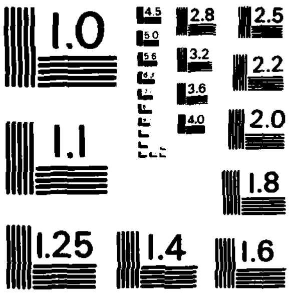
AD-A124 115 ESTIMATION IN THE PRESENCE OF NOISE OF A SIGNAL WHICH IS FLAT EXCEPT FOR..(U) MASSACHUSETTS INST OF TECH CAMBRIDGE STATISTICS CENTER Y VAD NOV 82 TR-ONR-25 1/1

UNCLASSIFIED N00014-75-C-0555 F/G 17/2 NL

END
DATE
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ESTIMATION IN THE PRESENCE OF NOISE OF A SIGNAL WHICH IS FLAT EXCEPT FOR JUMPS - PART I, A BAYESIAN STUDY

BY

YI-CHING YAO

MASSCHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL REPORT NO. ONR 25

NOVEMBER 1982

PREPARED UNDER CONTRACT

N00014-75-C-0555 (NR-609-001)

FOR THE OFFICE OF NAVAL RESEARCH

Estimation in the Presence of Noise of a Signal Which is Flat Except for Jumps - Part I, A Bayesian Study

Abstract

Consider the problem of estimating, in a Bayesian framework and in the presence of additive Gaussian noise, a signal which is a random step function. The best linear estimates, the Bayes estimates and the estimates with known change points are derived, evaluated and compared analytically and numerically. A characterization of the Bayes estimates is presented. This characterization has a reasonable interpretation and also provides a way to compute the Bayes estimates with a number of operations of the order of T^3 where T is the fixed time span. An approximation to the Bayes estimates is proposed which is reasonably good and reduces the total number of operations to the order of T.

Key words: Change points, nonlinear filtering, smoothing, Bayesian inference.

AMS 1980 subject classification: Primary 62M20, 93E14; Secondary: 62G05, 93E11, 62F15

1. Introduction

We consider the problem of estimating, in a Bayesian framework, a signal which is a step function when one observes the signal plus Gaussian noise. Optimal linear and nonlinear estimates are derived and compared.

This problem is a simplified version of a more general one, applications of which appear in many fields where the unknown underlying structure is a function, of one or more variab^ es, which is discontinuous or has discontinuous derivatives. Several examples follow.

(A) In seismology, the density of the sedimentary layers of the earth's crust can be locally approximated by a piecewise constant function.

(B) In tomography, the density of the contents of the head may have discontinuities due to tumors as well as those due to the skull.

(C) In image processing, the light intensity of a picture changes from object to object.

(D) In econometric modelling, ARMA processes are commonly used and their parameters may be subjected to changes due to sudden shifts of governmental policies and international relationships.

(E) In regression analysis, regression curves may be made up of broken straight lines.

(F) In tracking problem, a target may be liable to make sudden changes in direction.

In the above cases, it is desired to estimate the signal processes, i.e. the density functions, the intensity of light, the parameters of ARMA processes, the regression curves and the path of the moving target. They can be measured either directly with measurement error (in (C), (E)) or indirectly through various transformations (in (A), (B), (D), and (F)). There are two important and relevant problems:

(1) Can one estimate such signals efficiently?

(2) Can one detect whether or when (or where) a process changes its character?

The second problem is particularly interesting in quality control, and in the engineering literature it is called detection. Paradoxically, the first problem is called smoothing when it is applied to smooth signals. Smoothing is used in contrast to filtering where the estimate of the signal at time t is based on the observed process only up to time t and not beyond.

We shall restrict ourselves to the simple case where the signal processes are flat except for jumps and can be measured directly. In other words, in discrete time denote the signal process by $\mu_1, \mu_2, \ldots, \mu_T$ and let $\mu_{n+1} = \mu_n$ except for occasional changes. Let the observations $X_n = \mu_n + \epsilon_n$, $n = 1, 2, \ldots, T$ where the ϵ_n are measurement noise. We shall concentrate on estimating the signal process and pay little attention to detecting change

points.

If the change points were known, we could estimate μ_n by the average of the data points between the two surrounding change points. If jumps are not large, it is hard to tell when jumps take place and to take appropriate action. Moreover, if measurement noise has a heavy-tailed distribution, outliers may be disguised as jumps.

In order to develop insight for estimating the signal from the observations, we take a Bayesian point of view and consider a simple model. To be specific, we will characterize the underlying problem through the following special assumptions, which form the discrete time version of a model of Duncan (See Barnard (1959), p. 255).

(1) The sequence of the change points forms a discrete renewal process with identically geometrically distributed interarrival times.

(2) The distinct heights of the singal are mutually independent from a common Gaussian distribution.

(3) The measurement noise is Gaussian white noise.

Barnard (1959) and Chernoff and Zacks (1964) studied a similar model where the number of operations required to compute the Bayes solution is of the order of 2^T. Here T is the fixed time span. In contrast, we will see that in our case the Bayes solution can be computed with a number of operations of the order of T^3.

With respect to the above three assumptions, some *basic questions arise:*

(Qa) How well can we do if we know the change points?

(Qb) How well can we do if we use the best linear estimate?

(Qc) How well can we do with the best nonlinear estimate?

(Qd) If the parameters are not known, can we estimate them from the empirical data?

(Qe) What if the model is not satisfied?

(Qf) What about the analogous continuous time problem?

The first three of these questions are studied in great detail here. Results on the others will be presented in a forthcoming report. In Section 2, the Bayesian model is formulated more precisely. In Section 3, the minimum variance linear estimates of the signal are derived and their average mean squared error is expressed in a closed form. In Section 4, a characterization of the Bayes solution is presented which has a reasonable interpretation. In Section 5, a good approximation to the Bayes solution is proposed. In Section 6, the estimates based on the additional knowledge of the change points are considered and their asymptotic average mean squared error as $T \to \infty$ is found. In Section 7, four

types of estimates are evaluated and compared. Finally, there is an appendix where detailed derivations are presented for some of the less obvious results presented in the previous five sections.

2. The Special Bayesian Model

The three assumptions of the model are described more precisely below.

(1) Let $J = (J_1, J_2, \ldots, J_{T-1})$ be a Bernoulli sequence indicating when changes take place. i.e.

(2.1) $$J_n = 1 \text{ if there is a change between } n \text{ and } n+1, \quad = 0 \text{ otherwise.}$$

where $\Pr(J_n = 1) = p$, for $1 \le n \le T-1$. For convenience, define $J_0 \equiv J_T \equiv 1$.

(2) Let $Y_1, Y_2, \ldots, Y_T$ be i.i.d. $N(\theta, \sigma^2)$. Define the signal process $\{\mu_n\}$ recursively as follows.

(2.2) $$\mu_1 = Y_1$$

$$\mu_{n+1} = (1-J_n)\mu_n + J_n Y_{n+1}, \quad n=1,2,\ldots,T-1$$

(3) Let the observation process $\{X_n\}$ be given by

(2.3) $$X_n = \mu_n + \varepsilon_n, \quad n = 1,2,\ldots,T.$$

where the noise $\{\varepsilon_n\}$ is i.i.d. $N(0, \sigma_\varepsilon^2)$. The processes $\{J_n\}$, $\{Y_n\}$ and $\{\varepsilon_n\}$ are mutually independent.

Throughout this paper, the parameters p, θ, σ^2, and σ_ε^2 are assumed known and without loss of generality, θ and σ_ε^2 are set equal to 0 and 1, respectively.

3. The Minimum Variance Linear Estimates (MVLE).

The minimum variance linear estimates of the signal depend only on the first and second moments of the signal and observation processes. We have $E\mu_n = 0$, $EX_n = 0$, $\text{Cov}(\mu_i,\mu_j) = \sigma^2\rho^{|i-j|}$, $\text{Cov}(X_i,X_j) = \delta_{ij} + \sigma^2\rho^{|i-j|}$, and $\text{Cov}(\mu_i,X_j) = \sigma^2\rho^{|i-j|}$, where $\rho = 1 - p$.

The process $\{\mu_n\}$ has the same covariance structure as an AR(1) with parameter ρ. In other words, we may regard the linear estimation problem as the estimation of an AR(1) in the presence of white noise.

It is easy to derive that $\tilde{\mu}_n$, the MVLE of μ_n, satisfies

$$(3.1) \qquad \tilde{\mu}_n = \underset{\sim}{e}_n'(\underset{\sim}{I}-\underset{\sim}{M}^{-1})\underset{\sim}{X}$$

where $\underset{\sim}{e}_n = (0,0,\ldots,1,0,\ldots,0)'$ is the n-th natural coordinate vector,

$\underset{\sim}{I}$ = the $T\times T$ identity matrix,

$\underset{\sim}{M} = (M_{ij})_{T\times T}$, $M_{ij} = \delta_{ij} + \sigma^2\rho^{|i-j|}$,

and

$\underset{\sim}{X} = (X_1,X_2,\ldots X_T)'$.

Several explicit expressions have been derived in Snyders (1972) for the asymptotic behavior of the minimum mean squared errors as $T\to\infty$ in linear filtering, prediction and interpolation of weakly stationary discrete time processes corrupted by additive noise under very general conditions. In contrast, for finite T, explicit expressions have seldom been found. The following proposition presents a closed-form representation for $\text{AMSE}(\tilde{\mu}_n)$, the average of the mean squared errors of $\tilde{\mu}_n$. The proof can be found in Yao (1981).

Proposition 3.1

$$\begin{aligned}\text{AMSE}(\tilde{\mu}_n) &\equiv T^{-1}\sum_{n=1}^{T} E(\tilde{\mu}_n-\mu_n)^2 \\ &= 1 + T^{-1}\sigma^{-2}f_T'(-\sigma^{-2})/f_T(-\sigma^{-2}) \\ &= 1 + \sigma^{-2}u_+'(-\sigma^{-2})/u_+(-\sigma^{-2}) + o(1), (T\to\infty).\end{aligned}$$

where

$$f_T(\lambda) = a(\lambda)\ (u_+(\lambda))^T + b(\lambda)(u_-(\lambda))^T,$$

$$u_\pm(\lambda) = [1-\rho^2-\lambda(1+\rho^2)\pm\sqrt{(1-\rho^2-\lambda(1+\rho^2))^2-4\rho^2\lambda^2}]/2,$$

$$a(\lambda) = [(1-\lambda)^2-\rho^2-(1-\lambda)u_-]/(u_+^2-u_+u_-),$$

$$b(\lambda) = [(1-\lambda)u_+-(1-\lambda)^2+\rho^2]/(u_+u_--u_-^2).$$

4. The Bayes Solution - the Minimum Variance Nonlinear Estimates.

The Bayes solution can be computed by brute force with a number of operations of the order of 2^T. In this section, we present a characterization which has a reasonable interpretation and also provides a way to compute the solution with $O(T^3)$ operations.

In the following, we consider the conditional distributions of μ_n based on (1) the past and present data, $L(\mu_n|X_1,\dots,X_n)$, (2) the future data, $L(\mu_n|X_{n+1},\dots,X_T)$, and (3) all of the data, $L(\mu_n|X_1,X_2,\dots,X_T)$. We will see that $L(\mu_n|X_1,\dots,X_n)$ and $L(\mu_n|X_{n+1},\dots,X_T)$ can be computed recursively and $L(\mu_n|X_1,X_2,\dots,X_T)$ can be computed by use of $L(\mu_n|X_1,\dots,X_n)$ and $L(\mu_n|X_{n+1},\dots,X_T)$.

Here are convenient notations:

(1) $X_i^j \equiv (X_i,X_{i+1},\dots,X_j)$ $(i \le j)$.

(2) $S_o \equiv 0$, $S_n \equiv \sum_{k=1}^{n} X_k$ (cumulative sums).

(3) $L(Y) \equiv$ the distribution of random variable Y.

(4) $f_{\mu_n}(z|X_i^j) \equiv$ the conditional probability density of μ_n at z given X_i^j.

(5) "$f(x,z) \propto g(x,z)$ in z" means that there exists $c(x)$ such that $f(x,z) = c(x)g(x,z)$ for all x,z.

4.1 An Expression for $L(\mu_n|X_1^n)$.

Proposition 4.1

$$f_{\mu_{n+1}}(z|X_1^{n+1}) \propto \phi\ (X_{n+1}-z)[(1-p)f_{\mu_n}(z|X_1^n)+pf_\mu(z)]$$

in z, $1 \le n \le T-1$.

where ϕ is the standard normal density and

$$(4.1)\qquad f_\mu(x) = (2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$$

is the density of the prior signal distribution

The proof of Proposition 4.1 appears in Appendix A.1 and it is a simple application of Bayes' theorem. This proposition is an undating formula for computing $L(\mu_n|X_1^n)$, $n = 1,2,\dots,T$. Since $L(\mu_1|X_1) = N(S_1(1+\sigma^{-2})^{-1}, (1+\sigma^{-2})^{-1})$, we can demonstrate by use of induction and Proposition 4.1

Proposition 4.2

$$L(\mu_n | X_1^n) = \sum_{k=1}^{n} A_k^{(n)} \cdot N\left[\frac{S_n - S_{n-k}}{k+\sigma^{-2}}, \frac{1}{k+\sigma^{-2}}\right]$$

where

$$A_k^{(n)} = \frac{p(1-p)^{k-1}\alpha_{n-k+1}}{\sqrt{1+k\sigma^2}\,\alpha_{n+1}} \exp\left[\frac{(S_n-S_{n-k})^2}{2(k+\sigma^{-2})}\right]$$

$$(n=1,2,\ldots,T;\ k=1,\ldots,n)$$

and $\alpha_n (n=1,2,\ldots,T+1)$ are defined recursively by $\alpha_1 = 1$, and

$$\alpha_{n+1} = \sum_{k=1}^{n} \alpha_{n-k+1} \frac{p(1-p)^{k-1}}{\sqrt{1+k\sigma^2}} \exp\left[\frac{(S_n-S_{n-k})^2}{2(k+\sigma^{-2})}\right].$$

Barnard (1959) noticed that $L(\mu_n | X_1^n)$ is a mixture of n normal distributions. However, he did not present an explicit expression for the coefficients $A_k^{(n)}$.
From Proposition 4.2, we have

Proposition 4.3

$$E(\mu_n | X_1^n) = \sum_{k=1}^{n} A_k^{(n)} \cdot \frac{S_n - S_{n-k}}{k+\sigma^{-2}}$$

Remarks:

(1) Since $L(\mu_n | X_1^n, J_{n-k}=1, J_{n-k+1}=\ldots=J_{n-1}=0) = N\left[\frac{S_n-S_{n-k}}{k+\sigma^{-2}}, \frac{1}{k+\sigma^{-2}}\right]$, one may readily see from Proposition 4.2 that

$A_k^{(n)} = \Pr(J_{n-k} = 1, J_{n-k+1} = \ldots = J_{n-1} = 0 | X_1^n)$ is the conditional probability distribution of the last change point before time n.

(2) $E(\mu_n | X_1^n) = \sum_{k=1}^{n} A_k^{(n)} (S_n - S_{n-k})/(k+\sigma^{-2}) = \sum_{\ell=1}^{n} a_\ell^{(n)} X_\ell$

where

$$a_\ell^{(n)} = \sum_{k=n-\ell+1}^{n} A_k^{(n)}/(k+\sigma^{-2}).$$

Obviously, $0 < a_1^{(n)} < a_2^{(n)} < \ldots < a_n^{(n)}$, and

$$\sum_{\ell=1}^{n} a_\ell^{(n)} = \sum_{\ell=1}^{n} \sum_{k=n-\ell+1}^{n} A_k^{(n)}/(k+\sigma^{-2}) = \sum_{k=1}^{n} \frac{k}{k+\sigma^{-2}} A_k^{(n)} < \sum_{k=1}^{n} A_k^{(n)} = 1.$$

Thus, the (one-sided) Bayes solution, $E(\mu_n | X_1^n)$, is a sample dependent weighted average of the observations X_ℓ and the prior mean 0. There is a "shrinkage" toward the prior mean of the signal.

(3) The number of operations to compute α_n, given α_k, $k<n$, is $O(n)$ according to Proposition 4.2. The total number of operations to compute $E(\mu_n | X_1^n)$ for all n is $O(T^2)$.

4.2 An Expression for $L(\mu_n | X_{n+1}^T)$

As for Proposition 4.1, we can derive

Proposition 4.4

$$f_{\mu_{n-1}}(z | X_n^T) \propto (1-p)\,\phi(X_n - z) f_{\mu_n}(z | X_{n+1}^T)$$

$$+ p f_\mu(z) \int_{-\infty}^{\infty} \phi(X_n - z') f_{\mu_n}(z' | X_{n+1}^T) dz' \quad \text{in } z,$$

$$T \geq n \geq 2.$$

Since $L(\mu_T) = N(0, \sigma^2)$, we can derive by use of backward induction and Proposition 4.4

Proposition 4.5

$$L(\mu_n | X_{n+1}^T) = \sum_{k=0}^{T-n} B_k^{(T-n)} \cdot N\left[\frac{S_{n+k} - S_n}{k + \sigma^{-2}}, \frac{1}{k + \sigma^{-2}}\right]$$

where

$$B_k^{(T-n)} = \frac{p(1-p)^k \beta_{T-n-k}}{\sqrt{1 + k\sigma^2}\,\beta_{T-n}} \exp\left[\frac{(S_{n+k} - S_n)^2}{2(k + \sigma^{-2})}\right] + (1-p)\delta_{nT}$$

$$(T-n = 0, 1, \ldots, T-1; \quad k = 0, \ldots, T-n)$$

and β_{T-n} $(n = T, T-1, \ldots, 1)$ are defined recursively by $\beta_0 = 1$, and

$$\beta_{T-n} = \sum_{k=0}^{T-n-1} \beta_{T-n-1-k} \frac{p(1-p)^k}{\sqrt{1 + (k+1)\sigma^2}} \exp\left[\frac{(S_{n+k+1} - S_n)^2}{2(k + 1 + \sigma^{-2})}\right]$$

4.3 Expressions for $L(\mu_n | X_1^T)$ and $E(\mu_n | X_1^T)$

Proposition 4.6

$$f_{\mu_n}(z | X_1^T) \propto f_{\mu_n}(z | X_1^n) \quad f_{\mu_n}(z | X_{n+1}^T) / f_\mu(z) \quad \text{in } z.$$

This states that the "two-sided" conditional density of the signal is proportional to the product of the two "one-sided" conditional densities divided by its prior density. The idea of using forward and backward recursions has been introduced in the engineering literature. See Mayne (1966), Fraser (1967) and Forney (1973), Appendix.

Proof of Proposition 4.6

By Bayes' Theorem

$$f_{\mu_n}(z | X_1^T = x_1^T) = f_{X_1^T}(x_1^T | \mu_n = z) f_\mu(z) / f_{X_1^T}(x_1^T)$$

$$\propto f_{X_1^T}(x_1^T | \mu_n = z) f_\mu(z) \quad \text{in } z.$$

From the Markov property of the process $\{\mu_n\}$,

$$f_{X_1^T}(x_1^T|\mu_n=z) = f_{X_1^n}(x_1^n|\mu_n=z)\cdot f_{X_{n+1}^T}(x_{n+1}^T|\mu_n=z)$$

$$= \frac{f_{\mu_n}(z|X_1^n=x_1^n)f_{X_1^n}(x_1^n)}{f_\mu(z)} \quad \frac{f_{\mu_n}(z|X_{n+1}^T=x_{n+1}^T)f_{X_{n+1}^T}(x_{n+1}^T)}{f_\mu(z)}$$

$$\propto \frac{f_{\mu_n}(z|X_1^n=x_1^n)f_{\mu_n}(z|X_{n+1}^T=x_{n+1}^T)}{[f_\mu(z)]^2} \quad \text{in } z.$$

So,

$$f_{\mu_n}(z|X_1^T=x_1^T) \propto f_{\mu_n}(z|X_1^n=x_1^n)f_{\mu_n}(z|X_{n+1}^T=x_{n+1}^T)/f_\mu(z) \quad \text{in } z. \quad \square$$

From Propositions 4.2, 4.5, and 4.6, we can derive

Proposition 4.7

$$L(\mu_n|X_1^T) = \sum_{1\le i\le n\le j\le T} C_{ij}\cdot N\left[\frac{S_j-S_{i-1}}{j-i+1+\sigma^{-2}},\ \frac{1}{j-i+1+\sigma^{-2}}\right]$$

where

$$C_{ij} = C'_{ij}/D$$

$$D = \sum_{1\le i\le n\le j\le T} C'_{ij}$$

$$C'_{ij} = \alpha_i\beta_{T-j}\ \frac{(1-p)^{j-i}}{\sqrt{1+(j-i+1)\sigma^2}}\ \exp\left[\frac{(S_j-S_{i-1})^2}{2(j-i+1+\sigma^{-2})}\right]$$

Note: D is independent of n and therefore equal to α_{T+1}/p. See Appendix A.2.

Therefore, we have

Proposition 4.8

$$E(\mu_n|X_1^T) = \sum_{1\le i\le n\le j\le T} C_{ij}(S_j-S_{i-1})/(j-i+1+\sigma^{-2})$$

Remarks: These remarks are extensions to the two-sided case of the remarks after Proposition 4.3.

(1) Since

$$L(\mu_n|X_1^T, J_{i-1}=1, J_i=J_{i+1}=\ldots=J_{j-1}=0, J_j=1) = N\left[\frac{S_j-S_{i-1}}{j-i+1+\sigma^{-2}},\ \frac{1}{j-i+1+\sigma^{-2}}\right],$$

$$(i \le n \le j)$$

one can see from Proposition 4.7 that

$$C_{ij} = \Pr(J_{i-1}=1, J_i=\dots=J_{j-1}=0, J_j=1 \mid X_1^T).$$

Thus $\{C_{(i+1)j}: 0 \le i < n \le j \le T\}$ represents the conditional distribution of the two change points surrounding time n. So, $\Pr(J_n=1 \mid X_1^T)$ can be computed by

$$\Pr(J_n=1 \mid X_1^T) = \sum_{k=0}^{n-1} \Pr(J_k=1, J_{k+1}=\dots=J_{n-1}=0, J_n=1 \mid X_1^T) = \sum_{k=0}^{n-1} C_{(k+1)n}$$

In particular,

$$\Pr(\text{No change in } [1,T] \mid X_1^T) = C_{1T}$$

can be used to test whether changes have ever happened.

(2) $$E(\mu_n \mid X_1^T) = \sum_{1 \le i \le n \le j \le T} C_{ij}(S_j - S_{i-1})/(j-i+1+\sigma^{-2})$$

$$= \sum_{k=1}^{T} d_k^{(n)} X_k$$

where

$$d_k^{(n)} = \sum_{\substack{1 \le i \le \min(n,k) \\ \max(n,k) \le j \le T}} C_{ij}/(j-i+1+\sigma^{-2}), \quad 1 \le k \le T.$$

So, $0 < d_1^{(n)} < d_2^{(n)} < \dots < d_{n-1}^{(n)} < d_n^{(n)} > d_{n+1}^{(n)} > \dots > d_T^{(n)} > 0$

and

$$\sum_{k=1}^{T} d_k^{(n)} = \sum_{1 \le i \le n \le j \le T} \frac{j-i+1}{j-i+1+\sigma^{-2}} C_{ij} < 1.$$

Thus, the Bayes estimate $E(\mu_n \mid X_1^T)$ is a sample dependent weighted average of the observations X_k and the prior mean 0, and the weights $d_k^{(n)}$ attain their maximum at $k = n$ and decrease strictly as k moves away from n on either side.

(3) The number of operations required to compute $\alpha_n, \beta_{T-n}, C_{ij}$ $(1 \le n \le T, 1 \le i \le j \le T)$ is $O(T^2)$. The number of operations to compute $E(\mu_n \mid X_1^T)$ is $O(n(T-n))$. So the total number of operations to compute $E(\mu_n \mid X_1^T)$ for all n is $O(T^3)$.

5. An Approximation to the Bayes Solution

Harrison and Stevens (1976) proposed, for the filtering problem, an approximation technique for computing the posterior distributions of states in multi-process models. Their basic idea is to apply the following step recursively in time. First, the (estimated) posterior distribution of the state at time t is approximated by a normal distribution with the same first two moments. Next, this normal approximation is used together with the observation at $t + 1$ to estimate the posterior distribution of the state at $t + 1$.

Applying this idea, we can approximate $L(\mu_n | X_1^n)$ as follows. Suppose that $L(\mu_n | X_1^n)$ approximately equals $N(\theta_n, \tau_n^2)$. By use of Proposition 4.1, we are led to the following recursion with initial conditions $\theta_1 = X_1(1+\sigma^{-2})^{-1}$ and $\tau_1^2 = (1 + \sigma^{-2})^{-1}$.

$$= \frac{p}{1-p}\left(\frac{1+\tau_n^2}{1+\sigma^2}\right)^{1/2} \exp\left[\frac{(X_{n+1}-\theta_n)^2}{2(1+\tau_n^2)} - \frac{X_{n+1}^2}{2(1+\sigma^2)}\right]$$

(5.1) $$\theta_{n+1} = \frac{X_{n+1}+\tau_n^{-2}\theta_n}{(1+\gamma_{n+1})(1+\tau_n^{-2})} + \frac{\gamma_{n+1}X_{n+1}}{(1+\gamma_{n+1})(1+\sigma^{-2})}$$

$$\tau_{n+1}^2 = \frac{1}{1+\gamma_{n+1}}\left[\left(\frac{X_{n+1}+\tau_n^{-2}\theta_n}{1+\tau_n^{-2}}\right)^2 + \frac{1}{1+\tau_n^{-2}}\right]$$

$$+ \frac{\gamma_{n+1}}{1+\gamma_{n+1}}\left[\left(\frac{X_{n+1}}{1+\sigma^{-2}}\right)^2 + \frac{1}{1+\sigma^{-2}}\right] - \theta_{n+1}^2$$

$$n = 1,2,\ldots, T - 1$$

Similarly, suppose that $L(\mu_n | X_n^T)$ approximately equals $N(\omega_n, \delta_n^2)$. Since the system $\{\mu_n, X_n\}$ is time reversible,

ω_n, δ_n^2 satisfy the reversed-time version of (5.1).

Now, we extend Harrison-Stevens approximation to the smoothing case. We need the following variation of Proposition 4.6, the proof of which is similar to that of Proposition 4.6.

Proposition 5.1

$$f_{\mu_n}(z|X_1^T) \propto \phi(X_n-z)[(1-p)f_{\mu_{n-1}}(z|X_1^{n-1})+pf_\mu(z)]$$

$$\times [(1-p)f_{\mu_{n+1}}(z|X_{n+1}^T)+pf_\mu(z)]/f_\mu(z) \quad \text{in } z,\ 2 \le n \le T-1$$

Since we approximate $L(\mu_{n-1} | X_1^{n-1})$ and $L(\mu_{n+1} | X_{n+1}^T)$ by $N(\theta_{n-1}, \tau_{n-1}^2)$ and $N(\omega_{n+1}, \delta_{n+1}^2)$, respectively, we are naturally led by use of Proposition 5.1 to the following approximation $\hat{\mu}_n$ to $E(\mu_n | X_1^T)$, the mean of $L(\mu_n | X_1^T)$.

(5.2) $$\hat{\mu}_n = h_1^{(n)} \frac{X_n+\tau_{n-1}^{-2}\theta_{n-1}+\delta_{n+1}^{-2}\omega_{n+1}}{1+\tau_{n-1}^{-2}+\delta_{n+1}^{-2} - \sigma^{-2}} + h_2^{(n)} \frac{X_n+\tau_{n-1}^{-2}\theta_{n-1}}{1+\tau_{n-1}^{-2}}$$

$$+ h_3^{(n)} \frac{X_n+\delta_{n+1}^{-2}\omega_{n+1}}{1+\delta_{n+1}^{-2}} + h_4^{(n)} \frac{X_n}{1+\sigma^{-2}},\ 2 \le n \le T-1$$

where

$$h_i^{(n)} = g_i^{(n)}/G^{(n)}, \quad i = 1, 2, 3, 4$$

$$G^{(n)} = \sum_{i=1}^{4} g_i^{(n)}$$

$$g_1^{(n)} = (1-p)^2 F(\theta_{n-1}, \tau_{n-1}^2, \omega_{n+1}, \delta_{n+1}^2, X_n)$$

$$g_2^{(n)} = p(1-p) F(\theta_{n-1}, \tau_{n-1}^2, 0, \sigma^2, X_n)$$

$$g_3^{(n)} = p(1-p) F(0, \sigma^2, \omega_{n+1}, \delta_{n+1}^2, X_n)$$

$$g_4^{(n)} = p^2 F(0, \sigma^2, 0, \sigma^2, X_n)$$

and

$$F(A,B,C,D,X) = (BD(1-\sigma^{-2}) + B + D)^{-1/2}$$

$$\times \exp\left[\frac{(X+AB^{-1}+CD^{-1})^2}{2(1+B^{-1}+D^{-1}-\sigma^{-2})} - \frac{1}{2}(A^2B^{-1}+C^2D^{-1})\right]$$

Remarks:

(1) When $\sigma^2 < 1$, it can happen with a very small probability that $1 + \tau_{n-1}^{-2} + \delta_{n+1}^{-2} - \sigma^{-2} \le 0$. When this event happens, the right side of the formula in Proposition 5.1 does not converge to 0 as $|z|$ goes to ∞ where $f_{\mu_{n-1}}(z|X_1^{n-1})$ and $f_{\mu_{n+1}}(z|X_{n+1}^T)$ are replaced by the corresponding normal approximations. Therefore, (5.2) is no longer a good approximation. One suggestion is to replace (5.2) by $\hat{\mu}_n = \frac{X_n}{1+\sigma^{-2}}$ when $1 + \tau_{n-1}^{-2} + \delta_{n+1}^{-2} - \sigma^{-2} \le 0$.

(2) The number of operations required to compute $\hat{\mu}_n$ for all n is $0(T)$. Hence, the computational requirements of $\hat{\mu}_n$ are much smaller than those of the exact Bayes solution.

(3) We will see in Section 7 that the $\hat{\mu}_n$ are close to $E(\mu_n|X_1^T)$ in the sense of mean squared error. Thus the $\hat{\mu}_n$ are nearly optimal.

(4) It is not clear how one can apply Harrison-Stevens' idea to approximate efficiently the conditional change probabilities $Pr(J_n=1|X_1^T)$ and $Pr(\text{No change}\,|\,X_1^T)$.

6. $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$: The Estimates of μ_n Given the Change Points

In this section, we study $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ which can be used to see how much additional information for estimating μ_n is obtained from the knowledge of the change points.

Define $[r_n(\underset{\sim}{J}), s_n(\underset{\sim}{J})]$ to be the largest integral interval containing n which contains no change $(1 \le r_n(\underset{\sim}{J}) \le n \le s_n(\underset{\sim}{J}) \le T)$. Since $\underset{\sim}{X}$ and $\underset{\sim}{\mu}$ are Gaussian conditional on $\underset{\sim}{J}$, the minimum variance estimate of μ_n given $\underset{\sim}{X}$ and $\underset{\sim}{J}$ is the linear estimate

$$(6.1)\quad E(\mu_n|\underset{\sim}{X},\underset{\sim}{J}) = \sum_{k=r_n(\underset{\sim}{J})}^{s_n(\underset{\sim}{J})} X_k/(s_n(\underset{\sim}{J}) - r_n(\underset{\sim}{J}) + 1 + \sigma^{-2})$$

and

$$(6.2)\quad E((E(\mu_n|\underset{\sim}{X},\underset{\sim}{J}) - \mu_n)^2|\underset{\sim}{J}) = 1/(s_n(\underset{\sim}{J}) - r_n(\underset{\sim}{J}) + 1 + \sigma^{-2})$$

and

(6.3) $\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})) = \frac{1}{T}\sum_{n=1}^{T} E(s_n(\underset{\sim}{J}) - r_n(\underset{\sim}{J}) + 1 + \sigma^{-2})^{-1}$

The following proposition gives an explicit expression for the asymptotic behavior of $\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X},\underset{\sim}{J}))$ as $T \to \infty$. The proof can be found in Yao (1981).

Proposition 6.1

$$\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})) \equiv T^{-1}\sum_{n=1}^{T} E(E(\mu_n|\underset{\sim}{X},\underset{\sim}{J}) - \mu_n)^2$$

$$= p - p^2\sigma^{-2}(1-p)^{-\sigma^{-2}-1}\int_0^{1-p}\frac{x^{\sigma^{-2}}}{1-x}\,dx + o(1), \quad (T\to\infty).$$

7. Comparison Among Four Types of Estimates.

In this section, the performance of $\tilde{\mu}_n$, $\hat{\mu}_n$, $E(\mu_n|\underset{\sim}{X})$ and $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ is compared in terms of their average mean squared errors for $T = 20$. Sixty cases are considered where $p \in \{0.05, 0.1, 0.2, 0.4, 0.6, 0.8\}$ and $\sigma \in \{0.3, 0.5, 1, 2, 3, 4, 5, 7, 10, 15\}$.

The AMSE of $\tilde{\mu}_n$ is calculated from Proposition 3.1 while that of $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ is estimated by simulation with 2000 replications for each one of the 60 cases.

Since $E(\hat{\mu}_n-\mu_n)^2 = E[E(\mu_n|\underset{\sim}{X})-\mu_n]^2 + E[\hat{\mu}_n-E(\mu_n|\underset{\sim}{X})]^2$,

we have

(7.1) $\mathrm{AMSE}(\hat{\mu}_n) = \mathrm{AMSE}(E(\mu_n|\underset{\sim}{X})) + \mathrm{AMSE}(\hat{\mu}_n-E(\mu_n|\underset{\sim}{X}))$

where $\mathrm{AMSE}(\hat{\mu}_n - E(\mu_n|\underset{\sim}{X})) \equiv T^{-1}\sum_{n=1}^{T} E[\hat{\mu}_n-E(\mu_n|\underset{\sim}{X})]^2$. The AMSE of $E(\mu_n|\underset{\sim}{X})$ and $\hat{\mu}_n-E(\mu_n|\underset{\sim}{X})$ are estimated by simulation with 400 replications for each one of the 60 cases. The AMSE of $\hat{\mu}_n$ is estimated by use of (7.1).

The simulation results are summarized in Table 7.1 and Figure 7.1 where either p or σ^2 is fixed.

It is also interesting to compare $\hat{\mu}_n$ with $E(\mu_n|\underset{\sim}{X})$ for small p. We consider six cases where $T = 100$, $p \in \{0.01, 0.03, 0.05\}$ and $\sigma \in \{1, 3\}$. Only $E[\hat{\mu}_{100}-E(\mu_{100}|\underset{\sim}{X})]^2$ and $E[E(\mu_{100}|\underset{\sim}{X}) - \mu_{100}]^2$ are estimated. The simulation is done with 900 replications for each one of the 6 cases. The results are presented in Table 7.2.

Remarks:

(1) It can be shown that the AMSE of $\tilde{\mu}_n$ and $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ are increasing as p or σ^2 increases. So is the AMSE of $E(\mu_n|\underset{\sim}{X})$ as p increases. (See Appendix A.3). However, from the simulation results, it appears that as σ^2 increases, $\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X}))$ first increases and then decreases and eventually approaches $\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X},\underset{\sim}{J}))$. One explanation is that when σ^2 is large enough $\underset{\sim}{J}$ can be well estimated from $\underset{\sim}{X}$, and this information can offset the loss of the relatively small prior information about μ_n.

(2) From the simulation results, it appears that $E(\mu_n|\underset{\sim}{X})$ is only slightly worse than $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ in every case. However, $\tilde{\mu}_n$ is very poor when σ^2 is moderately large and p is small. Actually, if we allow $T \to \infty$ and fix σ^2, it is not difficult to show that

$$\mathrm{AMSE}(\tilde{\mu}_n) = \sqrt{\frac{\sigma^2}{2}}\, p^{1/2} + o(p^{1/2}) \qquad (p \to 0^+)$$

$$\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})) = p + o(p) \qquad (p \to 0^+) .$$

In other words, $\tilde{\mu}_n$ is very inefficient compared with $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ when p is small. The asymptotic behavior of $\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X}))$ as $p \to 0^+$ is not known.

(3) Since σ^2 can be regarded as the signal to noise ratio, it is interesting to consider relative mean squared error, i.e. mean squared error divided by the energy of the signal. In other words, σ^{-2}AMSE is used to replace AMSE. As a matter of fact, the σ^{-2} AMSE for the case that $L(\mu_n) = N(0,\sigma^2)$ and $L(\epsilon_n) = N(0,1)$ is the same as the AMSE for the case that $L(\mu_n) = N(0,1)$ and $L(\epsilon_n) = N(0,\sigma^{-2})$. Therefore, it is not hard to show that the σ^{-2} AMSE of $\tilde{\mu}_n$ and $E(\mu_n|\underset{\sim}{X})$ and $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ are decreasing as the signal to noise ratio σ^2 increases. (See Appendix A.3)

(4) It appears that $\mathrm{AMSE}(\hat{\mu}_n - E(\mu_n|\underset{\sim}{X}))$ is decreasing as p

increases. Equivalently, the larger p is, the better Harrison-Stevens' approximation is. $\mathrm{AMSE}(\hat{\mu}_n - E(\mu_n|\underset{\sim}{X}))$ is at most about 10 % of $\mathrm{AMSE}(E(\mu_n|\underset{\sim}{X}))$ in our simulation cases. Since the cost of computing $\hat{\mu}_n$ is much less than that of $E(\mu_n|\underset{\sim}{X})$, it may be desirable to substitute $\hat{\mu}_n$ for $E(\mu_n|\underset{\sim}{X})$ when $p \geq 0.05$.

(5) According to Tables 7.2, when p is very small, say about 0.01 or less, $\hat{\mu}_n$ is no longer close to optimal. In other words, the more complicated approximation is preferable for small *p*.

References

[1] Barnard, G.A. (1959). Control charts and stochastic processes. *J. Roy. Stat. Soc.* B 21, 239-271 (with discussion).

[2] Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* 35, 999-1018.

[3] Forney, G.D. (1973). The Viterbi algorithm. *IEEE Proc.* 61, 268-278.

[4] Fraser, D.C. (1967). *A New Technique for the Optimal Smoothing of Data*. Sc. D. Dissertation, MIT, Cambridge, MA.

[5] Harrison, P.J. and Stevens, C.F. (1976). Bayesian Forecasting. *J. Roy. Stat. Soc.* B 38, 205-247 (with discussion).

[6] Mayne, D.Q. (1966). A solution of the smoothing problem for linear dynamic systems. *Automatica* 4, 73-92.

[7] Snyders, J. (1972). Error formulae for optimal linear filtering prediction and interpolation of stationary time series. *Ann. Math. Statist.* 43, 1935-1943.

[8] Yao, Y.C. (1981). *The Linear Theory of Estimating Means in Time Series Subjected to Changes in Geometrically Distributed Time Intervals.* Tech. Rept. ONR21, Statist. Center, MIT, Cambridge, MA.

APPENDIX

A.1 Proof of Proposition 4.1

(1)
$$
\begin{aligned}
f_{\mu_{n+1}}(z|X_1^n=x_1^n) &= \int f_{\mu_n}(z'|X_1^n=x_1^n) f_{\mu_{n+1}}(z|X_1^n=x_1^n,\mu_n=z')dz' \\
&= \int f_{\mu_n}(z'|X_1^n=x_1^n) f_{\mu_{n+1}}(z|\mu_n=z')dz' \\
&= \int f_{\mu_n}(z'|X_1^n=x_1^n)[(1-p)\delta(z'-z)+pf_\mu(z)]dz' \\
&= (1-p)f_{\mu_n}(z|X_1^n=x_1^n) + pf_\mu(z) .
\end{aligned}
$$

(2)
$$
\begin{aligned}
f_{(\mu_{n+1},X_{n+1})}((z,x_{n+1})|X_1^n=x_1^n) \\
&= f_{(\mu_{n+1},\varepsilon_{n+1})}((z,x_{n+1}-z)|X_1^n=x_1^n) \\
&= f_{\mu_{n+1}}(z|X_1^n=x_1^n)\ \phi(x_{n+1}-z)
\end{aligned}
$$

$$
\begin{aligned}
f_{\mu_{n+1}}(z|X_1^{n+1}=x_1^{n+1}) &= f_{(\mu_{n+1},X_{n+1})}((z,x_{n+1})|X_1^n=x_1^n)/ \\
&\qquad f_{X_{n+1}}(x_{n+1}|X_1^n=x_1^n) \\
&\propto f_{(\mu_{n+1},X_{n+1})}((z,x_{n+1})|X_1^n=x_1^n) \\
&= \phi(x_{n+1}-z)\ f_{\mu_{n+1}}(z|X_1^n=x_1^n) \quad \text{(from (2))} \\
&= \phi(x_{n+1}-z)[(1-p)f_{\mu_n}(z|X_1^n=x_1^n) + pf_\mu(z)] \\
&\qquad \text{(from (1))}
\end{aligned}
$$
$\square$

A.2 Proof of D Being Independent of n in Proposition 4.7

From the recursive definitions of the α and β, for $1 \le n < n+1 \le T$

$$\beta_{T-n} \sum_{i=1}^{n} (1-p)^{i-1} \frac{\alpha_{n-i+1}}{\sqrt{1+i\sigma^2}} \exp\left[\frac{(S_n - S_{n-i})^2}{2(i+\sigma^{-2})}\right] =$$

$$= \alpha_{n+1} \sum_{j=0}^{T-n-1} (1-p)^{j} \frac{\beta_{T-n-1-j}}{\sqrt{1+(j+1)\sigma^2}} \exp\left[\frac{(S_{j+n+1} - S_n)^2}{2(j+1+\sigma^{-2})}\right]$$

i.e.

$$\sum_{i=1}^{n} (1-p)^{n-i} \frac{\alpha_i \beta_{T-n}}{\sqrt{1+(n-i+1)\sigma^2}} \exp\left[\frac{(S_n - S_{i-1})^2}{2(n-i+1+\sigma^{-2})}\right] =$$

$$= \sum_{j=n+1}^{T} (1-p)^{j-n-1} \frac{\alpha_{n+1} \beta_{T-j}}{\sqrt{1+(j-n)\sigma^2}} \exp\left[\frac{(S_j - S_n)^2}{2(j-n+\sigma^{-2})}\right]$$

i.e.

$$\sum_{i=1}^{n} C'_{in} = \sum_{j=n+1}^{T} C'_{n+1,j}$$

So,

$$\sum_{1 \le i \le n \le j \le T} C'_{ij} = \sum_{1 \le i \le n+1 \le j \le T} C'_{ij} \qquad \square$$

A.3 Proof of Monotonicity of the AMSE of $\tilde{\mu}_n$, $E(\mu_n|\underset{\sim}{X})$ and $E(\mu_n|\underset{\sim}{X},\underset{\sim}{J})$ in p and/or in σ^2.

Proposition A3.1

$E(\tilde{\mu}_n-\mu_n)^2$ is increasing as p increases for all n.

Proof of Proposition A3.1

Throughout the proof, σ^2, p and q $(0<p<q<1)$ are fixed. Without loss of generality, let the time index run from $-T'$ to T'' $(T', T'' \ge 0)$ and $n = 0$. Denote by $V(p)$ (or $V(q)$) the MSE of the minimum variance linear estimate at time 0 when the rate of change is p (or q, respectively). We want to show $V(p) \le V(q)$.

Let $\{\xi_i ; -T' \le i \le T''\}$ be a stationary Gaussian AR(1) process with parameter $1-p$ and mean 0 and variance σ^2. Let $\{\xi'_i ; i \ge 1\}$ and $\{\xi''_i ; i \ge 1\}$ be two stationary Gaussian AR(1) processes with common parameter $1-q$ and common mean 0 and common variance σ^2. Let α and β be i.i.d. with $Pr(\alpha=i) = r(1-r)^{i-1}$, $i=1,2,\ldots$ where $r=(q-p)/(1-p)$. Let $\{\varepsilon_i; -T' \le i \le T''\}$, $\{\varepsilon'_i ; i \ge 1\}$ and $\{\varepsilon''_i; i \ge 1\}$ be three independent Gaussian white noises with common variance 1. All the processes $\{\xi_i\}$, $\{\xi'_i\}$, $\{\xi''_i\}$, α, β, $\{\varepsilon_i\}$, $\{\varepsilon'_i\}$ and $\{\varepsilon''_i\}$ are mutually independent.

For i between $-T'$ and T'', define

$$X_i \equiv \xi_i + \epsilon_i \,, \qquad -T' \le i \le T''$$

$$X_i' \equiv \begin{cases} 0 \,, & i = 0 \\ \xi_i' + \epsilon_i' \,, & i \ge 1 \\ \xi_{-i}'' + \epsilon_{-i}'' \,, & i \le -1 \end{cases}$$

$$\eta_i \equiv \begin{cases} \xi_i \,, & -\beta < i < \alpha \\ \xi_i' \,, & i \ge \alpha \\ \xi_{-i}'' \,, & i \le -\beta \end{cases}$$

and

$$X_i'' \equiv \begin{cases} X_i \,, & -\beta < i < \alpha \\ X_i' \,, & \text{otherwise} \end{cases}$$

Obviously, $\{\xi_i, X_i\}$ has the same covariance structure as the special Bayesian model with the rate of change equal to p. Suppose that $\{\eta_i, X_i''\}$ has the same covariance structure as the model with the rate of change equal to q. Since $\{\xi_i, X_i\}$ is Gaussian, the minimum variance linear

estimate of ξ_0 based on $\{X_i\}$ is $E(\xi_0 | X_i, -T' \le i \le T'')$. Also note $\xi_0 = \eta_0$. So,

$$\begin{aligned} V(p) &= E[E(\xi_0 | X_{-T'}^{T''}) - \xi_0]^2 \\ &= E[E(\xi_0 | X_{-T'}^{T''}, \alpha, \beta, (X')_{-T'}^{T''}) - \xi_0]^2 \\ &\le E[E(\xi_0 | (X'')_{-T'}^{T''}) - \xi_0]^2 \\ &= E[E(\eta_0 | (X'')_{-T'}^{T''}) - \eta_0]^2 \le V(q) \end{aligned}$$

It is not hard to see that $V(p) < V(q)$ except for $T' = T'' = 0$.

The only thing left is to show that for $i \le j$

$$E\eta_i \eta_j = \sigma^2 (1-q)^{j-i}$$

$$E\eta_i X_j'' = \sigma^2 (1-q)^{j-i}$$

and

$$EX_i'' X_j'' = \delta_{ij} + \sigma^2 (1-q)^{j-i}$$

The last two equations can be easily derived from the first one.

Now consider the following cases:

(1) $0 \le i \le j$,

$$E\eta_i\eta_j = E(\eta_i\eta_j \mid \alpha \le i)\Pr(\alpha \le i) + E(\eta_i\eta_j \mid i<\alpha\le j)\Pr(i<\alpha\le j)$$

$$+ E(\eta_i\eta_j \mid j<\alpha)\Pr(j<\alpha)$$

$$= \sigma^2(1-q)^{j-i}(1-(1-r)^i) + 0 + \sigma^2(1-p)^{j-i}(1-r)^j$$

$$= \sigma^2(1-q)^{j-i}$$

(2) $i \le j \le 0$, the same as (1).

(3) $i < 0 < j$

$$E(\eta_i\eta_j) = E(\eta_i\eta_j \mid \alpha \le j)\Pr(\alpha \le j) + E(\eta_i\eta_j \mid \alpha > j, -\beta < i) \times$$

$$\Pr(\alpha > j, -\beta < i) + E(\eta_i\eta_j \mid \alpha > j, -\beta \ge i)\Pr(\alpha > j, -\beta \ge i)$$

$$= 0 + \sigma^2(1-p)^{j-i}(1-r)^j(1-r)^{-i} + 0$$

$$= \sigma^2(1-q)^{j-i}$$

o

<u>Proposition A3.2</u>

$E[E(\mu_n \mid \underset{\sim}{X}) - \mu_n]^2$ <u>is increasing as</u> p <u>increases for all</u> n.

<u>Proof of Proposition A3.2</u>

Throughout the proof, σ^2, p and q $(0 < p < q < 1)$ are fixed. Without loss of generality, let the time index run from $-T'$ to T'' and $n = 0$. Also, denote by $V(p)$ (or $V(q)$) the MSE of the Bayes estimate at time 0 when the rate of change is p (or q, respectively). We want to show $V(p) \le V(q)$.

Let $\{\xi_1 < \xi_2 < \xi_3 < \ldots\}$ and $\{\eta_1 < \eta_2 < \eta_3 < \ldots\}$ be two independent discrete time Poisson processes with parameter p. In other words, $\xi_1, \xi_2 - \xi_1, \ldots, \eta_1, \eta_2 - \eta_1, \ldots$ are i.i.d. with $\Pr(\xi_1 = i) = p(1-p)^{i-1}$, $i = 1, 2, \ldots$. Let $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ be i.i.d. $N(0, \sigma^2)$. Define

$$\mu_i \equiv \begin{cases} Y_j, & \xi_j \le i < \xi_{j+1} \\ Y_0, & -\eta_1 < i < \xi_1 \\ Y_{-k}, & -\eta_{k+1} < i \le -\eta_k \end{cases}$$

Let $X_i = \mu_i + \epsilon_i$ where $\{\epsilon_i\}$ is Gaussian white noise with $E\epsilon_1^2 = 1$. The processes $\{\epsilon_i\}$, $\{\eta_i\}$, $\{Y_i\}$ and $\{\epsilon_i\}$ are mutually independent. Thus $\{\mu_i, X_i;\ -T' \le i \le T''\}$ satisfies all the assumptions of the special Bayesian model with the rate of change equal to p.

Now, we will generate a system independent of the previous one. Let $\alpha_1, \alpha_2-\alpha_1, \alpha_3-\alpha_2, \dots, \beta_1, \beta_2-\beta_1, \beta_3-\beta_2, \dots$ be mutually independent where

$$\Pr(\alpha_1=i) = \Pr(\beta_1=i) = r(1-r)^{i-1} \quad i=1,2,\dots$$

$$\Pr(\alpha_{k+1}-\alpha_k=i) = \Pr(\beta_{k+1}-\beta_k=i) = q(1-q)^{i-1}, k=1,2,\dots; i=1,2,\dots$$

and

$$r = (q-p)/(1-p).$$

Let $\dots, Z_{-1}, Z_0, Z_1, \dots$ be i.i.d. $N(0,\sigma^2)$. Define

$$\mu_i' \equiv \begin{cases} Z_j, & \alpha_j \le i < \alpha_{j+1} \\ Z_0, & -\beta_1 < i < \alpha_1 \\ Z_{-k}, & -\beta_{k+1} < i \le -\beta_k \end{cases}$$

Let $X_i' = \mu_i' + \epsilon_i'$ where $\{\epsilon_i'\}$ is Gaussian white noise with $E\epsilon_i'^2 = 1$. The processes $\{\alpha_i\}$, $\{\beta_i\}$, $\{Z_i\}$, and $\{\epsilon_i'\}$ are mutually independent.

Now we define the third system in terms of the previous two. Define

$$\mu_i'' \equiv \begin{cases} \mu_i, & -\beta_1 < i < \alpha_1 \\ \mu_i', & \text{otherwise} \end{cases}$$

and

$$X_i'' \equiv \begin{cases} X_i, & -\beta_1 < i < \alpha_1 \\ X_i', & \text{otherwise} \end{cases}$$

Suppose that $\{\mu_i'', X_i'';\ -T' \le i \le T''\}$ satisfies all the assumptions of the model with the rate of change equal to q. Since $\mu_0'' = \mu_0$,

$$E[E(\mu_0|\underset{\sim}{X}'')-\mu_0]^2 = V(q)$$

Using the independence of the first two systems

$$\begin{aligned} V(p) &= E[E(\mu_0|\underset{\sim}{X})-\mu_0]^2 \\ &= E[E(\mu_0|\underset{\sim}{X},\alpha_1,\beta_1,\underset{\sim}{X}')-\mu_0]^2 \\ &\le E[E(\mu_0|\underset{\sim}{X}'')-\mu_0]^2 \\ &= V(q) \end{aligned}$$

It is easy to see that $V(p) < V(q)$ except for $T' = T'' = 0$.

Now, the only thing left is to show $\{\mu_i'', X_i''\}$ satisfies all the assumptions of the model with the rate of change equal to q. Actually we need only show that $\zeta_1, \zeta_2 - \zeta_1, \ldots$ are i.i.d. with $\Pr(\zeta_1 = i) = q(1-q)^{i-1}$, $i = 1, 2, \ldots$ where

$$u \equiv \sup\{k; \xi_k < \alpha_1\} \cup \{0\}$$

and

$$\zeta_i \equiv \begin{cases} \xi_i , & \text{if } 1 \le i \le u \\ \alpha_{i-u} , & \text{for } i > u \end{cases}$$

Let $\zeta_0 \equiv 0$. For every $k \geq 0$,

$$\Pr(\zeta_{k+1} - \zeta_k \ge i \mid \zeta_\ell = m_\ell, 0 \le \ell \le k, u = j < k) = \Pr(\alpha_{k+1-j} - \alpha_{k-j} \ge i \mid \zeta_\ell = m_\ell, 0 \le \ell \le k, u = j < k)$$

$$= (1-q)^{i-1}$$

$$\Pr(\zeta_{k+1} - \zeta_k \ge i \mid \zeta_\ell = m_\ell, 0 \le \ell \le k, u \ge k) = \Pr(\zeta_{k+1} - \zeta_k \ge i \mid \xi_\ell = m_\ell, 0 \le \ell \le k, \alpha_1 > \xi_k)$$

$$= \Pr(\min(\xi_{k+1}, \alpha_1) - \xi_k \ge i \mid \xi_\ell = m_\ell, 0 \le \ell \le k, \alpha_1 > \xi_k)$$

$$= \Pr(\min(\xi_1, \alpha_1) \ge i)$$

$$= (1-p)^{i-1}(1-r)^{i-1}$$

$$= (1-q)^{i-1}$$

So,

$$\Pr(\zeta_{k+1} - \zeta_k = i \mid \zeta_1, \ldots, \zeta_k) = q(1-q)^{i-1}. \qquad \square$$

Proposition A3.3

$E[E(\mu_n \mid \underset{\sim}{X}, \underset{\sim}{J}) - \mu_n]^2$ is increasing as p or σ^2 increases for all n.

Proof of Proposition A3.3

Throughout the proof σ_1^2, σ_2^2, p_1 and p_2 ($\sigma_1^2 < \sigma_2^2$, $0 < p_1 < p_2 < 1$) are fixed. Denote by $V(p, \sigma^2)$ the MSE of the estimate at time n with known change points when the rate of change is p and the variance of the signal is σ^2. We want to show

$$V(p_1, \sigma_1^2) \le V(p_2, \sigma_1^2) \quad \text{and} \quad V(p_2, \sigma_1^2) \le V(p_2, \sigma_2^2).$$

Let $\{J_i ; 1 \le i \le T-1\}$ and $\{J_i' ; 1 \le i \le T-1\}$ be two independent Bernoulli sequences with $\Pr(J_1 = 1) = p_2 = 1 - \Pr(J_1 = 0)$ and $\Pr(J_1' = 1) = p_1/p_2 = 1 - \Pr(J_1' = 0)$. Let $Y_1, \ldots, Y_T$ be i.i.d. $N(0,1)$. Define $\{\mu_i^{(1)}\}$, $\{\mu_i^{(2)}\}$, and $\{\mu_i^{(3)}\}$ by

$$\mu_1^{(1)} \equiv \sigma_1 Y_1$$

$$\mu_{n+1}^{(1)} \equiv (1-J_n)\mu_n^{(1)} + J_n \sigma_1 Y_{n+1}, \quad n=1,2,\ldots,T-1$$

$$\mu_n^{(2)} = (\sigma_2/\sigma_1)\mu_n^{(1)} \quad , \quad n=1,2,\ldots,T$$

$$\mu_1^{(3)} = \sigma_1 Y_1$$

$$\mu_{n+1}^{(3)} = (1-J_n J_n')\mu_n^{(3)} + J_n J_n' \sigma_1 Y_{n+1}, n=1,2,\ldots,T-1$$

Let $X_n^{(\alpha)} = \mu_n^{(\alpha)} + \epsilon_n$ $\alpha=1,2,3;$ $n=1,2,\ldots,T$ where $\{\epsilon_n\}$ is Gaussian white noise with $E\epsilon_1^2 = 1$. From the construction of these processes, we have

$$E[E(\mu_n^{(\alpha)} | \underset{\sim}{X}^{(\alpha)}, \underset{\sim}{J}) - \mu_n^{(\alpha)}]^2 = V(p_2, \sigma_\alpha^2), \ \alpha = 1,2$$

$$E[E(\mu_n^{(3)} | \underset{\sim}{X}^{(3)}, \underset{\sim}{JJ'}) - \mu_n^{(3)}]^2 = V(p_1, \sigma_1^2)$$

where $\underset{\sim}{JJ'}$ is the componentwise product of $\underset{\sim}{J}$ and $\underset{\sim}{J}'$.

From (6.2)

$$E([E(\mu_n | \underset{\sim}{X}, \underset{\sim}{J}) - \mu_n]^2 | \underset{\sim}{J}) = (S_n(\underset{\sim}{J}) - r_n(\underset{\sim}{J}) + 1 + \sigma^{-2})^{-1},$$

we can easily see that

$$E([E(\mu_n^{(3)} | \underset{\sim}{X}^{(3)}, \underset{\sim}{JJ'}) - \mu_n^{(3)}]^2 | \underset{\sim}{JJ'}) \le E([E(\mu_n^{(1)} | \underset{\sim}{X}^{(1)}, \underset{\sim}{J}) - \mu_n^{(1)}]^2 | \underset{\sim}{J})$$

$$E([E(\mu_n^{(1)} | \underset{\sim}{X}^{(1)}, \underset{\sim}{J}) - \mu_n^{(1)}]^2 | \underset{\sim}{J}) < E([E(\mu_n^{(2)} | \underset{\sim}{X}^{(2)}, \underset{\sim}{J}) - \mu_n^{(2)}]^2 | \underset{\sim}{J})$$

Therefore,

$$V(p_1, \sigma_1^2) \le V(p_2, \sigma_1^2)$$

$$V(p_2, \sigma_1^2) < V(p_2, \sigma_2^2)$$

It is easy to see that $V(p_1, \sigma_1^2) < V(p_2, \sigma_i^2)$ unless the time span $T = 1$. □

<u>Proposition A3.4</u>

<u>The ASME of</u> $\tilde{\mu}_n$ <u>is increasing as</u> σ^2 <u>increases.</u> This proposition can be easily proved by use of the results in Yao (1981).

<u>Proposition A3.5</u>

<u>The</u> σ^{-2} <u>AMSE of</u> $\tilde{\mu}_n$, $E(\mu_n | \underset{\sim}{X})$ <u>and</u> $E(\mu_n | \underset{\sim}{X}, \underset{\sim}{J})$ <u>are decreasing as</u> σ^2 <u>increases.</u>

From the remark (3) in Section 7 one can readily see that Proposition A3.5 is a consequence of the following lemma.

Lemma A3.1

Let $\{\mu_n;\ 1 \le n \le T\}$ *be a stochastic signal sequence with finite second moments. Let the observations* $X_n = \mu_n + \epsilon_n,\ 1 \le n \le T,$ *where* $\{\epsilon_n\}$ *is i.i.d.* $\mathcal{N}(0,\sigma_\epsilon^2)$ *and is independent of* $\{\mu_n\}$. *Then* $E[E(\mu_n|X_1^T)-\mu_n]^2$ *is increasing as* σ_ϵ^2 *increases.*

Proof of Lemma A3.1

Denote by $V_n(\sigma_\epsilon^2)$ the MSE of the Bayes estimate of μ_n when $E\epsilon_n^2 = \sigma_\epsilon^2$. We want to show that $V_n(\sigma_1^2) \le V_n(\sigma_2^2)$ for $0 \le \sigma_1^2 < \sigma_2^2 < \infty$.

Let $\{\epsilon_n\}$ and $\{\epsilon_n'\}$ be two mutually independent i.i.d. Gaussian sequences with common mean 0 and variances σ_1^2 and $\sigma_2^2 - \sigma_1^2$, respectively. $\{\mu_n\}$, $\{\epsilon_n\}$ and $\{\epsilon_n'\}$ are mutually independent. Let $X_n = \mu_n + \epsilon_n$, $X_n' = \mu_n + \epsilon_n + \epsilon_n'$, $1 \le n \le T$. Thus,

$$V_n(\sigma_1^2) = E[E(\mu_n|X_1^T)-\mu_n]^2$$

$$= E\{E[\mu_n|X_1^T,(\epsilon')_1^T]-\mu_n\}^2$$

$$\le E\{E[\mu_n|(X')_1^T]-\mu_n\}^2 = V_n(\sigma_2^2) \qquad \square$$

Table 7.1 The AMSE of Four Types of Estimates as Functions of p and σ for T = 20.

(a) $\sigma = .3$

AMSE \ p	.05	.1	.2	.4	.6	.8
Linear	.0494	.0584	.0682	.0768	.0804	.0821
Bayes	.0463 (.0027)	.0577 (.0027)	.0648 (.0025)	.0761 (.0020)	.0801 (.0019)	.0819 (.0016)
Known Change Points	.0415 (.0002)	.0481 (.0002)	.0581 (.0002)	.0694 (.0001)	.0756 (.0001)	.0798 (.0001)
(H-S)-Bayes*	.0002 (.0001)	.0001 (.0000)	.0000 (.0000)	.0000 (.0000)	.0000 (.0000)	.0000 (.0000)

(b) $\sigma = .5$

AMSE \ p	.05	.1	.2	.4	.6	.8
Linear	.0880	.1119	.1415	.1732	.1894	.1975
Bayes	.0801 (.0041)	.1125 (.0050)	.1410 (.0047)	.1741 (.0044)	.1853 (.0039)	.1957 (.0034)
Known Change Points	.0622 (.0004)	.0793 (.0005)	.1062 (.0005)	.1425 (.0004)	.1669 (.0003)	.1858 (.0002)
(H-S) - Bayes	.0017 (.0002)	.0012 (.0001)	.0007 (.0001)	.0001 (.0000)	.0000 (.0000)	.0000 (.0000)

*: The AMSE of (H-S)-Bayes = the AMSE of $\hat{\mu}_n - E(\mu_n|X)$

(c) $\sigma = 1.$

AMSE \ p	.05	.1	.2	.4	.6	.8
Linear	.1733	.2320	.3100	.4050	.4603	.4904
Bayes	.1305	.2053	.2623	.3780	.4580	.4859
	(.0068)	(.0085)	(.0077)	(.0086)	(.0088)	(.0079)
Known Change Points	.0831	.1159	.1734	.2734	.3573	.4340
	(.0008)	(.0010)	(.0011)	(.0012)	(.0010)	(.0007)
(H-S)- Bayes	.0090	.0071	.0034	.0007	.0001	.0000
	(.0005)	(.0004)	(.0003)	(.0001)	(.0000)	(.0000)

(d) $\sigma = 2.$

AMSE \ p	.05	.1	.2	.4	.6	.8
Linear	.3231	.4301	.5583	.6914	.7575	.7901
Bayes	.1561	.2484	.3683	.5639	.7033	.7648
	(.0088)	(.0108)	(.0115)	(.0133)	(.0137)	(.0116)
Known Change Points	.0930	.1345	.2161	.3741	.5183	.6625
	(.0010)	(.0013)	(.0016)	(.0020)	(.0013)	(.0015)
(H-S)- Bayes	.0158	.0131	.0077	.0016	.0003	.0000
	(.0007)	(.0005)	(.0003)	(.0001)	(.0000)	(.0000)

(e) $\sigma = 3.$

AMSE \ p	.05	.1	.2	.4	.6	.8
Linear	.4504	.5793	.7116	.8245	.8721	.8937
Bayes	.1553	.2331	.3891	.6002	.7637	.8435
	(.0092)	(.0103)	(.0134)	(.0138)	(.0155)	(.0142)
Known Change Points	.0947	.1423	.2298	.4072	.5685	.7366
	(.0010)	(.0014)	(.0018)	(.0022)	(.0021)	(.0017)
(H-S)- Bayes	.0160	.0144	.0091	.0023	.0005	.0001
	(.0009)	(.0006)	(.0005)	(.0001)	(.0000)	(.0000)

(f) $\sigma = 4.$

AMSE \ p	.05	.1	.2	.4	.6	.8
Linear	.5550	.6866	.8039	.8901	.9229	.9371
Bayes	.1462	.2302	.3852	.6016	.7796	.8765
	(.0090)	(.0105)	(.0143)	(.0144)	(.0161)	(.0148)
Known Change Points	.0967	.1411	.2331	.4113	.5890	.7658
	(.0010)	(.0014)	(.0019)	(.0022)	(.0022)	(.0018)
(H-S)- Bayes	.0153	.0141	.0093	.0027	.0006	.0001
	(.0008)	(.0006)	(.0004)	(.0001)	(.0000)	(.0000)

(g) $\sigma = 5$.

AMSE \ P	.05	.1	.2	.4	.6	.8
Linear	.6387	.7623	.8606	.9257	.9489	.9587
Bayes	.1492	.2316	.3521	.6191	.7771	.8931
	(.0094)	(.0105)	(.0125)	(.0143)	(.0152)	(.0156)
Known Change Points	.0957	.1440	.2370	.4180	.6062	.7837
	(.0010)	(.0014)	(.0019)	(.0023)	(.0023)	(.0019)
(H-S)-	.0155	.0140	.0086	.0028	.0006	.0001
Bayes	(.0011)	(.0007)	(.0005)	(.0001)	(.0000)	(.0000)

(h) $\sigma = 7$.

AMSE \ P	.05	.1	.2	.4	.6	.8
Linear	.7563	.8546	.9211	.9601	.9731	.9785
Bayes	.1333	.2006	.3385	.5826	.7800	.9067
	(.0085)	(.0102)	(.0129)	(.0145)	(.0161)	(.0153)
Known Change Points	.0968	.1448	.2390	.4220	.6078	.7977
	(.0010)	(.0015)	(.0019)	(.0024)	(.0023)	(.0019)
(H-S)-	.0136	.0125	.0086	.0029	.0008	.0001
Bayes	(.0010)	(.0007)	(.0004)	(.0001)	(.0000)	(.0000)

(i) $\sigma = 10$.

AMSE \ P	.05	.1	.2	.4	.6	.8
Linear	.8547	.9200	.9589	.9799	.9866	.9893
Bayes	.1270	.2011	.3282	.5874	.7610	.8965
	(.0079)	(.0099)	(.0131)	(.0154)	(.0163)	(.0157)
Known Change Points	.0973	.1466	.2373	.4266	.6145	.8039
	(.0011)	(.0014)	(.0019)	(.0024)	(.0024)	(.0020)
(H-S)-	.0113	.0120	.0071	.0030	.0009	.0001
Bayes	(.0008)	(.0008)	(.0006)	(.0002)	(.0001)	(.0000)

(j) $\sigma = 15$.

AMSE \ P	.05	.1	.2	.4	.6	.8
Linear	.9266	.9619	.9811	.9909	.9940	.9952
Bayes	.1206	.1768	.3017	.5350	.7344	.8971
	(.0075)	(.0084)	(.0114)	(.0142)	(.0159)	(.0154)
Known Change Points	.0966	.1460	.2390	.4263	.6190	.8086
	(.0011)	(.0015)	(.0019)	(.0024)	(.0024)	(.0019)
(H-S)-	.0099	.0099	.0068	.0030	.0009	.0001
Bayes	(.0010)	(.0007)	(.0005)	(.0002)	(.0001)	(.0000)

Tables 7.2 The MSE of Four Types of Estimates of μ_{100} as Functions of p and σ for T = 100.

(a) σ = 1.

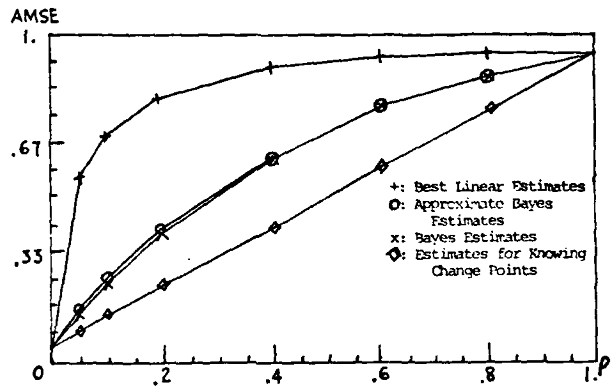
MSE \ P	.01	.03	.05
Linear	.1236	.1956	.2380
Bayes	.0731	.1399	.1845
	(.0082)	(.0125)	(.0144)
Known Change Points	.0366	.0829	.1139
	(.0022)	(.0035)	(.0040)
(H-S)-	.0248	.0175	.0120
Bayes	(.0022)	(.0011)	(.0009)

(b) σ = 3.

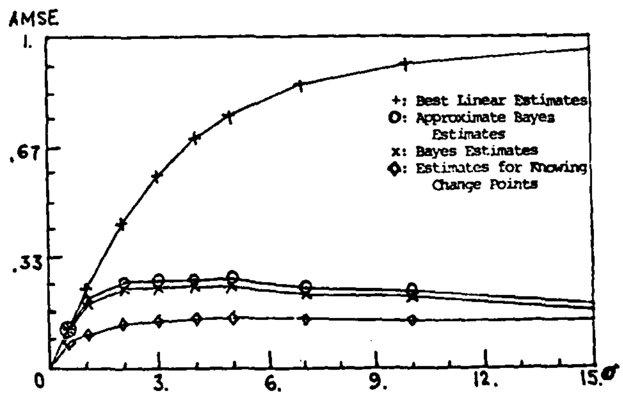
MSE \ P	.01	.03	.05
Linear	.3378	.5008	.5841
Bayes	.0982	.2054	.2489
	(.0187)	(.0215)	(.0265)
Known Change Points	.0497	.1049	.1472
	(.0040)	(.0061)	(.0067)
(H-S)-	.0279	.0314	.0254
Bayes	(.0016)	(.0018)	(.0016)

Figure 7.1 AMSE as a function of p and σ .

(a) σ = 4



(b) p = 0.1



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 25	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Estimation in the Presence of Noise of a Signal Which is Flat Except for Jumps - Part I, A Bayesian Study		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Yi-Ching Yao		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0555
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Center, Massachusetts Institute of Technology, Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-609-001)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Statistics and Probability Code 436, Arlington, VA 22217		12. REPORT DATE November 10, 1982
		13. NUMBER OF PAGES 48
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Change points, nonlinear filtering, smoothing, Bayesian Inference

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Consider the problem of estimating, in a Bayesian framework and in the presence of additive Gaussian noise, a signal which is a random step function. The best linear estimates, the Bayes estimates and the estimates with known change points are derived, evaluated and compared analytically and numerically. A characterization of the Bayes estimates is presented. This characterization has a reasonable interpretation and also provides a way to compute the Bayes estimates with a number of operations of the order of T^3 where T is the fixed time span. An approximation to the Bayes estimates is proposed which is reasonably good and reduces the total number of operations to the order of T.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

DATE
FILMED

3-83

DTIC