ESTIMATING THE CURRENT MEAN OF A NORMAL DISTRIBUTION WHICH IS SUBJECTED TO CHANGES IN TIME

BY

H. CHERNOFF AND S. ZACKS

TECHNICAL REPORT NO. 91
OCTOBER 31, 1963

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE,
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
Nens-225(52) (NR 342-022)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATING THE CURRENT MEAN OF A
NORMAL DISTRIBUTION WHICH IS
SUBJECTED TO CHANGES IN TIME

by

H. Chernoff and S. Zacks

TECHNICAL REPORT NO. 91
October 31, 1963

PREPARED UNDER CONTRACT Nonr-225(52)
(NR-342-022)
FOR
OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATING THE CURRENT MEAN OF A NORMAL DISTRIBUTION WHICH IS SUBJECTED TO CHANGES IN TIME

by

H. Chernoff and S. Zacks

1. Introduction

The present study was motivated by consideration of a "tracking" problem. Observations are taken on the successive positions of an object traveling on a path, and it is desired to estimate its current position. If the path is smooth, regression estimates seem appropriate. If, however, the path is subject to occasional changes in direction, regression will give misleading results long after a naive observer would have made corrections. Our objective is to arrive at a simple formula which implicitly accounts for possible changes in direction and discount observations taken before the latest change.

To develop insight into the nature of a reasonable procedure, we study a simpler problem. In this problem successive observations are taken on \( n \) independently and normally distributed random variables \( X_1, X_2, \ldots, X_n \) with means \( \mu_1, \mu_2, \ldots, \mu_n \) and variance \( 1 \). Each mean \( \mu_i \) is equal to the preceding mean \( \mu_{i-1} \) except when an occasional change takes place. The object is to estimate the current mean \( \mu_n \).

We shall study this problem from a Bayesian point of view. First, we assume that the time points of change obey an arbitrary specified a-priori probability distribution appropriate to the special case being
studied. Second, we assume that the amounts of change in the means, when changes take place, are independently and normally distributed random variables, with mean 0 and variance $\sigma^2$. Third, we assume that the current mean $\mu_n$ is a normally distributed random variable with mean 0 and variance $\tau^2$. Letting $\tau^2$ approach infinity, we derive according to these assumptions a Bayes estimator of $\mu_n$ for an a-priori uniform distribution on the whole real line and a quadratic loss function. This estimator is invariant under translations of $X_i$. The minimum variance linear unbiased estimator (M.V.L.U.) of $\mu_n$ is also derived. The M.V.L.U. estimator is considerably simpler than the Bayes estimator. However, when the expected number of changes in the means is neither zero nor $n-1$ the Bayes estimator is more efficient than the M.V.L.U. one. Generally, the Bayes estimator is very difficult for applications, since it requires many involved computations. A considerable simplification is attainable in the formula for the general Bayes estimator by letting the a-priori variance of the changes, $\sigma^2$, approach zero. This simplified estimator might not be an efficient one in cases of large changes. As an alternative, we consider the problem where the a-priori distribution of time points of change is such that there is at most one change. This problem leads to a relatively simple Bayes estimator, called A.M.O.C. Bayes estimator. However, difficulties may arise if this estimator is applied when there are actually two (or more) changes. If the first change is larger than the second one, the method tends to act as though the latter change had not taken place. We shall describe an "ad-hoc" estimator, which applies a combination of the A.M.O.C. Bayes estimator and a sequence of tests designed
to locate the last time point of change. The various estimators are then compared by a Monte Carlo study of samples of size 9.

Our Bayesian approach seems to be more appropriate for the related problem of testing whether a change in mean has occurred. This problem was studied by Page [3,4]. The test procedure obtained by our approach is simpler than that used by Page. The power functions of the two procedures are compared.
2. The statistical model and distribution theory.

Let \( X = (X_1', \ldots, X_n')' \) be the column vector representing the \( n \) observations, and \( \mu = (\mu_1, \ldots, \mu_n)' \) the column vector representing the corresponding means. Then,

\[
(2.1) \quad X = \mu + \epsilon
\]

where \( \epsilon = (\epsilon_1', \ldots, \epsilon_n')' \) is the vector of observation errors. The successive means are related by:

\[
(2.2) \quad \mu_i = \mu_{i+1} + J_i Z_i \quad (i = 1, \ldots, n-1)
\]

where \( J_i (i = 1, \ldots, n-1) \) is a random variable which assumes the value 1 if there is a change between time points \( i \) and \( i+1 \), and the value 0 otherwise. \( Z_i (i = 1, \ldots, n-1) \) is a random variable representing the amount of change, when a change takes place.

Let \( J = (J_1, J_2, \ldots, J_{n-1}, 0)' \) and \( Z = (Z_1, \ldots, Z_{n-1}, 0)' \). We assume that \( \epsilon, J, Z, \) and \( \mu_n \) are independently distributed with

\[
(2.3) \quad \mathcal{L}(\epsilon) = \mathcal{N}(0, I)
\]

\[
(2.4) \quad \mathcal{L}(Z) = \mathcal{N}(0, \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & I \end{pmatrix})
\]

and

\[
(2.5) \quad \mathcal{L}(\mu_n) = \mathcal{N}(0, \tau^2)
\]

where \( \mathcal{L}(\ ) \) designates the distribution law of the variable in the bracket; \( \mathcal{N}(\theta, \Sigma) \) denotes the normal distribution law with mean \( \theta \).
and covariance matrix \( \Sigma \); and where \( I \) is the \( n \times n \) identity matrix. Without specifying it further at present we represent \( \mathcal{L}(J) \) by its density function:

\[
(2.6) \quad p(j) = P(J = j)
\]

where \( j \) is a point in an \( n \)-dimensional Euclidean space. We remark that although the \( X_i \)'s are independent for a given \( \mu \), the a-priori distribution of \( \mu \) leads to a resulting distribution for the \( X_i \)'s where they are no longer independent. Moreover, given \( \{J=j\} \), \( X \) and \( \mu \) are linear functions of the \( 2n \) linearly independent normal variables in \( \epsilon, Z \) and \( \mu_n \). Thus, the conditional joint distribution of \( X \) and \( \mu \) given \( \{J=j\} \), and other such conditional distributions, may be determined by classical normal multivariate analysis techniques. In particular, since

\[
(2.7) \quad X_i = \begin{cases} 
\mu + \epsilon_i + \sum_{k=i}^{n-1} J_k Z_k, & \text{if } i=1, \ldots, n-1 \\
\mu + \epsilon_n, & \text{if } i=n
\end{cases}
\]

we obtain from (2.3) and (2.4)

\[
(2.8) \quad \mathcal{L}(X|\mu, J) = \mathcal{N}(\mu, \epsilon, \Sigma(J))
\]

where \( \epsilon \) is the \( n \times 1 \) column vector whose elements are all 1, and

\[
(2.9) \quad \Sigma(J) = I + \sigma^2 J_J J_T
\]

in which \( J_T \) is the upper triangular \( n \times n \) matrix.
\[ J_T = \begin{bmatrix}
  J_1 & J_2 & \cdots & J_{n-1} & 0 \\
  J_2 & \cdots & J_{n-1} & 0 \\
  \vdots & & \ddots & \vdots & \vdots \\
  0 & \cdots & J_{n-1} & 0 \\
  0 & & & 0 & 0
\end{bmatrix} \]

Furthermore, since the distribution law of \( \mu_n \) is normal we derive from (2.5) and (2.8)

\[ \mathcal{L}(x|J) = \mathcal{N}(0, \Sigma^*(J)) \]

where

\[ \Sigma^*(J) = \Sigma(J) + \tau^2 ee' \]

and,

\[ \mathcal{L}(x, \mu_n|J) = \mathcal{N} \left( 0, \begin{bmatrix}
  \Sigma^*(J) & \tau^2 e^2 \\
  \tau^2 e e' & \tau^2
\end{bmatrix} \right) \]

Hence, the a-posteriori distribution law of \( \mu_n \), given \( X \) and \( J \), is

\[ \mathcal{L}(\mu_n|X,J) = \mathcal{N} \left( \frac{e' \sum_{-1} \Phi(J)X}{e' \sum_{-1} \Phi(J)e + \tau^2}, \frac{1}{e' \sum_{-1} \Phi(J)e + \tau^2} \right) \]

The distribution law \( \mathcal{L}(X|J) \) together with \( \mathcal{L}(J) \) can be combined to give \( \mathcal{L}(J|X) \) which, according to Bayes theorem, can be expressed by

\[ p(j|x) = p(j=x|X) = \frac{p(j)n(x|0, \Sigma^*(j))}{\sum_{(j)} p(j)n(x|0, \Sigma^*(j))} \]

6
where \( n(x|\theta, \Sigma) \) represents the normal density with mean \( \theta \) and covariance matrix \( \Sigma \). Finally, the a-posteriori distribution law of \( \mu_n \), given \( X \), can be represented schematically by

\[
(2.16) \quad \mathcal{L}(\mu_n | X) = \sum_{\{j\}} p(j|X) \mathcal{L}(\mu_n | X, j=j)
\]
3. **The minimum variance linear unbiased estimator.**

In the present section we derive the formula for the M.V.L.U. estimator of $\mu_n$. This estimator is a weighted average of $X_1, \ldots, X_n$ whose weights do not depend on the observations, but only on the a-priori assumptions concerning the distributions of $J$ and $Z$. We expect therefore the M.V.L.U. estimator to be efficient only in cases where the available information on $X$ does not substantially affect the a-posteriori distribution $\mathcal{L}(J|X)$. That is, when $\mathcal{L}(J|X)$ is approximately $\mathcal{L}(J)$. This is the case when changes in the means occur almost always; or when there are almost no changes.

Consider a linear estimator $\tilde{\mu}_n = a'X$, $a' = (a_1, \ldots, a_n)$. Applying (2.8) the conditional distribution of $\tilde{\mu}_n$, given $\mu_n$, has conditional mean

$$E(\tilde{\mu}_n | \mu) = a' E(X | \mu_n) = \mu_n a'e$$

and variance

$$\text{Var}(\tilde{\mu}_n | \mu) = a' V a$$

where

$$V = E(\sum (J)) .$$

Thus, $\tilde{\mu}_n$ is unbiased if, and only if, $a'e = 1$. Standard calculations, applying Lagrange Multipliers with the restrictions $a'e = 1$, yields the M.V.L.U. estimator,

$$\tilde{\mu}_n = \frac{e'^{-V^{-1}}}{e'^{-V^{-1}} e} X .$$
Let $\xi_i (i=1, \ldots, n)$ be the sum of the components of the $i^{th}$ column vector of $V^{-1}$. Then, we can write

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

(3.5)

Now we elaborate the formula for $\tilde{\mu}_n$. According to (2.9) and (3.3),

$$V = I + \sigma^2 E[J_T J_T']$$

(3.6)

Define the $n \times n$ matrix $W^{(k)}(k=1, \ldots, n)$ to be one whose upper left $k \times k$ submatrix consists of elements equal to 1, and all of whose other elements are 0. Then, as is readily verified,

$$V = I + \sigma^2 \sum_{k=1}^{n-1} J_k W^{(k)}$$

(3.7)

$$= I + \sigma^2 \sum_{k=1}^{n-1} p(j_k)W^{(k)}$$

where $p(j_k) = P(J_k = 1), k=1, \ldots, n-1$. Hence, according to (3.4) and (3.7),

$$\tilde{\mu}_n = \frac{e'[I+\sigma^2 \sum_{k=1}^{n-1} p(j_k)W^{(k)}]^{-1} X}{e'[I+\sigma^2 \sum_{k=1}^{n-1} p(j_k)W^{(k)}]^{-1} e}$$

(3.8)

In the special case where $p(j_k) = p$ for all $k=1, \ldots, n-1$ one obtains the following formula for $\xi_i$: 

\[ \xi_i = \begin{cases} \frac{1}{v_1 v_2 \cdots v_{n-2}(v_{n-1})}, & \text{if } i=1 \\ \frac{v_{i-1} - 1}{v_{i-1} \cdots v_{n-2}(v_{n-1})}, & \text{if } i=2,\ldots,n-1 \end{cases} \] 

where

\[ v_k = \begin{cases} 2+\sigma^2_p, & \text{if } k=1 \\ 2+\sigma^2_p - \frac{1}{v_{k-1}}^{-1}, & \text{if } k=2,\ldots,n-1 \end{cases} \] 

According to formulae (3.9)-(3.10), when \( \sigma^2 \) is large the weight assigned to \( X_i \) (i=1,...,n) by the M.V.L.U. estimator is of the order of magnitude of \( (\sigma^2)^{-(n-1)} \). Indeed, \( v_k = O(\sigma^2) \) as \( \sigma^2 \to \infty \), for all \( k=1,\ldots,n-1 \). Hence, \( \xi_i = O\left(\frac{1}{(\sigma^2)^{n-1}}\right) \) as \( \sigma^2 \to \infty \) for all \( i=1,\ldots,n \). On the other hand, if \( \sigma^2 \to 0 \) the weights \( \xi_i \) are almost equal. Indeed, in the case \( \sigma^2 = 0 \), \( \xi_i = \frac{1}{n} \) for all \( i=1,\ldots,n \).
4. The Bayes estimator.

Assuming a quadratic loss function, the Bayes estimator \( \hat{\mu}_n \) is the mean of the a-posteriori distribution of \( \mu_n \), given \( X \). Thus according to (2.14) and (2.16) this Bayes estimator is given by

\[
\hat{\mu}_n = \sum_{\{j\}} p(j | x) \frac{e' \sum^{-1}(j)^T x}{e' \sum^{-1}(j)e + \tau^{-2}}.
\]

Substituting (2.15) in (4.1) we obtain

\[
\hat{\mu}_n = D^{-1} \sum_{\{j\}} p(j)n(x | 0, \sum^*(j)) \frac{e' \sum^{-1}(j)^T x}{e' \sum^{-1}(j)e + \tau^{-2}}
\]

where

\[
D = \sum_{\{j\}} p(j)n(x | 0, \sum^*(j)).
\]

To derive a translation invariant estimator we shall let \( \tau^2 \to \infty \).

Then, the sum of the elements of \( e' \sum^{-1}(j)/[e' \sum^{-1}(j)e + \tau^{-2}] \) converges to 1. Hence \( \hat{\mu}_n \) converges, as \( \tau^2 \to \infty \), to a weighted average of linear functions of \( X \), each of which is a weighted average of the observations \( X_i (i=1, \ldots, n) \). Note that the coefficients of \( \hat{\mu}_n \) depend on \( X \). Hence, the Bayes estimator is generally non-linear. However, in the special case \( P(J=j_0) = 1 \), for some \( j_0 \), i.e., when all the points of change are known, the limit of \( \hat{\mu}_n \) coincides with the corresponding M.V.L.U. estimator \( \tilde{\mu}_n \).

In general, the limiting behavior of \( \hat{\mu}_n \) involves the asymptotic behavior of \( \sum^*-1(j) \) and of det. \( \sum^*(j) \), as \( \tau^2 \to \infty \).
Consider the \((n+1) \times (n+1)\) matrix,

\[
(4.4) \quad \mathbf{H} = \begin{bmatrix}
\sum (j) : \tau^2 e \\
\dotsc & \dotsc \\
\tau^2 e' & -\tau^2
\end{bmatrix}.
\]

According to a well-known formula (see, Anderson [1]) the determinant of \(\mathbf{H}\) is

\[
(4.5) \quad \det \mathbf{H} = -\tau^2 \det \left[ \sum (j) + \tau^2 ee' \right]
\]

\[
= -\tau^2 \det \sum^*(j)
\]

\[
= -\tau^2 \left[ 1 + \tau^2 e' \sum^{-1}(j)e \right] \det \sum (j).
\]

Hence,

\[
(4.6) \quad \det \sum^*(j) = \left[ 1 + \tau^2 e' \sum^{-1}(j)e \right] \det \sum (j)
\]

\[
\approx \tau^2 e' \sum^{-1}(j)e \det \sum (j), \quad \text{as } \tau^2 \to \infty.
\]

Similarly, one can prove that

\[
(4.7) \quad \sum'^{-1}(j) = \sum^{-1}(j) - \frac{\sum^{-1}(j)e e' \sum^{-1}(j)}{e' \sum e + \tau^{-2}}
\]

\[
\approx \sum^{-1}(j) - \frac{\sum^{-1}(j)e e' \sum^{-1}(j)}{e' \sum^{-1}(j)e}, \quad \text{as } \tau^2 \to \infty.
\]

The inversion of \(\sum (j)\), as well as the determination of \(\det \sum (j)\), may be facilitated by introducing the transformation represented by
\[(4.8)\]
\[A = \begin{bmatrix}
1 & -1 & 0 \\
1 & -1 & \ddots \\
0 & \ddots & \ddots & \ddots \\
& & 1 & -1 & 1
\end{bmatrix} .
\]

Then,
\[(4.9)\]
\[AJ_T = J_D = \begin{bmatrix}
J_1 & J_2 & 0 \\
0 & \ddots & \ddots \\
& & \ddots & J_{n-1} & 0
\end{bmatrix} .
\]

It is relatively easy to invert
\[(4.10)\]
\[A(\sum (J))A' = AA' + \sigma^2 J_D \]

which has zeros for all elements which are more than one position removed from the main diagonal. The inverse of \(\sum (J)\) is then
\[(4.11)\]
\[\sum^{-1}(J) = A'[AA' + \sigma^2 J_D]^{-1}A \]

and
\[(4.12)\]
\[\det \sum (J) = \det [AA' + \sigma^2 J_D] .\]

This determinant is called a **continuant** and the method of its evaluation is well known (see, Muir [2]).

Denoting by \(\hat{\mu_n}(j)\) the conditional Bayes estimator, given \(J=j\), for the uniform a-priori distribution of \(\mu_n\) on the whole real line, i.e.,
\[
\hat{\mu}_n(j) = \frac{e' \sum_{-1} X}{e' \sum_{-1} e}
\]

we get, according to \(4.2\), \(4.3\), \(4.6\) and \(4.7\) that the Bayes estimator for the uniform a-priori distribution on the whole real line is:

\[
\hat{\mu}_n = D^{-1} E_J \left\{ \frac{1}{\sqrt{e' \sum_{-1} e . \det \sum (J)}} \exp \left[ -\frac{1}{2} [X - \hat{\mu}_n(J)e]' \sum_{-1} [X - \hat{\mu}_n(J)e] \right] \sum_{-1} X \hat{\mu}_n(J) \right\}
\]

where

\[
D = E_J \left\{ \frac{1}{\sqrt{e' \sum_{-1} e . \det \sum (J)}} \exp \left[ -\frac{1}{2} [X - \hat{\mu}_n(J)e]' \sum_{-1} (X) \right] \right\}
\]

and where \(E_J\) designates the expectation operator with respect to the a-priori distribution of \(J\).
5. A reduction applied to cases of few changes.

In the present section we give another representation of the Bayes estimator, which is especially suitable for cases of a small number of changes in the mean. The derivation is parallel to that of the original Bayes estimator. The difference lies in the present emphasis of the fact that when the time points of change are known the sums of observations between these time points are sufficient statistics for \( \hat{\mu}_n \). This permits us to reduce the general \( n \times n \) matrices of section 4, when there are few changes, to lower order matrices which are easier to manipulate and incidentally add somewhat to the understanding of the nature of \( \hat{\mu}_n \). This approach will prove useful later.

Let \( J=j \) with \( j_1 = 1 \) at \( i = m_1, m_1 + m_2, \ldots, m_1 + m_2 + \cdots + m_r \) \((r=1,2,\ldots,n-1)\). In this case there are \( r \) changes in the mean, taking place at the time points \( m_1, m_1 + m_2, \ldots, m_1 + \cdots + m_r \). Let \( m = (m_1, \ldots, m_r, m_{r+1})' \) where \( m_{r+1} = n - (m_1 + \cdots + m_r) \). The vector and its dimension \( r+1 \) are determined by \( j \) and conversely \( j \) is determined by \( m \). In case \( J=j \) is such that there are \( r \) changes, we transform the random vector of observations \( X \) to an \((r+1)\)-dimensional vector \( Y \) as follows:

\[
Y_1 = X_{m_1} + \cdots + X_{m_1 + m_2} \\
Y_2 = X_{m_1 + m_2} + \cdots + X_{m_1 + m_2 + m_3} \\
\vdots \\
Y_{r+1} = X_{m_1 + \cdots + m_r} + \cdots + X_n
\]

(5.1)

The random vector \( Y \) is a function of \((X,J)\). Straightforward calculation yields,
\[(5.2) \quad \mathcal{L}(Y|\mu_n, J=j) = \mathcal{N}(\mu_n, \mathcal{G}(m))\]

where the covariance matrix \(\mathcal{G}(m)\) is given by,

\[(5.3) \quad \mathcal{G}(m) = m_D + \sigma^2 m_T m_T'\]

\(m_D\) being a diagonal matrix of order \((r+1) \times (r+1)\) whose diagonal elements are \(m_1, \ldots, m_r, m_{r+1}\); \(m_T\) being the \((r+1) \times (r+1)\) upper triangular matrix

\[(5.4) \quad m_T = \begin{bmatrix} m_1 & m_2 & \cdots & m_r & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & m_r & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}\]

In the following we prove that the Bayes estimator, for the a-priori uniform distribution on the whole real line, can be represented in terms of the variables \(m\) and \(Y\) in the following manner:

\[(5.5) \quad \hat{\mu}_n = D^{-1} \sum_{\{j\}} p(j) \psi(Y|j) \frac{m'}{m'} \mathcal{G}^{-1}(m)m \frac{\mathcal{G}^{-1}(m)y}{\mathcal{G}^{-1}(m)m}\]

where,

\[(5.6) \quad D = \sum_{\{j\}} p(j) \psi(Y|j)\]

and where

\[(5.7) \quad \psi(Y|j) = \left( \frac{m_1 \cdots m_{r+1}}{\text{det. } \mathcal{G}(m)} \right)^{1/2} \exp \left[ -\frac{1}{2} (Y-\hat{\mu}_n(m))' \mathcal{G}^{-1}(m) Y \right. \\
\left. \quad + \frac{1}{2} y' m_D^{-1} y \right]\]
in which \( \hat{\mu}_n(m) \) represents the conditional Bayes estimator, given \( J=j \), namely

\[
\hat{\mu}_n(m) = \frac{m' \mathbf{\Theta}^{-1}(m) Y}{m' \mathbf{\Theta}^{-1}(m)m}.
\]

The proof of (5.5) and (5.7) is similar to previous derivations. Since the a-priori distribution of \( \mu_n \) is normal, one obtains

\[
\mathcal{L}(Y|J=j) = \mathcal{N}(0, \mathbf{\Theta}^*(m))
\]

where

\[
\mathbf{\Theta}^*(m) = \mathbf{\Theta}(m) + \tau^2 \mathbf{I}_{mm'}.
\]

Similarly,

\[
\mathcal{L}(Y, \mu_n|J=j) = \mathcal{N}\left(0, \begin{bmatrix} \mathbf{\Theta}^*(m) & \tau^2 \mathbf{I}_m \\ \tau^2 \mathbf{I}_{m'}, & \tau^2 \end{bmatrix}\right).
\]

It follows that the a-posteriori distribution of \( \mu_n \), given \( (X, J) \) is

\[
\mathcal{L}(\mu_n|X, J=j) = \mathcal{N}\left(\frac{m' \mathbf{\Theta}^{-1}(m) Y}{m' \mathbf{\Theta}^{-1}(m)m+\tau^{-2}}, \frac{1}{m' \mathbf{\Theta}^{-1}(m)m+\tau^{-2}} \right).
\]

Let \( Y^* = (Y, Y^*_r+2, \ldots, Y^*_n)' \), where \( Y^*_r+2, Y^*_r+2, \ldots, Y^*_n \) orthonormal contrasts in \( X_1, X_2, \ldots, X_n \) and hence

\[
\sum_{i=1}^{r+1} \frac{Y_{i}^2}{n_i} + \sum_{i=r+2}^{n} Y_{i}^2 = X'X.
\]

Since these contrasts are independent of \( \mu_n \) and \( Z \) we have,
(5.13) \[ \mathcal{L}(\mu_n | Y, J) = \mathcal{L}(\mu_n | Y^*, J) = \mathcal{L}(\mu_n | X, J) \].

Hence, the conditional Bayes estimator, given \( J = j \), is
\[ m' \mathcal{G}^{-1}(m)Y / (m' \mathcal{G}^{-1}(m)m + \tau^{-2}) \]. This estimator converges, as \( \tau^2 \to \infty \), to (5.8). The conditional density of \( Y^* \), given \( J = j \), coincides with that of \( X \), given \( J = j \), except for a factor of \( (m_1 \cdots m_{r+1})^{1/2} \). Thus,

(5.14) \[ n(X|0, \sum^*(j)) = (m_1 \cdots m_{r+1})^{1/2} n(Y|0, \mathcal{G}^*(m)) \cdot \frac{1}{(2\pi)^{n-1-r}} \exp \left[ -\frac{1}{2} \sum_{i=r+2}^{n} y_i^2 \right] = \]

\[ = \frac{1}{(2\pi)^{n}} \exp \left\{ -\frac{1}{2} X'X \right\} \left( \frac{m_1 \cdots m_{r+1}}{\text{det. } \mathcal{G}^*(m)} \right)^{1/2} \exp \left\{ -\frac{1}{2} Y'(\mathcal{G}^{*-1}(m)-m_{r+1}^{-1})Y \right\} \]

Application of (2.15) and (5.14) yields (5.5)-(5.7) after substituting in (5.14)

(5.15) \[ \mathcal{G}^{*-1}(m) = \mathcal{G}^{-1}(m) - \mathcal{G}^{-1}(m)mm' \mathcal{G}^{-1}(m) \]

\[ m' \mathcal{G}^{-1}(m)m + \tau^{-2} \]

and

(5.16) \[ \text{det. } \mathcal{G}^*(m) = \text{det. } \mathcal{G}(m) [1 + \tau^2 m' \mathcal{G}^{-1}(m)m] \]

and letting \( \tau^2 \to \infty \).
6. The Bayes estimator for the case of at most one change.

Applying the formulae derived in the preceding section we obtain a relatively simple formula for the case of where the distribution of \( J \) is such that there is at most one change in the mean. The formula obtained sheds more light on the characteristics of the Bayes estimator. The Bayes estimator for at most one change will serve as a basis for an ad-hoc estimation procedure described in the next section, which can be applied in general.

Let \( p_0 = P(J=0) \) and \( p_m = P(J_m=1, J_{m'}=0 \text{ for all } m' \neq m) = P(J_m=1) \) \( m=1, \ldots, n-1 \). That is, \( p_0 \) is the a-priori probability of no change, and \( p_m (m=1, \ldots, n-1) \) is the a-priori probability of one change taking place between \( X_m \) and \( X_{m+1} \). In the case of one change taking place at time point \( m (m=1, \ldots, n-1) \) the covariance matrix \( \sigma(m) \) is given by:

\[
\sigma(m) = \begin{bmatrix}
m & 0 \\
0 & n-m
\end{bmatrix} + \sigma^2 \begin{bmatrix}
m^2 & 0 \\
0 & 0
\end{bmatrix} = \begin{bmatrix}
m(1+\sigma^2 m) & 0 \\
0 & n-m
\end{bmatrix}.
\]

Hence, the conditional Bayes estimator is, according to (5.8)

\[
\hat{\mu}_n(m) = \frac{n\bar{X}_n + \sigma^2 m (n-m) \bar{X}_{n-m}}{n + \sigma^2 m (n-m)}, \quad m=1, \ldots, n-1
\]

where

\[
\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{X}_{n-m} = \frac{1}{n-m} \sum_{i=m+1}^{n} X_i \quad (m=1, \ldots, n-1).
\]

Furthermore, as easily verified, formula (5.7) is reduced in the present case to

\[
\psi(y|J_m=1) = \frac{1}{(n+\sigma^2 m(n-m))^{1/2}} \exp \left[ \frac{1}{2} \frac{m^2(n-m)^2 \sigma^2}{n^2 + \sigma^2 m(n-m)} (\bar{X}_m - \bar{X}_{n-m})^2 \right]
\]
The conditional Bayes estimator and its corresponding weight for the case of no change can be obtained from (6.2) and (6.4) by substituting \( m=0 \). Thus, the formula of the Bayes estimator for at most one change is,

\[
\hat{\mu}_n^* = D^{*-1} \sum_{m=0}^{n-1} \frac{p_m}{\sqrt{n+\sigma^2 m(n-m)}} \exp\left[ \frac{1}{2} \cdot \frac{\sigma^2 m^2 (n-m)^2 (\bar{x}_m - \bar{x}_{n-m}^*)^2}{n^2 + \sigma^2 m(n-m)n} \right] \]

\[
\frac{n\bar{x}_n + \sigma^2 m(n-m)\bar{x}_{n-m}^*}{n+\sigma^2 m(n-m)}
\]

where

\[
D^* = \sum_{m=0}^{n-1} \frac{p_m}{\sqrt{n+\sigma^2 m(n-m)}} \exp\left[ \frac{1}{2} \cdot \frac{\sigma^2 m^2 (n-m)^2 (\bar{x}_m - \bar{x}_{n-m}^*)^2}{n^2 + \sigma^2 m(n-m)n} \right]
\]

A simplified form of the Bayes estimator for the case of at most one change is obtained by considering the asymptotic form of \( \hat{\mu}_n^* \) for \( \sigma^2 \rightarrow \infty \). The Bayes estimator is then approximated by

\[
\mu_n^{**} = (D^{**})^{-1} \left\{ \frac{p_o}{\sqrt{n}} \bar{x}_n + \frac{1}{\sigma} \sum_{m=1}^{n-1} \frac{p_m}{\sqrt{m(n-m)}} \exp\left[ \frac{1}{2} \cdot \frac{m(n-m)(\bar{x}_m - \bar{x}_{n-m}^*)^2}{n} \bar{x}_{n-m}^* \right] \right\}
\]

where

\[
D^{**} = \frac{p_o}{\sqrt{n}} + \frac{1}{\sigma} \sum_{m=1}^{n-1} \frac{p_m}{\sqrt{m(n-m)}} \exp\left[ \frac{1}{2} \cdot \frac{m(n-m)(\bar{x}_m - \bar{x}_{n-m}^*)^2}{n} \right]
\]
7. **An ad-hoc estimation procedure.**

The general form of the Bayes estimator, as given by (4.9) and (5.5) is very complex. Our objective is to construct an estimation procedure which is robust, efficient, and yet simple enough to be of practical use. The Bayes estimator based on the model of at most one change is relatively simple. However, computations show that, it is an inefficient estimator in cases where the expected number of changes in the mean is greater than one. As is to be expected, the "at most one change" estimator does not do well when there are several changes, for a large change may hide subsequent smaller ones. We present here an ad-hoc procedure which estimates the last time point of change by a sequence of tests. Then, the "at most one change" estimator is applied to the observations following the last estimated time point of change. As will be shown later the ad-hoc procedure improves upon the "at most one change" estimator in cases where the latter is inefficient.

To present the ad-hoc procedure in detail, define

\[
B_{n}(m,k) = D_{m}^{*l} \ p_{k}^{(m)} \ \psi_{k}(X^{(m)}) , \quad k=0,1,...,m-1
\]

where

\[
X^{(m)} = (X_{n-m+1} ,..., X_{n})',
\]

and

\[
p_{0}^{(m)} = (1-p)^{m-1}
\]

\[
p_{k}^{(m)} = p(1-p)^{m-2} \quad k=1,2,...,m-1,
\]
\[ \psi_k(x^{(m)}) = \frac{1}{[m+\sigma^2(k-m)]^{1/2}} \exp \left( \frac{1}{2} \frac{k^2(m-k)^2 \sigma^2}{m^2 + \sigma^2 mk(m-k)} \right) \left( \frac{x^{(m)} - x^{(m-k)}}{m-k} \right)^2 \]

for \( k=0,1,\ldots,n-1 \)

in which

\[ \overline{x}^{(m)}_k = \frac{1}{k} \sum_{i=n-m+1}^{n-m+k} x_i \]

and

\[ \overline{x}^{(m)}_{m-k} = \frac{1}{m-k} \sum_{i=n-m+k+1}^{n} x_i \]

and where

\[ p_k^{(m)} = \sum_{k=0}^{m-1} \psi_k(x^{(m)}) \]

For \( k > 0 \), \( B_n(m,k) \) is the a-posteriori probability, based on the data \( x^{(m)} \), that a change has taken place after the \( (n-k) \)th observation. For \( k=0 \), \( B_n(m,0) \) is the a-posteriori probability that no change has taken place. This computation assumes a-priori probabilities proportional to \( p_k^{(m)} \).

Let \( K_m \) be the value of \( k \) which maximizes \( B_n(m,k) \). \( K_m = 0 \) corresponds to a tentative Bayes decision that no change has taken place during the last \( m \) time points.

The ad-hoc procedure consists of computing \( K_2, K_3, \ldots \) until we reach a non-zero \( K_m \), say \( K \). Then we apply the "at most one change" estimator to the observations \( x_{n-K+1}, x_{n-K+2}, \ldots, x_n \).
The following example illustrates the procedure with $p=0.2$ and $\sigma=3$. Consider a sample of 9 observations: $X_1 = 2.6130$, $X_2 = 1.6610$, $X_3 = 1.8145$, $X_4 = 1.2737$, $X_5 = 2.6157$, $X_6 = -0.3256$, $X_7 = -2.4220$, $X_8 = -0.1186$, $X_9 = -0.0341$. The a-posteriori probabilities $B_9(m,k)$ and the Bayes estimators for at most one change, based on the last $m$ observations, are given in the following table:

This table gives a strong indication of a change taking place between $X_5$ and $X_6$. In fact, $K_7 = 4$ is the first non-zero $K$ and our ad-hoc procedure, computing the "at most one change estimate" based on the last four observations, yields the estimate $-0.6301$. 
<table>
<thead>
<tr>
<th>m</th>
<th>B(m,m-8)</th>
<th>B(m,m-7)</th>
<th>B(m,m-6)</th>
<th>B(m,m-5)</th>
<th>B(m,m-4)</th>
<th>B(m,m-3)</th>
<th>B(m,m-2)</th>
<th>B(m,m-1)</th>
<th>B(m,0)</th>
<th>Estim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0702</td>
<td>0.9298</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2474</td>
<td>0.0722</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0542</td>
<td>0.0954</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7149</td>
<td>0.0890</td>
<td>0.0088</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01753</td>
<td>0.6812</td>
<td>0.1378</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0103</td>
<td>0.0158</td>
<td>0.7682</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0037</td>
<td>0.0089</td>
<td>0.0154</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0025</td>
<td>0.0032</td>
<td>0.0086</td>
</tr>
</tbody>
</table>
8. **Comparison of estimators by Monte Carlo computations.**

Three methods of estimating the current mean, $\mu_n$, have been considered. These are the minimum variance linear unbiased procedure; the Bayes procedure; and an ad-hoc procedure. Accordingly we derived four relatively simple estimators. The M.V.L.U. estimator (3.7); the "at most one-change Bayes" (A.M.O.C. Bayes) estimator (6.5); the simplified A.M.O.C. Bayes estimator (6.7); and the ad-hoc estimator in the present section. The efficiency of these estimators and their robustness in small samples are studied numerically. For this purpose we used a computer (IBM 7090) to generate samples of $n=9$ observations, according to the following models:

**Model I:** A change between each time point. The parameter $\mu_9$ is assigned the value 0, and the observations $X_1, \ldots, X_9$ are generated according to the model:

$$X_i = \begin{cases} 
\sigma \sum_{k=1}^{8} \eta_k + \epsilon_i, & \text{for } i=1, \ldots, 8 \\
\epsilon_9, & \text{for } i=9 
\end{cases}$$

(8.1)

where $\epsilon_1, \ldots, \epsilon_9; \eta_1, \ldots, \eta_8$ are independent random variables, each having a $\mathcal{N}(0, \sigma^2)$ distribution law. In order to study the effect of the magnitude of changes in the mean on the estimators, the parameter $\sigma$ was assigned the values 2, 3, 4.

**Model II:** Binomial changes, $0 < p < 1$. In this model we assign a constant probability, $0 < p < 1$, of a change between any two consecutive observations.
The parameter \( \mu_9 \) is assigned the value \( 0 \), and the observations \( X_1, \ldots, X_9 \) are generated according to the model:

\[
X_i = \begin{cases} 
\sigma \sum_{k=1}^{8} J_k \eta_k^+ \epsilon_i, & \text{for } i=1, \ldots, 8 \\
\epsilon_9, & \text{for } i=9 
\end{cases}
\]  

(8.2)

where \( \epsilon_1, \ldots, \epsilon_9; \eta_1, \ldots, \eta_8 \) are independent random variables, each having a \( \mathcal{N}(0,1) \) distribution law; and \( J_1, \ldots, J_8 \) are independent binomial random variables, with

\[
P(J_k = 1) = p, \quad \text{for all } k=1, \ldots, 8.
\]

(8.3)

Nine cases were considered, corresponding to the combinations of \( \sigma = 2, 3, 4 \) and \( p = 0.1, 0.2, 0.3 \).

Model III: Assigned changes.

(i) III. A. No change.

All the means \( \mu_1 = \cdots = \mu_9 = 0 \) and,

\[
X_i = \epsilon_i \quad (i=1, \ldots, 9)
\]

(8.4)

where \( \epsilon_1, \ldots, \epsilon_9 \) are \( \mathcal{N}(0,1) \) independent random variables.

(ii) III. B. One assigned change.

The means \( \mu_1, \ldots, \mu_9 \) are given, for each \( m=1, \ldots, 8 \), by:

\[
\begin{cases} 
\mu_1 = \cdots = \mu_m = \sigma \\
\mu_{m+1} = \cdots = \mu_9 = 0 
\end{cases}
\]

(8.5)

and we consider each case \( m=1, \ldots, 8 \) with \( \sigma = 2, 3, 4 \).
(iii) III. C. Two assigned changes.

Case 1 - The means of the observations are given by

\[(8.6) \quad \mu_1 = \cdots = \mu_6 = 0 ; \quad \mu_7 = \sigma ; \quad \mu_8 = \mu_9 = 0\]

where \( \sigma = 2, 3, 4 \). This model is used to check the effect of two adjacent changes that cancel each other.

Case 2 - Two changes in the same direction. Here the means of \( X_1, \ldots, X_9 \) are:

\[(8.7) \quad \mu_1 = \mu_2 = 2\sigma ; \quad \mu_3 = \cdots = \mu_7 = \sigma ; \quad \mu_8 = \mu_9 = 0\]

where \( \sigma = 2, 3, 4 \).

100 samples of \( n=9 \) observations were generated for each model, and each of the following estimators was applied to each sample. In order to check the effect of substituting incorrect values of the parameters \( p \) and \( \sigma^2 \) in the formulae of the estimators, we considered the following seven estimators.

- Estimator 1 - M.V.L.U., \( \sigma^2 = 3, \quad p = 0.2 \)
- Estimator 2 - M.V.L.U., \( \sigma^2 = 20, \quad p = 0.2 \)
- Estimator 3 - A.M.O.C. Bayes, \( \sigma^2 = 3, \quad p = 0.2 \)
- Estimator 4 - A.M.O.C. Bayes, \( \sigma^2 = 20, \quad p = 0.2 \)
- Estimator 5 - A.M.O.C. simplified Bayes, \( \sigma^2 = 20, \quad p = 0.2 \)
- Estimator 6 - Ad-hoc procedure, \( \sigma^2 = 3, \quad p = 0.2 \)
- Estimator 7 - Ad-hoc procedure, \( \sigma^2 = 20, \quad p = 0.2 \)

The means and mean-square-errors (M.S.E.) of these seven estimators, over the 100 samples generated for most of the models, are represented in the following table.
Table 2: The means (underlined) and M.S.E.'s of the estimators over 100 samples

<table>
<thead>
<tr>
<th>Model</th>
<th>Est. 1</th>
<th>Est. 2</th>
<th>Est. 3</th>
<th>Est. 4</th>
<th>Est. 5</th>
<th>Est. 6</th>
<th>Est. 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>I, $\sigma=2$</td>
<td>-0.2718</td>
<td>-0.0595</td>
<td>-0.1866</td>
<td>-0.0726</td>
<td>-0.0425</td>
<td>-0.0827</td>
<td>-0.0818</td>
</tr>
<tr>
<td></td>
<td>2.1406</td>
<td>0.7333</td>
<td>3.3140</td>
<td>2.3794</td>
<td>2.2672</td>
<td>1.0235</td>
<td>0.9912</td>
</tr>
<tr>
<td>I, $\sigma=3$</td>
<td>0.2473</td>
<td>0.1251</td>
<td>0.2734</td>
<td>0.1366</td>
<td>0.2486</td>
<td>0.1379</td>
<td>0.1358</td>
</tr>
<tr>
<td></td>
<td>3.2079</td>
<td>0.9661</td>
<td>5.7730</td>
<td>3.8901</td>
<td>4.7529</td>
<td>1.1168</td>
<td>1.1202</td>
</tr>
<tr>
<td>I, $\sigma=4$</td>
<td>-0.0891</td>
<td>-0.0111</td>
<td>-0.2734</td>
<td>-0.2754</td>
<td>-0.3818</td>
<td>+0.0151</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>5.5779</td>
<td>1.4063</td>
<td>9.1200</td>
<td>6.1292</td>
<td>11.9777</td>
<td>1.2406</td>
<td>1.2441</td>
</tr>
<tr>
<td>II, $p=0.1$</td>
<td>0.0847</td>
<td>0.0665</td>
<td>0.0539</td>
<td>0.0469</td>
<td>0.0487</td>
<td>0.0525</td>
<td>0.0263</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>0.4460</td>
<td>0.6200</td>
<td>0.4337</td>
<td>0.4312</td>
<td>0.4408</td>
<td>0.4153</td>
<td>0.3965</td>
</tr>
<tr>
<td>II, $p=0.1$</td>
<td>-0.0649</td>
<td>-0.0458</td>
<td>-0.0917</td>
<td>-0.0988</td>
<td>-0.0989</td>
<td>-0.1139</td>
<td>-0.1351</td>
</tr>
<tr>
<td>$\sigma=3$</td>
<td>0.6394</td>
<td>0.7172</td>
<td>0.6022</td>
<td>0.5712</td>
<td>0.5809</td>
<td>0.5891</td>
<td>0.5305</td>
</tr>
<tr>
<td>II, $p=0.1$</td>
<td>-0.0528</td>
<td>-0.0853</td>
<td>0.0462</td>
<td>0.0328</td>
<td>0.0278</td>
<td>-0.0016</td>
<td>-0.0173</td>
</tr>
<tr>
<td>$\sigma=4$</td>
<td>0.7702</td>
<td>0.7489</td>
<td>0.7385</td>
<td>0.6283</td>
<td>0.6322</td>
<td>0.4210</td>
<td>0.4179</td>
</tr>
<tr>
<td>II, $p=0.2$</td>
<td>-0.0258</td>
<td>-0.0473</td>
<td>-0.0502</td>
<td>-0.0750</td>
<td>-0.0806</td>
<td>-0.0836</td>
<td>-0.0865</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>0.5462</td>
<td>0.7792</td>
<td>0.5262</td>
<td>0.5216</td>
<td>0.5331</td>
<td>0.6733</td>
<td>0.6470</td>
</tr>
<tr>
<td>II, $p=0.2$</td>
<td>0.0667</td>
<td>0.0380</td>
<td>0.0487</td>
<td>0.0040</td>
<td>-0.0060</td>
<td>0.0397</td>
<td>0.0495</td>
</tr>
<tr>
<td>$\sigma=3$</td>
<td>0.8594</td>
<td>0.7777</td>
<td>1.1348</td>
<td>0.8768</td>
<td>0.8632</td>
<td>0.5425</td>
<td>0.5032</td>
</tr>
<tr>
<td>II, $p=0.2$</td>
<td>0.0182</td>
<td>0.0386</td>
<td>-0.0426</td>
<td>-0.0013</td>
<td>-0.0156</td>
<td>0.0629</td>
<td>0.0669</td>
</tr>
<tr>
<td>$\sigma=4$</td>
<td>1.4807</td>
<td>1.0593</td>
<td>2.1400</td>
<td>1.5033</td>
<td>1.5460</td>
<td>0.6758</td>
<td>0.6892</td>
</tr>
<tr>
<td>II, $p=0.3$</td>
<td>-0.1233</td>
<td>-0.0980</td>
<td>-0.1418</td>
<td>-0.1386</td>
<td>-0.1369</td>
<td>-0.0380</td>
<td>-0.0436</td>
</tr>
<tr>
<td>$\sigma=2$</td>
<td>0.7718</td>
<td>0.7605</td>
<td>0.7945</td>
<td>0.7458</td>
<td>0.7522</td>
<td>0.6405</td>
<td>0.6326</td>
</tr>
<tr>
<td>II, $p=0.3$</td>
<td>-0.0429</td>
<td>0.0377</td>
<td>-0.1507</td>
<td>-0.1489</td>
<td>-0.0992</td>
<td>0.0154</td>
<td>-0.0094</td>
</tr>
<tr>
<td>$\sigma=3$</td>
<td>1.4082</td>
<td>0.9192</td>
<td>2.0073</td>
<td>1.5540</td>
<td>1.6971</td>
<td>0.7944</td>
<td>0.7989</td>
</tr>
<tr>
<td>II, $p=0.3$</td>
<td>0.2000</td>
<td>0.0607</td>
<td>0.3241</td>
<td>0.2491</td>
<td>0.3326</td>
<td>-0.0286</td>
<td>-0.0320</td>
</tr>
<tr>
<td>$\sigma=4$</td>
<td>2.3742</td>
<td>1.0009</td>
<td>3.5142</td>
<td>2.7221</td>
<td>3.2602</td>
<td>0.7747</td>
<td>0.8088</td>
</tr>
<tr>
<td>Model</td>
<td>Est. 1</td>
<td>Est. 2</td>
<td>Est. 3</td>
<td>Est. 4</td>
<td>Est. 5</td>
<td>Est. 6</td>
<td>Est. 7</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>III A</td>
<td>0.0255</td>
<td>0.0325</td>
<td>0.0027</td>
<td>-0.0016</td>
<td>-0.0007</td>
<td>-0.0122</td>
<td>-0.0070</td>
</tr>
<tr>
<td></td>
<td>0.3078</td>
<td>0.6112</td>
<td>0.1654</td>
<td>0.1746</td>
<td>0.1846</td>
<td>0.2679</td>
<td>0.2789</td>
</tr>
<tr>
<td>III B,1</td>
<td>-0.0194</td>
<td>-0.0218</td>
<td>0.0194</td>
<td>-0.0018</td>
<td>-0.0156</td>
<td>-0.0003</td>
<td>-0.0063</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.4069</td>
<td>0.8476</td>
<td>0.1728</td>
<td>0.1920</td>
<td>0.2039</td>
<td>0.5246</td>
<td>0.4974</td>
</tr>
<tr>
<td>III B,1</td>
<td>-0.0194</td>
<td>-0.0218</td>
<td>0.0658</td>
<td>-0.0158</td>
<td>-0.0396</td>
<td>-0.0647</td>
<td>-0.0604</td>
</tr>
<tr>
<td>$\sigma^4$</td>
<td>0.3142</td>
<td>0.5868</td>
<td>0.1289</td>
<td>0.1246</td>
<td>0.1256</td>
<td>0.2683</td>
<td>0.2458</td>
</tr>
<tr>
<td>III B,5</td>
<td>0.1059</td>
<td>0.0151</td>
<td>0.2658</td>
<td>0.1688</td>
<td>0.1332</td>
<td>0.0320</td>
<td>0.0464</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3169</td>
<td>0.5948</td>
<td>0.3397</td>
<td>0.3769</td>
<td>0.3775</td>
<td>0.4146</td>
<td>0.4265</td>
</tr>
<tr>
<td>III B,5</td>
<td>0.2047</td>
<td>-0.0080</td>
<td>0.3206</td>
<td>0.0722</td>
<td>0.0213</td>
<td>-0.0045</td>
<td>0.0128</td>
</tr>
<tr>
<td>$\sigma^4$</td>
<td>0.4394</td>
<td>0.7475</td>
<td>0.3387</td>
<td>0.2769</td>
<td>0.2793</td>
<td>0.4598</td>
<td>0.4330</td>
</tr>
<tr>
<td>III B,7</td>
<td>0.4415</td>
<td>0.0061</td>
<td>0.7190</td>
<td>0.5446</td>
<td>0.4751</td>
<td>0.1541</td>
<td>0.2048</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.5897</td>
<td>0.7647</td>
<td>1.0173</td>
<td>1.0892</td>
<td>1.0995</td>
<td>1.0570</td>
<td>1.1401</td>
</tr>
<tr>
<td>III B,7</td>
<td>0.9772</td>
<td>0.2558</td>
<td>0.7161</td>
<td>0.2405</td>
<td>0.1316</td>
<td>0.0493</td>
<td>0.0451</td>
</tr>
<tr>
<td>$\sigma^4$</td>
<td>1.2983</td>
<td>0.6901</td>
<td>1.0460</td>
<td>0.7467</td>
<td>0.7392</td>
<td>0.6589</td>
<td>0.6399</td>
</tr>
<tr>
<td>III B,8</td>
<td>0.9522</td>
<td>0.3266</td>
<td>1.1722</td>
<td>0.9864</td>
<td>0.9036</td>
<td>0.5665</td>
<td>0.5636</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.2950</td>
<td>0.8677</td>
<td>1.9536</td>
<td>1.9667</td>
<td>0.9452</td>
<td>2.0107</td>
<td>2.0013</td>
</tr>
<tr>
<td>III B,8</td>
<td>1.9223</td>
<td>0.7190</td>
<td>1.4325</td>
<td>0.6353</td>
<td>0.4205</td>
<td>0.1917</td>
<td>0.1775</td>
</tr>
<tr>
<td>$\sigma^4$</td>
<td>4.0753</td>
<td>1.3022</td>
<td>3.2118</td>
<td>2.3153</td>
<td>2.2392</td>
<td>1.6658</td>
<td>1.6777</td>
</tr>
<tr>
<td>III C,1</td>
<td>0.1903</td>
<td>0.0061</td>
<td>0.2633</td>
<td>0.2455</td>
<td>0.2423</td>
<td>0.1189</td>
<td>0.0955</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.3352</td>
<td>0.6211</td>
<td>0.2910</td>
<td>0.3402</td>
<td>0.3651</td>
<td>0.5251</td>
<td>0.4931</td>
</tr>
<tr>
<td>III C,1</td>
<td>0.4145</td>
<td>0.0242</td>
<td>0.6616</td>
<td>0.6606</td>
<td>0.6605</td>
<td>-0.0018</td>
<td>-0.0167</td>
</tr>
<tr>
<td>$\sigma^4$</td>
<td>0.5458</td>
<td>0.6719</td>
<td>0.8723</td>
<td>0.9941</td>
<td>1.0496</td>
<td>0.7509</td>
<td>0.7127</td>
</tr>
<tr>
<td>III C,2</td>
<td>0.4964</td>
<td>0.1553</td>
<td>0.9684</td>
<td>0.7926</td>
<td>0.4772</td>
<td>0.2862</td>
<td>0.3129</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.6005</td>
<td>0.7215</td>
<td>1.3203</td>
<td>1.1810</td>
<td>1.1534</td>
<td>0.9575</td>
<td>0.9985</td>
</tr>
<tr>
<td>III C,2</td>
<td>0.8920</td>
<td>0.0860</td>
<td>1.8091</td>
<td>1.4087</td>
<td>1.3213</td>
<td>-0.0229</td>
<td>-0.0196</td>
</tr>
<tr>
<td>$\sigma^4$</td>
<td>1.1480</td>
<td>0.6736</td>
<td>4.6786</td>
<td>4.0504</td>
<td>3.9797</td>
<td>0.6849</td>
<td>0.7083</td>
</tr>
</tbody>
</table>
Examination of the numerical results, which are represented in Table 2, leads to the following conclusions:

1) As anticipated, when there is no change in the means (Model III A) the "at most one change" (A.M.O.C.) Bayes estimator, with the smallest value of $\sigma^2$ (Est. 3) is the most efficient one (has the smallest M.S.E.) among the estimators studied. Moreover, the differences between the M.S.E.'s of estimators 3, 4, and 5 are negligible. This fact indicates that the A.M.O.C. Bayes estimator is not sensitive to incorrect values of the parameters. The ad-hoc estimators (Est. 6 and Est. 7) are slightly less efficient than the A.M.O.C. Bayes estimators. Less efficient than all, when there is no change, are the M.V.L.U. estimators.

2) When changes in the mean always occur, (Model I), we expect the A.M.O.C. Bayes estimators to be poor ones, and the M.V.L.U. estimators to be better. This is verified by the numerical results which also show that these estimators are sensitive to the specification of $\sigma$.

3) When the expected number of changes in the mean is about one (Model II, $p = 0.1$) the estimators based on the "at most one change" Bayes procedure are better than the M.V.L.U. estimators. The ad-hoc estimators performed most efficiently, and turned out to be robust against variations in the parameter $\sigma^2$ of the generated samples. The other estimators loose efficiency when $\sigma^2$ becomes large.

4) When the expected number of changes in the mean is greater than one (Models II, $p = 0.2$, $p = 0.3$) and $\sigma^2$ of the generated samples is large, only the ad-hoc estimator performed efficiently. If $\sigma^2$ of the
generated samples is not large (Model II, p = 0.2, \( \sigma = 2 \)), the A.M.O.C. Bayes (Est. 3, 4, 5) may perform as good as the ad-hoc procedure.

5) The results of applying the estimators on Model III B of one assigned change show that when the change takes place very close to the last observation (Models III B7, III B8) the A.M.O.C. Bayes estimator and the ad-hoc estimator loose their efficiency. These estimators are very efficient if the change takes place at the very beginning of the sequence. This is not the case with the M.V.L.U. Estimator 2 whose efficiency does not depend heavily on the place where the change has occurred. Indeed, Estimator 2 gives most weight to the last observation and attaches very little weight to previous observations. Therefore the actual place of change does not affect it significantly.

6) The experiment with Model III C1 shows that if two changes of equal magnitude but different directions occur successively and the parameter \( \sigma^2 \) of the generated samples is not large (\( \sigma^2 = 4 \)) the A.M.O.C. Bayes estimators are the best ones. The M.V.L.U. estimator with a small assigned value of \( \sigma^2_p \) is equally good. The ad-hoc estimators are slightly less efficient, and the M.V.L.U. estimator with a large assigned value of \( \sigma^2_p \) (Est. 2) is less efficient than all the estimators examined. The picture changes when the value of the parameter \( \sigma^2 \) is large. In this case the M.V.L.U. estimators are the most efficient. The ad-hoc estimators are less efficient and the A.M.O.C. Bayes estimators are least efficient. Model III C2 shows that when the two changes are in the same direction, and the time point of the second change is close to that of the last observation the only efficient estimators for small \( \sigma^2 \) are the M.V.L.U. ones. When \( \sigma^2 \) is large the only efficient estimators are the ad-hoc ones.

The Bayesian approach applied in the present study can be particularly useful for deriving test procedures in problems of testing whether a change in a location parameter of a distribution has taken place at an unknown time point. Problems of detecting a change in the location parameter occur in many different fields. Sampling inspection of the quality of products from a continuous production process is one example of a possible application.

The testing problem which will be considered here is the following, formulated by Page [4] who proposed an alternative test statistic. Given a finite sequence of normally distributed independent random variables \( X_1, \ldots, X_n \), having expected values \( \mu_1, \ldots, \mu_n \) and variance 1, we wish to test the hypothesis,

\[
\mathcal{H}: \mu_1 = \cdots = \mu_n
\]

against the alternatives

\[
\mathcal{A}: \mu_1 = \cdots = \mu_m ; \mu_{m+1} = \cdots = \mu_n ; 1 \leq m \leq n-1
\]

\[
\mu_{m+1} - \mu_m = \delta > 0
\]

where \( m \) is unknown.

We consider the two cases, \( \mu_1 \) known and \( \mu_1 \) unknown. (1) The method of deriving a test procedure consists of selecting a-priori distributions for the nuisance parameters, and characterizing the corresponding Bayes solution. The operating characteristics of the resulting test procedure are then studied.
Let $M$ be a random variable designating the point of change in $M=1,2,\ldots,n-1$, with p.d.f.

\[
p_{m} = P(M=m) = \begin{cases} \frac{1}{n-1}, & \text{if } m=1,2,\ldots,n-1 \\ 0, & \text{otherwise} \end{cases}
\]

and let $\Delta$ be a random variable representing the magnitude of change in the mean,

\[
\Delta = \mu_{m+1} - \mu_{m}, \quad 0 \leq \Delta < \infty.
\]

We assign $\Delta$ a semi-normal a-priori distribution, whose density is:

\[
h_{\Delta}(\delta) = \begin{cases} 0, & \text{if } \delta \leq 0 \\ \frac{1}{\sigma} \sqrt{2\pi} \exp \left[ -\frac{1}{2\sigma^2} \delta^2 \right], & \text{if } \delta > 0 \end{cases}
\]

In the case $\mu_1=0$, we arrive at the following likelihood-ratio

\[
L(X_1,\ldots,X_n) = \frac{\frac{1}{n-1} \sum_{m=1}^{n-1} f_1(X_1,\ldots,X_n|M=m)}{f_0(X_1,\ldots,X_n)}
\]

\[
\quad = \frac{2}{n-1} \sum_{m=1}^{n-1} \exp \left[ \frac{1}{2}(n-m+1) - S_{n-m}^2 \right] \phi \left( \frac{S_{n-m}^*}{\sqrt{n-m+\frac{1}{\sigma^2}}} \right)
\]

where $S_{n-m}^* = \sum_{i=m+1}^{n} X_i$; and $\phi(n)$ is the $\mathcal{N}(0,1)$ c.d.f.

The likelihood-ratio (9.6) can be written, for small values of $\sigma$, as

\[
L(X_1,\ldots,X_n) = \frac{1}{2} + \frac{\sigma}{\sqrt{2\pi}} \sum_{m=1}^{n-1} S_{n-m}^* \phi(\sigma), \text{ as } \sigma \to 0.
\]
A Bayes procedure is to reject $H_0$ whenever $L(X_1, \ldots, X_n)$ is greater than an appropriate constant. Accordingly we derive from (9.7) the following test statistic:

\begin{equation}
T(X_1, \ldots, X_n) = \sum_{m=1}^{n-1} S_{n-m}^* = \sum_{i=1}^{n} (i-1)X_i.
\end{equation}

A test statistic for the case $\mu_1$ is unknown is obtained in a similar fashion. We consider $\mu_1$ a random variable having a $\mathcal{N}(0, \tau^2)$ a-priori distribution, and then we let $\tau^2 \to \infty$. Computations similar to those of section 6 yield the test statistic:

\begin{equation}
T^*(X_1, \ldots, X_n) = \sum_{i=1}^{n} (i-1)(X_i - \bar{X}_n)
\end{equation}

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

A size-$\alpha$ test for the case $\mu_1 = 0$ is easily obtained from (9.8), since under $H$

\begin{equation}
\mathcal{L}(T(X_1, \ldots, X_n) | H) = \mathcal{N}(0, \frac{1}{6} \frac{n(n-1)(2n-1)}{n-1}).
\end{equation}

Thus, a size-$\alpha$ test criterion is

\begin{equation}
C_\alpha = u_{1-\alpha} \sqrt{\frac{n(n-1)(2n-1)}{6}}, \quad 0 < \alpha < 1
\end{equation}

where $u_{1-\alpha}$ is the $(1-\alpha)$-th fractile of $\mathcal{N}(0,1)$.

The corresponding power function, when $(M=m)$ and $(\Delta=\delta)$ is

\begin{equation}
\beta_m(\delta) = 1 - \Phi\left( u_{1-\alpha} - \sqrt{\frac{2n(n-1)}{4n-2}} \left[ 1 - \frac{m(m-1)}{n(n-1)} \right] \right)
\end{equation}

for $0 \leq \delta < \infty$. 

34
A size-\(\alpha\) test criterion for the case \(\mu_1\) is unknown can be derived from the fact that under \(H\) we have,

\[
(9.13) \quad \mathcal{A}(T^*(X_1, \ldots, X_n)|H) = \mathcal{G}(0, \frac{1}{12n(n-1)(n+1)}).
\]

Thus, one obtains the test criterion:

\[
(9.14) \quad c^*_\alpha = u_{1-\alpha} \sqrt{\frac{1}{12n(n^2-1)}} , \quad 0 < \alpha < 1
\]

and the power function:

\[
(9.15) \quad \beta^*_m(\delta) = 1 - \Phi \left( u_{1-\alpha-\delta} \sqrt{\frac{3m(n-m)}{n(n^2-1)}} \right)
\]

for \(0 \leq \delta < \infty\).

Expression (9.12) shows that the power-function of the test statistic \(T\) (9.8) is monotonically decreasing with \(m\). On the other hand the power function of \(T^*\) (9.9) attains its maximum, for a fixed \(\delta\), when \(m \sim \frac{n}{2}\). Page [4] formulates the testing problem considered here in more general terms, as follows: Let \(X_1, \ldots, X_n\) be independent random variables. It is required to test the null hypothesis that all the \(n\) random variables are identically distributed, with a c.d.f. \(F(x|\theta)\), against the alternative that \(X_1, \ldots, X_m\) (\(1 \leq m \leq n-1\)) have a distribution function \(F(x|\theta)\) and \(X_{m+1}, \ldots, X_n\) have a distribution function \(F(x|\theta')\) where \(\theta \neq \theta'\); \(m\) and \(\theta'\) are unknown but \(\theta\) is known. Page proposes the following test procedure for one-sided alternatives:

Record the cumulative sums

\[
(9.16) \quad S_r = \sum_{i=1}^{r} (x_i - \theta) \quad \text{for } i=1, \ldots, n
\]
and reject the null hypothesis when

\[
(9.17) \quad m = \max_{0 \leq r \leq n} \left\{ S_r - \min_{0 \leq i \leq r-1} S_i \right\}, \quad S_0 = 0
\]

is greater than a test criterion \( h \). In his paper (1955), Page evaluates the procedure for a class of distributions symmetric about \( \theta \). He considers the random variables \( Y_i = \text{sgn}(X_i - \theta) \) (\( i=1, \ldots, n \)) and studies the operating characteristics of his test procedure for the binomial random variables \( Y_1, \ldots, Y_n \), where, under the null hypothesis \( P(Y_i = 1) = \frac{1}{2} \) for all \( i=1, \ldots, n \), and under the alternatives \( P(Y_i = 1) = \frac{1}{2} \) for \( i=1, \ldots, m \) and \( P(Y_j = 1) = p, \quad p > \frac{1}{2} \), for \( j=m+1, \ldots, n \).

The test statistic (9.8) can also be applied to Page's problem with the binomially distributed random variables \( Y_1, \ldots, Y_n \). In the following table, we compare the power function of Page's test procedure (9.17) with the power function of the present Bayesian test procedure (9.3) for a sample of size \( n=20 \).

To achieve the significance level \( \alpha = 0.05 \) it was necessary to use randomized procedures. For the Page test we use \( h=9 \) with probability 0.4 and \( h=10 \) with probability 0.6. For the statistic (9.8) we use for the limits of \( T, 87 \) with probability \( \gamma = 0.71 \) and 88 with probability 0.29.
Table 3: The power of the Page test procedure and of the Bayesian test procedure for a sample of size \( n = 20, \alpha = 0.05 \).

<table>
<thead>
<tr>
<th>Proc.</th>
<th>( p )</th>
<th>( m )</th>
<th>0.5</th>
<th>0.55</th>
<th>0.60</th>
<th>0.65</th>
<th>0.70</th>
<th>0.75</th>
<th>0.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page</td>
<td>0</td>
<td>0</td>
<td>0.050</td>
<td>0.109</td>
<td>0.207</td>
<td>0.350</td>
<td>0.527</td>
<td>0.709</td>
<td>0.861</td>
</tr>
<tr>
<td>Bayes</td>
<td>0</td>
<td>0</td>
<td>0.050</td>
<td>0.105</td>
<td>0.194</td>
<td>0.322</td>
<td>0.482</td>
<td>0.655</td>
<td>0.811</td>
</tr>
<tr>
<td>Page</td>
<td>2</td>
<td>2</td>
<td>0.050</td>
<td>0.105</td>
<td>0.195</td>
<td>0.325</td>
<td>0.488</td>
<td>0.664</td>
<td>0.821</td>
</tr>
<tr>
<td>Bayes</td>
<td>2</td>
<td>2</td>
<td>0.050</td>
<td>0.104</td>
<td>0.191</td>
<td>0.316</td>
<td>0.473</td>
<td>0.644</td>
<td>0.801</td>
</tr>
<tr>
<td>Page</td>
<td>4</td>
<td>4</td>
<td>0.050</td>
<td>0.098</td>
<td>0.176</td>
<td>0.287</td>
<td>0.431</td>
<td>0.594</td>
<td>0.756</td>
</tr>
<tr>
<td>Bayes</td>
<td>4</td>
<td>4</td>
<td>0.050</td>
<td>0.101</td>
<td>0.184</td>
<td>0.301</td>
<td>0.450</td>
<td>0.616</td>
<td>0.775</td>
</tr>
<tr>
<td>Page</td>
<td>6</td>
<td>6</td>
<td>0.050</td>
<td>0.091</td>
<td>0.154</td>
<td>0.245</td>
<td>0.364</td>
<td>0.508</td>
<td>0.664</td>
</tr>
<tr>
<td>Bayes</td>
<td>6</td>
<td>6</td>
<td>0.050</td>
<td>0.098</td>
<td>0.172</td>
<td>0.279</td>
<td>0.416</td>
<td>0.572</td>
<td>0.730</td>
</tr>
<tr>
<td>Page</td>
<td>8</td>
<td>8</td>
<td>0.050</td>
<td>0.083</td>
<td>0.132</td>
<td>0.201</td>
<td>0.292</td>
<td>0.408</td>
<td>0.545</td>
</tr>
<tr>
<td>Bayes</td>
<td>8</td>
<td>8</td>
<td>0.050</td>
<td>0.093</td>
<td>0.158</td>
<td>0.250</td>
<td>0.369</td>
<td>0.510</td>
<td>0.661</td>
</tr>
<tr>
<td>Page</td>
<td>10</td>
<td>10</td>
<td>0.050</td>
<td>0.075</td>
<td>0.111</td>
<td>0.160</td>
<td>0.223</td>
<td>0.305</td>
<td>0.407</td>
</tr>
<tr>
<td>Bayes</td>
<td>10</td>
<td>10</td>
<td>0.050</td>
<td>0.087</td>
<td>0.141</td>
<td>0.215</td>
<td>0.312</td>
<td>0.430</td>
<td>0.565</td>
</tr>
<tr>
<td>Page</td>
<td>12</td>
<td>12</td>
<td>0.050</td>
<td>0.068</td>
<td>0.092</td>
<td>0.123</td>
<td>0.162</td>
<td>0.211</td>
<td>0.272</td>
</tr>
<tr>
<td>Bayes</td>
<td>12</td>
<td>12</td>
<td>0.050</td>
<td>0.080</td>
<td>0.122</td>
<td>0.177</td>
<td>0.249</td>
<td>0.337</td>
<td>0.442</td>
</tr>
<tr>
<td>Page</td>
<td>14</td>
<td>14</td>
<td>0.050</td>
<td>0.062</td>
<td>0.076</td>
<td>0.094</td>
<td>0.116</td>
<td>0.141</td>
<td>0.171</td>
</tr>
<tr>
<td>Bayes</td>
<td>14</td>
<td>14</td>
<td>0.050</td>
<td>0.073</td>
<td>0.102</td>
<td>0.139</td>
<td>0.185</td>
<td>0.241</td>
<td>0.308</td>
</tr>
<tr>
<td>Page</td>
<td>16</td>
<td>16</td>
<td>0.050</td>
<td>0.056</td>
<td>0.064</td>
<td>0.072</td>
<td>0.082</td>
<td>0.093</td>
<td>0.105</td>
</tr>
<tr>
<td>Bayes</td>
<td>16</td>
<td>16</td>
<td>0.050</td>
<td>0.065</td>
<td>0.083</td>
<td>0.104</td>
<td>0.128</td>
<td>0.156</td>
<td>0.180</td>
</tr>
<tr>
<td>Page</td>
<td>18</td>
<td>18</td>
<td>0.050</td>
<td>0.052</td>
<td>0.055</td>
<td>0.057</td>
<td>0.060</td>
<td>0.063</td>
<td>0.066</td>
</tr>
<tr>
<td>Bayes</td>
<td>18</td>
<td>18</td>
<td>0.050</td>
<td>0.057</td>
<td>0.065</td>
<td>0.074</td>
<td>0.082</td>
<td>0.092</td>
<td>0.101</td>
</tr>
<tr>
<td>Exact</td>
<td>10</td>
<td>10</td>
<td>0.050</td>
<td>0.099</td>
<td>0.167</td>
<td>0.261</td>
<td>0.382</td>
<td>0.524</td>
<td>0.679</td>
</tr>
<tr>
<td>Binomial</td>
<td>10</td>
<td>10</td>
<td>0.050</td>
<td>0.099</td>
<td>0.167</td>
<td>0.261</td>
<td>0.382</td>
<td>0.524</td>
<td>0.679</td>
</tr>
</tbody>
</table>

37
Table 3 shows that using the test statistic (9.8) gives slightly more power than using Page's test statistic (9.17), unless the change occurs near the very beginning (m small). The test statistic (9.8) seems to be better adapted to the problem Page formulated. On the other hand the Page procedure seems to be well devised to handle the variation of this problem where the initial value of the mean is known only to be below a certain level and it is desired to detect if it has changed to a value above that level.
10. **Discussion.**

As shown in sections 4 and 5, the method of deriving a translation invariant Bayes estimator leads to a fairly complicated expression. However, a considerable simplification is attained when we are ready to assume that at most one change in the means may occur. The results of the Monte Carlo experiments show that the Bayes estimator derived under the assumption of "at most one change" (A.M.O.C.) are very efficient even if more than one change may occur, provided that the expected number of changes is not greater than one. However, the A.M.O.C. Bayes estimators are not efficient given that there are two or more changes unless the last change is considerably larger than the previous ones. The ad-hoc estimation procedure was designed to use the relatively simple A.M.O.C. Bayes estimator in such a way as to avoid its shortcomings in the case of several changes. This procedure consists of "detecting" the last time point of change and applying the A.M.O.C. Bayes estimator to the following observations. It seems to be rather efficient although it can be improved upon in situations where there is available considerable information on the nature of the time points of change.

In this paper we have neglected the dynamic and compound nature of the problem. That is to say that in real applications, the problem is a sequence of estimations for, after each estimate we are given another observation and have to decide on a new estimate for the mean of the last observation.
Studying the ad-hoc procedure from this repeated point of view, letting \( \sigma \) become large and observing some of the numerical computations have suggested the following tentative procedure. As in the ad-hoc procedure one part consists of deciding if and where a change has taken place and discarding observations before the suspected change. The other part consists of using the weighted averages of the averages

\[
\overline{X}^*_n = \frac{1}{n-1} \sum_{r=i+1}^{n} X_r,
\]

as suggested by the ad-hoc procedure. The tentative procedure is formalized as follows. After \( n \) observations have been accumulated compute

\[
T_n = \max_{1 \leq i \leq n-1} \sqrt{\frac{n}{i(n-1)}} \exp \left\{ \frac{1}{2} \frac{(X_i - \overline{X}^*_n)^2}{\frac{1}{i} + \frac{1}{n-i}} \right\}
\]

and \( m \) which is that value of \( i \) for which \( T_n \) is attained. If \( T_n > L = 150 \), discard the first \( m \) observations. Then act as though only \( n-m \) observations have been accumulated and repeat the procedure.

If \( T_n \leq L \), compute

\[
\hat{\mu}_n^+ = \sum_{i=0}^{n-1} w_i \overline{X}^*_n \quad / \quad \sum_{i=0}^{n-1} w_i
\]

where \( w_0 = 1 \) and .

\[
w_i = (0.044) \sqrt{\frac{n}{i(n-1)}} \exp \left\{ \frac{1}{2} \frac{(X_i - \overline{X}^*_n)^2}{\frac{1}{i} + \frac{1}{n-i}} \right\} \quad i=1,2,\ldots,n-1.
\]
The tentative procedure involves the numbers 150 and 0.04 in the hopeful expectation that these numbers yield robust results. The authors believe in applications involving special cost functions, frequencies of change, amounts of change, or sample sizes, it may become desirable to change these numbers.

The technique of using Bayesian inference was applied as a technical device to yield insight leading to simple robust procedures. It worked remarkably well on the quality control problem formulated by Page where it led to a very simple test which compares favorably with that studied by Page. It did not work quite so well or easily on the simplified version of the tracking problem. It is to be hoped that the tentative procedure described above can be applied with minor modification in more realistic versions of the tracking problem.
References


STANFORD UNIVERSITY
TECHNICAL REPORTS DISTRIBUTION LIST
CONTRACT No N-225 (S2)

Commanding Officer
Engineering Research & Development Labs.
Fort Belvoir, Virginia
1
Commanding Officer
Frankford Arsenal
Library Branch, 0270, Bldg. 40
Bridge and Tacony Streets
Philadelphia 37, Pennsylvania
1
Commanding Officer
Rock Island Arsenal
Rock Island, Illinois
1
Commanding Officer
Redstone Arsenal (ORDW-QC)
Huntsville, Alabama
1
Commanding General
White Sands Missile Range
Attn: Tech. Library
New Mexico
1
Commanding General
Attn: Paul C. Cox, Ord. Mission
White Sands Proving Ground
Las Cruces, New Mexico
1
Commanding General
Attn: Technical Documents Center
Signal Corps Engineering Laboratory
Fort Monmouth, New Jersey
1
Commanding General
Ordnance Weapons Command
Attn: Research Branch
Rock Island, Illinois
1
Commanding General
U.S. Army Electronic Proving Ground
Fort Huachuca, Arizona
Attn: Technical Library
1
Commander
Wright Air Development Center
Attn: ARL Tech. Library, WCRR
Wright-Patterson Air Force Base, Ohio
1
Commander
Western Development Division, WDSIT
P.O. Box 262
Inglewood, California
1
Chief, Research Division
Office of Research & Development
Office of Chief of Staff
U.S. Army
Washington 25, D. C.
1
Chief, Computing Laboratory
Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland
1
Director
National Security Agency
Attn: REMP-1
Fort George G. Meade, Maryland
2
Director
Operations Research Analysis Div. AFOOP
Attn: U.S. Air Force
Washington 25, D. C.
1
Director
Snow, Ice & Permafrost Research Establishment
Corps of Engineers
1215 Washington Avenue
Wilmette, Illinois
1
Director
Lincoln Laboratory
Lexington, Massachusetts
1
Department of Mathematics
Michigan State University
East Lansing, Michigan
1

Document Library
U.S. Atomic Energy Commission
19th and Constitution Aves., N.W.
Washington 25, D. C.
1
Headquarters
Environmental Protection Agency
United States Air Force
Washington 25, D. C.
1
Institute of Statistics
North Carolina State College of A & E
Raleigh, North Carolina
1
Librarian
The RAND Corporation
1700 Main Street
Santa Monica, California
1
Library Division
Naval Missile Center Command
U.S. Naval Missile Center
Attn: J. L. Nickel
Point Mugu, California
1
Mathematics Division
Code 5077
U.S. Naval Ordnance Test Station
China Lake, California
1
NASA
Attn: Mr. R. B. Jackson, Office of Aero Intelligence
Washington 25, D. C.
1
National Applied Mathematics Lab.
National Bureau of Standards
Washington 25, D. C.
1
Office, Atty. Chief of Staff, 6-4
Research Branch, R & D Division
Department of the Army
Washington 25, D. C.
1
Superintendent
U.S. Navy Postgraduate School
Monterey, California
Attn: Library
1
Technical Information Office
Naval Research Laboratory
Washington 25, D. C.
6
Technical Information Service
Attn: Reference Branch
P.O. Box 62
Oak Ridge, Tennessee
1
Technical Library Branch
Code 234
U.S. Naval Ordnance Laboratory
Attn: Clayborn Groves
Corona, California
1
Institute for Defense Analyses
Communications Research Division
von Neumann Hall
Princeton, New Jersey
1
Mr. Irving B. Atman
Inspection & QC Division
Office, Atty., Secretary of Defense
Room 28870, The Pentagon
Washington 25, D. C.
1
Professor T. W. Anderson
Department of Statistics
Columbia College
New York 27, New York
1
Professor Robert Eichholzer
Dept. of Industrial and Engineering
Administration
Stevens School of Mechanical Engineering
Cornell University
Ithaca, New York
1

March, 1963
<table>
<thead>
<tr>
<th>Name</th>
<th>Position/Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Fred C. Andrews</td>
<td>Department of Mathematics, University of Oregon</td>
<td>Eugene, Oregon</td>
</tr>
<tr>
<td>Professor Z. W. Blumberg</td>
<td>Department of Mathematics, University of Washington</td>
<td>Seattle, Washington</td>
</tr>
<tr>
<td>Dr. David Blackwell</td>
<td>Department of Mathematical Sciences, University of California</td>
<td>Berkeley, California</td>
</tr>
<tr>
<td>Professor Julius Blum</td>
<td>Department of Mathematics, University of New Mexico</td>
<td>Albuquerque, New Mexico</td>
</tr>
<tr>
<td>Professor Ralph A. Bradley</td>
<td>Department of Statistics, Florida State University</td>
<td>Tallahassee, Florida</td>
</tr>
<tr>
<td>Dr. John W. Cell</td>
<td>Department of Mathematics, North Carolina State College</td>
<td>Raleigh, North Carolina</td>
</tr>
<tr>
<td>Professor William G. Cochran</td>
<td>Department of Statistics, Harvard University</td>
<td>Cambridge, Massachusetts</td>
</tr>
<tr>
<td>Arthur S. Martens</td>
<td>Bureau of Ships, Code 334</td>
<td>Room 3207, Main Navy</td>
</tr>
<tr>
<td>Dr. Walter L. Deemer, Jr.</td>
<td>Operations Analysis Div., DCE/USA</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Professor Cyrus Deran</td>
<td>Dept. of Industrial Engineering, Columbia University</td>
<td>New York, New York</td>
</tr>
<tr>
<td>Dr. Donald P. Gaver</td>
<td>Westinghouse Research Labs, Buhl Rd., Church Boro,</td>
<td>Pittsburgh, Pa.</td>
</tr>
<tr>
<td>Mr. Lewis A. Leake</td>
<td>Head, Operations Research Group, Code 01-2</td>
<td>Pacific Missile Range</td>
</tr>
<tr>
<td>Dr. Ivan Hershberger</td>
<td>Office, Chief of Research &amp; Dev., U.S. Army, Research Division 3E382</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Professor W. Hirsch</td>
<td>Institute of Mathematical Sciences, New York University</td>
<td>New York, New York</td>
</tr>
<tr>
<td>Mr. Eugene Hixson</td>
<td>Code 600.1, GSFC, NASA</td>
<td>Greenbelt, Maryland</td>
</tr>
<tr>
<td>Professor Harold Hotelling</td>
<td>Department of Statistics, University of North Carolina</td>
<td>Chapel Hill, North Carolina</td>
</tr>
<tr>
<td>Professor Solomon Kulback</td>
<td>Department of Statistics, George Washington University</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Professor W. K. Knuska</td>
<td>Department of Statistics, The University of Chicago</td>
<td>Chicago, Illinois</td>
</tr>
<tr>
<td>Professor Eugene Lukacs</td>
<td>Department of Mathematics, Catholic University</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Dr. Craig Moore</td>
<td>2954 Winchester Way, Rancho Cordova, California</td>
<td></td>
</tr>
<tr>
<td>Dr. Clifford Maloney</td>
<td>Biomechanics Division, Chemical Corps, U.S. Army</td>
<td></td>
</tr>
<tr>
<td>Professor G. W. McElrath</td>
<td>Department of Mechanical Engineering, University of</td>
<td></td>
</tr>
<tr>
<td>Dr. Knott T. Millets</td>
<td>Executive Director, Air Force Office of Scientific</td>
<td></td>
</tr>
<tr>
<td>Dr. William R. Paxst</td>
<td>Bureau of Weapons, Room 0306, Main Navy</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Mr. Edward Paulson</td>
<td>72-10 41 Ave. Woodside, New York</td>
<td></td>
</tr>
<tr>
<td>Dr. Walter Price</td>
<td>Chief Reliability Branch, 750, Diamond Ordnance Fuze Laboratory</td>
<td>Room 105, Building 13</td>
</tr>
<tr>
<td>Professor Ronald Pyke</td>
<td>Mathematics Department, University of Washington</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Dr. Paul Rider</td>
<td>Wright Air Development Center, WCRRM</td>
<td></td>
</tr>
<tr>
<td>Professor Herbert Robbins</td>
<td>Dept. of Mathematical Statistics, Columbia University</td>
<td>New York, New York</td>
</tr>
<tr>
<td>Professor Judith Rosenblatt</td>
<td>Department of Mathematics, University of New Mexico</td>
<td>Albuquerque, New Mexico</td>
</tr>
<tr>
<td>Professor Murray Rosenblatt</td>
<td>Department of Mathematics, Brown University</td>
<td>Providence, Rhode Island</td>
</tr>
<tr>
<td>Professor Herman Rubin</td>
<td>Department of Statistics, Michigan State University</td>
<td>East Lansing, Michigan</td>
</tr>
<tr>
<td>Professor J. S. Rustagi</td>
<td>College of Medicine, University of Cincinnati</td>
<td>Cincinnati, Ohio</td>
</tr>
<tr>
<td>Professor I. R. Savage</td>
<td>School of Business Administration, University of Minnesota</td>
<td>Minneapolis, Minnesota</td>
</tr>
<tr>
<td>Miss Marion M. Sandeir</td>
<td>2201 Cedar Street, Berkeley, Calif.</td>
<td></td>
</tr>
<tr>
<td>Professor L. J. Savage</td>
<td>Mathematics Department, University of Michigan</td>
<td>Ann Arbor, Michigan</td>
</tr>
<tr>
<td>Professor W. L. Smith</td>
<td>Statistics Department, University of North Carolina</td>
<td>Chapel Hill, North Carolina</td>
</tr>
<tr>
<td>Dr. Milton Sobel</td>
<td>Statistics Department, University of Minnesota</td>
<td>Minneapolis, Minnesota</td>
</tr>
<tr>
<td>Mr. G. P. Steck</td>
<td>Division 5511, Sandia Corp., Sandia Base</td>
<td>Albuquerque, New Mexico</td>
</tr>
<tr>
<td>Professor Donald Trux</td>
<td>Department of Mathematics, University of Oregon</td>
<td>Eugene, Oregon</td>
</tr>
<tr>
<td>Professor John W. Tukey</td>
<td>Department of Mathematics, Princeton University</td>
<td>Princeton, New Jersey</td>
</tr>
<tr>
<td>Mr. Mr. Watson</td>
<td>Department of Mathematics, University of Toronto</td>
<td>Toronto, Ontario, Canada</td>
</tr>
<tr>
<td>Dr. Harry Weinsgate</td>
<td>U.S. Army Combat and Disarmament Agency</td>
<td>State Department Bidg.</td>
</tr>
<tr>
<td>Professor S. S. Wilks</td>
<td>Office of Naval Research, Code 200</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Professor Silas Williams</td>
<td>Standards Branch, Proc. Div., Office, DC/5 for Logistics</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Professor Jacob Wolffitz</td>
<td>Department of Mathematics, Cornell University</td>
<td>Ithaca, New York</td>
</tr>
<tr>
<td>Mr. William W. Wolman</td>
<td>Code MER – Bldg., T-2 Room C301, 700 Jackson Place, N.W.</td>
<td>Washington, D.C.</td>
</tr>
<tr>
<td>Marvin Zelen</td>
<td>Mathematics Research Center, U.S. Army</td>
<td>Madison, Wisconsin</td>
</tr>
<tr>
<td>Additional copies for project leader and assistants and reserve for future requirements</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Contract No.: 225(52)
March, 1963

(171)
<table>
<thead>
<tr>
<th>Name</th>
<th>Title and Affiliation</th>
<th>Address</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr. Fred Frishman</td>
<td>Department of the Army, Chief of Research and Development</td>
<td>Room 3D442, Pentagon, Washington, D.C.</td>
<td>1</td>
</tr>
<tr>
<td>Mrs. Dorothy M. Gilford</td>
<td>Logistics and Mathematical Statistics Branch</td>
<td>Office of Naval Research, Washington, D.C.</td>
<td>1</td>
</tr>
<tr>
<td>Dr. Robert Lundegard</td>
<td>Logistics and Mathematical Statistics Branch</td>
<td>Office of Naval Research, Washington, D.C.</td>
<td>3</td>
</tr>
<tr>
<td>Mr. R. H. Noyes</td>
<td>Inst. for Exploratory Research, USASRD</td>
<td>Fort Monmouth, New Jersey</td>
<td>1</td>
</tr>
<tr>
<td>Major Oliver A. Shaw, Jr.</td>
<td>Mathematics Division</td>
<td>Air Force Office of Scientific Research, Washington 25, D.C.</td>
<td>2</td>
</tr>
<tr>
<td>Mr. Carl L. Schaniel</td>
<td>Code 122, U.S. Naval Ordnance Test Station</td>
<td>China Lake, California</td>
<td>1</td>
</tr>
<tr>
<td>Mr. J. Weinstein</td>
<td>Institute for Exploratory Research, USASRD</td>
<td>Fort Monmouth, New Jersey</td>
<td>1</td>
</tr>
</tbody>
</table>