Inference About the Change-Point in a Sequence
Of Random Variables: A Selection Approach*

by

TaChen Liang and S. Panchapakesan
Southern Illinois University at Carbondale

Technical Report #86-24

Department of Statistics
Purdue University

June 1986

* This research is supported partly by the Office of Naval Research Contract N00014-84-
C-0167 at Purdue University. Reproduction in whole or part is permitted for any purpose
of the U.S. Government.
Inference About the Change-Point in a Sequence
Of Random Variables: A Selection Approach*

by

TaChen Liang and S. Panchapakesan
Southern Illinois University at Carbondale

Abstract

A selection approach is made for making inference about the point in a sequence of random variables at which the underlying distribution changes. Three selection rules are derived: a Bayes rule, a gamma-minimax rule, and a locally optimal rule. While the first two rules select one point as the change-point, the third selects a subset (possibly empty) of points.

1. Introduction

Let $X_1, \ldots, X_k$ be a sequence of independent random variables where $X_i$ has the probability density function $p(x|\theta_i), i = 1, \ldots, k$. The $\theta_i$ are unknown. It is assumed that $\theta_1 \leq \ldots \leq \theta_k$ with exactly one shift; in other words, there is an unknown integer $m(1 \leq m \leq k - 1)$ such that $\theta_1 = \ldots = \theta_m < \theta_{m+1} = \ldots = \theta_k$. Then $m$ is called the change-point. Problems of inference about a change-point have been investigated by several authors using different approaches. Page (1955, 1957) and Bhattacharyya and Johnson (1968) have considered testing for a shift using nonparametric methods. Hinkley (1970) used asymptotic arguments based on maximum likelihood estimates and likelihood ratio tests. The problem has also been considered within a Bayesian framework by Chernoff and Zacks (1964), Kander and Zacks (1966), Mustafi (1968), Broemeling (1972, 1974), Sen and Srivastava (1973), Smith (1975), and Raftery and Akman (1986). Broemeling and Magalit (1975) have discussed parametric tests for a shift. Worsley (1986) has investigated confidence regions and tests for a change-point using maximum likelihood methods. Inference about change-points has been studied by econometrists under the general context of structural changes in a model. Some recent papers using a Bayesian approach are Booth and Smith (1982), Diaz (1982), Holbert (1982), and Salazar (1982), which appeared in a special issue of the Journal of Econometrics (Volume 19, 1982) edited by Lyle Broemeling. The emphasis and the objectives of the present paper are along the lines of the selection approach of multiple decision problems.

For selecting the true change-point we derive three different selection rules. Section 2 deals with a Bayes rule under a fairly general loss function assuming that the prior

* This research is supported partly by the Office of Naval Research Contract N00014-84-C-0167 at Purdue University. Reproduction in whole or part is permitted for any purpose of the U.S. Government.

Key Words: change-point inference, selection, Bayes rule, gamma-minimax, locally optimal.

distributions of the location of the change-point and the amount of change are independent. 
A gamma-minimax rule is derived in Section 3. These two rules select one point as the 
change-point, while the locally optimal rule derived in Section 4 selects a subset (possibly 
empty) of the $k - 1$ points.

2. Bayes Procedure

Let $\Omega_i = \{ \theta = (\theta_1, \ldots, \theta_k) | \theta_1 = \ldots = \theta_i < \theta_{i+1} = \ldots = \theta_k \}, i = 1, \ldots, k - 1$, and 
$\Omega = \bigcup_{i=1}^{k-1} \Omega_i$. If $\theta \in \Omega_i$, then $i$ is the change-point. Let $\tau_{ij} = \tau(\theta_i, \theta_j)$ be a measure of 
separation of $\theta_j$ from $\theta_i$, defined so that it is increasing in $\theta_j$ for a fixed $\theta_i$ and $\tau_{ij} = \tau^*$ for all $i$. For example, if $\theta$ is a location parameter, a natural choice is $\tau_{ij} = \theta_j - \theta_i$ with 
$\tau^* = 0$. Let $T_{ij}$ be a statistic based on $X_i$ and $X_j$, suitably defined as a sample measure 
of the separation $\tau_{ij}$.

Let $Y_i = T_{i,i+1}, i = 1, \ldots, k - 1$. We assume that $Y = (Y_1, \ldots, Y_{k-1})$ has a density $f(y|\tau)$ depending on the parameter $\tau = (\tau_{12}, \ldots, \tau_{k-1,k})$. For convenience, let $\alpha_i = \tau_{i,i+1}, i = 1, \ldots, k - 1$. For $\theta \in \Omega_i$, we have correspondingly $\alpha = (\alpha_1, \ldots, \alpha_{k-1})$ such that 
$\alpha_i = \tau > \tau^* = \alpha_j$ for all $j \neq i$, where $\tau$ is the amount of shift. Thus $\Omega_i = \{ \alpha | \alpha_i > \tau^* = \alpha_j 	ext{ for all } j \neq i \}$. In this case, we write $f(y|\alpha_i(\tau))$ for $f(y|\tau)$.

Now, our action space is $A = \{1, 2, \ldots, k - 1\}$. Action $i$ corresponds to the decision 
that $i$ is the change-point. For given $\alpha \in \Omega_m$ and action $a$, the associated loss function 
is defined by

$$L(\alpha, a) = \begin{cases} 
\sum_{i=a}^{m-1} L_{mi}^{(1)}(\alpha) & \text{if } a < m, \\
\sum_{j=m+1}^{a} L_{mj}^{(2)}(\alpha) & \text{if } a > m, \\
0 & \text{if } a = m, 
\end{cases} \quad (2.1)$$

where

$$L_{mi}^{(1)}(\alpha) \begin{cases} > 0 & \text{and is nonincreasing in } i \text{ for } i < m, \\
= 0 & \text{for } i \geq m, \end{cases} \quad (2.2)$$

and

$$L_{mj}^{(2)}(\alpha) \begin{cases} > 0 & \text{and is nondecreasing in } j \text{ for } j > m, \\
= 0 & \text{for } j \leq m. \end{cases} \quad (2.3)$$

Let $p(m) = \Pr\{ \alpha \in \Omega_m \}, m = 1, \ldots, k - 1$, and let $g(\tau)$ be the prior density of the 
amount of shift with support on $(\tau^*, \infty)$. It is assumed that the two prior distributions 
are independent.

A decision rule $\delta = (\delta_1, \ldots, \delta_{k-1})$ is a measurable mapping from $Y$, the sample space 
of $Y$, to $[0, 1]^{k-1}$ such that $0 \leq \delta_i(y) \leq 1$ and $\sum_{i=1}^{k-1} \delta_i(y) = 1$ for each $y \in Y$. The value 
of $\delta_i(y)$ is the probability of taking action $i$ given the observation $y$. 

2
Let \( r(\hat{\delta}|g, p) \) denote the Bayes risk associated with the rule \( \hat{\delta} \), where \( p = (p(1), \ldots, p(k-1)) \). Then

\[
r(\hat{\delta}|g, p) = \sum_{m=1}^{k-1} p(m) \int_{\Omega_m} \int_y \sum_{a=1}^{k-1} L(\varphi_m(r), a) \delta_a(y) f(y|\varphi_m(r)) g(r) \, dy \, dr.
\] (2.4)

Now, let

\[
f_m(y) = \int_r^{\infty} f(y|\varphi_m(r)) g(r) \, dr
\]
\[
g_m(r|y) = f(y|\varphi_m(r)) g(r)/f_m(y)
\]
\[
f(y) = \sum_{m=1}^{k-1} p(m) f_m(y)
\]
\[
h_m(y) = p(m) f_m(y)/f(y)
\]
\[
R(a|y) = \sum_{m=1}^{k-1} h_m(y) \int_r^{\infty} L(\varphi_m(r), a) g_m(r|y) \, dr.
\] (2.5)

By the usual interchanging of summations and integrals, it can be easily seen that

\[
r(\hat{\delta}|g, p) = \sum_{a=1}^{k-1} \int_y \delta_a(y) R(a|y) f(y) \, dy.
\] (2.6)

Then \( \hat{\delta} = (\delta_1, \ldots, \delta_{k-1}) \) is a Bayes rule if \( \sum_{a \in A(y)} \delta_a(y) = 1 \), where

\[
A(y) = \{a|R(a|y) = \min_{1 \leq a' \leq k-1} R(a'|y)\}.
\] (2.7)

In order to obtain more insight for implementation of this Bayes rule, let

\[
S_{m,i}^{(1)}(y) = \int L_{m,i}^{(1)}(\varphi_m(r)) g_m(r|y) \, dr, \quad i < m
\]
\[
S_{m,j}^{(2)}(y) = \int L_{m,j}^{(2)}(\varphi_m(r)) g_m(r|y) \, dr, \quad j > m
\]
\[
D(a|y) = R(a+1|y) - R(a|y).
\] (2.8)

First, note that using (2.1), we can express \( R(a|y) \) as

\[
R(a|y) = \sum_{m=a+1}^{k-1} \sum_{i=a}^{m-1} h_m(y) S_{m,i}^{(1)}(y) + \sum_{m=1}^{a-1} \sum_{j=m+1}^{a} h_m(y) S_{m,j}^{(2)}(y).
\] (2.9)

Lemma 2.1. For fixed \( m \) and \( y \),
(a) \( S_{m,i}^{(1)}(y) \) is nonincreasing in \( i \) for \( i < m \), and
(b) \( S_{m,j}^{(2)}(y) \) is nondecreasing in \( j \) for \( j > m \).

Proof: Omitted as it is obvious.

Lemma 2.2. For given \( y \), \( D(a|y) \) is nondecreasing in \( a \).

Proof: From (2.9), it is easy to see that

\[
D(a|y) = \sum_{m=1}^{a} h(m|y) \, S_{m,a+1}^{(2)}(y) - \sum_{m=a+1}^{k-1} h(m|y) \, S_{m,a}^{(1)}(y).
\]

Therefore, \( D(a+1|y) - D(a|y) = \sum_{m=1}^{a} h(m|y) \left[ S_{m,a+2}^{(2)}(y) - S_{m,a+1}^{(2)}(y) \right] + h(a+1,y)S_{a+1,a+2}^{(2)}(y) + \sum_{m=a+2}^{k-1} h(m|y) \left[ S_{m,a}^{(1)}(y) - S_{m,a+1}^{(2)}(y) \right] + h(a+1,y)S_{a+1,a+1}^{(1)}(y). \)

Using Lemma 2.1 and the nonnegativity of \( h(m|y) \) and \( S_{m,i}^{(r)} \), \( r = 1, 2 \), it can now be seen that \( D(a+1|y) - D(a|y) \geq 0 \).  \( \Box \)

Based on Lemma 2.2, the Bayes rule \( \hat{a} \) can be expressed as follows:

Randomize your decision over the set \( B(y) = \{a|D(a|y) = 0\} \). If the set \( B(y) \) is vacuous, then choose action \( b \) where \( b \) is the largest \( a \) for which \( D(a|y) < 0 \). If such a \( b \) does not exist, choose action 1.

Remark 2.1. Because of the monotonicity of \( D(a|y) \) in \( a \), the set \( B(y) \) is either vacuous or it consists of consecutive members of the set \( \{1, 2, \ldots, k-1\} \). We can define a nonrandomized rule, by taking the action corresponding to the smallest member of \( B(y) \).

3. \( \Gamma \)-Minimax Rule

In this section, we assume a uniform prior for the change-point, i.e. \( p(m) = \frac{1}{k-1}, m = 1, \ldots, k-1 \). Further, the loss function is taken to be

\[
L(g_{m}(r), a) = |m - a|L \, I_{[\tau_{0}, \infty)}(r)
\]

where \( L > 0 \) is a known constant, \( \tau_{0} > r^{*} \), and \( I_{B} \) denotes the indicator function of the set \( B \).

\( \Gamma \)-minimax selection rules for selecting the best population has been considered in the literature; see Gupta and Huang (1977), for example. For \( 0 < \pi_{0} < 1 \), let

\[
\Gamma = \{ G|\int_{\tau_{0}}^{\infty} dG(r) \leq \pi_{0} \}
\]
A rule $\delta^o$ is a $\Gamma$-minimax rule if, for any other rule $\delta$, \[ \sup_{G \in \Gamma} r(\delta, G) \geq \sup_{G \in \Gamma} r(\delta^o, G). \]

We now state and prove a theorem which gives a $\Gamma$-minimax rule under certain condition.

**Theorem 3.1.** Suppose that there exists a $\tau' \geq \tau_0$ such that, for each $m, 1 \leq m \leq k - 1$,

\[
\sup_{\tau \geq \tau_0} \int \sum_{j=1}^{k-1} L(\varrho_m(\tau), j) \delta_j^o(y) f(y|\varrho_m(\tau)) \, dy
= \int \sum_{j=1}^{k-1} L(\varrho_m(\tau'), j) \delta_j^o(y) f(y|\varrho_m(\tau')) \, dy
\]

(3.3)

where

\[
\delta_j^o(y) = \begin{cases} \frac{1}{|M|} & \text{if } L_j(y) \leq \min_{1 \leq i \leq k-1} L_i(y) \\ 0 & \text{otherwise,} \end{cases}
\]

(3.4)

\[
L_i(y) = \sum_{j=1}^{k-1} |i - j| Lf(y|\varrho_j(r')),
\]

and $|M|$ is the cardinality of the set

\[
M = \{j | L_j(y) = \min_{1 \leq i \leq k-1} L_i(y)\}.
\]

Then $\delta^o = (\delta_1^o, \ldots, \delta_{k-1}^o)$ is a $\Gamma$-minimax rule.

**Proof:** Let $G_0$ be any distribution such that $G_0(\tau') = 1, G_0(\tau_0-) = G_0(\tau'-) = 1 - \pi_0$. 

5
Then $G_0 \in \Gamma$. Therefore, for any rule $\delta$,

$$
\sup_{G \in \Gamma} r(\delta|G) \geq r(\delta|G_0)
$$

$$
= \sum_{m=1}^{k-1} \frac{1}{k-1} \int_y \sum_{a=1}^{k-1} |m-a| L \delta_a(y) \pi_0 f(y|\alpha_m(r')) \, dy
$$

$$
= \sum_{a=1}^{k-1} \frac{1}{k-1} \int_y \pi_0 \delta_a(y) L_a(y) \, dy
$$

$$
\geq \sum_{a=1}^{k-1} \frac{1}{k-1} \int_y \pi_0 \delta_a(y) L_a(y) \, dy
$$

$$
= \sum_{m=1}^{k-1} \frac{1}{k-1} \int_y \sum_{a=1}^{k-1} |m-a| L \delta_a^o(y) \pi_0 f(y|\alpha_m(r')) \, dy
$$

$$
\geq \sum_{m=1}^{k-1} \frac{1}{k-1} \int_y \sum_{a=1}^{k-1} |m-a| L \delta_a^o(y) f(y|\alpha_m(r')) \, dy \, dG(r), \text{ for any } G \in \Gamma
$$

$$
\geq \sum_{m=1}^{k-1} \frac{1}{k-1} \int_y \sum_{a=1}^{k-1} |m-a| L \delta_a^o(y) f(y|\alpha_m(r)) \, dy \, dG(r), \text{ because of (3.3)}
$$

$$
= r(\delta^o, G), \text{ for any } G \in \Gamma.
$$

Therefore, $\sup_{G \in \Gamma} r(\delta|G) \geq \sup_{G \in \Gamma} r(\delta^o|G)$. \qed

Now, let

$$
D^*(a|y) = L_{a+1}(y) - L_a(y), \quad a = 1, \ldots, k - 1.
$$

(3.5)

Since $D^*(a|y) = L \sum_{j=1}^a f(y|\alpha_j(r')) - L \sum_{j=a+1}^{k-1} f(y|\alpha_j(r'))$, $D^*(a|y)$ is nondecreasing in $a$.

Using this property, we can state the $\Gamma$-minimax rule in the form:

Randomize your decision over the set $B^*(y) = \{a|D^*(a|y) = 0\}$. If the set is vacuous, then choose action $b$, where $b$ is the largest $a$ for which $D^*(a|y) < 0$. If such a $b$ does not exist, choose action 1.

A remark similar to Remark 2.1 holds here.

4. Locally Optimal Rule

In this section, a selection rule $\delta$ is defined by the individual selection probabilities $\delta_i(y), i = 1, \ldots, k - 1$. This results in selection of a subset (possibly empty) of the $k-1$ points. Further, the selected points need not be consecutive. We restrict ourselves to the class of rules defined by

$$
D = \{\delta| \lim_{r \uparrow r^*} E_{\Omega_m(r)}(\delta_m(Y)) = P^* \text{ for } m = 1, \ldots, k - 1\}.
$$

(4.1)
In other words, considering the configurations of \( \varphi \) for any given change-point location, the limiting value of the probability of choosing the true location is \( P^* \) as the amount of shift goes to \( \tau^* \) (corresponding to equi-parameter configuration of \( \theta \)). This is similar to the class considered by Gupta, Huang and Nagel (1979) but for a different problem.

We seek a rule \( \delta \in \mathcal{D} \) which is locally optimal in the sense that it maximizes

\[
\mathcal{Q} = \sum_{m=1}^{k-1} \lim_{\tau \to \tau^*} \frac{\partial}{\partial \tau} E_{\varphi_m(\tau)} (\delta_m(\mathcal{Y})). \tag{4.2}
\]

The quantity \( \mathcal{Q} \) reflects the sensitivity of the rule in making the correct decision in a neighborhood of an equi-parameter configuration of \( \theta \). Such a measure in a different context has been used by Huang, Panchapakesan and Tseng (1984).

In deriving a locally optimal rule, we assume usual regularity conditions so that limit operations and differentiations under the integral signs can be carried out. Then, we can write

\[
\mathcal{Q} = \sum_{m=1}^{k-1} \int_y \delta_m(y) f_m(y|\tau^*) \, dy
\]

where \( f_m(y|\varphi_m(\tau)) = \frac{\partial}{\partial \tau} f(y|\varphi_m(\tau)) \) and \( \tau^* = (\tau^*, \ldots, \tau^*) \).

**Theorem 4.1.** Under regularity conditions, a rule belonging to \( \mathcal{D} \) which maximizes \( \mathcal{Q} \) is given by

\[
\delta_m^*(y) = \begin{cases} 
1 & \text{if } \frac{f_m(y|\tau^*)}{f_n(y|\tau^*)} > c_m \\
0 & \text{if } \frac{f_m(y|\tau^*)}{f_n(y|\tau^*)} < c_m
\end{cases}
\tag{4.3}
\]

where \( \lambda_m \) and \( c_m \) are determined such that

\[
\int_y \delta_m^*(y) f(y|\tau^*) \, dy = P^*, \ m = 1, \ldots, k - 1. \tag{4.4}
\]

**Proof:** The proof is straightforward, by noting that for any rule \( \delta \in \mathcal{D} \), \( \int_y \delta_m(y) f(y|\tau^*) \, dy = P^* \) for \( m = 1, \ldots, k - 1 \), and that

\[
\sum_{m=1}^{k-1} \int_y (\delta_m^*(y) - \delta_m(y))(f_m(y|\tau^*) - c_m f(y|\tau^*)) \, dy \geq 0.
\]

**Example.** Suppose we have a sequence of sample means (based on \( n \) independent observations) from normal distributions \( N(\theta_i, \sigma^2), i = 1, \ldots, k \), where \( \sigma^2 \) is known. We take \( Y_i = \bar{X}_{i+1} - \bar{X}_i \), \( i = 1, \ldots, k - 1 \). Then, for \( \theta \in \Omega_m \), \( Y \) has a \((k-1)\text{-variate normal distribution having mean vector with } \tau = \theta_{m+1} - \theta_m \) as the \( m \)-th component and zero
everywhere else, and covariance matrix $V = (\sigma_{ij})$, where $\sigma_{ii} = \frac{2\sigma^2}{n}$ and $\sigma_{ij} = -\frac{\sigma^2}{n}$ or 0 according as $|i - j|$ is $= or > 1$. It is easy to see that

$$\frac{f_m(y|x^*)}{f(y|x^*)} = \sum_{i=1}^{k-1} y_i \sigma^{im}$$

where $V^{-1} = (\sigma^{im})$. By noting that $\sigma^{i1} = \frac{(k-i)\sigma^2}{k\sigma^2}$ for $1 \leq i \leq k-1$, and $\sigma^{ij} = j\sigma^{i1}$ for $1 \leq j \leq i$, we get

$$\sum_{i=1}^{k-1} y_i \sigma^{im} = \frac{(k-m)mn}{k\sigma^2} \left[ \frac{\bar{X}_{m+1} + \ldots + \bar{X}_k}{k-m} - \frac{\bar{X}_1 + \ldots + \bar{X}_m}{m} \right].$$

This gives the intuitively appealing rule:

$$\delta_m(y) = \begin{cases} 1 & \text{if } \frac{\bar{X}_{m+1} + \ldots + \bar{X}_k}{k-m} - \frac{\bar{X}_1 + \ldots + \bar{X}_m}{m} > c_m; \\ 0 & \text{otherwise}; \end{cases}$$

where $c_m = \sigma \sqrt{\frac{k}{nm(k-m)}} \Phi^{-1}(1 - P^*)$.

REFERENCES


INFEERENCE ABOUT THE CHANGE-POINT IN A SEQUENCE OF RANDOM VARIABLES: A SELECTION APPROACH

TaChen Liang and S. Panchapakesan

Purdue University
Department of Statistics
West Lafayette, IN 47907

Office of Naval Research
Washington, DC

Approved for public release, distribution unlimited.

change-point inference, selection, Bayes rule, gamma-minimax, locally optimal

A selection approach is made for making inference about the point in a sequence of random variables at which the underlying distribution changes. Three selection rules are derived: a Bayes rule, a gamma-minimax rule, and a locally optimal rule. While the first two rules select one point as the change-point, the third selects a subset (possibly empty) of points.