STATISTIQUE ET ANALYSE DES DONNÉES

S. ZACKS

Classical and bayesian approaches to the change-point problem : fixed sample and sequential procedures

Statistique et analyse des données, tome 7, nº 1 (1982), p. 48-81

<http://www.numdam.org/item?id=SAD_1982__7_1_48_0>

Statistique et Analyse des Données
1982 - 1 - 48-81

CLASSICAL AND BAYESIAN APPROACHES TO THE CHANGE-POINT PROBLEM: FIXED SAMPLE AND SEQUENTIAL PROCEDURES*

S. ZACKS
Department of Mathematical Sciences
SUNY, Binghamton, NY

RESUME

Le problème du point de changement peut être décrit en ces termes. Considérons une suite de variables aléatoires indépendantes $X_1, X_2, \ldots$, et une suite de paramètres à valeurs entières positives $2 \leq \tau_1 < \tau_2 < \tau_3 \ldots$ Les points τ_j ($j = 1,2,\ldots$) sont les instants de changements dans les lois de probabilité des variables aléatoires, c'est-à-dire $X_1,\ldots,X_{\tau_1 - 1}$ ont une distribution identique F_1 ; $X_{\tau_1},\ldots,X_{\tau_2-1}$ ont une distribution F_2, etc. Les distributions $F_1, F_2, \ldots$ peuvent être connues ou partiellement connues, mais les points de changement τ_j sont inconnus. Le problème est d'estimer les paramètres inconnus τ_j ou de tester des hypothèses les concernant. Cette classe de problèmes est très vaste. Elle contient essentiellement tous les problèmes de tests de la stationnarité d'une suite de variables aléatoires contre la possiblité de changements brusques en localisation, échelle ou forme de la loi de probabilité. Ainsi tous les problèmes de contrôle statistique sont dans ce domaine. Il y a dans la littérature, diverses formulations du problème et différentes approches. Il existe des formulations statiques ou dynamiques du problème, avec la possibilité de un ou plusieurs points de changement. Les procédures d'échantillonnage sont fixes ou séquentielles. Les structures inférentielles sont soit classiques soit bayesiennes. Ce texte passe en revue les différentes formulations et approches et produit une bibliographie importante.

Partially supported by ONR Contract N00014-81-K-0407 at SUNY-Binghamton, and by ONR Contract N00014-75-C-0729 at the Program in Logistics, The George Washington Univeristy.

* Texte de la conférence invitée faite par le professeur S. ZACKS aux Journées de Statistique de l'Association des Statisticiens Universitaires tenues à Bruxelles, du 24 au 27 mai 1982.

ABSTRACT

The change-point problem can be described in the following terms. Consider a sequence of independent random variables $X_1, X_2, \ldots$ and a sequence of positive-integer valued parameters $2 \le \tau_1 < \tau_2 < \tau_3 < \ldots$ The points τ_j $(j = 1,2,\ldots)$ are epochs of change in the distribution laws of the random variables ; i.e., $X_1, \ldots, X_{\tau_1 - 1}$ have an identical distribution F_1 ; $X_{\tau_1}, \ldots, X_{\tau_2 - 1}$ have an identical distribution F_2 etc. The distributions $F_1, F_2, \ldots$ may be known or partially known, but the points of change, τ_j, are unknown. The problem is to estimate the unknown parameters τ_j or to test hypotheses concerning these points of change. This class of problems is a very broad one. It embraces essentially all problems which test the stationarity of a sequence of random variables versus the possibility of abrupt changes in the location, scale or shape of the distributions. Thus, all problems of statistical control fall in this domain. In the literature there are various formulations of the problem and different approaches. There are static or dynamic formulations of the problem ; with a possibility of only one point-of-change or many points of change. The sampling procedures are either fixed sample or sequential sampling. The inference framework is either classical or Bayesian. The present paper reviews the various formulations and approaches and provides an extensive bibliography.

I - INTRODUCTION

The change-point problem can be considered one of the central problems of statistical inference, linking together statistical control theory, theory of estimation and testing hypotheses, classical and Bayesian approaches, fixed sample and sequential procedures. It is very often the case that observations are taken sequentially over time, or can be intrinsically ordered in some other fashion. The basic question is therefore, whether the observations represent independent and identically disbributed random variables, or whether at least one change in the distribution law has taken place.
This is the fundamental problem of statistical control theory, testing the stationarity of stochastic processes, estimation of the current position of a time-series, etc... Accordingly, a survey of all the major developments in statistical theory and methodology connected with the very general outlook of the change-point problem, would require review of the field of statistical quality control, the switching regression problems, inventory and queueing contrôl, etc. This is, however, too broad to cover in a single revue paper.

The present review paper is therefore focused on methods developed during the last two decades for the estimation of the current position of the mean function of a sequence of random variables (or of a stochastic process) ; testing the null hypothesis of no change among given n observations, against the alternative of at most one change ; the estimation of the location of the change-point(s) and some sequential detection procedures. The present paper is composed accordingly of five major sections. Section 2 is devoted to the problem of estimating the current position of a sequence of random variables, specifically discussing the problem with respect to possible changes of the means of independent normally distributed random variables. We review the studies on this problem of Barnard [6], Chernoff and Zacks [14], Mustafi [45] and others. Section 3 is devoted to the testing problem in a fixed sample. More specifically, we consider a sample of n independent random variables. The null hypothesis is $H_o : F_1(x) = \ldots = F_n(x)$, against the alternative, $H_1 : F_1(x) = \ldots = F_\tau(x)$; $F_{\tau+1}(x) = \ldots = F_n(x)$, where $F_\tau \neq F_{\tau+1}$ and $\tau = 1,2,\ldots,n-1$ designates a possible unknown change point. The studies of Chernoff and Zacks [14], Kander and Zacks [36], Gardner [21], Bhattacharya and Johnson [9], Sen and Srivastava [57] and others are discussed. These studies develop test statistics in parametric and non-parametric, classical and Bayesian frameworks. Section 4 presents Bayesian and maximum likelihood estimation of the location of the shift points. The Bayesian approach is based on modeling the prior distribution of the unknown parameters, adopting a loss function and deriving the estimator which minimizes the posterior risk. This approach is demonstrated with an example of a shift in the mean of a normal sequence. The estimators obtained are generally non-linear complicated functions of the random variables. From the Bayesian point of view these estimators are optimal. If we ask, however, classical questions concerning the asymptotic behavior of such estimators, or their sampling distributions under repetitive sampling, the analytical problems become very difficult and untractable. The classical efficiency of such estimators is often estimated in some special cases by extensive simulations. The maximum likelihood estimation of the location parameter of the change point is an attractive alternative to the Bayes estimators. Hinkley [26-30] investigated the asymptotic behavior of these estimators. The derivation of the asymptotic distributions of these estimators is very complicated. We present in Section 4 Hinkley's approach for the determination of the sampling distributions of the maximum likelihood estimators. Section 5 is devoted to sequential detection procedures. We present the basic Bayesian and classical results in this area. The studies of Shiryaev [60,61] , Bather [7,8], Lorden [43] and Zacks and Barzily [69] are discussed with some details. The study of Lorden [43] is especially significant in proving that Page's CUSUM procedures [47-49] are asymptotically minimax.

The important area of switching regressions have not been reviewed here in any details. The relevance of the switching regression studies to the change-point problem is obvious. Regression relationship may change at unknown epochs (change points), resulting in different regression regimes that should be detected and identified. The reader is referred to the important studies of Quandt [51,52], Inselman and Arsenal [35], Ferreira [19], Maronna and Yohai [44] and others.

An annotated bibliography on the change-point problem was published recently by Shaban [59]. The reader can find there additional references to the seventy-one references given in the last section of the present paper.

2 - ESTIMATING THE CURRENT POSITION OF A PROCESS

G. Barnard, in his celebrated 1959 paper [6] on control charts and stochastic processes, suggested to consider the problem of estimating the current position of a process as a tool of statistical control. The problem of estimating the current mean of a process requires modeling of the possible change mechanism of the mean function. In the context of statistical control problems the mean, as function of time, is generally assumed to commence at an initial point, μ_o, known or unknown, and then change abruptly at unknown epochs, $\tau_1, \tau_2, \ldots$

Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables. We denote by μ_i $(i = 1, \ldots, n)$ a location parameter of the distribution of X_i. If the random variables are normally distributed then μ_i is the expected value (mean) of X_i.

Generally, neither the change points τ_1, $\tau_2, \ldots$ nor the size of changes are known, and the problem of estimating μ_n, after observing $X_1, X_2, \ldots, X_n$, might have no better solution than the trivial estimator $\hat{\mu}_n = X_n$, unless the phenomenon studied allows proper modeling. In the present paper we discuss the models adopted by Barnard [6] and by Chernoff and Zacks [14], and the estimators of the current position which they derived from these models. The related study of Mustafi [45] is also presented. As will be shown, time-series procedures of exponential smoothing are strongly related to linear unbiased estimator studied in [6] and [14].

2.1. - Barnard's Estimator of μ_n

Consider the given sequence of observations in a reversed time manner, i.e., $X_n, X_{n-1}, X_{n-2}, \ldots$ Barnard adopted the basic assumption that the corresponding random variables are independent and normally distributed, with the same known variance ($\sigma_x^2 = 1$). Suppose that the observations are taken at regular time intervals of 1 unit. Barnard's model assumes that the epochs of change $\tau_1, \tau_2, \ldots$ follow a Poisson process with intensity λ (per time unit). At each of the random change epochs $\tau_1, \tau_2, \ldots$ the size of the shift in the mean is a random variable, δ, following a normal distribution, $N(0,\sigma^2)$. Moreover, $\delta_1, \delta_2, \ldots$ are mutually independent, and the sequence $\{\delta\}$ is independent of $\{\tau\}$. Thus, if $J_1, J_2, \ldots$ designate the number of change epochs between, X_n and X_{n-1}, X_{n-1} and X_{n-2}, then $J_1, J_2, \ldots$ is a sequence of i.i.d. (independent and identically distributed) random variables having a Poisson distribution, $P(\lambda)$. The models is $X_n = \mu_n + E_n$

$$X_{n-i} = \mu_n + \sum_{k=1}^{i} S_k + E_{n-i} \qquad i=1,\ldots,n-1 \tag{2.1}$$

where $S_i = \sum_{j=1}^{J_i} \delta_j$ and $E_1,\ldots,E_n$ are i.i.d. $N(0,1)$. Assuming that λ and σ^2 are known, Barnard provided the general form of the minimum mean square error (MSE) linear estimator of μ_n, and that of its (formal) Bayes estimator (which is actually called by Barnard "the mean-likelihood estimator"). It is shown that the minimum MSE linear estimator, is the exponential smooting estimator

$$\hat{\mu}_n = B\, X_n + A\, \hat{\mu}_{n-1}. \tag{2.2}$$

The (formal) Bayes estimator of μ_n is of the form

$$\hat{\mu}_n = \sum_{j_n} \pi(\underset{\sim}{j}_n | \underset{\sim}{X}_n) \frac{\underset{\sim}{1}_n' \underset{\sim}{V}_n^{-1} \underset{\sim}{X}_n}{\underset{\sim}{1}_n' \underset{\sim}{V}_n^{-1} \underset{\sim}{1}_n} \tag{2.3}$$

where $\underset{\sim}{1}_n$ is an n-dimensional vector of 1's ; $\underset{\sim}{j}_n = (j_1,\ldots,j_{n-1})'$ is a particular realization of $J_1,\ldots,J_{n-1}$; $\underset{\sim}{X}_n = (X_n, X_{n-1},\ldots,X_1)'$; $\pi(\underset{\sim}{j}_n\ \underset{\sim}{X}_n)$ its posterior probability, and $\underset{\sim}{V}_n$ the covariance matrix of $\underset{\sim}{X}_n$ corresponding to a given realization $\underset{\sim}{j}_n$.

2.2. - Chernoff and Zacks' Model and BLUE of μ_n

Chernoff and Zacks assumed a model different from that of Barnard, although there are general similarities. According to their model, if $\mu_i = E\{X_i\}$ then

$$\mu_i = \mu_{i+1} + J_i \, \delta_i \, , \qquad i = 1,\ldots,n-1 \tag{2.4}$$

where J_i is a random variable assuming the value 1 if there is a shift in the mean between the ith and (i+1)st observations, and the value 0 otherwise. Furthermore, $\delta_1,\ldots,\delta_{n-1}$ are i.i.d. $N(0,\sigma^2)$, $J_1,\ldots,J_{n-1}$ are i.i.d., $P[J_i=1] = p$ $(i=1,\ldots,n)$. Let $\underset{\sim}{J} = (J_1,\ldots,J_{n-1})$ and $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_{n-1})$, $\underset{\sim}{J} \perp\!\!\!\perp \underset{\sim}{\delta}$. Chernoff and Zacks showed that the minimum variance linear unbiased estimator (BLUE) of μ_n is

$$\hat{\mu}_n = \frac{X_n + \sum_{i=1}^{n-1} \xi_i X_i}{1 + \sum_{i=1}^{n-1} \xi_i} \tag{2.5}$$

where

$$\xi_i = \begin{cases} (\nu_{i-1}-1)/\nu_{i-1}\cdots\nu_{n-2}(\nu_{n-1}-1) \, , & i = 2,\ldots,n-1 \\ 1/\nu_1\nu_2\cdots\nu_{n-2}(\nu_{n-1}-1), & i = 1 \end{cases} \tag{2.6}$$

and

$$\nu_k = \begin{cases} 2+\sigma^2 p \, , & \text{if } k = 1 \\ 2+\sigma^2_p-\nu^{-1}_{k-1} & \text{if } k = 2,\ldots,n-1 \end{cases} \tag{2.7}$$

In the following table we illustrate some of these weights :

Table 2.1. Weights for the BLUE
$\sigma^2 = 1$, $p = .1$

n \ i	1	2	3	4	5
2	.909	1.000			
3	.763	.840	1.000		
4	.606	.666	.793	1.000	
5	.464	.510	.735	.745	1.000

Notice that when $p\sigma^2 = 0$ then $\xi_i = 1$ for all $i=1,\dots,n-1$. In this case $\hat{\mu}_n = \bar{x}_n$ is the common sample mean. On the other hand, when $p\sigma^2 \to \infty$ then the weights ξ_i diminish to zero in a geometric rate, i.e. $\xi_1 = O((p\sigma^2)^{-(n-i)})$. Accordingly, as n increases, the weight given to observations at the beginning of the sequence is close to zero. In particular, if $p\sigma^2$ is large, it is sufficient to base the estimator only on the last m observations. Mustafi [45] investigated the characteristics of such estimators based on the last block of m observations.

Moreover, Mustafi showedthat, if the value of $c = p\sigma^2$ is unknown, it can be estimated consistently by

$$\hat{c} = \frac{6S_1^2 - 2S_2^2}{S_2^2 - 2S_1^2}, \tag{2.8}$$

where

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - X_{i+1})^2,$$
$$S_2^2 = \frac{1}{n-2}\sum_{i=1}^{n-2}(X_i - 2X_{i+1} + X_{i+2})^2. \tag{2.9}$$

Let $\hat{\mu}_{n,m}$ denote the UMVU estimator of μ_n, based on the last m observations in which c is replaced by its estimate $\hat{c}$. According to Mustafi's procedure, the first n-m observations are used to estimate c by (2.9), and the estimator $\hat{c}$ of c is substituted in (2.6)-(2.7) to obtain the corresponding weights $\hat{\xi}_{i,m}$. Notice that the estimator obtained in this manner is not BLUE anymore. Furthermore, $\hat{c}$ might be negative (with positive probability). In such a case, $\hat{\xi}_{i,m}$ is replaced by its positive part $\xi^+_{i,m} = \max(0, \hat{\xi}_{i,m})$. Mustafi established that

(i) $\qquad E\{\hat{\mu}_{n,m}\} = \mu_n \qquad$, for each n,m

(ii) $\qquad V\{\hat{\mu}_{n,m}\} \le 1 + \sigma^2 p(m-1)$,

and

(iii) $\qquad \lim_{n\to\infty} V\{\hat{\mu}_{n,m}\} = V\{\hat{\mu}_m\}$,

where $\hat{\mu}_m$ is the BLUE estimator based on the last m observations, with known c.

2.3. - Chernoff-Zacks Bayes Estimators of μ_n

Assuming that μ_n has a prior normal distribution $N(0,\tau^2)$, we obtain that the posterior distribution of μ_n, given $\underset{\sim}{X}_n$ and $\underset{\sim}{J}_n = (J_1,\dots,J_{n-1})$ is normal, with mean

$$\hat{\mu}_n(\underset{\sim}{J}_n) = \frac{\underset{\sim}{1}_n' \; \Sigma^{-1} \; (\underset{\sim}{J}_n) \; \underset{\sim}{X}_n}{\underset{\sim}{1}_n' \; \Sigma^{-1} \; (\underset{\sim}{J}_n) \; \underset{\sim}{1}_n} \tag{2.10}$$

and variance

$$\nu\{\underset{\sim}{J}_n\} = \frac{1}{\tau^{-2} + \underset{\sim}{1}_n' \; \Sigma^{-1} \; (\underset{\sim}{J}_n) \; \underset{\sim}{1}_n}, \tag{2.11}$$

where

$$\Sigma\,(\underset{\sim}{J}_n) = I + \sigma^2 \; J \; J' \qquad \text{and}$$

$$J = \left[\begin{array}{cccc|c} J_1 & J_2 & \cdots & J_{n-1} & \\ & J_2 & \cdots & J_{n-1} & \\ 0 & & & & \underset{\sim}{0} \\ & & & J_{n-1} & \\ \hline & & \underset{\sim}{0}' & & 0 \end{array}\right] \tag{2.12}$$

Let $p_n(j)$ be a prior probability function of $\underset{\sim}{J}_n$. The posterior probability function of $\underset{\sim}{J}_n$, given $\underset{\sim}{X}_n$, is then

$$p_n(\underset{\sim}{j}\,|\,\underset{\sim}{X}_n) = \frac{p_n(\underset{\sim}{j})\,n(\underset{\sim}{X}_n\,|\,\underset{\sim}{0}\;\Sigma^*(\underset{\sim}{j}))}{\sum_{\{j\}} p_n(\underset{\sim}{j})\,n(\underset{\sim}{X}_n\,|\,\underset{\sim}{0},\Sigma^*(\underset{\sim}{j}))}, \tag{2.13}$$

where $\Sigma^*(\underset{\sim}{j}) = \Sigma(\underset{\sim}{j}) + \tau^2 \underset{\sim}{1}_n \underset{\sim}{1}_n'$, and $n(\underset{\sim}{x}_n|\underset{\sim}{0},\Sigma)$ is the multivariate normal p.d.f. at $\underset{\sim}{X}_n$, with mean vector $\underset{\sim}{0}$ and covariance matrix Σ. Finally, the Bayes estimator of μ_n is

$$\mu_n^{(B)} = \sum_{\{j\}} p_n(\underset{\sim}{j}\,|\,\underset{\sim}{X}_n)\; \hat{\mu}_n(\underset{\sim}{j}). \tag{2.14}$$

This estimator is obviously non-linear, due to the non-linear structure of the posterior probabilites. The structure of the Bayes estimator (2.14) is the same as that of Barnard's mean-likelihood estimator (2.3). The problem with these estimators is in their degree of complexitv. The sample space of $\underset{\sim}{J}_n$

consists of 2^{n-1} different points and it is a very difficult matter to choose a proper prior distribution. Even if we ascribe, a priori, each of these 2^{n-1} points equal probabilities, we have to make a significantly large number of calculations to determine $\mu_n^{(B)}$. In many problems of interest it is unreasonable to assume that the mean is likely to shift between any two observations. If it is reasonable to assume that the number of possible shifts among a relatively small number of observations is at most one, the computations will be significantly simplified. The Bayes estimator based on the assumption of at most one change (AMOC) is presented in the next section.

2.4. - The AMOC-Bayes Estimator of μ_n

According to the AMOC model we assume that among the given n observations there is at most one change. Let τ be an integer valued parameter assuming the values $0,1,\ldots,n-1$. If $\tau = 1$, the first t random variables have the same mean $\mu_n+\delta$ and the last $n-t$ random variables have the mean μ_n. If $\tau = 0$ there was no shift in the mean among the n observations. Let $\pi(t)$ be the prior probability of $\{\tau = t\}$. The conditional Bayes estimator, for a given value of t, is

$$\hat{\mu}_n(t) = \frac{n\bar{X}_n + \sigma^2 t(n-t)\bar{X}^*_{n-t}}{n + \sigma^2\, t(n-t)}, \qquad t = 0,\ldots,n-1, \tag{2.15}$$

where

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{X}^*_{n-t} = \frac{1}{n-t}\sum_{j=t+1}^{n} X_j.$$

Furthermore, the posterior probability of $\{\tau=t\}$, given $\underset{\sim}{X}_n$, is

$$\pi(t\,|\,\underset{\sim}{X}_n) = \frac{\mu(t)\,\exp\{\frac{1}{2}\cdot\frac{\sigma^2 t^2(n-t)^2(\bar{X}_t-\bar{X}^*_{n-t})^2}{n^2+\sigma^2\, t(n-t)n}\}}{(n+\sigma^2 t(n-t))^{1/2}\; D_n}, \tag{2.16}$$

where

$$D_n = \sum_{j=0}^{n-1} \frac{\pi(j)}{(n+\sigma^2 j(n-j))^{1/2}}\, \exp\{\frac{1}{2}\cdot\frac{\sigma^2 j^2(n-j)^2\;(\bar{X}_j-\bar{X}^*_{n-j})^2}{n^2+\sigma^2 j(n-j)n}\} \tag{2.17}$$

The Bayes estimator of μ_n in the AMOC model is accordingly

$$\hat{\mu}_n = \sum_{j=0}^{n-1} \pi(j\,|\,\underset{\sim}{X}_n)\,\hat{\mu}_n(j). \tag{2.18}$$

2.4.1. - Adaptative AMOC-Bayes Estimation

The AMOC procedure can be applied on the last m observations sequentially, starting with m=2 and increasing it until a strong indication emerges that a shift has taken place. The procedure is then stopped and μ_n is estimated according to (2.18) on the basis of the last m observations. This process is summarized in algorithm :

Step 0. Set $m = 2$.

Step 1. Set $Y_1 = X_{n-m+1}, \ldots, Y_m = X_n$.

Step 2. Compute $\pi(t|\underset{\sim}{Y}_m)$, $t = 0,\ldots,m-1$.
If $\pi(0|\underset{\sim}{Y}_n) = \max\limits_{0\le j\le m-1} \pi(j|\underset{\sim}{Y}_m)$

go to Step 3 ; else go to Step 4.

Step 3. Set $m \leftarrow m+1$ and go to Step 1.

Step 4. Let $k^* =$ least $j = 1, m-1$ such that
$\pi(j|\underset{\sim}{Y}_m) = \max\limits_{0\le t\le m-1} \pi(t|\underset{\sim}{Y}_m)$.

Step 5. Apply estimator (2.18) on the last $m-k^*$ observations.

The following numerical example illustrates the adaptative estimation process according to the above algorithm. Consider the following n=9 observations on independent, normally distributed r.v.'s : $X_1 = 2.613$, $X_2 = 1.661$, $X_3 = 1.814$, $X_4 = 1.274$, $X_5 = 2.616$, $X_6 = -.326$, $X_7 = -2.422$, $X_8 = -.119$, $X_9 = -.034$. Assume that $\sigma^2 = 3$ and the prior distribution is

$$\pi_m(0) = (1-p)^{m-1}$$
$$\pi_m(t) = p(1-p)^{m-t-1}, \qquad t = 1,\ldots,m-1 ,$$

with $p = .2$. The posterior probabilities of the change points, given the last m observations, are given in the following table.

Table 2.2. Posterior Probabilities of the Shift Locations

m \ t	0	1	2	3	4
2	.9298	.0702			
3	.6804	.0722	.2474		
4	.7844	.0660	.0954	.0542	
5	.1765	.0107	.0088	.0890	.7149

According to these posterior probabilites there is a strong indication that a shift took place between the fifth and the sixth observation. The AMOC Bayes estimator based on the last four observations is $\hat{\mu}_{9,4} = -.6301$.
Experience with the application of this method on various data sets shows that it could be too sensitive as an estimator of the location of the shift points. Farley and Hinich [18] showed in a series of simulations that the above procedure leads to a high proportion of indication of change when there are none (false alarms). This problem can, however, be overcome by proper choice of the parameters p and σ^2.
σ^2 should be at least 3 or 4 times the variance of the random variables $E_1, \ldots, E_n$. As an estimator of the current position the above procedure performs very well. This was also reported by Farley and Hinich in [18]. We provide here some numerical comparisons of the characteristics of the UMVU, AMOC-Bayes and the adaptive AMOC-Bayes estimators of μ_n, based on some simulation experiments. These results are taken from Chernoff and Zacks [14]. In these experiments 100 replicas of samples of size n=9 were simulated from normal distributions, with means μ_i and variance 1. In all cases $\mu_9=0$. We compare the means and MSE, over the 100 replicas, of the following estimators :

$\hat{\mu}_1$: UMVU with $\sigma^2 = 3$, $p = .2$

$\hat{\mu}_2$: AMOC-Bayes, $\sigma^2 = 3$, $p = .2$

$\hat{\mu}_3$: Adaptive AMOC-Bayes, $\sigma^2 = 3$, $p = .2$.

The models of shifts in the means are :

Model I : A random change between every two observations, i.e. $\mu_i \sim N(0,2)$ $(i = 1,\ldots,8)$.

Model II : $\mu_i = \sigma^2 \sum_{k=i}^{8} J_k \eta_k$,

$J_1, J_2, \ldots$ are i.i.d. Bernouilli, with $p = .1$, $\sigma = 2$.
$\eta_1, \eta_2, \ldots$ are i.i.d. $N(0,1)$.

Model III : No change.

The simulation estimates are :

Table 2.3 - Simulation Characteristics of Three Estimators

Model	Estimates $\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	
I	-.2718	-.1866	-.0827	Mean
	2.1406	3.3140	1.0235	MSE
II	.0847	.0539	.0525	Mean
	.4460	.4337	.4135	MSE
III	.0255	.0027	-.0122	Mean
	.3078	.6112	.2679	MSE

The above results indicate that the adaptive AMOC-Bayes estimator is performing as well or better than the UMVU or the AMOC-Bayes, especially when the actual process of shifts in the means is different from the one assumed in the model.

3 - TESTING HYPOTHESES CONCERNING CHANGE POINTS

The problem of testing hypotheses concerning the existence of shift points was posed by Chernoff and Zacks [14] in the following form.
Let $X_1, \ldots, X_n$ be a sequence of independent random variables having normal distributions $N(\theta_i, 1)$, $i = 1, \ldots, n$. The hypothesis of no shift in the means, versus the alternative of one shift in a positive direction is

$$H_o : \theta_1 = \ldots = \theta_n = \theta_o$$

vs

$$H_1 : \theta_1 = \ldots = \theta_\tau = \theta_o \; ; \; \theta_{\tau+1} = \ldots = \theta_n = \theta_o + \delta \, ,$$

where $\tau = 1, \ldots, n-1$ is an unknown index of the shift point, $\delta > 0$ is unknown and the initial mean θ_o may or may not be known.

Chernoff and Zacks showed in [14] that a Bayes test of H_o versus H_1, for δ values close to zero, is given by the test statistic

$$(3.1) \qquad T_n = \begin{cases} \sum_{i=1}^{n-1} (i+1)X_i \quad, & \text{if } \theta_o \text{ is known} \\ \sum_{i=1}^{n-1} (i+1)(X_i-\bar{X}_n), & \text{if } \theta_o \text{ is unknown,} \end{cases}$$

where $\bar{X}_n$ is the overage of all the n observations. It is interesting to see that this test statistic weighs the current observations (those with index close to n) more than the initial ones. However, the weight is linear rather than geometric (as in the estimation of the current position). Since the above test statistic is a linear function of normal random variables T_n is normally distributed and it is easy to obtain the critical value for a size α test and the power function. These functions are given in the paper of Chernoff and Zacks [14] with some numerical illustrations.
The above results of Chernoff and Zacks were later generalized by Kander and Zacks [36] to the case of the one-parameter exponential family, in which the density functions are expressed, in the natural parameter form as $f(x;\theta) = h(x) \exp\{\theta U(x)+\psi(\theta)\}$ (see Zacks [70, pp. 95]). Again, Kander and Zacks established that the Bayes test of H_o, for small values of δ when θ_o is known, is of the form (3.1), where X_i are replaced by $U(X_i)$ $(i = 1,\dots,n)$. The exact determination of the critical levels might require a numerical approach, since the exact distribution of T_n is not normal, if $U(X_i)$ are not normal. Kander and Zacks showed how the critical levels and the power functions can be determined exactly, in the binomial and the negative-exponential cases. If the samples are large, the null distribution of T_n converges to a normal one, according to the Lapunov version of the Central Limit Theorem (see Fisz [20, pp. 202]). Kander and Zacks [36] provided numerical comparisons of the exact asymptotic power functions of T_n, in the binomial and the negative-exponential cases.
It is often the case that the sample size is not sufficiently large for the normal approximation to yield results close to the true ones. For this reason, Kander and Zacks tried to approximate the exact distribution of T_n by the Edgeworth expansion

$$(3.2) \qquad F_n(Z) \doteq \Phi(Z) - \frac{\gamma_{1,n}}{3!}\Phi^{(3)}(Z) + \frac{\gamma_{2,n}}{4!}\Phi^{(4)}(Z) + (10\gamma_{1,n}^2/6!)\Phi^{(6)}(Z),$$

where $F_n(Z)$ is the exact distribution of the standardized test statistic $Z_n = (T_n - E\{T_n\})/(\mathrm{Var}\ \{T_n\})^{1/2}$; $\Phi(Z)$ is the standard normal c.d.f. ; $\Phi^{(\nu)}(Z)$ is the ν-th derivative of $\Phi(Z)$ and $\gamma_{1,n} = \mu_{3,n}/(\mu_{2,n})^{3/2}$, $\gamma_{2,n} = \mu_{4,n}/\mu_{2,n}^2 - 3$ where $\mu_{j,n}$ is the j-th central moment of T_n.

It was shown that when the samples are not large (n=10) the Edgeworth expansion of the c.d.f. of Z_n, under the alternative hypothesis H_1, provides power function approximation better than those of the normal approximation. Hsu [34] utilized the above test for testing whether a shift occurred in the variance of a normal distribution.

Gardner [21] considered the testing problem of H_o versus H_1 for the normally distributed random variables, but with $\delta \neq 0$ unknown. He showed that the Bayes test statistics, with prior probabilities Π_t, $t = 1,2,\ldots,n-1$, is

$$
\begin{aligned}
Q_n &= \sum_{t=1}^{n-1} \Pi_t \left[\sum_{j=t}^{n-1} (X_{j+1} - \bar{X}_n)\right]^2 \\
&= \sum_{t=1}^{n-1} \Pi_t (n-t)^2 (\bar{X}^*_{n-t} - \bar{X}_n)^2,
\end{aligned} \tag{3.3}
$$

where $\bar{X}^*_{n-t}$ is the mean of the last n-t observations and $\bar{X}_n$ is the mean of all n observations. Gardner investigated the exact and the asymptotic distributions of Q_n, under the null hypothesis H_o and under the alternative H_1, for the case of equal prior probabilities. Scaling Q_n, so that its expected value will be 1 for each n, by the transformation $Y_n = (6n/(n^2-1))Q_n$, $n = 2,3,\ldots$, we obtain that, under H_o, Y_n is distributed like $\sum_{k=1}^{n-1} \lambda_k U_k^2$, where $U_1,\ldots,U_{n-1}$ are i.i.d. standard normal r.v.'s and

$$
\lambda_k = \frac{6n^2}{\pi^2(n^2-1)k^2}\left[\frac{2n}{k\pi}\cos(k\pi/2n)\right]^{-2}, \quad k = 1,\ldots,n-1. \tag{3.4}
$$

Thus, as $n \to \infty$, the asymptotic distribution of Y_n, under H_o, is like that of

$$
Y = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} U_k^2 . \tag{3.5}
$$

The distribution of Y is that of the asymptotic distribution of Smirnov's statistic ω_n^2, normalized to have mean 1. Smirnov's statistic compares the empirical c.d.f. of a sample of continuous random variables to a particular distribution, $F_o(x)$. More specifically, if $X_{(1)} \leq \ldots \leq X_{(n)}$ is the order statistic, corresponding to n i.i.d. random variables, and if $F_n(x)$ is the corresponding empirical c.d.f., i.e.,

$$F_n(x) = \sum_{j=1}^{n} I\{X_{(j)} \leq x < X_{(j+1)}\}\frac{j}{n},$$

Smirnov's statistic is

$$(3.6) \qquad \omega_n^2 = \frac{1}{2n} + 6[\sum_{j=1}^{n} F_o(X_{(j)}) - F_n(x) + \frac{1}{2n}]^2$$

Gardner refers the reader to Table VIII of VonMises [66] for the critical values of Y_n, for large n. Critical values $c_n(\alpha)$, for $\alpha = .10, .05$ and $.01$ and various values of n, can be obtained from Figure 1 of Gardner's paper.
Gardner showed also that, under H_o, the p.d.f. of Y_n is

$$(3.7) \qquad f_n(y) = \frac{1}{\pi} \int_o^\infty \prod_{k=1}^{n-1} (1+t^2\alpha_k^{-2})^{-1/4} \cos(ty - \frac{1}{2}\sum_{k=1}^{n-1} \tan^{-1} t\alpha_k^{-1})dt,$$

where $\alpha_k = \frac{n-1}{3}\cos^2(k\pi/2n)$, $k = 1,\ldots,n-1$. The integration of $f_n(y)$ for the determination of its $(1-\alpha)$th fractile, $c_n(\alpha)$, requires special numerical techniques. The power function of the test was determined by Gardner in some special cases by simulation.
Sen and Srivastava [56] discussed the statistic

$$(3.8) \qquad U_n = \frac{1}{n^2}\sum_{i=1}^{n-1} (\sum_{j=i}^{n-1} X_{j+1})^2$$
$$= \frac{1}{n^2}\sum_{i=1}^{n-1} (n-i)^2(\bar{X}^*_{n-i})^2,$$

for testing H_o versus H_1 with $\delta \neq 0$, when the initial mean, μ_o, is known.
They showed that the asymptotic distribution of U_n, under H_o, has the c.d.f.

$$(3.9) \qquad F(Z) = \frac{4}{\sqrt{\pi}}\sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(1/2+j)}{j!} (1 - \Phi(\frac{1/2+2j}{\sqrt{2Z}})).$$

In addition, they derived the c.d.f. of U_n for finite values of n, and provided a table in which these distributions are presented for n=10, 20, 50 and ∞ (asymptotic).
In addition, Sen and Srivastava proposed test statistics which are based on the likelihood ratio test. More specifically, for testing H_o versus H_1, with $\delta > 0$, when μ_o is unknown, the likelihood function, when the shift is at a point t, is

$$(3.10) \qquad L_t(\underset{\sim}{X}_n) = \frac{1}{(2\pi)^{n/2}} \exp\{-1/2[\sum_{i=1}^{t} (X_i - \bar{X}_t)^2 + \sum_{i=t+1}^{n} (X_i - \bar{X}^*_{n-t})^2]\}$$

It can be easily shown that the likelihood ratio test statistic is then

$$\text{(3.11)} \qquad \Lambda_n = \sup_{1 \le t \le n-1} (\bar{X}_t - \bar{X}^*_{n-t}) / (\tfrac{1}{t} + \tfrac{1}{n-t})^{1/2}.$$

Power comparisons of the Chernoff and Zacks Bayesian statistic T_n and the likelihood ratio statistic Λ_n are given for some values of n and point of shift τ. These power comparisons are based on simulations, which indicate that the Chernoff-Zacks Bayesian statistic is generally more powerful than the Sen - Srivastava likelihood ratio statistic when $\tau \simeq n/2$. On the other hand, when τ is close to 1 or to n, the likelihood ratio test statistic is more powerful.

Bhattacharyya and Johnson [9] approached the testing problem in a non-parametric fashion. It is assumed that the random variables $X_1, X_2, \ldots, X_n$ are independent and have continuous distributions F_i $(i=1,\ldots,n)$. Two types of problems are discussed. One in which the initial distribution, F_o, is known and is symmetric around the origin. The other one is that in which the initial distribution is unknown and not necessarily symmetric. The hypotheses corresponding to the shift problem when F_o is known is $H_o : F_o = \ldots = F_n$, for some specified F_o in $\mathcal{F}_o = \{F : F \text{ continuous and symmetric about } 0\}$.
versus

$$H_1 = F_o = F_1 = \ldots = F_\tau > F_{\tau+1} = F_n, \quad \text{some } F_o \in \mathcal{F}_o.$$

τ is an unknown shift parameter. $F_\tau > F_{+1}$ indicates that the random variables after the point of shift are stochastically greater than the ones before it. For the case of known initial distribution $F_o(x)$, the test is constructed with respect to a translation alternative of the form $F_{\tau+1}(x) = F_o(x-\Delta)$, where $\Delta > 0$ is an unknown parameter. The problem is invariant with respect to the group of all transformations $x_i' = g(x_i)$, $i = 1,\ldots,n$, where $g(x)$ is continuous, odd and strictly increasing. The maximal invariant statistic is $(R_1,\ldots R_n)$ and $(J_1,\ldots,J_n)$, where R_i = rank of $|X_i|$ $(i = 1,\ldots,n)$, and $J_i = 0$ if $\text{sgn}(X_i) = -1$, $J_i = 1$ if $\text{sgn}(X_i) = 1$.

The average power of a test is thus

$$\bar{\psi}(\Delta) = \sum_{i=1}^{n} q_i \psi(\Delta | i-1),$$

where $\psi(\Delta|t)$ is the power at Δ, when the shift occurs after t observations, $q_1,\ldots,q_n$ are given probability weights ($q_i \geq 0$, $\Sigma q_i = 1$). Bhattacharyya and Johnson proved that, under some general smoothness conditions on the p.d.f. $f_o(x)$, the form of the invariant test statistic, maximizing the derivative of the average power $\bar{\psi}(\Delta)$ at $\Delta = 0$, is

$$(3.12) \qquad T_n = \sum_{i=1}^{n} Q_i \; \mathrm{sgn}(X_i) \; E\{-f_o' \; (V^{(R_i)})/f_o \; (V^{(R_i)})\},$$

where $V^{(1)} \leq \ldots \leq V^{(n)}$ is an ordered statistic of n i.i.d. random variables having a distribution $F_o(x)$, and $Q_i = \sum_{j=1}^{n} q_j$. More specific formulae for the cases of double-exponential, logistics and normal distributions are given. The null hypothesis H_o is rejected for large values of T_n. It is further proven that, any test of the form $T = \sum_{i=1}^{n} Q_i \; \mathrm{sgn}(X_i) \; U(R_i)$, where U is a strictly increasing function, is unbiased. Moreover, if the system of weights $q_{n,i}$; $i=1,\ldots,n$ satisfies the condition.

$$(3.13) \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Q_{n,i}^2 = b^2 \quad , \qquad 0 < b^2 < \infty \; ,$$

then, the distribution of $T_n/(nb^2 (\int_0^1 \psi^2(u)\; dn))^{1/2}$, as $n \to \infty$, converges to the standard distribution, where

$$(3.14) \qquad \psi(u) = -f_o' \; (F_o^{-1} \; (\frac{1}{2} \; (u+1))/f_o \; (F_o^{-1}(1/2(u+1))).$$

Similar analysis is done for the case of unknown initial distribution F_o. In this case the test statistic is a function of the maximal invariant $(S_1,\ldots,S_n)$, which are the ranks of $(X_1,\ldots,X_n)$. The test statistic in this case is of the general form

$$(3.15) \qquad T_n^* = \sum_{i=1}^{n} Q_i \; E\{-f'(V^{(S_i)})/f(V^{(S_i)})\}.$$

In the normal case, for example, with equal weights for $t=2,\ldots,n$ and weight 0 for $t=1$, the test statistic is $T_n^* = \sum_{i=1}^{n} (i-1)S_i$.

Notice the similarity in structure between the statistic T_n^* and that of Chernoff and Zacks, T_n. The difference is that the actual values of X_i are replaced by their ranks, S_i.

Hawkins [23] also considered the normal case, with two sided hypothesis, both θ_o and δ unknown. Like Sen and Srivastava, he considered the test statistic $U_n = \max_{1\le k\le n-1} |T_k|$, where

$$(3.16) \qquad T_k = \sqrt{\frac{n}{k(n-k)}} \sum_{i=k}^{n} (X_i - \bar{X}_n), \quad k = 1,\dots,n-1.$$

The statistics $T_1,\dots,T_{n-1}$ are normally distributed, having a correlation function

$$(3.17) \qquad \ell(T_m, T_k) = \sqrt{\frac{m(n-k)}{k(n-m)}}, \quad m \le k$$

Hawkins provides recursive formulae for the exact determination of the distribution of U. Conservative testing can be made by applying the Banferroni inequality

$$P\{\max_{1\le k\le n-1} |T_k| > c\} \le (n-1)\, P\, |T_1| > c\}$$

$$= 2(n-1)\, \Phi(-c)$$

Hence, a conservative α level test of H_o can be on the critical level $Z_{1-\alpha/(2n-2)}$, where Z_γ is the γ-fractile of the standard normal distribution.
A numerical example is given to compare the exact and the Banferroni approximation to the critical values of the test statistic U_n. In an attempt to understand the asymptotic properties of U_n, Hawkins considered the behavior of the maximum of a Gaussian process having the same covariance structure as that of $T_1, T_2, \dots$ The asymptotic results are still not satisfactory.
Pettitt [50] discussed non-parametric tests different from those of Bhattacharyya and Johnson. He defined for each $t=1,\dots,n$, $U_{t,n} = \sum_{i=1}^{t} \sum_{j=t+1}^{n} \operatorname{sgn}(X_i - X_j)$ and studied the properties of the test statistic

$$(3.18) \qquad K_n = \max_{1\le t\le n} |U_{t,n}| \quad .$$

The distribution of K_n was studied for Bernouilli random variables.

4 - ESTIMATING THE LOCATION OF THE SHIFT POINT

Two types of estimators of the location of the shift point τ, appear in the literature : Bayesian and maximum likelihood. El-Sayyad [17], Smith [62], Broemeling [1], Zacks [70; pp. 311] and others, give the general Bayesian framework for inference concerning the location of the shift point , τ, in an AMOC model.

Hinkley [28] studied the maximum likelihood estimator. We stard with an example concerning the Bayesian estimation and proceed then to present Hinkley's results.

4.1. - Bayesian Estimation of the Change Point

The Bayesian procedure is to derive the posterior distribution of the change point τ, and determine the estimator which minimizes the posterior risk, for a specified loss function.

If the loss function for estimating τ by $\hat{\tau}$ is $L(\hat{\tau},\tau) = |\hat{\tau}-\tau|$, then the Bayes estimator of the change point is the median of the posterior distribution of τ, given $\underset{\sim}{X}_n$. For example, suppose that $X_1,\dots,X_n$ are independent random variables having normal distributions $N(\theta_i,1)$, where

$$\theta_1 = \dots = \theta_\tau = \theta_o, \quad \theta_{\tau+1} = \dots = \theta_n = \theta_o + \delta,$$

with θ_o known ($\theta_o = 0$ say). Furtherfore, assume that the prior distribution of δ is normal, $N(0,\sigma^2)$, independently of τ, and τ has prior probabilities $\Pi(t) = P\{\tau=t\}$, $t = 1,\dots,n$. Here $\{\tau = n\}$ indicates the event of no change.

The posterior probabilities of τ for this model are

$$(4.1) \quad \Pi(t|X_n) = \frac{\Pi(t)(1+(n-t)\sigma^2)^{1/2} \exp\left\{\dfrac{(\bar{X}^*_{n-t})^2(n-t)^2\sigma^2}{2(1+(n-t)\sigma^2)}\right\}}{\sum\limits_{j=1}^{n} \Pi(j)(1+(n-j)\sigma^2)^{1/2} \exp\left\{\dfrac{(\bar{X}^*_{n-j})^2}{2}\,\dfrac{(n-j)^2\sigma^2}{(1+(n-j)\sigma^2)}\right\}}$$

where $\bar{X}^*_{n-t} = \frac{1}{n-t}\sum\limits_{i=t+1}^{n} X_i$ is the average of the last $(n-t)$ observations.

The median of the posterior distribution is then the Bayes estimator of τ, namely

(4.2) $\hat{\tau}(\underset{\sim}{X}_n)$ = least positive integer t, such that

$$\sum_{i=0}^{t} \Pi(i|\underset{\sim}{X}_n) \geq .5.$$

In the following table we present the posterior probabilites (4.1) computed for the values of four simulated samples. Each sample consists of n=20 normal variates with means θ_i and variance 1. In all cases $\theta_o = 0$. Case I consists of a sample with no change in the mean, $\delta = 0$. Cases II-IV have a shift in the mean at $\tau = 10$, and $\delta = .5$, 1.0 and 2.0. Furthermore, the prior probabilities of τ are $\Pi(t) = p(1-p)^{t-1}$ for $t=1,\ldots,n-1$ and $\Pi(n) = (1-p)^{n-1}$, with p = .01 ; and the prior variance of δ is $\sigma^2 = 3$.

Table 4.1 Posterior Probabilities of $\{\tau=t\}$

t \ δ	0	.5	1.0	2.0
1	0.002252	0.012063	0.003005	0.000000
2	0.004284	0.016045	0.002885	0.000000
3	0.004923	0.016150	0.002075	0.000000
4	0.006869	0.022634	0.002193	0.000000
5	0.006079	0.008002	0.002202	0.000001
6	0.004210	0.006261	0.002291	0.000050
7	0.004020	0.006735	0.001954	0.000026
8	0.002867	0.015830	0.001789	0.000015
9	0.003534	0.015914	0.001959	0.001087
10	0.002972	0.011537	0.002228	0.068996
11	0.003033	0.019014	0.002708	0.908434
12	0.003070	0.010335	0.002661	0.016125
13	0.003395	0.006026	0.002996	0.005237
14	0.003087	0.003201	0.003017	0.000009
15	0.004064	0.003461	0.003096	0.000011
16	0.003355	0.002709	0.002820	0.000009
17	0.004991	0.002899	0.003078	0.000000
18	0.009664	0.003486	0.004004	0.000000
19	0.007255	0.006106	0.012432	0.000000
20	0.916077	0.811593	0.940607	0.000000

We see in Table 4.1. that Bayes estimator for Cases I-III is $\hat{\tau} = 20$ (no change), while in Case IV it is $\hat{\tau}=11$. That is, if the magnitude of change in the mean is about twice the standard deviation of the random variables, the posterior distribution is expected to have its median close to the true change point. In many studies (for example, Smith [62]) the Bayesian model is based on the assumption of equal prior probabilities of $\{\tau=t\}$. Such prior probabilities yield in the above cases the following posterior probabilities.

Table 4.2. Posterior Probabilities of $\{\tau=t\}$, for Equal Prior Probabilities

t \ δ	0	.5	1.0	2.0
1	0.023329	0.030693	0.019734	0.000001
2	0.020209	0.031206	0.037996	0.000006
3	0.024060	0.028078	0.125330	0.000035
4	0.023996	0.024921	0.081694	0.000149
5	0.023063	0.026290	0.083705	0.002859
6	0.022546	0.030888	0.111434	0.005653
7	0.022951	0.042321	0.079959	0.001071
8	0.029850	0.036347	0.059293	0.005238
9	0.043298	0.030515	0.026376	0.029615
10	0.043976	0.031933	0.069415	0.931462
11	0.052939	0.033107	0.020594	0.014332
12	0.059540	0.037187	0.034396	0.008651
13	0.065588	0.048819	0.033543	0.000431
14	0.037356	0.040960	0.052289	0.000457
15	0.060050	0.049399	0.043785	0.000037
16	0.055957	0.055566	0.048865	0.000004
17	0.049753	0.069433	0.022328	0.000000
18	0.050994	0.085113	0.034621	0.000000
19	0.156117	0.092993	0.012691	0.000000
20	0.134429	0.174230	0.001955	0.000000

As seen in Table 4.2, the Bayes estimator $\hat{\tau}$ when $\delta=2$ is exactly at the true point of change $\tau=10$. On the other hand, when $\delta=0$ the estimate is $\hat{\tau}=16$. Smith derived formulae of the Bayes estimators for cases of sequences of Bernouilli trials [62], and for switching linear regression problems [63]. Bayesian estimators for the location of the shift parameter for switching regression problems are given also by Ferriera [19], Holbert and Broemeling [32], Tsurumi [65] and others.

4.2. - Maximum Likelihood Estimators

Let $X_1,X_2,\ldots,X_n$ be a sequence of independent random variables. As before, assume that

$$X_1,X_2,\ldots,X_\tau \sim F_0(x)$$

and

$$X_{\tau+1},\ldots,X_n \sim F_1(x),$$

where $F_o(x)$ and $F_1(x)$ are specified distributions. τ is the unknown point of shift. The maximum likelihood estimator (MLE) of τ is

(4.3) $\hat{\tau}_n$ = least positive integer t

t = 1,...,n, maximizing $S_{n,t}$, where

(4.4.)
$$S_{n,t} = \begin{cases} \sum_{i=1}^{t} \log f_o(X_i) + \sum_{i=t+1}^{n} \log f_1(X_i) & , \text{ if } t = 1,\dots,n-1 \\ \sum_{i=1}^{n} \log f_o(X_i) & , \text{ if } t = n. \end{cases}$$

$f_o(x)$ and $f_1(x)$ are the p.d.f.'s corresponding to $F_o(x)$ and $F_1(x)$. We present here the method of deriving the asymptotic distribution of $\hat{\tau}_n$, as n and $\tau \to \infty$, following the development of Hinkley [28].

Let $U_i = \log f_o(X_i) - \log f_1(X_i)$, $i = 1,2,\dots,n$.

Since $S_{n,t} = \sum_{i=1}^{t} U_i + \sum_{i=1}^{n} \log f_1(X_i)$, it readily follows that $\hat{\tau}_n$ is the least positive integer maximizing $V_t = \sum_{i=1}^{t} U_i$ $(t = 1,\dots,n)$. Consider the sequence $W_t = V_t - V_\tau$, where τ is the true point of shift. For very large value of τ ($\tau \to \infty$) consider the backward and forward sequences

$$W = \{0, -U_\tau, -U_\tau - U_{\tau-1}, \dots, -\sum_{j=0}^{k} U_{\tau-j}, \dots\}$$

and

$$W' = \{0, U_{\tau+1}, \dots, \sum_{j=0}^{k} U_{\tau+j}, \dots\}.$$

Let $M = \sup_{0 \le k \le \tau} \{-\sum_{j=0}^{k} U_{\tau-j}\}$ and $M' = \sup_{0 \le k \le \infty} \{\sum_{j=0}^{k} U_{\tau+j}\}$,

and $Y_j = -U_{\tau-j+1}$, $Y'_j = U_{\tau+j}$, $j = 1,\dots$. Thus,

$$W = \{0, Y_1, Y_1+Y_2, \dots\}, \qquad W' = \{0, Y'_1, Y'_1+Y'_2, \dots\}.$$

Let $\hat{\tau}$ be the point at which Max{M,M'} occurs. Notice that

$$\{\hat{\tau} = \tau\} \equiv \{M = M' = 0\},$$

(4.5)
$$\{\hat{\tau} = \tau+k\} \equiv \{M' > 0 \text{ and } M' > M\},$$

$$\{\hat{\tau} = \tau-k\} \equiv \{M > 0 \text{ and } M > M'\}.$$

Accordingly, since the sequences W and W' are independent,

$$P\{\hat{\tau} = \tau\} = P[M=0]P[M'=0] \quad ,$$

(4.6) $$P\{\tau = \tau+k\} = P\{M' > 0,\ M' > M,\ I'=k\} \quad ,$$

and

$$P\{\hat{\tau} = \tau-k\} = P\{M > 0,\ M \geq M',\ I = k\},$$

where

$$I = \inf \{k \ ; \ M = \sum_{j=1}^{k} Y_j\},$$

$$I' = \inf \{k \ ; \ M' = \sum_{j=1}^{k} Y_j'\}.$$

Thus let $\beta_k(x)dx = P\{I=k \ , \ x \leq M \leq x + dx\}$ and $\beta_k'(x)dx = P\{I'=k,\ x \leq M' \leq x + dx\}$. Furthermore, let $\alpha(x)$ and $\alpha'(x)$ be the c.d.f. of M and M', respectively.
Then

$$P\{\hat{\tau} = \tau\} = \alpha(0)\ \alpha'(0) \ ,$$

(4.8) $$P\{\hat{\tau} = \tau+k\} = \int_0^\infty \beta_k'(x)\alpha(x)dx \ ,$$

and $$P\{\hat{\tau} = \tau-k\} = \int_0^\infty \beta_k(x)\ \alpha'(x)dx.$$

5 - DYNAMIC CONTROL PROCEDURES

There are numerous papers on dynamic control problems, all of which deal in one way or another with the problem of shift at unknown time points. In particular we mention here the papers of Girshick and Rubin [2], Bather [7,8] Lorden [43], Yadin and Zacks [68], Shiryaev [60,61], and Zacks and Barzily [69].

We present first the Bayesian theory, followed by discussed of the CUSUM procedure. Again, we consider a sequence of independent random variables $X_1, X_2, \ldots, X_{m-1}, X_m, \ldots$ Let τ be the point of shift, $\tau = 0,1,\ldots$ If $\tau \leq 1$, all the observations are from $F_1(x)$. If $\tau = t$ ($t = 2,3,\ldots$) then the first t-1 observations are from $F_o(x)$ and X_t, $X_{\tau+1}, \ldots$ are from $F_1(x)$. Let $f_o(x)$ and $f_1(x)$ be the p.d.f. corresponding to $F_o(x)$ and $F_1(x)$, respectively.
The random variables $X_1, X_2, \ldots$ are observed sequentially and we wish to apply a stopping rule which will stop soon after the shift occurs, without too many "false alarms". The following objectives are considered in the selection of a stopping variable N :

1°) If $\Pi(\tau)$ denotes the prior distribution of τ, then the prior risk

$$(5.1) \qquad R(\Pi,N) = P_{\Pi}(N < \tau) + c\, P_{\Pi}(N \geq \tau)\, E_{\Pi}\{N-\tau \mid N \geq \tau\}$$

is minimized, with respect to all stopping rules.

2°) To minimize $E_{\Pi}\{N-\tau \mid N \geq \tau\}$ subject to the constraint $P_{\Pi}(N < \tau) \leq \alpha$, $0 < \alpha < 1$.

5.1. - The Bayesian Procedures

The shift index, τ, is considered a random variable, having a prior p.d.f. $\Pi(t)$, concentrated on the non-negative integers. Shiryaev [60] postulated the following prioir distribution

$$(5.2) \qquad \Pi(t) = \begin{cases} \Pi & , \text{ if } t = 0 \\ (1-\Pi)p(1-p)^{t-1} & , \text{ if } t = 1,2,\ldots \end{cases}$$

for $0 < \Pi < 1$, $0 < p < 1$. $(\Pi+(1-\Pi)p)$ is the prior probability that the shift has occurred before the first observation, and p is the prior probability of a shift occurring between any two observations.
After observing $X_1,\ldots,X_n$, the prior p.d.f. $\Pi(t)$ is converted to a posterior probability function on $\{n,n+1,\ldots\}$, namely,

$$(5.3) \qquad \Pi_n(t) = \begin{cases} \Pi_n & ; t = n \\ (1-\Pi_n)p(1-p)^{t-1} & , t = n+1,\ldots \end{cases}$$

where Π_n is the posterior probability that the shift took place before the n-th observation. This posterior probability is given by $\Pi_n = 1-q_n$, where

$$(5.4) \qquad q_n = \frac{(1-\Pi)(1-p)^n}{D_n} \prod_{i=1}^{n} f_o(X_i),$$

and

$$(5.5) \qquad \begin{aligned} D_n = {} & (\Pi+(1-\Pi)p) \prod_{i=1}^{n} f_1(X_i) + \\ & (1-\Pi)p \sum_{j=1}^{n-1} (1-p)^{j-1} \prod_{i=1}^{j} f_o(X_i) \prod_{i'=j+1}^{n} f_1(X_i) + \\ & (1-\Pi)(1-p)^n \prod_{i=1}^{n} f_o(X_i)\ . \end{aligned}$$

Let $R(X_i) = f_1(X_i)/f_o(X_i)$, $i = 1,2,\ldots$, then

$$(5.6) \qquad q_{n+1} = \frac{(1-\pi)(1-p)^{n+1}}{R(X_{n+1})[D_n-(1-\Pi)(1-p)^n]+B_{n+1}}$$

where

$$B_{n+1} = R(X_{n+1})(1-\Pi)(1-p)^n p + (1-\Pi)(1-p)^{n+1} .$$

But $(1-\Pi)(1-p)^n = q_n D_n$. Hence,

$$(5.7) \qquad q_{n+1} = \frac{q_n(1-p)}{R(X_{n+1})(1-q_n(1-p))+q_n(1-p)} ,$$

or

$$(5.8) \qquad \Pi_{n+1} = \frac{(\Pi_n+(1-\Pi_n)p-R(X_{n+1})}{(\Pi_n+(1-\Pi_n)p)R(X_{n+1})+(1-\Pi_n)(1-p)} ,$$

$n = 0,1,\ldots$ with $\Pi_o = \Pi$ and $q_o = 1-\Pi$. Accordingly, the sequence of posterior probabilities $\{\Pi_n ; n \geq 0\}$ is Markovian, i.e., the conditional distribution of Π_{n+1} depends on the first n observations $X_1,\ldots,X_n$, only through Π_n. This can lead immediatly to the construction of recursive determination of the distribution of any stopping variable depending only on Π_n (see Zacks [71]). Shiryaev [60] has shown that when F_o and F_1 are known, the optimal stopping variable, with respect to the above objectives, is to stop at the smallest n for which $\Pi_n \geq A^*$, for some $0 < A^* < 1$.

Bather [7] has shown that for the constraint of bounding the expected number of false alarms by η, $A^* = (\eta+1)^{-1}$ is the optimal stopping boundary.

When the distributions F_o and F_1 are not completely specified, the above problem of finding optimal stopping variables becomes much more complicated. Zacks and Barzily [69] studied Bayes procedures for detecting shifts in the probability of success, θ, of Bernouilli trials, when the values θ_o, before the shift, and the value θ_1 after it, are unknown. The Bayesian model assumed that θ_o and θ_1 have a uniform prior distribution over the simplex $\{(\theta_o,\theta_1) ; 0 < \theta_o \leq \theta_1 < 1\}$ and the point of shift, τ, has the prior distribution (5.2). In this case, the posterior probability Π_n depends on the whole vector of observations $X_1,\ldots,X_n$, and not only on Π_{n-1} and X_n. It is shown that this posterior probability is a function of $\underset{\sim}{X}_n = (X_1,\ldots,X_n)'$ given by

$$(5.9) \qquad \Pi_n(\underset{\sim}{X}_n) = 1-(1-\Pi)(1-p)^{n-1}B(T_n+1,\ n-T_n+2)/D_n(\underset{\sim}{X}_n)$$

where $B(p,q)$ is the beta-function ;

$T_j = \sum_{i=1}^{j} X_i$ and

$$D_n(\underset{\sim}{X}_n) = \Pi\, B(T_n+2,\ n-T_n+1) +$$

$$(1-\Pi)p \sum_{j=1}^{n-1} (1-p)^{j-1}\ B(T_{n-j}^{(n)}+1\ ,\ n-j-T_{n-j}^{(n)}+1).$$

(5.10)

$$\sum_{i=0}^{T_{n-j}^{(n)}} \binom{n-j-1}{i} B(T_n+1,\ n-T_n+2)$$

$$+\ (1-\Pi)(1-p)^{n-1}\ B(T_n+1,\ n-T_n+2).$$

Here, $T_{n-j}^{(n)} = T_n - T_j$ $(j = 0,\dots,n)$. The sequence $\{\Pi_n(\underset{\sim}{X}_n)\ ;\ n \geq 1\}$ is not Markovian, but is submartingale. Zacks and Barzily considered the problem of determining the optimal stopping rule under the following cost conditions :

After each observation we have the option to stop observations and declare that a shift has occurred. The process is then inspected. If the shift has not yet occurred a penalty of 1 unit is imposed. If, on the other hand, the shift has already occurred, a penalty of C units per delayed observation (or time unit) is imposed. It is shown then that the optimal stopping variable is

(5.11) $\qquad N^\circ$ = least $n \geq 1$, such that $\Pi_n(\underset{\sim}{X}_n) \geq b_n(\underset{\sim}{X}_n)$,

where the stopping boundary $b_n(\underset{\sim}{X}_n)$ is given implicitly, as the limit for $j \to \infty$, of

$$b_n^{(j)}(\underset{\sim}{X}_n) = \min\left(\Pi^* - \frac{M_n^{(j-1)}(\underset{\sim}{X}_n)}{C+p}\ ,1\right), \tag{5.12}$$

with $\Pi^* = p/(C+p)$ are the functions $M_n^{(j)}(\underset{\sim}{X}_n)$ can be determined recursively, according to the formula

$$M_n^{(j)}(\underset{\sim}{X}_n) = E\{\min\ (0, C\ \Pi_{n+1}(\underset{\sim}{X}_n, X_{n+1})\} \tag{5.13}$$

$$-p(1-\Pi_{n+1}(\underset{\sim}{X}_n, X_{n+1})) + M_{n+1}^{(j-1)}\ (\underset{\sim}{X}_n, X_{n+1}))\,|\,\underset{\sim}{X}_n\}.$$

It is very difficult, if not impossible, to determine these functions explicitly, for large values of j. The authors therefore considered a suboptimal procedure bases on $b_n^{(2)}(\underset{\sim}{X}_n)$ only. Numerical simulations illustrate the performance of the suboptimal procedure.

5.2 - Asymptotically Minimax Rules and the CUSUM Control

Lorden [42,43] considered the sequential detection procedure from a non-Bayesian point of view and proved that the well known CUSUM procedures of Page [47,48,49] are asymptotically minimax.

Let $X_1, X_2, \ldots$ be a sequence of independent random variables. The distributions of $X_1, \ldots, X_{m-1}$ is $F_o(x)$ and that of $X_m, X_{m+1}, \ldots$ is $F_1(x)$. The point of shift m is unknown , $F_o(x)$ and $F_1(x)$ are known. The family of probability measures is $\{P_m\ ;\ m = 1,2,\ldots\}$, where $P_m(\underset{\sim}{X}_n)$ is the joint p.d.f. of $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$, in which X_m is the first random variable with a c.d.f. $F_1(x)$.

It is desired to devise a sequential procedure with a (possibly) extended stopping variable, N, (i.e., $\lim_{n\to\infty} P_m[N>n] \geq d > 0$, $m = 0,1,\ldots$) which minimizes the largest possible expectation of delayed action, and does not lead to too many false alarms. More precisely, if $P_o(\underset{\sim}{X})$ denotes the c.d.f. under the assumption that all observations have $F_o(X)$ as a c.d.f. ; and if $E_m\{.\}$ denotes expectation under $P_m(.)$, the objectife is to minimize

$$E_1\{N\} = \sup_{m\geq 1} \ \text{ess sup}\ E_m\{(N-m-1)^+ | \mathcal{F}_{m-1}\} \tag{5.14}$$

subject to the constraint

$$E_o\{N\} \geq \gamma^* , \qquad 1 < \gamma^* < \infty . \tag{5.15}$$

$E_m\{.|\mathcal{F}_{m-1}\}$ denotes the conditional expectation given the σ-field generated by $(X_1, \ldots, X_{m-1})$. It is proven by Lorden [43] that an asymptotically minimax procedure, as $\gamma \to \infty$, is provided by Page's procedure, which is described below.

Let $R(X_i) = f_1(X_i)/f_o(X_i)$, $i = 1,2,\ldots$ where $f_i(x)$ is the p.d.f. corresponding to $F_i(X)$, $i = 0,1$. Let

$S_k = \sum_{i=1}^{k} \log R(X_i)$, $k=1,2,\ldots$ and $T_n = S_n - \min_{k\leq n} S_k$. Then for $\gamma = \log\ \gamma^*$

$$N^* = \text{least } n \geq 1 \text{ such that } T_n \geq \gamma \tag{5.16}$$

is Page's (extended) stopping variable.

The statistic T_n can be computed recursively by the formula

$$T_{n+1} = (T_n + \log R(X_{n+1}))^+ , \qquad n = 0,1,\ldots \tag{5.17}$$

$$T_o \equiv 0.$$

The above detection procedure can be considered as a sequence of one-sided Wald's SPRT with boundaries $(0,\gamma)$. Whenever the T_n statistic hits the lower boundary, 0, the SPRT is recycled, and all the previous observations can be discarded. On the other hand, for the first time $T_n \geq \gamma$ the sampling process is stopped. The repeated cycles are independent and identically distributed. Thus, Wald's theory of SPRT can be used to obtain the main results of the present theory. Let α and β be the error probabilities in each such independent cycle of Wald's SPRT ; i.e., $\alpha = P_o[T_n \geq \gamma]$ and $\beta = P_1[T_n = 0]$. Let N_1 be the length of a cycle. Accordingly,

(5.18) $$E_o\{N^*\} = \frac{1}{\alpha} E_o\{N_1\}$$
and
$$E_1\{N^*\} = \frac{1}{1-\beta} \; E_1\{N_1\}.$$

Set $\gamma^* = \frac{1}{\alpha}$, then the constraint (5.15) is satisfied, since $E_1\{N_1\} \geq 1$. Moreover, Lorden proved that $E_1\{N^*\} = \bar{E}_1\{N^*\}$. Finally, applying well known results on the expected sample size in Wald's SPRT, we obtain

(5.19) $$E_1\{N^*\} \simeq \frac{\log \alpha}{I_1}, \text{ as } \alpha \to 0,$$

where $I_1 = E_1 \{\log \frac{f_1(X)}{f_o(X)}\}$ is the Kull back-Leibler information for discriminating between F_o and F_1.
The right hand size of (5.19) was shown to be the asymptotically minimum expected sample size. Thus, Page's procedure is asymptotically minimax.
in [42] Lorden and Eisenberg applied the theory presented here to solve a problem of life testing for a reliability system. It is assumed that the life length of the system is distributed exponentially, with intensity (failure-rate) λ. At an unknown time point, θ, the failure rate shifts from λ to $\lambda(1+\eta)$, $0 < \eta_1 \leq \eta \leq \eta_2 < \infty$. Approximations to the formulae of $E_o\{N^*\}$ and $E_\eta\{N^*\}$ are given, assuming that λ is known. By proper transformations of the statistics the detection procedure can be applied also to cases of unknown λ. It is interesting to present some of the numerical results of this study. For the case of $\lambda = 1$ and $\alpha = 1/\gamma$. The expected number of observations required is

η	γ	$E_o\{N\}$	$\bar{E}_\eta\{N\}$
.4	20	422	48
.6	50	676	36
.9	40	342	20

Page's CUSUM procedure is thus very conservative, relative to the Bayes procedures, which detect the shifts fast, but have also small $E_o\{N\}$.

6 - REFERENCES

[1] BAGSHAW, M. and JOHNSON, R.A. (1975) The influence of reference values and estimated variance on the ARL of CUSUM tests. J.R.S.S., B, 37, 413-420.

[2] BAGSHAW, M. and JOHNSON, R.A. (1975) Sequential detection of a drift change in a Wiener process. Commu. Statist. 4, 787-796

[3] BAGSHAW, M. and JOHNSON, R.A. (1975) The effect of serial correlation on the performance of CUSUM tests II. - Technometrics, 17, 73-80

[4] BALMER, D.W. (1975) On a quickest detection problem with costly information. J. Appl. Prob., 12, 87-97.

[5] BALMER, D.W. (1981) On quickest detection problem with variable monitoring. J. Appl. Prob., 18, 760-767

[6] BARNARD, G.A. (1959) Control charts and stochastic processes. J. Roy. Statist. Soc. B, 21, 239-271.

[7] BATHER, J.A. (1967) On a quickest detection problem. Ann. Math. Statist., 38, 711-724

[8] BATHER, J.A. (1976) A control chart model and a generalized stopping problem for Brownian motion. Math. Oper. Res., 1, 209-224.

[9] BHATTACHARYYA, G.K. and JOHNSON, R.A. (1968) Non-parametric tests for shift at an unknown time point. Ann. Math. Statist., 39, 1731-1743.

[10] BROEMELING, L. (1974) Estimating the future values of changing sequences. Commu. Statist., A, 6, 87-102.

[11] BROEMELING, L.D. (1974) Bayesian inferences about a changing sequence of random variables. Commun. Statist., 3, 234-255.

[12] BROWN, R.L., DURBIN, J. and EVANS, J.M. (1975) Techniques for testing the constancy of regression relationships over time. J. Royal Statist. Soc., B 37, 149-192.

[13] BROWN, R.L. and DURBIN, J. (1968) Methods of investigating whether a regression relationship is constant over time. Selected Statistical Papers I. Amsterdam : Math. Centrum, European Meeting, 37-45.

[14] CHERNOFF, H. and ZACKS, S. (1964) Estimating the current mean of a normal distribution which is subjected to changes in time. Ann. Math. Statist., 35, 999-1028.

[15] COOB, G.W. (1978) The problem of the Nile : conditional solution to a change point problem. Biometrika, 62, 243-251

[16] DARKHOVSHK, B.S. (1976) A non-parametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. Theory of Prob. Appl., 21, 178-183.

[17] EL-SAYYAD, G.M. (1975) A Bayesian analysis of the change-point problem. Egypt. Statist.J.; 19, 1-13.

[18] FARLEY, J.U. and HINICH, M.J. (1970) Detecting "small" mean shifts in time series. Management Science, 17, 189-199.

[19] FERRIERA, P.E. (1975) A Bayesian analysis of switching regression model: Known number of regimes. J. Amer. Statist. Assoc., 70, 370-374

[20] FISZ, M. (1963) Probability Theory and Mathematical Statistics. 3rd Edition. John Wiley and Sons, New-York.

[21] GARDNER, L.A. Jr. (1969) On detecting changes in the mean of normal variates. Ann. Math. Statist., 40, 114-115.

[22] GIRSHICK, M.A. and RUBIN, H. (1952) A Bayes appoach to a quality control model. Ann. Math. Statist. 23, 114-115.

[23] HAWKINS, D.M. (1977) Testing a sequence of observations for a shift in location. J. Amer. Statist. Assoc., 72, 180-186.

[24] HAWKINS, D.M. (1980) A note on continuous and discontinuous segmented regression. Technometrics, 22, 443-444.

[25] HINES, W.G.S. (1976) A simple monitor of a system with sudden parameter changes. IEEE Trans. Inf. Theory, IT, 210-216

[26] HINKLEY, D.V. and HINKLEY, E.A. (1970) Inference about the change-point in a sequence of Binomial random variables. Biometrika, 57, 477-488.

[27] HINKLEY, D.V. (1969) Inference about the intersection in two-phase regression. Biometrika, 56, 495-504.

[28] HINKLEY, D.V. (1970) Inference about the change-point in a sequence of random variables. Biometrika, 57, 1-16.

[29] HINKLEY, D.V. (1971) Inference in two phase regression. J. Amer. Statist. Assoc., 66, 736-743.

[30] HINKLEY, D.V. (1971) Inference about the change-point from a cumulative sum tests. Biometrika, 58, 509-523.

[31] HINKLEY, D.V. (1972) Time-ordered classification. Biometrika, 59, 509-523.

[32] HOLBERT, D. and BROEMELING, L. (1977) Bayesian inferences related to shifting sequences and two-hase regression. Commun. Statist. A, 6, 265-275.

[33] HSU, D.A. (1977) Tests for variance shift at an unknown time point. Applied Statist., 26, 279-284.

[34] HSU, D.A. (1979) Detecting shifts of parameter in Gamma sequences with applications to stock price and air traffic flow analysis. J. Amer. Statist. Asso., 74, 31-40

[35] INSELMANN, E.H. and ARSENAL, F. (1968) Tests for several regression equations. Ann. Math. Statist., 39, 1362.

[36] KANDER, A. and ZACKS, S. (1966) Test procedures for possible changes in parameters of statistic at distributions occuring at unknown time points. Ann. Math. Statist., 37, 1196-1210.

[37] KHAN, R.A. (1975) A sequential detection procedure. Tech. Rep. n° 17, Depart. of Statistics, CWRU.

[38] KHAN, R.A. (1979) Some first passage problems related to CUSUM procedures. Stoch. Proc. Appl., 9, 207-216.

[39] LAI, T.L. (1973) Gaussian processes, moving averages and quickest detection problems. Ann. Prob., 1, 825-837

[40] LAI, T.L. (1974) Control charts based on weighted sums. Ann. Statist. 2, 134-147.

[41] LEE, A.F.S. and HEGHINIAN, S.M. (1977) A shift of the mean level in a sequence of independent normal random variables - a Bayesian approach. Technometrics, 19, 503-506.

[42] LORDEN, G. and EISENBERG, I. (1973) Detection of failure rate increases. Technometrics, 15, 167-175

[43] LORDEN, G. (1971) Procedures for reacting to a change in distribution. Ann. Math. Statist., 42, 1897-1908.

[44] MARONNA, R. and YOHAI, V.J. (1978) A bivariate test for the dectection of a systematic change in mean. J. Amer. Statist. Assoc., 73, 640-645.

[45] MUSTAFI, C.K. (1968) Inference problems about parameters which are subjected to changes over time. Ann. Math. Statist., 39, 840-854.

[46] NADLER, J. and ROBBINS, N.D. (1971) Some characteristics of Page's two sided procedure for detecting a change in the location parameter. Ann. Math. Statist., 42, 538-551.

[47] PAGE, E.S. (1954) Continuous inspection schemes. Biometrika 41, 100-115.

[48] PAGE, E.S. (1955) A test for a change in a parameter occuring at an unknown point. Biometrika, 42, 523-526.

[49] PAGE, E.S. (1957) On problem in which a change in a parameter occurs at an unknown point. Biometrika, 44, 248-252

[50] PETTITT, A.N. (1979) A non parametric approach to the change-point problem. Appl. Statist., 28, 126-135.

[51] QUANDT, R.E. (1958) The estimation of the parameters of a linear regression system obeys two separate regimes. J. Amer. Statis. Assoc.

[52] QUANDT, R.E. (1960) Tests of the hypothesis that a linear regression system obeys two separate regimes. J. Amer. Statist. Assoc., 55, 324-330.

[53] RAO, P.S.E.S. (1972) On two phase regression estimator. Sankhya, 34, 473-476

[54] SCHWEDER, T. (1976) Some 'optimal' methods to detect structural shift or outliers in regression. J. Amer. Statist. Assoc., 71, 491-501.

[55] SEN,A. and SRIVASTAVE, M. (1975) On tests for detecting change in the mean when variance is unknown. Ann. Inst. Statist. Math., 27, 593-602

[56] SEN, A. and SRIVASTAVA, M.S. (1975) On tests for detecting changes in means. Annals of Statist., 3, 98-108

[57] Sen, A. and SRIVASTAVA, M.S. (1975) On one sided tests for change in level. Technometrics, 17, 61-64.

[58] SEN, A.K. and SRIVASTAVA, M.S. (1973) On multivariate tests for detecting change in mean. Sankhya, 35, 173-186.

[59] SHABAN, S.A. (1980) Change point problem and two-hase regression : an annotated bibliography. Inst. Statist. Review, 48, 83-93

[60] SHIRYAEV, A.N. (1973) On optimum methods in quickest detection problems. Theory Prob. Appl., 8, 22-46

[61] SHIRYAEV, A.N. (1973) Statistical Sequential Analysis: Optimal Stopping Rules. Translations of Mathematical Monographs, Vol. 38, American Mathematical Society, Providence RI.

[62] SMITH, A.F.M. (1975) A Bayesian approach to inference about change-point in sequence of random variables. Biometrika, 62, 407-416.

[63] SMITH, A.F.M. (1977) A Bayesian analysis of some time varying models. Recent Developments in Statistics, J.R. Barra, et.al., Editors, North Holland, New York.

[64] SWAMY, P.A.V.B. and MEHTA, J.S. (1975) Bayesian and non-Bayesian analysis of switching regressions and of random coefficient regression. J. Amer. Statist. Assoc., 70, 593-602

[65] TSURUMI, H. (1977) A Bayesian test of a parameter shift and an application. Jour. of Econometrics, 6, 371-380

[66] VON MISES, R. (1964) Mathematical Theory of Probability and Statistics Academic Press, New York.

[67] WHICHERN, D.W., MILLER, R.B. and HSU D.A. (1976) Change of variance in first-order autoregressive time series models with an application. Appl. Statist. 25, 248-256.

[68] ZACKS, S. and YADIN, M. (1978) Adaptation of the service capacity in a queueing system which is subjected to a change in the arrival rate... Adv. Appl. Prob., 10, 666-681.

[69] ZACKS, S. and BARZILY, Z. (1981) Detecting a shift in the probability of success in a sequence of Bernouilli trials. J. Statist. Planning and Inf., 5, 107-119.

[70] ZACKS, S. (1981) Parametric Statistical Inference : Basic Theory and Modern Approaches. Pergamon Press, Oxford.

[71] ZACKS, S. (1981) The probability distribution and the expected value of a stopping variable associated with one-sided CUSUM procedures... Commum. Statist., A, 10, 2245-2258.