainties in civil engineering practice including water resources engineering. However, it imposes unnecessary restrictive requirements on inferential problems. Concerns thus arise about the effectiveness of using Bayesian theory in dealing with more general inferential problems. The recently developed Dempster-Shafer theory appears to be able to surmount the limitations of Bayesian theory. The new theory was originally proposed as a pure mathematical theory. A reasonable amount of work has been done in trying to adopt this new theory in practice, most of this work being related to inexact inference in expert systems and all of the work still remaining in the fundamental stage. The purpose of this research is first to compare the two theories and second to try to apply Dempster-Shafer theory in solving real problems in water resources engineering.

In comparing Bayesian and Dempster-Shafer theory, the equivalent situation between these two theories under a special situation is discussed first. The divergence of results from Dempster-Shafer and Bayesian approaches under more general situations where Bayesian theory is unsatisfactory is then examined. Following this, the conceptual difference between the two theories is argued. Also discussed in the first part of this research is the issue of dealing with evidence including classifying sources

of evidence and expressing them through belief functions.

In attempting to adopt Dempster-Shafer theory in engineering practice, the Dempster-Shafer decision theory, i.e. the application of Dempster-Shafer theory within the framework of conventional decision theory, is introduced. The application of this new decision theory is demonstrated through a water resources engineering design example.

## Table of Contents

Abstract .....	ii
List of Tables .....	vii
List of Figures .....	viii
Acknowledgements .....	ix
1. Introduction .....	1
2. Review of Bayesian Theory and Its Application .....	7
2.1 Conventional Bayesian Theory .....	7
2.1.1 Discrete Form .....	8
2.1.2 Continuous Form .....	10
2.2 The Application of Bayesian Theory to Water Resources Engineering .....	14
2.3 Comments on Bayesian Theory .....	20
2.4 Other Forms of Bayesian Theory .....	21
2.5 Summary .....	24
3. Introduction to Dempster-Shafer Theory .....	25
3.1 Representation of Evidence vs. Belief Function ..	26
3.2 Dempster's Rule of Combination .....	31
3.3 Special Classes of Belief Functions .....	36
3.3.1 Vacuous Belief Function .....	36
3.3.2 Simple Belief Function .....	38
3.3.3 Bayesian Belief Function .....	39
3.3.4 Consonant Belief Function .....	40
3.4 Summary .....	41
4. The Equivalence between Bayesian and Dempster-Shafer Theory .....	42
4.1 Statistical Specification Model and Bayesian Theory .....	42
4.2 Dempster-Shafer Approach to Statistical Specification Model .....	46

4.3 A Further Example .....	53
4.4 Summary .....	55
5. Divergence of Results from Dempster-Shafer and Bayesian Theories .....	57
5.1 Introduction .....	57
5.2 Discounting A Belief Function .....	60
5.3 Divergence of Results of Two Theories and Sensitivity Analysis .....	63
5.3.1 Discounting the Prior Bayesian Belief Function .....	64
5.3.2 Discounting the Consonant Belief Function ..	67
5.4 Numerical Example .....	69
5.5 Summary .....	76
6. Conceptual Difference Between the Two Theories and Representation of Evidence .....	81
6.1 Introduction .....	81
6.2 Conceptual Difference Between the Two Theories ..	82
6.2.1 Two Types of Uncertainties .....	82
6.2.2 Conceptual Difference Between the Two Theories .....	84
6.3 General Consideration of Evidence .....	90
6.4 Expressing Evidence through a Belief Function ...	93
6.5 Summary .....	99
7. Dempster-Shafer Decision Making in Water Resources Engineering .....	102
7.1 Introduction .....	102
7.2 Dempster-Shafer Decision Theory .....	104
7.3 The Application of Dempster-Shafer Decision Theory .....	108
7.3.1 Description Of The Original Problem .....	108
7.3.2 Dempster-Shafer Decision Analysis .....	111

7.4 Summary .....	116
8. Conclusions .....	118
REFERENCES .....	126

List of Tables

Table	Description	Page
7.1	Prior probabilities and utilities (after R.J. McAniff et al.)	109
7.2	Sample likelihood of $Z_k=10.5\%$ (after R.J. McAniff et al.)	109
7.3	Expected costs based on Bayesian posteriors	110
7.4	Basic probability assignment of a consonant belief function derived from inferential evidence	111
7.5	Combined basic probability assignments	112
7.6	Belief and plausibility values of singletons	113
7.7	The upper and lower expected costs	115

## List of Figures

Figure	Description	Page
2.1	The prior probability density function	11
2.2	Procedures for using Bayesian decision theory	19
3.1	An illustration of set of propositions of $\Theta$	27
5.1(a)	$m(A)$ vs. $a$ when the two belief functions are mildly conflicting and only the prior belief function is discounted	78
5.1(b)	The same as in Fig. 5.1(a), but only the consonant belief function is discounted	78
5.1(c)	The same as in Fig. 5.1(a), but the two belief function are discounted simultaneously	79
5.2(a)	$m(A)$ vs. $a$ when the two belief functions are highly conflicting and only the prior belief function is discounted	79
5.2(b)	The same as in Fig. 5.2(a), but only the consonant belief functionis discounted	80
5.2(c)	The same as in Fig. 5.2(a), but the two belief function are discounted simultaneously	80

### Acknowledgements

The author wishes to express his special gratitude to his supervisor, Dr. W.F. Caselton, for his invaluable encouragement, guidance and suggestions throughout the process of this research, and his efforts made in reviewing this thesis. The author also thanks Dr. A.D. Russell for his thorough review of this thesis and his valuable comments.

The financial support in the form of scholarship provided by Chinese government is gratefully acknowledged.

The author also wishes to thank his parents and friends for their help and encouragement which are important to the completion of this research.

## 1. INTRODUCTION

Uncertainties are always involved in engineering design problems. Dealing inappropriately with uncertainties may result in the engineering project being underdesigned or overdesigned, and this in turn may cause unnecessary extra project costs, unexpected damage and other negative impacts. There is therefore a professional responsibility to deal with uncertainties in the most responsible and insightful way possible. Engineering design under uncertainty consists of describing the critical random design variables and then performing decision analysis based on these descriptions using conventional decision theory. The first part of this process includes dealing with the uncertainties of design variables. Because the specification of design variables are very important for the whole process, significant efforts have been made over the last two decades[26] in trying to find a convincing method to deal explicitly with the uncertainties involved. The most conventional approach to this problem utilizes point estimation procedures[26]. This approach requires one to choose a probability model which best describes the random design variable and then estimate the parameters of this model using historical records or regional data. The decision analysis is then undertaken as a separate step based on the estimated probability model of the design variable. One weakness of this approach is that it is unable to subsequently incorporate any newly obtained information, such as from site investigation, in the

reestimation of the probability distribution of the design variable. Also, by considering the decision analysis as a separate step, this conventional approach fails to connect the estimation of design variables with the decision analysis in a systematic or logical way.

Bayesian decision theory has been successfully adopted in areas of civil engineering design as well as water resources engineering design. This involves using Bayes' equation in the inexact inference about the random design variable and then making decisions based on this inference using conventional decision theory. Bayesian theory makes it possible to incorporate all of the available prior information in the inexact inference of the design variable and subsequently to update the inferential results after obtaining new information. Furthermore, Bayesian decision theory provides a framework which relates the inferential results of the uncertain design variable with the final decision analysis, i.e. the subsequent updating of the inferential results for the design variable after obtaining new information is automatically reflected in the final optimal decision. A short review of Bayesian theory and its application in water resources engineering design is presented in Chapter 2.

In using Bayesian theory, the inferential problem has to be placed into the statistical specification model and both the prior probability assessments based on the prior information and the sample likelihoods based on the newly

obtained information have to be expressed in the form of conventional probability distributions. If the available sources of information are always specific enough so that they can explicitly specify the required probability distributions, Bayesian decision theory would be an ideal method for water resources engineering decision problems. Unfortunately, since decisions often have to be determined under a wide variety of circumstances, in many cases the available information will not be sufficient to specify a conventional probability distribution. Therefore a more general theory of decision is needed.

Dempster-Shafer theory, which was initially proposed by Dempster and subsequently advanced substantially by Shafer, provides an alternative to Bayesian theory. Compared with Bayesian theory, Dempster-Shafer theory was formalized from an entirely different point of view: it uses a belief function to represent a piece of evidence and uses Dempster's rule of combination in pooling all sources of information. Because of the greater flexibility of a belief function in representing evidence, one might anticipate that Dempster-Shafer theory offers at least the prospect of surmounting the limitations posed by Bayesian theory. The relevant Dempster-Shafer theory is introduced in Chapter 3.

An extensive comparision of Bayesian theory and Dempster-Shafer theory at the most fundamental level is undertaken in this research. The equivalence between the two theories is discussed in Chapter 4. In those general

situations where Dempster-Shafer theory is the appropriate scheme, Bayesian theory can only be used as an approximate approach and the results from the two theories differ. A general comparision of results from the two approaches is difficult because of the involvement of subjectivities in choosing the approximation. However, the divergence of results from Dempster-Shafer and Bayesian methods can be analyzed theoretically in a simple situation, a situation in which the two methods are initially identical and the results from Dempster-Shafer and Bayesian schemes then diverge as doubts about the sources of evidence are introduced. A discussion of the divergence of the results from the two approaches is presented in Chapter 5.

In Chapter 6, the conceptual difference between the two theories when expressing evidence is addressed. Since dealing with evidence is of very important interest in engineering practice, the issues of collecting and rearranging sources of evidence and expressing them in the forms of belief functions are also discussed in this chapter.

In recent years, a reasonable amount of work has been done by researchers in attempting to implement Dempster-Shafer theory in their own fields of interest, most of this work having been related to inexact inference in expert systems. Because many inferential problems which arise in civil engineering practice strongly resemble the inferential problems arising in expert systems,

Dempster-Shafer theory may also be considered as a candidate for a general method for inexact inference in civil engineering practice. Though considerable efforts have been made in trying to adopt the Dempster-Shafer theory in the real world, no significant applications have so far been reported, most of the work being still at the fundamental stage. Furthermore, little work has yet been done in applying this new theory in civil engineering practice, even at the fundamental level. The research reported in Chapter 7 therefore intended to make the first tentative step toward the application of the new theory into civil engineering practice specifically in connection with the representation of evidence. Quantitative decision making under uncertainty is an important by-product of inexact reasoning. The application of Dempster-Shafer theory in the decision analysis in water resources engineering, which is referred to as the Dempster-Shafer decision theory, is the last issue discussed. This, together with an example of application of Dempster-Shafer decision theory is presented in Chapter 7.

Finally the research results, general conclusions and suggestions for future research are summarized in Chapter 8.

It should be noted that this thesis is not intended to serve as an introduction to either Bayesian or Dempster-Shafer theory, its intent being to explain the common ground shared by the two theories and their points of departure. While the prospect is that of engineering application, the examples chosen are not intended to reflect

state of the art application of either theory but are for the purpose of comparision and contrast.

## 2. REVIEW OF BAYESIAN THEORY AND ITS APPLICATION

### 2.1 CONVENTIONAL BAYESIAN THEORY

The original Bayesian theory was aimed at dealing with the experimental problem. A typical problem may be described as follows

Suppose an experiment is designed so that its outcome may be represented by one of a set of elements  $X=\{x_i, i=1\dots m\}$ , and the outcome will be associated with one of these elements with certain probability. The probability distribution which governs the experimental outcomes  $X$  is called a chance density function and is denoted by  $\theta_j=q_{\theta_j}(x)$ . Because of the uncertainties associated with the experiment, the true chance density function is unknown. Suppose there is a set of chance density functions  $\Theta=\{q_{\theta_j}(x) j=1,\dots,n\}$ , one and only one of them being correct. Let the prior probability judgements on this set  $\Theta$  be  $p'(\theta_j)$ . The problem of updating  $p'(\theta_j)$  after observing some experimental outcome then arises. This kind of probability model is called a statistical specification model[18].

In this section, a brief review of Bayesian theory dealing with the problem described above is presented. More comprehensive and detailed descriptions of Bayesian theory can be found in various textbooks dealing with the application of probability and decision theory in civil engineering practice[1][3].

### 2.1.1 DISCRETE FORM

Consider the problem described above, but assume that both sets  $X$  and  $\Theta$  are discrete. After an observation of an experimental outcome  $x_i$ , the conditional probability that  $x_i$  will occur given each element  $\theta_j$  in  $\Theta$  being the true parameter (and therefore specifying the true chance density function) can then be obtained from the set of chance density functions. This is expressed as

$$p(x_i/\theta_j) = q_{\theta_j}(x_i) \quad j=1, 2, \dots, n \quad (2.1)$$

The updated probability distribution on  $\Theta$  after the observation of  $x_i$  can then be obtained from probability theory, i.e.

$$p''(\theta_j/x_i) = \frac{p(\theta_j, x_i)}{p(x_i)}$$

The prior probability judgements on  $\Theta$  are  $p'(\theta_j)$  hence  $p(\theta_j, x_i)$  can be expressed as  $p(x_i/\theta_j)p'(\theta_j)$  and the total probability of the occurrence of  $x_i$  can be replaced by  $p(x_i) = \sum p(x_i/\theta_j)p'(\theta_j)$ , the above equation can then be rewritten as

$$p''(\theta_j/x_i) = k p(x_i/\theta_j)p'(\theta_j) \quad (2.2)$$

where  $k = \frac{1}{\sum_j p(x_i/\theta_j)p'(\theta_j)}$

The above equation is known as Bayes' equation. On the right hand side of this equation, the term  $p'(\theta_j)$  is known and the conditional probability  $p(x_i/\theta_j)$  is given by Eq. 2.1. Following the observation of  $x_i$ , the updated probability distribution can then be obtained through the above Bayesian theory.

Since the probability  $p'(\theta_j)$  is obtained prior to the observation of  $x_i$ , it is called a prior probability. This prior probability will be obtained from the prior or old information. The probability  $p''(\theta_j/x_i)$  is termed the posterior probability, since it is obtained after the observation  $x_i$ . The conditional probability  $p(x_i/\theta_j)$ , which is a function of  $\theta_j$ , is referred to as the sample likelihood function. The constant  $k$  is a normalizing factor which ensures that the calculation of  $p''(\theta_j/x_i)$  yields a true probability distribution.

Bayesian theory can be used to progressively update the probability distribution on  $\Theta$  when a series of independent experiments are performed and the experimental outcomes are observed. One only needs to consider the calculated posterior probability as the prior probability and use the sample likelihood function obtained from a newly observed experimental outcome to get the new posterior probability distribution through Eq. 2.2. An alternative approach to the same problem is that, following a series of observations  $x=\{x_1, x_2, \dots, x_k\}$ , the total sample likelihood function is first assessed and then the posterior probability is

obtained from a simple implementation of Eq. 2.2. The two approaches are mathematically equivalent. The total sample likelihood function  $L(x/\theta_j)$  can be obtained through the following equation

$$L(x/\theta_j) = \prod_i p(x_i/\theta_j) \quad (2.3)$$

The posterior probability is then given as

$$p''(\theta_j) = k L(x/\theta_j) p'(\theta_j)$$

In engineering practice, the experimental outcomes  $x \in X$  and the parameter value  $\theta \in \Theta$  may be best described as continuous variables. The continuous form of Bayesian theory is described in Sec. 2.1.2.

#### 2.1.2 CONTINUOUS FORM

Consider again the problem posed at the beginning of this section and assume that the two sets are continuous. The prior probability density function is given in Fig. 2.1. The prior probability density function is given in Fig. 2.1. The probability that a  $\theta$  value lies within the interval  $(\theta_j, \theta_j + \Delta\theta)$  is given as  $p(\theta_j \leq \theta \leq \theta_j + \Delta\theta) = f(\theta_j) \Delta\theta$ .

After an observation of an experimental outcome  $x$ , the continuous sample likelihood function  $l(x/\theta_j)$  can then be

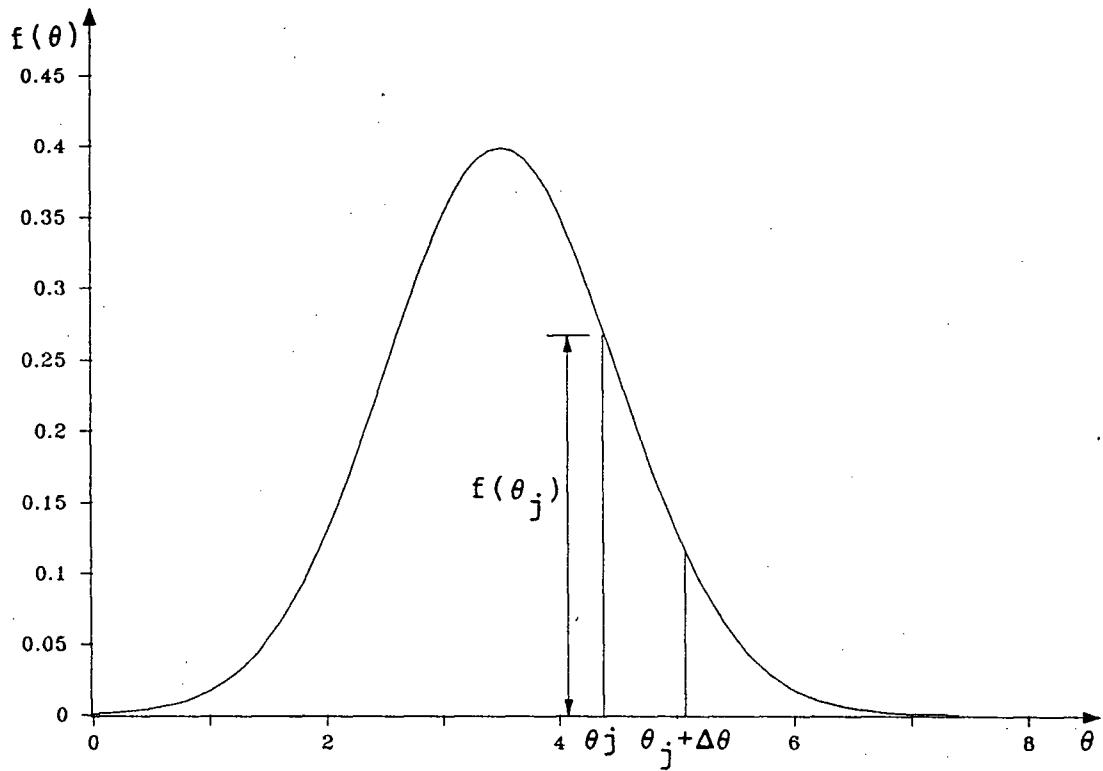


Figure 2.1 The prior probability density function

obtained from the continuous chance density function  $q_{\theta_j}(x)$  as follows

$$l(x/\theta_j) = q_{\theta_j}(x) \quad (2.4)$$

which is a function of  $\theta_j$ .

Consider the interval  $(\theta_j, \theta_j + \Delta\theta)$ , the use of Bayes' equation 2.2 yields

$$p''[(\theta_j \leq \theta \leq \theta_j + \Delta\theta)/x] = \frac{l(x/\theta_j)p'(\theta_j \leq \theta \leq \theta_j + \Delta\theta)}{\sum_k l(x/\theta_k)p(\theta_k \leq \theta \leq \theta_k + \Delta\theta)\Delta\theta}$$

using the probability density function to express probability, the above equation can be expressed as

$$f''(\theta_j/x)\Delta\theta = \frac{l(x/\theta_j)f'(\theta_j)\Delta\theta}{\sum_k l(x/\theta_k)f'(\theta_k)\Delta\theta}$$

Letting  $\Delta\theta$  tend to 0, the above equation becomes

$$f''(\theta/x) = k l(x/\theta) f'(\theta) \quad (2.5)$$

$$\text{where } k = \frac{1}{\int_{-\infty}^{+\infty} l(x/\theta) f'(\theta) d\theta}$$

The above equation is known as the continuous Bayes' equation. Comparing Eq. 2.5 with Eq. 2.2, it is seen that those two equations are of essentially the same form. Analogous to the discrete form,  $f'(\theta)$  in Eq. 2.5 is named the prior probability density function and  $f''(\theta/x)$  the posterior probability density function. The term  $l(x/\theta)$  is a continuous sample likelihood function.

As for the discrete case, if a set of observations  $x=\{x_1, x_2, \dots, x_k\}$  are obtained from a series of independent experiments, Eq. 2.5 can be used to incorporate all of those observations and obtain a final posterior probability density function. One way to deal with this problem is to use Eq. 2.5 sequentially with the previous calculated posterior probability density function as prior probability density function until all of the observations have been

considered. Alternatively, the total sample likelihood of  $x=\{x_1, \dots, x_k\}$  is calculated first and the Eq. 2.5 is then used to obtain the final posterior probability density function. The total sample likelihood function for  $x=\{x_1, \dots, x_k\}$  is given by the following equation

$$L(x/\theta) = \prod_i l(x_i/\theta) \quad (2.6)$$

which is a continuous function of  $\theta$ .

For continuous Bayesian theory, the computation can be simplified by using a particular form of prior distribution. This form of prior distribution is compatible with the sample likelihood function, and is called a conjugate of this function. By using the conjugate prior distribution, the posterior probability distribution has the same form as the prior probability distribution. It should be noted that a conjugate prior distribution is adopted primarily to simplify the mathematical computation. For a given form of likelihood function, its conjugate function can be used as prior probability distribution provided it does not conflict with any idea about the real prior distribution. Thus, if the prior information strongly supports a particular form of prior distribution, then that form of prior distribution should be used instead of the conjugate distribution.

## 2.2 THE APPLICATION OF BAYESIAN THEORY TO WATER RESOURCES ENGINEERING

Various uncertainties are always involved in water resources practice. These uncertainties can be divided into two types, natural uncertainty(NU) and informational uncertainty(IU)[22]. NU is the uncertainty inherent in the natural random process itself, such as the annual flooding on a river, and can be described by a statistical model. The IU is due to the lack of information about the natural random process. For example, one may rarely have sufficient data to find the exact statistical model and its parameters. It is seen that the IU can be divided into parameter uncertainty and model uncertainty. Because the NU is about the probabilistic phenomenon itself, it is independent of one's knowledge and information. The IU, on the other hand, can be eliminated if sufficient information is obtained.

Water resources design considers both the uncertainties about hydrological variables and the preferences of engineers towards some particular outcomes. For example, the crest elevation of flood protection dikes on a river has to be selected based on the predicted flood frequency curve and the subjective preferences of engineers towards economic decision alternatives. The water resources design then involves two procedures, the inference or prediction of a critical but random hydrological event, in this example the peak annual flood discharge, and the decision analysis based on the inference and engineer's preferences.

Because of the predominance of both short periods and imperfect data concerning the hydrological events, water resources design under uncertainty has long been recognized as a practical but challenging problem. The standard approach to this problem using classical statistical methods involves first choosing a probability model which one believes best describes the random character of the critical design event and evaluating the parameter values of that model by inferential methods using either historical records or regional data. The decisions are then made on the basis of some economic analysis as a separate step. It is seen that the standard approach fails to consider the uncertainties among the competing models. Furthermore, the uncertainties are not logically related to the final decision process. The water resources project based on such an approach may therefore be inadvertently underdesigned or overdesigned.

Bayesian decision theory, i.e. the Bayesian inferential framework together with decision theory, provides a more reasonable approach to the water resources design problem. It begins by first performing the Bayesian inference about the hydrological event. The decision analysis is then undertaken and automatically incorporates the uncertainties about the hydrological event, economic considerations and the decision maker's preferences in the final decision. The advantage of Bayesian decision theory is that it provides a methodology to pool together all of the available

information about the uncertainties of the hydrological event. This would include regional and historical data as well as subjective judgements. Bayesian decision theory also provides a theoretical way to consider both the parameter uncertainty and the model uncertainty in the final decision. Just a few of the many applications of Bayesian decision theory in water resources design problems are cited here. R.M. Shane *et al.*[22] first suggested the use of Bayesian theory to incorporate regional and historical data to reduce the uncertainties in the evaluation of a hydrological variable. G. Tschanne[24], D.R. Davis *et al.*[6] used Bayesian decision theory in designing hydrological structures in situations where the historical records are short and also examined the influence of obtaining new information on the final decision. However, their studies assumed only one model and considered only uncertainties about the parameters of that model. G.J. Vicens *et al.*[26] extensively discussed the inference about the parameter uncertainties. E.F. Wood *et al.*[27] and B. Bodo *et al.*[14] applied the more general Bayesian decision theory in water resources design in which the uncertainties about the parameter and model are considered together in the decision process.

The procedures for using Bayesian theory in water resources design is summarized as follows (and is also shown in Fig. 2.2)

1) Assume that  $I_r$  is a regional data set which acts as prior information and  $X$  is the historical record which acts as new information. According to Bayesian theory, the posterior probability density function of a set of parameters  $\Theta$  is then given by

$$f''(\Theta/I_r, X) = k l(X/\Theta) f'(\Theta/I_r) \quad (2.7)$$

where  $f'(\Theta/I_r)$  is the prior probability density function on  $\Theta$ ;  $l(X/\Theta)$  is the sample likelihood function which may be calculated by the method given in Sec 2.1.

2) After incorporating the parameter uncertainty within a probability density function for the hydrological event, the new probability density function which is called the *Bayesian distribution*[3] is given by

$$f(y) = \int_{\Theta} f(y/\Theta) f''(\Theta) d(\Theta) \quad (2.8)$$

where  $f''(\Theta)$  is the same as  $f''(\Theta/I_r, X)$  which is given by Eq. 2.7.

3) The model uncertainty is then considered in basically the same way as parameter uncertainty. Assuming  $p'(\beta_i)$  is the prior probability that model  $i$  is correct, the posterior probability is then given by

$$p''(\beta_i) = (K_i/K^*) p'(\beta_i) \quad (2.9)$$

where  $K^*$  is a normalizing factor;  $K_i$  is the marginal likelihood function of the observed data determining from the  $i^{th}$  model.

4) The final probability density function for the hydrological event which considers both types of uncertainties is known as a *composite Bayesian distribution*[3] and is expressed as

$$f(y) = \sum_i p''(\beta_i) f_i(y) \quad (2.10)$$

where  $f_i(y)$  is the Bayesian distribution of model  $i$  which is given in Eq 2.8.

5) The decision analysis is dependent upon the choice of decision strategy  $d_i$  from a set of alternatives  $D$  and the random outcome  $y_j$  out of  $Y$ . A utility function  $u=U(d_i, y_j)$  must be defined. Such a utility function expresses the decision maker's preferences when measured numerically.

6) After the utility function is obtained, the decision can then be made based on the criterion of maximizing the expected utility value. The expected utility value conditioned on decision  $d_i$  is given by

$$E(u/d_i) = \int_Y u(d_i, y_j) f(y) dy \quad (2.11)$$

where  $f(y)$  is given by Eq. 2.10. The Bayesian decision rule is then to choose  $d^*$  so that

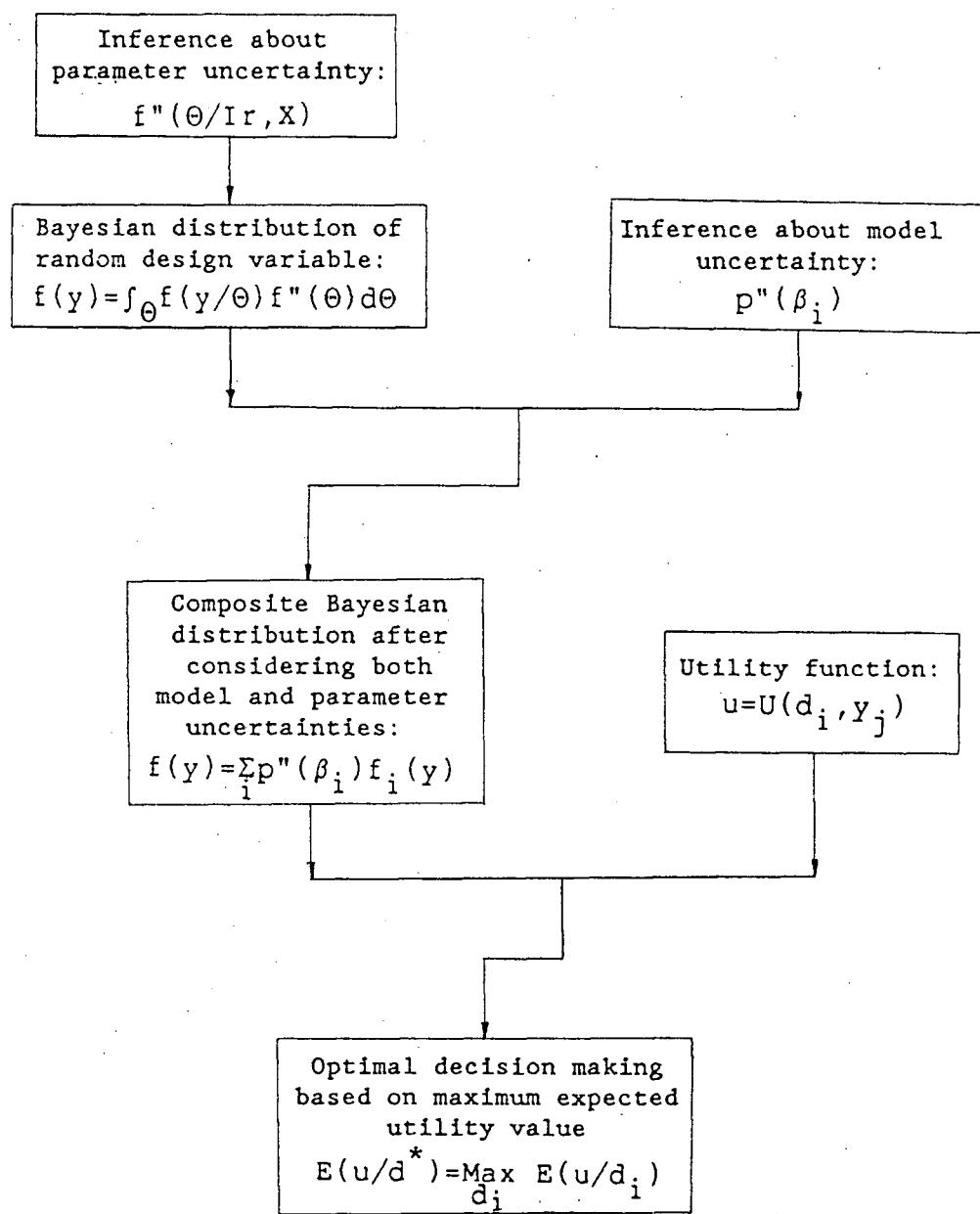


Figure 2.2 Procedures for using Bayesian decision theory

$$E(u/d^*) = \max_{d_i} E(u/d_i) \quad (2.12)$$

### 2.3 COMMENTS ON BAYESIAN THEORY

Bayesian theory has been well accepted in the literature but to a much lesser extent by practitioners. Its ability to deal with the uncertainties and to facilitate decision analysis in the face of these uncertainties is appealing. It provides a methodology which allows one to utilize all of the available information, both objective data and subjective judgements, in the inference process. Furthermore, the inference can be updated sequentially as new information is obtained. Bayesian decision theory considers the uncertainties simultaneously with the decision maker's preferences, and thus mediates the decision maker's two principle areas of concern.

However, there are some concerns about Bayesian theory. Clearly the theory is confined to the so called statistical specification model, which was originally designed to describe an experiment. In practice, not all inferential problems can be fitted into such a model. The available information has to be divided into the "old" and "new" evidence and the prior probability and sample likelihood function (i.e. conditional probability) are then based on this assignment[20]. The situation may arise where there is not sufficient information to obtain all of the needed probabilities. Even though experts may be consulted, one might still feel very uncomfortable about the probability

judgements if, for example, the number of these judgements becomes very large or divergent. Furthermore, as will be discussed in Chapter 3, Bayesian theory may not be able to express the ignorance appropriately.

Another very important concern about this theory is that it expresses both types of uncertainties with probability distributions on single elements. This is intuitively unacceptable because it uses the same approach to deal with the two entirely different types of uncertainty.

In the previous discussions about Bayesian theory, the new evidence, from which the sample likelihood function is derived, is assumed to be observed with certainty. In practice, however, the new evidence may not be obtained without some error or other uncertainty. For example, an experiment which is known to be accurate 70% of the time, gives an outcome  $x_i$ . It is still not known whether the experiment is, in this particular instance, operating accurately or inaccurately. Some of these concerns can be addressed by modifying the traditional Bayesian theory. This issue will be discussed in Sec. 2.4.

#### 2.4 OTHER FORMS OF BAYESIAN THEORY

In this section, two ways of modifying the original Bayesian theory to take into account some of the concerns mentioned above will be given. The first modification bases the Bayesian theory on the concept of likelihood ratio and

thereby reduces the task of assessing the required probabilities. The second one considers the situation in which the new evidence is incomplete evidence.

1) Bayesian theory based on likelihood ratio

Assuming  $\Theta = \{\theta_1, \dots, \theta_n\}$  is a set of hypotheses and  $x = \{x_1, \dots, x_k\}$  a set of observations, then the Bayesian theory can be rewritten as follows[20]

$$\frac{p(\theta_i/x)}{p(\theta_j/x)} = \frac{p(\theta_i)}{p(\theta_j)} \cdot \frac{p(x/\theta_i)}{p(x/\theta_j)}$$

or:

$$\frac{p(\theta_i/x)}{p(\theta_j/x)} = \frac{p(\theta_i)}{p(\theta_j)} \cdot L(x/\theta_i : \theta_j) \quad (2.14)$$

where  $L(x/\theta_i : \theta_j)$  is called the likelihood ratio favoring  $\theta_i$  over  $\theta_j$ .

Since the probabilities  $p(\theta_k/x)$ ,  $k=1, 2, \dots, n$ , must add to 1.0, they are entirely determined by their ratios  $\frac{p(\theta_i/x)}{p(\theta_j/x)}$ . Therefore, one only needs to assess the likelihood ratios of one hypothesis over another to evaluate the posterior probabilities, instead of evaluating the absolute conditional probabilities. The task of evaluating the probabilities is thus reduced.

2) Bayesian theory based on incomplete evidence

If the new evidence on which the sample likelihood is obtained is incomplete, i.e. there are some uncertainties about the evidence itself, the original Bayesian theory cannot be used directly, and some modification is necessary. The modified formulation, following R.C. Jeffrey[13] and R.O. Duda[9] is

$$p(\theta/x') = p(\theta/x)p(x/x') + p(\theta/\tilde{x})p(\tilde{x}/x') \quad (2.15)$$

where  $x'$  represents incomplete evidence. Here,  $x$  is the evidence known to be true with certainty and  $\tilde{x}$  is the evidence known to be certainly false while  $p(x/x')$  and  $p(\tilde{x}/x')$  are the probabilities of the evidence being true or false respectively after the relevant observation  $x'$ . Note that  $p(\theta/x)$  and  $p(\theta/\tilde{x})$  are simply the Bayesian posterior probabilities of  $\theta$  conditioned on  $x$  and  $\tilde{x}$  respectively.

The validity of Eq. 2.15 can be demonstrated at the extremes by the following

(1) if  $p(x/x')=1.0$ , then  $p(\tilde{x}/x')=0.0$ . Eq. 2.15 yields

$$p(\theta/x') = p(\theta/x)$$

(2) if  $p(\tilde{x}/x')=1.0$ , then  $p(x/x')=0.0$ . Eq. 2.15 yields

$$p(\theta/x') = p(\theta/\tilde{x})$$

(3) if  $p(x/x')=p(x)$ , then  $p(\tilde{x}/x')=p(\tilde{x})$ . Eq. 2.15 yields

$$p(\theta/x') = p(\theta)$$

The third conclusion means that if the new evidence  $x'$  is no better than the prior knowledge, the prior probabilities will remain unchanged.

## 2.5 SUMMARY

Even though there are some serious concerns about Bayesian theory, it should still be considered as an effective method in dealing with inference about uncertainties and with decision analysis if the inferential problem can be fitted into the so called statistical specification model and the available information (including subjective information) is sufficient to judge all of the required probabilities. For the more general situations where the inferential problem is hard to fit into the statistical specification model, or the available information is not sufficient for one to judge the required probabilities, or both, a more general theory is necessary. The Dempster-Shafer theory seems to be a possible candidate to meet these requirements. This theory will be introduced in Chapter 3.

### 3. INTRODUCTION TO Dempster-SHAFER THEORY

A brief review of Bayesian theory and its application was given in Chapter 2. At the end of that chapter, a recently developed new theory called Dempster-Shafer theory was mentioned. This theory, originally developed by Dempster[7] and subsequently broadened substantially by Shafer[18], aims at finding a more general and convincing method for the inference of uncertainties. Even though Bayesian theory can be proved to be a special case of Dempster-Shafer theory (see Chapter 4), it was formalized from an entirely non Bayesian point of view. It introduces the concept of using a belief function to represent evidence instead of the classical probability distribution used in Bayesian theory. The pooling of different sources of information in Dempster-Shafer theory is through Dempster's rule of combination which is quite different from the Bayesian method based on the Bayes' equation. In recent years, considerable interest has arisen in the use of Dempster-Shafer theory for inference in expert systems. Because inferential problems in civil engineering practice strongly resemble inferential problems in expert systems, Dempster-Shafer theory appears to be also worthy of serious consideration as a general method of dealing with uncertainties in engineering practice.

Dempster-Shafer theory provides a systematic framework for dealing with uncertainties. The essential parts of this theory are the concept of a belief function by which

evidence is represented, and Dempster's rule of combination by which different bodies of evidence are pooled. In this chapter, a brief introduction of this new theory is presented. More detailed descriptions and properties of Dempster-Shafer theory can be found in Shafer[18].

### 3.1 REPRESENTATION OF EVIDENCE VS. BELIEF FUNCTION

Suppose there is a set of possible answers  $\Theta$  to some question, one and only one of them being the correct answer. The set  $\Theta$  is then exhaustive and mutually exclusive, and it is termed the *frame of discernment* by Shafer. Any subset  $A$  in  $\Theta$  is called a *proposition*. The set of all propositions of  $\Theta$  corresponds to the set of subsets of  $\Theta$  and is denoted by  $2^\Theta$ . Among the propositions are the whole frame  $\Theta$  itself, the single elements and the empty set  $\emptyset$ . Example 3.1 demonstrates the concept of a frame of discernment.

#### *Example 3.1. Finding the type of material*

It is necessary to find out the type of material on some site where a dam will be built. Assume the material can only be one of three possible types: {sand}, {rock} and {soil}. The frame of discernment is then

$$\Theta = \{\text{sand, rock, soil}\}$$

The set of propositions  $2^\Theta$  can be expressed by a tree as

shown in Fig. 3.1. The proposition at each node is so arranged that it implies its ancestors.

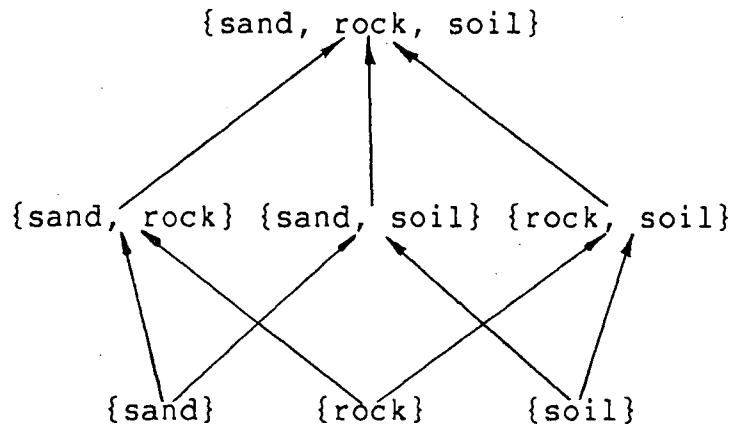


Figure 3.1 An illustration of set of propositions of  $\Theta$

The logical relationship between propositions such as conjunction, disjunction, implication and negation can be translated into the more graphical set-theoretic relationship between two subsets such as intersection, union, inclusion and complementation. Manipulation of logic within an expert system is another advantage of the Dempster-Shafer theory which enriches its capabilities. This topic is not, however, a concern of this thesis.

If a body of evidence is obtained, it assigns probability mass over the propositions. The probability number assigned to some proposition A is called a *basic probability assignment* and denoted by  $m(A)$ . It is obvious

that the basic probability assignment should satisfy the following conditions

$$m(\emptyset) = 0.$$

$$0 \leq m(A) \leq 1.0$$

$$\sum_A m(A) = 1.0$$

where  $\emptyset$  represents the null or empty set. The propositions on which the basic probability numbers are not zero are called *focal elements*. The belief value  $m(A)$  which is assigned to proposition A is also committed to any proposition which A implies. The total belief committed to A is therefore the summation of all basic probability numbers to the propositions which imply A. In other words, all belief committed to subproposition which fall within the joint proposition A are considered to contribute to the belief in A. Thus,

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (3.1)$$

$Bel(A): 2^\Theta [0,1]$  is called a belief function over  $\Theta$ . A belief function is sometimes also called a *support function*. The two concepts are different from the pure mathematical point of view. This difference is discussed in Shafer[18]. Here it is sufficient to use the term of belief function only. A

belief function should satisfy the following conditions

$$\text{Bel}(\emptyset) = 0$$

$$\text{Bel}(\Theta) = 1.0$$

$$0.0 \leq \text{Bel}(A) \leq 1.0 \text{ for } A \in \Theta$$

The value  $\text{Bel}(A)$  is only the support which is provided by the evidence on proposition A. To describe the impact of the evidence on A, one also needs to acknowledge the belief value, implied by the evidence which is not against A, i.e. not supporting the negation  $A^-$ . This is called *the degree of plausibility* on A and is denoted by  $\text{Pl}(A)$ . Accordingly

$$\text{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (3.2)$$

$$\text{or } \text{Pl}(A) = 1 - \text{Bel}(A^-)$$

The degree of plausibility is analogous to the concept of *upper probability*. The distinction between the two concepts can be found in Shafer[18]. The formulation given in Eq. 3.2 is called *the plausibility function*. The *relative plausibilities* of singletons  $\text{Rpl}(\theta)$  of belief function  $\text{Bel}(\theta)$  is expressed as

$$\text{Rpl}(\theta) = c \text{Pl}(\theta) \quad (3.3)$$

where c is a non-zero constant which is independent of  $\theta$ .

Similar to the belief function, the plausibility value is between 0 and 1.0; it will be 0.0 when the evidence is wholly against A and 1.0 when there is no evidence against A at all. The plausibility value for the empty set  $\emptyset$  is obviously zero. It should be noted that the three concepts, basic probability assignment, belief function and plausibility function, essentially represent the same thing, i.e. once one of them is determined, the other two will be fixed accordingly. For example, if the belief function is known, the basic probability assignment can be obtained from

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \text{Bel}(B)$$

This interdependence should not in any way diminish the importance of the three concepts as they considerably enrich the descriptive capabilities of the Dempster-Shafer scheme and enhance its properties for useful application in the real world. In effect, the belief value  $\text{Bel}(A)$  is the minimum support a body of evidence can provide with 100% confidence, while the plausibility value  $\text{Pl}(A)$  is the maximum possible support the evidence can conceivably provide for proposition A. The Dempster-Shafer theory therefore uses a band  $[\text{Bel}(A), \text{Pl}(A)]$  to express the full impact of a piece of evidence on A, in contrast to the Bayesian theory which considers only the singletons and is constrained to use probabilities on singletons to represent the impact of the evidence. One may intuitively conclude[28] that the probability of a proposition A, namely  $P(A)$ , should

be somewhere between the belief value and the plausibility value, i.e.

$$\text{Bel}(A) \leq P(A) \leq \text{Pl}(A) \quad (3.4)$$

though this probability value may never be known. If  $\text{Pl}(A)=\text{Bel}(A)$ , then  $\text{Bel}(A)$  is the same as probability value  $P(A)$ . The difference between  $\text{Pl}(A)$  and  $\text{Bel}(A)$  can be interpreted as the degree of uncertainty about proposition A, i.e. the ambiguity in the evidence (or whether some of the information is providing the beliefs) concerning the belief in A. The belief value assigned to the whole frame  $\Theta$  is called the ignorance of the evidence as this component of belief is unable to resolve the difference in belief between any propositions.

A belief function can be obtained, once a piece of evidence is obtained, using several techniques, some of which will be discussed in Chapter 6. If several sources of evidence are obtained, Dempster's rule of combination can be used to combine all of the belief functions to yield a resultant belief function. Dempster's rule of combination is introduced in Sec. 3.2.

### 3.2 DEMPSTER'S RULE OF COMBINATION

As was discussed in Sec. 3.1, the basic probability assignment expresses exactly the same information as the belief function. Since Dempster's rule of combination is

simpler to describe in terms of the basic probability assignment, the discussions in this section will be based mainly on this assignment.

Assume there are two distinct (or independent) bodies of evidence bearing on a frame  $\Theta$ . The corresponding basic probability assignments are denoted as  $m_1(A)$  and  $m_2(B)$ . Dempster's rule of combination then gives a new basic probability assignment  $m(C)$  which results from the combination of the two bodies of evidence. It is expressed as

$$m(C) = m_1(A) + m_2(B) = (1-k)^{-1} \sum_{A \cap B = C} m_1(A)m_2(B) \quad (3.5)$$

$$m(\emptyset) = 0.0$$

$$\text{where } k = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)$$

$k$  is the basic probability number assigned to empty proposition  $\emptyset$ . Because  $m(\emptyset)$  should have zero value, the basic probability numbers on the other non-empty propositions should therefore be inflated by a factor  $(1-k)^{-1}$ . The focal elements of the combined belief function given by  $m(C)$  are the non-empty intersections of the focal elements given by  $m_1(A)$  and  $m_2(B)$ . If several pieces of evidence are obtained, the corresponding belief functions can then be combined by using Eq. 3.5 repeatedly. The final

basic probability assignment will not depend on the order in which the basic probability assignments are combined. Similar to the case of combining two bodies of evidence, the focal elements of the combined basic probability assignment are the intersections of the focal elements coming from each basic probability assignment.

It is recognized that the value  $k$  is the basic probability number assigned to the empty proposition. This value therefore indicates the conflict between the two belief functions. In fact, Shafer suggests that the degree of conflict between two belief functions be calculated from

$$\text{con}(\text{Bel}_1, \text{Bel}_2) = \log K = -\log(1-k) \quad (3.6)$$

which is simply a transformation of  $k$ . Eq. 3.6 indicates that if the two belief functions are completely consistent, then  $k=0.0$  and the degree of conflicting  $\log K=0.0$ ; if the two belief functions flatly contradict each other, then  $k=1.0$  and the degree of conflict  $\log K=\infty$ . In this latter extreme case, the combination of the two belief functions does not yield a valid belief function. The following example illustrates the application of Dempster's rule of combination in combining belief functions

*Example 3.2 (Continuation of Example 3.1)*

Suppose there is one piece of evidence which suggests that there is 40% chance that the material is rock, 20% chance that it is sand and 40% chance that it may be any of the three possibilities. The basic probability assignment then is

$$m_1(\{\text{rock}\})=0.4$$

$$m_1(\{\text{sand}\})=0.2$$

$$m_1(\Theta)=0.4$$

and the corresponding belief function is

$$\text{Bel}_1(\{\text{rock}\})=0.4$$

$$\text{Bel}_1(\{\text{sand}\})=0.2$$

$$\text{Bel}_1(\Theta)=1.0$$

Assume that there is another piece of evidence which gives rise to

$$m_2(\{\text{sand}\})=0.3$$

$$m_2(\{\text{sand, rock}\})=0.5$$

$$m_2(\{\text{sand, soil}\})=0.1$$

$$m_2(\Theta)=0.1$$

and the corresponding belief function is

$$\text{Bel}_2(\{\text{sand}\})=0.3$$

$$\text{Bel}_2(\{\text{sand, rock}\})=0.8$$

$$\text{Bel}_2(\{\text{sand, soil}\})=0.4$$

$$\text{Bel}_2(\Theta)=1.0$$

The combination of  $m_1(A)$  with  $m_2(B)$ , using Eq. 3.5, yields:

$$m_3(\{\text{sand}\})=0.381$$

$$m_3(\{\text{rock}\})=0.286$$

$$m_3(\{\text{sand, soil}\})=0.048$$

$$m_3(\{\text{sand, rock}\})=0.238$$

$$m_3(\Theta)=0.047$$

and the corresponding belief function is

$$\text{Bel}_3(\{\text{sand}\})=0.381$$

$$\text{Bel}_3(\{\text{rock}\})=0.286$$

$$\text{Bel}_3(\{\text{sand, soil}\})=0.429$$

$$\text{Bel}_3(\{\text{sand, rock}\})=0.524$$

$$\text{Bel}_3(\Theta)=1.0$$

### 3.3 SPECIAL CLASSES OF BELIEF FUNCTIONS

In Sec. 3.1, the general concept of belief function was introduced. In this section, some special classes of belief functions will be presented. These special classes of belief functions are not only simple in expression but also, as will be discussed in detail in Chapter 6, meaningful when representing evidence in practice.

#### 3.3.1 VACUOUS BELIEF FUNCTION

The *vacuous belief function* is featured by assigning one's total belief to the whole frame  $\Theta$  and none to any singletons or joint propositions in  $\Theta$ . It has the expression

$$\text{Bel}(\Theta) = 1.0$$

$$\text{Bel}(A) = 0.0 \quad A \in \Theta$$

From this, it can be easily shown that the plausibility value for any proposition  $A$  is 1.0, i.e.  $\text{Pl}(A) = 1.0$ .

The vacuous belief function can be used to describe complete ignorance. By contrast, in Bayesian theory the ignorance is represented by assigning equal probabilities to all singletons. Interpreted from the belief function point of view, this treatment of ignorance suggests that one has specific belief on each singleton proposition. To say the least, this is difficult to reconcile with the idea of complete ignorance. As a matter of fact, it can be easily shown that the combination of a belief function with a

vacuous belief function will not change the original belief function. This is consistent with a conventional view of the influence of belief characterized as the "ignorance". By contrast, the combination of the same belief function with the Bayesian representation of ignorance will definitely change the original belief function. This point may better be illustrated by a simple example.

*Example 3.3*

Consider a belief function  $\text{Bel}_1: 2^\Theta$ ,  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . Its basic probability assignment is given as

$$m_1(\{\theta_1\}) = 0.1$$

$$m_1(\{\theta_1, \theta_2\}) = 0.3$$

$$m_1(\{\theta_1, \theta_3\}) = 0.2$$

$$m_1(\Theta) = 0.4$$

The combination of  $m_1(A)$  with a vacuous belief function on  $\Theta$

$$m_2(A) = 0 \text{ for all } A \in \Theta$$

$$m_2(\Theta) = 1.0$$

will keep  $m_1(A)$  unchanged. The combination of  $m_1(A)$  with the Bayesian representation of ignorance

$m_3(\theta_1) = m_3(\theta_2) = m_3(\theta_3) = 1/3$  will yield

$$m(\{\theta_1\})=0.348 \quad m(\{\theta_2\})=0.391 \quad m(\{\theta_3\})=0.261$$

which is obviously different from the original basic probability assignment  $m_i(A)$ . One might reasonably conclude from this that Dempster-Shafer theory provides a more appropriate way to represent ignorance than does Bayesian theory.

### 3.3.2 SIMPLE BELIEF FUNCTION

A belief function  $\text{Bel}: 2^\Theta [0,1]$  is called a *simple belief function* if there exists a non-empty subset  $A$  of  $\Theta$  such that

$$m(A)=s \quad A \in \Theta$$

$$m(B)=0.0 \quad B \in \Theta \quad B \neq A$$

$$m(\emptyset)=1-s$$

The belief function is then expressed as

$$\text{Bel}(B)=\begin{cases} 0.0 & \text{if } B \text{ does not contain } A \\ s & \text{if } B \text{ contains } A \text{ and } B \neq \emptyset \\ 1.0 & \text{if } B=\emptyset \end{cases}$$

When two simple belief functions have the same focal element  $A$ , the combination of them will give another simple belief function again with the same focal element  $A$ . If they do not

have the same focal element, this combination will not lead to a simple belief function.

### 3.3.3 BAYESIAN BELIEF FUNCTION

If the focal elements of a belief function are all singletons, the corresponding belief function is named a *Bayesian belief function*. It is so named simply because it is identical to the probability distribution on singletons used in Bayesian theory. This type of belief function is consistent with intuitive frequency interpretation of probability. In fact, if one has enough observations or experience about some unknown parameter, he can give the frequency (or chance) of each possibility being the true parameter and by doing this, he is actually giving a Bayesian belief function. A detailed discussion of the properties of this belief function and the evidence corresponding to such type of belief function can be found in Chapter 6. It is interesting to note that for each singleton, the belief value, plausibility value and basic probability number are the same for Bayesian belief function. The combination of Bayesian belief function with any belief function is still a Bayesian belief function. The same result can be obtained by combining this Bayesian belief function with the relative plausibilities of the singletons of the other general belief function. i.e.

$$\text{Bel}(\theta) = k \text{Bel}_0(\theta) \text{Rpl}(\theta) \quad (3.7)$$

This concept will also be discussed again in Chapter 4.

### 3.3.4 CONSONANT BELIEF FUNCTION

If the focal elements of a belief function can be ordered so that they are nested, this type of belief function is referred to as a *consonant belief function*. Assuming the focal elements of such a belief function are  $A_1 = \{\theta_1\}$ ,  $A_2 = \{\theta_1, \theta_2\}$  ...  $A_n = \{\theta_1, \dots, \theta_n\}$  the belief function and the plausibilities for singletons then must satisfy the following conditions[5]

$$\text{Bel}(A_1) \leq \text{Bel}(A_2) \leq \dots \leq \text{Bel}(A_n)$$

$$\text{Pl}(\theta_1) \geq \text{Pl}(\theta_2) \dots \geq \text{Pl}(\theta_n)$$

and

$$\text{Pl}(\theta_1) = 1.0$$

which in turn yields

$$\text{Bel}(\{\theta_1\}) = 1 - \text{Pl}(\theta_2)$$

$$\text{Bel}(\{\theta_1, \theta_2\}) = 1 - \text{Pl}(\theta_3)$$

....

$$\text{Bel}(\{\theta_1, \dots, \theta_{n-1}\}) = 1 - \text{Pl}(\theta_n)$$

$$\text{Bel}(\emptyset) = 1.0$$

### 3.4 SUMMARY

An entirely new theory which is known as Dempster-Shafer theory is introduced in this chapter. This theory is based on the concept of a belief function and Dempster's rule of combination. This is entirely different from the basis of Bayesian theory. With an understanding of this new theory, one might reasonably expect that it would be used as a more general approach to the problem of inference about uncertainty. This seems particularly true when the limitations posed by Bayesian theory are considered. Also of some significance after almost three decades of application in civil engineering, Bayesian theory can be shown to be a special case of Dempster-Shafer theory.

While the general belief function may be complicated, some special classes of belief functions, which have simple mathematical properties and are potentially meaningful in representing evidence, have been presented. As will be discussed in detail in Chapter 6, these classes of belief functions can play a major role in representing evidence. The consonant belief function is of special interest when investigating the relationship between Bayes' and Dempster-Shafer methods. In fact all of the belief functions in the classes, except the Bayesian type of belief function, are special cases of consonant belief function. It will be seen in Chapter 6 that there is an even stronger case for using consonant belief functions to represent evidence.

## 4. THE EQUIVALENCE BETWEEN BAYESIAN AND DEMPSTER-SHAFER

### THEORY

#### 4.1 STATISTICAL SPECIFICATION MODEL AND BAYESIAN THEORY

The concept of a statistical specification model has been mentioned at the beginning of Chapter 2. A more formal definition of this model is now given[18].

Suppose the frame of discernment  $\Theta$  consists of the possible values of a parameter  $\theta$  and suppose that an experiment is governed by one of a class  $\{q_\theta(x)\}$ ,  $\theta \in \Theta$ , of chance densities on a set of experimental outcomes  $X$ , the correct value  $\theta$  corresponding to the correct chance density function  $q_\theta(x)$  over  $X$ . The specification of sets  $\Theta$  and  $X$ , together with the class of chance density functions  $\{q_\theta(x)\}$ ,  $\theta \in \Theta$ , is defined as a *statistical specification model*.

In a statistical specification model, an observed outcome  $x$  of the experiment is a piece of evidence about the density function  $q_\theta(x)$ , and hence the value of  $\theta$ , being correct. Since the possibilities contained in  $\Theta$  can be construed as the causes and the experimental outcomes  $X$  as the effects, the evidence given by outcome  $x$  is then referred to as *inferential evidence*, i.e. the evidence of cause that is provided by an effect.

As was discussed in Chapter 2, Bayesian theory is based on the statistical specification model. If the prior information about the frame of discernment  $\Theta$  can be expressed by a prior probability distribution  $p'(\theta_j)$  on  $\Theta$ ,

then the updated probability distribution on  $\Theta$  after observing an outcome  $x_i$  can be obtained from Bayes' equation

$$p''(\theta_j/x_i) = kp'(\theta_j)p(x_i/\theta_j) \quad (2.2)$$

where  $k$  is a normalizing factor as defined previously. According to Eq. 2.1 in Chapter 2, the sample likelihood  $p(x_i/\theta_j)$  can be expressed as

$$p(x_i/\theta_j) = q_{\theta_j}(x_i) \quad (2.1)$$

Therefore, the Bayes' formula can be expressed as

$$p''(\theta_j/x_i) = kp'(\theta_j)q_{\theta_j}(x_i) \quad (4.1)$$

where  $q_{\theta_j}(x_i)$  is a function of  $\theta_j$ .

The following simple example illustrates the procedures of using Bayesian theory for the inferential problem which can be put into the statistical specification model.

*Example 4.1 (From Tang and Ang Vol. 1 pp. 332 Example 8.1)*

Piles for a foundation were initially designed for the capacity of 250 tons each. However, on some rare occasions, it is estimated that some of the piles may be subjected to load as high as 300 tons. The task is to determine the probability of the piles failing under 300 tons load.

Suppose the probability of failure ranges from 0.2 to 1.0 with interval 0.2. The frame of discernment  $\Theta$  then is  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$ . The prior probability distribution on  $\Theta$  is specified as

$$\begin{aligned} p'(\theta_1) &= 0.3 & p'(\theta_2) &= 0.4 & p'(\theta_3) &= 0.15 & p'(\theta_4) &= 0.1 \\ p'(\theta_5) &= 0.05 \end{aligned}$$

A test of the pile under the load of 300 tons is conducted and the possible results of the test form another set  $X = \{x_1 = \text{fail}, x_2 = \text{survive}\}$ . The chance density function  $q_\theta(x)$  for each  $\theta$  which governs the experimental outcomes  $X = \{x_1, x_2\}$  is given as

$q_\theta(x)$	$x_1$	$x_2$
$q_{\theta_1}(x)$	0.2	0.8
$q_{\theta_2}(x)$	0.4	0.6
$q_{\theta_3}(x)$	0.6	0.4
$q_{\theta_4}(x)$	0.8	0.2
$q_{\theta_5}(x)$	1.0	0.0

According to the definition, this inferential problem can be fitted into the statistical specification model. Since the prior probability distribution and the chance density functions are known, Eq. 4.1 can then be used to

update the probability distribution on  $\theta$  after the observation of outcome  $x_i$ . For example, if one test indicates that the pile fails to survive the 300 tons load, i.e.  $x_i=x_1$ , the updated posterior probabilities conditioned on  $x_1=\text{fail}$  then is

$$\begin{aligned} p''_1(\theta_1) &= 0.06k \quad p''_1(\theta_2) = 0.16k \quad p''_1(\theta_3) = 0.09k \quad p''_1(\theta_4) = 0.08k \\ p''_1(\theta_5) &= 0.05k \end{aligned}$$

The normalizing factor  $k$  should be  $k = \frac{1}{0.44} = 2.273$ , therefore, the posterior probabilities are:

$$\begin{aligned} p''_1(\theta_1) &= 0.136 \quad p''_1(\theta_2) = 0.364 \quad p''_1(\theta_3) = 0.205 \quad p''_1(\theta_4) = 0.182 \\ p''_1(\theta_5) &= 0.113 \end{aligned}$$

Because the test shows that the pile fails under the 300 tons load, the revised probability distribution reflects this information by shifting its supports towards the elements which indicate higher failure probabilities.

Another test is made and it indicates that the pile survives the 300 tons test load. The previously obtained posteriors conditioned on  $x_1=\text{fail}$  can then be used as priors and the new posteriors obtained the same way as before, yielding

$$p''_2(\theta_1)=0.244 \quad p''_2(\theta_2)=0.490 \quad p''_2(\theta_3)=0.184 \quad p''_2(\theta_4)=0.082$$
$$p''_2(\theta_5)=0.0$$

Since this test indicates that the pile succeeded in holding the 300 tons load, the new posteriors reflect this information by shifting to provide more support to those elements with lower failure probabilities. Note that  $p''_2(\theta_5)=0.0$ . In this situation, this probability value will remain at zero even though the piles fail to survive the 300 tons load in a set of succeeding tests. Since Bayesian theory becomes ineffective in dealing with the situation like this, care should be taken to avoid the occurrence of zero probability. The best way to do this is to substitute a close-to-zero value for a zero probability and a close-to-unit value for unity whenever these arise.

#### 4.2 DEMPSTER-SHAFER APPROACH TO STATISTICAL SPECIFICATION MODEL

Recalling the discussion in Chapter 3, applying Dempster-Shafer theory requires one to first construct belief functions from the available pieces of evidence and then combine them by using Dempster's rule of combination. In the special case in which the inferential problem can be put into the statistical specification model, the Dempster-Shafer approach involves constructing belief functions from the prior evidence and the inferential

evidence (i.e. experimental outcome) and combining them by the combination rule. For the purpose of comparison with the discussions in section 4.1, the prior evidence is assumed to provide a Bayesian belief function which is the same as the prior probability distribution in Bayesian theory. (The prior evidence, of course, does not necessarily give rise to a Bayesian belief function). For such a Bayesian belief function, the belief value, basic probability number and probability for each singleton are the same, i.e.

$$\text{Bel}_o(\theta_j) = m_o(\theta_j) = p'(\theta_j)$$

The calculation of the belief function from the inferential evidence is not so direct and needs some assumptions and analyses[18]. If a piece of inferential evidence  $x_i$  is obtained about the statistical specification model  $\{q_\theta(x)\}$ ,  $\theta \in \Theta$ , one may intuitively feel that evidence  $x_i$  favors those elements of  $\Theta$  that provide greater chance about the occurrence of  $x_i$ . Accordingly, it may be assumed that the plausibility of singleton  $Pl(\theta_j)$  is proportional to the chance that  $q_{\theta_j}(x)$  assigns  $x_i$ , i.e.

$$Pl(\theta_j) = c q_{\theta_j}(x_i) \quad \theta_j \in \Theta \quad (4.2)$$

where  $c$  is a constant which is independent of  $\theta_j$ . The above intuitive idea also leads to another reasonable assumption: that the inferential evidence should best be described by a

consonant belief function. If the elements  $\theta_j \in \Theta$  are ordered so that the element  $\theta_{j+1}$  has less tendency to produce the evidence  $x$  than  $\theta_j$ , then  $\theta_1$ , having the greatest tendency to produce  $x$ , will obtain a degree of support. The element  $\theta_2$  will not deserve any positive support as this would conflict with support for  $\theta_1$ . The set  $\{\theta_1, \theta_2\}$  however should have more support than  $\theta_1$  alone since  $\{\theta_1, \theta_2\}$  has greater tendency to produce  $x$  than  $\theta_1$  alone. Accordingly, a consonant belief function derived from inferential evidence  $x$  will have ranked beliefs, i.e.

$$\text{Bel}(\theta_1, \dots, \theta_n) \geq \text{Bel}(\theta_1, \dots, \theta_{n-1}) \geq \dots \geq \text{Bel}(\theta_1, \theta_2) \geq \text{Bel}(\theta_1)$$

The two assumptions, i.e. the belief function being consonant and the plausibility of a singleton being proportional to the chance value that  $q_\theta(x)$  assigns  $x$ , completely determines the consonant belief function[18]

$$\text{Bel}(A) = 1 - \frac{\max_{\theta \subseteq A} q_\theta(x)}{\max_{\theta \subseteq \Theta} q_\theta(x)} \quad (4.3)$$

The corresponding plausibility function  $Pl(A)$  and basic probability assignment  $m(A)$  can be calculated by the methods given in Chapter 3.

Once the belief functions based on the prior information and inferential evidence are calculated, they can be combined by Dempster's rule of combination. In the case where the prior evidence can be represented by a

Bayesian belief function, the combination of those belief functions will yield a new Bayesian belief function. In fact, as was mentioned in Sec. 3.3, (also see Ref. [18]) this combination can be simplified by combining the prior Bayesian belief function with the relative plausibilities of singletons of the consonant belief function derived from the inferential evidence, i.e.

$$\text{Bel}(\theta_j) = k \text{Belo}_0(\theta_j) \text{Rpl}(\theta_j) \quad (4.4)$$

The relative plausibilities of singletons  $\text{Rpl}(\theta_j)$  have the expression

$$\text{Rpl}(\theta_j) = c \text{Pl}(\theta_j) \quad (3.3)$$

The plausibility function  $\text{Pl}(\theta_j)$  taking the form given by Eq. 4.2, the combined Bayesian belief function  $\text{Bel}(\theta)$  therefore can be expressed in the final form

$$\text{Bel}(\theta_j) = k \text{Belo}_0(\theta_j) q_{\theta j}(x_i) \quad (4.5)$$

Comparing Eq. 4.5 with the Bayesian formula Eq. 4.1, it can be seen that the two equations are essentially the same once one realizes that the prior Bayesian belief function  $\text{Belo}_0(\theta_j)$  in Eq. 4.5 is exactly the same as the prior probability distribution  $p'(\theta_j)$  in Eq. 4.1. It can then be concluded that, for this special case of the statistical

specification model with the prior evidence expressed by a Bayesian belief function, Dempster-Shafer theory yields the same result as Bayesian theory.

Since Dempster-Shafer theory is not necessarily confined to the statistical specification model, and since not all of the inferential problems encountered in practice can be fitted into the statistical specification model, and, furthermore, not all of the prior evidence can be satisfactorily expressed by a Bayesian belief function, Dempster-Shafer theory can therefore be used to deal with more general inferential problems. In other words, Bayesian theory is only a very special case of Dempster-Shafer theory. Example 4.2 illustrates the Dempster-Shafer approach to the inferential problem given in Example 4.1. It demonstrates that the Dempster-Shafer approach will yield the same result as the Bayesian approach to this particular problem.

*Example 4.2 (Continuation of Example 4.1)*

The belief function  $\text{Bel}_0(\theta)$ , based on the prior evidence, is a Bayesian belief function which is identical to the prior probability distribution. The basic probability assignment for this prior Bayesian belief function is

$$m_0(\theta_1)=0.30 \quad m_0(\theta_2)=0.40 \quad m_0(\theta_3)=0.15 \quad m_0(\theta_4)=0.10$$

$$m_0(\theta_5)=0.05$$

A test is conducted and its result indicates  $x_1$ , i.e. the pile fails to survive the 300 tons load. This test result will give rise to a consonant belief function which can be calculated by Eq.4.3. The resulting consonant belief function and its corresponding basic probability assignment are

$$\text{Bel}_1(\theta_5) = 0.2$$

$$m_1(\theta_5) = 0.2$$

$$\text{Bel}_1(\theta_5, \theta_4) = 0.4$$

$$m_1(\theta_5, \theta_4) = 0.2$$

$$\text{Bel}_1(\theta_5, \theta_4, \theta_3) = 0.6$$

$$m_1(\theta_5, \theta_4, \theta_3) = 0.2$$

$$\text{Bel}_1(\theta_5, \theta_4, \theta_3, \theta_2) = 0.8$$

$$m_1(\theta_5, \theta_4, \theta_3, \theta_2) = 0.2$$

$$\text{Bel}_1(\Theta) = 1.0$$

$$m_1(\Theta) = 0.2$$

The combination of  $m_0(A)$  with  $m_1(A)$  through Dempster's rule of combination yields a new Bayesian belief function with its basic probability assignment given as

$$m_2(\theta_1) = 0.136 \quad m_2(\theta_2) = 0.364 \quad m_2(\theta_3) = 0.204 \quad m_2(\theta_4) = 0.182$$

$$m_2(\theta_5) = 0.114$$

which is identical to the Bayesian posterior probabilities after the observation of  $x_1$ . As was mentioned earlier in this section, the same result can be obtained by combining the prior Bayesian belief function with the relative plausibilities of singletons of the consonant belief

function Bel<sub>1</sub>(A).

If another test yields the result  $x_2=\text{survive}$ , the consonant belief function Bel<sub>3</sub>(A) and the corresponding basic probability assignment  $m_3(A)$  derived from  $x_2$  are given as

$$\text{Bel}_3(\theta_1)=0.25$$

$$m_3(\theta_1)=0.25$$

$$\text{Bel}_3(\theta_1, \theta_2)=0.5$$

$$m_3(\theta_1, \theta_2)=0.25$$

$$\text{Bel}_3(\theta_1, \theta_2, \theta_3)=0.75$$

$$m_3(\theta_1, \theta_2, \theta_3)=0.25$$

$$\text{Bel}_3(\theta_1, \theta_2, \theta_3, \theta_4)=1.0$$

$$m_3(\theta_1, \theta_2, \theta_3, \theta_4)=0.25$$

The combination of  $m_3(A)$  with  $m_2(A)$  yields:

$$m_4(\theta_1)=0.244 \quad m_4(\theta_2)=0.491 \quad m_4(\theta_3)=0.183 \quad m_4(\theta_4)=0.082$$

$$m_4(\theta_5)=0.0$$

which is the same as the Bayesian posterior probabilities after the two observations  $x_1$  and  $x_2$ . Again, this result can also be obtained by combining the Bayesian belief function given by  $m_2(A)$  with the relative plausibilities of singletons given by the consonant belief function Bel<sub>3</sub>(A).

#### 4.3 A FURTHER EXAMPLE

The following example shows once more the whole procedure for the Dempster-Shafer approach to the inferential problem, based on the statistical specification model.

*Example 4.3 (From Shafer[18] pp. 243)*

Considering a coin-tossing experiment. The result is either Heads or Tails. The chance density function which governs the experiment is unknown and is assumed to be

$$q_\theta(H) = \theta/10 \text{ and } q_\theta(T) = 1 - \theta/10$$

where  $\theta \in \Theta$  and  $\Theta = \{0, 1, 2, \dots, 10\}$ .

If the toss result is head, it will produce a consonant belief function which can be calculated from the following equation derived from Eq. 4.3

$$\text{Bel}_H(A) = 1 - \frac{\max_{\theta \in A} q_\theta(H)}{\max_{\theta \in \Theta} q_\theta(H)}$$

Note that  $q_\theta(H) = \theta/10$ , therefore  $\max_{\theta \in \Theta} q_\theta(H) = 1.0$ , and the above equation becomes

$$\text{Bel}_H(A) = 1 - \max_{\theta \in A} (\theta/10)$$

Furthermore, the focal elements  $A$  of  $\text{Bel}_H$  are:  $\{10\}$ ,  $\{9, 10\}$ ,  $\{8, 9, 10\}, \dots, \{0, 1, \dots, 10\}$ . The evaluation of  $\text{Bel}_H(A)$  according

to above equation and the corresponding basic probability assignment  $m_H(A)$  is

$$\begin{array}{ll}
 \text{Bel}_H(\{10\}) = 1/10 & m_H(\{10\}) = 1/10 \\
 \text{Bel}_H(\{9, 10\}) = 2/10 & m_H(\{9, 10\}) = 1/10 \\
 \text{Bel}_H(\{8, 9, 10\}) = 3/10 & m_H(\{8, 9, 10\}) = 1/10 \\
 \dots\dots\dots & \\
 \text{Bel}_H(\{1, 2, \dots, 10\}) = 1 & m_H(\{1, 2, \dots, 10\}) = 1/10
 \end{array}$$

Similarly, a toss resulting in tail will give another consonant belief function which is

$$\text{Bel}_T(A) = 1 - \frac{\max_{\theta \subseteq A} q_\theta(T)}{\max_{\theta \subseteq \Theta} q_\theta(T)}$$

The focal elements for  $\text{Bel}_T(A)$  are  $\{0\}$ ,  $\{0, 1\}$ , ...  $\{0, 1, \dots, 9\}$ . The calculations of  $\text{Bel}_T(A)$  and the corresponding basic probability assignment are given as

$$\begin{array}{ll}
 \text{Bel}_T(\{0\}) = 1/10 & m_T(\{0\}) = 1/10 \\
 \text{Bel}_T(\{0, 1\}) = 2/10 & m_T(\{0, 1\}) = 1/10 \\
 \text{Bel}_T(\{0, 1, 2\}) = 3/10 & m_T(\{0, 1, 2\}) = 1/10 \\
 \dots\dots\dots & \\
 \text{Bel}_T(\{0, 1, \dots, 9\}) = 1 & m_T(\{0, 1, \dots, 9\}) = 1/10
 \end{array}$$

If  $n$  tosses are recorded, and  $k$  out of  $n$  are heads, then the final belief function  $\text{Bel}(A)$  can be obtained by combining  $k$  consonant belief functions  $\text{Bel}_H(A)$  with  $n-k$  the

consonant belief functions  $\text{Bel}_T(A)$  using Dempster's combination rule. It should be noted that, though each of these individual belief functions is consonant, the combined belief function  $\text{Bel}(A)$  will be disonant unless  $k=0$  or  $k=n$ , i.e. unless all the records are either tails or heads. If prior information is available, the belief function derived from the prior information can be combined with  $\text{Bel}(A)$  obtained above to yield a new belief function. In the case where the prior belief function can be expressed as a Bayesian belief function, the result of the combination with  $\text{Bel}(A)$  is still a Bayesian belief function.

#### 4.4 SUMMARY

Bayesian theory and Dempster-Shafer theory can both be used to deal with the inferential problem. But these two theories are different in the sense that Bayesian theory can only be used to deal with the inferential problem which can be put into the statistical specification model and in which the prior evidence is specifically expressed by a Bayesian prior probability distribution. Dempster-Shafer theory, on the other hand, can be used to deal with more general inferential problems which do not necessarily fit into the statistical specification model. In the case where Bayesian theory is appropriate, Dempster-Shafer theory can still be used and yields the same results as Bayesian theory. The conclusion therefore is that Bayesian theory is a very special scheme in dealing with inferential problem which is

contained within the more general Dempster-Shafer theory.

The concept of a statistical specification model is very attractive theoretically and reveals both similarities and difference between Bayesian theory and Dempster-Shafer theory. It is also attractive in practice since a great number of problems can be fitted or interpreted into the scheme of such model.

This chapter has concentrated largely on the discussions about the equivalence between Bayesian theory and Dempster-Shafer theory. In the following two chapters, significant differences between the two theories will be discussed. The discussions in Chapter 5 will concentrate on the divergence of the results of Dempster-Shafer and Bayesian theories. In Chapter 6, the conceptual difference between the two theories and the representation of evidence through belief functions are discussed. The advantages of Dempster-Shafer theory over Bayesian theory can subsequently be viewed from these discussions.

## 5. DIVERGENCE OF RESULTS FROM DEMPSTER-SHAFER AND BAYESIAN THEORIES

### 5.1 INTRODUCTION

The condition for equivalence between Bayesian theory and Dempster-Shafer theory was discussed in Chapter 4. It was concluded that for the special case of a statistical specification model, in which the prior evidence can be expressed as a Bayesian belief function, the two theories will yield the same results. In the case where the prior evidence can only be satisfactorily expressed as a non-Bayesian belief function, Dempster-Shafer theory is clearly the more appropriate approach to the inferential problem. In this latter case, there is no equivalent Bayesian approach. Dempster-Shafer theory can also be used in the more general situation in which the inferential problem can not be fitted into the statistical specification model. In that situation also, as one might expect, there is no suitable rigorous Bayesian approach.

For the situation where Dempster-Shafer theory is the only appropriate approach to an inferential problem, Bayesian theory might still be adopted but recognized to be only an approximate approach. For example, in a case where the inferential problem can be fitted into the statistical specification model but the "true" prior evidence can clearly be expressed only as a non-Bayesian belief function, one may adopt a surrogate conventional probability

distribution, based as clearly as possible on this prior evidence, and then use the Bayesian theory for the inference. In doing so, subjective judgements are clearly involved.

For the more general case where the inferential problem does not fit into the statistical specification model, one may force the problem into the context of the Bayesian probability model. Doing this involves assigning some of the evidence as "prior" evidence and the remaining evidence as "inferential" evidence. A prior probability distribution has to be then constructed from the "prior" evidence and a sample likelihood function constructed from the "inferential" evidence. This process involves further subjective judgements. As one would expect, this approximate Bayesian approach will be quite different from a more appropriate Dempster-Shafer approach and would eventually produce different results. A further discussion about the difference between the two theories when dealing with inferential problems can be found in Chapter 6.

A general comparision between the Dempster-Shafer and Bayesian approaches in the situations where Dempster-Shafer theory is the appropriate approach and there is no true equivalent Bayesian approach, may be very difficult because of the subjective elements involved in the necessarily approximate Bayesian approach. But it is possible to access the divergence of the two approaches in a simplified situation. Assume that an inferential problem can be fitted

into the statistical specification model and the prior evidence can be expressed by a Bayesian belief function, then the Bayesian theory can be properly used. At this point, the results obtained by either Dempster-Shafer approach or Bayesian approach will agree. Suppose now that some doubts arise concerning the prior evidence and the inferential evidence and should be taken into account by modifying the prior Bayesian belief function and the consonant belief function. This modification can be made using Shafer's *discounting method*[18] which will be introduced in Sec. 5.2. The Dempster-Shafer results based on the discounted belief functions will diverge from the original Bayesian results. As will be seen in latter sections, this divergence will depend on the degree of the doubt one casts on the two belief functions and the degree of conflict between the two belief functions.

A theoretical comparision between the two approaches described above is presented in Sec. 5.3. The discussion will be confined to a two element frame of discernment to simplify the analysis. In Sec. 5.4, numerical examples are given to further illustrate the conclusions obtained in Sec. 5.3 as well as to provide a sensitivity view of the divergence of results.

## 5.2 DISCOUNTING A BELIEF FUNCTION

Assume a piece of evidence can be represented by a belief function  $\text{Bel}(A)$ , its corresponding basic probability assignment being denoted as  $m(A)$ . Assume further that one has some doubts about the truth of the evidence. Then he can modify the belief function using Shafer's discounting method.

Shafer's discounting method involves first determining a discount rate  $a$ , where  $0 \leq a \leq 1.0$ . The  $a$  value is a measure of the degree of doubt one has on the evidence: the greater the  $a$  value is, the higher the degree of doubt about the evidence. When  $a$  is equal to zero, it indicates that one has no doubt at all about the evidence. In the case where  $a$  reaches 1.0, it means that one has no confidence at all about the evidence. This implies that the evidence conveys nothing but complete ignorance of the situation. The value  $1-a$  therefore can be considered as *the degree of trust* one has on the evidence.

The discounting of a belief function based on the discount rate  $a$  is done by multiplying the belief function  $\text{Bel}(A)$  by the degree of trust  $1-a$  and by assigning the displaced belief fraction  $a$  (which is reassigned to ignorance) to the whole frame  $\Theta$ . The discounted belief function  $\text{Bel}'(A)$  can then be expressed as

$$\text{Bel}'(A) = (1-a)\text{Bel}(A) \quad A \in \Theta \quad (5.1)$$

$$\text{Bel}'(\Theta) = 1.0$$

and the corresponding basic probability assignment is

$$m'(A) = (1-a)m(A) \quad A \in \Theta \quad (5.2)$$

$$m'(\Theta) = (1-a)m(\Theta) + a$$

Shafer's discounting method can be used to discount any belief function when one feels less than completely confident about that belief function. This method can also be used in the situation where one wants to reduce the influence of a belief function. For example, assuming a situation with a given set of belief functions, where one of them is strongly conflicting with the others while the other belief functions only mildly conflict amongst themselves. Under some conditions, such as where the source of conflicting belief is low in credibility, one may wish to discount the odd belief function to reduce its effect on the final combination. Two simple examples are presented to illustrate Shafer's discounting method.

#### *Example 5.1*

Suppose there is a Bayesian belief function  $\text{Bel}(\theta): 2^\Theta$ ,  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , and its basic probability assignment is

given by

$$m_1(\theta_1)=0.5 \quad m_1(\theta_2)=0.3 \quad m_1(\theta_3)=0.2$$

If one is 80% confident about the belief function, then the belief function can be discounted by a discount rate  $a=0.2$  through Eq. 5.2. The basic probability assignment of the discounted belief function is

$$m'_1(\theta_1)=0.4 \quad m'_1(\theta_2)=0.24 \quad m'_1(\theta_3)=0.16 \quad m'_1(\Theta)=0.2$$

Note that a discounted Bayesian belief function becomes a non-Bayesian belief function for all values of  $a$  greater than 0.

### *Example 5.2*

Suppose there is a consonant belief function  $\text{Bel}(A)$  on  $\Theta=\{\theta_1, \theta_2, \theta_3\}$  and its basic probability assignment is given by

$$m_2(\theta_1)=0.5 \quad m_2(\theta_1, \theta_2)=0.3 \quad m_2(\Theta)=0.2$$

If one is 80% confident about this belief function then the belief function can be discounted by a discount rate  $a=0.2$ . The discounted belief function will be

$$m'_2(\theta_1)=0.4 \quad m'_2(\theta_1, \theta_2)=0.24 \quad m'_2(\Theta)=0.36$$

Note that discounting a consonant belief function still results in a consonant belief function.

### 5.3 DIVERGENCE OF RESULTS OF TWO THEORIES AND SENSITIVITY ANALYSIS

Assume that there is an inferential problem, the two element frame of discernment being  $\Theta = \{\theta_1, \theta_2\}$ , and the inferential problem being able to fit into the statistical specification model. Suppose the prior evidence can be expressed by a Bayesian belief function with its basic probability assignment

$$m_1(\theta_1) = p_1, \quad m_1(\theta_2) = p_2 \quad (5.3)$$

and the sample likelihoods of the inferential evidence  $x$  are given as

$$l(x/\theta_1) = l_1, \quad l(x/\theta_2) = l_2, \quad l_1 \geq l_2$$

Shafer's consonant belief function based on the inferential evidence  $x$  can then be expressed as

$$m_2(\theta_1) = 1 - l_2/l_1, \quad m_2(\theta_1, \theta_2) = l_2/l_1, \quad (5.4)$$

In this case both Bayesian theory and Dempster-Shafer theory can be used properly to perform the inference. The inferential result will be a posterior Bayesian belief

function

$$m(\theta_1) = \frac{p_1 l_1}{p_1 l_1 + p_2 l_2} \quad (5.5)$$

$$m(\theta_2) = \frac{p_2 l_2}{p_1 l_1 + p_2 l_2}$$

Assume further that one has some doubts about the prior evidence on which the Bayesian belief function is based and about the inferential evidence on which the consonant belief function is based. The two belief functions can be discounted using Shafer's discounting method as given in Sec. 5.2. Dempster-Shafer belief combination based on these discounted belief functions will produce results which diverge from the original Bayesian results. The divergence of results from the Dempster-Shafer and Bayesian approaches under progressively increasing discounting will be assessed in this section. Two separate cases will be considered. One where only the prior Bayesian belief function is discounted and the other where only the consonant belief function is discounted.

### 5.3.1 DISCOUNTING THE PRIOR BAYESIAN BELIEF FUNCTION

As the first case, consider that the Bayesian belief function  $m_1(\theta)$  is discounted by a discount rate  $a$ . The discounted prior Bayesian belief function then is

$$m'_1(\theta_1) = p_1(1-a) \quad m'_1(\theta_2) = p_2(1-a) \quad m'_1(\theta_1, \theta_2) = a$$

The combination of  $m'_1(A)$  with  $m_2(A)$  yields

$$\begin{aligned} m(\theta_1) &= \frac{p_1(1-a)+a(1-l_2/l_1)}{1-p_2(1-a)(1-l_2/l_1)} \\ m(\theta_2) &= \frac{p_2(1-a)l_2/l_1}{1-p_2(1-a)(1-l_2/l_1)} \\ m(\theta_1, \theta_2) &= \frac{l_2/l_1 a}{1-p_2(1-a)(1-l_2/l_1)} \end{aligned} \quad (5.6)$$

Note that when  $a=0.0$ , Eq. 5.6 will be identical to Eq. 5.5; when  $a>0.0$ , Eq. 5.6 will diverge from Eq. 5.5; when  $a=1.0$ , Eq. 5.6 will be identical to Eq. 5.4. Therefore the greater the  $a$  value is, the higher the degree of divergence of Eq. 5.6 from Eq. 5.5. The rate of divergence is given by the first derivatives of Eq. 5.6, i.e.

$$\frac{d[m(\theta_1)]}{da} = \frac{-p_1+k-k^2p_2}{[1-p_2(1-a)k]^2} \quad (5.7.1)$$

$$\frac{d[m(\theta_2)]}{da} = \frac{-p_2(1-k)}{[1-p_2(1-a)k]^2} \quad (5.7.2)$$

$$\frac{d[m(\theta_1, \theta_2)]}{da} = \frac{(1-k)(1-p_2k)}{[1-p_2(1-a)k]^2} \quad (5.7.3)$$

where  $k=1-l_2/l_1$ .

Note that once  $p_1$ ,  $p_2$  and  $k$  are determined, Eq. 5.7 will be functions of the discount rate  $a$ . Eq. 5.7 therefore gives the sensitivity information of the divergence of Dempster-Shafer and Bayesian results with change in the discount rate  $a$ .

An interesting situation to consider is the rate of divergence of the Dempster-Shafer and Bayesian results when  $a$  is close to 0.0. Setting  $a$  equal to zero, Eq. 5.7 will give sensitivity information about the rate of divergence with a slight increase of  $a$  from zero, i.e.

$$\frac{d[m(\theta_1)]}{da} \Big|_{a=0} = \frac{-p_1 + k - k^2 p_2}{[1-p_2 k]^2} \quad (5.8.1)$$

$$\frac{d[m(\theta_2)]}{da} \Big|_{a=0} = \frac{-p_2(1-k)}{[1-p_2 k]^2} \quad (5.8.2)$$

$$\frac{d[m(\theta_1, \theta_2)]}{da} \Big|_{a=0} = \frac{(1-k)(1-p_2 k)}{[1-p_2 k]^2} \quad (5.8.3)$$

Note that the term  $p_2 k$  in the denominator of the above equations is the degree of conflict between the original two belief functions  $m_1(A)$  and  $m_2(A)$ . Therefore, the rate of the divergence of the two theories close to  $a=0$  will depend on the degree of conflict of the two original belief functions. When the degree of conflict is not too high, i.e. the value  $p_2 k$  substantially less than 1.0, Eq. 5.8.1 will give a finite value, indicating that the Dempster-Shafer approach will diverge from the Bayesian approach gradually as  $a$  increases. When the degree of conflict is very high, i.e. the value  $p_2 k$  is very close to 1.0, Eq. 5.8.1 will give a very large value. This indicates that in a situation where the two belief functions are strongly conflicting, the Dempster-Shafer approach will diverge significantly from the Bayesian approach when only the slightest amount of

discounting is introduced. In other words, when the two belief functions are strongly conflicting, Bayesian theory will give an inferential result which is not indicative of the slightest doubt concerning the prior evidence or the experimental result. While it might be concluded that the Bayesian approach should be avoided in this situation, and Shafer's discounting method with Dempster's rule of combination should be considered as the appropriate approach, the question of assessing the correct choice of a value for  $a$  is still unresolved.

The discussions and conclusions in this section have been based on Eq. 5.8.1. Similar conclusions can also be drawn from Eq. 5.8.2 and Eq. 5.8.3.

### 5.3.2 DISCOUNTING THE CONSONANT BELIEF FUNCTION

Assume one is confident of the Bayesian belief function but has some doubts about the consonant belief function. The consonant belief function should then be discounted, according to the discount rate  $a$ , as follows

$$m'_2(\theta_1) = (1 - l_2/l_1)(1 - a) \quad m'_2(\theta_1, \theta_2) = (1 - a)l_2/l_1 + a$$

The combination of  $m_1(A)$  with  $m'_2(A)$  yields

$$m(\theta_1) = \frac{p_1}{p_1 + p_2[(1-a)l_2/l_1 + a]} \quad (5.9.1)$$

$$m(\theta_2) = 1 - \frac{p_1}{p_1 + p_2[(1-a)l_2/l_1 + a]} \quad (5.9.2)$$

The rate of the divergence of the Dempster-Shafer approach from the Bayesian approach, as a result of  $a$ , can then be obtained by the derivatives of Eq. 5.9

$$\frac{d[m(\theta_1)]}{da} = \frac{-p_1(1-l_2/l_1)}{[p_1+p_2(1-a)l_2/l_1+a]^2} \quad (5.10.1)$$

$$\frac{d[m(\theta_2)]}{da} = \frac{p_1(1-l_2/l_1)}{[p_1+p_2(1-a)l_2/l_1+a]^2} \quad (5.10.2)$$

Setting  $a=0.0$ , the sensitivity information about the rate of divergence with a slight increase of  $a$  from zero can be obtained i.e.

$$\frac{d[m(\theta_1)]}{da} \mid_{a=0} = \frac{-p_1(1-l_2/l_1)}{(p_1+p_2l_2/l_1)^2}$$

$$\frac{d[m(\theta_2)]}{da} \mid_{a=0} = \frac{p_1(1-l_2/l_1)}{(p_1+p_2l_2/l_1)^2}$$

Note that the denominator  $[p_1+p_2l_2/l_1]^2$  in the above equations can again be written as  $(1-p_2k)^2$ . The above equations can then be written as

$$\frac{d[m(\theta_1)]}{da} \mid_{a=0} = \frac{-p_1(1-l_2/l_1)}{(1-p_2k)^2} \quad (5.11.1)$$

$$\frac{d[m(\theta_2)]}{da} \mid_{a=0} = \frac{p_1(1-l_2/l_1)}{(1-p_2k)^2} \quad (5.11.2)$$

The comparison of Eq. 5.11 with Eq. 5.8 shows that they have the same denominator. This indicates that the rate of divergence of the Dempster-Shafer approach from the Bayesian

approach, in this situation, also depends on the degree of conflict between the two belief functions. Therefore the same conclusions as that in Sec. 5.3.2 can be obtained through the analysis of Eq. 5.11. This confirms the symmetrical treatments of the two sources of the belief in the Dempster-Shafer scheme.

The discounting of Bayesian belief function and consonant belief function has been considered separately in Sec. 5.3.1 and Sec. 5.3.2. In a situation where both belief functions are discounted simultaneously, the analysis becomes more difficult because of the two discount variables involved. However, this situation can be easily investigated through a numerical example.

In Sec. 5.4, numerical examples will be presented to illustrate the conclusions obtained in this section and to demonstrate the situation where the two belief functions are discounted at the same time.

#### 5.4 NUMERICAL EXAMPLE

In Sec. 5.3, it was concluded that the divergence of the results from the Dempster-Shafer and Bayesian approach, when discounting belief functions, will depend on the discount rate  $\alpha$  and the degree of conflict between the two original belief functions. Two numerical examples are presented in this section and address two different situations: where belief functions are only mildly conflicting and where belief functions are strongly

conflicting. In each of these two situations, three cases will be discussed, i.e. the discounting of the prior Bayesian belief function only, the discounting of the consonant belief function only and the discounting of both. The examples are confined to the two element frame of discernment. It is assumed that the inferential problem can be fitted into the statistical specification model consistent with the discussions in Sec. 5.3.

*Example 5.3 The belief functions are mildly conflicting*

Assume a two element frame of discernment  $\Theta = \{\theta_1, \theta_2\}$ . The prior evidence is expressed by a Bayesian belief function with its basic probability assignment

$$m_1(\theta_1) = 0.65 \quad m_1(\theta_2) = 0.35$$

The sample likelihood of a piece of inferential evidence  $x$  is known to be  $l(x/\theta_1) = 0.75$  and  $l(x/\theta_2) = 0.45$ . A consonant belief function can then be obtained based on the inferential evidence and its basic probability assignment is given as

$$m_2(\theta_1) = 0.4 \quad m_2(\theta_1, \theta_2) = 0.6$$

The degree of conflict between  $m_1(A)$  and  $m_2(A)$  is 0.14, which is substantially less than 1.0. This situation therefore represents a mildly conflicting situation.

Now consider the three different cases of discounting of the belief functions:

*Case 1.*  $m_1(A)$  is discounted by discount rate  $a$  while  $m_2(A)$  remains unchanged

The discounting of  $m_1(A)$  by a discount rate  $a$  will yield a new belief function which is given by

$$m'_1(\theta_1) = 0.65(1-a) \quad m'_1(\theta_2) = 0.35(1-a) \quad m'_{1,2}(\theta_1, \theta_2) = a$$

The combination of  $m'_{1,2}(A)$  with  $m_2(A)$  will yield

$$\begin{aligned} m(\theta_1) &= \frac{0.65 - 0.05a}{0.86 + 0.14a} \\ m(\theta_2) &= \frac{0.21(1-a)}{0.86+0.14a} \\ m(\theta_1, \theta_2) &= \frac{0.6a}{0.86+0.14a} \end{aligned} \tag{5.12}$$

*Case 2.*  $m_2(A)$  is discounted by discount rate  $a$  while  $m_1(A)$  remains unchanged

The discounting of  $m_2(A)$  by a discount rate  $a$  will yield a new belief function which is

$$m'_{2,2}(\theta_1) = 0.4(1-a) \quad m'_{2,2}(\theta_1, \theta_2) = 0.6 + 0.4a$$

Combining  $m_1(A)$  with  $m'_{2,2}(A)$  by using Dempster's rule of

combination, a new belief function will be obtained

$$\begin{aligned} m(\theta_1) &= \frac{0.65}{0.86+0.14a} \\ m(\theta_2) &= \frac{0.35(0.6+0.4a)}{0.86+0.14a} \end{aligned} \quad (5.13)$$

*Case 3.* both  $m_1(A)$  and  $m_2(A)$  are discounted simultaneously

Assume the two belief functions are discounted by the same discount rate  $a$ . The combination of the discounted belief functions  $m'_1(A)$  and  $m'_2(A)$  given in case 1 and case 2 will yield

$$\begin{aligned} m(\theta_1) &= \frac{0.65(1-a)+0.4a(1-a)}{1-0.14(1-a)^2} \\ m(\theta_2) &= \frac{0.35(1-a)+0.14(1-a)^2}{1-0.14(1-a)^2} \\ m(\theta_1, \theta_2) &= \frac{a-0.4a(1-a)}{1-0.14(1-a)^2} \end{aligned} \quad (5.14)$$

Figures 5.1(a) to 5.1(c) plots the relationship  $m(A)$  vs.  $a$  as specified by Eq. 5.12 through Eq. 5.14 respectively. From Fig. 5.1(a) to Fig. 5.1(c), it is seen that for the situation in which the degree of conflict between the two belief functions is not very high, the divergence of the results of Dempster-Shafer approach from the Bayesian approach will increase gradually as the discount rate  $a$  increases. In other words, the resultant beliefs after the discounting of the original belief

functions for all the three cases approximately linearly interpolate with  $a$  between the posteriors and the beliefs of the non-discounted belief function (the vacuous belief function in case 3). This observation is consistent with the conclusions obtained in Sec. 5.3.

*Example 5.4 The belief functions are strongly conflicting*

Assume in this example that the prior Bayesian belief function can be expressed as

$$m_1(\theta_1) = 0.01 \quad m_1(\theta_2) = 0.99$$

and the consonant belief function based on the inferential evidence can be expressed as

$$m_2(\theta_1) = 0.95 \quad m_2(\theta_1, \theta_2) = 0.05$$

The degree of conflict between  $m_1(A)$  and  $m_2(A)$  is 0.9405. This situation can be considered to be strongly conflicting. Three cases will again be considered as in example 5.3 and since the procedure is unchanged, only the resultant belief function after discounting and combining is listed for each case

*Case 1.  $m_1(A)$  is discounted by discount rate  $a$  while  $m_2(A)$  remains unchanged*

The combined belief function after discounting  $m_1(A)$  is given by

$$\begin{aligned}m(\theta_1) &= \frac{0.01+0.94a}{1-0.9405(1-a)} \\m(\theta_2) &= \frac{0.0495(1-a)}{1-0.9405(1-a)} \\m(\theta_1, \theta_2) &= \frac{0.05a}{1-0.9405(1-a)}\end{aligned}\tag{5.15}$$

*Case 2.*  $m_1(A)$  remains unchanged while  $m_2(A)$  is discounted by discount rate  $a$

The combination of discounted  $m_2(A)$  with  $m_1(A)$  will yield

$$\begin{aligned}m(\theta_1) &= \frac{0.01}{0.0595+0.9405a} \\m(\theta_2) &= \frac{0.0495+0.9405a}{0.0595+0.9405a}\end{aligned}\tag{5.16}$$

*Case 3.* Both  $m_1(A)$  and  $m_2(A)$  are discounted simultaneously

Again it will be assumed they are discounted by the same discount rate  $a$ . After combining the discounted belief functions by Dempster's rule of combination, the new belief function is

$$\begin{aligned}m(\theta_1) &= \frac{0.01(1-a)+0.95a(1-a)}{1-0.9405(1-a)^2} \\m(\theta_2) &= \frac{0.0495(1-a)^2+0.99a(1-a)}{1-0.9405(1-a)^2}\end{aligned}\tag{5.17}$$

$$m(\theta_1, \theta_2) = \frac{0.05a(1-a)+a^2}{1-0.9405(1-a)^2}$$

The relationship between  $m(A)$  and  $a$ , expressed by Eq. 5.15 through Eq. 5.17, are drawn in Figure 5.2(a) through Figure 5.2(c) respectively. It can be seen from Figure 5.2(a) to Figure 5.2(c) that, in the situation in which the two belief functions are highly conflicting, the divergence of Dempster-Shafer approach from Bayesian approach will be very sensitive to the change of  $a$  when  $a$  is very small. In other words, the Dempster-Shafer results vary with  $a$  substantially non-linearly when  $a$  is small. It is also seen that in the early stage when  $a$  is small, the Dempster-Shafer results converge more rapidly than the linear convergence, as was discussed in Example 5.3, on the non discounted belief function (or vacuous belief function in case 3). As  $a$  becomes larger, the divergence increases more gradually or linearly. Note that this result is true for all three cases.

The two entirely different situations described in Examples 5.3 and 5.4 also reveal the fact that when the two belief functions are discounted at the same time, the Dempster-Shafer results will converge on the complete ignorance (i.e. vacuous belief function) as  $a$  progressively increases. The two examples also indicate that the divergence curve is more complex in the case where the two belief functions are discounted simultaneously. This is especially true when the two belief functions are strongly conflicting. While it is difficult to assess a value for  $a$ ,

the observations from above two examples may be useful in judging the appropriate value for  $a$  in practice.

### 5.5 SUMMARY

When an inferential problem can be fitted into the statistical specification model, and the prior evidence can be expressed as a Bayesian belief function, both Dempster-Shafer theory and Bayesian theory are appropriate. The two theories will yield the same results. In the situation where one has some doubt concerning either the Bayesian belief function or the consonant belief function, the belief function may be discounted by a discount rate  $a$ . The results of a Dempster-Shafer approach based on the discounted belief functions will diverge from those obtained using a Bayesian approach. The degree of the divergence will depend on the value of the discount rate  $a$  and the degree of conflict between the prior and evidential belief functions. When these two belief functions are highly conflicting and  $a$  is close to zero, the divergence of the results from the two methods is very sensitive to the  $a$  value. This also indicates that the combination of two highly conflicting belief functions will be very sensitive to a slight error in either belief function. While Bayesian theory fails to deal with this problem at all, Shafer's discounting method together with Dempster's rule of combination is at least able to incorporate the influence of doubt about the evidence. As mentioned earlier, however, the correct value

of  $a$  may not be easy to establish in practice.

It should be noted that the discounting of a Bayesian belief function will become a non-Bayesian belief function while the discounting of a consonant belief function still remains a consonant belief function. Though the discussions in this chapter are based on the simple two element frame of discernment, it seems to be intuitively reasonable to extend the conclusions obtained from these discussions to more complex frames.

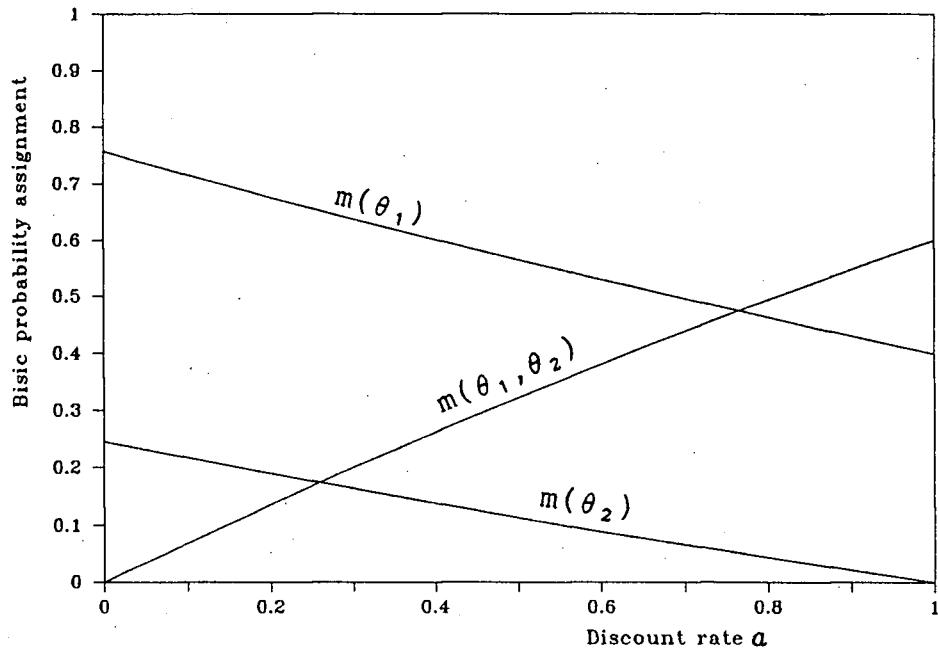


Figure 5.1(a)  $m(A)$  vs.  $\alpha$  when the two belief functions are mildly conflicting and only prior belief function is discounted

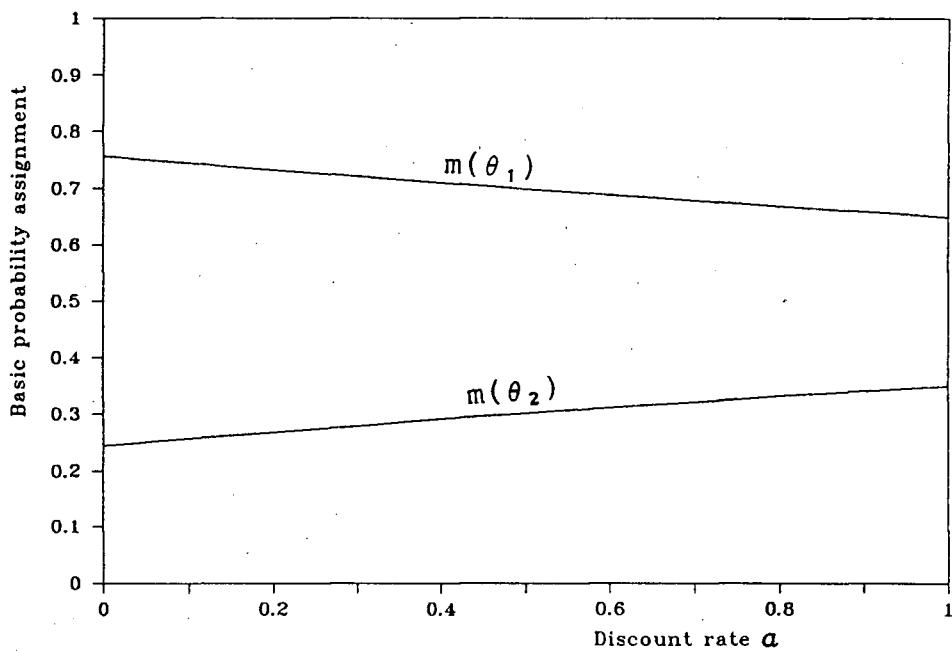


Figure 5.1(b) The same as in Fig. 5.1(a) but only the consonant belief function is discounted

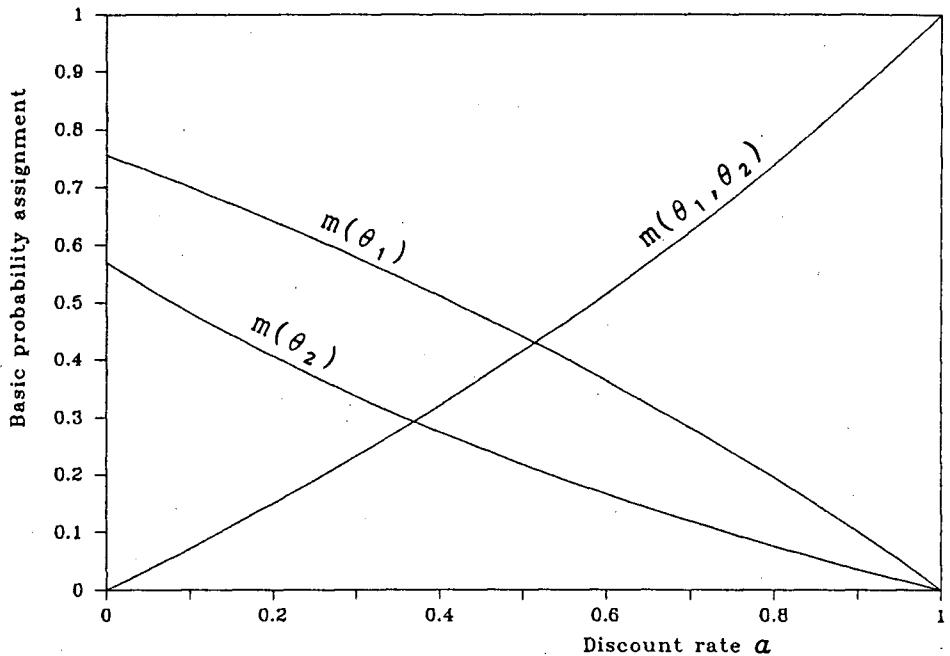


Figure 5.1(c) The same as in Fig. 5.1(a) but the two belief functions are discounted simultaneously

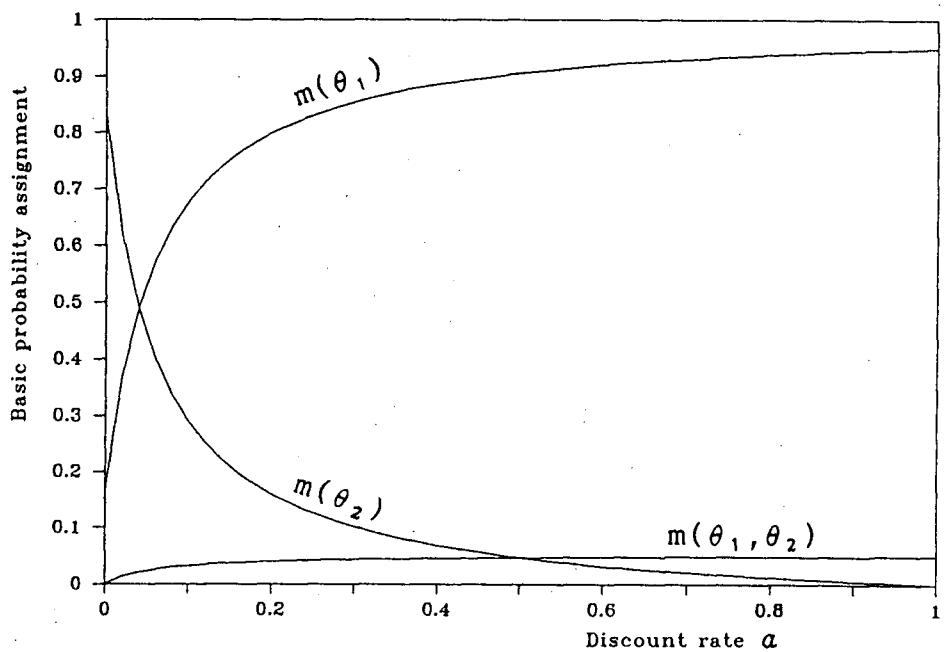


Figure 5.2(a)  $m(A)$  vs.  $\alpha$  when the two belief functions are highly conflicting and only prior belief function is discounted

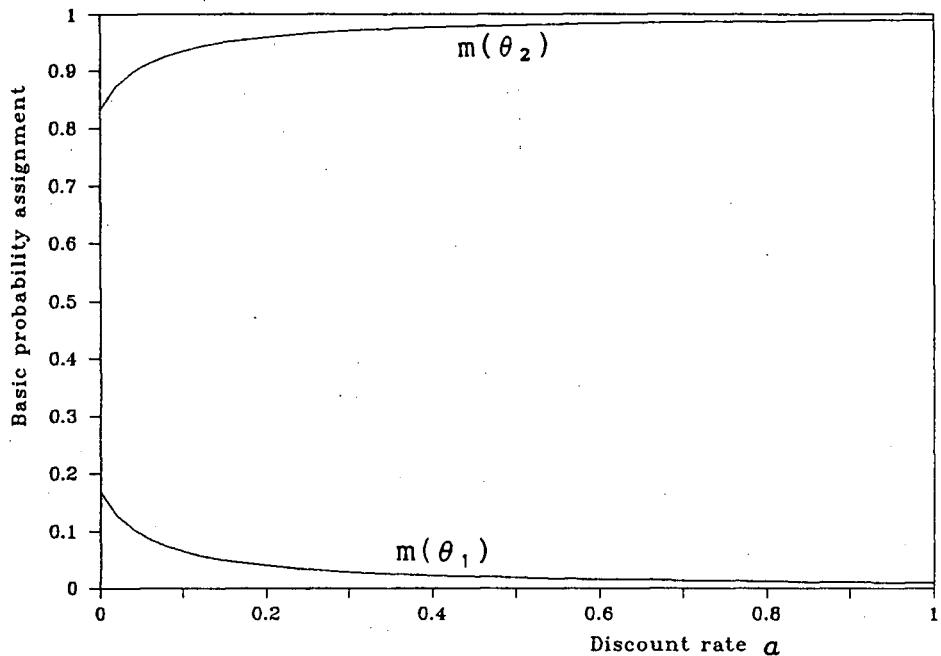


Figure 5.2(b) The same as in Fig. 5.2(a) but only the consonant belief function is discounted

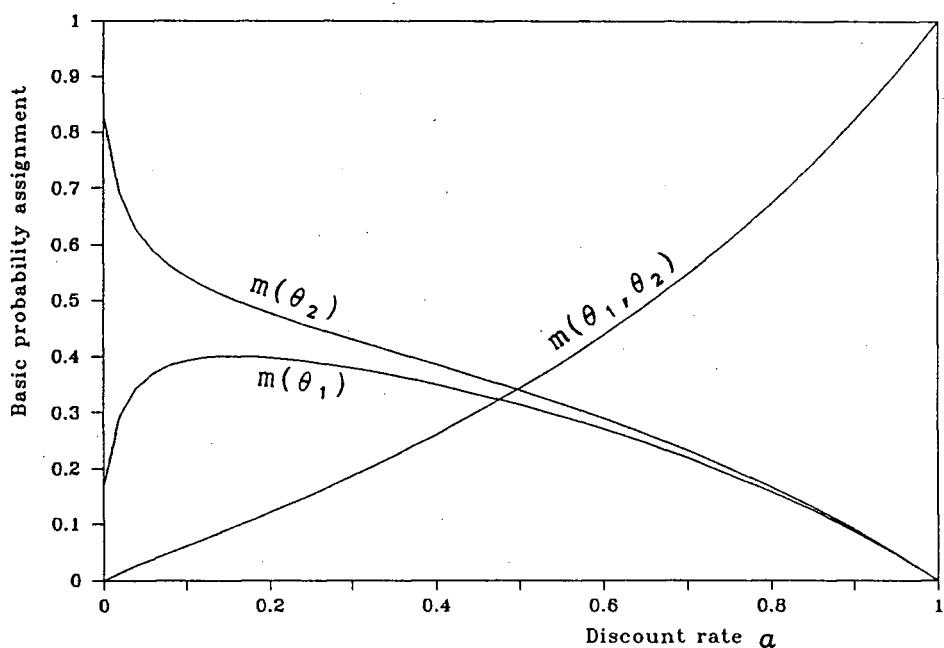


Figure 5.2(c) The same as in Fig. 5.2(a) but the two belief functions are discounted simultaneously

## 6. CONCEPTUAL DIFFERENCE BETWEEN THE TWO THEORIES AND REPRESENTATION OF EVIDENCE

### 6.1 INTRODUCTION

So far, the discussions about inference based on Bayesian and Dempster-Shafer theory have been under the assumption that the belief function is obtained perfectly once a piece of evidence is obtained. Little has been said about the evidence itself and the relationship between the types of evidence and the forms of belief functions. The question does not arise in a purely theoretical comparision of the two theories. However, in engineering practice, collecting, rearranging and correctly expressing the evidence through a mathematical belief function are very important steps in the inferential process. In practice, consideration must be given to the sources of evidence, to the evidence itself and to its expression through a belief function.

Before discussing evidence and its representation, a detailed discussion about the conceptual difference between Bayesian theory and Dempster-Shafer theory is presented in Sec. 6.2. The advantages of Dempster-Shafer theory over Bayesian theory are assessed from this viewpoint. In Sec. 6.3, general discussions about the characteristic of evidence are presented. In Sec. 6.4, the issue of representing evidence by belief functions will be discussed. Finally, in last section of this chapter, some general

conclusions are presented.

## 6.2 CONCEPTUAL DIFFERENCE BETWEEN THE TWO THEORIES

### 6.2.1 TWO TYPES OF UNCERTAINTIES

There are various kinds of uncertainties associated with a given question. Some examples may be

- a) What will be the weather tomorrow: rainy, sunny or cloudy?
- b) What will be the result of tossing a fair coin?
- c) What is the depth of the clay layer on the site of an earth dam?
- d) What is going to be the streamflow on a river for the next month?
- e) What are the correct values for some parameters in a hydrological model?

As was mentioned earlier in Chapter 2, the uncertainties associated with the above questions can be classified into two types, one is a *natural (or inherent) type of uncertainty (NU)* and the other is *informational (or statistical) type of uncertainty (IU)*[22]. NU is always encountered when one needs to predict randomly occurring, random magnitude events. IU, on the other hand, is entirely due to inadequacy of the sampling information concerning

historic events. In the above examples, uncertainties about a), b) and d) belong to the NU catagory while uncertainties about c) and e) belong to the IU catagory.

The inferential objectives arising for the two types of uncertainties are slightly different. For NU, one's best objective can only be to find the correct chance density function which governs the outcome of a future event. The possible outcomes will form a mutually exclusive and exhaustive set which may be called the *set for NU* and denoted as  $X=\{x_i, i=1, 2, \dots, m\}$ . If enough pieces of evidence are obtained, the final inferential result will be a correct chance density function (or probability mass function) on  $X$ . For the case of IU, the objective will be to try to find the truth from several possibilities, such as the specific depth of a sub-layer somewhere underground. These possibilities will also form a mutually exclusive and exhaustive set. This set may be called the *set for IU* and is denoted by  $\Theta=\{\theta_j, j=1, 2, \dots, n\}$ . For set  $\Theta$ , if a reasonable amount of evidence is available, the inferential result will tend to focus on a single element with high belief value.

Fortunately, the two types of uncertainties can both be expressed within the same framework of uncertainty. For example, in order to find the chance density function which governs the outcome  $x$  of a coin-toss, one may assume a set of possibilities  $\Theta=\{\theta_j, j=1, 2, \dots, m\}$  with each element  $\theta_j$  represents a possible chance density function  $q_{\theta_j}(x)$ . Because there is only one correct chance density function,

the set  $\Theta$  will be consistent with the IU. Thus, finding the correct chance density function  $q_\theta(x)$  which governs the experimental outcomes  $X$  is equivalent to finding the correct parameter  $\theta$  in set  $\Theta$ . Because of this relationship, the inferential problem about NU may be converted to the inferential problem about IU.

The distinction between the two types of uncertainties is not unimportant. In fact, as will be seen in Sec. 6.2.2, the distinction may help one to better understand the conceptual difference between the Bayesian theory and Dempster-Shafer theory.

#### 6.2.2 CONCEPTUAL DIFFERENCE BETWEEN THE TWO THEORIES

Recall that the original Bayesian theory is based on the statistical specification model  $\{q_\theta(x)\}$ , which is designed to find the correct chance density function for an experiment. In such a model, the set of experimental results  $X=\{x_i, i=1, 2, \dots, m\}$  has NU associated with it, while the set of possible chance density functions  $\Theta=\{\theta_j, j=1, 2, \dots, n\}$  has IU associated with it. Because the probability model provides the ability to incorporate both types of uncertainties, it is therefore possible to use Bayesian theory in both kinds of inferential situations. In fact, if the inferential problem is to find the true chance density function over set  $X$ , the Bayesian theory needs one to imagine a set  $\Theta=\{\theta_i\}$  with each single element  $\theta_i$  representing a different chance density function; if the

problem is to find the true element out of a possible set  $\Theta$ , the Bayesian theory needs one to think about a pretended experiment whose outcomes are the set  $X=\{x_i\}$ , and each single element in  $\Theta$  represents a chance density function on set  $X=\{x_i\}$ . Once the two sets  $\Theta$  and  $X$  are determined, Bayesian theory requires one to find a prior conventional probability distribution on set  $\Theta$  based on the prior evidence and a set of chance density functions on set  $X$  (from which the sample likelihood is derived). The inference based on Bayesian theory can then be undertaken after an experimental observation  $x$  has been obtained, as was discussed in Chapter 2.

While the original Bayesian theory is based on the statistical specification model, the inferential problems may not always fit into this probability model perfectly. In that case, if one chooses Bayesian theory for the inference, the inferential problem must be forced into the context of a statistical specification model. In other words, a decision has to be made as to which of the available pieces of information should be used as "prior" evidence and which as "new" evidence. As will be discussed in detail in Sec. 6.3, the collective evidence may include various forms, varying not only in their degree of accuracy but also in their relevance to the inferential problem of interest. Furthermore the choice of "new" evidence on which the sample likelihood function is to be based may often not be in the form of experimental outcomes as the original Bayesian

probability model requires.

Once the partitioning to "old" evidence and "new" evidence is determined, the "prior" probability distribution must be obtained from the "prior" evidence and sample likelihood must be obtained from the "new" evidence. Note that in both cases the evidence must be translated into the form of a conventional probability distribution. Since either the "old" evidence or the "new" evidence may not be sufficient to exactly specify these conventional probability distributions, the translation may therefore also involve subjective judgements. It is clear that Bayesian theory presumes a perfect translation is possible and thus ignores the quality of the evidence, i.e. it does not care whether or not the evidence is sufficient enough to provide a specific conventional probability distribution. Though the application of Bayesian theory in this case is quite different from its original sense, such an extension of the application of Bayesian theory should be considered appropriate only if all the probability judgements can be made with reasonable confidence. Since a great number of subjective judgements may be involved especially when the inferential problem becomes big, it may sometimes be very difficult to obtain the appropriate probability judgements[20][21].

Bayesian theory has long been the predominant method for inexact inference. Recall that, in Bayesian theory, the probability judgements on a set  $\Theta$  and on a set  $X$  both take

the form of frequency-like conventional probability distributions. A concern about the conventional probability distribution is that it expresses the probability judgements in an internally conflicting way, i.e. it supports disjoint conclusions at the same time. As was mentioned earlier, the set  $X$  has NU associated with it and the set  $\Theta$  has IU associated with it. For the set  $X$ , since each element may be the truth with some degree of possibility, it seems to be reasonable to express the probability judgements by using a frequency-like conventional probability distribution. However, for the set  $\Theta$ , since there is only one element which is the truth and the rest must be wrong, it seems to be intuitively more appropriate to express the probability judgements on set  $\Theta$  in a concordant form instead of the inherently internally conflicting conventional probability distribution. The probability judgements for both types of uncertainties have to be made in similar ways in Bayesian theory simply because there is no other choice. This may help to explain the fact that, in most of the literature dealing with uncertainty problems, the two types of uncertainties are mixed and treated identically (for example, see Ref. [25]).

As a suggested generalization of Bayesian theory[18], Dempster-Shafer theory does not need the inferential problem to be confined to the statistical specification model. As a special case, when the inferential problem can be put into this probability model, the Dempster-Shafer theory can still

be appropriately used. In fact, it has already been shown in Chapter 4 that if the Dempster-Shafer theory is used in this case, the prior belief function can be represented either as a Bayesian belief function or as a non-Bayesian belief function. If the former choice is made then it is equivalent to Bayesian theory.

In the more general case, the Dempster-Shafer theory requires one to calculate the belief functions from the available pieces of evidence. (The issue of expressing evidence by belief function will be discussed in Sec. 6.4) The inference is then undertaken by combining all of the belief functions together through the Dempster's rule of combination.

In Dempster-Shafer theory, the belief functions obtained from the available pieces of information may take various forms including the consonant belief functions and the internally conflicting conventional probability distributions (i.e. Bayesian belief functions). Since Dempster-Shafer's belief function is much more flexible than the conventional probability distributions in expressing evidence, it is therefore possible to represent various kinds of evidence in belief functions. In other words, the concept of a belief function makes it possible for one to represent the evidence more faithfully according to the quality of the evidence and its origin. Recalling that in Bayesian theory, any kind of evidence, regardless of its quality, must be represented in the form of a conventional

probability distribution, it should then be concluded that the Dempster-Shafer theory is a much more flexible and potentially convincing method.

The inference based on Dempster-Shafer theory is undertaken on the frame of discernment  $\Theta$ . Since there is only one element which is the truth and the rest must be wrong, the uncertainties associated with the frame  $\Theta$  should belong to the IU category according to the definitions given in Sec. 6.2.1. As was discussed earlier in this section, the probability judgements on the frame  $\Theta$  based on a piece of evidence should be made in a more consonant way. Therefore, it may be concluded that with Dempster-Shafer theory, the consonant belief function would be generally the more appropriate way to represent the information provided by evidence associated with  $\Theta$ .

Even though it seems attractive to use a consonant belief function to represent the evidence in Dempster-Shafer theory, some people may still prefer conventional, frequency-like probability judgements. As was mentioned earlier, the main concern about this kind of probability judgement is that it allows belief values derived from a single piece of evidence to be assigned to disjoint possibilities even though the single piece of evidence should itself be consonant. This enigma may be solved by interpreting such probability judgements in a different way: to interpret it using the odds concept[16]. For example, the depth of a sub-layer somewhere underground lies in one of

the three possibilities  $\theta_1=40$  ft,  $\theta_2=50$  ft,  $\theta_3=60$  ft. Assume that the conventional probability judgements are  $p(\theta_1)=0.3$ ,  $p(\theta_2)=0.6$ , and  $p(\theta_3)=0.1$ . The odds ratio can then be expressed as:  $p(\theta_1):p(\theta_2):p(\theta_3)=3:6:1$  which can be interpreted as: it is two times as likely that  $\theta_2$  will be the right value as it is  $\theta_1$ , and six times as likely that  $\theta_2$  will be the right value as it is  $\theta_3$ . Thought of this way, the conventional probability distribution can be considered as an effective and reasonable way to express evidence. Nevertheless, this interpretation would not satisfy those who insist on the consonant way of representing evidence.

### 6.3 GENERAL CONSIDERATION OF EVIDENCE

The foregoing discussions have concentrated on the conceptual difference between Bayesian theory and Dempster-Shafer theory. Little has been said about the evidence itself on which both the Dempster-Shafer's belief function and the Bayesian conventional probability distribution are based. In this and the following sections, general consideration of evidence and its representation through a belief function will be discussed.

Any information which is relevant to the question one is interested in resolving should be considered as evidence bearing on that question. For an inferential problem, the evidence comes from two different types of source. The first source is from practical observations such as observed historic data, experimental results etc.. The evidence thus

obtained may be referred to as *objective evidence*. The second source is from experts who provide the probability judgements, not only according to their past experience and their personal knowledge, but also somewhat according to their desire to collaborate and even to influence the final outcome. This type of evidence may be called *subjective evidence*. In practice, the evidence, either from some observations or from expert judgements, may not alone be sufficient to provide a belief function. In fact, it may be necessary that a single belief function be determined by an expert based jointly on observed data and his personal judgement. Furthermore, since the observed data will vary both in its relevance to the problem, and in its ability to provide numerical belief values, the determination of a belief function may have to be based on multiple observations instead of only one individual observation. In other words, if a set of pieces of evidence is obtained, it is not necessary that a specific belief function be obtained based on each of the observations. It is rather more likely that several pieces of evidence together support one belief function. As it will be seen later on in this section, the way of determining a belief function by personal judgements based on some practical observations plays an essential part in representing evidence by numerical belief functions.

Once a set of pieces of evidence are obtained, different theories may be chosen for the inference. The way the set of pieces of evidence will be dealt with varies

according to the chosen theory. As was mentioned in Sec.6.2, if one choose Bayesian theory, the evidence has to be grouped into two parts, one acts as "new" evidence and the other as "old" (or prior) evidence. The Bayesian prior probability judgements based on the "old" evidence and the sample likelihood function based on the "new" evidence are then made and the inference is then completed using Bayesian theory. In doing this, the "old" evidence should be independent of the "new" evidence. There are no restrictive rules to govern the division of the pieces of evidence, but the general somewhat self-serving principle[20] is that it should be easy and effective to judge the prior probabilities from the "old" evidence and to judge the sample likelihood function from the "new" evidence. It is clear that the division is rather subjective and different divisions are possible.

If the more general Dempster-Shafer theory is chosen, a set of belief functions should be obtained which correspond to the set of pieces of evidence. As was mentioned earlier in this section, each piece of evidence alone will not necessarily specify a belief function. It is therefore necessary to group the pieces of evidence so that each group is sufficiently strong and specific enough for one to build a belief function with resonable confidence. The way to group the pieces of evidence is more subjective than theoretical. However some general rules should be followed. The pieces of evidence, which are dependent should be

grouped together. The more vague and less relevant evidence should be grouped with more specific and more relevant pieces of evidence. The grouped items of evidence should be independent of each other. Once the belief functions are obtained from the grouped items of evidence, the inference should then be completed by combining the belief functions through Dempster's rule of combination. Choosing different approaches for the same inferential problem and different treatments of the collective evidence will lead to different inferential results. Instead of adopting a single approach, a comparative approach might be worth considering if it is not computationally prohibitive.

The general considerations of the evidence have been discussed in this section. In the next section, the specific problem of expressing evidence through a belief function will be discussed.

#### 6.4 EXPRESSING EVIDENCE THROUGH A BELIEF FUNCTION

Once an item of evidence is obtained, subjective judgements will be needed to establish the belief function which represents this item of evidence. According to the theory of constructive probability proposed by Shafer and others[14][19][20] the expert, who makes the subjective probability judgements, needs to compare the evidence with a *scale of canonical examples* and pick the canonical example which matches the evidence best. Because the form of the belief function for the choice of the canonical example has

already been determined, the belief function which represents the evidence is thus specified. It is seen that comparing an item of evidence with a canonical example is not only qualitative but also quantitative. It should be noted that the canonical examples are only devices by which one can effectively obtain the belief function to represent some evidence. It is not necessary that the evidence be similar to the canonical example in all respects. For a given item of evidence, a different expert may choose different types of canonical examples, or make different numerical judgements even within the same choice of canonical example. This will obviously lead to different inferential results, further reflecting the inevitable subjectivities in expressing evidence by belief function.

There are many possible forms of belief functions for a given frame of discernment, this makes it very difficult to choose the right belief function to express the evidence. This difficulty, however, can be minimized by recognizing that, in engineering practice, the engineers or other experts would prefer the evidence to be expressed in the more meaningful and simpler forms of belief functions. Indeed, it seems to be difficult to interpret the practical implications of a general belief function where belief values are assigned to (potentially at least) all joint propositions and singletons. It would be very hard to relate such a general belief function to an item of evidence. In fact, if there is an item of evidence which is so complex as

to require a general belief function, it might be preferable to decompose it into several simpler and easily understandable pieces of evidence, each of them being represented by a simpler form of belief function. A final complex general belief function can be obtained only after combining these simpler belief functions. It appears to be highly desirable therefore to have classes of belief functions which, on the one hand, have simple forms of expression and, on the other hand, are meaningful in representing types of evidence encountered frequently in engineering practice.

In Chapter 3, several classes of belief functions have already been mentioned. Further characteristics of the evidence which may be expressed by these classes of belief functions will be discussed in the following.

#### 1. The frequency form of evidence and conventional probability judgement

Recall that in Bayesian theory, both the prior probabilities and the sample likelihood function have the form of conventional probability distributions. Since a conventional probability distribution can be interpreted by the frequency concept, the evidence which supports a conventional probability distribution may be referred to as the *frequency form of evidence*.

The general characteristics of the frequency form of evidence, is that it provides odds for each element in the

frame of discernment. A canonical example for this type of evidence is given as follows

Assume that the frame of discernment for the depth of a sub-layer somewhere underground is  $\Theta = \{\theta_1=50\text{ft}, \theta_2=60\text{ft}, \theta_3=70\text{ft}\}$ . A piece of evidence may tell you that there is  $o_1$  times more likely that  $\theta_2$  is the truth than  $\theta_1$ , and there is  $o_2$  times more likely that  $\theta_3$  is the truth than  $\theta_1$ , then, the Bayesian probability assignments might be

$$P(\theta_1) = \frac{1}{1+o_1+o_2} \quad P(\theta_2) = \frac{o_1}{1+o_1+o_2} \quad P(\theta_3) = \frac{o_2}{1+o_1+o_2}$$

Another canonical example applies where an event is directly governed by some chance. By comparing the evidence with this canonical example, the probability judgement  $P(\theta_i)=p_i$  may be interpreted as evidence which supports  $\theta_i$  by probability  $p_i$ . This is equivalent to some knowledge which supports the truth in the canonical example exactly  $p_i$  of the time[20].

Any item of evidence which is compatible with the above canonical examples should be represented by a conventional probability distribution. The probability values should be determined only after choosing the canonical example which matches the evidence best.

## 2. Complete ignorance and the vacuous belief function

The evidence which corresponds to the vacuous belief function is *complete ignorance*. The characteristic of such evidence is that it confirms that the truth is in the frame of discernment but it does not say any more. While this appears to be a more satisfactory representation of ignorance than is provided by the uniform Bayesian prior, care in defining the frame of discernment is still essential. For the sub-layer example presented above, if the evidence can only tell that the true depth lies in the frame  $\Theta$ , but nothing more, then it can be represented by the vacuous belief function.

### 3. Clear but doubtful evidence and the simple belief function

The simple belief function is featured by assigning part of one's belief to a single proposition  $A \in \Theta$  and the rest to the whole frame  $\Theta$ . A canonical example for the evidence which corresponds to the simple belief function is as follows

Using the same sub-layer example as before, suppose a site investigation is conducted and it tells you that the true depth of the sub-layer is within the range 50ft to 60ft i.e. lies in a subset  $A$  of  $\Theta$ . Suppose also that this investigation is only reliable a proportion  $p$  of the time. Then the belief function corresponding to an observation of the experiment result is given by a simple belief function and can be expressed as

$$m(A)=p \quad m(\Theta)=1-p$$

Any evidence which is compatible with the above canonical example can be represented by a simple belief function. In practice, the feature of the evidence which may be represented by a simple belief function is that the measurement and its implication are clear, but the reliability of the measurement is in question[20].

#### 4. Concordant evidence and the consonant belief function

The most attractive feature of a consonant belief function is that it allows one to assign probabilities without any internal disagreement or conflict. The evidence, which can be expressed by a consonant belief function therefore may be referred to as *concordant evidence*. As was discussed in Chapter 3 (also see Ref. [5]) one determination of a consonant belief function requires only the specification of a belief value on one single element and the plausibility values for all elements. Therefore, a canonical example for the concordant evidence may take the following form

Assume the same sub-layer example as above, i.e. the depth of a sub-layer somewhere underground is within the frame of discernment  $\Theta=\{\theta_1, \theta_2, \theta_3\}$ . Suppose one can tell, based on a source of evidence, that  $\theta_1$  is the most likely true element,  $\theta_2$  is less likely and  $\theta_3$  is the least likely true element. Suppose one can also tell, based on the same

source of evidence, that the maximum possibility (i.e. the plausibility) of each element being the truth is  $P_l(\theta_i)$  for each  $\theta_i$  respectively. Since  $\theta_1$  is the most likely candidate for the truth, the maximum possibility  $P_l(\theta_1)$  (but not the probability) should be 1.0. Obviously,  $P_l(\theta_i)$  should satisfy  $P_l(\theta_1) \geq P_l(\theta_2) \geq P_l(\theta_3)$ . This source of evidence can then be represented through a consonant belief function as follows

$$Bel(\theta_1) = 1 - P_l(\theta_2)$$

$$Bel(\theta_1, \theta_2) = 1 - P_l(\theta_3)$$

$$Bel(\theta_1, \theta_2, \theta_3) = 1.0$$

Any evidence, which is compatible to the above example can be expressed through a consonant belief function.

### 6.5 SUMMARY

The conceptual difference between Bayesian theory and Dempster-Shafer theory is that Bayesian theory, which requires the inferential problem to be fitted into the statistical specification model, stresses more the satisfaction of the probability model and pays less attention to the quality of the evidence while Dempster-Shafer theory emphasizes more the quality of evidence itself. That is, Bayesian theory requires the collective pieces of evidence be expressed either as the prior probabilities or the sample likelihoods in the form of conventional probability distributions while Dempster-Shafer

belief function concept makes it possible to represent various forms of evidence according to the quality of the evidence. Since the pieces of evidence in practice do not necessarily explicitly specify conventional probability distributions, Dempster-Shafer theory appears to be a more realistic and more convincing method for engineering practice than does the Bayesian method.

The distinction between the IU and NU types of uncertainties were discussed. This discussion leads to the choice of choosing the consonant belief function as a belief structure in the Dempster-Shafer framework. While there is a strong case for adopting the consonant belief function, it still does not preclude the adoption of a Bayesian belief function (i.e. a conventional probability distribution) to represent evidence if warranted.

The evidence in practice varies both in its quality and in its relevance to the inferential problem of interest. While a general belief function is often far too complex to be used to represent evidence, the constrained and simplified classes of belief functions discussed above seem to be pertinent in representing much of the evidence which occurs in practice. Often, a single piece of evidence may not be substantial or complete enough to explicitly specify a belief function, the question of grouping of evidence then arises. Vague and incomplete pieces of evidence should, wherever possible, be combined into single item of evidence and then be expressed by the corresponding class of belief

functions. Conversely, the more complicated piece of evidence should also be decomposed into several simpler items of evidence which can then be expressed by the corresponding classes of belief functions.

## 7. DEMSTER-SHAFER DECISION MAKING IN WATER RESOURCES

### ENGINEERING

#### 7.1 INTRODUCTION

A detailed comparison between Dempster-Shafer theory and Bayesian theory has been the principal topics in Chapters 4,5, and 6. A general conclusion down from this comparison is that Bayesian theory, which confines itself to a special case of the statistical specification model, is a special case of the more general Dempster-Shafer theory. The generalizing aspect of Dempster-Shafer theory resolves various aspects of Bayesian theory which are overly restrictive in many situations commonly arising in practice.

Despite its limitations, Bayesian theory has been successfully used in dealing with uncertainties in various practical situations in civil engineering including water resources engineering. The use of Bayesian theory in water resources engineering, as was discussed in Chapter 2, has largely been to reduce the informational uncertainties associated with the evaluation of some parameters of a model and the selection of the model, and making engineering design (i.e. decision making under uncertainty) based on the inferential results. As a more general theory of inference, the Dempster-Shafer method appears to be a serious candidate as an alternative to Bayesian method in dealing with uncertainties in practice. In fact, renewed attention has been given in recent years by researchers in attempting to

apply Dempster-Shafer theory in their own fields of interest, most of the work being stimulated by the application of expert systems. A few of the applications of this new theory are given in references [2][5][10][11][12][15][25]. In spite of these efforts, no major breakthrough has been reported so far, and most of the work is still at the stage of solving elementary problems. In civil engineering, Caselton *et al.*[5] first paid attention to Dempster-Shafer theory. They have tried to utilize the new theory in construction engineering within an expert system framework. But this work still remains at the fundamental stage.

In this Chapter, the use of Dempster-Shafer theory as an alternative to Bayesian theory in decision making in water resources engineering will be discussed. The issue of decision making based on Dempster-Shafer scheme was first mentioned by Dempster[8]. Such decision making, since it is based on Dempster-Shafer theory, will be referred to here as *Dempster-Shafer decision theory*. The basic idea of Dempster-Shafer decision theory is presented in Sec. 7.2. In Sec 7.3, a real, if elementary, problem from water resources engineering practice is presented to demonstrate an application of this new theory. A summary of the discussions of Dempster-Shafer decision theory and its application is presented in Sec. 7.4.

## 7.2 DEMPSTER-SHAFER DECISION THEORY

Suppose there is a decision problem involving a random outcome which has possible values which form a frame of discernment  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The decision set is represented by  $D$ ,  $D = \{d_1, d_2, \dots, d_m\}$ . The utility function, which gives the loss value for a selected decision  $d_j$  and for a possible outcome  $\theta_i$ , is defined as  $u = U(\theta_i, d_j)$ . The problem is to minimize the expected utility. If the probability distribution  $p(\theta)$  over  $\Theta$  is known, the expectation can be calculated by

$$E(u/d_j) = \sum_i U(\theta_i, d_j) \cdot p(\theta_i) \quad (7.1)$$

Note that the probability distribution  $P(\theta_i)$  in Eq. 7.1 is assumed to be known. Bayesian decision theory involves choosing the optimal decision  $d^*$  so that:

$$E(u/d^*) = \min_{d_j} E(u/d_j)$$

Assume now that there is a belief function  $Bel(A)$  instead of a conventional probability distribution on  $\Theta$ . The true conventional probabilities of singleton elements are not known in this situation but the definitions of belief and plausibility contain the probabilities as follows[28]

$$Bel(\theta_i) \leq P_i \leq Pl(\theta_i) \quad (7.2)$$

Where  $P_i$  is the probability of  $\theta_i$  and  $\sum P_i = 1.0$ . Note that the probabilities  $\{P_i, i=1,2,\dots,n\}$  are a set of unknown variables which satisfy the constraints given in Eq. 7.2. Assuming the same utility function  $U(\theta_i, d_j)$ , applies the expected utility for a decision  $d_j$  can still be expressed along the same lines as Eq. 7.1, i.e.

$$E(u/d_j) = \sum_i U(\theta_i, d_j) \cdot P_i \quad (7.3)$$

Since the  $P_i$ ,  $i=1,2,\dots,n$ , in Eq. 7.3 are subject to known constraints, but might otherwise be considered as random variables, it is impossible to obtain a determinate expected value for any given decision. But it is possible to obtain both maximum and minimum possible expected utility values for decision  $d_j$ . The maximum is called the *upper expected value* and is denoted as  $E^*(u/d_j)$ ; the minimum is called the *lower expected value* and is denoted as  $E_*(u/d_j)$ . The calculations of upper and lower expected values involve finding the sets of  $P_i$  values which yield these extremes while meeting the above specified constraints. Therefore, this calculation can be formulated as two conventional linear programs which have the same form of objective function and constraints but one is to maximize and another minimize the objective function. In this formulation, the  $P_i$  are the unknown variables of the linear program. The formulations of the two linear programs for calculating the upper and lower expected values are

### Objective

$$\text{Min/Max } E(u/d_j) = \sum U(\theta_i, d_j) \cdot P_i$$

Subject to

$$Bel(\theta_i) \leq P_i \leq Pl(\theta_i) \quad (7.4)$$

$$\sum_i P_i = 1.0$$

and

$$P_i \geq 0 \quad i=1, 2, \dots, n$$

Since the decision problem is to minimize the expected loss, the appropriate decision criterion might be to choose the decision  $d_j \in D$  that yields a minimum upper expected utility. Dempster called this the *miniupper decision against Bel* [8]. The decision making  $d_j$  based on such criterion is not necessarily the globally optimum decision. In fact, the true expected value determined by the true probabilities, which are not known, is somewhere between the upper and lower bounds. It is possible, therefore, to have another decision action  $d_k$  which will give a smaller expected loss than that given by  $d_j$ . Nevertheless, the decision choice  $d_j$  based on the miniupper decision criterion can be considered as minimizing the maximum limit cost which the system may face. In other words, all the possible expected costs will fall below this upper limit. Such a decision is the safest or most conservative one, and might be suitable where the expected cost must be minimized under all circumstances. For the situation where the decision problem is to maximize the expected benefits, a logical decision criterion might be to

choose the action  $d_j$  which gives the maximum lower expected value. This expected value is called the *maxilower decision against Bel*.

Dempster-Shafer decision theory is the use of Dempster-Shafer theory within the framework of conventional decision theory. The advantages of Dempster-Shafer theory over Bayesian theory are essentially the advantages of Dempster-Shafer decision theory over Bayesian decision theory. Bayesian decision theory requires the inferential problem to be confined to a special case of the statistical specification model in which the prior evidence has to be expressed by a conventional probability distribution and the new evidence expressed by its sample likelihood values. As decisions must be made in civil engineering practice under a wide range of circumstances, the Bayesian state of affairs appears to be unduly restrictive. On the other hand, Dempster-Shafer decision theory does not require the inferential problem to be confined to the statistical specification model. In fact, it only requires one to establish a belief function from each piece of evidence and then combine them according to Dempster's rule of combination to produce a resultant belief function. The decision analysis can then be undertaken by applying the Dempster-Shafer decision theory in conjunction with the resultant belief function. Just as Bayesian theory is a special case of Dempster-Shafer theory, it can be demonstrated that Bayesian decision theory is a special case

of the more general Dempster-Shafer decision theory. Under the same circumstance of Bayesian belief, both will yield identical results. That is, the miniupper and the maxilower decisions are equal and coincide with the optimal Bayesian decision.

### 7.3 THE APPLICATION OF DEMPSTER-SHAFER DECISION THEORY

The following water resources example is taken from R.J. McAniff *et al.* [17]. The problem was originally designed to demonstrate the use of Bayesian decision theory. It will be used here to show how Dempster-Shafer decision theory can be used to improve the decision making in a less restrictive situation.

#### 7.3.1 DESCRIPTION OF THE ORIGINAL PROBLEM

An agricultural producer is concerned about the future cost of an irrigation system. A decision set  $D=\{d_j, j=1,..5\}$  represents the possible irrigation systems between which the producer can choose. The cost will depend on the choice of irrigation system and a future energy price level. This energy price level is a random variable and its possible values are denoted by  $\Theta=\{\theta_i, i=1,2,...10\}$ . The prior probabilities  $P(\theta)$  over  $\Theta$  are known and are given in Table 7.1. New information is obtained from estimating the future energy price levels using a forecasting model. The estimated future energy level is denoted as  $z_k$ . The sample likelihood of  $z_k$  given various parameters  $\theta_i, i=1,2...10$ , is given in

Table 7.2. The utility function  $U(\theta_i, d_j)$  specifies the cost of choosing irrigation system  $d_j$

Table 7.1 Prior probabilities and utilities  
(after R.J. McAniff et al.)

Possi- ble price level $\theta_i$	Range of price in- crease %	Average price in- crease %	Prior proba- bilities $p(\theta_i)$	Utilities (costs) in dollars				
				Centre pivot $d_1$	Travel- ing trickle $d_2$	Gated pipe with return $d_3$	Open ditch $d_4$	Dead level $d_5$
$\theta_1$	<0	<0	0.11	268,190	269,750	273,780	287,430	299,780
$\theta_2$	0-3	1.5	0.07	280,020	280,930	284,440	306,540	313,300
$\theta_3$	3-6	4.5	0.09	308,620	307,540	310,050	331,370	330,720
$\theta_4$	6-9	7.5	0.11	344,760	341,510	342,810	371,150	358,800
$\theta_5$	9-12	10.5	0.12	382,080	385,320	385,060	421,980	394,810
$\theta_6$	12-15	13.5	0.12	452,790	441,870	439,530	488,280	441,610
$\theta_7$	15-18	16.5	0.11	531,180	515,060	509,990	573,170	501,540
$\theta_8$	18-21	19.5	0.09	646,880	609,830	601,250	683,540	579,410
$\theta_9$	21-24	22.5	0.07	764,400	726,960	719,420	826,540	680,290
$\theta_{10}$	>24		0.11	903,500	862,160	844,350	977,600	787,020

Table 7.2 Sample likelihood of  $Z_k = 10.5\%$   
(after R.J. McAniff et al.)

$\theta_i$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$
$l(z_k/\theta_i)$	0.031	0.041	0.071	0.143	0.408
$\theta_i$	$\theta_6$	$\theta_7$	$\theta_8$	$\theta_9$	$\theta_{10}$
$l(z_k/\theta_i)$	0.143	0.071	0.041	0.031	0.020

and obtaining energy price level  $\theta_i$  is also given in Table 7.1. The problem is to choose the irrigation system so that the expected cost will be minimized.

Since the prior probability distribution  $P(\theta_i)$  and sample likelihood  $L(z_k/\theta_i)$  are given explicitly, Bayesian posteriors can be obtained directly. The decision analysis using Bayesian decision theory, based on these posteriors, can then be undertaken. The expected costs based on the posteriors for various decision alternatives are given in Table 7.3.

Table 7.3 Expected costs based on Bayesian posteriors

Expected costs (dollars)	Decision Choice				
	Center pivot $d_1$	Traveling trickle $d_2$	Gated pipe with return $d_3$	Open ditch $d_4$	Dead level $d_5$
418,881	410,129	409,046*	451,278	416,102	

It indicates that the irrigation system choice "Gated Pipe with Return" should be chosen to minimize the costs.

Now, consider the situation in which there is some doubt about the prior probability distribution. This distribution has to be discounted according to Shafer's discounting method described in Chapter 5. Bayesian decision theory is no longer appropriate, and Dempster-Shafer decision theory has to be used. This is demonstrated in Sec.

### 7.3.2

### 7.3.2 DEMPSTER-SHAFER DECISION ANALYSIS

Like the Bayesian decision theory, Dempster-Shafer decision theory needs one first to pool all the available information together and calculate a final belief function. In our problem, the prior belief function is obtained from prior information by discounting the prior probability distribution  $P(\theta_i)$  by a factor  $a$ . Therefore, the prior belief function can be expressed by its basic probability assignment as

$$m_1(\theta_i) = (1-a)P(\theta_i) \quad m_1(\theta) = a \quad i=1, 2, \dots, 10$$

As was discussed in Chapter 4, the Dempster-Shafer approach to dealing with the inferential evidence  $Z_k$  is to derive a consonant belief function from such evidence. The basic probability assignment  $m_2(A)$  of the consonant belief function, based on the sample likelihood given in Table 2, is given in Table 7.4.

Table 7.4 Basic probability assignment of a consonant belief function derived from inferential evidence

A	$\theta_5$	$\theta_4 \dots \theta_6$	$\theta_3 \dots \theta_7$	$\theta_2 \dots \theta_8$	$\theta_1 \dots \theta_9$	$\theta_1 \dots \theta_{10}$
$m_2(A)$	0.650	0.175	0.075	0.025	0.025	0.050

Once the prior belief function expressed by  $m_1(A)$  and the consonant belief function expressed by  $m_2(A)$  are obtained, they can then be combined to give a final belief

function  $\text{Bel}(A)$ . The combination will obviously depend on the discounting factor  $a$ . In our example,  $a$  is ranged over several values, i.e.  $a=0.0, 0.2, 0.8, 1.0$ . The larger the  $a$  is, the greater the doubt being expressed about the prior probability distribution. Thus  $a=0.0$  indicates that the prior probability distribution is based on solid information, and  $a=1.0$  indicates that the prior information is no better than complete ignorance. In this latter case,

Table 7.5 Combined basic probability assignments

	Discounting factor $a$			
	0.0	0.2	0.8	1.0
$\theta_1$	0.031	0.016	0.002	0
$\theta_2$	0.026	0.013	0.002	0
$\theta_3$	0.058	0.030	0.004	0
$\theta_4$	0.142	0.074	0.009	0
$\theta_5$	0.444	0.543	0.637	0.650
$\theta_6$	0.155	0.081	0.010	0
$\theta_7$	0.071	0.037	0.005	0
$\theta_8$	0.033	0.017	0.002	0
$\theta_9$	0.019	0.010	0.001	0
$\theta_{10}$	0.020	0.011	0.001	0
$\theta_4 \dots \theta_6$	0	0.084	0.164	0.175
$\theta_3 \dots \theta_7$	0	0.036	0.070	0.075
$\theta_2 \dots \theta_8$	0	0.012	0.023	0.025
$\theta_1 \dots \theta_9$	0	0.012	0.023	0.025
$\theta_1 \dots \theta_{10}$	0	0.024	0.047	0.050

the combined belief function is determined solely by the consonant belief function, i.e. the inferential evidence is the only effective information source. The resulting basic probability assignments  $m(A)$  for different  $a$  values are

given in Table 7.5. The belief and plausibility values for singletons derived from these results are given in Table 7.6.

Table 7.6 Belief and plausibility values of singletons

	$\alpha=0.0$		$\alpha=0.2$		$\alpha=0.8$		$\alpha=1.0$	
	Bel(A)	pl(A)	Bel(A)	pl(A)	Bel(A)	pl(A)	Bel(A)	pl(A)
$\theta_1$	0.031	0.031	0.016	0.052	0.002	0.072	0	0.075
$\theta_2$	0.026	0.026	0.013	0.062	0.002	0.095	0	0.100
$\theta_3$	0.058	0.058	0.030	0.114	0.004	0.168	0	0.175
$\theta_4$	0.142	0.142	0.074	0.242	0.009	0.337	0	0.350
$\theta_5$	0.444	0.444	0.543	0.711	0.637	0.965	0.650	1.000
$\theta_6$	0.155	0.155	0.081	0.249	0.010	0.338	0	0.350
$\theta_7$	0.071	0.071	0.037	0.121	0.005	0.168	0	0.175
$\theta_8$	0.033	0.033	0.017	0.065	0.002	0.096	0	0.100
$\theta_9$	0.019	0.019	0.010	0.046	0.001	0.071	0	0.075
$\theta_{10}$	0.020	0.020	0.011	0.035	0.001	0.048	0	0.050

Once the beliefs and plausibilities are obtained for the single elements, the upper and lower expected utilities can be computed for each decision action  $d_j$ . This was repeated for each value of the discounting factor  $\alpha$ . As was discussed in Sec. 7.1, the computation involves formulation and solution of a linear program. For example, consider the situation in which  $\alpha=0.2$  and the decision choice  $d_j=d_1=\text{"Centre Pivot"}$ . The linear program formulation for calculating the upper and lower expected losses are

Objective:

$$\text{Min/Max } Z = 268190P_1 + 280020P_2 + 308620P_3 + 344760P_4 + 382080P_5$$

$$+452790P_6 + 531180P_7 + 646880P_8 + 764400P_9 + 903500P_{10}$$

Subject to:

(1) The upper and lower limits of the probabilities of singletons defined by the supports and plausibilities are

$$0.016 \leq P_1 \leq 0.052$$

$$0.013 \leq P_2 \leq 0.062$$

$$0.030 \leq P_3 \leq 0.114$$

$$0.074 \leq P_4 \leq 0.242$$

$$0.543 \leq P_5 \leq 0.711$$

$$0.081 \leq P_6 \leq 0.249$$

$$0.037 \leq P_7 \leq 0.121$$

$$0.017 \leq P_8 \leq 0.065$$

$$0.010 \leq P_9 \leq 0.046$$

$$0.011 \leq P_{10} \leq 0.035$$

(2) The constraint on the summation of the probabilities

$$\sum P_i = 1.0$$

(3) The nonnegativity constraints

$$P_i \geq 0 \quad i=1, 2, \dots, 10$$

The calculated results for the various combined  $a$  values and different decision choices  $d_j$  are given in Table 7.7. Both miniupper and maxilower expected costs are shown for all decision choices.

From Table 7.7, it is seen that the greater the discount factor  $a$ , the larger the difference between the expected upper and lower values.

Table 7.7 The upper and lower expected costs

Discount rate $a$	Decision set					
	Center pivot $d_1$		Travelling trickle $d_2$		Gated pipe with return $d_3$	
	$E_1^*$	$E_{1*}$	$E_2^*$	$E_{2*}$	$E_3^*$	$E_{3*}$
0.0	418881	418881	410129	410129	409046*	409046
0.2	448906	375115	442286	374340	440159*	374513
0.8	478430	353430	470135	355206	367121	356039
1.0	481905	350480	473419	352605	470303	353529

Table 7.7 (continuation)

Discount rate $a$	Decision set			
	Open ditch $d_4$		Dead level $d_5$	
	$E_4^*$	$E_{4*}$	$E_5^*$	$E_{5*}$
0.0	451278	451278	416102	416102
0.2	488780	409673	441943	386083
0.8	512281	387505	464895*	370444
1.0	525117	384492	467604*	368319

When  $a=1.0$ , i.e. the Bayesian situation, the expected miniupper and maxilower costs coincide as there is no "ignorance" associated with the prior distribution. When  $a=1.0$ , i.e. the prior information is equivalent to complete ignorance, the calculated upper and lower expected values are entirely dependent on the "new" inferential evidence  $Z_k$  provided by the forecasting model. Decision  $d_3$  yields the

smallest miniupper expected cost and, in this context, would be the appropriate choice.

When the discounting rates  $a$  is 0.2, suggesting that the decision maker still has reasonable confidence in the prior probability distribution, the agriculture producer should still choose decision action  $d_3$ . When the discounting rate  $a$  increases to 0.8 or 1.0 i.e. when the decision maker expresses strong doubts about the prior probability distribution, the decision action  $d_5$  becomes the optimal choice. Maxilower expected costs are also shown in Table 7.7 for completeness but does not have any bearing on the optimal decision choice in this example.

#### 7.4 SUMMARY

Similar to Bayesian decision theory, Dempster-Shafer decision theory is the application of Dempster-Shafer theory within the framework of conventional decision theory. Therefore, the advantages of Dempster-Shafer decision theory over Bayesian decision theory are essentially the advantages of Dempster-Shafer theory over Bayesian theory. Since decision making in civil engineering practice is undertaken under a wide variety of circumstances, Dempster-Shafer decision theory seems to be much more attractive in practical application than Bayesian decision theory.

An example of applying Dempster-Shafer decision theory in solving a decision problem in water resources engineering shows that it does improve the decision results under

circumstances which Bayesian decision theory cannot accommodate. While little work has been done in applying Dempster-Shafer theory in solving real world practical problems, the preceding discussion of the Dempster-Shafer decision theory together with the example given, indicates a very promising new starting point in the application of this new theory in practice. Obviously, full acceptance and adoption of Dempster-Shafer theory in practice still requires a large amount of work.

## 8. CONCLUSIONS

This research has concentrated on the comparison of Bayesian theory and Dempster-Shafer theory and, to a limited extent, the application of Dempster-Shafer theory to some elementary water resources engineering problems. The equivalence between the two theories was discussed in Chapter 4. The investigation of divergence of results from Bayesian and Dempster-Shafer approaches and the sensitivity analysis were presented in Chapter 5. In Chapter 6, the conceptual difference between the two theories and the issues of sorting collective sources of evidence and representing them through belief functions were argued. Finally in Chapter 7, the Dempster-Shafer decision theory was presented. Its application in water resources engineering practice were demonstrated through a real water resources design example. The conclusions of this research can be summarized as follows.

Bayesian theory, as a special case of Dempster-Shafer theory, requires one to think of the inferential problem in terms of the statistical specification model, though clearly not all of the inferential problems satisfy the requirements of this model. This in turn requires that the evidence be grouped into two parts: namely the "old" evidence and the "new" evidence. Bayesian theory then requires one to specify explicitly the prior probabilities from the "old" evidence and the sample likelihoods from the "new" evidence. If all of this faithfully reflects the real situation then Bayesian

theory can be considered to be the proper approach. But often the evidence may not be specific enough to specify or support these probabilities. In this case, though subjective judgements may be used to resolve the difficulties, one may not feel comfortable or confident about inferential results based on those probability judgements.

Dempster-Shafer theory addresses the more general case and pays much more attention to the quality and character of the evidence than the probability model itself. It does not need the inferential problem to be fitted into the statistical specification model. But if the inferential problem does fit into the statistical specification model, then Dempster-Shafer theory accommodates the situation properly. In fact through the use of discounting, it also accommodates small deviation from the statistical specification model with ease. The high sensitivity of the results to discounting under certain condition, which has been demonstrated in Chapter 5, is disconcerting in the face of the necessarily precise view taken by the Bayesian approach. The prior probability judgements may be expressed either in the form of a Bayesian belief function or as a general belief function, the former being identical to a Bayesian approach. The more general Dempster-Shafer theory requires one to construct a belief function from each piece of evidence and then combine all of the belief functions by the Dempster's rule of combination. It is therefore seen that there are at least two advantages of Dempster-Shafer

theory over Bayesian theory: the Dempster-Shafer theory does not need the inferential problem to be confined to the statistical specification model; and it provides the concept of a belief function which may express the evidence in a much more realistic and faithful way than the very restrictive conventional expression of probability in Bayesian theory.

There are essentially two types of uncertainties, informational uncertainty or IU and natural uncertainty or NU. The probability judgements based on a piece of evidence for the set associated with NU are realistically expressed by the frequency-like conventional probability distribution while for the set associated with IU these are more realistically expressed by a consonant belief function. In Dempster-Shafer theory, since the uncertainties associated with the frame of discernment  $\Theta$  belong to the IU category, the more appropriate belief structure on the frame  $\Theta$  is the consonant type of belief function. Though it is more reasonable to adopt a consonant belief function, the conventional probability distribution is not precluded as an effective way of expressing evidence.

Collecting and rearranging pieces of evidence, and expressing them through belief functions, parallel very important aspects of engineering practice. So far, little or no work has been undertaken to formalize the processes involved. The discussions in Sec. 6.3 and Sec. 6.4 did suggest some possible components of a formalized struture.

With the variety of the evidence typically bearing on a frame of discernment, it may not be necessary or appropriate that each piece of evidence be expressed by its own explicit belief function and the evidence may have to be grouped. Some possible guidelines are that: the vague pieces of evidence should be put together with the more specific evidence; the interdependent pieces of evidence should be grouped together; the more complicated evidence should be decomposed into several simple pieces of evidence if it is difficult to express such evidence by a complex general belief function. In addition, just a few classes of belief functions appear to represent the majority types of evidence occurring in engineering practice. The obtained pieces of evidence should be grouped into the types of evidence which correspond to the classes of belief functions. The implications of the guidelines mentioned here are not yet fully understood and further investigation is essential to the development of a suitable scheme for expressing domain specific evidence in practice.

Expressing evidence through a belief function can also be assisted by comparing the evidence with a scale of canonical examples and adopting the example which matches the evidence best. The framework for the belief function provided for this example can then be used as the belief function for the natural piece of evidence. Subjectivity of course enters when an expert is asked to choose the canonical example and make the numerical belief judgements.

Clearly the choices of different experts will lead to different inferential results for the same inferential problem. Bounding and sensitivity methods may therefore be essential ingredients in a practical belief entry system.

Dempster-Shafer decision theory, i.e. the application of Dempster-Shafer theory in conjunction with conventional decision theory, fully exploits the advantages of Dempster-Shafer theory over Bayesian theory. It can be regarded as a generalization of Bayesian decision theory also. Under those conditions in which Dempster-Shafer theory and Bayesian theory are identical, it is reasoning that Dempster-Shafer decision theory and Bayesian decision theory are also identical. However, Dempster-Shafer decision theory can be used in the more general situation in which Bayesian decision theory is overly restrictive and unsatisfactory. The simple example of an application of this decision theory in resolving a realistic water resources engineering design problem demonstrates the potential of the Dempster-Shafer decision theory to improve the decision analysis framework.

While Dempster-Shafer theory gains some substantial advantages over Bayesian theory, these are not achieved without cost and we may anticipate some difficulties in its application. The difficulties may be summarised in the following categories.

Firstly, in Bayesian theory, both the prior probabilities based on prior evidence and the sample likelihood based on the inferential evidence are expressed

in the form of familiar, conventional probability distributions. In implementing Dempster-Shafer theory, belief is expressed in the unfamiliar format of a belief function. Also, since the belief value can be assigned to any grouping of propositions, one faces a much greater number of potential choices of belief assignment in Dempster-Shafer theory. This increased choice will make it difficult for one to choose the appropriate level of propositions and the appropriate belief assignments for any given piece of evidence. Even though special classes of belief functions can be adopted, as discussed in Chapter 6, the difficulty still remains about which type of belief function should be chosen and how the evidence should be expressed by a numerical belief function. In Chapter 6, the idea of canonical example was presented in order to facilitate the process of expressing evidence by a belief function. Even though such an idea is theoretically attractive, further work still needs to be done to facilitate its adoption in engineering practice.

Secondly, the Dempster-Shafer belief function is less manageable than the conventional probability distribution in Bayesian theory. In Bayesian theory, the probability distribution is often mathematically well expressed and the implications of such a probability distribution broadly understood. In Dempster-Shafer theory, however, the expression of general belief function is more complex, since it is defined on the whole frame of discernment the full

implications of the belief function are much more difficult to interpret, this being especially true when the frame becomes large.

Finally, recall that Bayesian theory can be expressed in continuous form. Since the variables in many situations in practice are continuous variables, the continuous expression of Bayesian theory is often preferred. However, most of the discussion of Dempster-Shafer theory so far has been confined to the discrete form and very little work has been done for the continuous situation. T.M. Strat[23] studied this problem based on the assumption that, for a continuous random variable, only the proposition which is formed by a set of contiguous values is of significance. For example, a random variable  $x$  has a set of possible, ascendingly ordered, discrete values  $x_1, \dots, x_n$ . Then propositions like  $\{x_3, x_4, x_5\}$  will be of significance and propositions like  $\{x_2, x_4, x_5\}$  will be considered to be impossible. The adoption of this assumption reduces the size of the frame of discernment to less than half of its original size. The continuous frame of discernment is obtained by reducing the interval of the discrete element to its limit. A continuous belief function based on a continuous frame of discernment can then be defined if a piece of evidence is obtained. Strat did derive the general mathematical formulas for the continuous belief function and Dempster's rule of combination. These formulas turned out to involve very complex and intractible expressions.

Nevertheless, it seems to provide a good starting point. It is worthwhile to note the way the assumption is incorporated and the way the problem is stretched to accommodate the continuous frame.

In conclusion, the Dempster-Shafer scheme appears to offer many significant advantages beyond the Bayesian scheme under the conditions of uncertainty experienced in civil engineering. As the already established Bayesian scheme occupies the position of special case within the Dempster-Shafer scheme, in both inference and decision analysis, no loss of capability is experienced when opting for the Dempster-Shafer scheme (except possibly in some cases where continuous variables are involved). While implementation of the Dempster-Shafer scheme will clearly involve overcoming a number of difficulties, which can range from lack of familiarity with the less conventional theory involved to numerical intractability with large frames, none would appear serious enough to obstruct the implementation. The substantial improvements promised by the Dempster-Shafer scheme in the rationalization and formalization of practical decision making under uncertainty would appear to provide more than adequate incentive to overcoming these difficulties.

## REFERENCES

1. Ang, A. H-S. and Tang, W.H., *Probability Concepts in Engineering Planning and Design*, Wiley, New York, 1975.
2. Barneyy, J.A., "Computational Methods for A Mathematical Theory of Evidence," *Proceedings Seventh National Joint Conference on Artificial Intelligence*, Vancouver, B.C., pp. 319-325, Aug. 1981.
3. Benjamin, J.R. and Cornell, C.A., *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill, New York, 1970.
4. Bodo, B. and Unny, T.E., "Model Uncertainty in Flood Frequency Analysis and Frequency-Based Design," *Water Resour. Res.*, 12(6), pp. 1109-1117, 1976.
5. Caselton, W.F., Froese, T.M., Russell, A.D. and Luo, W., "Belief Input Procedures for Dempster-Shafer Based Expert System," *Artificial Intelligence in Engineering: Robotics and Processes*, Computational Mechanics Publications, Southampton, pp. 351-370, 1988.
6. Davis, D.R., Kisiel, C.C. and Duckstein, L., "Bayesian Decision Theory Applied to Design in Hydrology," *Water Resour. Res.*, 8(1), 33-41, 1972.
7. Dempster, A.P., "A Generalization of Bayesian Inference," *Journal of the Royal Statistical Society, Series B*, 30, pp. 205-247.
8. Dempster, A.P. and Kong, A., "Comment," *Statistical Science*, Vol. 2, No. 1, pp. 32-36.

9. Duda, R.O., Hart, P.E. and Nilsson, N.J., "Subjective Bayesian Methods for Rule-based Inference Systems," *AFIPS Conf. Proc.*, 45, pp. 1075-1082, 1976.
10. Garvey, T.D., Lowrance, J.D. and Fischler, M.A., "An Inference Technique for Intelligence Knowledge from Disparate Sources," *Proceedings Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B.C., pp. 319-325, Aug. 1981.
11. Gordon, J. and Shortliffe, E.H., "A Method for Managing Evidential Reasoning in A Hierachical Hypothesis Space," *Artificial Intelligence*, Vol. 26, pp. 323-357, July 1985.
12. Gordon, J. and Shortliffe, E.H., "The Dempster-Shafer Theory of Evidence," in: B.G. Buchanan and E.H. Shortliffe (eds.), *Rule-based Expert Systems: The MYCIN Experiments of the Stanford-Heuristic Programming Project*, (Addison-Wesley, Reading, MA), pp. 272-292, 1984.
13. Jeffrey, R.C., *The Logic of Decision*, McGraw-Hill, New York, 1965.
14. Krantz, D.H. and Miyamoto, J., "Prors and Likelihood Ratios as Evidence," *Journal of the American Statistical Association*, 78, pp. 418-423, 1983.
15. Lu, S.Y. and Stephanou, H.E., "A Set-theoretic Framework for the Processing of Uncertain Knowledge," *Proceedings Fourth National Conference on Artificial Intelligence*, Austin, TX, pp. 308-313, 1984.
16. Moore, P.G. and Thomas, H., "Measuring Uncertainty," *Omega*, Vol. 3, No. 6, pp. 657-672, 1975.

17. McAniff, R.J., Flug, M. and Wade, J., "Bayesian Analysis of Energy Prices on Irrigation," *Journal of the Water Resources Planning and Management Division*, ASCE, 106(WR2), pp. 401-408.
18. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, New Jersey, 1976.
19. Shafer, G., "Belief Functions and Parametric Models," *Journal of the Royal Statistical Society, Series B*, 44, pp. 322-352, 1982.
20. Shafer, G. and Tversky, A., "Languages and Designs for Probability Judgement," *Cognitive Science*, 9, pp. 309-339, 1983.
21. Shafer, G., "Probability Judgement in Artificial Intelligence and Expert System," *Statistical Science*, Vol. 2, No. 1, pp. 3-44, 1987.
22. Shane, R.M. and Gaver, D.P., "Statistical Decision Theory Techniques for the Revision of Mean Flood Flow Regression Estimates," *Water Resour. Res.*, 6(6), 1649-1654, 1970.
23. Strat, T.M., "Continuous Belief Functions for Evidential Reasoning," *Proceedings Fourth National Conference on Artificial Intelligence*, Austin, Texas, pp. 308-313, Aug. 1984.
24. Tschannerl, G., "Designing Reservoirs with Short Streamflow Records," *Water Resour. Res.*, 7(4), 827-833, 1971.
25. Tversky, A. And Kahneman, D., "Judgement under Uncertainty: Heuristics and Biases," *Science*, Vol. 185, pp. 1124-131, 1974.

26. Vicens, G.J., Rodriguez-Iturbe, I. and Schaake, J.C. "A Bayesian Framework for the Use of Regional Information in Hydrology," *Water Resour. Res.*, 11(3), 405-414, 1972.
27. Wood, E.F. and Rodriguez-Iturbe, I., "Bayesian Inference and Decision Making for Extreme Hydrologic Events," *Water Resour. Res.*, 11(4), 533-542, 1975.
28. Yager, R.R., "On the Dempster-Shafer Framework," *Artificial Sciences*, Vol. 41, pp. 93-137, 1987.