Advanced Signal Processing:

The fundamentals of detection theory

Index of contents:

Advanced Signal Processing: The fundamentals of detection theory

1 Problem Statements

The technical development is a fast and continuous process. Special new Technologies step in our daily life, video mobile phoning, positioning systems, wireless networks and so on.
All these are based on the exchange of information and as we know information is transmitted by numerous Signals.
"Modern detection theory is fundamental to the design of electronic signal processing systems for decision making and information extraction."
All these systems share the same goal of being able to decide when an event interest occurs and then to determine more information about that event. There are two components: first the detection and then the decision. The information extraction is not subject to this topic.

These components will be described by mathematical formulation and techniques. You often cannot predict whether a signal is correctly transmitted or not. Many factors are in such a system that will interfere with the real signal. Therefore we will approximate if a signal can be transmitted and further more received. For this we will need the statistic and its distribution functions. With the help of the statistic we will find out what a good detector is and which is not.

The decision theory we find out something about the basic statistical groundwork for the design of detectors of signals in noise. The approaches follow directly from the theory of hypothesis testing.

2 Detection Theory in Signal Processing

As mentioned before detection theory is fundamental to electronic signal processing systems. These systems include:

1. Radar
2. Communications
3. Speech
4. Sonar
5. Image Processing
6. Biomedicine
7. Control
8. Seismology

All these systems share a common goal of being able to decide when an event of interest occurs. To do so we will need the decision and detection theory. To illustrate the problems of detection the first three systems will be described.

2.1 The radar System

In radar we are interested in determining the presence or absence of an approaching aircraft. For this we transmit an electromagnetic pulse, which if reflected by a large moving object, that will indicate the presence of an aircraft. If an aircraft is present, the received waveform will consist of the reflected pulse and noise due to ambient radiation and the receiver electronics. Now it is the function of the signal processor to decide whether the received waveform consists of noise only or a signal in noise. When an echo is present, the character of the received waveform will be somewhat different because of propagation loss and interaction of reflection. If it is detected we are interested in the aircrafts bearings, speed and range. So after the first task of the signal processing system the second task will start: the information extraction which is called estimation theory.

2.2 A digital information System

An example for a digital information system is the binary phase shift keyed system (BPSK). It is used to communicate the output of a digital data source that emits a "0" or "1". The data bit is first modulated, then transmitted, and at a receiver, demodulated and detected. The modulator converts a "0" into a waveform $s0(t) = \cos 2\pi F_0 t$ and a "1" in $s1(t) = \cos(2\pi F_0 t+\pi) = -\cos 2\pi F_0 t$ to allow transmission through a bandpass channel whose center frequency is F_0.

Now there is the function of the detector to decide between the two possibilities, although now we always have a signal present, the question is which signal. The signal is usually distorted and corrupted, therefore we'd be in need of filters.

2.3 Speech recognition

Here we wish to determine which word was spoken from among a group of possible words. A simple example is to discern among the digits "0", "1" to “9”. To recognize a spoken digit using a digital computer we would need to match the spoken digit with some stored digit. The big problem is that the waveform changes lightly for each utterance of the same word. We think of noise but this is the natural variability of speech.

2.4 Central Problem

In all of these systems , we are faced with the problem of making a decision based on a continuous-time waveform. Modern-day signal processing systems utilize digital computers to sample the continuous-time waveform and store the samples. So as a result, we have the equivalent problem of making a decision based on a discrete time waveform or data set. Mathematically, we assume the N-point data set {x[0], x[1], x[2],… x[N-1]} is available. We first form a function of the data or T(x[0], x[1], x[2],… x[N-1]) and then make a decision based on its value. Determining the function T and mapping it into a decision is the central problem addressed in detection theory.
The future trend is based on discrete-time signals or sequences and digital circuitry. With this transition the detection problem has evolved into one of making a decision based on the observation of a time series, which is just a discrete time process. Therefore our problem has now evolved into decision making based on data, which is subject to the statistical hypothesis testing.

3 The detection Problem

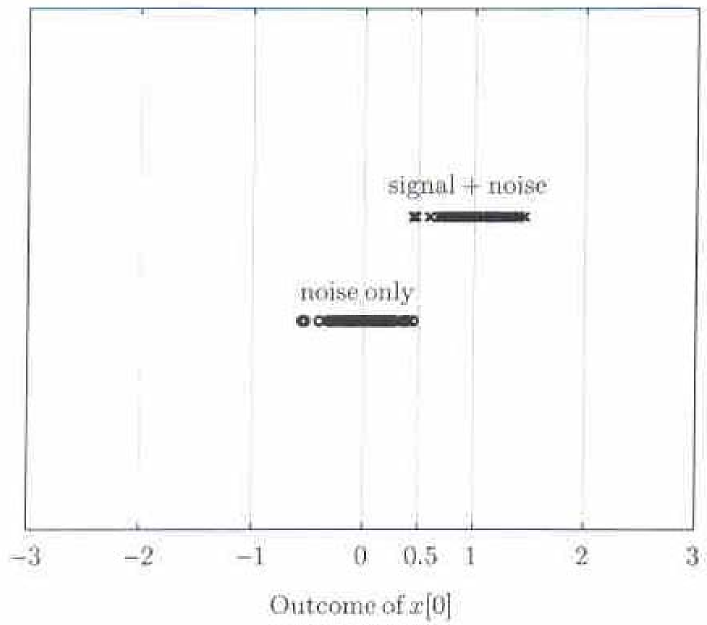
The simplest detection problem is to decide whether a signal is present, which, as always, is embedded in noise or if only noise is present. (Radar example)

Since we wish to decide between two possible hypotheses, signal and noise present vs. noise only present, we term this the *binary hypothesis testing problem.* The goal is to use the data as efficiently as possible in making our decision. This will be more general if we take a look at the communication problem. There we have to decide which of two possible signals was transmitted.
On the other hand it occurs that we wish to decide among more than two hypotheses like in speech communication systems. Such problems for example, where we have to determine which digit among ten possible ones was spoken, are referred to as the *multiple hypothesis testing problem*.
All these problems are characterized by the need to decide among two or more possible hypotheses based on observed data sets.

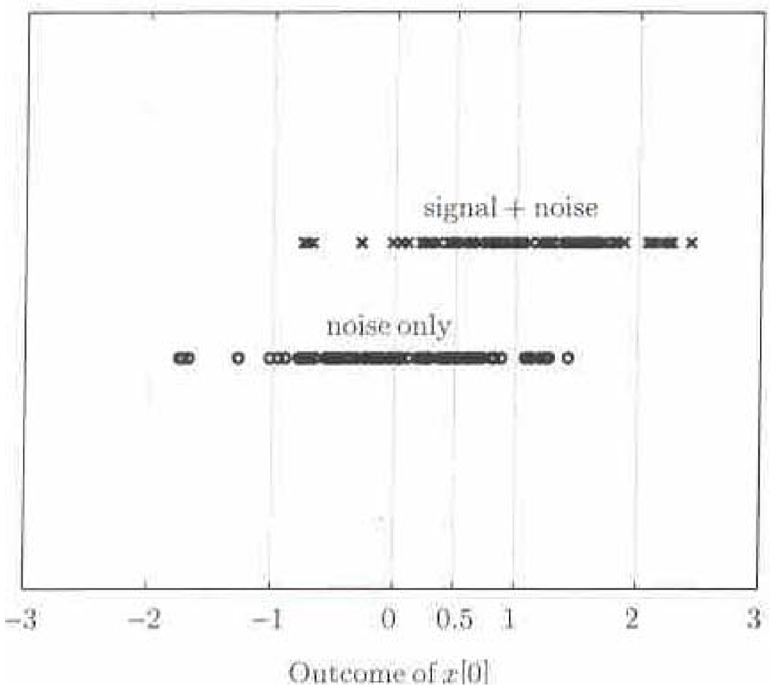
4 The Mathematical Detection Problem

We consider the detection of a DC level of Amplitude A=1 embedded in white Gaussian Noise $\omega[n]$ with variance σ^2. To simplify the discussion we assume that only one sample is available on which to base the decision.
So now we wish do decide between two hypotheses:
H0: $x[0] = \omega[n]$ und H1: $x[0] = 1 + \omega[n]$. Since the noise is assumed to have zero mean, me might decide that a signal is present if $x[0]>½$ and noise only is present if $x[0] <½$ since $E(x[0]) = 0$ if noise only is present and $E(x[0]) = 1$ if a signal in noise is present.
We will always be in error if a signal is present and $\omega[n] < -½$ or whenever only noise is present and $\omega[n] > ½$. It is also not possible to make correct decisions all the time but hopefully most of the time. For better understanding we consider what would happen if we repeated the experiment a number of times. For example observe x[0] for 100

realisations of ω[n] when a signal in noise is present and when it is not. You will get typical results for $\sigma^2 = 0.05$ as shown in the Figure:



The "o"'s denote the outcome when no signal is present and the "x"'s when a signal is present. You can see that we may make an incorrect decision but only rarely. If we change $\sigma^2 = 0.5$, then our chances of making an error increase dramatically which is shown in the next figure on the next page:



This is due to the increasing spread of the realisations of ω[n] as σ² increases.

4.1 The Probability Density Function

The PDF of noise is:

$$p(\omega[0]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\omega^2[0]\right)$$

Properties of a PDF:

- shows distribution graphically
- only at continuous Probability distributions

4.1.1 The PDFs for noise only:

PDF for σ²= 0.05:

$$p(x[0]) = \frac{1}{\sqrt{0,1\pi}} \exp\left(-10x^2[0]\right)$$

PDF for σ²= 0.5:

$$p(x[0]) = \frac{1}{\sqrt{\pi}} \exp\left(-x^2[0]\right)$$

4.1.2 The PDFs for Signal in noise:

PDF for σ²= 0.05:

$$p(x[0]) = \frac{1}{\sqrt{0,1\pi}} \exp\left(-10(x[0]-1)^2\right)$$

PDF for $\sigma^2 = 0.5$:

$$p(x[0]) = \frac{1}{\sqrt{\pi}} \exp\left(-(x[0]-1)^2\right)$$

The performance improves as the distant between the PDF's increases or the SNR increases A^2/σ^2

4.2 *Hypothesis*

Detection Performance depends on the discrimination between two hypotheses or PDF's. We model the detection problem as one of choosing between H0 and H1 The PDF's are denoted by p(x[0];H0) and p(x[0],H1) We ask whether x[0] has been generated according to p(x[0];H0) or p(x[0],H1)?
Alternatively more general: p(x[0];A) for A = 1 or 0 which is parameterized by A.
So the detection Problem could be seen as parameter test. We obtain p(x[0];H0) if A = 0 and p(x[0],H1). when A =1. Given the observation x[0] we whish to test if A = 1 or A = 0 or symbolically:

- H0: A = 0
- H1: A = 1

It is also convenient to assign prior Probabilities to the possible occurrences of H0 and H1. This is useful for On Off Keyed Systems where you want to transmit either a ZERO by sending no Pulse or a ONE by sending a Signal with the Amplitude A = 1. In an OOK System we will transmit a steady stream of data bits. Since the data bits are equally likely to be generated by the source, we would expect H0 to be true half the time and H1 the other half. The Hypotheses so appear as random events with the probability ½. When we do so the notation of the PDF will be p(x[0]|H0) and p(x[0]|H1)

4.3 The Gaussian PDF

Is also referred as normal PDF; for a scalar random variable x it is defined for –infinity < x < +infinity
x~N(μ,σ²) → „~" means "is distributed according to"

Cumulative distribution function CDF → μ=0 and σ²=1 → the PDF for CDF is called standard normal PDF
CDF is defined as PHI:

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)dt$$

Right tail Probability → Q(x) 1 – Φ(x)
Defined as:

$$Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)dt$$

4.4 Right Tail Probability

Is also referred as Complementary cumulative distribution function and can be sometimes approximated as

$$Q(x) \approx \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)$$

→ is the Probability of exceeding a given value → approximation useful for x > 4

4.5 CHI – SQUARED PDF

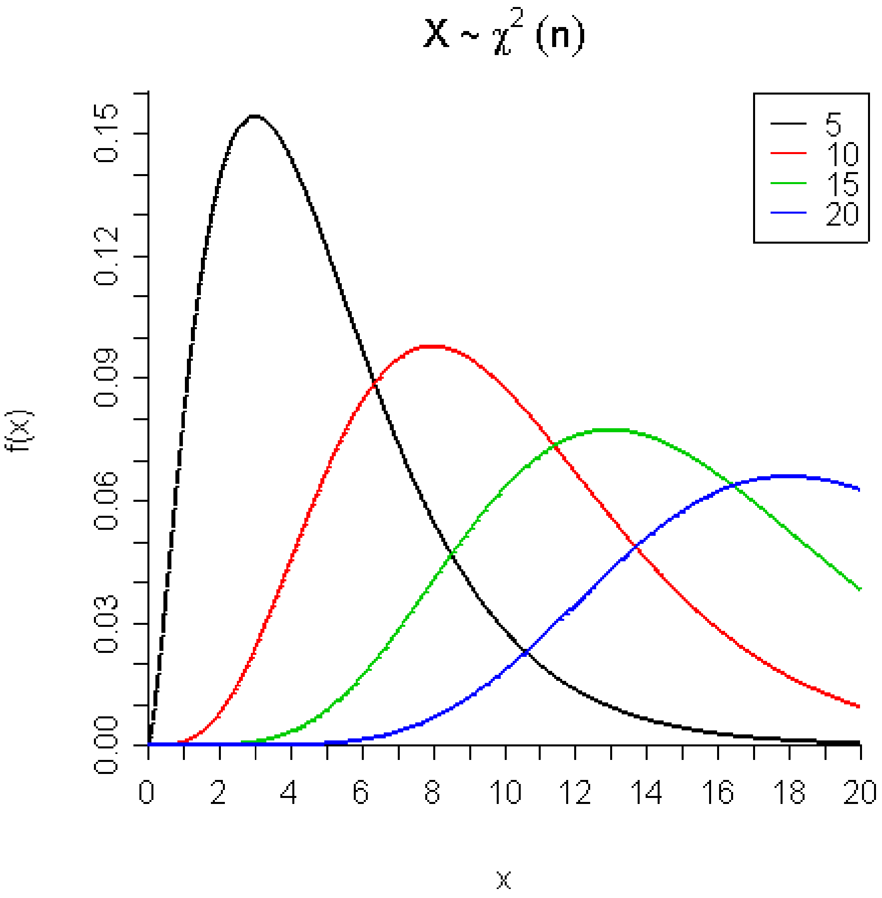
Is the distribution of a sum of $X=Z_1^2+\ldots+Z_n^2$ of n independent Z squared standard normal random variables → or if $Z_k \sim N(0,1)$ for k = 1 to n is independent then $X \sim X^2(n)$.

For n>100 X is approximately normal distributed.

The graph on the next page shows the CHI-Squared distribution for some n

If n is 0 the PDF is infinite at x=0

As n rises the chi squared becomes Gaussian.



A specific case of interest occurs when n=2:

This is referred to as an exponential PDF for x>0

$$p(x) = \frac{1}{2}\exp\left(-\frac{1}{2}x\right)$$

The Right tail probability of a CHI Squared random variable is:

$$Q(x) = \int_x^\infty p(t)dt$$

5 Decision Theory

We lay basic statistical groundwork for the design of detectors of signals in noise we address simple hypothesis testing problem where the PDF of each assumed Hypothesis is known.
The primary approaches to simple hypothesis testing are the classical approach based on NEYMAN PEARSON theorem and the BAYESIAN approach based on minimization of BAYES risk.

SONAR and RADAR systems → NEYMAN PEARSON criterion
Communication and Pattern recognition → BAYES risk

5.1 Neyman-Pearson Theorem (1)

The Detector that maximizes the probability of detection for a given probability of false alarm is the likelihood ratio test. as specified by the NEYMAN PEARSON theorem. The threshold is found from the false alarm constraint
In discussing the NP approach we center our discussion around a simple example of hypothesis testing. We must determine if μ is 0 or 1 based on only a single observation x[0].

binary hypothesis test → choose between two hypotheses

5.1.1 Error Types

On the basis of a single sample it is difficult to determine which PDF generated it.
Reasonable: decide H1 if x[0] > ½ → then x[0] is more likely if H1 true
Then we can say that p(x[0],H1)>p(x[0],H0)
Detector compares the observed value with ½ → threshold
Notation: P(Hi;Hj) → deciding Hi when Hj is true
P(H1;H0)=Pr(x[0]>1/2;H0) → darker area
Failure unavoidable but we should trade them off against each other → change threshold
It is not possible to reduce both error probabilities simultaneously

5.1.2 Probabilities

We choose to constrain P(H1;H0) to a fixed value α, then distinguish between two hypotheses H0 und H1 → we have the signal detection problem.

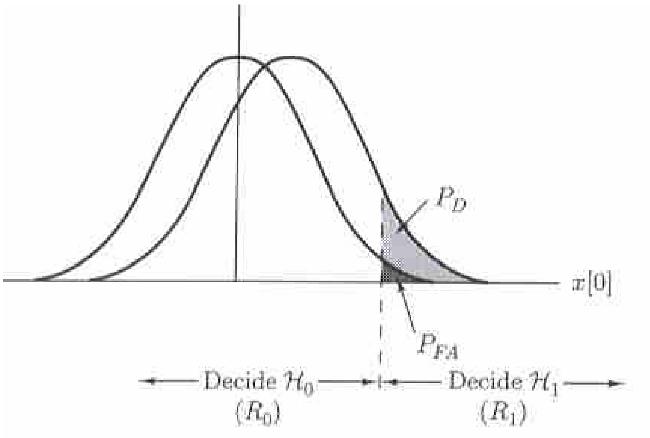
Define a probability of false alarm → decide H1 when H0 is true → Pfa and is small
Now we want to minimize the other error P(H0;H1) or minimize 1-P(H1;H0)=P(H1;H1) → Probability of detection Pd
This setup is termed the NEYMAN PEASON approach to hypothesis testing → maximize Pd and constraint Pfa

$$P_{FA} = P(H_1;H_0) = \Pr\{x[0] > \gamma; H_0\} = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt = Q(\gamma)$$

$$P_D = P(H_1;H_1) = \Pr\{x[0] > \gamma; H_1\} = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(t-1)^2\right) dt = Q(\gamma - 1)$$

The goal of a detector is to decide either H0 or H1 based on an observed data set {x[0]-x[N-1]}

This is a mapping from each possible data set into a decision. The decision regions are shown below:.



R1 is the set of values in R^N that map into a decision H1:

R1={x: decide H1 or reject H0}→ critical region

The set of Points in R^N that map into decision H0 is the complement set of R1:

R0={x: decide H0 or reject H1}

R1 u R0 = R^N

The Pfa becomes then:

$$P_{FA} = \int_{R_1} p(x; H_0)dx = \alpha$$

And the Pd:

$$P_D = \int_{R_1} p(x; H_1)dx$$

Pd is called the *power* of the test and the critical region that attains the maximum power is the *best critical region*
The NP theorem tells us how to choose R1 if we are given p(x;H0), p(x;H1) and alpha

5.2 Neyman-Pearson Theorem (2)

Function L(x) is termed Likelihood ratio since it indicates for each value of x the likelihood of H1 versus the likelihood of H0
The entire test is called the likelihood ratio Test LRT

$$L(x) = \frac{p(x;H1)}{p(x;H0)} > \gamma$$

Where the threshold can be found from:

$$P_{FA} = \int_{\{x;L(x)>\gamma\}} p(x;H0)dx = \alpha$$

5.3 Minimum Probability of Error

In detection Problems one can reasonably assign probabilities to the various hypotheses.
In doing so we expect a prior believe in the likelihood of the hypotheses.
For example: in digital communication the transmission of 0 or 1 is equally likely
Then it is reasonable to assign equal probabilities to H0 (ZERO) and H1 (ONE)
Not possible in SONAR and RADAR
This type of approach, where we assign prior probabilities, is the BAYESIAN approach to hypothesis testing.

With the BAYESIAN paradigm we can define a probability of error Pe

Pe = Pr{decide H0,H1 true} + Pr{decide H1,H0 true}

Pe = P(H0|H1)P(H1) + P(H1|H0)P(H0)

Using the Pe criterion, the two errors are weighted appropriately to yield an overall error measure. Similar to the NP test we compare the conditional likelihood ratio to a threshold.

Equivalently we choose the Hypothesis with the larger conditional likelihood or the one that maximizes p(x|Hi) for i=0,1

This is called the **Maximum Likelihood detector** ML

Decide H1 if:

P(H1|x) > P(H0|x)

With the Bayses rule:

$$p(H_i \mid x) = \frac{p(x \mid H_i) * P(H_i)}{p(x)}$$

we come for p(x) is a constant value to:

p(x|H1)P(H1)> p(x|H0)P(H0)

and finally:

$$\frac{p(x \mid H1)}{p(x \mid H0)} > \frac{P(H0)}{P(H1)} = \gamma$$

The probability of error decreases monotonically with NA²/ σ² which is the deflection coefficient.

$$P_e = Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right)$$

5.4 MAP Detector

We can also choose the hypothesis whose a posteriori (after data are observed) probability is maximum. This detector which minimizes Pe for any prior probability, is termed the maximum a posteriori detector → **MAP**

For equal prior probabilities the MAP detector reduces to the ML detector.

6 Cross-Reference of Statistical Terms for Binary Hypothesis Testing:

Statisticians	**Engineers**
Test statistic (T(x) and threshold (γ))	Detector
Null hypothesis (H0)	Noise only hypothesis
Alternative Hypothesis (H1)	Signal + noise hypothesis
Critical Region	Signal present decision Region
Type I error	False Alarm (FA)
Type II error	Miss (M)
Level of significance (α)	Probability of FA (PFA)
Probability of Type II error (β)	Probability of miss (PM)
Power of test (1- β)	Probability of detection (PD)