STATISTICAL SIGNAL PROCESSING

SIGNAL DETECTION
(DETERMINISTIC SIGNALS)

P.R. WHITE

GENERAL INTRODUCTION

BAYES' RULE

BAYES' CRITERIA

LIKELIHOOD RATIO TESTS

GAUSSIAN RANDOM VARIABLES

EXAMPLES

ALTERNATIVE DETECTION CRITERIA

 © ISVR 1999

THREE RELATED PROBLEMS

Three different forms of task:

Signal Detection: Is a signal, of a specified type, present or absent in the measured data?

Signal Classification: Does a signal belong to class 1, class 2,…,class *N*?

Estimation: What is the value of a parameter of a model describing the data?

A common framework for tackling these three problems is provided by hypothesis testing.

Detection is a binary hypothesis test, *i.e.* choosing between two competing hypotheses, H_0 – signal absent, H_1 – signal present.

Classification is a *N*-ary hypothesis test, *i.e.* choosing between *N* competing hypotheses, H_1 – signal belongs to class 1, H_2 – signal belongs to class 2, …, H_N – signal belongs to class *N*.

Estimation can be regarded as testing between an infinite number of hypotheses, such that H_θ - the model parameter is θ.

 © ISVR 1999

Background Material:

BAYES' RULE FOR CONDITIONAL PROBABILITIES

Consider two events *A* and *B*.

The probability of these two events occurring, together, is denoted

$$P(A,B)$$

e.g. The probability that some one has a height between 1.8m and 1.9m and weighs between 70kg and 80kg.

The conditional probability defines the probability of *A* given that *B* has occurred and is denoted:

$$P(A|B)$$

e.g. If I know that some one has a weight between 70kg and 80kg what is the probability that their height is between 1.8m and 1.9m.

Bayes' rule states that:

$$P(A|B) = \frac{P(A,B)}{P(B)}, \quad P(B|A) = \frac{P(A,B)}{P(A)}$$

Eliminating $P(A,B)$ from the two equations above gives

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

 © ISVR 1999

SIGNAL DETECTION

From the measured data $\underline{x}$, one has to decide between the two hypotheses:

H_0: Signal absent

H_1: Signal present

Regardless of how this is achieved there are 4 possible outcomes:

1. Detector says the signal is present, *i.e.* H_1 is chosen, when the signal is actually there. (Good, detection)
2. Detector says the signal is absent, *i.e.* H_0 is chosen, when the signal is actually there. (Bad, missed detection)
3. Detector says the signal is present, *i.e.* H_1 is chosen, when in reality the signal isn't there at all. (Bad, false alarm)
4. Detector says the signal is absent, *i.e.* H_0 is chosen, when the signal really isn't there. (Good)

Each of these events can be ascribed a probability. The notation we shall use is of the following form:

> The probability of outcome 2, for example, is denoted $P(D_0|H_1)$ this is the probability of deciding upon H_0 (hence the notation D_0), given that H_1 is true.

 © ISVR 1999

DESIGN CRITERIA

When constructing a detection system there are several goals one may aim to achieve, for example

"To maximise the probability of detection, $P(D_1|H_1)$".

The detector which does this is one which ignores the data (!) and always chooses H_1. This is evidently useless.

Or

"To minimise the false alarms, $P(D_1|H_0)$".

The detector that achieves this also ignores the data and always chooses H_0. This is equally useless.

To obtain reasonable detector designs one has to consider a combination of probabilities.

Note that if H_0 is true, the detector must select either H_0 or H_1, so that the two probabilities $P(D_0|H_0)$ and $P(D_1|H_0)$ are related via:

$$P(D_0|H_0)+P(D_1|H_0)=1 \qquad (1a)$$

Similarly

$$P(D_0|H_1)+P(D_1|H_1)=1 \qquad (1b)$$

This means that whilst we have defined 4 probabilities there are only 2 that can be freely selected.

 © ISVR 1999

BAYES CRITERION FOR SIGNAL DETECTION

One approach to designing a detector is to assign costs to each of the 4 possible detection outcomes.

These costs we denote C_{ij} which defines the cost associated with selecting the hypothesis H_i when H_j is true.

The *risk*, *R*, defined as the average total cost, is:

$$R=C_{00}P(D_0,H_0)+C_{01}P(D_0,H_1)+C_{10}P(D_1,H_0)+C_{11}P(D_1,H_1)$$

where, for example, $P(D_0,H_1)$ denotes the **joint** probability of the detector saying there is no signal and there being a signal[1].

Each of these joint probabilities can be written in terms of conditional probabilities using Bayes' rule, for example,

$$P(D_0,H_1)=P(D_0|H_1)P(H_1)$$

Thus the risk becomes

$$R=C_{00}P(H_0)P(D_0|H_0)+C_{01}P(H_1)P(D_0|H_1)+C_{10}P(H_0)P(D_1|H_0)$$
$$+C_{11}P(H_1)P(D_1|H_1)$$

[1] Note the difference between this and the conditional probability $P(D_0|H_1)$ which is the probability assuming that H_1 is true.

 © ISVR 1999

BAYES CRITERION FOR SIGNAL DETECTION (CONT'D)

Using equations (1) we can substitute for $P(D_1|H_0)$ and $P(D_1|H_1)$ into the expression for the risk, to obtain:

$$R=C_{10}P(H_0)+C_{11}P(H_1)-\{C_{10}-C_{00}\}P(H_0)P(D_0|H_0)+\{C_{01}-C_{11}\}P(H_1)P(D_0|H_1)$$

This depends upon the 2 probabilities $P(D_0|H_0)$ and $P(D_0|H_1)$, which are the probabilities of saying a signal is not there, when isn't, and saying a signal is not there, when it is.

The first 2 terms in R are fixed. Selecting the detection scheme can only affect the last 2 terms.

The costs, C_{ij}, are fixed and the probabilities $P(H_0)$ and $P(H_1)$ are beyond the control of the system designer. To minimise the risk one must minimise a function of the form:

$$A_1P(D_0|H_1)-A_2P(D_0|H_0)$$

where the constants A_1 and A_2 are given by:

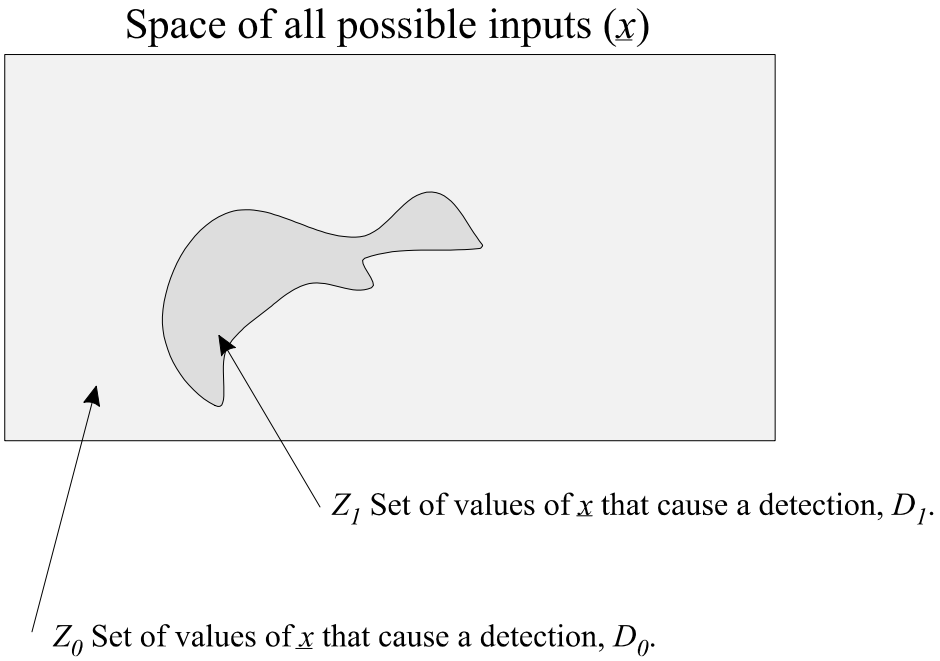
$$A_1=\{C_{10}-C_{00}\}P(H_0)$$

$$A_2=\{C_{01}-C_{11}\}P(H_1)$$

 © ISVR 1999

CONSTRUCTING THE DETECTOR

The probability $P(D_0|H_0)$ can be evaluated by integration of $p(\underline{x}|H_0)$ over the region Z_0, where Z_0 is depicted below:



$p(\underline{x}|H_0)$ is the probability density function for the data, $\underline{x}$, assuming that H_0 is true. So that:

$$P\left(D_0 \mid H_0\right) = \int_{Z_0} p\left(\underline{x} \mid H_0\right) d\underline{x}$$

Similarly

$$P\left(D_0 \mid H_1\right) = \int_{Z_0} p\left(\underline{x} \mid H_1\right) d\underline{x}$$

Allowing one to write:

$$A_1 P(D_0|H_1) - A_2 P(D_0|H_0) = \int_{Z_0} A_1 p\left(\underline{x} \mid H_0\right) - A_2 p\left(\underline{x} \mid H_1\right) d\underline{x}$$

 © ISVR 1999

LIKELIHOOD RATIO TESTS

In the integral:

$$\int_{Z_0} A_1 p\left(\underline{x} \mid H_0\right) - A_2 p\left(\underline{x} \mid H_1\right) d\underline{x}$$

the system designer can only control the region Z_0, *i.e.* which data sets result in the detector saying no signal is present.

Assuming A_1 and A_2 are positive[2], then this integral is minimised when Z_0 is selected so that the integrand is negative, *i.e.*

$$A_1 p\left(\underline{x} \mid H_0\right) < A_2 p\left(\underline{x} \mid H_1\right)$$

or when

$$\frac{p\left(\underline{x} \mid H_1\right)}{p\left(\underline{x} \mid H_0\right)} > \frac{A_1}{A_2} = \frac{\left\{C_{10}\text{-}C_{00}\right\} P\left(H_0\right)}{\left\{C_{01}\text{-}C_{11}\right\} P\left(H_1\right)}$$

This gives us a criterion for detection. If the data is such that

$$\Lambda\left(\underline{x}\right) = \frac{p\left(\underline{x} \mid H_1\right)}{p\left(\underline{x} \mid H_0\right)} > \lambda$$

where λ is the threshold, defined above, then a detection is made.

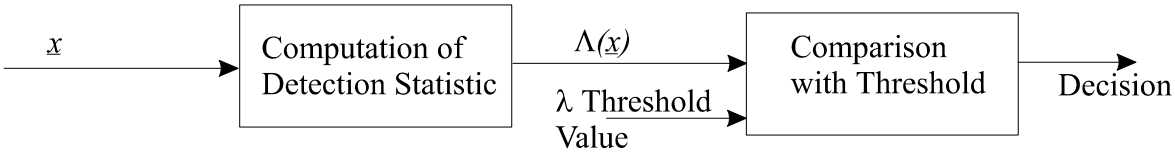
Such a scheme is termed a Likelihood Ratio (LR) test.

[2] This implies that $C_{10} > C_{00}$ and $C_{01} > C_{11}$, *i.e.* that the cost of an error is greater than the cost of a success.

 © ISVR 1999

COMMENTS ON LR TESTS

General structure of a LR based detector:



A LR based detector requires two design elements:

1) The quantity $\Lambda(\underline{x})$ must be computed. This is a positive scalar value, being the ratio of 2 probability density functions (pdfs), referred to as the *detection statistic*. The LR is the optimal detection statistic for a wide range of problems (if not all).
2) The threshold level must be selected. There are several strategies which maybe adopted to achieve this. The value A_1/A_2, based on the Bayesian cost is only one option.

In order to compute the LR one must know the pdf of the noise only signals ($p(\underline{x}|H_0)$) and the signal plus noise pdf, ($p(\underline{x}|H_1)$).

One can apply a monotonic function to both sides of the LR test without changing its effect. For example:

$$\log\{\Lambda(\underline{x})\} = \log\left\{\frac{p(\underline{x}\,|\,H_1)}{p(\underline{x}\,|\,H_0)}\right\} = \log\{p(\underline{x}\,|\,H_1)\} - \log\{p(\underline{x}\,|\,H_0)\} > \log\{\lambda\}$$

 © ISVR 1999

LIKELIHOODS

The quantities $p(\underline{x}|H_0)$ and $p(\underline{x}|H_1)$ are referred to as *likelihoods*.

Hence the name of the LR test, since its test statistic is formed from the ratio of 2 likelihoods.

Note that $p(\underline{x}|H_0)$ is a conditional pdf. Whilst one would like to think of it as describing the probability of observing the data, under the assumption of H_0, strictly speaking this is not valid.

Assuming $\underline{x}$ is a continuous random variable then the probability of observing any $\underline{x}$ is zero!

It is to side-step such technical inaccuracies the term ‘likelihood’ is introduced.

Likelihoods also play a central role in the area of signal estimation.

 © ISVR 1999

ASSESSMENT OF A DETECTOR'S PERFORMANCE

The most commonly used method for measuring a detector's performance is via the *Receiver Operating Characteristic (ROC) curves*.

The ROC curves are a performance measure that is insensitive to the threshold level used and so are a measure of the quality of the test statistic, $\Lambda(\underline{x})$.

The ROC curves are constructed by computing the probabilities of detection and probability of false alarm.

For a given threshold level the probability of detection, P_d, is given by

$$P_d = \int_{\lambda}^{\infty} p(\Lambda \mid H_1) d\Lambda$$

and the probability of false alarm, P_f, by

$$P_f = \int_{\lambda}^{\infty} p(\Lambda \mid H_0) d\Lambda$$

where $p(\Lambda|H_k)$ is the pdf of the detection statistic under assumption H_k.

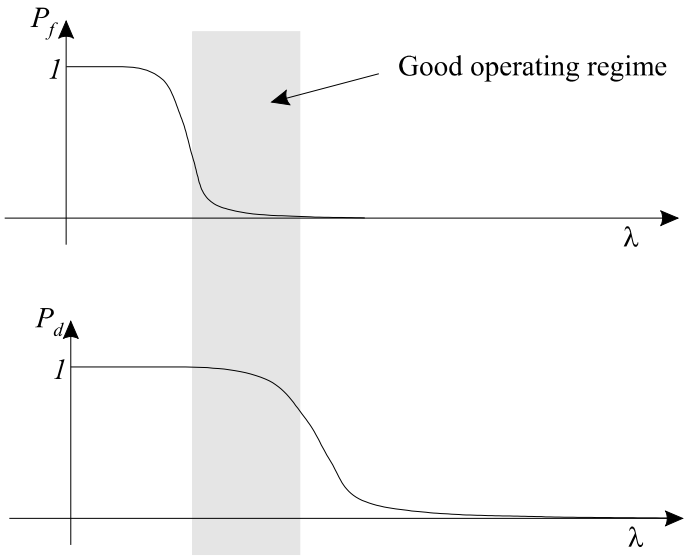
Clearly P_f and P_d depend upon the threshold λ.

If $\lambda=0$ then $P_d = P_f = 1$, whilst if $\lambda=\infty$ then $P_d = P_f = 0$.

 © ISVR 1999

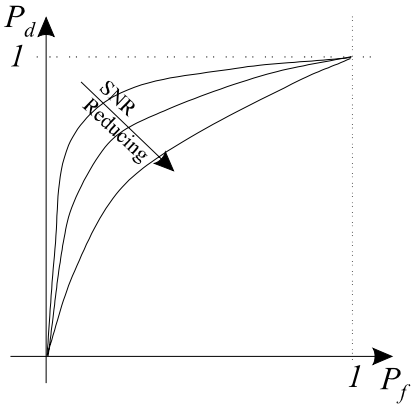
RECEIVER OPERATING CHARACTERISTICS

Below are typical plots of P_f and P_d as functions of λ.



For values of λ in the shaded region the detector has a high probability of detection and a low probability of false alarm.

The ROC curve is obtained by amalgamating the 2 above plots into a single curve, by plotting P_d against P_f as λ is varied. This is usually repeated for a range of SNRs.



 © ISVR 1999

Background Material:

GAUSSIAN RANDOM VARIABLES

If a single variable has a Gaussian (or Normal) distribution, then its pdf is given by:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

where μ is the mean and σ^2 is the variance.

The joint pdf of N independent Gaussian random variables, with a common mean and variance is:

$$p(\underline{x}) = \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x(n)-\mu)^2/(2\sigma^2)} = \frac{1}{\sigma^N (2\pi)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N} (x(n)-\mu)^2}$$

where $\underline{x}$ is a column vector containing the N random variables $x(n)$.

The joint pdf of N dependent, zero mean, Gaussian random variables is

$$p(\underline{x}) = \frac{1}{|R|^{1/2} (2\pi)^{N/2}} e^{-\underline{x}^t R^{-1} \underline{x}/2}$$

where R is the correlation matrix, defined as:

$$R = E[\underline{x}\underline{x}^t]$$

 © ISVR 1999

EXAMPLE 1

Consider the problem of trying to detect a non-zero mean from a single sample. So that

$$x(n)=w(n) \qquad \{\text{Under } H_0\}$$
$$x(n)=A+w(n) \qquad \{\text{Under } H_1\}$$

where $x(n)$[3] is the measured signal, A is the d.c. offset and $w(n)$ is an additive Gaussian noise.

Under H_0 the signal's pdf (likelihood function) is:

$$p\left(x(n)\,|\,H_0\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-x(n)^2/\left(2\sigma^2\right)}$$

Under H_1 the likelihood is

$$p\left(x(n)\,|\,H_1\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x(n)-A)^2/\left(2\sigma^2\right)}$$

The test statistic is

$$\Lambda\left(\underline{x}\right)=e^{-\left(-2Ax(n)+A^2\right)/\left(2\sigma^2\right)}=e^{-A\left(-x(n)+A/2\right)/\left(\sigma^2\right)}$$

The log of the LR test is

$$\frac{A}{\sigma^2}\left\{x(n)-A/2\right\}>log\left\{\lambda\right\}$$

or

$$x(n)>\frac{\sigma^2}{A}log\left\{\lambda\right\}+\frac{A}{2}$$

[3] The argument n is introduced in anticipation of the next example.

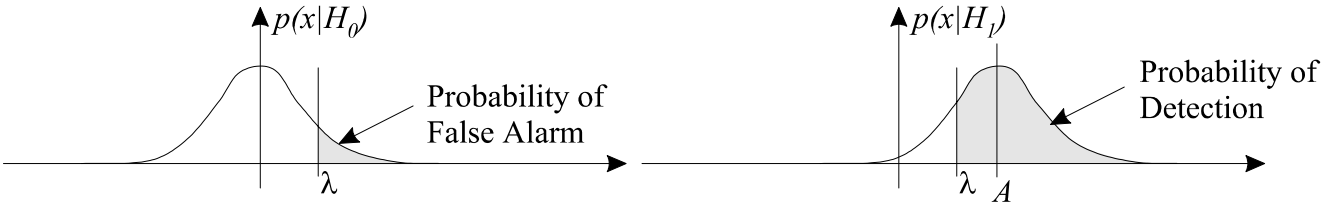
 © ISVR 1999

DISCUSSION OF EXAMPLE 1

This LR test involves simply comparing the data value, $x(n)$, with a threshold.

The threshold is defined as

$$\frac{\sigma^2}{A} log\{\lambda\} + \frac{A}{2}, \text{ where, } \lambda = \frac{\{C_{10}\text{-}C_{00}\} P(H_0)}{\{C_{01}\text{-}C_{11}\} P(H_1)}$$



If one assumes that the signal is equally likely to be present as absent, *i.e.* $P(H_0)=P(H_1)=0.5$ and the costs are $C_{01}=C_{10}$ and $C_{00}=C_{11}$, so that $\lambda=1$, then the threshold is $A/2$.

The threshold $A/2$ can be viewed as the one which is minimally committal, favouring neither H_0 nor H_1.

The additional term $\frac{\sigma^2}{A} log\{\lambda\}$ can be viewed as a bias away from $A/2$, which reflects either prior information about the probabilities $P(H_0)$ and $P(H_1)$ or weighting introduced to reflect the increased importance of, say, false alarms.

 © ISVR 1999

EXAMPLE 2

Now consider the case where one has *N* measurements from the data as defined in Example 1. Now

$$p\left(\underline{x}\,|\,H_0\right)=\frac{1}{\sigma^N\left(2\pi\right)^{N/2}}e^{-\sum_{n=1}^{N}x(n)^2/\left(2\sigma^2\right)}$$

and

$$p\left(\underline{x}\,|\,H_1\right)=\frac{1}{\sigma^N\left(2\pi\right)^{N/2}}e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-A\right)^2}$$

The log of the test statistic is thus

$$\Lambda\left(\underline{x}\right)=\frac{1}{2\sigma^2}\left[\sum_{n=1}^{N}x(n)^2-\sum_{n=1}^{N}\left(x(n)-A\right)^2\right]$$

$$\frac{1}{2\sigma^2}\left[2A\sum_{n=1}^{N}x(n)-NA^2\right]=\frac{NA}{\sigma^2}\left[\frac{1}{N}\sum_{n=1}^{N}x(n)-\frac{A}{2}\right]$$

So to conduct a LR test one performs the comparison:

$$\frac{1}{N}\sum_{n=1}^{N}x(n)>\frac{\sigma^2}{AN}log\left\{\lambda\right\}+\frac{A}{2}$$

 © ISVR 1999

DISCUSSION OF EXAMPLE 2

Observations:

Setting N=1 for this example gives the same result as obtained in example 1.

The threshold is the same threshold as for example 1, but with σ^2 replaced by σ^2/N.

This can be justified by realising that the average of N Gaussian random variables is distributed as a Gaussian random variable, with a variance of σ^2/N.

The additional data values are incorporated in the test by forming their average.

The threshold is set at $A/2$ if one is minimally committal.

 © ISVR 1999

EXAMPLE 3

Consider the same problem as Example 2, but now assume that the N measurements are independent random variables drawn from a Laplacian distribution:

$$p\left(x\left(n\right)\right)=\frac{1}{2\beta}e^{-\left|x(n)\right|/\beta}$$

So that

$$p\left(\underline{x}\,|\,H_0\right)=\frac{1}{\left(2\beta\right)^N}e^{-\sum_{n=1}^{N}\left|x(n)\right|/\beta} \quad \text{and} \quad p\left(\underline{x}\,|\,H_1\right)=\frac{1}{\left(2\beta\right)^N}e^{-\sum_{n=1}^{N}\left|x(n)-A\right|/\beta}$$

Giving the log of the test statistic as

$$\log\left\{\Lambda\left(\underline{x}\right)\right\}=\frac{1}{\beta}\left\{\sum_{n=1}^{N}\left|x(n)\right|-\sum_{n=1}^{N}\left|x(n)-A\right|\right\}$$

Thus the LR test is

$$\left\{\sum_{n=1}^{N}\left|x(n)\right|-\sum_{n=1}^{N}\left|x(n)-A\right|\right\}>\beta\log\left\{\lambda\right\}$$

 © ISVR 1999

EXAMPLE 4 : DETERMINISTIC SIGNALS

Consider the detection problem with the 2 hypotheses

$$x(n)=w(n) \qquad \{H_0\}$$
$$x(n)=s(n)+w(n) \quad \{H_1\}$$

where a set of N measurements are available, the noise is independent and Gaussian and the signal $s(n)$ is deterministic (contains no random elements and is known *a priori*).

$p(\underline{x}|H_0)$ is a given in example 2 and

$$p\left(\underline{x}\,|\,H_1\right)=\frac{1}{\sigma^N\left(2\pi\right)^{N/2}}e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(x(n)-s(n)\right)^2}$$

The signal $s(n)$ can be viewed as a time varying mean.

The log of the test statistic is:

$$\frac{1}{2\sigma^2}\left[\sum_{n=1}^{N}x(n)^2-\sum_{n=1}^{N}\left(x(n)-s(n)\right)^2\right]=\frac{1}{2\sigma^2}\left[2\sum_{n=1}^{N}x(n)s(n)-\sum_{n=1}^{N}s(n)^2\right]$$

So the LR test can be written:

$$\sum_{n=1}^{N}x(n)s(n)>\sigma^2\log\{\lambda\}+\frac{1}{2}\sum_{n=1}^{N}s(n)^2$$

 © ISVR 1999

DISCUSSION OF EXAMPLE 4

The test statistic in this example, $\sum_{n=1}^{N} x(n)s(n)$, can be regarded as an estimate of the cross-correlation between $x(n)$ and $s(n)$.

In applications such as radar and sonar $x(n)$ is an incoming signal of much longer duration than $s(n)$ and one seeks to apply the detection algorithm to each N point window of $x(n)$, *i.e.* to compute

$$\lambda(m) = \sum_{n=1}^{N} s(n)x(m+n) = \sum_{k=m+1}^{m+N} s(k-m)x(k)$$

Thus generating a time dependent detection statistic.

By noting that

$$s(m) * x(m) = \sum_{k} s(m-k)x(k)$$

one can see that

$$\lambda(m) = s(-m) * x(m)$$

i.e. the test statistic can be computed by convolving the input by a time reversed version of $s(n)$ or equivalently filter $x(n)$ using an FIR filter whose taps are the time reversed version of $s(n)$.

Such a detector is referred to as a *matched filter* or a *replica correlator* and is the basis of many systems used in practice.

 © ISVR 1999

MINIMUM PROBABILITY OF ERROR DETECTORS

Discussions so far have centred on how one determines the test statistic for use in a detection system. The only method discussed so far for selecting the threshold has been the Bayes' criterion.

One simplification to the Bayes' approach, widely used in communications systems, is to assume that the cost of a correct decision is zero, *i.e.* $C_{00}=C_{11}=0$ and that the cost of each form of error is the same, *i.e.* $C_{01}=C_{10}=\gamma$.

The value of γ is not important and without loss of generality we can assume $\gamma=1$.

The risk becomes:

$$R=P(H_1)P(D_0|H_1)+P(H_0)P(D_1|H_0)$$

and the LR test is:

$$\frac{p(\underline{x}|H_1)}{p(\underline{x}|H_0)} > \frac{P(H_0)}{P(H_1)} = \frac{1-P(H_1)}{P(H_1)}$$

This only requires the prior probabilities and since only errors are penalised, the system optimally minimises them.

 © ISVR 1999

NEYMAN-PEARSON CRITERION

In applications such as sonar and radar, assigning *a priori* probabilities and costs is unrealistic.

The Neyman-Pearson criterion allows one to set thresholds without recourse these quantities.

The philosophy is to choose a false alarm rate, to fix it, and then to maximise the probability of detection.

It can be shown that adopting this philosophy results in a LR test in which the threshold, λ, is determined by the false alarm rate, namely, by the solution of

$$\int_{\Lambda=\lambda}^{\infty} p\left(\Lambda\left(\underline{x}\right)|H_0\right)d\Lambda = \alpha$$

where α is the desired false alarm rate and $p\left(\Lambda\left(\underline{x}\right)|H_0\right)$ is the pdf of the test statistic under H_0.

 © ISVR 1999