Change detection

Niki Serakiotou

ABSTRACT

Detection theory first became famous because of its radar application success, but electrical engineers have since used it in a variety of other problems ranging from detection of failure in a machine or device, to detection of blood vessels in stereographic images of the brain, to detection of targets in automatic driving. In geophysical research it has already been used for parameter estimation. For all its impressive applications detection theory is established on a few simple ideas from hypothesis testing.

INTRODUCTION

In the middle of the ocean it is the wind and distant passing ships that generate the noise known as *background noise* which is Gaussian and almost stationary. In the seas where the traffic becomes heavier, the so called *merchant ship* noise is still Gaussian but less stationary (its variance changes from time to time). An even worse acoustic nuisance comes from a species known as *snapping shrimp*. The sound that a single shrimp makes (with a part of its body) can be heard several miles away, making the noise created by a whole group difficult to ignore — it is also highly non-stationary. These noises bother the submarines that are trying to detect acoustic signals coming from other submarines. To make matters worse, *cracking ice* in the Arctic sea causes high intensity impulsive bursts that are added to the Gaussian background.

At the same time, on land, a geophysicist using the Maximum-Likelihood Deconvolution method deals with a reflectivity sequence that consists of a low-variance, white, Gaussian backscatter noise and a train of occasional spikes (whose amplitude follows a Gaussian distribution while their location is assumed to follow a Bernoulli distribution).

In all of the above examples we have to deal with non-stationary data records — that is some of our data samples can be represented with one model and some with another. We may or may not exactly know what models to use for each part of our data record but we usually need to be able to **detect *when* a change in our data model occurs.**

In this paper I review some well known problems and detection techniques (all of whom can be considered as Electrical Engineering topics). My purpose is to

1. Show how change detection can be considered as a more general problem (instead of assessing each model-change case as a different entity).
2. Revisit hypothesis testing and detection theory since it is a possible alternative to estimation that is widely used in Geophysics (the Maximum Likelihood Deconvolution method that I mentioned above is an example of this possibility, and perhaps it is no coincidence that it was developed by electrical engineers (Mendel, 1986)).

The problem

The problem of change detection can be viewed as a **hypotheses testing** problem. Indeed we are dealing with a record of data $(x_1, \ldots, x_n)$ and we have to decide which of the two hypotheses is true

$H_0 : x_i, i = 1, \ldots, n$ follow the model M_0
or
$H_1 : x_i, i = 1, \ldots, r$ follow the model M_0
while $x_i, i = r+1, \ldots, n$ follow the model M_1

and also **estimate the moment** r at which the change happened (if it did).

At this point, I present a short review on hypothesis testing. The reader may skip this paragraph (if he or she is already familiar with the material) without loss of context.

A reminder on Hypothesis Testing

Suppose that we can find ourselves in either of two situations, but only one can be happening at a time. We then gather some *observations* or *data* and using them as clues we try to *decide* which of the two is actually taking place. Of course, we will have to use a *criterion* to make our decision. This criterion will then lead us to a procedure or algorithm into which we will be feeding our observations and take the decision as the output.

For example, suppose that our observations are n samples of an acoustic signal. For these samples $x_1, \ldots, x_n$ we know that they are *statistically independent* from each other and we want to decide which of the following *hypotheses* is true about them

H_0 : The samples contain only zero mean, Gaussian noise.
or
H_1: The samples contain also a dc part s

This is actually the simple problem of detecting a known, constant signal in Gaussian noise. It can be equivalently expressed as

$H_0 : x_i, i = 1, \ldots, n$ follow $\mathcal{N}(0, \sigma_0)$
$H_1 : x_i, i = 1, \ldots, n$ follow $\mathcal{N}(s, \sigma_0)$

Now if our data record is actually a part of a signal where the dc component is turned on and off at arbitrary moments, then we might want to formulate the two hypotheses somewhat differently — the way we did in the introduction. That is

$H_0 : x_i, i = 1, \ldots, n$ follow $\mathcal{N}(0, \sigma_0)$
$H_1 : x_i, i = 1, \ldots, r$ follow $\mathcal{N}(0, \sigma_0)$ while $x_i, i = r+1, \ldots, n$ follow $\mathcal{N}(s, \sigma_0)$

in which case we will have to estimate the change time r.

There is a number of criteria for the testing available in the literature. *Bayes* criterion tries to minimize the cost of making wrong decisions. *Neyman - Pearson* criterion tries to maximize the probability of detection while keeping small the probability of a false alarm. Both criteria yield the same procedure for the decision making:

- Compute the **Likelihood Ratio** , which is a function of the observations defined as

$$\Lambda(x_1, \ldots, x_n) = \frac{P[x_1, \ldots, x_n / H_1]}{P[x_1, \ldots, x_n / H_0]} \tag{1}$$

- Compare this to a **threshold** T and decide as follows

$$\Lambda(x_1, \ldots, x_n) > T \longrightarrow H_1 \tag{2}$$

$$\Lambda(x_1, \ldots, x_n) < T \longrightarrow H_0 \tag{3}$$

The threshold in both criteria is defined by considerations about the cost of wrong decisions. Figure 1 below shows the procedure for the decision making.

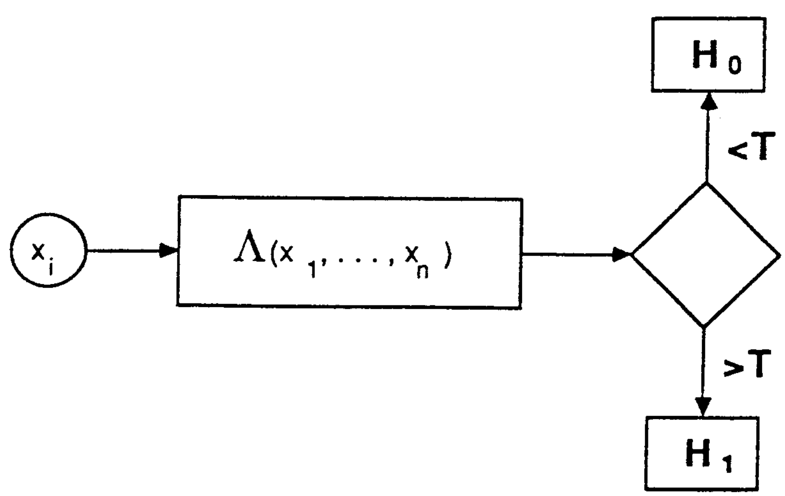

FIG.1 *Bayes* and*Neyman-Pearson* criteria lead to this procedure for deciding whether a change has occured inside the data record $x_1, \ldots, x_n$ that we examine.

Continuing on the Change Detection Problem ...

... we note that when we talk about a single change there are two models involved. That is, we have the "before" model M_0 and the "after" model M_1.

Of course, sometimes more than one type of change is possible; then we have more than two possible models involved and we consider the various possibilities of switching from one model to another.

For the moment consider a single change. As a rule, one of the two models, say the M_0, describes the regular situation, the normal status, and thus it can be assumed completely known.(Otherwise it can be identified with an adaptive algorithm.)

For the second model M_1 one of the following can be true.

1. M_1 is also completely known. In most cases a rather unrealistic assumption.
2. M_1 is known to belong to a known set of possible models. (This is exactly the case when more than one type of change can occur in our data.)
3. A few prior information is available about M_1.
4. Nothing is known about M_1.

The following sections contain specific examples of change detection procedures.

EXAMPLES OF CHANGE DETECTION

1. ALL MODELS ARE KNOWN.

EXAMPLE 1.1: Detection of impulses contaminating Gaussian noise.

The statististical characteristics of the impulses are known. Three algorithms are proposed. The two first are based on the Neyman-Pearson detector; the third one uses the Page-Hinkley detection scheme.

In this example we are examining a sequence $\{ x_i \}$ of data. They consist of zero mean Gaussian noise with (low) variance σ_0 disturbed by sudden and brief intervals of zero mean Gaussian noise with (high) variance σ_1. We want to detect when an impulsive interval occurs and when it ends.

1.1.1 The Neyman-Pearson detector.

We formulate our hypotheses as follows.

$$H_0 : x_i \text{ follows } \mathcal{N}(0, \sigma_0)$$
$$H_1 : x_i \text{ follows } \mathcal{N}(0, \sigma_1) \qquad (4)$$

Using the Neyman-Pearson criterion, we derive the likelihood ratio

$$\Lambda(x_i) = \frac{probability\ that\ x_i\ follows\ \mathcal{N}(0,\sigma_1)}{probability\ that\ x_i\ follows\ \mathcal{N}(0,\sigma_0)} = \frac{f_1(x_i)}{f_0(x_i)} = \frac{e^{\frac{x_i^2}{2\sigma_1^2}}}{e^{\frac{x_i^2}{2\sigma_0^2}}} \tag{5}$$

Then the decision rule is

$$\Lambda(x_i) < T \longrightarrow H_0 \tag{6}$$

$$\Lambda(x_i) > T \longrightarrow H_1 \tag{7}$$

We can equivalently use the logarithm of the ratio $\lambda(x_i) = log\Lambda(x_i)$. Then the rule becomes

$$\lambda(x_i) = x^2 \cdot (constant) < logT \longrightarrow H_0$$

$$\lambda(x_i) = x^2 \cdot (constant) > logT \longrightarrow H_1$$

and finally

$$| x_i | < t \longrightarrow H_0 \tag{8}$$

$$| x_i | > t \longrightarrow H_1 \tag{9}$$

In order to avoid "false alarms", that is deciding that an impulse has occurred though it actually has not, we usually apply some algorithm to provide for "smoother" switchings. So instead of deciding on a single sample basis according to equations (8),(9), we take into account the results in a set of samples that we call a "voting window". The final decision for the x_i sample is then the voting result among preliminary decisions taken for each sample of the window that surrounds x_i (or starts with it). Since smoothing is generally interesting I present two such smoothing algorithms below.

1.1.2 Czarnecki's algorithm.

This algorithm is presented and analyzed by Czarnecki (1983). A window of width $m+1$ samples, is shifted along the data samples. To decide for the i-th sample, we apply the rule given by equations (8),(9) to each of the previous m and each of the following m samples as well as to the i-th sample. If the majority of these samples "vote" for, say, H_1 then we decide that the i-th sample is part of an impulsive interval and so on.

1.1.3 My algorithm.

Here I introduce another "smoothing" algorithm whose theoretical analysis I omit since it is complicated and of little practical value.

We apply the simple decision rule of equations (8), (9) to the i-th sample. If the rule decides for H_0 we simply accept the decision and restart the procedure for the next sample. Otherwise we suspect that this i-th sample might be the beginning of an impulse, so we take a vote in a window starting with x_i and including the following $2m$ samples. If the result of this vote goes for H_0 then we accept that this hypothesis holds for x_i and restart the procedure for the next sample. If the result of the vote in this "i-th" window goes for H_1, then we are pretty sure there is an impulse starting at, or very close to, the i-th

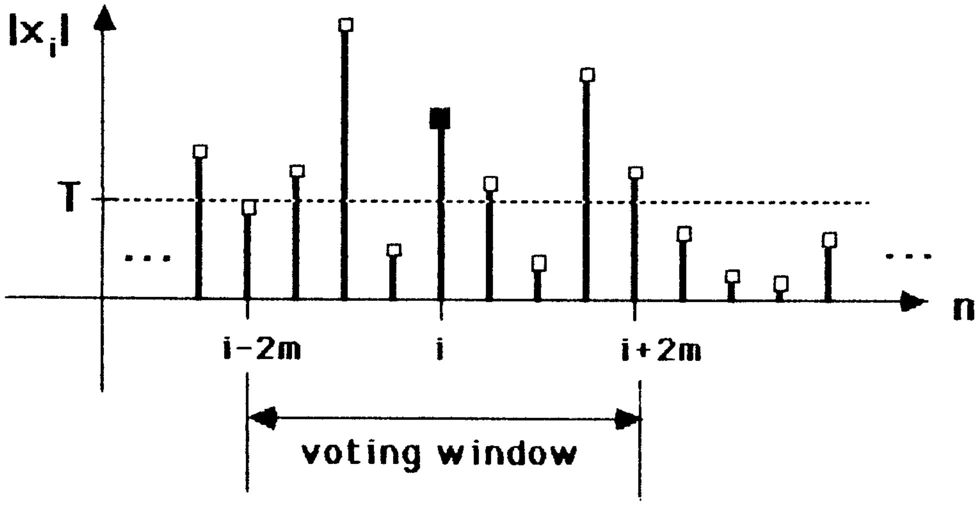

FIG.2 Czarnecki's smoothing algorithm: The decision between H_0 and H_1 for the i-th sample is based on the voting result among preliminary decisions inside the "window" consisting of samples $x_{i-2m}, \ldots, x_{i+2m}$.

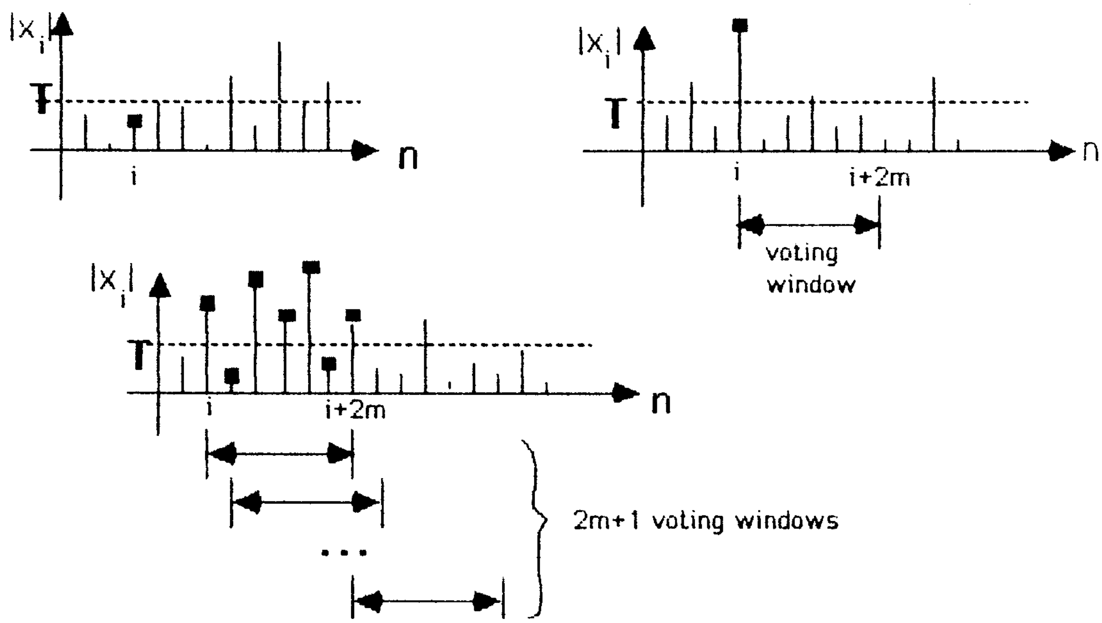

FIG.3 My algorithm: There are three possible ways to decide between H_0 and H_1 for each sample x_i. In (a) we decide immediately for H_0. In (b) we decide for H_0 if the voting result in the window $(x_1, \ldots, x_{i+2m})$ is for H_0, or else we decide for x_i as well as for $x_{i+1}, \ldots, x_{i+2m}$ using the $2m+1$ voting windows of (c).

sample. Some of the following samples must then also belong to the impulse but we have to be careful as to where the impulse ends. For this purpose the decisions for the next $2m$ samples are taken in this way: For the i-th sample the decision is H_1, the voting result in the window starting at the i-th sample. For the $i+j$-th sample, $j = 1, \ldots, 2m$, the decision takes into consideration the voting results in the windows that start at the i-th, $i+1$-th,...,$i+j$-th samples, that is, for the $i+j$-th sample we accept as true the hypothesis that gets the most votes in these $j+1$ number of windows.

1.1.4 Comparison of the two smoothing algorithms.

A classical criterion of the performance of a detection procedure is the *probability of error* that the procedure yields. Probability of error is by definition the sum of two probabilities:

1. probability of a miss, that is the probability that we do not detect an impulse though it has actually occured
2. probability of a false alarm, that is the probability with which our procedure decides that an impulse has occured though it actually has not.

Obviously the names come from the radar literature. For our example I computed the error probability or to be more accurate *the percentage of wrong decisions* that each of the two algorithms made when I applied them on synthetic data.

It turned out that the second algorithm had a better performance than the first for all window widths and for all the thresholds I used. I will not present my results here so as not to distract the reader with technical details.

1.1.5 The Page-Hinkley or cumulative-sum detection scheme.

Instead of formulating our problem as in equation (4) we can equivalently use the following hypotheses

$H_0 : x_i, i = 1, \ldots, n$ follows $\mathcal{N}(0, \sigma_0)$
$H_1 : x_i, i = 1, \ldots, r-1$ follows $\mathcal{N}(0, \sigma_0)$
while $x_i, i = r, \ldots, n$ follows $\mathcal{N}(0, \sigma_1)$

Then the likelihood ratio becomes

$$\begin{aligned} \Lambda &= \frac{P[(x_1, \ldots, x_n)/H_1]}{P[(x_1, \ldots, x_n)/H_0]} = \\ &= \frac{\prod_{i=1}^{r-1} f_0(x_i) \prod_{i=r}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)} = \\ &= \prod_{i=r}^{n} \frac{f_1(x_i)}{f_0(x_i)} = \\ &= \prod_{i=r}^{n} \frac{\sigma_0}{\sigma_1} e^{\frac{x_i^2}{2}(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2})} \end{aligned} \tag{10}$$

and its logarithm

$$\lambda(n, r) = \sum_{i=r}^{n} \frac{\sigma_0}{\sigma_1} + \frac{1}{2}(\frac{1}{\sigma_0^2} - \frac{1}{\sigma 1^2}) x_i^2 \tag{11}$$

Since r is unknown its *maximum likelihood estimate under the hypothesis* H_0 must be used inside the $\lambda(n, r)$ formula to be able to apply the decision rule.

$$\hat{r} = arg \max_{1 \leq r \leq n} \prod_{i=1}^{r-1} f_0(x_i) \prod_{i=r}^{n} f_1(x_i) \tag{12}$$

The denominator of Λ is *independent* of r, so we can equivalently write

$$\hat{r} = arg \max_{1 \leq r \leq n} \lambda(n, r) \tag{13}$$

Then our decision rule becomes

$$\lambda(n, \hat{r}) \quad < \quad T \longrightarrow H_0 \tag{14}$$

$$\lambda(n, \hat{r}) \quad > \quad T \longrightarrow H_1 \tag{15}$$

or equivalently

$$\max_{1 \leq r \leq n} \lambda(n, r) \quad < \quad T \longrightarrow H_0 \tag{16}$$

$$\max_{1 \leq r \leq n} \lambda(n, r) \quad > \quad T \longrightarrow H_1 \tag{17}$$

We can either process records of n samples finding the maximum of the n sums and comparing it to the threshold — or — we can process each coming data sample using an "alarm" method to detect the impulse, as follows.

First of all, note that we can manipulate the likelihood ratio writing it as

$$\begin{aligned}
\lambda(n, r) &= (n - r + 1) \log \frac{\sigma_0}{\sigma_1} + \frac{1}{2}(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}) \sum_{i=r}^{n} x_i^2 = \\
&= k_1(n - r + 1) + k_2 \sum_{i=r}^{n} x_i^2 = \\
&= k_1 n + k_2 \sum_{i=1}^{n} x_i^2 - k_1(r-1) - k_2 \sum_{i=1}^{r-1} x_i^2 = \\
&= F(n) - F(r-1)
\end{aligned} \tag{18}$$

where

$$F(n) = k_1 n + k_2 \sum_{i=1}^{n} x_i^2$$

Then we have

$$\max_{1 \leq r \leq n} \lambda(n, r) = F(n) - \min_{1 \leq r \leq n} F(r-1) \tag{19}$$

If we think a little about the decision rule of equations (16),(17) and what equation (19) says we can easily produce the following sample-by-sample (on-line) decision procedure.

1. Assign initial value $j = 0$ to the counting parameter j.

2. Each time a data sample comes in, do

- $j = j + 1$
- compute $F(j)$
- $F^j_{min} = \min_{1 \le r \le j} F(r-1)$

3. If $F(j) - F^j_{min} < T$ go to step 2 and continue
4. If $F(j) - F^j_{min} > T$ "set the alarm" indicating that a change happened at the moment k for which $F(k) = F^j_{min}$ and then go to step 1, continue.(Note that we detected the change with delay $(j-k)$.)

Figure 4 describes the Page-Hinkley detection procedure for this example, while Figure 5 shows the behaviour of this procedure indicating the detection times and the delays for a typical sequence of data samples.

EXAMPLE 1.2: Detection of a jump in the mean of Gaussian noise.

We are going to use the Page-Hinkley detector and for simplicity assume that our noise is zero mean. We formulate our problem as

$H_0 : x_i, i = 1, \ldots, n$ follow $\mathcal{N}(0, \sigma_0)$
$H_1 : x_i, i = 1, \ldots, r-1$ follow $\mathcal{N}(0, \sigma_0)$
while $x_i, i = r, \ldots, n$ follow $\mathcal{N}(s, \sigma_0)$

In exactly the same fashion as in equation (10), the likelihood ratio is

$$\Lambda = \prod_{i=r}^{n} \frac{f_1(x_i)}{f_0(x_i)} = \tag{20}$$

$$= \prod_{i=r}^{n} \frac{e^{\frac{(x_i - s)^2}{2\sigma_0^2}}}{e^{\frac{x_i^2}{2\sigma_1^2}}} \tag{21}$$

and its logarithm

$$\lambda(n, r) = const \cdot \sum_{i=r}^{n} x_i - \frac{s}{2} \tag{22}$$

and with the same reasoning as before, we use the maximum likelihood estimate of the unknown time of change r. So our decision rule will be

$$\max_{1 \le r \le n} \lambda(n, r) \quad < \quad T \longrightarrow H_0 \tag{23}$$

$$> \quad T \longrightarrow H_1 \tag{24}$$

We may write the likelihood ratio as

$$\lambda(n, r) = \sum_{i=1}^{n} x_i - \frac{s}{2} - \sum_{i=1}^{r-1} x_i - \frac{s}{2} \tag{25}$$

$$= S(n) - S(r-1) \tag{26}$$

Then it is obvious how we can use an on-line decision procedure following an algorithm like the one we saw in the end of section 1.1.4.

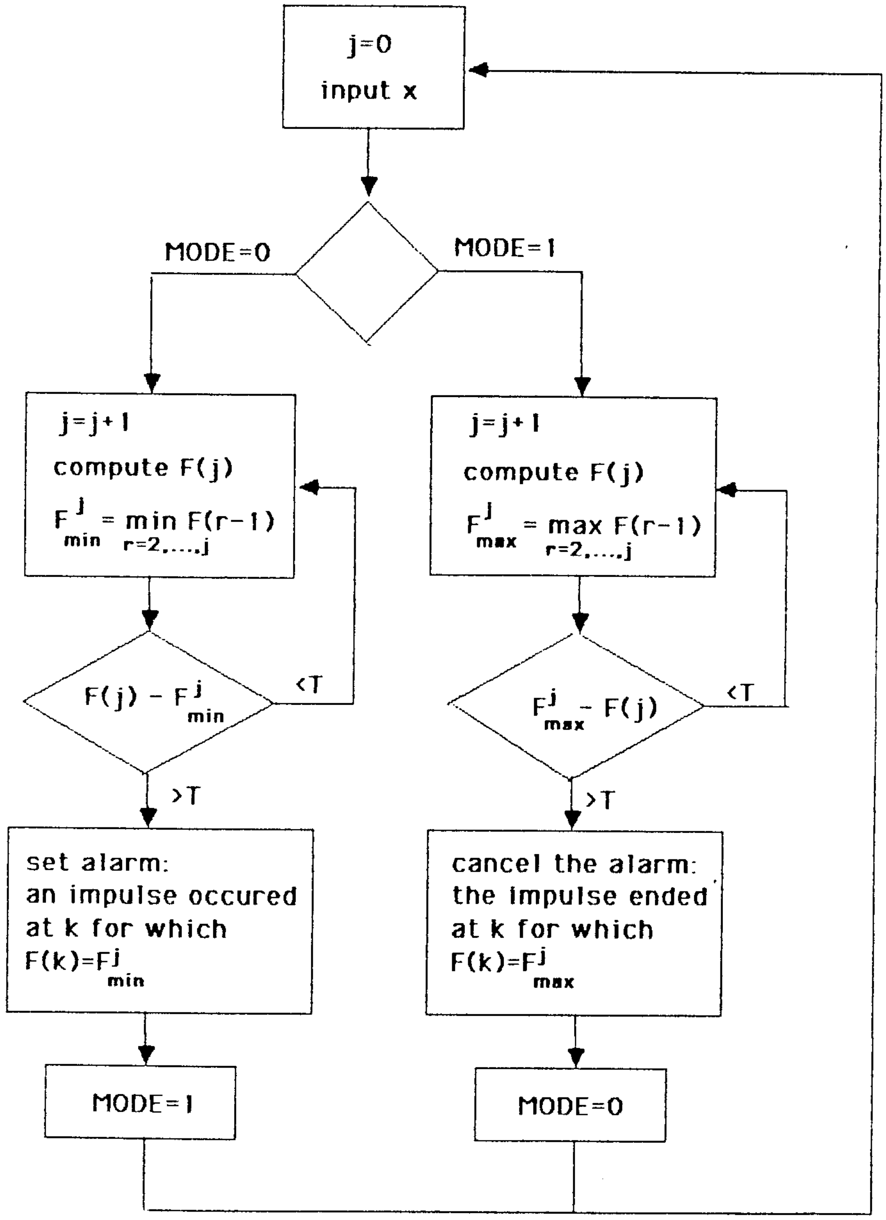

FIG. 4: The Page-Hinkley detection procedure for EXAMPLE 1.1.

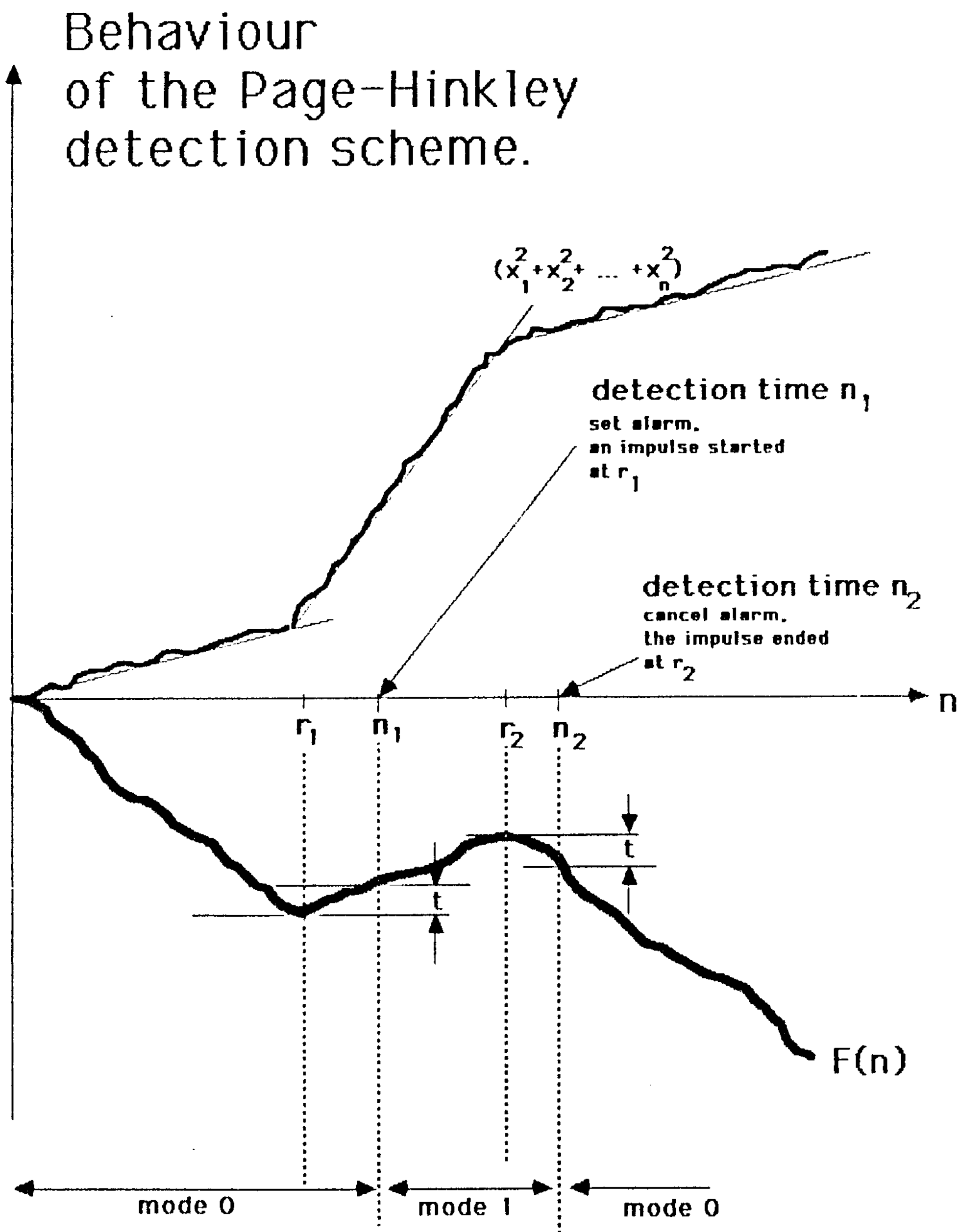

FIG. 5: The horizontal axis is (discrete) time and the evolution of the sum of the squares of the samples as well as of the function $F(n) = k_1 n + k_2 \sum_{i=1}^{n} x_i^2$ are indicated by continuous lines for simplicity.

2. THE MODELS THAT THE SYSTEM FOLLOWS ARE ELEMENTS OF A KNOWN SET.

EXAMPLE 2.1: Detection of a known signal in impulsively contaminated Gaussian noise.

A more than two hypothesis problem has to be formed here. Indeed there are two possible changes that can happen to a low variance Gaussian noise signal; either a dc part can suddenly be added to it or a high variance Gaussian impulse. (We conveniently assume that it is highly unlikely for both of them to occur at the same moment!)

Obviously this problem is simply a combination of examples 1.1 and 1.2, but we must not forget that now we have to form a **three hypothesis problem**, namely

$H_0 : x_i, i = 1, \ldots, n$ follow $\mathcal{N}(0, \sigma_0)$
$H_1 : x_i, i = 1, \ldots, r_1 - 1$ follow $\mathcal{N}(0, \sigma_0)$ *while* $x_i, i = r_1, \ldots, n$ follow $\mathcal{N}(s, \sigma_0)$
$H_2 : x_i, i = 1, \ldots, r_2 - 1$ follow $\mathcal{N}(0, \sigma_0)$ *while* $x_i, i = r_2, \ldots, n$ follow $\mathcal{N}(0, \sigma_1)$

So now, we have two likelihood ratios to deal with

$$\Lambda_1 = \frac{P[x_1, \ldots, x_n/H_1]}{P[x_1, \ldots, x_n/H_0]} \quad with \quad \log \Lambda_1 = \lambda_1 \tag{27}$$

$$\Lambda_2 = \frac{P[x_1, \ldots, x_n/H_2]}{P[x_1, \ldots, x_n/H_0]} \quad with \quad \log \Lambda_2 = \lambda_2 \tag{28}$$

Bayes criterion leads to the following decision rule

$$\lambda_1 \; < \; t_1 \longrightarrow H_0 \; or \; H_2 \tag{29}$$

$$> \; t_1 \longrightarrow H_1 \; or \; H_2 \tag{30}$$

$$\lambda_2 \; < \; t_2 \longrightarrow H_0 \; or \; H_1 \tag{31}$$

$$> \; t_2 \longrightarrow H_2 \; or \; H_1 \tag{32}$$

$$\lambda_2 - \lambda_1 \; < \; t_3 \longrightarrow H_1 \; or \; H_0 \tag{33}$$

$$> \; t_3 \longrightarrow H_2 \; or \; H_0 \tag{34}$$

This problem becomes even more interesting if we want to use the on-line Page-Hinkley detection schemes that we saw in sections 1.1.5 and 1.2. Then we need to "update" the order of the three hypotheses after each detection event according to the results of our detection. Our decision procedure now obtains an adaptive element. The algorithm we use is shown in Figure 7. In this figure we use i, j, k to symbolize the three possible situations, that is

1. zero-mean Gaussian noise with low variance σ_0

2. a dc part s is present, the variance is σ_0

3. zero-mean Gaussian noise with high variance σ_1

The expression $H_1 : i \rightarrow j$, *at* r_1 for example, means: *Let* H_1 *be the hypothesis that a change from situation* i *to situation* j *happened at time* r_1. I hope that Figure 6 will be easy to understand after that. It sketches the detection procedure for this three-hypothesis problem. This procedure is actually an adaptive procedure where the three hypotheses are updated after each iteration depending on the detection result of this iteration.

Comments on Examples 1.1 , 1.2 and 2.1

Of course the problems presented and solved in these examples are extremely simplified to the point that they are unrealistic. We assume that we know too many things about the statistics of the noises (variances etc.) and in 1.2 our task is to detect the very simple dc signal. Nevertheless one should think of these examples as being the stepping stones to more complicated and realistic problems concerning communication channels, real life detection tasks (as in submarines or radar) or even a reflectivity sequence ...

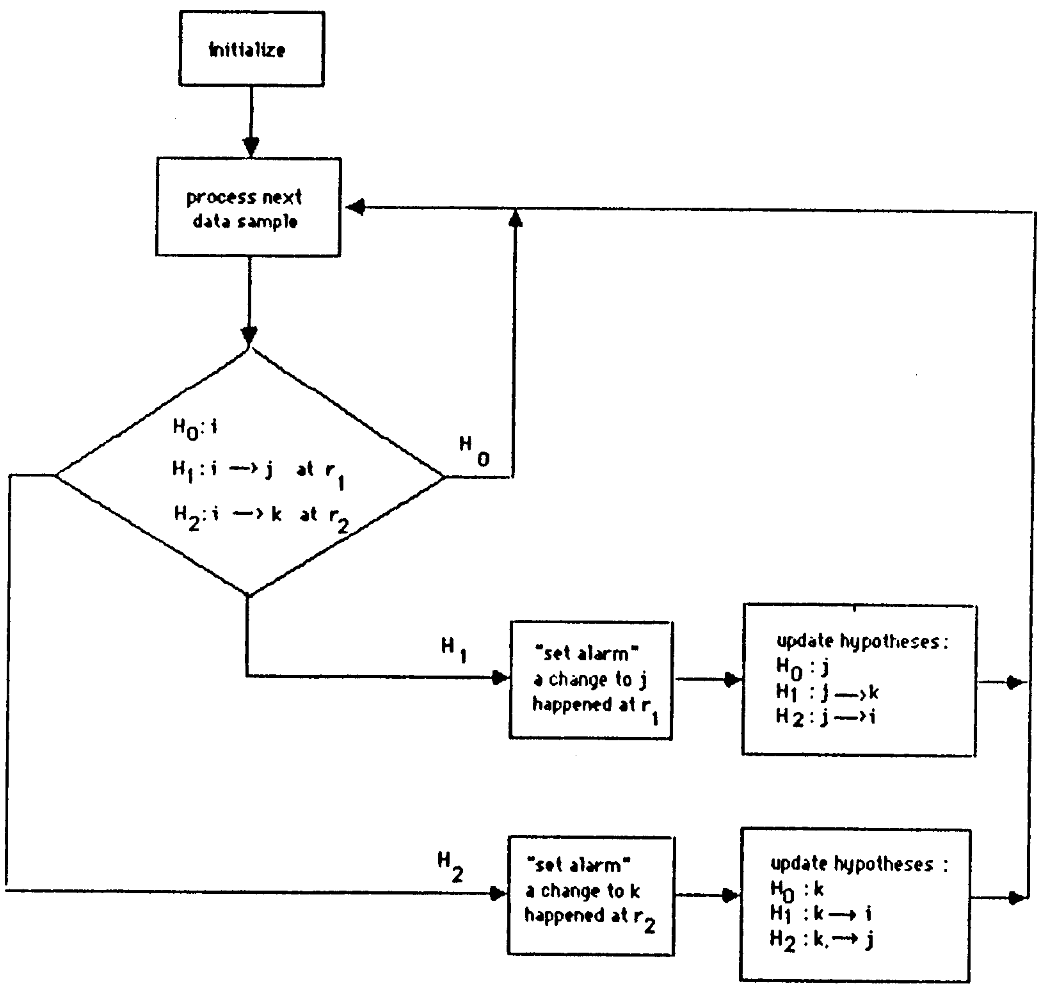

FIG. 6: The (on-line) decision procedure for Example 2.1.

EXAMPLE 2.2: Detection of abrupt changes in linear stochastic systems.

A deterministic (discrete time) linear system is described by the following model

$$x(k+1) = A(k)x(k) + B(k)u(k)$$
$$y(k) = C(k)x(k)$$
$$x(0) = x_0 \, , \, k > 0$$

A stochastic linear system is described by

$$x(k+1) = A(k)x(k) + B(k)u(k) + w(k) \quad (35)$$
$$y(k) = C(k)x(k) + v(k) \quad (36)$$

where $w(k)$,$v(k)$ are zero-mean Gaussian white noise processes independent of each other. Our problem here is to detect a sudden change in the model that a linear stochastic system follows. Before presenting a likelihood ratio method to attack this problem we need some basic knowledge of the famous Kalman filter. Kalman filtering is a rather complicated story, but I will present the main idea as briefly as possible.

About Kalman filter: We have the system of equations (35),(36). Let $\hat{x}(k+1/k+1)$ be the *one-step predicted estimate* of x(k+1) *based* on $u(0), \ldots, u(k), y(0), \ldots, y(k+1)$. The Kalman filter gives $\hat{x}(k+1/k+1)$ by

$$\hat{x}(k+1/k) = A(k)\hat{x}(k/k) + B(k)u(k) \quad (37)$$
$$\hat{x}(k+1/k+1) = \hat{x}(k+1/k) + K(k+1)\gamma(k+1) \quad (38)$$
$$\gamma(k+1) = y(k+1) - C(k)\hat{x}(k+1/k) \quad (39)$$

where $\gamma(k)$ is the *innovation process.* Under the assumptions we made for the system, this process is zero-mean Gaussian, white and with covariance matrix $V(k)$ which is also computed by the Kalman filter.

We may now address our problem.

2.2.1 The Generalized Likelihood Ratio method.

We examine a linear stochastic system and form the following **multiple** hypothesis problem. $Under H_0$ the system remains unchanged and is described by the model M_0:

$$x(k+1) = A(k)x(k) + B(k)u(k) + w(k) \quad (40)$$
$$y(k) = C(k)x(k) + v(k) \quad (41)$$

$Under H_i$, $i = 1, \ldots, N$ at the moment θ an abrupt change takes place. After that the system is described by one of the following N models $M_i, i = 1, \ldots, N$

$$x(k+1) = A(k)x(k) + B(k)u(k) + w(k) + f_i(k, \theta) \quad (42)$$
$$y(k) = C(k)x(k) + v(k) + g_i(k, \theta) \quad (43)$$

We want to detect the change, decide which of the possible N changes actually occured and finally we want to estimate the time θ of its occurrence. For this purpose we shall use a likelihood ratio approach slightly different from the one we used in the previous examples.

A Kalman filter is placed after our system. As long as the system follows M_0, that is *under*H_0 the filter gives

$$\gamma(k+1) = y(k+1) - C(k)\hat{x}(k+1/k) = \gamma_0(k+1) \tag{44}$$

by equation (39). This is a zero-mean Gaussian white process with covariance $V_0(k)$ as we already saw above. Since both our system models and the filter are **linear** as soon as the model changes to M_i, that is *under*H_i the filter gives

$$\begin{aligned} \gamma(k+1) &= y(k+1) - C(k)\hat{x}(k+1/k) :+ \rho_i(k+1,\theta) & (45)\\ &= \gamma_0(k+1) + \rho_i(k+1,\theta) & (46) \end{aligned}$$

where ρ_i can be recursively computed from f_i, g_i . Note that *under*H_i the innovation process $\gamma(k)$ is white, Gaussian with mean ρ_i and covariance matrix $V(k)$.

So we see that *a change in the model resulted in a change in the mean of the innovations process. Thus to detect this change we could use the innovations process γ as our observations instead of using the output process $y(k)$ of our system.* This will result to a likelihood ratio that we call *Generalized Likelihood Ratio.* This is computed in the same way as in our previous examples.

$$\begin{aligned} \Lambda_i &= \frac{P[\gamma(1),\ldots,\gamma(k)/H_i]}{P[\gamma(1),\ldots,\gamma(k)/H_0]} = \\ &= \frac{\prod_{j=1}^{\theta-1} const\cdot e^{-\frac{1}{2}\gamma_0^\top(j)V^{-1}(j)\gamma_0(j)} \prod_{j=\theta}^{k} const\cdot e^{-\frac{1}{2}(\gamma_0(j)-\rho_i(j,\theta))^\top V^{-1}(j)(\gamma_0(j)-\rho_i(j,\theta))}}{\prod_{j=1}^{k} const\cdot e^{-\frac{1}{2}\gamma_0^\top(j)V^{-1}(j)\gamma_0(j)}} \\ &= \prod_{j=\theta}^{k} e^{-\frac{1}{2}(\gamma_0(j)-\rho_i(j,\theta))^\top V^{-1}(j)(\gamma_0(j)-\rho_i(j,\theta))+\frac{1}{2}\gamma_0^\top(j)V^{-1}(j)\gamma_0(j)} \qquad (47) \end{aligned}$$

Again, taking its logarithm we get

$$\begin{aligned} \lambda_i &= \sum_{j=\theta}^{k} -\frac{1}{2}(\gamma_0(j)-\rho_i(j,\theta))^\top V^{-1}(j)(\gamma_0(j)-\rho_i(j,\theta)) + \frac{1}{2}\gamma_0^\top(j)V^{-1}(j)\gamma_0 \qquad (48)\\ &= \sum_{j=\theta}^{k} \frac{1}{2}\gamma_0^\top(j)V^{-1}(j)\rho_i(j,\theta) + \frac{1}{2}\rho_i^\top(j,\theta)V^{-1}(j)\gamma_0(j) - \frac{1}{2}\rho_i^\top(j,\theta)V^{-1}(j)\rho_i(j,\theta) \end{aligned}$$

Both V and V^{-1} are symmetric matrices and the first two terms inside the sum are equal. So

$$\lambda_i(k,\theta) = \sum_{j=\theta}^{k} \gamma_0^\top(j)V^{-1}(j)\rho_i(j,\theta) - \sum_{j=\theta}^{k} \rho_i^\top(j,\theta)V^{-1}(j)\rho_i(j,\theta) \tag{49}$$

As in previous examples we have to use the maximum likelihood $\hat{\theta}$ of the unknown time of change θ. This will finally give us the following decision rule

$$\begin{aligned} \max_{1\le\theta\le k} \lambda_i(k,\theta) &< T_i \longrightarrow H_0 & (50)\\ &> T_i \longrightarrow H_i & (51) \end{aligned}$$

Of course this is the decision rule for the "first" change that happens to the system. If in real life our system is a mechanical system and we try to detect when one of N possible types of failure occurs, this decision rule is all we need — after that we stop, repair our system and then start all over again.

If on the other hand we have to continuously watch our system switching between different models then we can use a procedure like the on-line adaptive algorithm described in section 2.1.1.

In any case the important feature of the *Generalized Likelihood Ratio* method is that it uses a single Kalman filter at the output of the system. This is noted here in contrast to the next method presented in section 2.2.2, which uses N Kalman filters to solve the same problem, that is detect when and which of the N possible changes happened to the system.

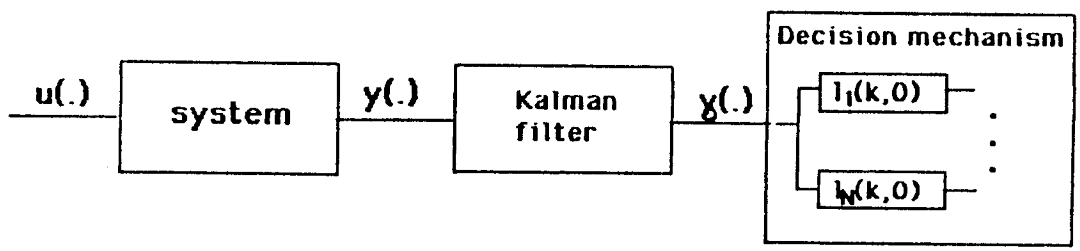

FIG. 7: The Generalized Likelihood Ratio Method.

2.2.2 The Multiple Model method.

The *Generalized Likelihood Ratio* method of the previous section was designed to detect *additive* changes in our stochastic system. The *Multiple Model* method detects changes between a known set of possible models but the changes need not be additive.

The basic idea of this method is to use as many Kalman filters as the possible models, say N. Each of the filters is based on one of the N possible models. (To make this clear: By "model" we mean of course a set of matrices A,B,C and noise processes w,v. Then N Kalman filters are N sets of equations like equations (37),(38),(39).) Each of the N innovation processes that are produced will be white, zero-mean Gaussian (with the respective covariance matrix given by the filter) *if and only if* the system actually follows the respective model. The rest $N-1$ innovation processes will deviate from this behaviour, thus providing us with a way to know what is happening to our system by simply watching these processes.

For example a quantitative way to do this is to keep computing and comparing the probabilities

$$p_i(k) = P[H_i/u(0), \ldots, u(k-1), y(1), \ldots, y(k)] \quad (52)$$

which can be computed *recursively* from the innovation processes coming as the outputs of the N Kalman filters.

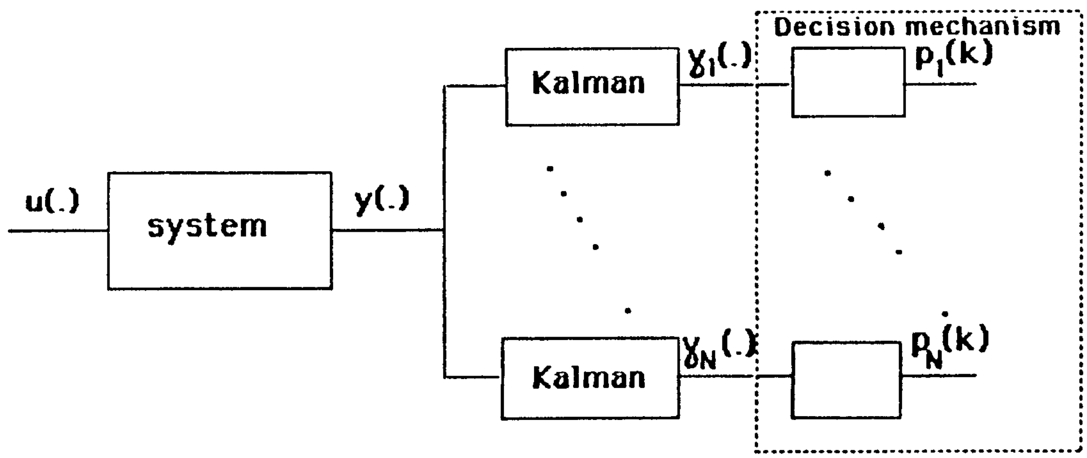

FIG. 8: The Multiple Model method.

3. ONE MODEL IS KNOWN AND THE OTHER IS COMPLETELY UNKNOWN.

EXAMPLE 3: Detection of impulses contaminating Gaussian noise.

When the M_0 model is known but the statistical characteristics of M_1 are completely unknown, the obvious thing to do is to *use an estimate for whatever is unknown.* We note that in a likelihood ratio approach the unknown parameters are then in the numerator of the ratio only. Then maximum likelihood estimation proves very convenient since maximization under H_1 is equivalent to maximization of the ratio itself, as we already saw in equations (11),(12).

For example in the detection of impulses of unknown variance in Gaussian noise the decision rule can be easily derived by equations (5),(6),(7) to be

$$\max_{\sigma_1} \Lambda(x_i, \sigma_1) \quad < \quad T \longrightarrow H_0 \quad (53)$$

$$> \quad T \longrightarrow H_1 \quad (54)$$

In the same way, if we wanted to use the Page-Hinkley detection scheme using equations (16),(17) and maximum likelihood estimation of the impulse variance would give the following decision rule

$$\max_{1 \leq r \leq n} \max_{\sigma_1} \quad < \quad T \longrightarrow H_0 \tag{55}$$

$$> \quad T \longrightarrow H_1 \tag{56}$$

CONCLUSION

Old fashioned though it may be called, the theory of detection and estimation still has a lot to offer to a variety of problems and a variety of disciplines. The problem of change detection is only one such example. In geophysics the issue could also be *detection vs. estimation.* Indeed when dealing with a huge amount of data some of which are different from the others it could be a lot easier to *detect* "which are which" than using estimation to do it. I believe it is exactly this fact that could make detection theory appealing to a geophysicist.

REFERENCES

Basseville,M. and Benveniste,A., 1984, Lecture notes in control and information sciences, Vol. 77: Detection of abrupt changes in signals and dynamical systems, Springer-Verlag.

Bouvet, M. and Schwartz, S., Underwater noises: Statistical modeling, detection and normalization, Dept. of Electrical Engineering, Princeton University, Princeton NJ 08544.

Bouvet,M. and Schwartz, S.C., Comparison of adaptive receivers for signal detection in ambient underwater noise, Dept. of Electrical Engineering, Princeton University, Princeton NJ 08544.

Czarnecki, S.V., 1983, Nearly optimal detection of signals in non-Gaussian noise, Ph.D. dissertation, Dept. of Electrical Engineering, Princeton University.

Hines, W.G.S., A simple monitor of a system with sudden parameter changes, IEEE Transactions on Information Theory, Vol. IT-22, No. 2, March 1976.

Hines, W.G.S., Improving a simple monitor of a system with sudden parameter changes, IEEE Transactions on Information Theory, July 1976.

Mendel, J., Some modeling problems in reflection seismology, IEEE ASSP Magazine, April 1986.

Miller, J. and Thomas, J.B., Detectors for discreet-time signals in non-Gaussian noise, IEEE Transactions on Information Theory, Vol. IT-18, No. 2, March 1972.

Miller, J. and Thomas, J.B., The detection of signals in impulsive noise modeled as a mixture process, IEEE Transactions on Communications, May 1976.

Shiryaev, A.N., (translated by R. A. Silverman), On optimum methods in quickest detection problems, Theory of probability and its applications, 1963.

Veitch, J. and Wilks, A., A characterization of Arctic undersea noise, J. Acoust. Soc. Am., 77(3), March 1985.