ESTIMATION IN THE PRESENCE OF NOISE OF A SIGNAL WHICH IS FLAT EXCEPT FOR...

(U) MASSACHUSETTS INST OF TECH

CAMBRIDGE STATISTICS CENTER Y VAO JAN 03 TR-OMR-28

UNCLASSIFIED NO0014-79-C-0889

F/G 12/1 NL
ESTIMATION IN THE PRESENCE OF NOISE OF A SIGNAL WHICH IS FLAT EXCEPT FOR JUMPS - PART II, THE EMPIRICAL BAYES APPROACH

By
Yi-Ching Yao
Massachusetts Institute Of Technology

TECHNICAL REPORT NO. ONR 28
JANUARY 1983

PREPARED UNDER CONTRACT
N00014-75-C-0555 (NR-609-001)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in whole or in part is permitted for any purpose of the United States Government

This document has been approved for public release and sale; its distribution is unlimited
Estimation in the Presence of Noise of a Signal Which is Flat Except for Jumps - Part II. The Empirical Bayes Approach

Abstract

This is the second of a two-part paper. In the first part Yao (1982), a special Bayesian Model A is studied in detail. In this part, a more general model is proposed and studied in an empirical Bayes framework. The results for Model A are applied to step-function signals using the ideas of empirical Bayes and maximum likelihood applied to the parameters of the Bayesian Model A. An efficient computational method is proposed to approximate the likelihood function under Model A. Several empirical Bayes estimators of the unknown step-function signal are compared by simulation.

Key words: Change points, nonlinear filtering, smoothing, empirical Bayes, maximum likelihood, pseudo maximum likelihood, Kullback Information.

AMS 1980 subject classification: Primary 62M20, 93E14; secondary 62C12, 62G95, 93E11

1. Introduction

This is the second of a two-part paper. We consider the problem of estimating a signal which is a step function when one observes the signal plus noise. In other words, in discrete time denote the signal process by \( \psi_1, \psi_2, \ldots, \psi_T \) and let \( \psi_n = \psi_{n-1} + \epsilon_n \) except for occasional changes. Let the observations \( X_i = \psi_i + c_i, 1 \leq n \leq T \) where the \( c_i \) are noise. We are interested in estimating \( \psi_n \) based on \( X_i, 1 \leq i \leq T \). In the first part Yao (1982), we studied this problem in a Bayesian framework. A special Bayesian model (to be called Model A) was proposed there and the corresponding Bayes solution was derived and evaluated analytically and numerically. In the second part, we will invoke the idea of empirical Bayes to attack more general cases where not all of the assumptions of Model A are satisfied.

In the next section, a generalization of Model A is proposed. In Section 3, partial results are obtained on identifying the underlying distributions and estimating optimally the step-function signal. In Section 4, the results for Model A are applied to more general step-function signals using the ideas of empirical Bayes and maximum likelihood. In Section 5, an approximation to the likelihood is proposed and this approximation is evaluated in terms of the Kullback information. In
Section 6, several empirical Bayes estimators are studied by use of simulation.

2. A General Bayesian Model

In this section we propose the following model.

1. The time intervals between successive changes in the signal are i.i.d.
2. The successive heights of the signal are i.i.d.
3. The additive noise is an i.i.d. sequence.

To be more specific,

1'(i) Let \( \xi, \xi_1, \xi_2, \xi_3, \ldots \) be i.i.d. (\( F_\xi \)), positive integer valued with finite first moment. Let \( \xi' \) be independent of \( (\xi_n) \) and

\[
\Pr(\xi' = \xi) = \Pr(\xi' = \xi)/\xi \quad \xi = 1, 2, \ldots
\]

Define the sequence of change points \( \{n_k\} \) by

\[
n_0 = 0, n_1 = \xi', n_2 = \xi'_1, \ldots, n_k = \xi'_{k-1} \quad \xi'_{k-1} = \xi'_{k-2} \ldots = \xi'_0 = n_0 = 0, \ldots
\]

Note: The random variable \( \xi' \) is introduced in order that the 0-1 sequence generated by \( (n_1, n_2, \ldots) \) (ones at the \( n_k \) and zeros elsewhere) be stationary. This is a matter of convenience and is not essential as far as asymptotic results are concerned.

(2') Let \( Y, Y_0, Y_1, Y_2, \ldots \) be i.i.d. (\( F_y \)) and represent successive heights of the signal. i.e., define the signal process \( (u_n) \) to be

\[
u_n = Y_k \quad \text{for} \quad n \leq n_k + 1
\]

(3') Let the additive noise \( \xi, \xi_1, \xi_2, \ldots, \xi_T \) be i.i.d. (\( F_\xi \)). Assume \( E\xi = 0 \). Let the observations be

\[
X_n = u_n + \xi_n \quad n = 1, 2, \ldots,T
\]

Note: Model A is a special case of this general model when \( F_\xi \) is geometrical and \( F_y \) and \( F_\xi \) are normal. Model A can be described by four parameters \( p, a, b, c \) where \( \Pr(\xi' = 1) = p, \Pr(\xi' = 2) = F_y(0, a^2) \) and \( F_\xi(0, b, c) \).

Suppose \( F_y, F_\xi, \) and \( F_\xi \) are unknown. Two natural questions arise:

01. Are they identifiable?
02. Can \( \nu_n \) be estimated "optimally"?

We designate the subsequence \( (X_n, 1 \leq n \leq T) \) by \( \xi_n^T \) and we shall call estimators \( \hat{u}_n(\xi_n^T) \) of \( u_n \) uniformly asymptotically optimal relative to \( F_y, F_\xi, \) and \( F_\xi \) if
The proofs of (3.1) and (3.2) appear in the Appendix.

We have

$$\lim_{T \to \infty} E[(R_{m(T)} - R_n)] = 0$$

where 

$$E_{(r)} \text{ means expectation according to the probability structure determined by } F_r, T_r, \text{ and } F_x.$$}

This definition is consistent in essence with that in Robbins (1964).

The partial answers to (21) and (22)

Proposition 3.1 (Strong Consistency of \( \theta \))

Assume that 

$$\text{Var}(y) > 0, \text{ and } E[r^2] < \infty.$$ Then there is an estimate \( \theta \) such that 

$$\theta_c \to \theta \text{ w.p.1.}$$

Proof of Proposition 3.1

Although \((X_n)\) is not a weakly dependent sequence, the following are still true.

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t = E(\theta) \text{ a.s., } 1 \leq \theta \leq 4.$$
Using (3.1) through (3.11) we shall show that $\sigma^2$ can be consistently estimated. First it may be seen that

$$\omega(\tau^2) = (\tau^2)^{2(-2)(\lambda^2-\lambda_1^2)} (\tau^2) (\lambda_2, \lambda_2^2, \lambda_2^3) \omega \sigma$$

The case where $\sigma^2 = 0$ is special, for then $\omega = 1$ and a change takes place at each time point. Then there is no way to distinguish the signal from the noise without additional information. For this reason we consider two cases:

1. When $\lambda_2 - \lambda_1^2 > \tau^{-1/2}$, estimate $\sigma^2$ by $\hat{\sigma}^2 := \sigma^2$ and therefore estimate $p_{\theta}$ by $\hat{p}_{\theta}$, the distribution with unit mass at $\theta$. 

(3.12) $B_4 := \frac{1}{n} \sum_{n=1}^{T-1} X_{n+1}^2 - \sigma^2 (EY^2 + 2EC^2 + (C^2)^2)$
When $\mathcal{B}_1 - A_2^2 \geq T^{-1/3}$, estimate $E_T^2$ by the larger solution (denoted by $E_T^2$) of the quadratic equation $\eta(E_T^2) = 0$. Since $\eta(A_1^2) = 0$, $E_T^2 \geq A_2^2$. It will be shown $ET^2 = ET^2 (T^{-\infty})$ a.s. when $\mathcal{B}_1 > 0$. By (3.5) and (3.9), estimate $\mathcal{B}_1$ by $\tilde{B}_1 = \min(B_1, A_2)$. By (3.4), estimate $E_k$ by $E_k = \frac{1}{1-\lambda}$. By (3.4) and (3.3), estimate $\mathcal{B}_1$ by $\tilde{B}_1 = \frac{1}{1-\lambda} \frac{1}{1-\lambda} - \frac{1}{1-\lambda} (E_T^2 - A_2^2)$ for $2 \leq k \leq \ceil{\log T} + 1$. Applying (3.4) which relates $\mathcal{B}_1$ and $\mathcal{B}_k$, we are led to introduce $\tilde{\mathcal{B}}_k$ recursively by

\[ \tilde{\mathcal{B}}_k = 2 - \delta_k(1-\tilde{\mathcal{B}}_k) \]

and

\[ \tilde{\mathcal{B}}_k = \frac{b_k-1}{b_k} \left( \frac{1}{\delta_k+1} \left( \frac{1}{\delta_k} \frac{1}{\delta_k} \right) - \tilde{\mathcal{B}}_k(1-\tilde{\mathcal{B}}_k) \right) \]

Let us estimate $\mathcal{F}_1$, $\mathcal{F}_2$, ... by

\[ \mathcal{F}_k = \min(\tilde{\mathcal{B}}_k, 1) \]

\[ \tilde{\mathcal{F}}_k = \max(\tilde{\mathcal{B}}_k(0), \tilde{\mathcal{F}}_k(0), \tilde{\mathcal{F}}_k(0), \tilde{\mathcal{F}}(0), \tilde{\mathcal{F}}(0)) \]

\[ \mathcal{F}(\log T) = \frac{1}{\log T} \tilde{\mathcal{F}}(\log T) \]

Now, we show that the estimate of $\mathcal{F}_k$ is consistent. There are two different cases:

1. If $Pr(\mathcal{F}_1 = 1) = 1$, i.e. $\mathcal{B}_1 = 0$, then

\[ A_2^2 = (E_T^2) (T^{-\infty}) \text{ a.s.} \]

\[ A_1^2 = (E_T^2) (T^{-\infty}) \text{ a.s.} \]

Since $\{X_n\}$ is i.i.d., we can apply the law of the iterated logarithm (Philipp and Stout (1975), p. 26),

\[ A_1^2 - B_1 = o(\sqrt{\log \log T}) \text{ a.s.} \]

Thus, $\mathcal{B}_1 - A_2^2 < T^{-1/3}$ for $T$ large enough. So $\mathcal{F}_k = A_1$ for $T$ large enough. This proves the consistency when $\mathcal{F}_k = A_1$.

2. If $Pr(\mathcal{F}_1 = 1) > 1$, i.e. $\mathcal{B}_1 > 0$, then

\[ \lim_{T \to \infty} A_1^2 = (E_T^2) (T^{-\infty}) \text{ a.s.} \]

\[ \lim_{T \to \infty} A_1^2 = (E_T^2) (T^{-\infty}) \text{ a.s.} \]

using the assumption $\text{Var}(T) > 0$. So $\mathcal{B}_1 - A_2^2 < T^{-1/3}$ for
T large enough. Since $G(1Y2) = 0$ (T-- and C1(A12,0)
Finally, since list
and $A2$ is
bounded away from (and below) $Ly_2$ for $T$
important to
note that with the use of Equations (3.5) - (3.12) we can
show that $EY^2, EY^3, EY^4, EY^5, \text{and } EY^6$ can be
consistently estimated when $\sigma_1 > 0$.

To approach question 02, we derive

Proposition 3.2

Assume that $EY^2 = \Pr((-1) < 1, \sigma_2 = \mu(\eta, \sigma^2)$ and
$F_\eta = \eta(\sigma_2^2)$ where $F_\eta, \eta, \sigma^2$ and $\sigma_2$ are unknown. Then
for all $k \geq 1$ there exist $\nu(k) (x_1, T)$ such that

$$\lim_{T \to \infty} E(\nu(k) (x_1, T) - \nu_n^2) = E(E(\nu_k (x_1, T) - \nu_n^2)$$

uniformly for $n=1, \ldots, T-c(T)$

where $c(T)$ is an arbitrary positive integer valued,
increasing unbounded function.

Proof of Proposition 3.2

From Proposition 3.1 there exist consistent estimates
$\hat{\alpha}_k (x_1, T)$, $\sigma(\eta, \Omega), \sigma_2 (x_1, T)$, $\sigma_2(\eta, \Omega)$ and
$\sigma_2 (x_1, T)$ of $F_\eta, \sigma_2$ and $\sigma_2$, respectively. In
particular, $\hat{\alpha}$ can be chosen to be
$\frac{1}{I_{(1)} x_{(1)}}$ by stationarity,

$$\lim_{T \to \infty} E(\nu(k) (x, T) - \nu_n^2) = E(E(\nu_k (x_1, T) - \nu_n^2)$$
Remark 2:
Although in Proposition 3.2 $F_Y$ and $F_z$ are required to be Gaussian, the same result can be proved if they belong to (regular) parametric families of distributions whose parameters can be estimated consistently.

4. An Empirical Bayes Estimate Using Model A with Unknown Parameters

In general, a step-function signal can be either deterministic or stochastic and therefore Model A, or even the general model, can fail to be satisfied. Why then should we consider these models? The basic idea is that it is hoped the unknown signal would resemble a "typical" realisation of these models with properly assigned parameters or distributions. Indeed, this is a possible interpretation of the empirical Bayes idea. The most famous example is the James-Stein estimate which shows some superiority to the classical estimate of the mean of a multivariate normal distribution.

It is almost impossible to produce a sensible estimate of the signal without any information about the structure of the signal and/or the noise. Hence, our first assumption is that the noise is Gaussian white noise. One main reason to have the Gaussian assumption is that it is hard to distinguish outliers from jumps if the noise has a heavy tailed distribution. Furthermore, if the step-function...
Siqnal has many jumps, the noise variance cannot be well estimated. Indeed, the noise variance in Model A is not identifiable without further information. For instance, the observation process \( (X_n) \) is i.i.d. \( N(0,1) \) when \( (p,\sigma,\sigma_c) = (1,0,1,0) \) or \( (p,0,0,1) \). So, we make the second assumption that the rate of jump in the signal is at most \( p_0 \) where \( p_0 \) is a specified number between 0 and 1.

As the next step in generalizing our estimation procedure, let us assume that Model A applies with unknown parameters \( p,\sigma,\sigma_c \) and apply maximum likelihood to estimate these parameters. To be more precise, we estimate the signal \( u_n \) as follows. First, fit Model A to the observations \( X_1(1,2,\ldots,T) \) by finding the maximum likelihood estimates (MLE) \( \hat{p}, \hat{\sigma}, \hat{\sigma}_c \) with the constraint that \( p \leq p_0 \). Next, estimate \( u_n \) by

\[
(4.1) \quad \hat{u}_n^{EB} = \mathbb{E}(p,\sigma,\sigma_c) (u_n | X_1^T)
\]

where \( EB \) stands for empirical Bayes.

Since the MLE satisfy, for constants \( a \neq 0, c \),

\[
\hat{p}(ax_1^c,\ldots,ax_T^c) = \hat{p}(x_1^c,\ldots,x_T^c)
\]

(4.2) \( \hat{\sigma}(ax_1^c,\ldots,ax_T^c) = a \hat{\sigma}(x_1^c,\ldots,x_T^c) + c \)

and since Model A is time reversible, we have Proposition 4.1

The empirical Bayes estimator of \( u_n, \hat{u}_n^{EB} \), is translation invariant, scale invariant and time reversible. That is,

\[
\hat{u}_n^{EB}(ax_1^c,\ldots,ax_T^c) = a \hat{u}_n^{EB}(x_1^c,\ldots,x_T^c) + c
\]

\[
\hat{u}_n^{EB}(x_1,\ldots,x_T) = \hat{u}_{T+1-n}^{EB}(x_T,\ldots,x_1)
\]

The computation of the MLE can be very time-consuming. A naive method may require \( O(T^2) \) operations to compute the likelihood for each quadruple \( (p,0,0,1) \). We present in Proposition 4.2 a representation of the likelihood function which reduces the number of operations to the order of \( T \). Since the log likelihood \( L(p,\sigma,\sigma_c) \) satisfies

\[
\hat{\sigma}(ax_1^c,\ldots,ax_T^c) = a \hat{\sigma}(x_1^c,\ldots,x_T^c)
\]
\begin{align}
(4.3) \quad & L(p, \theta, \sigma, X_1^T) = L(p, \theta, \sigma, 1; X'_{-1}) = -T \log \sigma \\
& \text{where } X'_{-1} = (X_{-1}, \theta_1) \text{, we need only consider } L(p, \theta, \sigma, X'_{-1}) \text{.} \\
& \text{Let } S_n = 0 \text{ and } S_n = \frac{1}{T} \Sigma_{k=1}^{T} X_k \text{ for } 1 \leq n \leq T. \\
\text{Proposition 4.2:} \\
& L(p, \theta, \sigma, X_1^T) = \log f_{X_1}(x_1) + \sum_{n=1}^{T-1} \log f_{X_n}(x_n | x_{n-1}^T) \\
& \text{where} \\
& L(X_1) = N(0, \sigma^2 + 1), \\
& \text{and } A^{(n)} \text{ are defined in Proposition 4.2 of Yao (1982).} \\
\text{Proof of Proposition 4.2:} \\
& \text{We need only derive (4.4). However, this is a simple} \\
& \text{consequence of Proposition 4.2 of Yao (1982) and the} \\
& \text{following identity.} \\
& L(K_{n+1} | x_1^T) = L(n_{n+1} + t_{n+1} | x_1^T) \\
& = L(n_{n+1} | x_1^T) \Theta (0.1) \\
& = [(1-p)L_n(x_1^T) + p N(0, \sigma^2)] \Theta (0.1) \\
& = (1-p)L_n(x_1^T) \Theta (0.1) + p N(0, \sigma^2 + 1) \\
& \text{where } I_1 \Theta I_2 \text{ is the convolution of law } I_1 \text{ with } I_2. \\
5. \text{ An Approximation to the Likelihood and the Pseudo MLE} \\
& \text{Even though Proposition 4.2 suggests a way to compute} \\
& \text{the likelihood with } O(T^2) \text{ operations, it is still time-} \\
& \text{consuming to compute the MLE without further reduction in} \\
& \text{computation. Therefore it is desired to find a more} \\
& \text{efficient way to approximate the likelihood. We will make} \\
& \text{use of an idea of Harrison and Stevens (1976) to develop} \\
& \text{an approximation procedure which reduces the number of} \\
& \text{operations to the order of } T. \text{ This idea has been used and} \\
& \text{justified in Yao (1982).} \\
& \text{We approximate } L(K_{n+1} | x_1^T) \text{ as follows. Again, assume} \\
& \theta = 0 \text{ and } \sigma = 1 \text{ for simplicity. In Section 5 of Yao} \\
& (1982), N(\eta_n, \tau_2^2) \text{ is introduced to approximate } L(n_{n+1} | x_1^T) \\
& \text{where } \eta_n \text{ and } \tau_2^2 \text{ are defined recursively. Since} \\
& L(K_{n+1} | x_1^T) = (1-p)N_{n+1}(x_1^T) \Theta (0.1) + p N(0, \sigma^2 + 1) \\
& \text{we are naturally led to approximate } L(K_{n+1} | x_1^T) \text{ by} \\
& (1-p)N_{n+1}(x_1^T) \Theta (0.1) + p N(0, \sigma^2 + 1). \\
& \text{Now we can approximate the log likelihood } L(p, \theta, \sigma, X_1^T) \\
& \text{by use of Proposition 4.2 and the above approximation and} \\
& \text{denote this approximate log likelihood by } \Sigma(p, \theta, \sigma, X_1^T). \\
& \text{It should be noted that this approximation is exact}
when \( p \) is 0 or 1, for Model A is a Gaussian system when \( p \) is 0 or 1. Now we propose to measure this approximation in terms of the Kullback information between \( \exp(L) \) and \( \exp(L) \) under Model A. More precisely, we will treat

\[
(5.1) \quad I_k(p, \theta, \sigma_x^2) = \frac{\partial}{\partial p} \{ \exp(L(p, \theta, \sigma_x^2)) - L(p, \theta, \sigma_x^2) \}
\]

as a measure of how well \( L \) is approximated by \( L \). Note that

\[
(5.2) \quad I_k(p, \theta, \sigma_x^2) = I_k(p, 0, 0/\sigma_x^2, 1)
\]

We considered 63 cases where

\[
p(0.02, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8), \sigma(0.5, 1, 2, 3, 4, 5, 7, 10, 15),
\theta = 0 \text{ and } \sigma_x = 1.
\]

The \( I_k \) were estimated by use of simulation with a computer (HP 3000) where 400 samples of size \( T = 20 \) were generated for each case. The results are presented in Table 5.1.

According to Table 5.1, \( E(L-L) < 0.14 \), and \( SD(L-L) < 0.48 \).

Here \( SD(Y) \) is the standard deviation of random variable \( Y \). So,

-1.44 \( \leq E(L-L) - 3 \cdot SD(L-L) < E(L-L) + 3 \cdot SD(L-L) \leq 1.58
\]

The probability that the likelihood ratio \( \exp(L-L) \) satisfies

0.24 \( \leq \exp(-1.44) \cdot \exp(L-L) \cdot \exp(1.58) \) = 4.85

is very high in the worst case under Model A. This suggests that the approximation will yield reasonably good results.

We shall define the pseudo MLE \( \hat{p}^{\text{PSE}}, \hat{\theta}^{\text{PSE}}, \hat{\sigma}^{\text{PSE}} \) as the values of the parameters which maximize \( L \) subject to \( p \leq p_0 \). Then we estimate \( \hat{p}^{\text{PSE}} \) by

\[
(5.3) \quad \hat{p}^{\text{PSE}} = \text{argmax}_{p \leq p_0} \{ L(p, \hat{\theta}^{\text{PSE}}, \hat{\sigma}^{\text{PSE}}) \} (u_X, x_T^X).
\]

6. Simulation on Empirical Bayes Estimators

In the last section, we have introduced, for the sake of computer, \( \hat{u}_n \) which is an approximation to \( \hat{u}_n \). In order to evaluate the performance of \( \hat{u}_n \), we carried out the following computer simulations on an HP 3000.

We considered 21 deterministic signal sequences \( u_{n(1)} \) of length \( T = 20 \) (1, 2, 3, 4, 5, 7, 10, 15). For each signal sequence, we generated 100 samples of Gaussian white noise of variance 1.

In defining \( \hat{u}^{\text{EB}} \), we estimated the parameters of Model A by use of pseudo maximum likelihood. It is interesting to see how well the method of moments can be compared to the pseudo maximum likelihood method. It is also interesting to see how much the additional information \( \sigma_c = 1 \) can contribute to estimating \( \hat{u}_n \).
Hence, we considered the following four estimators of \( \theta_n \).

(i) **Estimator 1** \( \bar{\theta}_n \), \( \beta_0 = 0.2 \)

(ii) **Estimator 2** – This is defined in the same way as Estimator 1 except with one more constraint \( \gamma_e = 1 \) in the pseudo maximum likelihood estimation of the parameters.

(iii) **Estimator 3** – \( \hat{\theta}_n \), \( \beta_1 = \bar{\theta} \) (the sample mean), \( \gamma_1 = \max(\gamma_2,0) \), \( \gamma_2 = \max(\gamma_3,0) \) and \( \beta_2,\beta_2 \) satisfy

\[
\sum_{i=1}^{T-1} \bar{X}_i T = \bar{X}^2 + \gamma_2^2 + \gamma_2^2
\]

\[
\sum_{i=1}^{T-2} \bar{X}_i T = \bar{X}^2 + (1-\gamma_2)\gamma_2^2
\]

(iv) **Estimator 4** – \( \bar{\theta}_n \), \( \beta_3 = \max(\beta_4,0.2),0 \), \( \beta_4 = \bar{\theta}_n \) and \( \beta_5,\beta_5 \) satisfy

\[
\sum_{i=1}^{T-1} \bar{X}_i T = \bar{X}^2 + \gamma_2^2 + 1
\]

\[
\sum_{i=1}^{T-2} \bar{X}_i T = \bar{X}^2 + (1-\gamma_2)\gamma_2^2
\]

We use the average of mean squared errors (AMSE) as the criterion. The simulation results are presented in Table 6.1 where we also present the mean and standard deviation of \( \bar{\theta}_n \), the pseudo MLZ of \( \theta_e \).

**Note:**

All the four estimators have one common property. That is, they first estimate \( \beta_p,\beta_e,\beta_e \) and then estimate \( \bar{\theta}_n \) by the corresponding Bayes estimate \( \hat{\theta}_n \) \((p,\beta,\beta,\beta) \), \((p,\beta,\beta,\beta) \), \((p,\beta,\beta,\beta) \), \((p,\beta,\beta,\beta) \).

**Remarks:** (Based on Table 6.1)

1. Roughly speaking, when the number of jumps increases, the AMSE of \( \bar{\theta}_n \) increases. When the size of jumps increases, the AMSE of \( \bar{\theta}_n \) first increases and then decreases. For when the size of jumps is moderate (i.e.
compatible with the noise) it is hard to tell where jumps take place and to take appropriate action. This property is similar to that of the Bayes estimator. (See Remark 1 of Section 7 in Yao (1982)).

(2) Estimator 1 \( \hat{\theta}_n \) is better than Estimator 2. This implies that the method of pseudo maximum likelihood is significantly better than the method of moments in finding suitable parameter values.

(3) Estimator 1 is just slightly worse than Estimator 2. So the information about the noise variance is not very important for estimating the signal unless the rate of change in the signal is high. In that case, it is hard to estimate \( \sigma \) well.

(4) The empirical Bayes estimator, \( \tilde{\theta}_n \), is robust against the signals' behavior. However, it is not known how to deal with cases involving non-Gaussian noise which may introduce outliers under the veil of jumps.

(5) When the prior information, the rate of change \( \leq P_0 \), is not correct, \( \tilde{\theta}_n \) may be misleading, although our limited simulations do not indicate so.

(6) It is interesting that \( \tilde{\sigma}_n \), the pseudo MLE of \( \sigma \), estimates \( \sigma \) well with small bias. This is essentially due to the information \( P \leq P_0 \).

References


Appendix

Proof of (3.1) and (3.2) in Proposition 3.1

It is not difficult to see by applying the Gauss-Koksma strong law of large numbers (Philipp and Stout (1975), Appendix 1) that we need only show

\[
\lim_{T \to \infty} \frac{1}{n} \sum_{n=1}^{T} u_n^2 = E u_n^2 \text{ a.s.} \quad 1 \leq a \leq 4
\]

and

\[
\lim_{T \to \infty} \frac{1}{n} \sum_{n=1}^{T} v_n^2 = E v_n^2 \text{ a.s.} \quad 1 \leq a \leq 4
\]

For example, since

\[
\frac{1}{n} \sum_{n=1}^{T} \sum_{n=1}^{T} (u_n^2 - E v_n^2) = 0 \quad \text{a.s. by the Gauss-Koksma strong law of large numbers,}
\]

and

\[
\lim_{T \to \infty} \frac{1}{n} \sum_{n=1}^{T} u_n^2 = E u_n^2 \text{ a.s.}
\]

if

\[
\lim_{T \to \infty} \frac{1}{n} \sum_{n=1}^{T} v_n^2 = E v_n^2 \text{ a.s.}
\]

In the following we only treat one case, i.e.

\[
\lim_{T \to \infty} \frac{1}{n} \sum_{n=1}^{T} u_n^2 = E u_n^2 \text{ a.s. for } t \geq 1
\]

and the rest can be established similarly.

\[
U_T = 0 \quad \text{if there exists } k \text{ such that } n \leq k \leq n+1
\]

\[
U_T = \frac{1}{n} \sum_{n=1}^{T} u_n^2 - E u_n^2 \quad \text{otherwise}
\]

\[
V_T = \frac{1}{n} \sum_{n=1}^{T} v_n^2 - (E v_n^2) \quad \text{if there exists } k \text{ such that } n \leq k \leq n+1
\]

\[
= 0 \quad \text{otherwise}
\]

\[
(U_T) \text{ and } (V_T) \text{ are stationary.}
\]

\[
E \left( \frac{1}{n} \sum_{n=1}^{T} u_n^2 \right) = \frac{n}{E u_n^2} + 2 \sum_{k=1}^{n} (n-k) E u_n^2
\]

\[
= \frac{n}{E u_n^2} + 2 \sum_{k=1}^{n} (n-k) E u_n^2 
\]

\[
\leq n \left( E u_n^2 + 2 E v_n^2 \right)
\]
\[ E \left( \frac{\sum_{i=1}^{N} V_i^2}{n} \right)^2 = N E \left( V_i^2 \right)^2 + 2 \sum_{k=1}^{N-1} (N-k) E V_k^2 \]

\[ \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} u_n \right) = 0 \text{ a.s.} \]

So,

\[ \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} u_n^2 \right) \leq \frac{1}{T} \left( \sum_{n=1}^{T} \left( u_n \right)^2 \right) \leq \frac{1}{T} \left( \sum_{n=1}^{T} \left( E V_k \right)^2 \right) \]

\[ \sum_{k=1}^{N-1} (N-k) E V_k^2 \]

where

\[ c = \max \left( k : \eta_k < T \right) \]

Now, by the strong law of large numbers.

\[ \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} u_n \right) = 0 \text{ a.s.} \]

\[ \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} \left( u_n \right)^2 \right) = 0 \text{ a.s.} \]

\[ \sum_{k=1}^{N-1} (N-k) E V_k^2 \]

\[ \sum_{k=1}^{N-1} (N-k) E V_k^2 \leq \left( \sum_{k=1}^{N-1} (N-k) \right) E (\sum_{k=1}^{N-1} V_k^2) \]

\[ \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} \left( u_n \right)^2 \right) \leq \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} \left( E V_k \right)^2 \right) \leq \lim_{T \to \infty} \frac{1}{T} \left( \sum_{n=1}^{T} \left( u_n \right)^2 \right) \]

Therefore, by the strong law of large numbers.
Table 5.1 Kullback Information between $\exp(L)$ and $\exp(L)$
under Model A

The upper and lower figures are $E(L-L)$ and $SD(L-L)$, respectively.

<table>
<thead>
<tr>
<th>n</th>
<th>.02</th>
<th>.05</th>
<th>.2</th>
<th>.4</th>
<th>.6</th>
<th>.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>.022</td>
<td>.012</td>
<td>.001</td>
<td>.001</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.174</td>
<td>.106</td>
<td>.072</td>
<td>.038</td>
<td>.012</td>
<td>.004</td>
</tr>
<tr>
<td>1</td>
<td>.055</td>
<td>.044</td>
<td>.033</td>
<td>.006</td>
<td>.003</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>.324</td>
<td>.302</td>
<td>.255</td>
<td>.151</td>
<td>.074</td>
<td>.020</td>
</tr>
<tr>
<td>2</td>
<td>.099</td>
<td>.117</td>
<td>.040</td>
<td>.021</td>
<td>.012</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.427</td>
<td>.478</td>
<td>.360</td>
<td>.265</td>
<td>.109</td>
<td>.046</td>
</tr>
<tr>
<td>3</td>
<td>.105</td>
<td>.121</td>
<td>.092</td>
<td>.045</td>
<td>.008</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>.620</td>
<td>.627</td>
<td>.397</td>
<td>.282</td>
<td>.124</td>
<td>.054</td>
</tr>
<tr>
<td>4</td>
<td>.103</td>
<td>.136</td>
<td>.075</td>
<td>.052</td>
<td>.009</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>.418</td>
<td>.448</td>
<td>.409</td>
<td>.325</td>
<td>.141</td>
<td>.062</td>
</tr>
<tr>
<td>5</td>
<td>.090</td>
<td>.101</td>
<td>.093</td>
<td>.057</td>
<td>.012</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>.426</td>
<td>.419</td>
<td>.385</td>
<td>.302</td>
<td>.157</td>
<td>.075</td>
</tr>
<tr>
<td>7</td>
<td>.055</td>
<td>.071</td>
<td>.089</td>
<td>.027</td>
<td>.021</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.376</td>
<td>.369</td>
<td>.370</td>
<td>.298</td>
<td>.155</td>
<td>.066</td>
</tr>
<tr>
<td>10</td>
<td>.057</td>
<td>.069</td>
<td>.101</td>
<td>.043</td>
<td>.029</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.345</td>
<td>.360</td>
<td>.384</td>
<td>.308</td>
<td>.191</td>
<td>.072</td>
</tr>
<tr>
<td>15</td>
<td>.051</td>
<td>.042</td>
<td>.058</td>
<td>.029</td>
<td>.012</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>.352</td>
<td>.374</td>
<td>.414</td>
<td>.304</td>
<td>.144</td>
<td>.076</td>
</tr>
</tbody>
</table>
### Table 6.1: The ANSE of the Estimators over 100 Samples

<table>
<thead>
<tr>
<th>Signal Successive Points of Change</th>
<th>Est. 1</th>
<th>Est. 2</th>
<th>Est. 3</th>
<th>Est. 4</th>
<th>Given C.P.</th>
<th>$\text{SE}(\hat{\theta}_e)^1$</th>
<th>$\text{SE}(\hat{\theta}_e)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0</td>
<td>0.071 (0.012)</td>
<td>0.059 (0.008)</td>
<td>0.124 (0.025)</td>
<td>0.067 (0.008)</td>
<td>0.943</td>
<td>0.166</td>
<td></td>
</tr>
<tr>
<td>2 0.1</td>
<td>0.228 (0.012)</td>
<td>0.216 (0.011)</td>
<td>0.352 (0.024)</td>
<td>0.251 (0.009)</td>
<td>1.949</td>
<td>1.196</td>
<td></td>
</tr>
<tr>
<td>3 0.3</td>
<td>0.254 (0.018)</td>
<td>0.241 (0.017)</td>
<td>0.670 (0.058)</td>
<td>0.395 (0.033)</td>
<td>0.930</td>
<td>0.180</td>
<td></td>
</tr>
<tr>
<td>4 0.5</td>
<td>0.187 (0.019)</td>
<td>0.185 (0.019)</td>
<td>0.486 (0.050)</td>
<td>0.199 (0.018)</td>
<td>1.970</td>
<td>1.165</td>
<td></td>
</tr>
<tr>
<td>5 0.1</td>
<td>0.204 (0.008)</td>
<td>0.197 (0.008)</td>
<td>0.314 (0.027)</td>
<td>0.197 (0.007)</td>
<td>0.961</td>
<td>0.162</td>
<td></td>
</tr>
<tr>
<td>6 0.3</td>
<td>0.301 (0.024)</td>
<td>0.272 (0.019)</td>
<td>0.908 (0.054)</td>
<td>0.385 (0.036)</td>
<td>1.944</td>
<td>2.041</td>
<td></td>
</tr>
<tr>
<td>7 0.2, 4</td>
<td>0.370 (0.017)</td>
<td>0.361 (0.016)</td>
<td>0.802 (0.054)</td>
<td>0.630 (0.040)</td>
<td>0.936</td>
<td>0.220</td>
<td></td>
</tr>
<tr>
<td>8 0.3, 0</td>
<td>0.407 (0.030)</td>
<td>0.371 (0.025)</td>
<td>0.947 (0.071)</td>
<td>0.642 (0.060)</td>
<td>0.916</td>
<td>0.258</td>
<td></td>
</tr>
<tr>
<td>9 0.3, 10</td>
<td>0.395 (0.031)</td>
<td>0.380 (0.027)</td>
<td>0.952 (0.078)</td>
<td>0.700 (0.067)</td>
<td>0.913</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td>10 0.4, 0</td>
<td>0.356 (0.022)</td>
<td>0.350 (0.022)</td>
<td>0.793 (0.070)</td>
<td>0.440 (0.031)</td>
<td>0.944</td>
<td>0.198</td>
<td></td>
</tr>
</tbody>
</table>

* The number in parentheses next to an entry is the estimated standard error for that entry.

* This column is the ANSE of the estimator using the averages of the data points between successive time points of change.

* The estimated standard error of the estimated $\hat{\theta}_e$ is $\text{SE}(\hat{\theta}_e)/10$.

### Table 6.1 - Continued

<table>
<thead>
<tr>
<th>Signal Successive Points of Change</th>
<th>Est. 1</th>
<th>Est. 2</th>
<th>Est. 3</th>
<th>Est. 4</th>
<th>Given C.P.</th>
<th>$\text{SE}(\hat{\theta}_e)^1$</th>
<th>$\text{SE}(\hat{\theta}_e)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 0.4, 6 6.13</td>
<td>0.305 (0.020)</td>
<td>0.302 (0.020)</td>
<td>0.600 (0.043)</td>
<td>0.375 (0.026)</td>
<td>0.964</td>
<td>0.189</td>
<td></td>
</tr>
<tr>
<td>12 0.1, 2, 3 6.10, 16</td>
<td>0.348 (0.023)</td>
<td>0.308 (0.017)</td>
<td>0.729 (0.041)</td>
<td>0.493 (0.032)</td>
<td>0.997</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td>13 0.3, 6, 9 6.10, 16</td>
<td>0.477 (0.027)</td>
<td>0.466 (0.025)</td>
<td>0.764 (0.051)</td>
<td>0.504 (0.031)</td>
<td>0.947</td>
<td>0.254</td>
<td></td>
</tr>
<tr>
<td>14 0.5, 10, 15 6.10, 16</td>
<td>0.321 (0.032)</td>
<td>0.314 (0.031)</td>
<td>0.728 (0.046)</td>
<td>0.305 (0.032)</td>
<td>0.908</td>
<td>0.185</td>
<td></td>
</tr>
<tr>
<td>15 0.1, 0.1, 6.10, 16</td>
<td>0.280 (0.007)</td>
<td>0.265 (0.006)</td>
<td>0.365 (0.026)</td>
<td>0.259 (0.008)</td>
<td>0.918</td>
<td>0.180</td>
<td></td>
</tr>
<tr>
<td>16 0.3, 0, 6.10, 16</td>
<td>0.335 (0.038)</td>
<td>0.334 (0.022)</td>
<td>0.994 (0.052)</td>
<td>0.693 (0.058)</td>
<td>0.890</td>
<td>0.314</td>
<td></td>
</tr>
<tr>
<td>17 0.1, 3, 6, 3.7, 12, 10</td>
<td>0.426 (0.018)</td>
<td>0.415 (0.017)</td>
<td>0.907 (0.081)</td>
<td>0.578 (0.038)</td>
<td>1.016</td>
<td>0.238</td>
<td></td>
</tr>
<tr>
<td>18 0.5, 3, 6, 3.7, 12, 10</td>
<td>0.378 (0.022)</td>
<td>0.366 (0.023)</td>
<td>0.994 (0.036)</td>
<td>0.364 (0.022)</td>
<td>0.961</td>
<td>0.200</td>
<td></td>
</tr>
<tr>
<td>19 0.3, 3, 6, 4.2, 3.7, 12, 10</td>
<td>0.369 (0.021)</td>
<td>0.352 (0.022)</td>
<td>0.875 (0.063)</td>
<td>0.729 (0.042)</td>
<td>0.997</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td>20 0.3, 3, 6, 4.2, 3.7, 12, 10</td>
<td>0.369 (0.021)</td>
<td>0.352 (0.022)</td>
<td>0.875 (0.063)</td>
<td>0.729 (0.042)</td>
<td>0.997</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td>21 0.3, 5, 6, 7, 8, 9, 10, 11</td>
<td>0.394 (0.024)</td>
<td>0.390 (0.023)</td>
<td>1.037 (0.045)</td>
<td>1.190 (0.035)</td>
<td>1.230</td>
<td>0.280</td>
<td></td>
</tr>
</tbody>
</table>

1: Signal 19 is the following. $u_n = 0.5(n-1), 1 \leq n \leq 20$

2: Signal 20 is the following. $u_n = n-1, 1 \leq n \leq 11; u_n = 21 - n, 12 \leq n \leq 20$

3: Signal 21 is the following. $u_n = 10 - 0.1 (n-1)^2, 1 \leq n \leq 20$