CHAPTER 5
SEQUENTIAL DETECTION OF
DISORDER TIME

5.1 Introduction.

Suppose that during the time a sequence of random variables is being observed, the probability distribution governing them undergoes a change, called disorder or disruption, at some unknown instant. Assume that it is desirable to detect the occurrence of such a change 'quickly' and that the object is to obtain a rule which minimises a loss associated with the delay. This sums up, broadly, the 'disorder' problem (in sequential sampling). The word disorder is used here to mean a change in the distribution of the observations either through one or more parameters or through its form. In addition to detection, other relevant objects of interest could be a study of the effect of disorder on the observations, and inference concerning the pre- and post-disorder distributions, or the shift epoch itself. In the course of observation of any phenomenon, it is not uncommon to find ourselves faced with the problem of disorder. See, e.g., Hsu (1979) who gives interesting applications such as the ones occurring in stock market and air traffic analysis.
The fixed-size sample version, which amounts to a retrospective analysis, is called the change-point problem and has been studied extensively in a number of special cases. Chernoff and Zacks (1964), Hinkley (1970) and Lee and Heghinian (1977) are concerned, among others, with changes in the normal mean, and Menzenfricke (1981) about changes in the variance; Bhattacharya and Johnson (1968) and Pettitt (1979) propose nonparametric tests; MacNeill and Duong (1982) study the effect of parameter changes in a time series model. Darkhouskii and Brodskii (1980) address the problem when observations are dependent but satisfy a mixing condition. They give a consistent estimator of the disorder instant.

When we are called on to detect changes in law during the evolution of the process, the problem is called as the sequential detection of disorder (or quickest detection problem). In engineering parlance, this may be termed 'online detection'. Quality surveillance problems can be modelled in the present framework. Thus, the work of Shewhart (1931) and Page (1954) may be regarded as precursors of the optimal sequential detection formulation. Shewhart's control-chart technique has a wide practical and intuitive appeal while Page's cumulative sum (cusum) procedures are statistically more satisfying and
widely studied theoretically.

The cusum statistic of Page is extended to the detection of a shift in the Wiener model by Bagshaw and Johnson (1975). Khan (1979) develops a cusum type procedure for detection of disorder in an independent sequence so as to permit control of the 'false-alarm' probability.

Some of the early attempts to determine optimal strategies for detecting shifts in the Wiener model were by Antleman and Savage (1965) and by Bather (1963). Their criterion of optimality was income per unit. Antleman and Savage considered the case of observations being available free of cost as well as of costly information.

In Shiryaev (1963, 1965), the sequential detection problem has been investigated for an independent sequence as well as for Wiener process using a Bayesian approach with a function of the average detection delay as the loss. If 'no change in law' is labelled as a state of 'order' (= '0'), a 'change' as 'disorder' (= '1'), and if $\xi_t$ denotes the state at time $t$, then the process of interest is $(\xi_t : t \geq 0)$. It makes only one transition, 0 $\rightarrow$ 1, and gets absorbed. Once absorption takes place we are required to put off the source. But $(\xi_t)$ is
not directly observable; a signal indicating absorption is obtainable from evidence inherent in the realization of the observation process. It is generally assumed that a signal can be verified instantly to know whether there really was any change in the physical factors that determined the probabilistic behaviour of the realization. This setup, put in a Bayesian framework, leads to a stopping problem. Shiryaev has shown that it is sufficient to base the decision rule on the process \( \Pi_t : t \geq 0 \), where \( \Pi_t = P(\xi_t = 1 \mid \mathcal{F}_t) \) is the posterior probability of absorption of \( \xi_t \) given \( \mathcal{F}_t \), the history at time \( t \) of the observation process. Bather (1967) gives further results. Balnor (1975) obtains optimal strategies for detecting a change in the drift rate of a Wiener process when information has to be obtained at a cost.

For disorder in point processes, Davis and Wan (1977) give a stochastic filtering formulation and solve it by applying the innovation process approach. They get results similar to those of Shiryaev.

A different objective function, that of maximising the probability of stopping within a specified number of observations from the disorder epoch was introduced by Bojdecki (1979).
The remainder of this Chapter consists of the following. In Section 5.2, we give the Bayesian formulation of the problem in its different versions. We also introduce there a further variant of the problem, namely, minimisation of the probability of an 'excessive' delay while the probability of a 'false signal' is controlled to a predetermined level. This version is solved in Section 5.3. In Sections 5.4 and 5.5, we consider the disorder problem in two Markov-dependent sequences: a shift in the mean of an autoregressive process, and a change in the regression coefficient when the 'errors' are serially correlated.

5.2 Disorder Detection in an Independent Sequence

A. Review.

It is the purpose of this section to present and compare various versions of the disorder problem differing in their loss structure but under the common assumption of independence of observations. This includes some new versions also.

5.2.1 Formal Statement of the Problem.

Suppose that a process \( X = \{X_n : n \in \mathbb{N}\} \) of independent random variables defined on a measurable space \((\Omega, \mathcal{F})\) is being sequentially observed.
and that it consists of
\[ X_n = X_n' \quad (n < \theta), \]
\[ = X_n'' \quad (n \geq \theta), \]
where \( \theta \) is the disorder instant. The probability
distribution of \( X_1', X_2', \ldots, X_{\theta-1}' \) is \( P_0 \) while it
is \( P_1 \) for \( X_\theta', X_{\theta+1}', \ldots \). Thus, disorder at epoch \( \theta \) is
assumed to cause an instantaneous switch to \( P_1 \). The
process \( (\xi_n : n \in \mathcal{F}) \) is such that \( \xi_n \) indicates the
probability distribution of \( X_n \). Its initial state \( \xi_0 \)
specifies whether, prior to experimentation, the
source is in order or not. Thus,
\[ \xi_n = 0 \quad (n < \theta) \]
\[ = 1 \quad (n \geq \theta). \]
We assume that the disorder moment \( \theta \) is random and
has a known distribution. Also \( \theta, \xi_n \) and the other
relevant functions to follow are all assumed to be
\( \mathcal{A} \)-measurable. Let \( (\mathcal{F}_n) \) denote the sequence of
histories induced by \( X \), and \( \mathcal{F} \) the class of \( \mathcal{F}_n \)-times
for which the relevant risks are defined and the
risks are greater than \( -\infty \). Based on the realization
of \( X \) upto the decision epoch, we wish to stop the
process \( (\xi_n) \) soon after it reaches the disorder
state so as to minimise an associated risk. The
decision problem is thus a stopping problem. In fact,
A sequential detection rule is defined to be a Markov time relative to \( \{ \mathcal{F}_n \} \). Thus, given \( \tau \in \mathcal{T} \), the event \( \{ \tau = n \} \) leads to termination of \( \mathcal{X} \) at time \( n \) and declaration of disorder.

A detection rule may lead, on the one hand to a delayed detection of disorder and on the other it may call for stopping when actually no change has occurred—false or premature signal. To each of these, there corresponds a loss. The two being in opposition to each other, a detection rule must ideally, balance them.

Suppose that a certain fixed probability measure \( p^\pi \) specifies the prior distribution of \( \theta \) as

\[
(1) \quad p^\pi (\theta = 0) = \pi, \quad p^\pi (\theta = k) = p_k \quad (k \in \mathbb{N}, \pi \in [0, 1])
\]

for sets \( A_k = \{ \omega : X_{k-1} \leq x_k, \ldots, X_n \leq x_n \} \) \( (k, n \in \mathbb{N}) \), where \( k \) is the disorder moment, it is the mixture of \( P_0 \) and \( P_1 \) given by

\[
(k=1) p^\pi (A_{1n}) = (\pi + \rho_1)P_1(A_{1n}) + \sum_{i=2}^{n} p_i P_0(A_{1,i-1})P_1(A_{1n}) \]

\[
(2) \quad + \sum_{j=n+1}^{\infty} p_j P_1(A_{1n}),
\]

\[
(1 < k \leq n) \quad p^\pi (A_{kn}) = \sum_{i=k}^{n} p_i P_0(A_{1,i-1})P_1(A_{1n}) + \sum_{j=n+1}^{\infty} p_j P_1(A_{1n})
\]
Thus (1) and (2) together describe any finite dimensional probability distribution of the process \( X \). We shall evaluate all our risk elements wrt \( P^\pi \).

If the loss under a \( \tau \) is denoted by \( L(\tau; \theta, \xi_\tau) \), the problem may be stated as

\[
(3) \quad \inf \ E^\pi L(\tau; \theta, \xi_\tau).
\]

### 5.2.2 Expressions for Posterior Probabilities

We represent \( \pi_n \) and other posterior probabi-

\[Q \text{ilities in a form convenient for use later. Similar expressions are obtained in Section 5.4 for Markov-}
\]
dependent normal sequence \( X \).

Let \( P_0(.) \) and \( P_1(.) \) denote respectively the densities of \( P_0 \) and \( P_1 \) wrt some \( \sigma \)-finite measure, say \( \mu \). Define for \( k, n \in \mathbb{N} \),

\[
\nu_k = \frac{P_1(x_k)}{P_0(x_k)},
\]

\[
\nu_k^n = \prod_{j=k}^{n} \nu_j \quad (1 < k < n)
\]

\[
= 1 \quad (k > n)
\]

and

\[
p_k^n = P(\theta = k | \mathcal{F}_n),
\]

\[
\pi_k^n = P(\theta \leq k | \mathcal{F}_n).
\]
We shall write \( \pi_n \) as \( \pi_n \). Note that \( \xi_n = 1 \) iff \( \theta \leq n \). When \( \theta \) has the (general) distribution in (1), we evaluate the conditional probabilities \( p^\theta_k \) and \( \pi_n^\theta \) by Bayes' formula to get

(1) \[ w_n = \frac{\sum_{j=1}^n (p_j / P(\theta > n)) \, v_j^n}{1 + \frac{\pi}{1-\pi} v_1^n + \sum_{j=1}^n (p_j / P(\theta > n)) \, v_j^n} \]

(ii) \[ p_k^n = (1-\pi_n) (p_k / P(\theta > n)) \, v_k^n \]

(iii) \[ \pi_k^n = 1 - (1-\pi_n) \sum_{j=k+1}^\infty (p_j / P(\theta > n)) \, v_j^n \]

When the prior distribution of \( \theta \) is as in (5) below

(5) \[ P^\theta(\theta = 0) = \pi \quad (0 \leq \pi \leq 1) , \]

(5) \[ P^\theta(\theta = j \mid \theta > 0) = p q^{j-1} \quad (0 < q < 1, \ q = 1-p, \ j \in \mathbb{N}) , \]

the following simpler representations are possible

(1) \[ w_n = \frac{p \sum_{k=1}^n y_k^n}{1 + \frac{\pi}{1-\pi} y_1^n + p \sum_{k=1}^n y_k^n} \]

(ii) \[ p_k^n = (1-\pi_n) \rho y_k^n \quad (1 \leq k \leq n) \]

(6) \[ = (1-\pi_n) p q^{k-n-1} \quad (k > n) , \]

(iii) \[ \pi_k^n = 1 - (1-\pi_n) (p \sum y_j^n + 1) \quad (k < n) \]

\[ = 1 - (1-\pi_n) q^{k-n} \quad (k \geq n) , \]

where \( y_k \) and \( y_k^n \) have the definitions:
\[ y_k = \frac{1}{\mathbb{P}_0(x_k)} \frac{\mathbb{P}_0(x_k)}{\mathbb{P}_0(x_k)} , \]

\[ y_k^n = \prod_{j=k}^{n} y_j \quad (1 \leq k \leq n) \]

\[ = 1 \quad (k > n). \]

5.2.3 Various Versions—Their Interrelations.

Given a detection rule \( \tau \), \( \{ \tau < \theta \} \) and \( (\tau - \theta)^+ \) denote respectively the event 'false alarm' and the random variable 'detection delay in the presence of disorder'. Let \( \alpha(\tau) \) denote the probability of a false alarm and \( N(\tau) \) the expected number of false signals before the appearance of disorder. Let \( \mathcal{F}(\alpha) \) consist of rules, each of which has a false-alarm probability that does not exceed a specified level \( \alpha \), i.e.,

\[ \mathcal{F}(\alpha) = \{ \tau \in \mathcal{F} : \alpha(\tau) \leq \alpha \} \quad (0 < \alpha < 1). \]

Similarly, \( \mathcal{F}(N) \) for a fixed \( N(\in N) \). We now state Versions I and II (Shiryaev, 1963).

\begin{align*}
\text{Version I:} & \quad \inf_{\mathcal{F}(N)} \mathbb{E} (\tau - \theta)^+ \\
\text{Version II:} & \quad \inf_{\mathcal{F}(\alpha)} \mathbb{E} (\tau - \theta)^+ 
\end{align*}

These two are equivalent if it is known a priori that \( \theta \) follows the law in (5). In that case
\( \alpha(\tau)/[1 - \alpha(\tau)] \). Another simplicity arising out of this prior is time-homogeneity of the process \( \Pi_H \) (compare expressions in (4) and (6)).

We add a third version which turns out to be equivalent to the first two under (5). For a general prior, this equivalence no longer holds. Define 
\((\tau - \theta)^+ = (\tau - \theta)^+ \) \( \text{if } \theta < \tau \), and let \( \mathcal{T}(\gamma) \) denote the class \( \{ \tau \in \mathcal{T} : \Pi(\tau - \theta)^+ \leq \gamma \} \) for a preassigned positive constant \( \gamma \). This gives another variant of the problem, referred to as

**Version III.** \( \inf \mathbb{E}^\pi (\tau - \theta)^+ \)

This is comparable to a somewhat standard approach, namely, seeking a procedure that minimises \( \hat{\text{ARL}}_1 \) \( \text{def} = \mathbb{E}_1 \tau \) in the class of procedures with \( \hat{\text{ARL}}_0 \) \( \text{def} = \mathbb{E}_0 \tau \) at least as large as some specified number (cf. Khan 1979; Munford, 1980).

**Lemma 5.1** \( \mathbb{E}^\pi (\tau - \theta)^+ = \mathbb{E}^\pi \left[ \Sigma_{k \geq \tau} P(\theta > k) \right] \)

which, when \( \theta \) is geometric \((p)\), reduces to \( \alpha(\tau)/p \).

**Proof** We use \( \mathcal{F}_\tau \)-measurability of \( 1_{\tau = \eta} \) and the expression for \( (1 - \pi^n_k) \) from (4).

\( \mathbb{E}^\pi (\tau - \theta)^+ = \mathbb{E}^\pi \Sigma_{\eta = 1}^{\infty} 1_{\tau = \eta} (n - \theta)^+ \)
When $\theta$ has a geometric prior with 'success' probability $p$, then

$$
\text{BIR}(1-n) = P(\theta > k \; \text{I} \; n)
$$

and

$$
\text{R}^T(t) = \sum_{n=1}^{\infty} P(\theta > k) \frac{1}{n^T}
$$

It is clear from the above that when $p$ is known, to each $\mathcal{I}^T(Y)$ there is a $\mathcal{I}^T(\alpha)$ corresponding to it. Immediately, we have the equivalence of Versions III and I (or II).

**Corollary.** Given a known geometric prior for $\theta$, the optimal detection rule in $\mathcal{I}^T(Y)$ is identical to the one optimal in $\mathcal{I}^T(\alpha)$ or $\mathcal{I}^T(N)$ if $\alpha$ and $N$ are chosen to be $pY$ and $pY/(1-p)$ (assuming $\gamma < p^{-1}$).

When one wishes to maximise the probability of detecting disorder within a reasonable time say $m(> 0)$ instants from the moment of disorder, the problem assumes
Version IV (Bojdecki, 1979). \( \inf \ P(\tau - \Theta \geq m) \),
where \( P \) is any fixed probability measure on \((\Omega, \mathcal{F})\).
In relation to the earlier versions, this one takes, apparently, quite a different view of the problem.
Even so, it may be restated in terms of a standard optimal stopping problem and a solution to it can be shown to exist. Indeed, it is equivalent to

\[
\inf \mathbb{E}( \Pi_{\tau+m}^\tau - \Pi_{\tau-m}^\tau ),
\]

The optimal rule for this decides about the presence or otherwise of disorder based on only the latest \( m \) observations.

5.2.4 A New Version.

We may note that Version IV does not allow a differential consideration of the two types of signals, true and false. Such a distinction may just be a natural requirement in a given situation or it may be desirable as a technical means of controlling the costs relating to false signals. Accordingly, we consider, analogous to the decision problem formulation of hypothesis testing, the loss structure implicit in (9) below.

Assume that \((X_n)\) consists of independent observations, but take the prior distribution of \( \Theta, \)
for convenience, to be
\[ P_k^*(\theta = k) = pq^{k-1} \quad (k \in \mathbb{N}, 0<p<1, q=1-p) \]
\[(8)\]
= 0 elsewhere,
which corresponds to (5) with \( \tau = 0 \). Let \( P \) be any fixed probability measure on \((\mathcal{X}, \mathcal{F})\). The loss elements concerned will be averaged wrt this measure (see (9) and Version V). We are lead to consider
\[ (9) \quad \inf_{\mathcal{F}} \left[ \int w P(\tau < \theta) + P((\theta-\tau)^+ \geq m) \right] \]
for a positive constant \( w \), and then control \( \zeta(\tau) \) to arrive at a new version, namely,
\[ \text{Version V.} \quad \inf_{\mathcal{F}} P((\theta-\tau)^+ \geq m). \]
Section 5.3 concerns itself with this.

5.2.5 On the Optimal Solution.

Having framed disorder problem as a stopping problem, the optimal sequential detection rule may be obtained by application of the theory outlined in Section 1.3. Consider the problem as posed in (3), and the \( \Pi_n \)-process referred to in the Introduction. It has been shown in Shiryaev (1964) that \( (\Pi_n) \) is Markov-sufficient (equivalently, Bahadur-transitive, see Irle, 1980) for \( (\mathcal{F}_n) \) under certain choices of \( L \). That is, if \( \mathcal{F} \) is the class of stopping times
relative to the sequence of histories \( (\xi_n^\tau) \) (say)
induced by \((\Pi_n)\), then, the following holds,

\[
\inf \mathbb{E}^\pi L(\tau ; \theta, \xi_n^\tau) = \inf \mathbb{E}^\pi L(\tau; \theta, \xi_n^\tau).
\]

Given a geometric prior for \( \theta \), the process
\((\Pi_n, \xi_n^\tau)\) is homogeneous Markov, and the optimal
sequential detection rule for Versions I-III is of
the form

\[
\inf \{ n : \tau_n \geq \gamma \}
\]

where \( \gamma \) is a constant, \( 0 < \gamma < 1 \). In general \( \gamma \) is
not easily determined (cf. boundaries for Wald SPRT).
But when \( X_n \)'s are discrete with a lattice distribution,
we may use the results of Zacks (1980) who has derived
recursive relations for the distribution of the optimal
rule stated in (10). He has also provided computational
algorithms and their diffusion approximations.

For Versions IV and V, \((\Pi_n)\) does not play
the same role. The rule optimal for these is based
on the latest \( m \) observations only (see Bojdecki,
1979; and Section 5.3).

When \( \theta \) has a general distribution, the form
of the optimal rule will still be the same (e.g., (10)
for Versions I-III); the Markov process we are
required to construct would however be nonhomogeneous.
5.3 Minimisation of 'Delayed' Detection Probability in an Independent Sequence.

Let us define a lapse of \( m \) (a positive constant) or more sampling epochs posterior to disorder as an unsafe delay or just delay. Now consider the problem posed in Section 5.2.4.

5.3.1 False-alarm Probability Not Restricted.

We are required to obtain the optimal sequential detection rule \( \tau^* \) for the problem in (9), i.e., in a manner which minimises

\[
wp(\tau < \theta) + P(\tau-\theta)^+ \geq m \]

where \( m > 0 \) and \( w \) weighs the relative importance of the two consequences of a declaration of disorder: false signal or delayed detection. We may also choose \( w \) so as to convert the sum \( wp(\tau < \theta) + P(\tau - \theta)^+ \geq m \) into an average cost.

Lemma 5.2  Suppose \( \mathcal{T}_m = \{ \tau \in \mathcal{T} : \tau \geq m \} \). Then,

\[
\inf_{\mathcal{T}_m}\{ wp(\tau < \theta) + P(\tau-\theta)^+ \geq m \} = \inf_{\mathcal{T}}\{ wp(\tau < \theta) + P(\tau-\theta)^+ \geq m \}.
\]

Proof. If \( \tau' = \max(\tau, m) \), it is then easily seen that
\begin{align*}
wP(\tau < \theta) + P((\tau - \theta)^+ \geq m) & \geq wP(\tau' < \theta) + P((\tau' - \theta)^+ \geq m) \\
\text{for every } \tau \in \mathcal{T}. \text{ This immediately proves the claim.} \end{align*}

If the optimal rule relative to the prior in (5) (instead of (8)) is sought, we would be required to consider \( \mathcal{T}_m \) and \( \{\tau : \tau < m\} \) separately.

For \( \tau \in \mathcal{T}_m \), the risk in (9) may be re-represented in terms of \( \pi_k^n \)'s as

\begin{align*}
wP(\tau < \theta) + P((\tau - \theta)^+ \geq m) \\
&= \mathbb{E} \left[ \sum_{n \geq m} 1_{\tau = n} \left( w \ 1_{n < \theta} + 1_{(n - \theta)^+ \geq m} \right) \right] \\
&= \mathbb{E} \left[ \sum_{n \geq m} 1_{\tau = n} \right] \mathbb{E} \left[ \sum_{k=0}^{n-m} \pi_k^n \right] \\
&= \mathbb{E} \left[ w(1 - \Pi_\tau) + \Pi_{\tau-m} \right].
\end{align*}

Now, substitute for \( \pi_k^n \) from (6). This enables us to restate the problem in (9) as

\begin{align*}
\inf_{\mathcal{T}_n} \mathbb{E} \left[ (1 - \Pi_\tau) (w - 1 - p \sum_{j=m+1}^{\tau} y_{ij}) \right] + 1
\end{align*}

where \( y_{ij} \) is defined by (7).
Optimal detection rule for \( m=2 \). We derive the optimal rule for online monitoring of \( X \) by taking \( m=2 \) in the problem specified by (11). This may be extended in a straightforward way to the case of a general \( m \). Results for the case \( m=1 \) being simpler to obtain, they are stated in Theorem 5.1 (\( m=1 \)). Our approach consists of constructing a Markov process appropriate to (11) and then using results stated in Section 1.3.

Let \( Y_n \) be defined as before, namely,

\[
Y_n = \frac{1}{q} \frac{p_1(X')}{p_0(X')} \quad (n \in \mathbb{N}).
\]

Then, the process \( Z = (Y_n, Y_{n-1}, \mathcal{F}_n : n=2,3,...) \) is Markov relative to \( (\mathcal{F}_n) \) and with a geometric \( \theta \), it is also time-homogeneous. This is so on account of the facts:

(i) since

\[
\pi_{n+1} = \frac{1 - q(1-\pi_n) Y_{n+1}}{1 - q(1-\pi_n) Y_{n+1} + (1 - \pi_n)},
\]

the stochastic sequence \( (Z_n, \mathcal{F}_n) \) has the property that, given \( X_{n+1} \), a fixed measurable function maps \( Z_n \) to \( Z_{n+1} \) (such a sequence has been called as transitive by Shiryaev, 1978);

(ii) \( X_{n+1} \) is independent of \( \mathcal{F}_n \) \( \forall n \).

Lemma 2.17 of Shiryaev (1978) shows that \( (Z_n, \mathcal{F}_n) \)
is Markov. Let $Z_n = z$. Take $m = 2$ in (11) and define
\[
h_2(z) = (1 - \pi_n) \left\{ w^{-1} - p(y_n + y_n y_{n-1}) \right\} + 1.
\]
Then, the problem before us is specified by
\[
\inf \mathbb{E} h_2(Z_n),
\]
a stopping problem for the Markov process $Z$.

Since $h_2(z_n) = \mathbb{E} \left[ w(1 - \pi_n) + \pi_n \mathbb{E}_{n-2} \right]$, $h_2$
is uniformly bounded on the two sides by $0$ and $\max(1, w)$.

A solution to (14) exists and is obtainable from the theory outlined in Section 1.3.3. Let $T$ and $Q$ be as defined there. Suppose $\nu$ denotes the probability measure induced by $Y$ under $P_0$ (i.e., by $p_1(X')/q_0(X')$).

Lemma 5.3 \[ T(1 - \pi_n) g(y_n) = q(1 - \pi_n) \int g(y) \, (dy) \]
for every $\mathcal{A}$-measurable function $g$, for which the expected value exists.

Proof. By partitioning the sample space as \[ \{ \theta \leq n+1 \} \, \{ \theta > n+1 \} \] and using independence, Bojdecki (1979) has shown that for a measurable $g : R_+ \rightarrow R_+$,
\[
T_g = \mathbb{E} \left[ g(Y_{n+1}) \cdot I_{\pi_n} \right] = q \int g(y) \left\{ y(1 - q(1 - \pi_n)) + (1 - \pi_n) \right\} \, \nu(dy),
\]
But it does hold for any $\mathcal{A}$-measurable function $g$.
the expected value exists finitely (this does not require a separate proof). Together with (13), this gives the desired result.

\[ \text{Theorem 5.1 (m=2)} \]

The sequential detection rule

\[ \tau^* = \inf \left\{ n \geq 2 : y_n(1+y_{n-1}) + \psi(y_n) \geq \frac{w-1}{p} \right\} \]

is optimal for the problem in (14) with a risk of \( 1 + r^*(0,0) \) provided that \( P( \tau^* < \omega) = 1 \).

(Functions \( y, r^* \) and \( \psi \) are respectively defined in (12), (15) and (17).)

\[ \text{Proof.} \] Let \( T \) and \( Q \), operate on \( h_2 \). Then

\[ h_2^*(z) = \lim_{N \to \infty} Q^N h_2(z) \]

is the infinal risk and the rule

\[ \tau^0 = \inf \{ n \geq 2 : h_2(z) = h_2^*(z) \} \]

is optimal for (14) provided that \( P( \tau^0 < \omega) = 1 \).

The loss of the \( \tau_n \)-time \( \tau_1 \) is

\[ \{ (1-x_n) (y_n(1+y_{n-1})) + 1 \} \]

which is uniformly bounded for every \( n \). Hence, Lemma 5.3 may be employed, taking \( 1-w-wp y_n(1+y_{n-1}) \), with \( y_{n-1} \) given, as the function \( g \). We then get

\[ Qh_2 = 1 + (1-x_n) \min \{ w-1-p y_n(1+y_{n-1}), q \int (w-1-p(1+y_n)) \nu(dy) \} \]

Set \( r_1(u,v) = w-1-pu(1+v) \) and, for \( k \in \mathbb{N} \),

\[ r_{k+1}(u,v) = \min \{ r_1, q \int r_k(y,u) \nu(dy) \} \]
It is then known that \[ r^*(y_n, y_{n-1}) = \lim_{n \to \infty} Q^n r_1(y_n, y_{n-1}) \]

is a maximal submartingale relative to \( r_1(y_n, y_{n-1}) \)
(i.e., submartingale \( r^* \) majorises all submartingales
which are majorised by \( r_1 \)) and satisfies uniquely the
equation
\[ r^*(u, v) = \min \left\{ r_1(u, v), q \int r^*(y, u) \, \nu(dy) \right\} \]

Hence, \( h^*_2(y_n, y_{n-1}, \tau_n) = 1 + (1-\tau_n) r^*(y_n, y_{n-1}) \)
and \( \tau^0 \) becomes
\[ \tau^0 = \inf \left\{ n \geq 2 : r_1(y_n, y_{n-1}) \leq q \int r^*(y, u) \, \nu(dy) \right\} \]

On setting
\[ \psi(u) = \psi(u ; w, p) = \frac{q}{p} \int r^*(y, u) \, \nu(dy) \]
we get
\[ \tau^0 = \inf \left\{ n \geq 2 : y_n(1+y_{n-1}) + \psi(y_n) \geq \frac{w-1}{p} \right\} \]

Note that \( \psi \) is free from \( \tau_n \) and is a fixed function
(i.e., it does not vary with \( n \)). Finiteness of \( \tau^0 \)
is established under a condition on \( w \) in Lemma 5.4
below. Assume that \( P(\tau^0 < \infty) = 1 \). Then, \( \tau^0 \)
is Bayes and its risk is obtained by evaluating
\[ \mathbb E h^*_2(Y_2, Y_1, \Pi_2) \]. Proceeding on lines parallel to
Bojdecki's yields \( 1 + r^*(0, 0) \) for it. This completes
the proof of the theorem.

**Lemma 5.4** Suppose that \( s \) is defined by

\[ s = \inf \left\{ s' : P \left( \frac{P_1(X')}{P_0(X')} \leq s' \right) = 1 \right\} \]
Then, \( P(\tau^0 < \omega) = 1 \) if
\[
w < 1 + \frac{p(q+q+1)}{q}
\]

**Proof.** It is easily shown using the definition of \( a \) that \( \nu \) is also concentrated on \([0, a] \).

Assume \( a < \omega \). Then, for every \( n \),
\[
Y_n + Y_n Y_{n-1} \leq \frac{a}{q} + \frac{\alpha^2}{q^2} \quad (w.p. 1)
\]
and by induction,
\[
q \int r^*(y,u) \nu(dy) \leq q \left\{ w-1-p(\frac{a}{q} + \frac{\alpha^2}{q^2}) \right\}
\]
Let \( \sigma \) be defined by
\[
\sigma = \inf \left\{ n \geq 2 : Y_n + Y_n Y_{n-1} \geq 1 \left[ w-1-q \{ w-1-p(\frac{a}{q} + \frac{\alpha^2}{q^2}) \} \right] \right\}
\]
\[
= \inf \left\{ n \geq 2 : Y_n + Y_n Y_{n-1} \geq w-1+q \left( \frac{a}{q} + \frac{\alpha^2}{q^2} \right) \right\}
\]
Observe that \( \sigma \in \mathcal{F} \) and by virtue of (19),
\( \sigma(\omega) > \tau^0(\omega) \) so that it is enough to show a.s. finiteness of \( \sigma \). The latter is accomplished by using Levy's extension of Borel-Cantelli lemma to arbitrarily dependent processes (see \( \text{Corollary 5.29 in Breiman, 1968} \)). Define, for every \( n \in \mathbb{N} \),
\[
A_n = \left\{ \omega : Y_n(1 + Y_{n-1}) \geq (w-1) + q(\frac{a}{q} + \frac{\alpha^2}{q^2}) \right\}
\]
Then, \( A_n \in \mathcal{F}_n \) and
\[
P(A_{n+1} \mid \mathcal{F}_n) = P \left[ Y_{n+1} (1 + Y_n) \geq w-1+q \left( \frac{a}{q} + \frac{\alpha^2}{q^2} \right) \mid \mathcal{F}_n \right]
\]
\[
\geq P \left[ Y_{n+1} \geq w-1+q \left( \frac{a}{q} + \frac{\alpha^2}{q^2} \right) \right]
\]
which is positive if

\[ w^{-1} - q \left( \frac{s}{q} + \frac{s^2}{q^2} \right) < \frac{q}{q} , \]

i.e., if \( w < 1 + \frac{s(s+q+1)}{q} \). Thus, \( P(A_{n+1} | \mathcal{F}_n) > 0 \) (\( \forall n \)) and hence

\[ P \left( \sum_{n=1}^{\infty} P(A_{n+1} | \mathcal{F}_n) = \infty \right) = 1 \]

whenever (18) holds. The result quoted above now gives \( P(A_n \text{ occurs } 1.o.o.) = 1 \), thus establishing a.s. finiteness of \( \sigma \) and hence of \( \tau^0 \).

When \( a = \infty \), the result \( P(\tau^0 < \infty) = 1 \) is true without any condition and can be proved by repeating the corresponding arguments in Bojdecki(1979). □

Remark 1. An alternative (a weaker one) to (18) is

\[ P \left( \frac{p_1(X'_1)}{p_0(X'_n)} > q \frac{w^{-1}}{p} \right) > 0 \]

which ensures that \( r_1 (Y_n, Y_{n-1}) < 0 \) on a set of positive probability, denote this set by \( A \). Then (19) gives,

(21) \[ w^{-1} - p Y_n(1 + Y_{n-1}) \geq w^{-1} - p(\frac{s}{q} + \frac{s^2}{q^2}) \quad (s < \infty) , \]

and for sufficiently large \( Y \) values in \( A \), the IHS of (21) is negative and arbitrarily close to the bound there. In particular, it is larger than \( q(w^{-1} - p(\frac{s}{q} + \frac{s^2}{q^2})) \).
which in view of (20) implies the required result. The case \( a = \infty \) can be dealt with as stated earlier.

**Remark 2.** In the generally more important case of \( w \leq 1 \) (this is the case when it is 'cheaper' to check whether a signal is false than to go on with the process in disorder state), the condition of the Lemma is automatically met. Indeed, from practical viewpoint, the condition may hardly be considered restrictive.

**Theorem 5.1 \( (m=1) \).** The sequential detection rule

\[
\tau^* = \inf \{ n \geq 1 : y_n \geq \frac{w-1-\gamma^*}{p} \}
\]

where the constant \( \gamma^* \) is the unique solution of

\[
\gamma^* = q \int \min (w-1-qy, \gamma^*) \nu(dy)
\]

minimises

\[
wP(\tau < \theta) + P(\tau > \theta)
\]

with \( 1 + \gamma^* \) as its risk if \( P(\tau^* < \infty) = 1 \).

**Proof** proceeds exactly on the lines of the case \( m=2 \) with some of the calculations simpler. \( \square \)

The condition \( P(\tau^* < \infty) = 1 \) in Theorem 5.1 \( (m=1) \) is satisfied if \( w > 1-a \). Beginning with the method used in proving Lemma 5.4 and noting that \( Y_n \)'s are independent,
we may apply Borel–Cantelli Lemma to complete the proof in the \( (m=1) \)-case.

**Remark.** The problem specified by (9) may be represented, in the case \( w=1 \) and \( m=1 \), as

\[
\sup \ P(1 \tau - \theta) < 1
\]

which corresponds to maximising the probability of instantaneous detection. In this particular case therefore, our problem coincides with Bojdecki's. This observation is verified in the solution of the example in Section 5.3.3.

**Optimal rule for a general \( m \).** Extension to any \( m(\geq 1) \) is apparent; \( \tau^* \) is given by

\[
\tau^* = \inf \left\{ n \geq m : \sum \psi(y_j, \ldots, y_{n-m+1}) \geq \frac{w-1}{p} \right\}
\]

where \( \psi_m \) is defined analogous to \( \psi \) in (17). It is worthwhile noting that the optimal rule for our problem (given \( m \)) is essentially myopic in the sense that at any instant, stopping (or declaration of disorder) is decided on the basis of only the latest \( m \) observations.

**5.3.2 False-alarm Probability Controlled.**

We now seek the optimal rule in the smaller but often more important class \( \mathcal{F}(\alpha) \) of detection.
rules \( \mathcal{T}(\alpha) = \{ \tau \in \mathcal{T} : \alpha(\tau) \leq \alpha \} \) for a specified \( \alpha, 0 < \alpha < 1 \). Evidently, such a restriction keeps the total cost of search due to false signals to a desired level. We have already noted that the work of Bojdecki (1979) does not incorporate this provision. Procedures based on cusum also do not in general allow such a restriction. However, Khan (1979) achieves this through a generalised cusum procedure. In an important special case, his procedure is equivalent (Page, 1964) to

\[
\tau' = \inf \{ n \geq 1 : W_n \geq h \}
\]

where \( h(> 0) \) is a constant, \( W_0 = 0 \) and

\[
W_n = \max \left\{ 0, W_{n-1} + \log \frac{p_1(x_n)}{p_0(x_n)} \right\} \quad (n \geq 1),
\]

In his paper, Khan establishes an optimal property of \( \tau' \). Note that \( E_0 \tau' \) and \( E_1 \tau' \) are respectively ARL of \( \tau' \) under \( p_0 \) and \( p_1 \) (by definition of ARL).

Employing the well known Wald approximations, he gets

\[
\text{ARL}_0 \approx \frac{e^h - h - 1}{I(p_0 \mid p_1)} = \delta_0(h), \text{ say},
\]

\[
\text{ARL}_1 \approx \frac{e^{-h} + h - 1}{I(p_1 \mid p_0)} = \delta_1(h), \text{ say}
\]

where

\[
I(p_0 / p_1) = \int p_1(x) \log \frac{p_1(x)}{p_0(x)} \, d\mu(x),
\]
and defined similarly is $I(P_1 | P_0)$ (these are Kullback-Liebler information numbers). Optimality implies that if $h$ is chosen to satisfy $A_0(h) \geq \gamma$, $A_1(h)$ is minimum subject to $A_0(h) \geq \gamma$ ($\gamma$ is a preassigned constant).

Version III introduced in this thesis is in a similar vein (but in a Bayesian setup). As remarked earlier, when $\theta$ can be assumed, a priori, to be geometric, Version III is equivalent to restricting $\alpha(\tau)$. The optimal rule for this version is: $\inf \left\{ n : \pi_n \geq \gamma \right\}$, and $\inf \left\{ n : \pi_n \geq \gamma(\alpha) \right\}$ in the $\alpha$-restricted case. But this procedure bristles with the difficulty of determining $\gamma$ exactly.

In the following, we take up the design of rules which minimise delayed detection probability in the $\alpha$-restricted class.

The 'm = 1' case. The problem

\begin{equation}
\inf_{\mathcal{Y}(\alpha)} P \left( (\tau - \theta)^+ \geq 1 \right)
\end{equation}

amounts to maximising the probability of instantaneous detection with false-alarm probability kept within a preassigned level $\alpha$. The solution $\tau^*_\alpha$ comes out surprisingly simple. In seeking to determine $\tau^*_\alpha$
from the knowledge of $\tau^*$ (see (22)), where
\[
P(\tau^* = \theta) = \sup_{\mathcal{F}} P(\tau = \theta), \quad P(\tau^*_\alpha = \theta) = \sup_{\mathcal{F}(\alpha)} P(\tau = \theta),
\]
we are prompted to check whether the form of $\tau^*$ is preserved by restriction to $\mathcal{F}(\alpha)$. The answer is 'yes', as seen in Theorem 5.2 below. It may be observed that just limiting $\alpha(\tau)$ to $\alpha$ or less eliminates the need to solve equation (23).

**Theorem 5.2** The solution to (25), for every fixed $\alpha(<\theta)$ is
\[
\tau^*_\alpha = \inf \{ n \geq 1 : y_n \geq k(\alpha) \}
\]
where $k(\alpha)$ is the smallest value satisfying
\[
P_0\left( Y_1 \geq k(\alpha) \right) = \frac{\alpha p}{q(1-\alpha)}.
\]

Proof. Consider the set of rules \{$\tau_k; \ k > 0$\},
\[
\tau_k = \inf \{ n \geq 1 : y_n \geq k \}.
\]
The false-alarm probability of $\tau_k$, for a fixed $k$, is
\[
\alpha(\tau_k) = P(\tau_k < \theta)
\]
\[
= P( Y_1 \geq k \cap \theta > 1) P(\theta > 1)
\]
\[
+ \sum_{n=2}^{\infty} \sum_{j=1}^{n-1} P(Y_n \geq k|\theta > n) P(Y_j < k|\theta > n) P(\theta > n)
\]
\[
= q P_0(Y_1 \geq k) + \sum_{n=2}^{\infty} \left( P_0(Y_1 \geq k) q^n \right) P_0(Y_1 < k)^{n-1}
\]
On simplification,
The probability of instantaneous detection which is to be maximised may be shown to be

\[ P(\tau_k = 0) = \frac{p P_0(\tau_k Y_1 \geq k)}{1 - q P_0(Y_1 < k)} \]

Since the LHS of (26) can not exceed unity, the specified value of \( \alpha \) must satisfy the condition \( \alpha < q \). The following observations complete the proof.

(i) \( t^* \) and rules of the form of \( \tau_k \) satisfy (26),

(ii) the RHS of (27) is a strictly increasing function of \( \alpha \) (for given \( p \)),

(iii) the LHS of (27) is at least left continuous and nonincreasing in \( k \),

(iv) as is implied by (ii) and (iii), corresponding to a fixed \( \alpha \), we can find the smallest \( k(\alpha) \) satisfying

\[ \alpha(\tau_k(\alpha)) = \alpha. \]

Remark. The condition \( \alpha < q \) does not limit the scope of the theorem in a serious way since \( q \) is generally 'close' to 1.
The 'm=2' case. We seek a solution for

\[
\inf_{\mathcal{F}(\alpha)} P(\tau - \theta)^{+} \geq 2
\]

having known \( \tau^* \), the optimal rule in \( \mathcal{F} \), stated in Theorem 5.1(m=2). The false-alarm probability of \( \tau^* \) is

\[
\alpha(\tau^*) = \sum_{n \geq 2} P(\tau^* = n, \theta > n)
\]

\[= P(Y_2(1+Y_1) + \psi(Y_2) \geq \frac{1-w}{p}, \theta > 2)\]

\[+ \sum_{n \geq 3} P(Y_k(1+Y_{k-1}) + \psi(Y_k) \leq \frac{1-w}{p}, 2 \leq k \leq n-1,\]

\[Y_n(1+Y_{n-1}) + \psi(Y_n) \geq \frac{1-w}{p}, \theta > n)\]

Assume that (i) \( \tau^* \) can be written as

\[\tau^* = \inf \{ n \geq 2 : \phi(Y_n, Y_{n-1}) \geq c^*(w) \}\]

where \( \phi \) is a function not depending on \( n \), and (ii) by limiting \( \alpha(\tau) \) to at most \( \alpha \), we can find a \( w_\alpha \) as the smallest \( w \) satisfying the equation obtained by replacing \( \alpha(\tau^*) \) by \( \alpha \) in (31). (Assumption (ii) is not essential since, just as in the discussion of optimal property 2 in Section 2.6, we can establish that the rule \( \tau_{c_0} = \inf \{ n \geq 2 : \phi(Y_n, Y_{n-1}) \geq c_0 \} \) minimises \( P(\tau = \theta)^{+} \geq 2 \) in the class of rules whose false-alarm probability does not exceed that of \( \tau_{c_0} \).) The arguments in Shiryaev (1963, 1978) now apply giving the optimal rule for (30). The threshold \( c^* \) of \( \tau^* \)
depends in general on \( w \). The risk
\[
\inf \{ w \tau \leq 0 \} + \mathbb{P}( (\tau - \theta)^+ \geq 2) \]
is obviously concave in \( w \) and zero for \( w = 0 \). Hence \( c^*(w) \) is continuous and nonincreasing. It is then possible to find a \( w^*(c^*) \) as the smallest \( w \) for which \( c^*(w) = c^* \). In particular, we may find \( w_{\alpha} \) as the smallest \( w \) such that

\[
\inf \{ n \geq 2 : \phi(y_n, y_{n-1}) \geq c \} \quad \text{such that} \quad \alpha(\tau_c) = \alpha. 
\]

By virtue of Theorem 5.1\( (m=2) \), \( \tau_{\alpha}^* \) is optimal in \( T_2 \) (i.e., for the problem in (33)) provided \( w \) satisfies the condition of Lemma 5.4. We shall assume this to hold. Furthermore, since \( \tau_{\alpha}^* \in T_2(\alpha) \subset T \), it is optimal in \( T(\alpha) \). Thus,

\[
 w_{\alpha} \mathbb{P}(\tau_{\alpha}^* > \theta) + \mathbb{P}( (\tau_{\alpha}^* - \theta)^+ \geq 2) \\
\leq w_{\alpha} \mathbb{P}(\tau > \theta) + \mathbb{P}( (\tau - \theta)^+ \geq 2),
\]

Since \( \mathbb{P}(\tau_{\alpha} > \theta) = \alpha \geq \mathbb{P}(\tau > \theta) \), we must necessarily
have \( P((\tau_\alpha^* - \theta)^+ \geq 2) \leq P((\tau - \theta)^+ \geq 2), \quad \forall \tau \in \mathcal{T}(\alpha). \)

Thus, \( \tau_\alpha^* \) defined analogous to \( \tau^* \) of Theorem 5.1 \((m=2)\) with \( w \) replaced by \( w_\alpha \) determined by solving (32) minimises \( P((\tau - \theta)^+ \geq 2) \) while its false-signal probability does not exceed \( \alpha \).

### 5.3.3. Example: Shift in the Mean of an Exponential Law

Suppose that the parameter of an exponential distribution changes from \( \lambda_0 \) to \( \lambda_1 \) (both known) at an unknown instant \( \theta \) during sequential observation. Let our object be to devise a rule that minimises \( w \, P(\tau < \theta) + P(\tau > \theta), \) \( w \) given. This amounts to choosing \( m=1 \) in (9). Here,

\[
y = \frac{1}{q} \frac{p_1(x)}{p_0(x)} = \frac{\lambda_1}{\lambda_0} \exp\{-(\lambda_1 - \lambda_0)x\} \quad (x > 0),
\]

and if \( f \) denotes the pdf of \( p_1(x')/q \, p_0(x') \), standard methods of getting \( f \) yield

\[
f(y) = \frac{1}{c-1} \left( \frac{q}{c} \right)^{\frac{1}{c-1}} y^{\frac{1}{c-1}} - 1 \quad (0 < y < \frac{q}{c}, \ c > 1)
\]

\[
f(y) = \frac{1}{1-c} \left( \frac{q}{c} \right)^{\frac{1}{c-1}} y^{\frac{1}{c-1}} - 1 \quad \left( \frac{q}{c} < \frac{q}{\alpha}, \ c < 1 \right)
\]

where we have written \( c \) for \( \lambda_1/\lambda_0 \). Note that \( f \) is the density of the \( \nu \)-measure introduced earlier. As per Theorem 5.1 \((m=1)\), the Bayes sequential detection rule (wrt a geometric \( \theta \)) is given by
that is,
\[ \tau^* = \inf \left\{ n \geq 1 : \gamma_n \geq \frac{w-1-\gamma^*}{p} \right\} \]

where, by (23), \( \gamma^* \) is the unique solution of
\[ \gamma = q \left( (w-1-py) f(y) dy + q \int \gamma f(y) dy \right) \]
\[ \{ y > \frac{w-1-\gamma}{p} \} \quad \{ y \leq \frac{w-1-\gamma}{p} \} \]

**Case (i) :** \( c > 1 \) (i.e., there is an increase in \( \lambda \)).

On evaluating the integrals in (37) using the \((c > 1)\)-expression of (34), we get
\[ \gamma = q(w-1) - p - p(\frac{q}{c}) \frac{c}{c-1} \left( \frac{w-1-\gamma}{p} \right) \frac{c}{c-1} \]
which, on setting \( k = (w-1-\gamma)/p \), becomes
\[ k = w + (c-1)(\frac{qk}{c}) \frac{c}{c-1} \quad (c > 1). \]

This equation possesses a unique solution that directly gives us the threshold \((w-1-\gamma^*)/p\) for \( \tau^* \) (see (35)).

**Case (ii) :** \( c < 1 \). Integration in (37) employing the \((c < 1)\)-expression from (34) and substitution of \( k \) for \((w-1-\gamma)/p \) gives
\[ k = \frac{w-1}{p} + \frac{1}{p} \frac{1-c}{c} \left( \frac{qk}{c} \right) \frac{1}{c-1} \quad (c < 1). \]

Its unique solution provides the threshold in (35) for the \((c < 1)\)-case.
As remarked at the end of Section 5.3.1, in the \((w=1, \alpha=1)\)-case, our solution should match Bojdecki's. This is easily verified. For instance, when \(c \in [q, 1)\), we get from (35) and (38)

\[ \tau^* = \inf \{ n \geq 1 : y_n \geq k \} = \inf \{ n \geq 1 : x_n \geq \frac{1}{\lambda} \log \frac{q(1-c)}{c(1-q)} \} \]

whose critical value \(\frac{1}{\lambda} \log \frac{q(1-c)}{c(1-q)}\) is nonpositive, thus implying that \(\tau^* = 1\) for a disorder such that \(q \leq (\lambda_1/\lambda_0) < 1\). It may be noted that stopping after one observation would not be worse than \(\tau^* = 0\) since the loss being minimised excludes sampling costs.

In the \(\alpha\)-restricted case, equations (26) and (28) need to be evaluated. Using (34),

\[ P_0 \left( X_1 \geq k \right) = 1 - \left( \frac{qk}{c} \right)^{\frac{1}{c-1}} \quad (c > 1), \]
\[ = \left( \frac{qk}{c} \right)^{\frac{1}{c-1}} \quad (c < 1), \]

Repeating the calculations made in deriving (34), now under \(P_1\), we get

\[ P_1 \left( X_1 \geq k \right) = 1 - \left( \frac{qk}{c} \right)^{\frac{c}{c-1}} \quad (c > 1), \]
\[ = \left( \frac{qk}{c} \right)^{\frac{c}{c-1}} \quad (c < 1). \]

When substituted into (26) and (28), we obtain
\[ \alpha(\tau_k) = q \left( \frac{ak}{c} \right)^{\frac{1}{c-1}} \cdot \left[ p + q \left( \frac{ak}{c} \right)^{\frac{1}{c-1}} \right] \]

(39)

\[ P(\tau_k = 0) = p \left( \frac{ak}{c} \right)^{\frac{c}{c-1}} \cdot \frac{c}{1 - q \left( \frac{ak}{c} \right)^{\frac{1}{c-1}}} \]

The advantage of controlling $\alpha$ is illustrated below.

**Numerical Illustration.** Let $q = \frac{3}{4}$ and $c = \frac{1}{2}$. Let us take $w = 1$ in (24) and determine the corresponding optimal rule $\tau^*$. For the threshold $\gamma^*$ we get

\[ \gamma^* = \left( \frac{1-c}{p} \right)^{1-c} \cdot \frac{c}{\left( \frac{a}{q} \right)} = \frac{2}{\sqrt{3}} \]

giving $\tau^* = \inf n \geq 1 : \gamma_n \geq \frac{2}{\sqrt{3}}$ as the optimal rule when $\alpha(\cdot)$ is not controlled. Substituting for $q$ and $c$ in (39), we see that $\alpha(\tau^*) = \frac{1}{2}$ so that $N(\tau^*) = 1$, and $P(\tau^* = 0) = 0.1925$. Now suppose that it is not costly to verify false signals and therefore we may allow $\alpha(\tau)$ to be as high as $\frac{2}{3}$ ($= \kappa$). Then $N(\tau_{\kappa}) \leq 2$ and the critical value $k(\alpha)$ of $\tau^*_\kappa$ is $\frac{\sqrt{2}}{3}$ (obtained by solving (29)) and $P(\tau^* = 0) = 0.4093$.

Thus, when a higher $\kappa$ can be tolerated, the 'reward' can be more than double and the corresponding rule is determinable more quickly by Theorem 5.1 (m=1) than, e.g., the rule given by Bojdacki.
5.4 Detecting a Shift in the Mean of an Autoregressive Process.

Most of the work concerning optimal online detection of disorder in a random sequence assumes independence of observations. In this section and in the one following it, we consider disorder in a certain dependent sequence and derive the optimal detection rule taking the risk function to be of the type described in Version II. The equivalence to Version I of Versions II and III continues to hold as long as the distribution of the disorder moment $\theta$ is assumed to be geometric.

5.4.1 The Model.

Let the process $(X_n)^0$ defined on $(\Omega, \mathcal{A})$ which is being observed sequentially be a normal Markov sequence with a change in mean likely to occur sometime during the evolution or initially. Let $\theta$ have the prior distribution stated in (5). Suppose $\theta = k$. Assume that $X_0 = 0$ and that the model is given by

when $k = 0$, \[ X_1 = \mu + \sigma + Y_1 \]

\[ X_n = \mu + \sigma + \rho(X_{n-1} - \mu - \sigma) + Y_n \]

\[ (n=2,3,\ldots) \],
when \( k > 0 \)

\[
X_n = \mu + \rho(X_{n-1} - \mu) + \gamma_n \quad (n=1,2,\ldots,k-1)
\]

(40)

\[
= \mu + \sigma + \rho(X_{n-1} - \mu) + \gamma_n \quad (n=k)
\]

\[
= \mu + \sigma + \rho(X_{n-1} - \mu - \sigma) + \gamma_n \quad (n=k+1,\ldots),
\]

where (i) \( \gamma_n \sim \mathcal{N}(0,1) \) \((n \in \mathbb{N})\) are independent

(ii) \( |\rho| < 1 \) and \( -\infty < \mu + \sigma \). Thus the model assumes that disorder results from a shift of magnitude \( \sigma \) (known) in the mean of an autoregressive process of order 1 (given \( \theta = k \), \( \mathbb{E} X_n = \mu \) for \( n < k \) and \( \mu + \sigma \) for \( n \geq k \)). The parameters \( \mu \) and \( \rho \) are assumed known.

As is frequently done for a Markov model like the one above, we transform \( X \) to \( Z \) by taking

\[
Z_1 = X_1, \quad Z_n = X_n - \rho X_{n-1} \quad (n = 2,3,\ldots).
\]

Let \( Z(n \mid k) \) denote the sample \((Z_1, Z_2, \ldots, Z_n)'\) when \( \theta = k \). Since the Jacobian of transformation is unity

(41) \[ Z(n \mid k) \sim \mathcal{N}(\mu(n \mid k), I(n)) \],

i.e., it has a multinormal distribution with mean vector

(42) \[ \mu(n \mid k) = (\mu, \rho\mu, \ldots, \rho^{k-1}\mu, \rho^k(\mu + \sigma), \ldots, \rho^k(\mu + \sigma))' \]

\[ (k > 1) \]

\[ = (\mu + \sigma, \rho(\mu + \sigma), \ldots, \rho^k(\mu + \sigma))' \]

\[ (k=0 \text{ or } 1) \]

and dispersion matrix \( I(n) \), the \( n \times n \) identity matrix.

In (42), we have written \( \overline{\rho} \) for \( 1 - \rho \). Evidently
$Z_1, Z_2, \ldots, Z_n$ are mutually independent for every $n$, and each is normal with a mean that depends upon the value of $\theta$.

5.4.2 The Posterior Process.

Let $P^\theta$ be as described by (1) and (2), and let $\Pi_n = P^\theta(\theta \leq n | \mathcal{F}_n)$. Given $X(n) = x(n)$, we can compute $z(n)$ and use Bayes formula to get $\tau_n$ (see (43) and (44) below). Let $f(z(n)|\theta)$ stand for the joint density of $Z(n|\theta)$. Then, for $1 \leq k \leq n$,

\[
P^\theta(\theta = k | \mathcal{F}_n) = \frac{(1-\pi)pq^{k-1} f(z(n)|\theta)}{\pi f(z(n)|\theta) + (1-\pi) \sum_{i=1}^{n} pq^{i-1} f(z(n)|\theta) + \sum_{j=n+1}^{\infty} pq^{j-1} f(z(n)|\theta > n)}
\]

and when $k = 0$, $P^\theta(\theta = 0 | \mathcal{F}_n) \propto \pi f(z(n)|\theta)$, Notice that $f(z(n)|\theta) = f(z(n)|\theta > n)$, $\forall n \geq 1$. Now, set

\[
y_k^n = q^{k-n-1} \frac{f(z(n)|\theta)}{f(z(n)|\theta > n)} (1 \leq k \leq n),
\]

\[
y_k^n = 1 (k > n).
\]

In terms of $y_k^n$'s,

\[
P^\theta(\theta = k | \mathcal{F}_n) = \frac{py_k^n}{1 + \frac{\pi}{1-\pi} y_1^n + p \sum_{j=1}^{n} y_j^n} (1 \leq k < n)
\]

\[
P^\theta(\theta = k | \mathcal{F}_n) = \frac{pq^{k-n-1}}{1 + \frac{\pi}{1-\pi} y_1^n + p \sum_{j=1}^{n} y_j^n} (k \geq n)
\]
and on summing,

\[ 1 - \pi_n = \frac{1}{1 + \frac{r}{1-\pi} y_1^n + p \sum_{j=1}^{n} y_j^n} \quad (n \geq 1). \]

Lemma 6.5 The process \( \{ \Pi_n \} \) is Markov wrt \( \{ \mathcal{F}_t \} \), and further \( \{ \Pi_n \} \) is homogeneous.

Proof From (44), we may write, for \( n \geq 1 \),

\[ (1 - \pi_{n+1})^{-1} = 1 + \frac{r}{1-\pi} y_1^{n+1} + p \sum_{j=1}^{n} y_j^{n+1} + p y_{n+1}^{n+1}. \]

Now, denote by \( \phi(1, 1, \theta, \gamma) \) the density of \( \mathcal{M}(\theta, \gamma) \). Then, \( y_k^n \)'s defined in (43) may be expressed, in the light of (41), as

\[ y_1^1 = \frac{1}{q} \frac{\phi(z_1 1 \mu + \delta, 1)}{\phi(z_1 1 \mu, 1)}, \]

and when \( n \geq 2 \),

\[ y_{k}^{n+1} = \frac{1}{q} y_k^{n} \frac{\phi(z_{n+1} 1 \beta (\mu + \delta), 1)}{\phi(z_{n+1} 1 \beta \mu, 1)} \quad (1 \leq k \leq n) \]

\[ = \frac{1}{q} \frac{\phi(z_{n+1} 1 \beta \mu + \delta, 1)}{\phi(z_{n+1} 1 \beta \mu, 1)} \quad (k = n+1) \]

\[ = q^{k-n-2} \quad (k > n+1). \]

Employing this in (45), we get
\[(1-\pi_{n+1})^{-1} = 1 + \left( \frac{x}{1-\pi} y_{1}^{n} + p \sum_{j=1}^{n} y_{j}^{n} \right) \frac{\phi(z_{n+1} \mid \bar{\rho}(\mu + \delta), 1)}{q(z_{n+1} \mid \bar{\rho}, 1)} + p \frac{\phi(z_{n+1} \mid \bar{\rho}(\mu + \delta), 1)}{q \phi(z_{n+1} \mid \bar{\rho}, 1)} \]
\[
= 1 + \frac{\epsilon_{n}}{1-\pi_{n}} \frac{\phi(z_{n+1} \mid \bar{\rho}(\mu + \delta), 1)}{\phi(z_{n+1} \mid \bar{\rho}, 1)} + p \frac{\phi(z_{n+1} \mid \bar{\rho}(\mu + \delta), 1)}{q \phi(z_{n+1} \mid \bar{\rho}, 1)}
\]

Also,
\[(1-\pi_{1})^{-1} = 1 + (\frac{x}{1-\pi} + p) \frac{1}{q} \frac{\phi(z_{1} \mid \bar{\rho}(\mu + \delta), 1)}{\phi(z_{1} \mid \bar{\rho}, 1)}
\]

Let \(\mathcal{G}_{n}\) be the \(\sigma\)-field induced by \(Z(n)\), \(n \geq 1\).
Define \(\mathcal{G}_{0} = (\Omega, \phi)\). Since the transformation of \(X(n)\) to \(Z(n)\) is 1-1 and Borel measurable, \(\mathcal{G}_{n} = \mathcal{F}_{x}(\forall n)\).

We are now in a position to infer that

(i) since \(\pi_{n+1}\) can be computed using \(\pi_{n}\) and
\[z_{n+1}, (\Pi_{n}, \mathcal{F}_{n})_{0}^{\omega}\] is a transitive sequence,
equivalently, \((\Pi_{n}, \mathcal{F}_{n})_{0}^{\omega}\) is transitive
(see the comment following (13)),

(ii) independence of \(Z_{n+1}\) and \(\mathcal{F}_{n}\) \((\forall n \geq 0)\) gives
\[P_{\pi}(Z_{n+1} \in B \mid \mathcal{F}_{n}) = P_{\pi}(Z_{n+1} \in B \mid \pi_{n})
\]
for every Borel set \(B\) on the state space of \(Z\).
These two observations imply, by virtue of Lemma 2.17 of Shiryaev (1978), that \((\Pi_n, \mathcal{F}_n)^\infty_1\) is Markov. Further, since a fixed function yields \(\pi_n\) from \(\pi_{n-1}\) (see (48)) for \(n \geq 2\), \((\Pi_n)^\infty_1\) is homogeneous.

**Remark.** The recursive expressions for \(\pi_n\) in (48) above are also useful in successive computation of the posterior probability.

5.4.3 The \(\alpha\)-Unrestricted Case.

Having proved that \((\Pi_n, \mathcal{F}_n)^\infty_1\) is a Markov sequence, the problem of online detection of a shift of mean in an autoregressive process can be solved analogous to disorder in an independent sequence (cf. Shiryaev, 1978). The problem before us is specified by

\[
\inf \left\{ w P^\tau (\tau < \theta) + E^\tau (\tau - \theta)^+ \right\}
\]

Expressed in terms of \(\pi_n\)'s, it is

\[
\inf \{ E^\tau w(1 - \Pi_\tau) + \sum_{k=0}^{\tau-1} \Pi_k \} \quad (\tau > 0),
\]

This is a stopping problem with 'sampling costs':

\(\pi_n\) is the 'sampling cost' at the \(n\)-th stage and on \(\{\tau = n\}\) the 'terminal decision loss' is \(w(1-\pi_n)\).
Theorem 5.3  The rule
\[ t^* = \inf \{ n \geq 2 : \pi_n > \gamma \} \]
is optimal for the problem in (50) with \( T \) replaced by
\[ T_2 = \{ t \in T : t \geq 2 \text{ if } P^w (t^* < \omega) = 1 \}. \]

Remark. The requirement that at least two observations be taken is to avoid nonhomogeneity at \( n = 1 \). If we consider the problem
\[ \inf P^w \left[ w(1 - \Pi_t) + \sum_{k=0}^{\tau-1} \Pi_k \right] \]
where \( P^w \) is the probability measure determined by \( P^w \) when \( \Pi_1 = \pi_1 \) is given, then stopping at \( n = 1 \) may be included.

Proof. Define \( g(\pi) = w(1-\pi) \) and \( Qg(\pi) = \min \{ g(\pi), \pi + Tg(\pi) \} \). Then, as on earlier such occasions, \( \lim_{N} Q^N g(\pi) \) is the infimal 'risk' for the problem in (50) and
\[ t^0 = \inf \{ n \geq 2 : \lim_{N} Q^N g(\pi_n) = g(\pi_n) \} \]
would be optimal in the class \( T_2 \). Concavity and continuity of \( \lim_{N} Q^N g(\pi) \) and of \( g(\pi) \) on \((0,1)\) permit us to write
\[ t^0 = \inf \{ n \geq 2 : \pi_n \geq \gamma \} , \]
where \( 0 < \gamma < 1 \). Finiteness of \( t^0 \) may now be
established by showing that \( r_n \) can take values arbitrarily close to 1 for some \( n < \infty \). Towards this, define

\[
A_n = \{ \emptyset \leq n \}, \quad Y_n = 1_{A_n}.
\]

Then, Levy's theorem (see Cor. 1 to Lemma 7.4.2 in Chow and Teicher, 1979) applies since \( Y_n \) is a bounded random variable \( (Y_n) \) with \( Y_n \to 1_n \) (a.s. - \( P^\omega \)) as \( n \to \infty \). This gives

\[
E^\omega (Y_n | \mathcal{F}_n) \to E(1_n | \mathcal{F}_0)
\]

where \( \mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n) \). Hence \( \tau^* \) is optimal for (50) with \( \mathcal{T} \) replaced by \( \mathcal{T}_2 \).

Wiener process approximations may be used for determining \( \gamma \), the threshold of \( \tau^* \). In the \( \alpha \)-restricted case an upper bound on \( \gamma \) continues to be available (just as in the independent sequence case).

Smith (1975) suggests, while his main concern is with the fixed-sample version, an "informal sequential procedure" which consists of "... stopping if the posterior odds became too small, in some appropriate sense". Our study above formalises this suggestion for the particular disorder detection problem and makes clear the kind of optimality achievable with it.
5.4.4 The $\alpha$-Restricted Case.

Let $\mathcal{F}_2(\alpha) = \{ \tau \in \mathcal{T} : t \geq 2, \alpha(\tau) \leq \alpha \}, 0 < \alpha < 1$.

Suppose we limit our search for the optimal rule for the problem in (50) to $\mathcal{F}_2(\alpha)$. Our object then is to solve

$$\inf_{\mathcal{F}(\alpha)} \mathbb{E}^\tau (\tau - \theta)^+$$

We claim that the solution $\tau^*_\alpha$ of (51) has the same form as $\tau^*$, the solution of (50) in $\mathcal{T}_2$ with $\gamma$ determined so as to satisfy

$$\alpha(\tau^*) = \alpha.$$

This assertion does not need a separate proof since the corresponding claim in the 'independent case' does not rely on the independence as such (cf. Shiryaev, 1978). We may now state

Theorem 5.4. Suppose that a shift occurs in the mean of an autoregressive process as described in (40). Then the sequential detection rule

$$\tau^*_\alpha = \inf \{ n \geq 2 : \tau_n \geq \gamma(\alpha) \}$$

where $(\alpha)$ satisfies $\alpha(\tau^*) = \alpha$, is such that

$$\mathbb{E}^\tau (\tau^*_\alpha - \theta)^+ = \inf_{\mathcal{F}_2(\alpha)} \mathbb{E}^\tau (\tau - \theta)^+$$

with $\mathcal{F}_2(\alpha)$ defined as $\{ \tau \in \mathcal{T} : t \geq 2, \alpha(\tau) \leq \alpha \}$.  

Again, there is no simple way of determining $\gamma$ exactly. It has the following upper bound which can serve as an approximate value. For every $\tau \in \mathcal{J}_2(\alpha)$,

$$\alpha(\tau) = \mathbb{E}^{\tau} (1 - \Pi_{\tau}) \leq 1 - \gamma(\alpha) \Rightarrow \gamma(\alpha) \leq 1 - \alpha.$$ 

5.5 Shift in the Parameter of a Regression Model with Serially Correlated Errors.

Suppose that we are observing a bivariate sequence $(X_n, Y_n)$ and the assumed regression model gets disrupted at a random instant $\theta(\in J)$. Let (52) below describe the model when $\theta = k$.

$$Y_n = \beta X_n + \varepsilon_n \quad (n < k)$$

$$= \beta X_n + \varepsilon_n \quad (n \geq k),$$

where $(\varepsilon_n)$ is a first-order linear autoregressive process with parameter $\rho$, $|\rho| < 1$ and $X_n$'s are iid $\mathcal{N}(\mu, 1)$ random variables. Our problem is to detect the shift epoch so as to minimise $\mathbb{E}^{\tau}(\tau - \theta)^+$. The approach is identical to the one followed in Section 5.4 (most of the present notation is drawn from there).

Box and Tiao (1965) consider a similar problem but their purpose is to test for a change at a specified point in a non-stationary time series when a fixed-size sample up to the hypothesised instant of change and another
fixed-size sample from the change-point onwards are available. That is, they are testing for change at a specified point.

The transformation

\[ Z_1 = Y_1, \quad Z_n = Y_n - \rho Y_{n-1} \quad (n=2,3,\ldots) \]
gives, as shown by Box and Tiao, an independent sequence \((Z_n)\). The conditional distribution of \(Z(n I k)\) given \(X(n I k) = x(n I k)\) is \(\mathcal{N}(\mu(x(n I k)), I(n))\), where \(I(n)\) is the \(n \times n\) identity matrix, and for \(k \geq 1\)

\[
\mu(x(n I k)) = \begin{pmatrix}
\beta x_1 \\
\vdots \\
\beta(x_{k-1} - \rho x_{k-2}) \\
\beta x_k - \beta \rho x_{k-1} \\
\beta(x_{k+1} - \rho x_k) \\
\vdots \\
\beta(x_n - \rho x_{n-1})
\end{pmatrix} \quad (x_0 = 0).
\]

When \(k=0\), \(\mu(x(n I k)) = (\beta_1 x_1, \beta_1 (x_2 - \rho x_1), \ldots, \beta_1 (x_n - \rho x_{n-1}))'\)

Unconditionally, \(Z(n I k)\) has the \(\mathcal{N}(\mu(n I k), I(n))\) distribution with

\[
\mu(n I k) = (\beta \mu \bar{\rho}, \ldots, \beta \mu \bar{\rho}, \mu(\alpha - \beta \rho), \ldots, \beta \mu \bar{\rho})' \quad (k \geq 1)
\]

\[
= (\beta \mu \bar{\rho}, \ldots, \beta \mu \bar{\rho})' \quad (k = 0,1).
\]

Expression for \(\pi_n\) (see (44)) remains the same but
\[ y_1^1 = \frac{1}{q} \frac{\phi(z_1 \mid \beta_1 \mu \bar{\rho}, 1)}{\phi(z_1 \mid \hat{\nu} \mu \bar{\rho}, 1)}, \]

\[ y_{nk}^{n+1} = \frac{1}{q} \frac{\phi(z_{nk+1} \mid \beta_1 \mu \bar{\rho}, 1)}{\phi(z_1 \mid \hat{\nu} \mu \bar{\rho}, 1)} \quad (1 \leq k \leq n) \]

\[ = \frac{1}{q} \frac{\phi(z_{nk+1} \mid \mu(\beta_1 - \hat{\nu} \rho), 1)}{\phi(z_1 \mid \hat{\nu} \mu \bar{\rho}, 1)} \quad (k = n+1) \]

\[ = q^{k-n-2} \quad (k > n+1). \]

Thus,

\[ (1 - \pi_1)^{-1} = 1 + \left( \frac{\pi}{1-\pi} + p \right) \frac{\phi(z_1 \mid \beta_1 \mu \bar{\rho}, 1)}{q \phi(z_1 \mid \hat{\nu} \mu \bar{\rho}, 1)}, \]

and for \( n \geq 1, \)

\[ (1 - \pi_{n+1})^{-1} = 1 + \frac{\pi_n}{1-\pi_n} \frac{\phi(z_{n+1} \mid \beta_1 \mu \bar{\rho}, 1)}{q \phi(z_{n+1} \mid \mu \bar{\rho}, 1)} \]

\[ + \frac{p \phi(z_{n+1} \mid (\beta_1 - \hat{\nu} \rho) \mu, 1)}{\phi(z_1 \mid \hat{\nu} \mu \bar{\rho}, 1)} \]

It is clear that again \((\prod_n)_{\omega}\) is Markov relative to \((\mathcal{F}_n)_{\omega}\), the histories of the sequence \((y_n)\). The remainder of the discussion in Section 5.4, in particular, Theorems 5.3 and 5.4 may be carried over to the present situation.