ESTIMATING THE CURRENT MEAN OF A NORMAL DISTRIBUTION WHICH IS SUBJECTED TO CHANGES IN TIME[1]

By H. Chernoff and S. Zacks

Stanford University and Technion, Israel Institute of Technology

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 discounts 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 $X_1, X_2, \cdots, X_n$ with means $\mu_1, \mu_2, \cdots, \mu_n$ and variance 1. Each mean μ_i is equal to the preceding mean μ_{i-1} except when an occasional change takes place. The object is to estimate the current mean μ_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 σ^2. Third, we assume that the current mean μ_n is a normally distributed random variable with mean 0 and variance τ^2. Letting τ^2 approach infinity, we derive, according to these assumptions, a Bayes estimator of μ_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 (MVLU) of μ_n is also derived. The MVLU 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 MVLU 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, σ^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

Received 31 December 1963; revised 2 April 1964.

[1] This work was supported in part by the Office of Naval Research, Contract Nonr-225(52) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

is such that there is at most one change. This problem leads to a relatively simple Bayes estimator, called AMOC (at most one change) 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 AMOC 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.

In a study on "Control Charts and Stochastic Processes" [2], G. A. Barnard has tackled the control problem as a problem of estimating the current mean of a process. The general forms of the Bayes estimator and the MVLU estimator were derived in Barnard's paper in a somewhat different manner, but the general result is essentially the same. Since the objective in the present study is different from Barnard's, the study of the structure and properties of the Bayes estimation procedure is more detailed.

Our Bayesian approach seems to be more appropriate for the related problem of testing whether a change in mean has occurred than for our estimation problem. 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 $X = (X_1, \cdots, X_n)'$ be the column vector representing the n observations, and $\mu = (\mu_1, \cdots, \mu_n)'$ the column vector representing the corresponding means. Then,

$$X = \mu + \epsilon \tag{2.1}$$

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

$$\mu_i = \mu_{i+1} + J_i Z_i \qquad (i = 1, \cdots, n-1) \tag{2 2}$$

where J_i $(i = 1, \cdots, n-1)$ 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 $(i = 1, \cdots, n-1)$ is a random variable representing the amount of change, when a change takes place.

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

$$\mathcal{L}(\epsilon) = \mathfrak{N}(0, I) \tag{2.3}$$

$$\mathcal{L}(Z) = \mathfrak{N}\left(0, \sigma^2 \begin{bmatrix} 1 & & & & \\ & 1 & & 0 & \\ & & \ddots & & \\ & 0 & & 1 & \\ & & & & 0 \end{bmatrix}\right) \tag{2.4}$$

and

$$\mathfrak{L}(\mu_n) = \mathfrak{N}(0, \tau^2) \tag{2.5}$$

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

$$p(j) = P\{J = j\} \tag{2.6}$$

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

$$\begin{aligned} X_i &= \mu_n + \epsilon_i + \sum_{k=i}^{n-1} J_k Z_k, && \text{if } i = 1, \cdots, n-1 \\ &= \mu_n + \epsilon_n, && \text{if } i = n \end{aligned} \tag{2.7}$$

we obtain from (2.3) and (2.4)

$$\mathfrak{L}(X \mid \mu_n, J) = \mathfrak{N}(\mu_n e, \Sigma(J)) \tag{2.8}$$

where e is the $n \times 1$ column vector whose elements are all 1, and

$$\Sigma(J) = I + \sigma^2 J_T J_T' \tag{2.9}$$

in which J_T is the upper triangular $n \times n$ matrix

$$J_T = \begin{bmatrix} J_1 & J_2 & \cdots & J_{n-1} & 0 \\ & J_2 & \cdots & J_{n-1} & 0 \\ & & \ddots & & \\ & 0 & & J_{n-1} & 0 \\ & & & & 0 \end{bmatrix}. \tag{2.10}$$

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

$$\mathfrak{L}(X \mid J) = \mathfrak{N}(0, \Sigma^*(J)) \tag{2.11}$$

where

$$\Sigma^*(J) = \Sigma(J) + \tau^2 ee' \tag{2.12}$$

and,

$$\mathfrak{L}(X, \mu_n \mid J) = \mathfrak{N}\left(0, \begin{bmatrix} \Sigma^*(J) & \vdots & \tau^2 e \\ \cdots & \vdots & \cdots \\ \tau^2 e' & \vdots & \tau^2 \end{bmatrix}\right). \tag{2.13}$$

Hence, the *a posteriori* distribution law of μ_n, given X and J, is

$$\mathcal{L}(\mu_n \mid X, J) = \mathfrak{N}\left(\frac{e'\Sigma^{-1}(J)X}{e'\Sigma^{-1}(J)e + \tau^{-2}}, \frac{1}{e'\Sigma^{-1}(J)e + \tau^{-2}}\right). \tag{2.14}$$

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

$$p(j \mid X) = P\{J = j \mid X\} = p(j)n(X \mid 0, \Sigma^*(j))/\sum_{\{j\}} p(j)n(X \mid 0, \Sigma^*(j)) \tag{2.15}$$

where $n(X \mid \theta, \Sigma)$ represents the normal density with mean θ and covariance matrix Σ. Finally, the *a posteriori* distribution law of μ_n, given X, can be represented *schematically* by

$$\mathcal{L}(\mu_n \mid X) = \sum_{\{j\}} p(j \mid X)\mathcal{L}(\mu_n \mid X, J = j). \tag{2.16}$$

3. The minimum variance linear unbiased estimator. We present in slightly more detail than given by Barnard [2] the formula for the MVLU estimator of μ_n. This estimator is a weighted average of $X_1, X_2, \cdots, X_n$ whose weights do not depend on the observations but only on the *a priori* assumptions concerning the distribution of J and Z. Therefore, we expect the MVLU estimator to be efficient only in cases where the available information on X does not substantially affect the *a posteriori* distribution $\mathcal{L}(J \mid X)$. That is, when $\mathcal{L}(J \mid X)$ is approximately $\mathcal{L}(J)$. This is the case when changes in the means occur almost always, or when there are almost no changes.

A standard argument shows that if $\mathcal{L}(Y) = \mathfrak{N}(\mu e, V)$, the linear function of Y which is the minimum variance unbiased estimator of μ is $e'V^{-1}Y/e'V^{-1}e$. In our application, the conditional distribution of X given μ_n is easily obtained from (2.8), and we have the MVLU estimator $\tilde{\mu}_n$ of μ,

$$\tilde{\mu}_n = e'V^{-1}X/e'V^{-1}e \tag{3.1}$$

where

$$V = E\{\Sigma(J)\} = I + \sigma^2\sum_{k=1}^{n-1} p_k W_n^{(k)}, \tag{3.2}$$

$$p_k = P\{J_k = 1\} \qquad k = 1, 2, \cdots, n-1, \tag{3.3}$$

and $W_n^{(k)}$ is an $n \times n$ matrix whose upper left $k \times k$ submatrix consists of elements equal to 1 and all of whose other elements are zero.

Let ξ_i $(i = 1, 2, \cdots, n)$ be the sum of the components of the ith column vector of V^{-1}. Then we can write

$$\tilde{\mu}_n = \left[\sum_{i=1}^{n-1} \xi_i + 1\right]^{-1}\left[\sum_{i=1}^{n-1} \xi_i X_i + X_n\right]. \tag{3.4}$$

In the special case where $p_k = p$ for all $k = 1, 2, \cdots, n-1$ one obtains the following formula for ξ_i :

$$(3.5)\qquad \begin{aligned}\xi_i &= 1/\nu_1\nu_2 \cdots \nu_{n-2}(\nu_{n-1} - 1), && \text{if } i = 1\\ &= (\nu_{i-1} - 1)/\nu_{i-1} \cdots \nu_{n-2}(\nu_{n-1} - 1), && \text{if } i = 2, \cdots, n-1\end{aligned}$$

where

$$(3.6)\qquad \begin{aligned}\nu_k &= 2 + \sigma^2 p, && \text{if } k = 1\\ &= 2 + \sigma^2 p - \nu_{k-1}^{-1}, && \text{if } k = 2, \cdots, n-1.\end{aligned}$$

According to formulae (3.4)–(3.6), when $p\sigma^2$ is large the weight assigned to X_i $(i = 1, \cdots, n)$ by the MVLU estimator is of the order of magnitude of $(p\sigma^2)^{-(n-i)}$. Indeed, $\nu_k = O(p\sigma^2)$ as $p\sigma^2 \to \infty$, for all $k = 1, \cdots, n-1$. Hence, $\xi_i = O(1/(p\sigma^2)^{n-i})$ as $p\sigma^2 \to \infty$ for all $i = 1, \cdots, n$. On the other hand, if $p\sigma^2 \to 0$ the weights ξ_i are almost equal. Indeed, in the case $p\sigma^2 = 0$, $\xi_i = 1/n$ for all $i = 1, \cdots, n$.

4. The Bayes estimator. In this section we first outline a derivation of the Bayes estimator obtained by a slightly different approach by Barnard [2]. We follow this by a more leisurely development of a reduced form of the estimator which is of particular value in our applications where few changes of the mean are anticipated.

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

$$(4.1)\qquad \hat{\mu}_n = \sum_{\{j\}} p(j \mid X) e'\Sigma^{-1}(j)X/[e'\Sigma^{-1}(j)e + \tau^{-2}].$$

Substituting (2.15) in (4.1) we obtain

$$(4.2)\qquad \hat{\mu}_n = D^{-1}\sum_{\{j\}} p(j)n(X \mid 0, \Sigma^*(j))e'\Sigma^{-1}(j)X/[e'\Sigma^{-1}(j)e + \tau^{-2}]$$

where

$$(4.3)\qquad D = \sum_{\{j\}} p(j)n(X \mid 0, \Sigma^*(j)).$$

To derive a translation invariant estimator we shall let $\tau^2 \to \infty$. Then, the sum of the elements of $e'\Sigma^{-1}(j)/[e'\Sigma^{-1}(j)e + \tau^{-2}]$ 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 $(i = 1, \cdots, n)$. Note that the coefficients of $\hat{\mu}_n$ depend on X. Hence, the Bayes estimator is generally nonlinear. 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 MVLU estimator $\tilde{\mu}_n$.

In general, the limiting behavior of $\hat{\mu}_n$ involves the asymptotic behavior of $\Sigma^{*-1}(J)$ and of det $\Sigma^*(j)$, as $\tau^2 \to \infty$. It is not difficult to prove, applying matrix equalities in [1], that as $\tau \to \infty$

$$\det \Sigma^*(j) \sim \tau^2 e'\Sigma^{-1}(j)e \cdot \det \Sigma(j), \tag{4.4}$$

and

$$\Sigma^{*-1}(j) \to \Sigma^{-1}(j) - \Sigma^{-1}(j)ee'\Sigma^{-1}(j)/e'\Sigma^{-1}(j)e. \tag{4.5}$$

Denoting by $\hat{\mu}_n(j)$ the conditional Bayes estimator, given $\{J = j\}$, for the uniform *a priori* distribution of μ_n on the whole real line,

$$\hat{\mu}_n(j) = e'\Sigma^{-1}(j)X/e'\Sigma^{-1}(j)e \tag{4.6}$$

we get, according to (4.2), (4.3), (4.4) and (4.5) that the Bayes estimator for the uniform *a priori* distribution on the whole real line is:

$$\begin{aligned}\hat{\mu}_n = D^{-1}E_J\{[e'\Sigma^{-1}(J)e \cdot \det \Sigma(J)]^{-\frac{1}{2}} \\ \cdot \exp[-\tfrac{1}{2}[X - \hat{\mu}_n(J)e]'\Sigma^{-1}(J)X]\hat{\mu}_n(J)\}\end{aligned} \tag{4.7}$$

where

$$D = E_J\{[e'\Sigma^{-1}(J)e \cdot \det \Sigma(J)]^{-\frac{1}{2}} \exp[-\tfrac{1}{2}[X - \hat{\mu}_n(J)e]'\Sigma^{-1}(J)X]\} \tag{4.8}$$

and where $E_J\{\ \}$ designates the expectation operator with respect to the *a priori* distribution of J. This general formula (4.7) has been given previously by Barnard [2].

We now 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 μ_n. This permits us to reduce the general $n \times n$ matrices of formula (4.7) 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 $j_i = 1$ at $i = m_1, m_1 + m_2, \cdots, m_1 + m_2 + \cdots + m_r$ $(r = 1, 2, \cdots, n - 1)$. In this case there are r changes in the mean, taking place at the time points $m_1, m_1 + m_2, \cdots, m_1 + \cdots + m_r$. Let $m = (m_1, \cdots, m_r, m_{r+1})'$ 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:

$$\begin{aligned} Y_1 &= X_1 + \cdots + X_{m_1} \\ Y_2 &= X_{m_1+1} + \cdots + X_{m_1+m_2} \\ &\vdots \\ Y_{r+1} &= X_{m_1+\cdots+m_r} + \cdots + X_n . \end{aligned} \tag{4.9}$$

The random vector Y is a function of (X, J). Straightforward calculation yields,

$$\mathfrak{L}(Y \mid \mu_n, J = j) = \mathfrak{N}(\mu_n m, \mathfrak{S}(m)) \tag{4.10}$$

where the covariance matrix $\mathfrak{S}(m)$ is given by,

$$\mathfrak{S}(m) = m_D + \sigma^2 m_T m_T' \tag{4.11}$$

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

$$m_T = \begin{bmatrix} m_1 & m_2 & \cdots & m_r & 0 \\ & \ddots & & \vdots & \vdots \\ & & & \vdots & \vdots \\ & 0 & & m_r & 0 \\ & & & & 0 \end{bmatrix}. \tag{4.12}$$

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:

$$\hat{\mu}_n = D^{-1} \sum_{\{j\}} p(j)\psi(Y \mid j) m'\mathfrak{S}^{-1}(m)Y/m'\mathfrak{S}^{-1}(m)m \tag{4.13}$$

where,

$$D = \sum_{\{j\}} p(j)\psi(Y \mid j) \tag{4.14}$$

and where

$$\begin{aligned} \psi(Y \mid j) &= \left(\frac{m_1 \cdots m_{r+1}}{[\det \mathfrak{S}(m)] m'\mathfrak{S}^{-1}(m)m}\right)^{\frac{1}{2}} \\ &\quad \cdot \exp\left[-\frac{1}{2}(Y - \hat{\mu}_n(m))'\mathfrak{S}^{-1}(m)Y + \frac{1}{2} Y' m_D^{-1} Y\right] \end{aligned} \tag{4.15}$$

in which $\hat{\mu}_n(m)$ represents the conditional Bayes estimator, given $J = j$, namely

$$\hat{\mu}_n(m) = m'\mathfrak{S}^{-1}(m)Y/m'\mathfrak{S}^{-1}(m)m. \tag{4.16}$$

The proof of (4.13) and (4.15) is similar to previous derivations. Since the *a priori* distribution of μ_n is normal, one obtains

$$\mathfrak{L}(Y \mid J = j) = \mathfrak{N}(0, \mathfrak{S}^*(m)) \tag{4.17}$$

where

$$\mathfrak{S}^*(m) = \mathfrak{S}(m) + \tau^2 mm'. \tag{4.18}$$

Similarly,

$$\mathfrak{L}(Y, \mu_n \mid J = j) = \mathfrak{N}\left(0, \begin{bmatrix} \mathfrak{S}^*(m) & \tau^2 m \\ \tau^2 m' & \tau^2 \end{bmatrix}\right). \tag{4.19}$$

It follows that the *a posteriori* distribution of μ_n, given (X, J) is

$$\mathfrak{L}(\mu_n \mid X, J = j) = \mathfrak{N}\left(\frac{m'\mathfrak{S}^{-1}(m)Y}{m'\mathfrak{S}^{-1}(m)m + \tau^{-2}}, \frac{1}{m'\mathfrak{S}^{-1}(m)m + \tau^{-2}}\right). \tag{4.20}$$

Let $Y^* = (Y, Y^*_{r+2}, \cdots, Y^*_n)'$ where $Y^*_{r+2}, Y^*_{r+2}, \cdots, Y^*_n$ are orthonormal contrasts in $X_1, X_2, \cdots, X_n$ and hence

$$\sum_{i=1}^{r+1} (Y_i^2/n_i) + \sum_{i=r+2}^{n} Y_i^{*2} = X'X.$$

Since these contrasts are independent of μ_n and Z we have,

$$\mathfrak{L}(\mu_n \mid Y, J) = \mathfrak{L}(\mu_n \mid Y^*, J) = \mathfrak{L}(\mu_n \mid X, J). \tag{4.21}$$

Hence, the conditional Bayes estimator, given $J = j$ is $m'\mathfrak{S}^{-1}(m)Y/(m'\mathfrak{S}^{-1}(m)m + \tau^{-2})$. This estimator converges, as $\tau^2 \to \infty$, to (4.16). The conditional density of Y^*, given $J = j$, coincides with that of X, given $J = j$, except for a factor of $(m_1 \cdots m_{r+1})^{\frac{1}{2}}$. Thus,

$$\begin{aligned} n(X \mid 0, \Sigma^*(j)) &= (m_1 \cdots m_{r+1})^{\frac{1}{2}} n(Y \mid 0, \mathfrak{S}^*(m)) \\ &\quad \cdot \frac{1}{(2\pi)^{\frac{1}{2}(n-1-r)}} \exp\left[-\frac{1}{2}\sum_{i=r+2}^{n} Y_i^{*2}\right] \\ &= \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2}X'X\right\}\left(\frac{m_1 \cdots m_{r+1}}{\det \mathfrak{S}^*(m)}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}Y'(\mathfrak{S}^{*-1}(m) - m_D^{-1})Y\right\}. \end{aligned} \tag{4.22}$$

Application of (2.15) and (4.22) yields (4.13)–(4.15) after substituting in (4.22)

$$\mathfrak{S}^{*-1}(m) = \mathfrak{S}^{-1}(m) - \mathfrak{S}^{-1}(m)mm'\mathfrak{S}^{-1}(m)/m'\mathfrak{S}^{-1}(m)m + \tau^{-2} \tag{4.23}$$

and

$$\det \mathfrak{S}^*(m) = \det \mathfrak{S}(m)[1 + \tau^2 m'\mathfrak{S}^{-1}(m)m] \tag{4.24}$$

and letting $\tau^2 \to \infty$.

5. 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 $p_m = P\{J_m = 1, J_{m'} = 0 \text{ for all } m' \neq m\} = P\{J_m = 1\}$ $m = 1, \cdots, n - 1$. That is, p_0 is the *a priori* probability of no change, and p_m $(m = 1, \cdots, n - 1)$ 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 $(m = 1, \cdots, n - 1)$ the covariance matrix $\mathfrak{S}(m)$ is given by:

$$\mathfrak{S}(m) = \begin{bmatrix} m & 0 \\ 0 & n - m \end{bmatrix} + \sigma^2 \begin{bmatrix} m^2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} m(1 + \sigma^2 m) & 0 \\ 0 & n - m \end{bmatrix}. \tag{5.1}$$

Hence, the conditional Bayes estimator is, according to (4.16)

$$\hat{\mu}_n(m) = [n\bar{X}_n + \sigma^2 m(n-m)\bar{X}^*_{n-m}]/[n + \sigma^2 m(n-m)], \quad m = 1, \cdots, n-1 \tag{5.2}$$

where

$$\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{X}^*_{n-m} = (n-m)^{-1}\sum_{i=m+1}^{n} X_i \quad (m = 1, \cdots, n-1). \tag{5.3}$$

Furthermore, as easily verified, formula (4.15) is reduced in the present case to

$$\psi(Y \mid J_m = 1) = \frac{1}{(n + \sigma^2 m(n-m))^{\frac{1}{2}}} \exp\left[\frac{1}{2}\frac{m^2(n-m)^2\sigma^2}{n^2 + \sigma^2 m(n-m)n} \cdot (\bar{X}_m - \bar{X}^*_{n-m})^2\right]. \tag{5.4}$$

The conditional Bayes estimator and its corresponding weight for the case of no change can be obtained from (5.2) and (5.4) by substituting $m = 0$. Thus, the Bayes estimator for at most one change is

$$\hat{\mu}^*_n = D^{*-1}\sum_{m=0}^{n-1}\frac{p_m}{[n + \sigma^2 m(n-m)]^{\frac{1}{2}}} \cdot \exp\left[\frac{1}{2}\cdot\frac{\sigma^2 m^2(n-m)^2(\bar{X}_m - \bar{X}^*_{n-m})^2}{n^2 + \sigma^2 m(n-m)n}\right]\frac{n\bar{X}_n + \sigma^2 m(n-m)\bar{X}^*_{n-m}}{n + \sigma^2 m(n-m)} \tag{5.5}$$

where

$$D^* = \sum_{m=0}^{n-1}\frac{p_m}{[n + \sigma^2 m(n-m)]^{\frac{1}{2}}} \cdot \exp\left[\frac{1}{2}\cdot\frac{\sigma^2 m^2(n-m)^2}{n^2 + \sigma^2 m(n-m)n}(\bar{X}_m - \bar{X}^*_{n-m})^2\right]. \tag{5.6}$$

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 $\sigma^2 \to \infty$. The Bayes estimator is then approximated by

$$\mu^{**}_n = (D^{**})^{-1}\left\{\frac{p_0}{n^{\frac{1}{2}}}\bar{X}_n + \frac{1}{\sigma}\sum_{m=1}^{n-1}\frac{p_m}{[m(n-m)]^{\frac{1}{2}}} \cdot \exp\left[\frac{1}{2}\frac{m(n-m)}{n}(\bar{X}_n - \bar{X}^*_{n-m})^2\right]\bar{X}^*_{n-m}\right\} \tag{5.7}$$

where

$$D^{**} = \frac{p_0}{n^{\frac{1}{2}}} + \frac{1}{\sigma}\sum_{m=1}^{n-1}\frac{p_m}{[m(n-m)]^{\frac{1}{2}}}\exp\left[\frac{1}{2}\frac{m(n-m)}{n}(\bar{X}_m - \bar{X}^*_{n-m})^2\right]. \tag{5.8}$$

6. An ad hoc estimation procedure. The general form of the Bayes estimator, as given by (4.7) and (4.13) 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. The consequent error in the location of the "one change" leads to a poor estimate of μ_n. 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

$$B_n(m, k) = D_m^{*-1} p_k^{(m)} \psi_k(X^{(m)}), \qquad k = 0, 1, \cdots, m-1 \tag{6.1}$$

where

$$X^{(m)} = (X_{n-m+1}, \cdots, X_n)' \tag{6.2}$$

and

$$\begin{aligned} p_0^{(m)} &= (1-p)^{m-1} \\ p_k^{(m)} &= p(1-p)^{m-2} \qquad k = 1, 2, \cdots, m-1' \end{aligned} \tag{6.3}$$

and

$$\begin{aligned} \psi_k(X^{(m)}) &= \frac{1}{[m + \sigma^2 k(m-k)]^{\frac{1}{2}}} \\ &\cdot \exp\left[\frac{1}{2}\frac{k^2(m-k)^2\sigma^2}{m^2 + \sigma^2 mk(m-k)}(\bar{X}_k^{(m)} - \bar{X}_{m-k}^{*(m)2}\right] \quad \textit{for } k = 0, 1, \cdots, n-1 \end{aligned} \tag{6.4}$$

in which

$$\bar{X}_k^{(m)} = k^{-1} \sum_{i=n-m+1}^{n-m+k} X_i \tag{6.5}$$

and

$$\bar{X}_{m-k}^{*(m)} = (m-k)^{-1} \sum_{i=n-m+k+1}^{n} X_i \tag{6.6}$$

and where

$$D_m^* = \sum_{k=0}^{m-1} p_k^{(m)} \psi_k(X^{(m)}). \tag{6.7}$$

For $k > 0$, $B_n(m, k)$ is the *a posteriori* probability, based on the data $X^{(m)}$, that a change has taken place at some point 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 $B_n(m, k)$. Then $K_m = 0$ 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, $\cdots$ until we reach a non-zero K_m, say K. Then we apply the "at most one change" estimator to the observations $X_{n-K+1}, X_{n-K+2}, \cdots, X_n$.

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 Table 1.

This table gives a strong indication of a change taking place between X_5 and X_6. In fact $K_5 = 4$ 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 1

A posteriori probabilities $B(m, k)$ and estimates of μ_n based on a sample of nine observations

m	$B(m, m-8)$	$B(m, m-7)$	$B(m, m-6)$	$B(m, m-5)$	$B(m, m-4)$	$B(m, m-3)$	$B(m, m-2)$	$B(m, m-1)$	$B(m, 0)$	Estim.
2								0.0702	0.9298	−0.0737
3							0.2474	0.0722	0.6804	−0.6229
4						0.0542	0.0954	0.0660	0.7844	*−0.6301*
5					*0.7149*	0.0890	0.0088	0.0107	0.1765	−0.5749
6				0.01753	0.6812	0.1378	0.0076	0.0089	0.1470	−0.5630
7			0.0103	0.0158	0.7682	0.1398	0.0037	0.0038	0.0584	−0.6232
8		0.0037	0.0089	0.0154	0.7907	0.1438	0.0025	0.0023	0.0328	−0.6460
9	0.0025	0.0032	0.0086	0.0137	0.8538	0.1094	0.0008	0.0006	0.0074	−0.6706

7. Comparison of estimators by Monte Carlo computations. Three methods of estimating the current mean, μ_n, have been considered. These are the minimum variance linear unbiased procedures; the Bayes procedure; and an *ad hoc* procedure. Accordingly we derived four relatively simple estimators. The MVLU estimator (3.1); the "at most one-change Bayes" (AMOC Bayes) estimator (5.5); the simplified AMOC Bayes estimator (5.7); 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 μ_9 is assigned the value 0, and the observations $X_1, \cdots, X_9$ are generated according to the model:

$$(7.1) \qquad \begin{aligned} X_i &= \sigma \sum_{k=i}^{8} \eta_k + \epsilon_i, && \text{for } i = 1, \cdots, 8 \\ &= \epsilon_9, && \text{for } i = 9 \end{aligned}$$

where $\epsilon_1, \cdots, \epsilon_9; \eta_1, \cdots, \eta_8$ are independent random variables, each having a $\mathfrak{N}(0, 1)$ distribution law. In order to study the effect of the magnitude of changes in the mean on the estimators, the parameter σ 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 μ_9 is assigned the value 0, and the observations $X_1, \cdots, X_9$ are generated according to the model:

$$X_i = \sigma \sum_{k=i}^{8} J_k \eta_k + \epsilon_i, \qquad \text{for } i = 1, \cdots, 8 \tag{7.2}$$
$$= \epsilon_9, \qquad \text{for } i = 9$$

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

$$P(J_k = 1) = p, \qquad \text{for all } k = 1, \cdots, 8. \tag{7.3}$$

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,

$$X_i = \epsilon_i \qquad (i = 1, \cdots, 9) \tag{7.4}$$

where $\epsilon_1, \cdots, \epsilon_9$ are $\mathfrak{N}(0, 1)$ independent random variables.

(ii) III. B. One assigned change.

The means $\mu_1, \cdots, \mu_9$ are given, for each $m = 1, \cdots, 8$, by:

$$\mu_1 = \cdots = \mu_m = \sigma, \qquad \mu_{m+1} = \cdots = \mu_9 = 0 \tag{7.5}$$

and we consider each case $(m = 1, \cdots, 8)$ with $\sigma = 2, 3, 4$.

(iii) III. C. Two assigned changes.

Case 1. The means of the observations are given by

$$\mu_1 = \cdots = \mu_6 = 0; \qquad \mu_7 = \sigma; \qquad \mu_8 = \mu_9 = 0 \tag{7.6}$$

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 $X_1, \cdots, X_9$ are:

$$\mu_1 = \mu_2 = 2\sigma; \qquad \mu_3 = \cdots = \mu_7 = \sigma; \qquad \mu_8 = \mu_9 = 0 \tag{7.7}$$

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 σ^2 in the formulae of the estimators, we considered the following seven estimators.

Estimator 1—MVLU, $\sigma^2 = 3$, $p = 0.2$
Estimator 2—MVLU, $\sigma^2 = 20$, $p = 0.2$
Estimator 3—AMOC Bayes, $\sigma^2 = 3$, $p = 0.2$
Estimator 4—AMOC Bayes, $\sigma^2 = 20$, $p = 0.2$
Estimator 5—AMOC simplified Bayes, $\sigma^2 = 20$, $p = 0.2$
Estimator 6—*Ad hoc* procedure, $\sigma^2 = 3$, $p = 0.2$
Estimator 7—*Ad hoc* procedure, $\sigma^2 = 20$, $p = 0.2$

The means and mean-square-errors (MSE) of these seven estimators, over the 100 samples generated for most of the models, are represented in Table 2.

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" (AMOC) Bayes estimator, with the smallest value of σ^2 (Est. 3) is the most efficient one (has the smallest MSE) among the estimators studied. Moreover, the differences between the MSE's of estimators 3, 4, and 5 are negligible. This fact indicates that the AMOC 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 AMOC Bayes estimators. Less efficient than all, when there is no change, are the MVLU estimators.

(2) When changes in the mean always occur, (Model I), we expect the AMOC Bayes estimators to be poor ones, and the MVLU estimators to be better. This is verified by the numerical results which also show that these estimators are sensitive to the specification of σ.

(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 MVLU estimators. The *ad hoc* estimators performed most efficiently, and turned out to be robust against variations in the parameter σ^2 of the generated samples. The other estimators lose efficiency when σ^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 σ^2 of the generated samples is large, only the *ad hoc* estimator performed efficiently. If σ^2 of the generated samples is not large (Model II, $p = 0.2$, $\sigma = 2$), the AMOC Bayes (Est. 3, 4, 5) may perform as well 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 II B7, III B8) the AMOC 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 MVLU 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

TABLE 2

The means (upper figures) and M.S.E.'s (lower figures) of the estimators over 100 samples

Model	Est. 1	Est. 2	Est. 3	Est. 4	Est. 5	Est. 6	Est. 7
I, $\sigma = 2$	−0.2718	−0.0595	−0.1866	−0.0726	−0.0425	−0.0827	−0.0818
	2.1406	0.7333	3.3140	2.3794	2.2672	1.0235	0.9912
I, $\sigma = 3$	0.2473	0.1251	0.2734	0.1366	0.2486	0.1379	0.1358
	3.2079	0.9661	5.7730	3.8901	4.7529	1.1168	1.1202
I, $\sigma = 4$	−0.0891	−0.0111	−0.2734	−0.2754	−0.3818	0.0151	0.0095
	5.5779	1.4063	9.1200	6.1292	11.9777	1.2406	1.2441
II, $p = 0.1$	0.0847	0.0665	0.0539	0.0469	0.0487	0.0525	0.0263
$\sigma = 2$	0.4460	0.6200	0.4337	0.4312	0.4408	0.4153	0.3965
II, $p = 0.1$	−0.0649	−0.0458	−0.0917	−0.0988	−0.0989	−0.1139	−0.1351
$\sigma = 3$	0.6394	0.7172	0.6022	0.5712	0.5809	0.5891	0.5305
II, $p = 0.1$	−0.0528	−0.0853	0.0462	0.0328	0.0278	−0.0016	−0.0173
$\sigma = 4$	0.7702	0.7489	0.7385	0.6283	0.6322	0.4210	0.4179
II, $p = 0.2$	−0.0258	−0.0473	−0.0502	−0.0750	−0.0806	−0.0836	−0.0865
$\sigma = 2$	0.5462	0.7792	0.5262	0.5216	0.5331	0.6783	0.6470
II, $p = 0.2$	0.0667	0.0380	0.0487	0.0040	−0.0060	0.0397	0.0495
$\sigma = 3$	0.8594	0.7777	1.1348	0.8768	0.8632	0.5425	0.5032
II, $p = 0.2$	0.0182	0.0386	−0.0426	−0.0013	−0.0156	0.0629	0.0669
$\sigma = 4$	1.4807	1.0593	2.1400	1.5033	1.5460	0.6758	0.6892
II, $p = 0.3$	−0.1233	−0.0980	−0.1418	−0.1386	−0.1369	−0.0380	−0.0436
$\sigma = 2$	0.7718	0.7605	0.7945	0.7458	0.7522	0.6405	0.6326
II, $p = 0.3$	−0.0429	0.0377	−0.1507	−0.1489	−0.0992	0.0154	−0.0094
$\sigma = 3$	1.4082	0.9192	2.0073	1.5540	1.6971	0.7944	0.7989
II, $p = 0.3$	0.2000	0.0607	0.3241	0.2491	0.3326	−0.0286	−0.0320
$\sigma = 4$	2.3742	1.0009	3.5142	2.7221	3.2602	0.7747	0.8088
III A	0.0255	0.0325	0.0027	−0.0016	−0.0007	−0.0122	−0.0070
	0.3078	0.6112	0.1654	0.1746	0.1846	0.2679	0.2789
III B, 1	0.0034	0.0267	0.0194	−0.0018	−0.0156	−0.0003	−0.0063
$\sigma = 2$	0.4069	0.8476	0.1728	0.1920	0.2039	0.5246	0.4974
III B, 1	−0.0194	−0.0218	0.0658	−0.0158	−0.0396	−0.0647	−0.0604
$\sigma = 4$	0.3142	0.5868	0.1289	0.1246	0.1256	0.2685	0.2458
III B, 5	0.1059	0.0151	0.2658	0.1688	0.1332	0.0320	0.0464
$\sigma = 2$	0.3169	0.5948	0.3397	0.3769	0.3775	0.4146	0.4265
III B, 5	0.2047	−0.0080	0.3206	0.0722	0.0213	−0.0045	0.0128
$\sigma = 4$	0.4394	0.7475	0.3387	0.2769	0.2793	0.4598	0.4330
III B, 7	0.4415	0.0061	0.7190	0.5446	0.4751	0.1541	0.2048
$\sigma = 2$	0.5897	0.7647	1.0173	1.0892	1.0995	1.0570	1.1401
III B, 7	0.9772	0.2558	0.7161	0.2405	0.1316	0.0493	0.0451
$\sigma = 4$	1.2983	0.6901	1.0460	0.7467	0.7392	0.6589	0.6399
III B, 8	0.9522	0.3266	1.1722	0.9864	0.9036	0.5665	0.5636
$\sigma = 2$	1.2950	0.8677	1.9536	1.9667	0.9452	2.0107	2.0013
III B, 8	1.9223	0.7190	1.4325	0.6353	0.4205	0.1917	0.1775
$\sigma = 4$	4.0753	1.3022	3.2118	2.3153	2.2392	1.6658	1.6777

TABLE 2—*Continued*

Model	Est. 1	Est. 2	Est. 3	Est. 4	Est. 5	Est. 6	Est. 7
III C, 1	0.1903	0.0061	0.2633	0.2455	0.2423	0.1189	0.0955
$\sigma = 2$	0.3352	0.6211	0.2910	0.3402	0.3651	0.5251	0.4931
III C, 1	0.4145	0.0242	0.6616	0.6606	0.6605	−0.0018	−0.0167
$\sigma = 4$	0.5458	0.6719	0.8723	0.9941	1.0496	0.7509	0.7127
III C, 2	0.4964	0.1553	0.9684	0.7926	0.7472	0.2862	0.3129
$\sigma = 2$	0.6005	0.7215	1.3203	1.1810	1.1534	0.9575	0.9985
III C, 2	0.8920	0.0860	1.8091	1.4087	1.3213	−0.0229	−0.0196
$\sigma = 4$	1.1480	0.6736	4.6786	4.0504	3.9797	0.6849	0.7083

magnitude but different directions occur successively and the parameter σ^2 of the generated samples is not large ($\sigma^2 = 4$) the AMOC Bayes estimators are the best ones. The MVLU estimator with a small assigned value of $\sigma^2 p$ is equally good. The *ad hoc* estimators are slightly less efficient, and the MVLU 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 σ^2 is large. In this case the MVLU estimators are the most efficient. The *ad hoc* estimators are less efficient and the AMOC 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 σ^2 are the MVLU ones. When σ^2 is large the only efficient estimators are the *ad hoc* ones.

8. A related testing problem. 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 $X_1, \cdots, X_n$, having expected values $\mu_1, \cdots, \mu_n$ and variance 1, we wish to test the hypothesis,

$$H: \mu_1 = \cdots = \mu_n \tag{8.1}$$

against the alternatives

$$\begin{aligned} A: \mu_1 = \cdots = \mu_m; \quad \mu_{m+1} = \cdots = \mu_n; \quad 1 \leqq m \leqq n-1 \\ \mu_{m+1} - \mu_m = \delta > 0 \end{aligned} \tag{8.2}$$

where m is unknown.

We consider the two cases, μ_1 known and μ_1 unknown. 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 $M = 1, 2, \cdots, n - 1$, with p.d.f.

$$(8.3) \qquad \begin{aligned} p_m = P\{M = m\} &= (n-1)^{-1}, \quad \text{if } m = 1, 2, \cdots, n-1 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

and let Δ be a random variable representing the magnitude of change in the mean,

$$(8.4) \qquad \Delta = \mu_{m+1} - \mu_m, \qquad 0 \leqq \Delta < \infty.$$

We assign Δ a semi-normal *a priori* distribution, whose density is:

$$(8.5) \qquad \begin{aligned} h_\Delta(\delta) &= 0, & \text{if } \delta \leqq 0 \\ &= \frac{1}{\sigma}\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\delta^2\right], & \text{if } \delta > 0. \end{aligned}$$

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

$$(8.6) \qquad \begin{aligned} L(X_1, \cdots, X_n) &= \frac{(n-1)^{-1} \sum_{m=1}^{n-1} f_1(X_1, \cdots, X_n \mid M = m)}{f_0(X_1, \cdots, X_n)} \\ &= \frac{2}{n-1} \sum_{m=1}^{n-1} \exp\left[\frac{1}{2}\left(n - m + \frac{1}{\sigma^2}\right)^{-1} S_{n-m}^{*2}\right] \Phi\left(\frac{S_{n-m}^*}{(n - m + 1/\sigma^2)^{\frac{1}{2}}}\right) \end{aligned}$$

where $S_{n-m}^* = \sum_{i=m+1}^n X_i$; and $\Phi(x)$ is the $\mathfrak{N}(0, 1)$ c.d.f. The likelihood-ratio (8.6) can be written, for small values of σ, as

$$(8.7) \qquad L(X_1, \cdots, X_n) = \frac{1}{2} + \frac{\sigma}{(2\pi)^{\frac{1}{2}}} \sum_{m=1}^{n-1} S_{n-m}^* + o(\sigma), \qquad \text{as } \sigma \to 0.$$

A Bayes procedure is to reject H_0 whenever $L(X_1, \cdots, X_n)$ is greater than an appropriate constant. Accordingly we derive from (8.7) the following test statistic:

$$(8.8) \qquad T(X_1, \cdots, X_n) = \sum_{m=1}^{n-1} S_{n-m}^* = \sum_{i=1}^n (i-1)X_i.$$

A test statistic for the case of μ_1 unknown is obtained in a similar fashion. We consider μ_1 a random variable having a $\mathfrak{N}(0, \tau^2)$ *a priori* distribution, and then we let $\tau^2 \to \infty$. Computations similar to those of Section 6 yield the test statistic:

$$(8.9) \qquad T^*(X_1, \cdots, X_n) = \sum_{i=1}^n (i-1)(X_i - \bar{X}_n)$$

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

A size α test for the case $\mu_1 = 0$ is easily obtained from (8.8), since under H

$$\mathfrak{L}(T(X_1, \cdots, X_n) \mid H) = \mathfrak{N}(0, \tfrac{1}{6}n(n-1)(2n-1)). \tag{8.10}$$

Thus, a size α test criterion is

$$c_\alpha = u_{1-\alpha}[n(n-1)(2n-1)/6]^{\frac{1}{2}}, \qquad 0 < \alpha < 1 \tag{8.11}$$

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

The corresponding power function, when $\{M = m\}$ and $\{\Delta = \Delta\}$ is

$$\beta_m(\delta) = 1 - \Phi\left(u_{1-\alpha} - \delta\left[\frac{3n(n-1)}{4n-2}\right]^{\frac{1}{2}}\left[1 - \frac{m(m-1)}{n(n-1)}\right]\right) \tag{8.12}$$

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

A size α test criterion for the case μ_1 is unknown can be derived from the fact that under H we have,

$$\mathfrak{L}(T^*(X_1, \cdots, X_n) \mid H) = \mathfrak{N}(0, \tfrac{1}{12}n(n-1)(n+1)). \tag{8.13}$$

Thus, one obtains the test criterion:

$$c_\alpha^* = u_{1-\alpha}[\tfrac{1}{12}n(n^2-1)]^{\frac{1}{2}}, \qquad 0 < \alpha < 1 \tag{8.14}$$

and the power function;

$$\beta_m^*(\delta) = 1 - \Phi\left(u_{1-\alpha} - \delta\,\frac{3^{\frac{1}{2}}m(n-m)}{[n(n^2-1)]^{\frac{1}{2}}}\right) \tag{8.15}$$

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

Expression (8.12) shows that the power-function of the test statistic T (8.8) is monotonically decreasing with m. On the other hand the power function of T^* (8.9) attains its maximum, for a fixed δ, when $m \sim n/2$. Page [3] formulates the testing problem considered here in more general terms, as follows: Let $X_1, \cdots, X_n$ 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. $F(x \mid \theta)$, against the alternative that $X_1, \cdots, X_m$ $(1 \leqq m \leqq n-1)$ have a distribution function $F(x \mid \theta)$ and $X_{m+1}, \cdots, X_n$ have a distribution function $F(x \mid \theta')$ where $\theta \neq \theta'$; m and θ' are unknown but θ is known. Page proposes the following test procedure for one-sided alternatives: Record the cumulative sums

$$S_r = \sum_{i=1}^{r} (X_i - \theta) \qquad \text{for } i = 1, \cdots, n \tag{8.16}$$

and reject the null hypothesis when

$$R = \max_{0 \leqq r \leqq n} \{S_r - \min_{0 \leqq i \leqq r-1} S_i\}, \qquad S_0 \equiv 0 \tag{8.17}$$

is greater than a test criterion h. In his paper [5], Page evaluates the procedure for a class of distributions symmetric about θ. He considers the random variables $Y_i = \operatorname{sgn}(X_i - \theta)(i = 1, \cdots, n)$ and studies the operating characteristics of his test procedure for the binomial random variables $Y_1, \cdots, Y_n$, where, under the null hypothesis $P\{Y_i = 1\} = \frac{1}{2}$ for all $i = 1, \cdots, n$, and under the alternatives $P(Y_i = 1\} = \frac{1}{2}$ for $i = 1, \cdots, m$ and $P\{Y_j = 1\} = p$, $p > \frac{1}{2}$, for $j = m + 1, \cdots, n$.

The test statistic (8.8) can also be applied to Page's problem with the binomially distributed random variables $Y_1, \cdots, Y_n$. In the following table, we compare the power function of Page's test procedure (8.17) with the power function of the present Bayesian test procedure (8.8) 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 (8.8) we use for the limits of T, 87 with probability $\gamma = 0.71$ and 88 with probability 0.29.

Table 3 shows that using the test statistic (8.8) gives slightly more power than using Page's test statistic (8.17), unless the change occurs near the very beginning (m small). The test statistic (8.8) 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

TABLE 3

The power of the Page test procedure and of the Bayesian test procedure for a sample of size $n = 20$, $\alpha = 0.05$

Test	m	p						
		0.5	0.55	0.60	0.65	0.70	0.75	0.80
Page	0	0.050	0.109	0.207	0.350	0.527	0.709	0.861
Bayes	0	0.050	0.105	0.194	0.322	0.482	0.655	0.811
Page	2	0.050	0.105	0.195	0.325	0.488	0.664	0.821
Bayes	2	0.050	0.104	0.191	0.316	0.473	0.644	0.801
Page	4	0.050	0.098	0.176	0.287	0.431	0.594	0.756
Bayes	4	0.050	0.101	0.184	0.301	0.450	0.616	0.775
Page	6	0.050	0.091	0.154	0.245	0.364	0.508	0.664
Bayes	6	0.050	0.098	0.172	0.279	0.416	0.572	0.730
Page	8	0.050	0.083	0.132	0.201	0.292	0.408	0.545
Bayes	8	0.050	0.093	0.158	0.250	0.369	0.510	0.661
Page	10	0.050	0.075	0.111	0.160	0.223	0.305	0.407
Bayes	10	0.050	0.087	0.141	0.215	0.312	0.430	0.565
Page	12	0.050	0.068	0.092	0.123	0.162	0.211	0.272
Bayes	12	0.050	0.080	0.122	0.177	0.249	0.337	0.442
Page	14	0.050	0.062	0.076	0.094	0.116	0.141	0.171
Bayes	14	0.050	0.073	0.102	0.139	0.185	0.241	0.308
Page	16	0.050	0.056	0.064	0.072	0.082	0.093	0.105
Bayes	16	0.050	0.065	0.083	0.104	0.128	0.156	0.188
Page	18	0.050	0.052	0.055	0.057	0.060	0.063	0.066
Bayes	18	0.050	0.057	0.065	0.074	0.082	0.092	0.101
Exact*	10	0.050	0.099	0.167	0.261	0.382	0.524	0.679

* The exact test is the one that would be applied by the statistician who is told when the change if any takes place.

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.

9. Discussion. As shown in Section 4, 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" (AMOC) 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 AMOC 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 AMOC 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 AMOC 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 σ 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 $\bar{X}^*_{n-i} = [1/(n-i)] \sum_{r=i+1}^{n} X_r$, as suggested by the *ad hoc* procedure. The tentative procedure is formalized as follows. After n observations have been accumulated compute

$$T_n = \max_{1 \leqq i \leqq n-1} \left(\frac{n}{i(n-i)}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}\frac{(\bar{X}_i - \bar{X}^*_{n-i})^2}{(1/i + 1/(n-i))}\right\} \tag{9.1}$$

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 $T_n \leqq L$, compute

$$\hat{\mu}^+_n = \sum_{i=0}^{n-1} w_i \bar{X}^*_{n-i} \Big/ \sum_{i=0}^{n-1} w_i \tag{9.2}$$

where $w_0 = 1$ and

$$w_i = (0.04)\left(\frac{n}{i(n-i)}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}\frac{(\bar{X}_i - \bar{X}^*_{n-i})^2}{(1/i + 1/(n-i))}\right\} \tag{9.3}$$

$$i = 1, 2, \cdots, n = 1.$$

The tentative procedure involves the numbers 150 and 0.04 in the hopeful expectation that these numbers yield robust results. The authors believe that 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

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] Barnard, G. A. (1959). Control charts and stochastic processes. *J. Roy. Statist. Soc. B* **21** 239–271.

[3] Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–116.

[4] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–526.