Term 081
King Fahd University of Petroleum and Minerals
EE 562 – Digital Signal Processing
Course Project

Fault Diagnosis: a Dempster-Shafer Theory Approach

By

Mohammad Shahab

227598

For

Dr. Mohammad Deriche

11 February 2009

Fault Diagnosis: a Dempster-Shafer Theory Approach

I. Introduction

Here in this report, a relatively new concept is introduced, namely, the Dempster-Shafer Theory of Evidence. This theory will be projected to a Fault Diagnosis application. This theory utilizes the concept of sensor (data) fusion. We can define sensor fusion as the combining of sensory data (or data derived from sensory data) from disparate sources such that the resulting information is 'better' than would be possible when these sources were used individually. There is no single sensor that can constantly obtain all the information required for fault diagnosis. As the development of sensor technology and signal processing methods progresses, more information can be obtained.

In section II, a detailed study will be given for the Dempster-Shafer Theory and in section III will briefly be about fault diagnosis. Section IV will give some literature survey of related work on Dempster-Shafer Theory and the use of it in Fault Diagnosis. Section V will give the work done by author with results. Section VI will conclude the report.

II. Dempster-Shafer Theory

Dempster-Shafer Theory is a mathematical theory of evidence. For discrete classes space, Dempster-Shafer theory can be interpreted as a generalization of probability theory where probabilities are assigned to *sets*. Traditionally in probability theory, evidence is associated with only one possible event. In Dempster-Shafer Theory, evidence can be associated with multiple possible events, e.g., sets of events. Where the evidence is sufficient to give assignments of probabilities to single events, the Dempster-Shafer model becomes a traditional probabilistic formulation. One of the most important features of Dempster-Shafer theory is that it has the ability to model narrowing hypothesis set as evidence builds up.

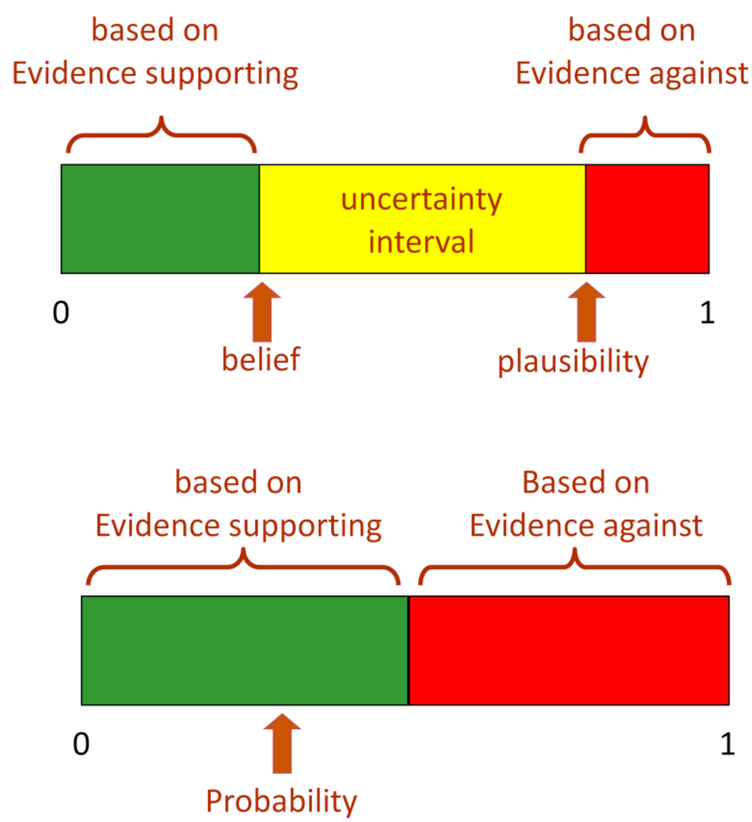
There are three major functions in Dempster-Shafer theory: the basic belief assignment function (BBA or m(.)), the Belief function (Bel), and the Plausibility function (Pl). The basic belief assignment (BBA) is a primitive of evidence theory. Generally speaking, the term "basic belief assignment" does not refer to probability in a classical way. The BBA, represented by m(.), defines a mapping of the power set to the interval between 0 and 1, where the BBA of the empty set is 0 and the summation of the BBA's of all the subsets of the power set is 1. The value of the BBA for a given set A (represented as m(A)), expresses the amount of all related and available evidence that supports the claim that a particular element of X (the universal set) belongs to the set A but to no particular subset of A. The value of m(A) is related only to the set A and makes no additional assumptions about any subsets of A. Any further evidence on the subsets of A would be represented by another BBA. The BBA can be shown by the equations:

$$m(X) \rightarrow [0,1] \quad (1)$$

$$m(\phi) = 0 \quad (2)$$

$$\sum_{\forall A \in X} m(A) = 1 \quad (3)$$

For any classification problem, discrete number of classes are defined. For a set of Θ={class 1, class 2, class 3, ... }, this is called the frame of discernment. The power set of Θ, 2 , contains all bossible subsets of Θ. Some it becomes, the power set, P(Θ)={ class 1, class 2, class 3, ..., {class 1, class 2}, {class 1, class 3},... {class 2, class 3}, ..., {class 1, class 2, class3}, ... {class 1, class 2, class 3, ...}}. If Θ has n elements, P(Θ) will have 2^n.



The above figure shows the difference between the two concepts of Probability versus the concept of Dempster-Shafer. You can see the for probability gives information about support of hypothesis and simultaneously gives information about the negation (compliment). However, in D-S theory, the ‘uncertainty’ or ‘not knowing’ is also modeled. We can see this mathematically as:

- $m(\Theta) \neq 1$
- m(class 1) + m(class 1') < 1

For data fusion, multiple sources of information is combined to give us 'better' judgment of situation. The purpose of accumulation of information is to summarize and simplify a collection of data whether the data is coming from a single source or multiple sources. Combination rules are the special types of accumulation methods for data obtained from multiple sources. These multiple sources provide different measures about the same frame of discernment and Dempster-Shafer theory is based on the assumption that these sources are independent. Demspter introduced his combination rule (for two sources) of:

$$m^{1,2}(C) = \frac{\sum_{A \cap B = C} m^1(A)\, m^2(B)}{\sum_{A \cap B \neq \varnothing} m^1(A)\, m^2(B)} = \frac{\sum_{A \cap B = C} m^1(A)\, m^2(B)}{1 - \sum_{A \cap B = \varnothing} m^1(A)\, m^2(B)} \tag{4}$$

III. <u>Fault Diagnosis</u>

Fault detection is a subfield of control engineering which concerns itself with monitoring a system, identifying when a fault has occurred and classifying the type of fault and its location. Two approaches can be distinguished: direct pattern recognition of sensor readings that indicate a fault and an analysis of the inconsistency between the sensor readings and expected values, derived from some model. Other terms describe the field can be seen in literature or industry are: Condition Monitoring, Health Monitoring, Fault Detection, etc. Condition Monitoring can be defined as monitoring a parameter of condition in machinery, such that a significant change is indicative of a failure. Mostly, uses of this are applied in predictive maintenance.

For a diagnosis system, it is essential requirement to have a good command about the diagnosed system knowledge. Applying data fusion technique to fault detection is how to use enormous data to find hidden new technology and revise former technology according to the system's motion station, enable the diagnosis system work more rapidly, accurately. In real information fusion system, multi-sensor's data are fused according to a ruler firstly, and then combining the fused information with the information coming from the machine's self and other relative regulation together to deduce fault position and fault reason according to specified method.

It is mostly used for rotating machines. Source of information is output of vibration analysis. Measurements can be taken on machine bearing with piezo-electric sensors to measure the vibrations. On the majority of critical machines, eddy-current sensors directly observe the rotating shafts to measure the radial vibration of the shaft. The level of vibration can be compared with history data.

IV. LITERATURE SURVEY

Here, we will go through some works in literature that dealt with the Dempster-Shafer Theory and the application of Dempster-Shafer theory in Fault Diagnosis.

In an interesting paper [1], presentation of a new combination technique based on the Dempster- Shafer theory of evidence is introduced. Also along with that, general overview of the concept of The Dempster-Shafer theory of evidence is explained in detail. In this paper, they propose an implementation which adapts to training data so that the overall mean square error is minimized.

In [2], the report gives a overview of both a Bayesian Fusion algorithm along with Dempster-Shafer methods. It begin with an outline of Bayes theory. Then it goes with approach of Dempster-Shafer theory. Also, a great section of the technical report is left to discuss several application of the Fusion in many domains.

In [3] paper, also comparison between Bayesian and Dempster-Shafer approaches is done. Other variations are also discussed. It shows that Dempster-Shafer combination do not require prior information unlike the Bayesian combination.

In [4], it is a book chapter that is comprises an early reference for the Dempster-Shafer Theory. It is highly cited. It goes in an educational way to give information about the theory. Also, they emphasize the explanation with a standard problem of Medical Diagnosis.

In [5], it explores the algorithmic point of view of the Dempster-Shafer Theory. It addresses two major problems with the theory: (i) understanding what the calculated values of belief mean; and (ii) the computational problems of Dempster's rule. Many algorithms are tested such as Monte-Carlo, etc.

The Thesis [6] is related to interesting application. It addresses the thought of having computers understand human users context information. The work proposes a context-sensing implementation method that can combine sensor outputs with subjective judgments. A case study with several simulated sensors using recorded data and artificially generated sensor outputs distributed over a LAN network is analyzed.

In the work [7], Fault diagnosis is studied as it requires reasoning and decision-making based on diagnostic knowledge and features extracted from raw data. In practice, fault features may be uncertain and imprecise due to sensor errors, fluctuating working conditions, and limitations of feature extraction methods. Features may not be apparent when a fault is in the early stages of development. In addition, diagnostic knowledge is not always accurate because most of it is obtained from experience. This study addresses an application of an improved Demspter–Shafer evidence theory in gearbox fault diagnosis and compared with conventional diagnostic methods.

In [8], paper is concerned with the use of Dempster-Shafer theory in fusion classifiers in a condition monitoring application. It addresses that the use of predictive accuracy for basic belief assignments (BBA) can improve the overall system performance when compared to traditional mass assignment techniques. The approach tests the effectiveness in a case study involving the detection of static thermostatic valve faults in a diesel engine cooling system.

In [9], paper addresses an approach for the fault diagnosis for induction motors by Dempster–Shafer theory. Features are extracted from motor stator current and vibration signals. The technique makes it possible for on-line application. The fusion of classification results from vibration and current increases the diagnosis accuracy. The efficiency of the proposed system is demonstrated by detecting motor electrical and mechanical faults originated from the induction motors.

In [10], the paper development of the problem is used in further sections of this report. In [10], it addresses an engine diagnostics which is a typical multi-sensor fusion problem. It involves the use of multi-sensor information such as vibration, sound, pressure and temperature, to detect and identify faults. From the viewpoint of evidence theory, information obtained from each sensor can be considered as a piece of evidence, and as such, multi-sensor based engine diagnosis can be viewed as a problem of evidence fusion. It also introduces two new methods for enhancing the effectiveness of mass functions in modeling and combining pieces of evidence. They also propose a rule for making rational decisions with respect to engine quality, and present a criterion to evaluate the performance of the proposed information fusion system. A case study is demonstrated to show the efficiency of this system in dealing with imprecise information and conflicts that may come up in the sensors.

V. WORK DONE

Here in this section, a specific example is computed as an application of the Dempster-Shafer Theory in a machine diagnosis situation. Most of work done assumes some of the method selected in [10] to analyze sensor values.

So, for a machine, certain states (or features) are measured as sensor outputs:

$$X = [x_1 \ x_2 \ \cdots \ x_n]$$

Each 'x' resemble some indication about the state of the machine. A table is constructed to tabulate the faults with corresponding machine state values.

$$\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_K \end{bmatrix} = \begin{bmatrix} x_{11}^f & x_{21}^f & \cdots & x_{n1}^f \\ x_{12}^f & \cdots & & \\ \vdots & & \ddots & \\ x_{1k}^f & \cdots & \cdots & x_{nk}^f \end{bmatrix}$$

With 'K' faults that could happen, each is marked by the corresponding set values of the states. These values are coming from analysis of the history of the system or from the analytical study of it.

So, for any time for any sensor, a measure should be developed to have indication about the fault. So, for sensor 'i', Fault 'k' is measure by the distance:

$$d_{ik} = \sqrt{\sum_n (x_n - x_{nk}^f)^2}$$

This is computed for all states at each sensor. From analyzing this 'distance', one can see that as the states measured are closer (i.e. low distance), the higher the chance that a specific fault is occurring. So, a nice measure can be seen as

$$p_{ik} = \frac{1}{d_{ik}}$$

This will contain information about faults at every sensor.

- Specific Problem

Let us assume that we have a machine with probable two faults. Only one state is indicative of the behavior of the system. Two sensors are measuring this state. The frame of discernment can be seen as

$$\Theta = \{F_0, F_1, F_2\}$$

With '0' means a healthy machine. So, the power set of the problem becomes

$$\{F_0, F_1, F_2, \{F_1, F_2\}\}$$

As we can not have any fault with 'no fault' class in the same time. The last element of the power set is concerned with a 'don't know' class about the machine. As two sensor is measure is sole state, two m(.)'s should be available.

Let us assume that the state of the machine 'x' have a range of [0, 2]. The classes that give us the information about the health of the system can be

$F_0 \rightarrow$ [0, 0.5]

$F_1 \rightarrow$ [0.5, 1]

$F_2 \rightarrow$ [1, 1.5]

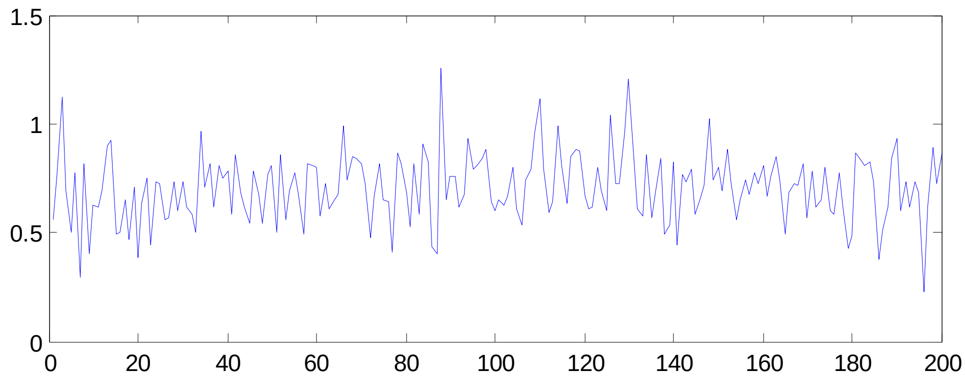
[1.5, 2] will indicate the 'don't know' class about the health. So, we can a measure from each sensor about each fault as

$$d_i = [\frac{1}{|0.25 - s_1|}, \frac{1}{|0.75 - s_1|}, \frac{1}{|1.25 - s_1|}, \frac{1}{|1.75 - s_1|}]$$

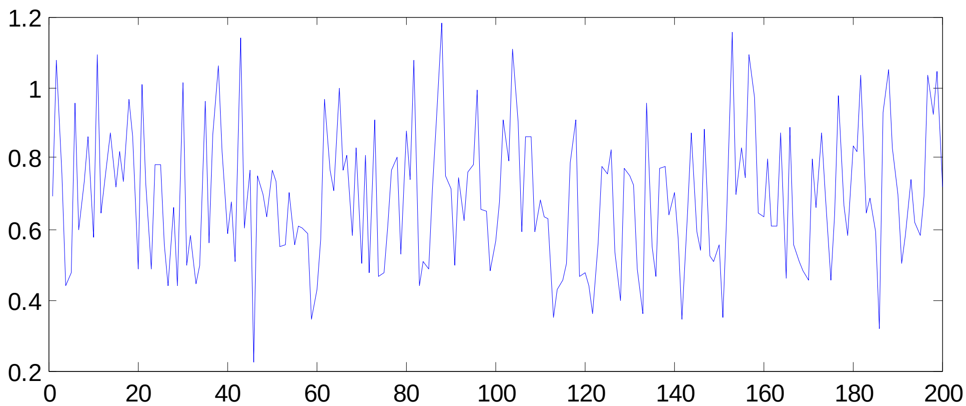
With 's' being the sensor reading. The values set are the middle of each of the intervals above. These measures should be also normalized to have the m(.) between 0 and 1.

To test the performance of the combination effect with the concept of Dempster-Shafer theory, let us force a machine fault. Let us assume a type 'fault 1' with actual state value of x = 0.7. As the two sensors are responsible of measuring this state, readings can be seen in figure below.

Sensor 1

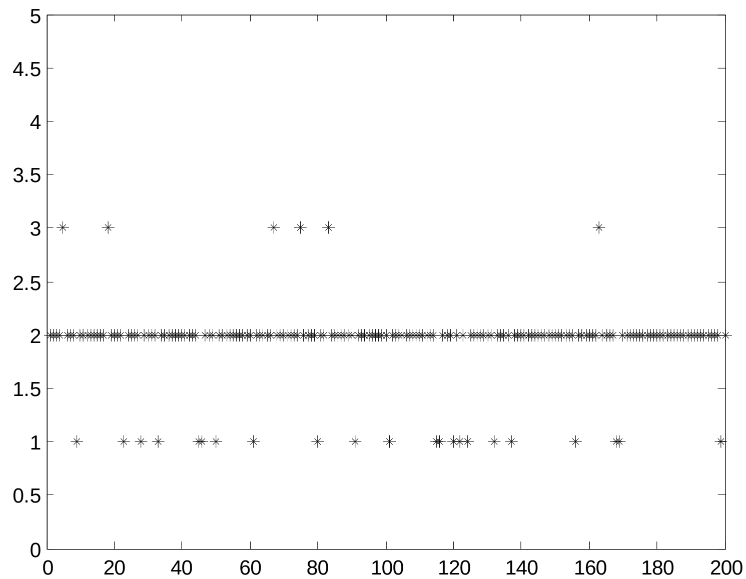


Sensor 2

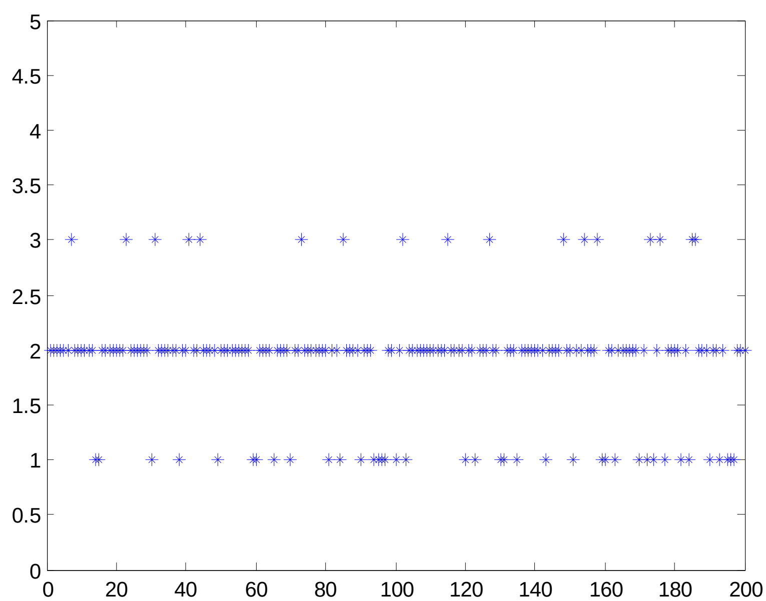


As you can see that it is expected that sensor reading are noisy with different amounts. Now, If the decision is left for each sensor to be done independently, below figure shows the decision at every time instant with ‘1’ being no fault, ‘2’ fault 1 (correct), ‘3’ fault 2, ‘4’ the ‘don’t know’.

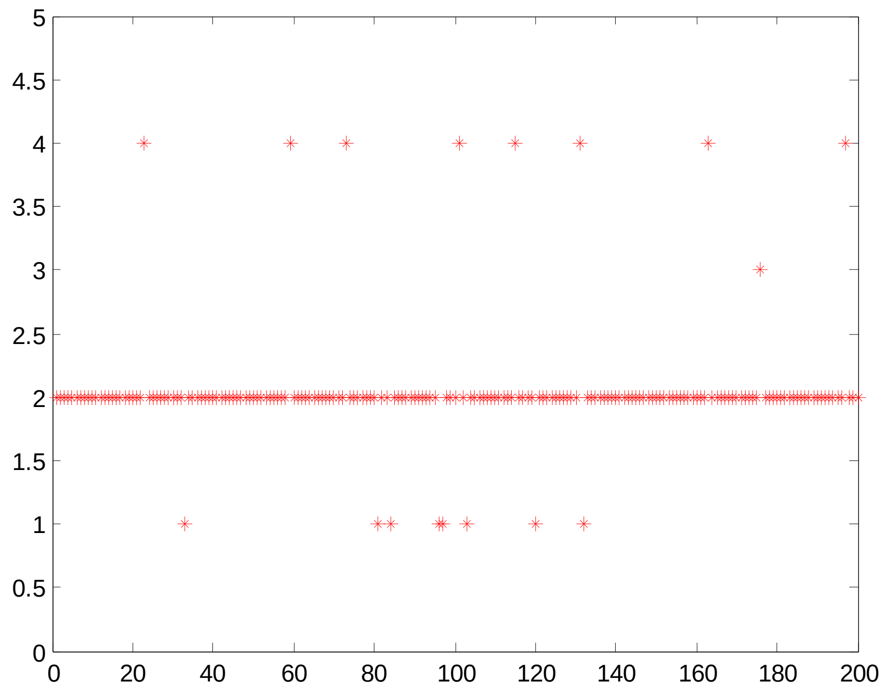
Sensor 1



Sensor 2



However, in combining the two sensors with eq. (4) and the computed measures m1(.) and m2(.), the decision is show in figure below.



With that same fault classes. You can see that more accuracy is shown in identifying the correct fault. Also, the 'don't know' class have been introduced unlike the sensors alone!

VI. Comments & Conclusion

Comments:

- Concern is about the computational complexity. Imagine if 3, 4 … classes, and/or if more sensors. Observe eq. (3).
- However, the 'don't know' class gave us good information rather than the assumption in the normal classical probability.
- A Future Direction of the work is to: investigate the algorithmic view of the problem to improve its performance.
- Another direction is to study the different effects of different noises.

Conceptually, the Dempster-Shafer Theory outperforms the classical Probability Theory. However, mathematically and practically, it is yet to be considered in mainstream works.

REFERENCES

[1] Al-Ani, A. and M. Deriche. A New Technique for Combining Multiple Classifiers using The Dempster- Shafer Theory of Evidence. Journal of Artificial Intelligence Research 17, p. 333-361, 2002.

[2] D. Koks and S. Challa. An introduction to bayesian and dempster-shafer data fusion. Technical Report DSTO-TR-1436, Defence Science and Tech Org, Edinburgh, Australia, Aug 2003.

[3] J. C. Hoffman and R. R. Murphy, "Comparison of Bayesian and Dempster-Shafer theory for sensing: A practitioner's approach," in SPIE Proc. on Neural and Stochastic Methods in Image and Signal Processing II, vol. 2032, July 1993, pp. 266–279.

[4] Gordon, J., & Shortliffe, E. H. (1984). The Dempster Shafer theory of evidence. In B. G. Buchanan & E. H. Shortliffe (Eds.), Rulebased expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project. Reading: MA: AddisonWesley.

[5] N. Wilson. Algorithms for Dempster-Shafer theory. In J. Kohlas and S. Moral, editors, Algorithms for Uncertainty and Defeasible Reasoning. Kluwer Academic Publishers, 1999.

[6] H. Wu. Sensor Data Fusion for Context-Aware Computing Using Dempster-Shafer Theory. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, December 2003.

[7] X.F. Fan and M.J. Zuo, Fault diagnosis of machines based on D–S evidence theory. Part 2: Application of the improved D–S evidence theory in gearbox fault diagnosis, Pattern Recognition Letters 27 (2006), pp. 377–385

[8] Parikh, C.R., Pont, M.J., Jones, N.B., 2001. Application of Dempster–Shafer theory in condition monitoring applications: a case study. Pattern Recognition Lett. 22 (6–7), 777–785.

[9] B.-S. Yang and K. J. Kim. Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals. Mechanical Systems and Signal Processing, 20:403–420, 2006.

[10] O. Basir and X.Yuan. Engine fault diagnosis based on multi-sensor information fusion using dempster-shafer evidence theory. Information Fusion, 2005.

APPENDIX: MATLAB PROGRAM

```
close all
clear all
clc
flt=4;
sen=2;

x=0.7*ones(1,200);

n1=0.15*(randn(1,200));
n2=0.2*(randn(1,200));

s1=x+n1;
s2=x+n2;

d10=1./abs(0.25-s1);
d11=1./abs(0.75-s1);
d12=1./abs(1.25-s1);
d112=1./abs(1.75-s1);

d1=[d10' d11' d12' d112'];
p1=zeros(size(d1));

for i=1:length(d10)
    p1(i,:)=d1(i,:)./sum(d1(i,:));
end

ds1=zeros(1,length(d10));

for i=1:length(p1)
    ds1(i)=find(p1(i,:)==max(p1(i,:)));
end

d20=1./abs(0.25-s2);
d21=1./abs(0.75-s2);
d22=1./abs(1.25-s2);
d212=1./abs(1.75-s2);

d2=[d20' d21' d22' d212'];
p2=zeros(size(d2));

for i=1:length(d20)
    p2(i,:)=d2(i,:)./sum(d2(i,:));
end

ds2=zeros(1,length(d20));

for i=1:length(p2)
    ds2(i)=find(p2(i,:)==max(p2(i,:)));
end

pf0=p1(:,1).*p2(:,1);
pf1=p1(:,2).*p2(:,2)+p1(:,2).*p2(:,4)+p2(:,2).*p1(:,4);
```

```
pf2=p1(:,3).*p2(:,3)+p1(:,3).*p2(:,4)+p2(:,3).*p1(:,4);
pf12=p1(:,4)+p2(:,4);


pf0n=zeros(size(pf0));
pf1n=zeros(size(pf0));
pf2n=zeros(size(pf0));
pf12n=zeros(size(pf0));
for i=1:length(pf0)
    pfs=(pf0(i)+pf1(i)+pf2(i)+pf12(i));
    pf0n(i)=pf0(i,:)./pfs;
    pf1n(i)=pf1(i,:)./pfs;
    pf2n(i)=pf2(i,:)./pfs;
    pf12n(i)=pf12(i,:)./pfs;
end

pf=[pf0n pf1n pf2n pf12n];

dsf=zeros(1,length(pf0n));



for i=1:length(pf)
    ds(i)=find(pf(i,:)==max(pf(i,:)));
end
```