*Change-point Problems*
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

# SOME ASPECTS OF CHANGE-POINT ANALYSIS

BY P. K. BHATTACHARYA
*University of California, Davis*

Change-points divide statistical models into homogeneous segments. Inference about change-points is discussed here in the context of testing the hypothesis of 'no change', point and interval estimation of a change-point, changes in nonparametric models, changes in regression, and detection of change in distribution of sequentially observed data.

**1. Introduction.** Suppose that in a linear array of independent observations $Y_1, \ldots, Y_n$, the distribution is subject to change after $Y_\tau$ for some $1 \leq \tau \leq n-1$. Detection and estimation of change-points which in this way divide statistical models into homogeneous segments is a fast-developing area of research in statistical theory and methods. We shall present here a brief account of some of the areas of change-point analysis.

The most basic problems are those of testing the hypothesis of "no change," and of estimating a change-point by a point estimator or a confidence set when the presence of one is suspected. In Sections 2–4, we shall discuss these problems and some nonparametric methods will be presented in Section 5. Change-point problems also occur in the context of regression when the nature of dependence of one variate on another may be different in two segments of the data, and in situations where the observations are obtained sequentially with the possibility of a change in distribution at any stage. Methods in these two areas will be discussed in Sections 6 and 7.

In presenting these accounts, our aim is to concentrate on the main issues rather than provide a comprehensive review of the available literature. There is an annotated bibliography compiled by Shaban (1980) and a survey article by Zacks (1983) where references to many of the early works on change-point analysis may be found. Our references will be mostly limited to the sources

---

AMS 1991 Subject Classification: Primary 62-02, 62E20; Secondary 62F03, 62F10, 62F25, 62G05, 62G10, 62J05, 62L15.

Key words and phrases: Change-point, two phase regression, quality control, two-sided random walk.

we draw upon in this article and will be given at the end of the sections to which they belong.

A common feature of all change-point methodologies is that due to the very nature of the problem, one must split the data in different ways and measure the divergence between the two segments by some criterion to look for heterogeneity. This naturally leads to the consideration of some stochastic processes, whose extrema form the basis of likelihood methods and whose integrals with respect to prior distributions form the basis of Bayesian methods of change-point analysis. This will be made explicit in some of our discussion, but will be implicit otherwise.

**2. Testing the Hypothesis of "No Change".** We shall restrict our discussion to the problem of change in mean of independent normal variables with common variance. Let $Y_i = \mu_i + \sigma Z_i$, $1 \leq i \leq n$, where $Z_1, \ldots, Z_n$ are iid $N(0,1)$. The problem is to test

$H : \mu_1 = \ldots = \mu_n$ against $A_+ : \mu_1 = \ldots = \mu_\tau < \mu_{\tau+1} = \ldots = \mu_n$ for some

$1 \leq \tau \leq n-1$, or against $A : \mu_1 = \ldots = \mu_\tau \neq \mu_{\tau+1} = \ldots = \mu_n$.

Let $\Delta = \mu_{\tau+1} - \mu_\tau$.

*2.1. Bayesian Approach.* We discuss the case of known $\sigma$ with $\sigma = 1$. If the initial mean $\mu_1$ is known, then taking $\mu_1 = 0$, we proceed from a prior in which $\tau$ is uniformly distributed on $\{1, \ldots, n-1\}$ and $\Delta$ is independent of $\tau$, having pdf $h$. The null hypothesis of "no change" is rejected for large values of the likelihood ratio

$$L(Y_1, \ldots, Y_n) = \frac{\frac{1}{n-1}\sum_{r=1}^{n-1} \int h(\delta) \exp[-\frac{1}{2}\sum_1^r Y_i^2 - \frac{1}{2}\sum_{r+1}^n (Y_i - \delta)^2] d\delta}{\exp[-\frac{1}{2}\sum_1^n Y_i^2]}$$

For the case of one-sided alternative $A_+$, taking $h$ to be the pdf of folded normal distribution with variance $\alpha^2$, we have

$$\frac{\alpha}{2} L = \frac{1}{n-1} \sum_{r=1}^{n-1} \exp\left[\frac{1}{2} \cdot \frac{S_{n-r}'^2}{n-r+\alpha^{-2}}\right] \Phi\left(\frac{S'_{n-r}}{n-r+\alpha^{-2}}\right)$$

where $S'_{n-r} = \sum_{r+1}^n Y_i$ and $\Phi$ is the cdf of $N(0,1)$. Now let $\alpha \to 0$, i.e., suppose that the change is small. Then $S'_{n-r}(n-r+\alpha^{-2})^{-1} \to 0$, so that for small $\alpha$,

$$\frac{\alpha}{2}L = \frac{1}{n-1}\sum_{r=1}^{n-1}\left[\frac{1}{2} + \frac{1}{\sqrt{2\pi}}\alpha^2 S'_{n-r}\right]\{1+o(1)\}$$
$$\simeq \frac{1}{2} + \frac{\alpha^2}{\sqrt{2\pi}} \cdot \frac{1}{n-1}\sum_{r=1}^{n-1} S'_{n-r}.$$

This leads to the test statistic

$$T_+ = \sum_{r=1}^{n-1} S'_{n-r} = \sum_{i=2}^{n-1}(i-1)Y_i$$

and the hypothesis H of "no change" is rejected for large values of $T_+$.

When $\mu_1$ is unknown, a prior for $\mu_1$ is taken to be $N(0,\beta^2)$ independent of $(\tau,\Delta)$ distributed as above, and then $\beta$ is allowed to tend to $\infty$. This leads to the test statistic

$$T_+^* = \sum_{i=2}^{n}(i-1)(Y_i - \bar{Y}_n),\ \bar{Y}_n = n^{-1}\sum_{1}^{n} Y_i$$

and $H$ is rejected for large values of $T_+^*$.

For two-sided alternatives, everything is carried out as above, except that the prior distribution of the amount of change $\Delta$ is taken to be $N(0,\alpha^2)$ (instead of folded normal) with $\alpha \to 0$. The test statistics obtained in this manner are

$$T = n^{-2}\sum_{i=2}^{n}\left[\sum_{j=i}^{n} Y_j\right]^2 \quad \text{for known initial mean,}$$

$$T^* = n^{-2}\sum_{i=2}^{n}\left[\sum_{j=i}^{n}(Y_j - \bar{Y}_n)\right]^2 \quad \text{for unknown initial mean.}$$

More general priors including priors for unknown $\sigma^2$ have also been considered in the literature without derivation of concrete test statistics as above.

The statistics $T_+$ and $T_+^*$ for the one-sided case are linear functions of the $Y_i$'s, so they are normally distributed. In the two-sided case, the test statistics are quadratic forms. Under H, the statistics $T$ and $T^*$ are distributed as $\sum_{r=1}^{n-1}\lambda_r Z_r^2$ and $\sum_{r=1}^{n-1}\lambda_r^* Z_r^2$, where $Z_1,\ldots,Z_{n-1}$ are independent $N(0,1)$, $\lambda_r = \{2n\sin(\frac{r\pi}{2n})\}^{-2}$ and $\lambda_r^* = \{2n\sin(\frac{(2r-1)\pi}{2(2n-1)})\}^{-2}$.

*2.2. Likelihood Ratio Tests.* Consider the hypothesis $A_r$ that the mean changes after $\tau = r$ and let

$$T_r = \frac{\bar{Y}'_{n-r} - \bar{Y}_r}{[r^{-1} + (n-r)^{-1}]^{1/2}} \tag{1}$$

where

$$\bar{Y}_r = \frac{1}{r}\sum_1^r Y_i, \; \bar{Y}'_{n-r} = \frac{1}{n-r}\sum_{r+1}^n Y_i.$$

Then the likelihood ratio test (LRT) for the hypothesis $H$ of "no change" against the alternative $A_r$ rejects $H$ for large values of $|T_r|$ when $\sigma$ is known and for large values of

$$W_r = |T_r| / \left[\sum_1^r (Y_i - \bar{Y}_r)^2 + \sum_{r+1}^n (Y_i - \bar{Y}'_{n-r})^2\right]^{1/2}$$

when $\sigma$ is unknown. Now the alternative hypothesis $A$ is the union of $A_1, \ldots,$ $A_{n-1}$. Hence if $\sigma$ is known, the LRT of $H$ against $A$ is based on the test statistic $U = \max_{1\le r\le n-1} |T_r|$ and if $\sigma$ is unknown, the LRT is based on the test statistic $W = \max_{1\le r\le n-1} W_r$, or equivalently, on $V = \max_{1\le r\le n-1} |T_r|/S$ with $S^2 = \sum_1^n (Y_i - \bar{Y}_n)^2$ since $W^2 = V^2/(1 - V^2)$.

The LRT's are modified for one-sided alternatives, rejecting $H$ in favor of $A_+$ for large values of $\max_{1\le r\le n-1} T_r$ when $\sigma$ is known and for large values of $\max_{1\le r\le n-1} T_r/S$ when $\sigma$ is unknown.

In the above discussion, the initial mean $\mu_1$ is unknown. If $\mu_1$ and $\sigma$ are both known, say $\mu_1 = 0$ and $\sigma = 1$, then the LRT rejects $H$ in favor of $A_+$ for large values of $\max_{1\le r\le n-1} \sqrt{n-r}\bar{Y}'_{n-r}$ and in favor of A for large values of $\max_{1\le r\le n-1} \sqrt{n-r}|\bar{Y}'_{n-r}|$.

The null distribution of the LRT statistic $U = \max_{1\le r\le n-1} |T_r|$ for the case of $\sigma = 1$ is obtained from the fact that $T_1, \ldots, T_{n-1}$ are jointly normal with $E(T_r) = 0$ and $\mathrm{Cov}(T_r, T_s) = \sqrt{\frac{r(n-s)}{s(n-r)}}$ for $1 \le r \le s \le n-1$. It follows that $(T_1, \ldots, T_{n-1})$ is Markovian and reversible. From these properties of $(T_1, \cdots, T_{n-1})$, the pdf of the null distribution of $U$ is obtained as

$$f_U(x) = 2\varphi(x)\sum_{r=1}^{n-1} g_r(x,x)g_{n-r}(x,x),$$

where $\varphi$ is the pdf of $N(0,1)$ and

$$g_r(x,y) = P[|T_i| < y, \; 1 \le i \le r-1 | T_r = x], \; x, y \ge 0$$

are given by a recursion formula. The null distribution of
$V = \max_{1 \le r \le n-1} |T_r|/S$ is more complicated.

*2.3. Comparison of Powers of Bayes and Likelihood Ratio Tests.* For the case of known $\sigma(=1)$, the following comparisons based on exact calculations for the Bayes test and on simulations for the LRT hold in the one-sided case as well as the two-sided case. The results mainly depend on the location of the change-point $\tau$ relative to the length $n$ of the observed sequence.

For the case of known initial mean, both tests have their best powers when $\tau = 1$. In their relative performance, the Bayes test is superior for $\tau/n \le 0.4$, the LRT is superior for $\tau/n \ge 0.75$ and for $0.4 < \tau/n < 0.75$, the Bayes test dominates the LRT for small $\Delta$ and vice versa.

For the case of unknown initial mean, both tests work at their respective best when $\tau/n = 1/2$. In comparison to one another, the Bayes test is superior for $|\frac{\tau}{n} - \frac{1}{2}| \le 0.1$, the LRT is superior for $|\frac{\tau}{n} - \frac{1}{2}| \ge 0.25$ and for $0.1 < |\frac{\tau}{n} - \frac{1}{2}| < 0.25$ the Bayes test dominates the LRT for small $\Delta$ and vice versa.

*2.4. Modifications of the Likelihood Ratio Test for the Case of Known $\sigma$.* The LRT statistic $U = \max_{1 \le r \le n-1} |T_r|$, where $T_r$ is given by (1), does not have an asymptotic distribution under $H$ as $n \to \infty$. To see this, let $X_n(t) = T_{[nt]}$, $0 < t < 1$ and note that $U = \max_{0<t<1} X_n(t)$. Now $X_n(\cdot)$ converges to a Gaussian process $X(\cdot)$, of which the transform $\zeta(t) = t^{1/2} X(\frac{t}{1+t})$, $0 < t < \infty$ is a standard Brownian motion. The law of iterated logarithm and the unboundedness of $\log|\log t|$ as $t \to 0$ or $\infty$ imply that $|\zeta(t)| > Ct^{1/2}$ in every neighborhood of 0 and $\infty$ and equivalently, $|X(t)| > C$ in every neighborhood of 0 and 1 with probability 1. Thus as $n \to \infty$, $U \to \infty$ and $n^{-1} \arg\max_r |T_r| \to 0$ or 1 with probability 1 under $H$.

Since $\{T_r\}$ becomes unstable at the two ends of the sequence for large $n$, the following modifications of $U$ are worth considering:

(1) $U_1 = \max_{m \le r \le n-m} |T_r|$ where $m/n = \alpha > 0$ but small. Such a test statistic has an asymptotic distribution under $H$, although it is not available in closed form.

(2) $U_2 = \max_{1 \le r \le n-1} n^{-1/2} |S_r - rn^{-1} S_n|$ converges in law to $\max_{0 \le t \le 1} |B_0(t)|$ under $H$, where $B_0(t) = B(t) - tB(1)$ is the Brownian bridge, and the distribution of $\max_{0 \le t \le 1} |B_0(t)|$ is well-known.

*2.5. Bibliographic Notes.* A Bayesian change-point model for a fixed sequence of normal means was formulated by Chernoff and Zacks (1964). Although this model included the possibility of any number of changes and the basic objective was to estimate the mean at the end of the sequence, one-sided tests for the null hypothesis of "no change" in the model of "at most one change" and the posterior distribution of a change-point followed naturally. The Bayesian point of view was pursued further by Kander and Zacks (1966) to extend these results to exponential families and by Smith (1975) to multiparameter models with possible constraints. Bayes tests for two-sided change in normal means were considered by Gardner (1969) who derived the asymptotic null distribution of the statistics. Exact null distributions of these statistics were obtained by Sen and Srivastava (1975) who also made a comparative study of the Bayes and the LRT's as mentioned above. Exact null distributions of the LRT statistics were expressed in terms of recursion formulas by Hawkins (1977) for the normal means problem and by Worsley (1986) in the more general context of an exponential family. An approximation to the tail probability of the null distribution of the studentized likelihood ratio statistic was obtained by James, James and Siegmund (1992).

## 3. Estimation of a Change-Point.

*3.1. Bayesian Approach.* As in the previous section, suppose that we observe $Y_i = \mu_i + Z_i$, $1 \le i \le n$, where $Z_1, \ldots, Z_n$ are independent $N(0,1)$ and either there exists $1 \le \tau \le n-1$ such that $\mu_1 = \cdots = \mu_\tau \ne \mu_{\tau+1} = \cdots = \mu_n$, or $\mu_1 = \cdots = \mu_n$ in which case we let $\tau = 0$. We now consider the problem of estimating $\tau$ (estimating $\tau = 0$ corresponds to accepting the hypothesis of "no change") and the current mean $\mu_n$.

Put a prior on $\tau, \mu_n$ and $\Delta = \mu_\tau - \mu_{\tau+1}$ (which takes effect if $\tau \ne 0$), letting $\tau, \mu_n, \Delta$ to be mutually independent with $P(\tau = j) = p(j), \mu_n$ distributed as $N(0, \beta^2)$ and $\Delta$ distributed as $N(0, \alpha^2)$. Since $(\bar{Y}_r, \bar{Y}'_{n-r})$ is sufficient under $\tau = r$, the conditional Bayes estimator of the current mean $\mu_n$ given $\tau = r$ is $E(\mu_n | \tau = r, \bar{Y}_r, \bar{Y}'_{n-r})$, which is approximated by

$$\hat{\mu}_n(r) = \frac{n\bar{Y}_n + \alpha^2 r(n-r)\bar{Y}'_{n-r}}{n + \alpha^2 r(n-r)}$$

as $\beta \to \infty$. Moreover, as both $\beta$ and $\alpha \to \infty$, the posterior distribution of $\tau$ is approximated by

$$P[\tau = r | \mathbf{Y}_1, \ldots, Y_n] = w(r) / \sum_{j=0}^{n-1} w(j)$$

where $w(0) = n^{-1/2} p(0)$ and for $1 \le r \le n-1$,

$$w(r) = \alpha^{-1} \{r(n-r)\}^{-1/2} p(r) \exp[T_r^2/2]$$

with $T_r$ given by (1). Averaging the conditional Bayes estimator $\hat{\mu}_n(r)$ with respect to the posterior distribution of $\tau$, we are led to the Bayes estimator

$$\hat{\mu}_n = \sum_{r=0}^{n-1} w(r)\hat{\mu}_n(r) / \sum_{r=0}^{n-1} w(r)$$

of the current mean.

*3.2. Maximum Likelihood Estimator.* Assume that there exists $1 \le \tau \le n-1$ such that the joint pdf of $(Y_1, \dots Y_n)$ is

$$\prod_1^{\tau} f(y_i, \theta_1) \prod_{\tau+1}^{n} f(y_i, \theta_2)$$

where $f$ is known and $\theta_1 \neq \theta_2$.

First let $\theta_1$, $\theta_2$ be known. Then the maximum likelihood estimator (MLE) of $\tau$ is

$$\begin{aligned}\hat{\tau} &= \arg \max_{1 \le r \le n-1} \left[ \sum_1^{r} \log f(Y_i, \theta_1) + \sum_{r+1}^{n} \log f(Y_i, \theta_2) \right] \\ &= \arg \max_{1 \le r \le n-1} \sum_1^{r} W_i\end{aligned}$$

where $W_i = \log[f(Y_i, \theta_1)/f(Y_i, \theta_2)]$ and any nonuniqueness in maximization is resolved by suitable convention. To think of the distribution of $\hat{\tau} - \tau$, note that

$$\hat{\tau} - \tau = \arg \max_j \sum_1^{\tau+j} W_i = \arg \max_j \xi(j),$$

where $\xi(0) = 0$ and

$$\xi(j) = \sum_1^{\tau+j} W_i - \sum_1^{\tau} W_i = \begin{cases} \sum_1^j W_{2i} = S_{2j}, & j > 0 \\ \sum_1^{-j} W_{1i} = S_{1,-j}, & j < 0, \end{cases}$$

with $W_{1i} = -W_{\tau-i+1}$ for $1 \le i \le \tau - 1$ and $W_{2i} = W_{\tau+i}$ for $1 \le i \le n - \tau$. In terms of the independent random walks

$$\left\{ S_{1j} = \sum_1^j W_{1i},\ 1 \le j \le \tau - 1 \right\},\quad \left\{ S_{2j} = \sum_1^j W_{2i},\ 1 \le j \le n - \tau \right\}$$

governed by $f(\cdot, \theta_1)$, $f(\cdot, \theta_2)$ respectively, and letting $S_{10} = S_{20} = 0$, we can express $\hat{\tau} - \tau$ as follows. Let

$$M_1 = \max_{0 \le j \le \tau-1} S_{1j},\ J_1 = \arg\max S_{1j} = \min\{j : S_{1j} = M_1\},$$

$$M_2 = \max_{0 \le j \le n-\tau} S_{2j}, \; J_2 = \arg\max S_{2j} = \min\{j : \; S_{2j} = M_2\},$$

Then (see Fig 1)

$$\hat{\tau} - \tau = J^* = \begin{cases} -J_1 & \text{if } M_1 \ge M_2 \\ J_2 & \text{if } M_1 < M_2 \end{cases}$$

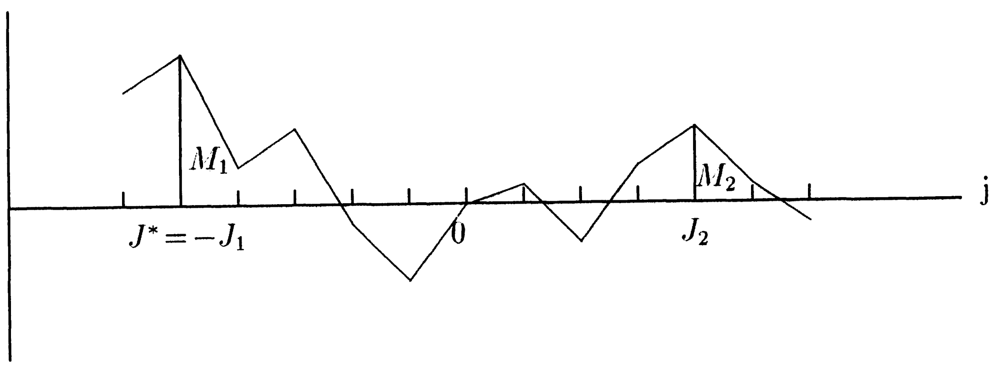


Figure 1: Representation of $\hat{\tau} - \tau$ in terms of two random walks

For unknown $\theta_1, \; \theta_2$,the MLE $(\hat{\theta}_1, \; \hat{\theta}_2, \; \hat{\tau})$ of $(\theta_1, \; \theta_2, \; \tau)$ is the maximizer of

$$L_n(\varphi_1, \; \varphi_2, \; r) = \sum_1^r \log f(Y_i, \; \varphi_1) + \sum_{r+1}^n \log f(Y_i, \; \varphi_2).$$

If the conditional MLE $(\hat{\theta}_{1r}, \; \hat{\theta}_{2r})$ of $(\theta_1, \; \theta_2)$ given $\tau = r$ is available in closed form, then the MLE $\hat{\tau}$ of $\tau$ is the maximizer of $L_n(\hat{\theta}_{1r}, \; \hat{\theta}_{2r}, \; r)$. In many situations,

$$\hat{\theta}_{1r} = \bar{Y}_r = r^{-1} \sum_1^r Y_i, \; \hat{\theta}_{2r} = \bar{Y}'_{n-r} = (n-r)^{-1} \sum_{r+1}^n Y_i.$$

Specifically, for change in mean of a normal distribution with known variance $\sigma^2$,

$$2\sigma^2 L_n(\hat{\theta}_{1r}, \; \hat{\theta}_{2r}, \; r) = -n\sigma^2 \log(2\pi) - \sum_1^n (Y_i - \bar{Y}_n)^2 + T_r^2,$$

where $T_r$ is given by (1). For convenience, let $Q_{nr} = -T_r$, i.e.,

$$\begin{aligned} Q_{nr} &= \frac{\bar{Y}_r - \bar{Y}'_{n-r}}{\{r^{-1} + (n-r)^{-1}\}^{1/2}} \\ &= \sqrt{\frac{n}{r(n-r)}}\left(\sum_1^r Y_i - rn^{-1}\sum_1^n Y_i\right) \\ &= \sigma\sqrt{\frac{n}{r(n-r)}}[(S_r^* - rn^{-1}S_n^*) + ng_n(r)\Delta\sigma^{-1}], \\ S_r^* &= \sigma^{-1}\sum_1^r (Y_i - EY_i), \quad \Delta = \theta_2 - \theta_1, \\ g_n(r) &= \begin{cases} -rn^{-1}(1 - \tau n^{-1}), & r \le \tau \\ -(1 - rn^{-1})\tau n^{-1}, & r > \tau. \end{cases} \end{aligned}$$

For $\Delta > 0$, $|Q_{nr}|$ is maximized with a negative value of $Q_{nr}$ with probability tending to 1 as $n \to \infty$, because $\max_r |S_r^* - rn^{-1}S_n^*| = O_p(n^{1/2})$ and $\max_r |ng_n(r)| = O(n)$. Thus asymptotically,

$$\hat{\tau} = \arg\max_r |Q_{nr}| = \arg\min_r Q_{nr} = \arg\min_r (Q_{nr} - Q_{n\tau}).$$

For large n and small $|r-\tau|$, $\{Q_{nr} - Q_{n\tau},\ r \le \tau\}$ and $\{Q_{nr} - Q_{n\tau},\ r > \tau\}$ behave approximately as independent random walks of which each step is $N(0,1)$ with a drift of $\Delta/(2\sigma)$.

Getting back to general parametric families, we consider the change-point problem in an asymptotic set up in which the length of the observed sequence $n \to \infty$ and the amount of change in the parameter is $\Delta = \theta_2 - \theta_1 = \delta\nu_n^{-1}$, where $\nu_n \to \infty$ slower than $n^{1/2}$. Then under regularity conditions, the following weak convergence holds for the log likelihood ratio with respect to uniform convergence on compact sets:

$$\begin{aligned} &\xi_n(u, v, t) = L_n(\theta_1 + un^{-1/2},\ \theta_2 + vn^{-1/2},\ \tau + \nu_n^2 t) - L_n(\theta_1, \theta_2, \tau) \\ &\xrightarrow{w} \frac{1}{2}(Z_1^2 + Z_2^2) - \frac{1}{2}\lambda I\left(u - \frac{Z_1}{\sqrt{\lambda I}}\right)^2 - \frac{1}{2}(1-\lambda)I\left(v - \frac{Z_2}{\sqrt{(1-\lambda)I}}\right)^2 \qquad (2) \\ &- \sqrt{I}|\delta|[B(t) + \frac{1}{2}|\delta|\sqrt{I}|t|], \end{aligned}$$

where $Z_1$, $Z_2$ are independent $N(0,1)$, $\{B(t), -\infty < t < \infty\}$ is a two-sided standard B.M. independent of $(Z_1,\ Z_2)$, $I = I(\theta_1)$ is the Fisher-information in the parametric family described by $f$ and $\lambda = \tau/n$.

To examine the asymptotics of the MLE $(\hat{\theta}_1,\ \hat{\theta}_2,\ \hat{\tau})$ of $(\theta_1,\ \theta_2,\ \tau)$, note that

$$\begin{aligned}(n^{1/2}(\hat{\theta}_1-\theta_1),\ n^{1/2}(\hat{\theta}_2-\theta_2),\ \nu_n^{-2}(\hat{\tau}-\tau))\\ =(U_n,\ V_n,\ T_n)=\arg\max \xi_n(u,\ v,\ t).\end{aligned} \tag{3}$$

Now the sequence $\nu_n^{-2}(\hat{\tau}-\tau)$ can be shown to be $O_p(1)$. From (2) and (3), it therefore follows that

$$\nu_n^{-2}(\hat{\tau}-\tau) \xrightarrow{L} T_{|\delta|\sqrt{I}},$$

where

$$T_\mu = \arg\min_t [B(t)+\frac{1}{2}\mu|t|], \qquad \mu>0.$$

These results extend in a natural manner to change-points in multiparameter families.

*3.3. Asymptotically Equivalent Estimator to the MLE.* For arbitrary $\varphi$, let $W_i = W_i(\varphi) = \frac{\partial \log f(Y_i,\varphi)}{\partial\theta}$ and define $\tilde{\tau}$ as the maximizer of $|\tilde{T}_r|$, where $\tilde{T}_r$ is obtained by replacing $\bar{Y}_r$ and $\bar{Y}'_{n-r}$ in (1) by $\bar{W}_r = r^{-1}\sum_1^r W_i$ and $\bar{W}'_{n-r} = (n-r)^{-1}\sum_{r+1}^n W_i$ respectively.

If $f$ belongs to an exponential family, then $\nu_n^{-2}(\tilde{\tau}-\tau)$ has the same asymptotic distribution as $\nu_n^{-2}(\hat{\tau}-\tau)$. For non-exponential family, $\tilde{\tau}$ is used as an initial estimator of $\tau$ to obtain a consistent estimator $\tilde{\theta}_1$ of $\theta_1$ and the procedure is repeated with $\varphi=\tilde{\theta}_1$. This leads to an estimator $\tau^*$ which has the same asymptotic distribution as that of the MLE.

For example, to estimate the change-point in the parameter $\theta$ of Bernoulli trials, let $Y_1,\ldots,Y_n$ be the original observations,

$$\begin{aligned} W_i &= \frac{\partial \log f(Y_i,\varphi)}{\partial\theta} = \frac{Y_i-\varphi}{\varphi(1-\varphi)},\\ \bar{W}_r-\bar{W}'_{n-r} &= \frac{\bar{Y}_r-\bar{Y}'_{n-r}}{\varphi(1-\varphi)},\\ \varphi(1-\varphi)|\tilde{T}_r| &= \sqrt{\frac{n}{r(n-r)}}|S_r-rn^{-1}S_n|,\end{aligned}$$

where $S_r=\sum_1^r Y_i$. Thus

$$\tilde{\tau} = \arg\max_r \sqrt{\frac{n}{r(n-r)}}|S_r-rn^{-1}S_n|.$$

In simulation studies, a comparison of $\tilde{\tau}$ and a simpler estimator

$$\tilde{\tilde{\tau}} = \arg\max_r |S_r-rn^{-1}S_n|$$

with the MLE

$$\hat{\tau} = \arg\max_{r}[\bar{Y}_r \log \bar{Y}_r + (1-\bar{Y}_r)\log(1-\bar{Y}_r)]$$

showed that

(i) $\tilde{\tau}$ has practically the same performance as $\hat{\tau}$ (actually, $\tilde{\tau}$ and $\hat{\tau}$ were identical in 86% to 100% of the cases);

(ii) $\tilde{\tilde{\tau}}$ is definitely inferior to $\hat{\tau}$ unless $\lambda = \tau/n \simeq 1/2$.

The reason for (ii) is that $\nu_n^{-2}(\tilde{\tilde{\tau}} - \tau)$ and $\nu_n^{-2}(\hat{\tau} - \tau)$ converge in law to

$$\tilde{T} = \arg\min_{t}[B(t) + \{1(t \le 0)(1-\lambda) + 1(t > 0)\lambda\} \cdot \frac{\delta|t|}{\sqrt{\theta(1-\theta)}}] = \arg\min_{t} \tilde{X}(t)$$

and

$$T = \arg\min_{t}[B(t) + \frac{1}{2} \cdot \frac{\delta|t|}{\sqrt{\theta(1-\theta)}}] = \arg\min_{t} X(t)$$

respectively, of which the former is stochastically larger for $\lambda \neq 1/2$ as can be seen from a comparison of the drift terms in Figure 2.

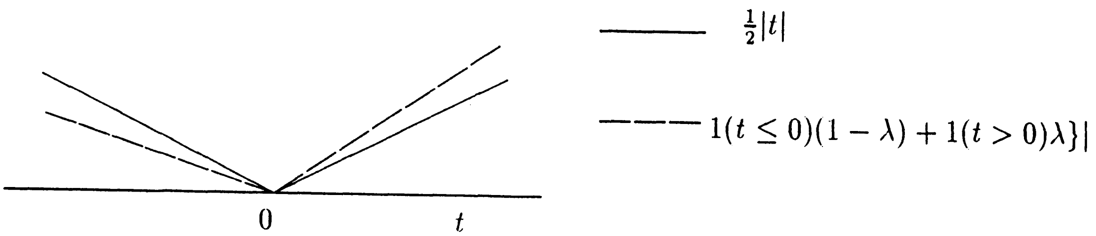


Figure 2: Comparison of drift terms of $X(t)$ and $\tilde{X}(t)$ for $\lambda > \frac{1}{2}$

The term $\sqrt{r(n-r)}$ in the denominator of the statistic $|\tilde{T}_r|$ to be maximized to obtain $\tilde{\tau}$ is quite important as can be seen from the above discussion. However, for values of $r$ near the two ends of the sequence, this term makes the statistic $\tilde{T}_r$ unstable. For this reason, $|\tilde{T}_r|$ should be maximized over $m \le r \le n - m$ for suitably chosen small $m$. Actually the asymptotic distribution referred to above holds when maximization is carried out over $n^{1/2}\nu_n \le r \le n - n^{1/2}\nu_n$ to obtain $\hat{\tau}$ and $\tilde{\tau}$. This restriction is asymptotically negligible and yet enough to avoid instability.

*3.4. Bibliographic Notes.* Bayes estimator of the change-point in a sequence of normal means was obtained by Chernoff and Zacks (1964). Ibragimov and Has'minskii (1981) have studied the change-point problem in continuous time, considering the onset of a drift in a standard Brownian motion. MLE of a change-point was first studied by Hinkley (1970, 1972) who recognized the key role of the extremum of a two-sided random walk in this context. Asymptotic distribution of the MLE of a change-point was derived by Bhattacharya and Brockwell (1976) in the normal means problem and by Bhattacharya (1987) in general multiparameter families. The comparison between MLE and two competing estimators in the Bernoulli case mentioned above was made by Pettit (1980). Cobb (1978) showed that the conditional distribution of the MLE given the ancillary data adjacent to the MLE is approximately the same as the Bayesian posterior corresponding to the uniform prior. In one of the very few papers on inference about multiple change-points, Yao (1988) has obtained a consistent estimator of the number of change-points using the Schwarz information criterion.

**4. Confidence Set for a Change-Point.** Even when the initial pdf $f_0$ and the changed pdf $f_1$ are completely specified, a confidence set ($CS$) for a change-point $\tau$ based on the distribution of the MLE $\hat{\tau}$ using $\hat{\tau}-\tau$ as a pivot is inefficient. This is because $\hat{\tau}$ is not sufficient and CS's based on the conditional distribution of $\hat{\tau}-\tau$ given the ancillary statistics $\{\sum_{j=1}^{\hat{\tau}+i} W_j - \sum_{j=1}^{\hat{\tau}} W_j,\ i = \pm 1, \pm 2, \ldots\}$ where $W_j = \log[f_1(Y_j)/f_0(Y_j)]$, perform better than those based on the distribution of $\hat{\tau}-\tau$.

We shall discuss here CS's for $\tau$ within the framework of single-parameter exponential families, obtained by inverting the acceptance regions of the LRT's for $\tau = k$ against $\tau \neq k$. The confidence coefficient is $1-\alpha$ throughout.

Let $Y_1, \cdots, Y_n$ be independent observations whose pdf changes from $f(\cdot, \theta_0)$ to $f(\cdot, \theta_1)$ after the $\tau$-th observation, where $f(y,\theta) = \exp[\theta y - \psi(\theta)]h(y)$. Then the LRT for $\tau = k$ against $\tau \neq k$ has acceptance region of the form

$$A(k,c) = \{\underline{y} : \max_r \Lambda_r(\underline{y}) - \Lambda_k(\underline{y}) \leq c^2\} \tag{4}$$

where

$$\begin{aligned}\Lambda_r(\underline{y}) = \Lambda_r\left(\sum_1^r y_i, \sum_1^n y_i\right) &= r \sup_\theta \left[\theta r^{-1} \sum_1^r y_i - \psi(\theta)\right] \\ &+ (n-r) \sup_\theta \left[\theta (n-r)^{-1} \sum_{r+1}^n y_i - \psi(\theta)\right].\end{aligned} \tag{5}$$

If we could find $c = c(\alpha, k)$ for each $k$ such that $P_{\tau=k}[A(k,c)] = \alpha$, then $\{k : \underline{y} \in A(k, c(\alpha, k$ $))\}$ would be a CS for $\tau$. Unfortunately, $P[A(k,c)]$ depends not only on $\tau$ but also on $\theta_0$, $\theta_1$. However, under $\tau = k$, $(S_k, S_n) = (\sum_1^k Y_i, \sum_1^n Y_i)$ is sufficient for $(\theta_0,\ \theta_1)$, so that $P_{\tau=k}[A(k,c)|S_k, S_n]$ is completely determined. This allows for the construction of a CS for $\tau$ by inverting the LRT's which are of size $\alpha$, conditionally given $(S_k, S_n)$.

Define $c(\alpha, k, \xi_1, \xi_2)$ such that

$$P_{\tau=k}[A(k, c(\alpha, k, \xi_1, \xi_2))|(S_k, S_n) = (\xi_1, \xi_2)] = 1 - \alpha. \tag{6}$$

On the conditioning set, using (5), (4) can be rewritten as

$$A(k, c) = \{\underline{y} : \max_r \Lambda_r(\underline{y}) \leq c^2 + \Lambda_k(\xi_1, \xi_2)\}.$$

Thus with $c^*(\alpha, k, \xi_1, \xi_2) = c^2(\alpha, k, \xi_1, \xi_2) + \Lambda_k(\xi_1, \xi_2)$, we rewrite (6) as

$$P_{\tau=k}[\max_r \Lambda_r(\underline{Y}) \leq c^*(\alpha, k, \xi_1, \xi_2)|(S_k, S_n) = (\xi_1, \xi_2)] = 1 - \alpha. \tag{7}$$

This provides an "in principle" construction of a CS for $\tau$, viz.,

$$C(\underline{y}) = \left\{k : \max_r \Lambda_r(\underline{y}) \leq c^*\left(\alpha, k, \sum_1^k y_i, \sum_1^n y_i\right)\right\}, \tag{8}$$

where $c^* = c^*(\alpha, k, \sum_1^k y_i, \sum_1^n y_i)$ is the $100(1-\alpha)$-percentile of the conditional cdf $F_k(\cdot|\sum_1^k y_i, \sum_1^n y_i)$ of $\max_r \Lambda_r(\underline{Y})$ given $(S_k, S_n) = (\sum_1^k y_i, \sum_1^n y_i)$ under $\tau = k$.

The actual evaluation of $c^*$ as a solution of (7) can be avoided by observing that $\max_r \Lambda_r(\underline{y}) \leq c^*$ is equivalent to

$$p(k, \underline{y}) = F_k\left(\max_r \Lambda_r(\underline{y})|\sum_1^k y_i, \sum_1^n y_i\right) \leq 1 - \alpha.$$

Moreover, note that under $\tau = k$, $T_1 = \max_{1 \leq r \leq k} \Lambda_r(\underline{Y})$ and $T_2 = \max_{k+1 \leq r \leq n-1} \Lambda_r(\underline{Y})$ are conditionally independent given $(S_k, S_n)$. Hence $p(k, \underline{y}) = p_1(k, \underline{y})p_2(k, \underline{y})$, where

$$p_i(k, \underline{y}) = P_{\tau=k}\left[T_i \leq \max_r \Lambda_r(\underline{y})|(S_k, S_n) = \left(\sum_1^k y_i, \sum_1^n y_i\right)\right].$$

The confidence set $C(\underline{y})$ given by (8) can now be written as

$$C(\underline{y}) = \{k : p_1(k, \underline{y})p_2(k, \underline{y}) \leq 1 - \alpha\}.$$

The main task in constructing this CS is to calculate $p_i(k, \underline{y})$, $i = 1, 2$ for each $k$ and for the observed $\underline{y}$. The boundary-crossing probabilities involved here are calculated by numerical integration and an asymptotic approximation is available for the case of change in mean of a normal distribution with known variance.

To avoid irregularities in $\Lambda_r(\underline{Y})$ at the very early and very late parts of the sequence, replacing $\max_{1 \leq r \leq n-1} \Lambda_r(\underline{Y})$ by $\max_{m \leq r \leq n-m} \Lambda_r(\underline{Y})$ is recommended, especially for large $n$, as in hypothesis testing and point estimation.

The CS's constructed above do not consist of consecutive integers in general, because $\Lambda_k(\underline{Y})$ does not decrease monotonically as $|k - \hat{\tau}|$ increases. Of course, one can construct confidence intervals $C^*(\underline{y}) = [k_1^*(\underline{y}), k_2^*(\underline{y}]$, where $k_1^*(\underline{y})$ and $k_2^*(\underline{y})$ are the smallest and the largest integers in the confidence set $C(\underline{y})$. This will satisfy the coverage probability condition conservatively.

*4.1. Bibliographic Notes.* Worsley (1986) constructed a confidence set of a change-point in exponential families. Asymptotic approximation of the boundary crossing probability needed in the case of normal means is due to Siegmund (1986). Some other approaches to confidence sets for change-points have been discussed by Siegmund (1988).

**5. Nonparametric Inference.** Let $Y_1, \ldots, Y_n$ be independent observations whose cdf changes from $F$ to $G$ after $Y_\tau$, where $\tau = n$ means "no change". So far we have only considered changes within a prescribed parametric family. In this section, we allow for more general types of changes from $F$ to $G$.

*5.1. Nonparametric Tests for the Hypothesis of "No Change".* Suppose that any possible change in distribution is a location shift. As in the parametric case, we consider the problem with known or unknown initial location. In the one-sided location shift alternative, the cdf changes from $F(\cdot)$ to $F(\cdot - \Delta)$, $\Delta > 0$, after $Y_\tau$ for some $1 \leq \tau \leq n-1$, where $F$ is unknown. When the initial location is known, we take it to be 0 and assume $F$ to be the cdf of a continuous distribution which is symmetric about 0. For unknown initial location, $F$ is an unknown continuous cdf. In each of these formulations, the problem of testing the null hypothesis of "no change" is invariant under certain transformations of the data.

When the initial distribution is continuous and symmetric about 0, the problem is invariant under all transformations

$$g_h(y_1, \ldots, y_n) = (h(y_1), \ldots, h(y_n)), \tag{9}$$

where $h$ is continuous, odd and strictly increasing. In this case, a maximal invariant is $(\underline{R}^+, \underline{S})$, where the $i$-th coordinates of $\underline{R}^+$ and $\underline{S}$ are $R_i^+ =$ rank of $|Y_i|$ and $S_i = sign(Y_i)$. All tests which are invariant under the above transformations are based on $(\underline{R}^+, \underline{S})$ and among these, we now find the test $\Psi$ which at a given level maximizes the average power $\bar{\beta}(\Psi; \Delta) = \sum_{i=1}^{n-1} q_i \beta(\Psi; \Delta, i)$ weighted by a prescribed set of weights $q_1, \ldots, q_{n-1} \geq 0$ with $\sum_1^{n-1} q_i = 1$ for all $\Delta \in (0, \epsilon)$, where $\epsilon$ is sufficiently small and $\beta(\Psi; \Delta, i)$ is the power of the test $\Psi$ against the alternative $H(\Delta, i)$ of a location shift of the amount $\Delta$ after $Y_i$.

Let $P_{\Delta,k}$ and $P_0$ denote respectively the probability distributions of $(\underline{R}^+, \underline{S})$ under $H(\Delta, k)$ and the null hypothesis. Under regularity conditions on $F$ (which includes the existence of an absolutely continuous pdf $f$), if $\Delta$ is sufficiently small, then the weighted likelihood ratios $\sum_{k=1}^{n-1} q_k P_{\Delta,k}(\underline{r}, \underline{s})/P_0(\underline{r}, \underline{s})$ for the $2^n \cdot n!$ points in the sample space of $(\underline{R}^+, \underline{S})$ are in the same ascending order as

$$T(\underline{s}, \underline{r}) = \sum_{k=1}^{n-1} q_k \sum_{i=k+1}^{n} s_i E[\phi^+(U_{n:r_i})] = \sum_{i=2}^{n} Q_i s_i E[\phi^+(U_{n:r_i})],$$

where $U_{n:i}$ is the $i$-th order statistic in a random sample of size $n$ from uniform $(0,1)$, $\phi^+(u) = -\frac{f' \circ F^{-1}}{f \circ F^{-1}}(\frac{1}{2} + \frac{1}{2}u)$ and $Q_i = q_1 + \ldots + q_{i-1}$. Hence the average power against local shift alternatives is maximized in the class of invariant tests by the test which rejects the null hypothesis for large values of $T(\underline{s}, \underline{r})$.

When the initial distribution is an arbitrary continuous cdf, the problem is invariant under all transformations $g_h$ of the form (9) where $h$ is continuous and strictly increasing. Now the vector $\underline{R}$ whose $i$-th coordinate $R_i$ is the rank of $Y_i$ is a maximal invariant and following the same line of argument as above, the average power against local shift alternatives is maximized in the class of tests based on $\underline{R}$ by rejecting the null hypothesis for large values of

$$T^*(\underline{r}) = \sum_{i=2}^{n} Q_i E[\phi(U_{n:r_i})],$$

where $Q_i$ and $U_{n:i}$ are as in the statistic $T(\underline{s}, \underline{r})$ and $\phi(u) = -\frac{f' \circ F^{-1}}{f \circ F^{-1}}(u)$.

The tests based on $T(\underline{s}, \underline{r})$ and $T^*(\underline{r})$ are obtained by a Bayesian approach analogous to the derivation of the tests based on $T_+$ and $T_+^*$ discussed in Section 2.1.

Nonparametric tests to detect changes other than location shift can be derived in the same manner.

*5.2. Estimation of Change-Point in a Nonparametric Model.* We now look at the problem of estimating $\tau$ after which the distribution of the sequence $Y_1, \ldots, Y_n$ changes from $F$ to $G$. Since we are going to be concerned with the asymptotics as $n \to \infty$, let $\tau = \tau_n = [n\theta]$, $0 < \theta < 1$, and consider the equivalent problem of estimating $\theta$. The main idea in all the methods in this area is to construct a measure of discrepancy between $(Y_1, \ldots, Y_{[nt]})$ and $(Y_{[nt]+1}, \ldots, Y_n)$ which estimates a function of $t$ on $(0,1)$ which has $\theta$ as its unique maximizer or minimizer. Notice the similarity between this and the minimization of the function $Q_{nr}$ in Section 3.2 and the role of $g_n(r)$ in that connection. In this section, we discuss methods of this nature at several levels of generality.

First suppose that $F, G$ are unknown continuous cdf's with $\int G\ dF - 1/2 = \lambda \neq 0$. Let

$$V_n(k) = \{k(n-k)\}^{-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \text{sign}(Y_i - Y_j),$$

rescale $V_n(\cdot)$ by defining

$$\xi_n(t) = V_n([nt]), \ \ t = 1/n, \dots, (n-1)/n,$$

and then extending $\xi_n(\cdot)$ to $[1/n, (n-1)/n]$ by linear interpolation. Consider the stochastic process $\{\xi_n(t), \ \ \alpha \leq t \leq 1-\alpha\}$ for some $0 < \alpha < 1/2$. Then

$$E[V_n(k)] = 1(k \leq \tau) \cdot 2(n-\tau)(n-k)^{-1} + 1(k > \tau) \cdot 2\tau k^{-1},$$

so that

$$\lim_{n \to \infty} E[\xi_n(t)] = 1(t \leq \theta) \cdot 2\lambda(1-\theta)(1-t)^{-1} + 1(t > \theta) \cdot 2\lambda\theta t^{-1}, \qquad (10)$$

which attains its unique maximum (minimum) at $\theta$ if $\lambda > 0 (\lambda < 0)$. It therefore seems reasonable to estimate $\theta$ by $\hat{\theta}_n$ which is a maximizer or a minimizer of $\xi_n(t)$ on the set $\alpha \leq t \leq 1-\alpha$, depending on whether $\lambda$ is known to be positive or negative. The following result shows that $\hat{\theta}_n$ is consistent and provides a rate of convergence.

Let $A_n$ and $B_n$ denote respectively the set of all minimizers and the set of all maximizers of $\xi_n(t)$ on $\alpha \leq t \leq 1-\alpha$. Then asymptotically, for every $\epsilon > 0, P_\theta[\max_{t \in B_n} |t-\theta| > \epsilon] = O(n^{-1})$ if $\lambda > 0$ and $P_\theta[\max_{t \in A_n} |t-\theta| > \epsilon] = O(n^{-1})$ if $\lambda < 0$.

Generalizing the above scheme, let

$$^t h_n(y) = (nt)^{-1} \sum_{1}^{nt} 1(Y_i \leq y) \ \text{ and } \ h_n^t(y) = (n-nt)^{-1} \sum_{nt+1}^{n} 1(Y_i \leq y) \quad (11)$$

denote the pre-t and post-t empirical cdf's respectively. In the construction of the above estimator, $\xi_n(t)$ is a particular measure of discrepancy between $^t h_n$ and $h_n^t$. A general class of measures of discrepancy can be defined by suitably combining the quantities $d_{ni}^t = |^t h_n(Y_i) - h_n^t(Y_i)|$, $1 \leq i \leq n$ to obtain

$$D_n(t) = \{t(1-t)\}^{1/2} S_n(d_{n1}^t, \dots, d_{nn}^t), \qquad (12)$$

where $S_n : \mathcal{R}^n \to \mathcal{R}$ satisfies $S_n(cy_1, \ldots, cy_n) = cS_n(y_1, \cdots, y_n)$ for $c \geq 0$ and $y_i \geq 0$. The factor $\{t(1-t)\}^{1/2}$ is introduced to control the variability of $S_n(\cdot)$, especially near $t = 0$ and 1. Note that

$$\begin{aligned} E[{}^t h_n(y)] &= {}^t h(y) = 1(t \leq \theta)F(y) + 1(t > \theta)t^{-1}\{\theta F(y) + (t-\theta)G(y)\}, \\ E[h_n^t(y)] &= h^t(y) = 1(t \leq \theta)(1-t)^{-1}\{(\theta - t)F(y) \\ &\quad + (1-\theta)G(y)\} + 1(t > \theta)G(y), \end{aligned} \tag{13}$$

and $d_{ni}^t$ estimates $\delta_{ni}^t = |{}^t h(Y_i) - h^t(Y_i)| = \rho(t)\delta_{ni}^\theta$, where

$$\rho(t) = 1(t \leq \theta)(1-\theta)(1-t)^{-1} + 1(t > \theta)\theta t^{-1}. \tag{14}$$

This $\rho(t)$ is the same as $(2\lambda)^{-1}\lim_{n\to\infty} E[\xi_n(t)]$ given by (10) and attains its unique maximum at $\theta$. Thus $D_n(t)$ given by (12) estimates

$$\begin{aligned} \Delta_n(t) &= \{t(1-t)\}^{1/2} S_n(\delta_{n1}^t, \ldots, \delta_{nn}^t) \\ &= \{t(1-t)\}^{1/2}\rho(t) S_n(\delta_{n1}^\theta, \ldots, \delta_{nn}^\theta), \end{aligned}$$

and $\{t(1-t)\}^{1/2}\rho(t)$ also attains its unique maximum at $\theta$. This motivates the estimator $\hat{\theta}_n = \arg\max D_n(t)$ over the set $\{1/n, \ldots, (n-1)/n\}$.

For functions $S_n(\underline{y})$ such as $n^{-1}\sum_1^n y_i$, $n^{-1}\sum_1^n y_i^2$, $\max_{1\leq i \leq n} y_i$ and in general, for a class of functions which are called "mean-dominant", the estimator $\hat{\theta}_n$ has the following properties provided that either $F$ or $G$ has positive probability on the set $\{y : F(y) \neq G(y)\}$:

(1) for arbitrary $\delta \in [0, 1/2)$, $n^\delta|\hat{\theta}_n - \theta| \to 0$ a.s.;

(2) there exist $C_1, C_2 > 0$ such that for $\epsilon > 0$ and $n \geq n(\epsilon)$,

$$P[|\hat{\theta}_n - \theta| > \epsilon] \leq C_1 n \exp[-C_2\epsilon^2 n].$$

The next level of generalization deals with observations taking values in an arbitrary space $(\mathcal{Y}, \mathcal{B})$ on which the probability changes from $P_n$ to $Q_n$ in a sequence of independent $Y_1, \ldots, Y_n$ after $\tau = n\theta$. We are actually considering a triangular array consisting of $\{Y_{ni}\}$ in which the change-point is $\tau_n = n\theta_n$. The pre-t and post-t empirical measures are now defined by

$${}^t P_n(B) = (nt)^{-1}\sum_1^{nt} \delta_{Y_i}(B),\ P_n^t(B) = (n - nt)^{-1}\sum_{nt+1}^{n} \delta_{Y_i}(B),$$

for $B \in \mathcal{B}$ where $\delta_y$ is the degenerate measure with unit mass at $y$, and their expected values are given by

$${}^t\Pi_n(B) = 1(t \leq \theta)P_n(B) + 1(t > \theta)t^{-1}\{\theta P_n(B) + (t-\theta)Q_n(B)\},$$

$$\Pi_n^t(B) = 1(t \leq \theta)(1-t)^{-1}\{(\theta - t)P_n(B) + (1-\theta)Q_n(B)\} + 1(t > \theta)Q_n(B)$$

respectively. Note that these formulas are obtained by substituting $\delta_{Y_i}(B)$, $P_n(B)$ and $Q_n(B)$ for $1(Y_i \leq y)$, $F(y)$ and $G(y)$ respectively in (11) and (13). Continuing this analogy, we see that $d_n^t(B) = \{t(1-t)\}^{1/2}[P_n^t(B) - {}^tP_n(B)]$ estimates $\delta_n^t(B) = \{t(1-t)\}^{1/2}\rho(t)[Q_n(B) - P_n(B)]$ where $\rho(t)$ is as in (14), so that $\theta$ is the unique maximizer of $\{t(1-t)\}^{1/2}\rho(t)$. This leads to the consideration of $\hat{\theta}_n = \arg\max S_n(\delta_n^t)$ over the set $\{1/n, \cdots, (n-1)/n\}$ as an estimator of $\theta$, where $S_n$ is a (possibly random) norm.

Suppose that $Q_n$ and $P_n$ are not too close to one another in the sense that $P[S_n(Q_n - P_n) > C\nu_n^{-1}] \to 1$ for a sequence $\{\nu_n\}$ satisfying $\nu_n(n^{-1}\log\log n)^{1/2} \to 0$. Then $\hat{\theta}_n - \theta = O_p(\nu_n^2 n^{-1})$ under a regularity condition on $S_n$. Results on limiting distribution of $\hat{\theta}_n$ for fixed change and contiguous change are also available. Notice the similarity of the rate of convergence of $\hat{\theta}_n$ with that of the MLE in the parametric model.

Another class of estimators in a similar spirit is based on

$$r_n(t) = n^{-1} \sum_{i=nt+1}^{n} \sum_{j=1}^{nt} K(Y_i, Y_j), \qquad 0 \leq t \leq 1$$

where $K$ is a bounded, measurable, antisymmetric kernel with

$$\lambda = \int\int K(x,y)d\,P(y)d\,Q(x) \neq 0.$$

Then for $0 \leq t \leq 1, r_n(t) \to \lambda r(t)$ a.s., where

$$r(t) = 1(t \leq \theta)(1-\theta)t + 1(t > \theta)\theta(1-t).$$

Like $\rho(t), r(t)$ also has its unique maximizer at $\theta$, which leads to the consideration of the maximizer (minimizer) of $r(t)$ over the set $\{1/n, \cdots, (n-1)/n\}$ when $\lambda$ is suspected to be $> 0$ ($< 0$). This estimator has an almost sure convergence rate of $n^{-1}\log n$.

*5.3. Bibliographic Notes.* The nonparametric tests for distributional change described above are due to Bhattacharyya and Johnson (1968). The estimator which maximizes the Mann-Whitney discrepancy between two parts of a series was proposed by Darkovskiy (1976). Carlstein (1988) generalized this estimator by maximizing other measures of discrepancy between pre-t and post-t empirical cdf's and Dümbgen (1991) extended this approach further by comparing pre-t and post-t empirical measures in an arbitrary space. Methods based on a U-statistic type comparison between pre-t and post-t empiricals have been proposed by Ferger (1991). An account of some nonparametric methods of change-point analysis can be found in Csörgö and Horváth (1988).

**6. Change-Points in a Regression Model.** There are two types of problems here, which we shall call Time-Varying Regression and Two-Phase Regression. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent random vectors. In the Time-Varying Regression model, we have

$$Y_i = \begin{cases} \beta_0 + \beta_1 X_i + \sigma Z_i, & i \le \tau \\ (\beta_0 + \Delta_0) + (\beta_1 + \Delta_1) X_i + \sigma Z_i, & i \ge \tau + 1 \end{cases} \tag{15}$$

where $\{X_i\}, \{Z_i\}$ are mutually independent iid sequences with $E(Z_i) = 0$, $E(Z_i^2) = 1$ and $(\Delta_0, \Delta_1) \ne (0, 0)$. If $1 \le \tau \le n-1$, then $\tau$ is a change-point and as before, $\tau = n$ means "no change". One can also consider the problem in its fixed design version where the regressors $x_1, \ldots, x_n$ are non-stochastic.

In a Two-Phase Regression model, the change in regression coefficients takes place not from the earlier to the later part of the observed sequence, but from the smaller to the larger values of the regressor (or more generally, according to the vector of regressors lying on one side or the other of a hyperplane). The change-point $\tau$ here is a point (or hyperplane) in the support of the $X$-distribution and in (15), the coefficient vector is $(\beta_0, \beta_1)$ for $X_i \le \tau$ and $(\beta_0 + \Delta_0, \beta_1 + \Delta_1)$ for $X_i > \tau$. In the Two-Phase Regression model with fixed design, $x_1, \ldots, x_{n\theta} \le \tau < x_{n\theta+1} < \cdots < x_n$ are non-stochastic and in (15), the coefficient vector changes after $i = n\theta$, so that either $\tau$ or $\theta$ can be regarded as the change-point.

We assume $Z_1, \ldots, Z_n$ to be independent $N(0,1)$, although much of the asymptotics of the procedures discussed here remain valid more generally.

*6.1. Time-Varying Regression.* For the most part, we shall consider this model for a fixed design (or argue conditionally if the regressors are stochastic). Suppose that there are $k$ regressors and let

$$\underline{x}_i' = (1, x_{i1}, \ldots, x_{ik}),\ \underline{\beta}' = (\beta_0, \beta_1, \ldots, \beta_k),$$

so that we have

$$Y_i = \underline{x}_i'\underline{\beta} + \sigma Z_i \text{ for } 1 \le i \le r \tag{16}$$

if and only if $r \le \tau$.

In the method of recursive residuals, we calculate the least squares estimate $\hat{\underline{\beta}}_{r-1}$ from $(\underline{x}_i', Y_i)$, $1 \le i \le r-1$, on the basis of (16) and then substitute this estimate for $\underline{\beta}$ in (16) for $i = r$ to calculate the residual $Y_r - \underline{x}_r'\hat{\underline{\beta}}_{r-1}$. Note that for $r \le \tau$, such a residual will behave like the true residual $\sigma Z_r$ plus an additional noise due to error in $\hat{\underline{\beta}}_{r-1}$ which is independent of $Z_r$.

For a more formal description of the method, let

$$X_r' = [\underline{x}_1, \ldots \underline{x}_r], \underline{Y}_r = (Y_1, \ldots, Y_r), \underline{Z}_r' = (Z_1, \ldots, Z_r)$$

and rewrite (16) as

$$\underline{Y}_r = X_r \underline{\beta}_r + \underline{Z}_r. \tag{17}$$

Assume that the $r \times (k+1)$ design matrix in (17) is of full rank for $r \geq k+1$. Then for $k+2 \leq r \leq \tau$, the residuals $Y_r - \underline{x}_r' \hat{\underline{\beta}}_{r-1}$ are easily seen to be independent normal variables with mean 0 and

$$\mathrm{Var}[Y_r - \underline{x}_r' \hat{\underline{\beta}}_{r-1}] = \sigma^2 V_r, V_r = 1 + \underline{x}_r' C_{r-1}^{-1} \underline{x}_r, C_{r-1} = X_{r-1}' X_{r-1}.$$

The recursive residuals defined as

$$\hat{Z}_r = V_r^{-1/2}(Y_r - \underline{x}_r' \hat{\underline{\beta}}_{r-1})$$

thus have the property of being independent $N(0, \sigma^2)$, as long as $r \leq \tau$.

In calculating the recursive residuals, the main difficulty is in inverting the matrix $C_{r-1} = X_{r-1}' X_{r-1}$ and calculating $\hat{\underline{\beta}}_{r-1}$ at each stage; but this is carried out by the recursion formulas:

$$\begin{aligned} C_r^{-1} &= C_{r-1}^{-1} - C_{r-1}^{-1} \underline{x}_r \underline{x}_r' C_{r-1}^{-1} / [1 + \underline{x}_r' C_{r-1}^{-1} \underline{x}_r], \\ \hat{\underline{\beta}}_r &= \hat{\underline{\beta}}_{r-1} + C_r^{-1} \underline{x}_r (Y_r - \underline{x}_r' \hat{\underline{\beta}}_{r-1}). \end{aligned}$$

Moreover, the successive residual sums of squares

$$RSS(r) = (\underline{Y}_r - X_r \hat{\underline{\beta}}_r)'(\underline{Y}_r - X_r \hat{\underline{\beta}}_r), \qquad r \geq k+2$$

are also obtained recursively as

$$RSS(r) = RSS(r-1) + \hat{Z}_r^2.$$

Under the hypothesis $H_0$ of "no change", $\hat{\sigma}^2 = RSS(n)/(n-k-1)$ is a consistent estimator of $\sigma^2$, and the sequence $\{W_t = \hat{\sigma}^{-1} \sum_{j=k+2}^{t} \hat{Z}_j,\ k+2 \leq t \leq n\}$ behaves approximately as a standard B.M. starting from 0 at time $t = k+1$. Motivated by this fact, a test for $H_0$ has been proposed in which $H_0$ is rejected if $\{W_t\}$ crosses a suitably constructed pair of linear boundaries.

Another proposal is to construct a test of $H_0$ based on the cumulative sum of squares of $\{W_t\}$. Let

$$Q_t = \sum_{j=k+2}^{t} \hat{Z}_j^2 / \sum_{j=k+2}^{n} \hat{Z}_j^2 = RSS(t)/RSS(n), \qquad k+2 \leq t \leq n.$$

For simplicity, let $n-k-1 = 2(N+1)$ be even and note that $T_j = (\hat{Z}_{k+2j}^2 + \hat{Z}_{k+2j+1}^2)/2$, $1 \leq j \leq N+1$ are independent Exp(1) under $H_0$. Hence under

$H_0$,

$$Q_{k+1+2i} = \sum_{j=1}^{i} T_j / \sum_{j=1}^{N+1} T_j, \qquad 1 \le i \le N$$

are jointly distributed as the order statistics $(U_{N:1}, \dots, U_{N:N})$ in a random sample from uniform $(0,1)$. Thus

$$D_N = \max_{1 \le i \le N} N^{1/2} |Q_{k+1+2i} - iN^{-1}|$$

can be used to test $H_0$, using the distribution of the empirical process as in the Kolmogorov-Smirnov test.

We now briefly discuss the MLE of $\tau$ in the time-varying regression model with stochastic regressors. For simplicity, we consider the case of $k = 1$ and assume that in (15), $\{X_i\}$ is a sequence of independent $N(\theta,\ \sigma_0^2)$ which is independent of $\{Z_i\}$. Suppose that the change in the regression parameters is contiguous. Specifically, let $(\Delta_0,\ \Delta_1) = (\delta_0 \nu_n^{-1},\ \delta_1 \nu_n^{-1})$, where $\nu_n \to \infty$ at a slower rate than $n^{1/2}$. Then from the general results for MLE of change-points in multi-parameter families, the asymptotic distribution of the MLE $\hat{\tau}$ of $\tau$ is obtained as

$$\nu_n^{-2}(\hat{\tau} - \tau) \xrightarrow{\mathcal{L}} T_{\sigma^{-1}\{(\delta_0 + \theta\delta_1)^2 + \sigma_0^2 \delta_1^2\}^{1/2}}$$

where $T_\mu = \arg\min[B(t) + \frac{1}{2}\mu|t|]$ as in Section 3.2. Note that $(\beta_0,\ \beta_1)$ changing under a linear constraint (in particular, $\beta_0$ changing with $\beta_1$ remaining unchanged) is included in this result.

*6.2. Two-Phase Regression.* We shall discuss the estimation problem under this model with fixed design involving one regressor. There are two cases to consider, depending on whether or not the regression lines in the two phases are required to meet at the change-point. We shall refer to these as the restricted case and the unrestricted case. Most of the literature in this area is devoted to the restricted case in which the MLE of the change-point is asymptotically normally distributed. However, the MLE of change-point behaves differently in the unrestricted case.

We first introduce some conditions on the limiting configuration of the design points. Let $F_n(x) = n^{-1} \sum_1^n 1(x_i \le x)$ and suppose that there is a function $F$ with $F(\tau) = \theta \in (0,\ 1)$ and $f(\tau) = F'(\tau) > 0$ such that as $n \to \infty$, $F_n(\tau) \to F(\tau)$ and $F_n(x) - F_n(\tau) = [F(x) - F(\tau)] \cdot [1 + o(1)]$ uniformly in a neighborhood of $\tau$. Suppose further that as $n \to \infty$,

$$(n\theta)^{-1} \sum_1^{n\theta} \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} \to \begin{bmatrix} 1 & \mu_1 \\ \mu_1 & \mu_1^2 + \sigma_1^2 \end{bmatrix} = A_1' A_1,$$

$$(n-n\theta)^{-1}\sum_{n\theta+1}^{n}\begin{bmatrix}1 & x_i\\ x_i & x_i^2\end{bmatrix}\to\begin{bmatrix}1 & \mu_2\\ \mu_2 & \mu_2^2+\sigma_2^2\end{bmatrix}=A_2'A_2, \tag{18}$$

where $\sigma_1^2$ and $\sigma_2^2$ are positive, and that $\max_{1\le i\le n}|x_i| = o(n^{1/2})$. In this limiting scheme, $x_{n\theta+1} - x_{n\theta} = \{nf(\tau)\}^{-1}[1+o(1)]$, so we can take $x_{n\theta} = \tau$ without affecting the asymptotics.

The MLE of the regression parameters $(\beta_0,\ \beta_1)$, $(\gamma_0,\ \gamma_1) = (\beta_0+\Delta_0,\ \beta_1+\Delta_1)$ and $n\theta$ are obtained as the minimizer of the loglikelihood ratio process

$$\begin{aligned}\xi_n(\varphi_0,\varphi_1,\psi_0,\psi_1,k) &= R_n(\varphi_0,\varphi_1,\psi_0,\psi_1,k) - R_n(\beta_0,\beta_1,\gamma_0,\gamma_1,n\theta)\\ R_n(\varphi_0,\varphi_1,\psi_0,\psi_1,k) &= \sigma^{-2}\sum_{1}^{k}(Y_i-\varphi_0-\varphi_1 x_i)^2\\ &\quad+\sigma^{-2}\sum_{k+1}^{n}(Y_i-\psi_0-\psi_1 x_i)^2.\end{aligned} \tag{19}$$

We now examine $\xi_n(\varphi_0,\varphi_1,\psi_0,\psi_1,k)$ near the true parameter values, i.e., at $\varphi_0 = \beta_0 + u_0 n^{-1/2}$, $\varphi_1 = \beta_1 + u_1 n^{-1/2}$, $\psi_0 = \gamma_0 + v_0 n^{-1/2}$, $\psi_1 = \gamma_1 + v_1 n^{-1/2}$ and $k = n\theta + j_n$. The difference between the restricted case and the unrestricted case lies in the role of $j_n$ in this minimization.

In the restricted case, the parameters $\beta_0,\beta_1,\gamma_0,\gamma_1,\tau$ must satisfy $\tau = (\gamma_0-\beta_0)/(\beta_1-\gamma_1) = -\Delta_0/\Delta_1$ with $\Delta_1\ne 0$. Hence the parameter space near the true parameters consists of $(un^{-1/2},\ vn^{-1/2},\ \tau_n(u,v))$ where $u' = (u_0,u_1)$, $v' = (v_0,v_1)$ and

$$\tau_n(u,v) = \{\tau + (n^{1/2}\Delta_1)^{-1}(u_0-v_0)\}/\{1-(n^{1/2}\Delta_1)^{-1}(u_1-v_1)\}.$$

In the minimization of $\xi_n(\cdot)$ given by (19), we therefore take $k = n\theta + j_n(u,v)$, where $j_n(u,v)$ is defined by $x_{n\theta+j_n(u,v)} = \tau_n(u,v)$. It now follows that $j_n(u,v) = O(n^{1/2})$ and $|x_i-\tau| = O(n^{-1/2})$ for $|i-n\theta|\le j_n(u,v)$.

We now consider the loglikelihood ratio process

$$\xi_n^*(u,v) = \xi_n(\beta_0+u_0n^{-1/2},\beta_1+u_1n^{-1/2},\gamma_0+v_0n^{-1/2},\gamma_1+v_1n^{-1/2},n\theta+j_n(u,v)).$$

Due to the order of magnitude of $j_n(u,v)$, the asymptotics of $\xi_n^*(u,v)$ becomes simplified and with respect to uniform convergence on compact sets:

$$\begin{aligned}\xi_n^*(u,v) \xrightarrow{w} &-(\|W_1\|^2+\|W_2\|^2)+\|\theta^{1/2}\sigma^{-1}A_1u-W_1\|^2\\ &+\|(1-\theta)^{1/2}A_2v-W_2\|^2,\end{aligned} \tag{20}$$

where $W_1$ and $W_2$ are independent bivariate normal vectors with mean vector 0 and covariance matrix I, and $A_1$ and $A_2$ are as in (18).

In the unrestricted case, the loglikelihood ratio has another term which is a two-sided random walk with drifts as in Section 3.2. For contiguous change in the regression parameters with $(\Delta_0, \Delta_1) = (\delta_0\nu_n^{-1}, \delta_1\nu_n^{-1})$, $\nu_n \to \infty$ slower than $n^{1/2}$, and with $k = n\theta + \nu_n^2 t$ in (19), the loglikelihood ratio has the following weak limit with respect to uniform convergence on compact sets:

$$\begin{aligned}\xi_n^*(u,v,t) &= \xi_n(\beta_0 + u_0 n^{-1/2}, \beta_1 + u_1 n^{-1/2}, \gamma_0 + v_0 n^{-1/2}, \gamma_1 + v_1 n^{-1/2}, n\theta + \nu_n^2 t) \\ &\xrightarrow{w} -(\|W_1\|^2 + \|W\|^2) + \|\theta^{1/2}\sigma^{-1}A_1 u - W_1\|^2 + \|(1-\theta)^{1/2}\sigma^{-1}A_2 v - W_2\|^2 \\ &\quad + \sigma^{-1}|\delta_1 + \delta_2\tau| \cdot \left[B(t) + \frac{1}{2}\sigma^{-1}|\delta_1 + \delta_2\tau| \cdot |t|\right],\end{aligned}$$

where $W_1, W_2, A_1, A_2$ are as in (20) and $B(\cdot)$ is a standard two-sided B.M. which is independent of $W_1, W_2$.

These weak convergence results indicate that in the restricted case, $n^{1/2}(\hat\beta_0 - \beta_0, \hat\beta_1 - \beta_1)$ and $n^{1/2}(\hat\gamma_0 - \gamma_0, \hat\gamma_1 - \gamma_1)$ are asymptotically independently distributed as $\theta^{-1/2}\sigma A_1^{-1}W_1$ and $(1-\theta)^{-1/2}\sigma A_2^{-1}W_2$ respectively. Consequently,

$$n^{1/2}(\hat\tau - \tau) = n^{1/2}\left(\frac{\hat\beta_0 - \hat\gamma_0}{\hat\gamma_1 - \hat\beta_1} - \frac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1}\right) \xrightarrow{\mathcal{L}} N(0, \alpha^2),$$

where the asymptotic variance $\alpha^2$ is obtained by the delta method.

In the unrestricted case, the estimators of the regression parameters behave in the same way as in the restricted case, but $\hat\tau$ is asymptotically independent of $\hat\beta_0, \hat\beta_1, \hat\gamma_0, \hat\gamma_1$ and

$$n\nu_n^{-2}(\hat\tau - \tau) \xrightarrow{\mathcal{L}} T^*/f(\tau)$$

where $T^* = \arg\min[B(t) + \frac{1}{2}\sigma^{-1}|\delta_0 + \delta_1\tau| \cdot |t|]$. Thus the limiting distribution of $\hat\tau$ is governed by the limiting density $f(\tau)$ of the design points at $\tau$, the jump $\Delta_0 + \Delta_1\tau = \nu_n^{-1}(\delta_0 + \delta_1\tau)$ in the regression at $\tau$, and the residual s.d. $\sigma$.

Similar results hold for the case of random designs.

*6.3. Bibliographic Notes.* Change-point analysis in a time-varying regression model was initiated by Quandt (1958, 1960). The method of recursive residuals to test for change in this model was developed by Brown, Durbin and Evans (1975). Estimation in segmented or multi-phase linear regression with unknown change-points and under continuity constraint was considered by Hudson (1966) and the asymptotic distribution of the estimated change-point was derived by Hinkley (1969). The problem was also treated in fairly general terms by Feder (1975). Distinction between the asymptotic properties of the estimated change-point with or without the continuity constraint was recognized by van de Geer (1988) in the random-design case and by Bhattacharya (1991) in the fixed-design case. Approximation to the tail probability of the LRT statistic for a change in only the intercept or in both slope and intercept was obtained by Kim and Siegmund (1989), who also constructed confidence sets for the parameters.

**7. Detecting Change in the Distribution of Sequentially Observed Data.** This is the on-line quality control problem in which manufactured items are sampled at regular intervals from a production process and an observation $X$ (possibly binary) is taken on each sampled item. This gives rise to independent random variables $\{X_i\}$, observed sequentially, whose distribution may change at any time $\tau$ (including $\tau = 1$) and the production process "goes out of control" when that happens. The problem is to stop as soon after $\tau$ as possible without too many false alarms.

The earliest techniques in this area are the Shewhart control charts: $\bar{X}$-chart to detect change in mean, R-chart or $s^2$-chart to detect change in variability and control chart for fraction detectives. In these charts, the mean $\bar{X}$, the range R, etc., of the samples are monitored and the process is stopped if and when some action limits are violated. These charts use only the current sample and fail to use accumulated evidence of change, thereby missing the sequential aspect of the problem. In an attempt to remedy this, one can replace the $\bar{X}$-chart by a moving average chart, using simple moving averages of several past observations or exponentially weighted (geometric) moving averages of all of the past observations. Average run lengths of moving average charts have been studied theoretically to derive bounds, and by simulation studies.

A method for detecting change in distribution of sequentially observed data $\{X_i\}$ is a stopping time. We shall discuss here a popular method known as Page-CUSUM procedure and its Bayesian counterpart known as Shiryaev-Roberts procedure.

The aim of the Page-CUSUM procedure is to detect change from a known pdf $f_0$ to another known pdf $f_1$. For this, at each stage of sampling, the null hypothesis of "no change yet" is tested against the alternative of change having occurred at 'some point in the past'. Specifically, at the $k$-th stage, the likelihood ratio test statistic for this purpose is:

$$T_k = \max\left\{\max_{1\le j\le k}\sum_{i=j}^{k} Z_i,\ 0\right\},\quad Z_i = \log[f_1(X_i)/f_0(X_i)],$$

which can be calculated by the recursion formula:

$$T_k = (T_{k-1} + Z_k)^+, \quad T_0 = 0. \tag{21}$$

The Page-CUSUM procedure stops at $N_P = \min\{k : T_k \geq \gamma\}$, where $\gamma$ is a prescribed boundary.

Note that this procedure can be thought of as the Wald SPRT($A$, $B$) for $f_0$ against $f_1$ with $\log B = 0$ and $\log A = \gamma$ which stops if the SPRT terminates with acceptance of $f_1$ but repeats from scratch if the SPRT terminates with acceptance of $f_0$. In particular, if $f_i$ is the pdf of $N(\mu_i, \sigma^2)$, $i = 0, 1$ with $\mu_1 > \mu_0$, then the Page-CUSUM procedure calculates

$$S_k = \max\left\{S_{k-1} + X_k - \frac{1}{2}(\mu_0 + \mu_1),\ 0\right\}, \quad S_0 = 0,$$

at each stage of sampling and stops at the first $k$ for which $S_k$ exceeds a prescribed boundary.

Let $E_\tau$ denote expectation under the hypothesis of "change from $f_0$ to $f_1$ after time $\tau$" and let $E_0$ denote expectation under the hypothesis of "no change from $f_0$". Then for a stopping time $N$, it would be desirable to have a large value of $E_0(N)$ so that false alarms are raised only at long intervals and to have

$$\bar{E}_1(N) = \sup_{\tau \geq 1} ess \sup E_\tau[(N - \tau + 1)^+ | X_1, \ldots, X_{\tau-1}]$$

to be small in order to reduce the delay $(N - \tau + 1)^+$ in detecting changes in a minimax sense. The Page-CUSUM stopping time $N_P = \min\{k : T_k \geq \gamma\}$ with $T_k$ given by (21) is minimax in the sense of minimizing $\bar{E}_1(N)$ subject to $E_0(N) \geq c$ for a preassigned $c > 1$.

In a Bayesian formulation of the problem, suppose that the prior distribution of $\tau$ is Geometric ($p$) and the loss in stopping at time $t$ is 1 if $t \leq \tau$ (penalty for false alarm) and is $c(t - \tau)$ if $t > \tau$ (penalty for delay), then the Bayes rule is to stop as soon as the posterior probability of a change having occurred exceeds a threshold. In the limit as $p \to 0$, the Bayes stopping rule takes the form:

$$N = \min\left\{k : \sum_{j=1}^{k} \prod_{i=j}^{k} \frac{f_1(X_i)}{f_0(X_i)} \geq A\right\}.$$

Such a procedure is called a Shiryaev-Roberts procedure.

*7.1. Bibliographic Notes.* After the introduction of the original control charts by Shewhart (1931), various modifications were introduced such as the use of warning lines (a succession of less serious departures from the norm calling for action) and moving averages to use past evidence together with current information. The sequential nature of the problem was recognized by Girshick and Rubin (1952) in a Bayesian formulation and in the CUSUM chart of Page (1954) motivated by the likelihood approach. The Bayesian procedure described here is due to Shiryaev (1963). This procedure was independently proposed by Roberts (1966) in a comparative study of various methods. Shiryaev (1963, 1978) showed that the Bayes rule stops as soon as the posterior probability of a change having already occurred exceeds a threshold. The minimax property of the Page-CUSUM procedure was first established in an asymptotic sense by Lorden (1971) and later, the exact minimax property was proved by Moustakides (1986). The asymptotic property of the Shiryaev-Roberts procedure mentioned above is due to Pollak (1985). Pollak and Siegmund (1985) compared the Page-CUSUM and the Shiryaev-Roberts procedures in the continuous time analogue of the problem, and found that neither of the two procedures is decidedly better than the other when the change had occurred before observations started or when the change occurs very late.

## REFERENCES

BHATTACHARYYA, G. K. and JOHNSON, R. A. (1968). Nonparametric tests for shift at an unknown time point. *Ann. Math. Statist.* **39**, 1731–1743.

BHATTACHARYA, P. K. and BROCKWELL, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrsch. verw. Gebiete* **37**, 51–75.

BHATTACHARYA, P. K. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case. *J. Multivariate Anal.* **23**, 183–208.

BHATTACHARYA, P. K. (1991). Weak convergence of the log likelihood process in the two-phase linear regression problem. In: *Probability Statistics and Design of Experiments.* 145–156. Wiley Eastern.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37**, 149–192.

CARLSTEIN, E. (1988). Nonparametric change-point estimation. *Ann. Statist.* **16**, 188–197.

CHERNOFF, H. and ZACKS, S. (1964) Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* **35**, 999–1028.

COBB, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika* **65**, 243–251.

CSÖRGŐ, M. and HORVÁTH, L. (1988). Nonparametric methods for changepoint problems. In: *Handbook of Statistics,* **7** 403–425. North Holland.

DARKOVSKIY, B. S. (1976). A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Probab. Appl.* **21**, 178–183.

DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric changepoint estimators. *Ann. Statist.* **19**, 1471–1495.

FEDER, P. I. (1975). On asymptotic distribution theory in segmented regression problems: Identified case. *Ann. Statist.* **3**, 49–83.

FERGER, D. (1991). *Nonparametric Changepoint-Detection Based on U-Statistics.* Ph.D. Dissertation. Giessen.

GARDNER, L. A., JR. (1969). On detecting changes in the mean of normal variates. *Ann. Math. Statist.* **40**, 116–126.

GIRSHICK, M. A. and RUBIN, H. (1952). A Bayes approach to a quality control model. *Ann. Math. Statist.* **23**, 114–125.

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.* **72**, 180–186.

HINKLEY, D. V. (1969). Inference about intersection in the two-phase regression problem. *Biometrika* **56**, 495–504.

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1–17.

HINKLEY, D. V. (1972). Time-ordered classification. *Biometrika* **59**, 509–523.

HUDSON, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *J. Amer. Statist. Assoc.* **61**, 1097–1129.

IBRAGIMOV, I. A. and HAS'MINSKII, R. (1981). *Statistical Estimation, Asymptotic Theory.* Springer-Verlag.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statist. Sinica* **2**, 69–90.

KANDER, A. and ZACKS, S. (1966). Test procedure for possible changes in parameters of statistical distribution occurring at unknown time points. *Ann. Math. Statist.* **37**, 1196–1210.

KIM, H.-J. and SIEGMUND, D. (1989). The likelihood ratio test for a changepoint in simple linear regression. *Biometrika* **76**, 409–423.

LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.***42**, 1897–1908.

MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14**, 1379–1387.

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115.

PETTIT, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika* **67**, 79–84.

POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13**, 206–227.

POLLAK, M. and SIEGMUND, D. (1985). A diffusion process and its application to detecting a change in the drift of a Brownian motion. *Biometrika* **72**, 267–280.

QUANDT, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *J. Amer. Statist. Assoc.* **50**, 873–880.

QUANDT, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.* **55**, 324–330.

ROBERTS, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8**, 411–430.

SEN, A. and SRIVASTAVA, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3**, 98–108.

SHABAN, S. A. (1980). Change-point problem and two-phase regression: An annotated bibliography. *Internat. Statist. Rev.* **48**, 83-93.

SHEWHART, W. A. (1931). *The Economic Control of the Quality of a Manufactured Product.* Van Nostrand.

SHIRYAEV, A. N. (1963). On optimal method in earliest detection problems. *Theory Probab. Appl.* **8**, 26–51.

SHIRYAEV, A. N. (1978). *Optimal Stopping Rules.* Springer-Verlag.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14**, 361–404.

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Internat. Statist. Rev.* **56**, 31–48.

SMITH, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**, 407–416.

VAN DE GEER, S. A. (1988). *Regression Analysis and Empirical Processes.* CWI Tract 45. Amsterdam.

WORSLEY, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika* **73**, 91–104.

YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett.* **6**, 181–189.

ZACKS, S. (1983). Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures in testing and estimation. In: *Recent Advances in Statistics*, 245–269. Academic Press.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616