Statistica Sinica **5**(1995), 319-331

A LOWER BOUND FOR ERROR PROBABILITY IN CHANGE-POINT ESTIMATION

Inchi Hu and Andrew L. Rukhin

Hong Kong University of Science & Technology and UMBC

Abstract. In the classical setting of the change-point estimation problem, where there is no consistent procedure, a lower bound on the limit of the maximum of error probabilities is established. This bound is attained by the maximum likelihood estimator when the two probability distributions before and after the change-point are known. The minimaxity of the maximum likelihood procedure in the sense of attaining the mentioned bound is proved for observations from an exponential family.

Key words and phrases: Change-point problem, error probability, exponential family, maximum likelihood procedure, minimaxity.

1. Introduction and Summary

In this paper the classical setting of the change-point estimation problem which has been studied by a number of authors (see Hinkley (1970), Cobb (1978)) is considered. It is well known that in this setting there is no consistent estimator of the change-point so that to study the asymptotic efficiency the setting is usually modified to allow the distributions to depend on the sample size in some fashion (Carlstein (1988), Ritov (1990)).

In Section 2, we establish a lower bound on the limit of the largest of error probabilities in estimating the change-point. This bound is attained by the maximum likelihood estimator when the two probability distributions before and after the change-point are completely known.

When the observations are from an exponential family it is shown in Section 3 that the maximum likelihood procedure is asymptotically minimax in that it attains the lower bound. The proof of this result reveals that the likelihood ratio process behave locally like random walks under both conditional and unconditional probability laws. (The proof make use of a result from Hu (1991) concerning conditional random walks). This fact helps to get a better insight into the performance of likelihood based procedures in the change-point problem and may be useful in deriving other results for procedures using the likelihood ratio method.

Numerical results in the case of normal observations with unknown means are also reported. These results give evidence of the accuracy of the lower bound for moderate sample sizes.

2. A Lower Bound for Error Probabilities

Let F and G be two different distribution functions with densities f and g and assume that the observed data

$$\mathbf{X} = \{X_1, \ldots, X_\nu, X_{\nu+1}, \ldots, X_m\},$$

consists of two independent parts, $\{X_1, \ldots, X_\nu\}$ being a random sample from distribution F, and the second random sample $\{X_{\nu+1}, \ldots, X_m\}$ coming from distribution G. In other words ν is the change-point, the parameter of interest. It is known (cf Hinkley (1970)) that there is no consistent estimator of ν. In our approach, asymptotic efficiency is defined by means of the error probabilities, which do not tend to zero as sample size increases, but which satisfy the inequality (2.1) in Lemma 1 analogous to the known one in the classical multiple decision problem (see Krafft and Puri (1974)).

Let $\hat{\delta}$ denote the maximum likelihood estimator, i.e.

$$\hat{\delta}\,(\mathbf{X}) = \arg\max\{\sum_{j=1}^{k} \log f(X_j) + \sum_{j=k+1}^{m} \log g(X_j)\} = \arg\max\left\{\sum_{j=1}^{k} \log \frac{f(X_j)}{g(X_j)}\right\}.$$

We accept here the usual convention that when the maximizer in this formula is not defined uniquely, the smallest value is chosen.

Lemma 1. *If both $m - \nu \to \infty$ and $\nu \to \infty$ as $m \to \infty$, then for any estimator $\delta = \delta(\mathbf{X})$*

$$\liminf \max_{\nu} P\Big(\delta(\mathbf{X}) \neq \nu\Big) \geq \lim \max_{\nu} P\Big(\hat{\delta}(\mathbf{X}) \neq \nu\Big) = 1 - \exp\{-d(F,G)\}, \tag{2.1}$$

where

$$d(F,G) = \sum_{k=1}^{\infty} k^{-1} \left[P(\sum_1^k Y_j \geq 0) + P(\sum_1^k Z_j > 0)\right].$$

Here $Y_j = \log g(X_j) - \log f(X_j)$ with X_j having the distribution F and $Z_j = \log f(X_j) - \log g(X_j)$ with X_j having the distribution G.

Proof of Lemma 1. The proof of Inequality (2.1) is based on the following fact. The maximum likelihood estimator is the Bayes estimator against the uniform prior for ν under the zero-one loss function. Therefore

$$\sum_{i=1}^{m} P\Big(\delta(\mathbf{X}) = i\Big) \leq \sum_{i=1}^{m} P\Big(\hat{\delta}(\mathbf{X}) = i\Big)$$

so that

$$\liminf_{m\to\infty} \min_{1\le i<m} P\Big(\delta(\mathbf{X})=i\Big) \le \lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} P\Big(\hat\delta(\mathbf{X})=i\Big).$$

One has

$$P\Big(\hat\delta(\mathbf{X})=i\Big) = P\Big(\sum_1^k Y_j \le 0, k=1,\dots,i-1, \sum_i^n Z_j < 0, n=i+1,\dots,m\Big)$$

$$= P\Big(\sum_1^k Y_j \le 0, k=1,\dots,i-1\Big) P\Big(\sum_1^k Z_j < 0, k=1,\dots,m-i\Big) = p_{i-1}q_{m-i}.$$

According to known results of random walks theory (cf. Siegmund (1985, Cor. 8.44) or Woodroofe (1982, Cor. 2.4)) as $i\to\infty$

$$p_i \to p = \exp\{-\sum_{k=1}^{\infty} k^{-1} P(\sum_1^k Y_j \ge 0)\}$$

and a similar formula holds for q_{m-i}. Inequality (2.1) now follows from the fact that if the sequences of positive numbers p_i and q_i converge to limits p and q respectively then

$$m^{-1}\sum_{i=1}^{m} p_{i-1}q_{m-i} \to pq$$

(cf. Knopp (1956, Sec. 2.4)).

The quantity $d(F,G)=d(G,F)$ defined by (2.1) provides a new "information -type" divergence between distributions F and G. Indeed, as is easy to see, $d(F,F)=\infty$, and $d(F,G)=0$, if F and G are singular.

If F and G are two normal distributions with the same, say, unit variance and means θ_1 and θ_2, then

$$d(F,G) = d_o(\Delta) = 2\sum_{k=1}^{\infty} k^{-1}\Phi(-\Delta\sqrt{k}) \tag{2.2}$$

with $\Delta = 0.5 \mid \theta_1-\theta_2 \mid$ and Φ denoting the standard normal distribution function. This function plays an important role in sequential analysis and renewal theory (see Siegmund (1985)). Its values are tabulated in Woodroofe (1982, p. 33). Some other properties of $d(F,G)$ are given in Rukhin (1994).

Lemma 1 shows that the maximum likelihood estimator is asymptotically minimax. Now we give an example of an estimator which is not asymptotically minimax, i.e. for which the inequality in Lemma 1 is strict. Pettitt (1980) suggested the following cusum type statistic for the estimation of the change point in a sequence of Bernoulli distributed observations, namely,

$$\tilde{\nu}(\mathbf{X}) = \arg\min \left\{ \sum_{j=1}^{k} \log \frac{g(X_j)}{f(X_j)} - \frac{k}{m} \sum_{j=1}^{m} \log \frac{g(X_j)}{f(X_j)} \right\}.$$

In James, James and Siegmund (1987) this statistic is studied for arbitrary distributions.

Assume that $\nu/m \to \rho$ with $0 < \rho < 1$ such that

$$\lambda = (1-\rho)E^G \log \frac{g(X)}{f(X)} - \rho E^F \log \frac{f(X)}{g(X)}.$$

Then almost surely as $m \to \infty$

$$T = T_m = \frac{1}{m} \sum_{j=1}^{m} \log \frac{g(X_j)}{f(X_j)} \to \lambda$$

and in the notation of Lemma 1

$$\begin{aligned} &P\left(\tilde{\nu}(\mathbf{X}) = \nu\right) \\ &= P\Big(\sum_1^k Y_j \le kT, k = 1, \ldots, \nu - 1, \sum_\nu^n Z_j < -(n - \nu + 1)T, n = \nu + 1, \ldots, m\Big) \\ &= E\Big[P\Big(\sum_1^k Y_j \le kT, k=1, \ldots, \nu - 1 | T\Big) P\Big(\sum_1^k Z_j < -kT, k=1, \ldots, m - \nu | T\Big)\Big]. \end{aligned}$$

As in the proof of Lemma 1 one obtains

$$\begin{aligned} P\left(\tilde{\nu}(\mathbf{X}) = \nu\right) &\to P\Big(k^{-1} \sum_1^k Y_j \le \lambda,\ k = 1, \ldots\Big) P\Big(k^{-1} \sum_1^k Z_j < -\lambda,\ k = 1, \ldots\Big) \\ &= \exp\left\{ -\sum_{k=1}^{\infty} k^{-1} [P(\sum_1^k Y_j \ge k\lambda) + P(\sum_1^k Z_j > -k\lambda)] \right\}. \end{aligned}$$

The quantity in the right-hand side of the last formula is maximized when $\lambda = 0$. Indeed for any k

$$P(\sum_1^k Y_j \ge k\lambda) + P(\sum_1^k Z_j > -k\lambda) \ge P(\sum_1^k Y_j \ge 0) + P(\sum_1^k Z_j > 0)$$

with strict inequality if $P(0 < \sum_1^k Y_j \le k|\lambda|) > 0$. This demonstrates the fact that $\tilde{\nu}(\mathbf{X})$ is not asymptotically minimax.

For example, when F and G are Bernoulli distributions with probabilities p_1 and p_2 respectively, one has

$$\lambda = [(1-\rho)p_2 + \rho p_1] \log \frac{p_2}{p_1} + [(1-\rho)(1-p_2) + \rho(1-p_1)] \log \frac{1-p_2}{1-p_1}$$

and

$$P(\sum_1^k Y_j \geq k\lambda) = P\left(n \log \frac{1-p_1}{1-p_2} + n|\lambda| > B_k \log \frac{p_2(1-p_1)}{p_1(1-p_2)} > n \log \frac{1-p_1}{1-p_2}\right),$$

where B_k is binomial random variable with parameters k and p_1. The function $d(F, G)$ for these distributions is tabulated in Rukhin (1995).

James, James and Siegmund (1987) studied five inference procedures for the change-point. They came to the conclusion that a modified maximum likelihood procedure and Pettitt's rule $\tilde{\nu}(\mathbf{X})$ were the best with the modified maximum likelihood procedure being slightly better than $\tilde{\nu}(\mathbf{X})$. Indeed, numerical results in James, James and Siegmund (1987) indicate that Pettitt's rule does not behave well when the change-point is close to the boundary. Our conclusion that $\tilde{\nu}(\mathbf{X})$ is not asymptotically minimax confirms the finding of these authors and adds new dimension to the comparison of these two alternative approaches.

3. Asymptotic Minimaxity of the Maximum Likelihood Estimator in Exponential Families

In this section we study the change-point estimation problem in an exponential family. Specifically we assume that F and G are members of an exponential family of the form

$$dF_\theta(x) = \exp\{\theta x - \psi(\theta)\} dF_0(x),$$

relative to some non-degenerate distribution function F_0. We shall treat only real x and θ, although the extension to the multivariate case is straightforward. However, the case of vector θ, where only part of the components change at ν, is considerably more difficult (see James, James and Siegmund (1987)).

It is well-known that $\mu(\theta) \equiv E_\theta(X) = \psi'(\theta)$ and $\mathrm{Var}_\theta(X) = \psi''(\theta) > 0$. Since $\mu(\theta)$ is strictly increasing one can write $\theta(\mu)$, that is, θ as a function of μ. For $\theta_0 \neq \theta_1$, let $\{X_1, \ldots, X_\nu\}$ and $\{X_{\nu+1}, \ldots, X_m\}$ be two random samples from F_{θ_0} and F_{θ_1} respectively.

Let $H(x) = \sup_\theta\{\theta x - \psi(\theta)\}$, $S_n = X_1 + \cdots + X_n$ and

$$\Lambda_n = nH\left(\frac{S_n}{n}\right) + (m-n)H\left(\frac{S_m - S_n}{m-n}\right).$$

The maximum likelihood estimator $\hat{\nu}$ of ν can be calculated as follows.

For fixed θ let $\ell_n(\theta) = \theta S_n - n\psi(\theta)$ and $\ell_{n,m}(\theta) = \theta(S_m - S_n) - (m-n)\psi(\theta)$ be the log likelihood ratio statistic of $\{X_1, \ldots, X_n\}$ and $\{X_{n+1}, \ldots, X_m\}$ respectively. Then

$$\max_n \sup_{\theta_0,\theta_1} [\ell_n(\theta_0) + \ell_{n,m}(\theta_1)] = \max_n \Lambda_n.$$

Hence

$$P_{\nu,\theta_0,\theta_1}\{\nu = \hat{\nu}\} = P_{\nu,\theta_0,\theta_1}\{\max_{1\leq n\leq m} \Lambda_n = \Lambda_\nu\}. \tag{3.1}$$

Let $\{U_k\}$ be a random walk with increments having the same distribution as $\ell_1(\theta_1) - \ell_1(\theta_0) = X(\theta_1 - \theta_0) - [\psi(\theta_1) - \psi(\theta_0)]$ where X has distribution function F_{θ_0}. Hence random walk $\{U_k\}$ has drift $E_{\theta_0}[\log \frac{dF_{\theta_1}}{dF_{\theta_0}}(X)] < 0$. Define $\tau_+ = \inf\{k : U_k > 0\}$.

Similarly let $\{V_k\}$ be a random walk with increments having the same distribution as $\ell_1(\theta_0) - \ell_1(\theta_1) = X(\theta_0 - \theta_1) - [\psi(\theta_0) - \psi(\theta_1)]$ where X has distribution function F_{θ_1}. Clearly random walk V_k also has negative drift $E_{\theta_1}[\log \frac{dF_{\theta_0}}{dF_{\theta_1}}(X)] < 0$. Define $\tau'_+ = \inf\{k : V_k \geq 0\}$.

The next theorem shows that $\hat{\nu}$ is asymptotically minimax for estimating the change point in an exponential family. Indeed according to (3.2) for any different parametric values θ_0, θ_1 its error probability behaves asymptotically as that of the maximum likelihood estimator (which uses the knowledge of θ_0 and θ_1). In this sense estimator $\hat{\nu}$ exhibits adaptive behavior, i.e. if θ's are viewed as nuisance parameters it is performing in an asymptotically optimal way for any unknown (but fixed) values of nuisance parameters.

Theorem 3.1. *Assume the distributions of* $\max_{n\geq 1} S_n/n$ *and* $\min_{n\geq 1} S_n/n$ *have no atom at* $(\psi(\theta_1) - \psi(\theta_0))/(\theta_1 - \theta_0)$ *under probability laws* P_{θ_0} *and* P_{θ_1}. *(This condition is always satisfied if* F_{θ_1} *and* F_{θ_0} *are atomless.) If* $\nu \sim m\rho$ *for some* $0 < \rho < 1$, *then*

$$\begin{aligned}&\lim_{m\to\infty} P_{\nu,\theta_0,\theta_1}\{\nu = \hat{\nu}\} = P\{\tau_+ = \infty\}P\{\tau'_+ = \infty\} \\ &= \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k}\left[P_{\theta_0}\left(\sum_1^k \log \frac{dF_{\theta_1}}{dF_{\theta_0}}(X_j) \geq 0\right) + P_{\theta_1}\left(\sum_1^k \log \frac{dF_{\theta_0}}{dF_{\theta_1}}(X_j) > 0\right)\right]\right\} \\ &= \exp\{-d(P_{\theta_0}, P_{\theta_1})\}.\end{aligned} \tag{3.2}$$

Proof of Theorem 3.1. The starting point is the following lemma whose proof will be given in the Appendix.

Lemma 2. *Let* $\hat{\nu}_0 = \mathrm{argmax}_{1\leq n\leq \nu}\Lambda_n$, $\hat{\nu}_1 = \mathrm{argmax}_{\nu\leq n\leq m}\Lambda_n$. *For any* a *satisfying* $1/2 > a > 0$, *let* $B_m = \{\nu - m^a \leq \hat{\nu}_0\}$ *and* $B'_m = \{\hat{\nu}_1 \leq \nu + m^a\}$. *Then*

$$\lim_{m\to\infty} P_{\nu,\theta_0,\theta_1}\{B_m\} = 1 = \lim_{m\to\infty} P_{\nu,\theta_0,\theta_1}\{B'_m\}. \tag{3.3}$$

Lemma 2 states that the maximum of $\{\Lambda_n, n=1,\dots,m\}$ occurs in a relatively small set $\{\nu-m^a, \nu-m^a+1,\dots,\nu+m^a-1,\nu+m^a\}$, with probability close to 1 for any a satisfying $1/2>a>0$ and m sufficiently large. Thus

$$\begin{aligned}\lim_{m\to\infty} P_{\nu,\theta_0,\theta_1}\{\nu=\hat\nu\} &= \lim_{m\to\infty} P_{\nu,\theta_0,\theta_1}\{\max_{\nu-m^a\le n\le \nu+m^a}\Lambda_n=\Lambda_\nu\}\\ &= \int P_{\nu,\theta_0,\theta_1}\{\max_{\nu-m^a\le n\le \nu+m^a}\Lambda_n=\Lambda_\nu|S_\nu,S_m\}P\{dS_\nu,dS_m\}. \qquad (3.4)\end{aligned}$$

By sufficiency, $P_{\nu,\theta_0,\theta_1}\{\max_{\nu-m^a\le n\le\nu+m^a}\Lambda_n=\Lambda_\nu|S_\nu,S_m\}$ does not depend on θ_0,θ_1. Furthermore, given (S_ν,S_m) random variables $\{\Lambda_n, n=1,2,\dots,\nu\}$ and $\{\Lambda_n, n=\nu,\dots,m\}$ are conditionally independent. Hence the left hand side of (3.4) equals

$$\int P\{\max_{\nu-m^a\le n\le\nu}\Lambda_n=\Lambda_\nu|S_\nu,S_m\}P\{\max_{\nu\le n\le\nu+m^a}\Lambda_n=\Lambda_\nu|S_\nu,S_m\}P_{\nu,\theta_0,\theta_1}\{dS_\nu,dS_m\}. \qquad (3.5)$$

By the law of large numbers $S_\nu\sim\nu\mu_0$ and $S_m\sim m[\rho\mu_0+(1-\rho)\mu_1]$ almost surely. Let $\rho\mu_0+(1-\rho)\mu_1=\mu_\rho$. It suffices to approximate

$$P\{\max_{\nu-m^a\le n\le\nu}\Lambda_n=\Lambda_\nu|S_\nu/\nu=x, S_m/m=y\}, \qquad (3.6)$$

and

$$P\{\max_{\nu\le n\le\nu+m^a}\Lambda_n=\Lambda_\nu|S_\nu/\nu=x, S_m/m=y\} \qquad (3.7)$$

for x,y in an arbitrarily small neighborhood around μ_0 and μ_ρ respectively. Since the expansions and approximations below hold uniformly over an appropriate neighborhood, for simplicity only the approximation of $x=\mu_0, y=\mu_\rho$ will be given.

We consider only the approximation of (3.6). That of (3.7) can be obtained similarly. For $\nu-m^a\le n\le\nu$, a Taylor's expansion and the fact that $H'(x)=\theta(x)$, $xH'(x)-H(x)=\psi(x)$ give

$$\begin{aligned}&nH\Big(\frac{S_n}{n}\Big)-\nu H\Big(\frac{S_\nu}{\nu}\Big)\\ &=(n-\nu)H\Big(\frac{S_\nu}{\nu}\Big)-(n-\nu)\frac{S_\nu}{\nu}H'\Big(\frac{S_\nu}{\nu}\Big)+(S_n-S_\nu)H'\Big(\frac{S_\nu}{\nu}\Big)+\frac{n}{2}\Big(\frac{S_n}{n}-\frac{S_\nu}{\nu}\Big)^2H''(\mu_n^*)\\ &=(S_n-S_\nu)\hat\theta_n-(n-\nu)\psi(\hat\theta_n)+R_{1,n}, \qquad (3.8)\end{aligned}$$

where $\hat\theta_n=\theta(S_n/n)$ and $R_{1,n}=n(S_n/n-S_\nu/\nu)^2H''(\mu_n^*)/2$, with some μ_n^* between S_n/n and S_ν/ν.

Similarly

$$(m-n)H\Big(\frac{S_m-S_n}{m-n}\Big)-(m-\nu)H\Big(\frac{S_m-S_\nu}{m-\nu}\Big)=-(S_n-S_\nu)\hat\theta_{m,n}+(n-\nu)\psi(\hat\theta_{m,n})+R_{2,n}, \qquad (3.9)$$

where $\hat{\theta}_{m,n} = \theta((S_m - S_n)/(m-n))$ and $R_{2,n} = (m-n)[(S_m - S_n)/(m-n) - (S_m - S_\nu)/(m-\nu)]^2 H''(\mu^*_{m,n})/2$ is the remainder in the Taylor's expansion.

From (3.8) and (3.9), it follows that

$$\Lambda_n - \Lambda_\nu = (S_\nu - S_n)(\hat{\theta}_{m,n} - \hat{\theta}_n) - (\nu - n)[\psi(\hat{\theta}_{m,n}) - \psi(\hat{\theta}_n)] + R_{1,n} + R_{2,n}. \quad (3.10)$$

By Lemma 4.1 of Hu (1991), the behavior of $S_\nu - S_n$ for $n = \nu, \ldots, \nu - m^a$ under $P\{\cdot|S_\nu/\nu = \mu_0\}$ is asymptotically the same as that under P_{θ_0}. Thus with $C = \{S_\nu/\nu = \mu_0, S_m/m = \mu_\rho\}$ one has

$$\lim_{m\to\infty} P\{\max_{\nu - m^a \le n \le \nu} |\hat{\theta}_n - \theta_0| > \epsilon | C\} = 0, \quad (3.11)$$

and

$$\lim_{m\to\infty} P\{\max_{\nu - m^a \le n \le \nu} |\hat{\theta}_{m,n} - \theta_1| > \epsilon | C\} = 0. \quad (3.12)$$

Furthermore

$$\begin{aligned} & P\{\max_{\nu - m^a \le n \le \nu} \frac{n}{2}(\frac{S_n}{n} - \frac{S_\nu}{\nu})^2 > \epsilon | C\} \\ &= P\{\max_{\nu - m^a \le n \le \nu} n[\frac{\nu(S_n - S_\nu) + (\nu - n)S_\nu}{n\nu}]^2 > 2\epsilon | C\} \\ &\le P\{\max_{\nu - m^a \le n \le \nu} \frac{(S_n - S_\nu)^2}{n} + \frac{m^{2a} S^2}{(n\nu)^2} > \epsilon | C\}. \end{aligned} \quad (3.13)$$

Since $m^{2a}/n \to 0$, for m sufficiently large, (3.13) is dominated by

$$P\{\max_{\nu - m^a \le n \le \nu} \frac{(S_n - S_\nu)^2}{n} \ge \frac{\epsilon}{3} | C\} \le P\{\frac{\sum_{i=1}^{m^{2a}+1} |X_i|^2}{n} > \frac{\epsilon}{3} | C\} \le \frac{m^{2a} E[X_i^2 | C]}{n\epsilon}. \quad (3.14)$$

By Lemma 4.5 of Hu (1991), the last term of (3.14) behaves like $m^{2a} E_{\theta_0}(X_i^2)/n\epsilon$, which tends to 0. Hence

$$P\{\max_{\nu - m^a \le n \le \nu} |R_{1,n}| > \epsilon | C\} \to 0. \quad (3.15)$$

By a similar argument, one can show that

$$P\{\max_{\nu - m^a \le n \le \nu} |R_{2,n}| > \epsilon | C\} \to 0. \quad (3.16)$$

From (3.10),(3.11), (3.12), (3.15), (3.16), and Lemma 4.1 of Hu (1991), it follows that

$$\lim_{m\to\infty} P\{\max_{\nu - m^a \le n \le \nu} \Lambda_n = \Lambda_\nu | C\} = P\{\tau_+ = \infty\}, \quad (3.17)$$

if the distribution of $\max_{n\geq 1} S_n/n$ ($\min_{n\geq 1} S_n/n$) for $\theta_1 > \theta_0$ ($\theta_1 < \theta_0$) under P_{θ_0} has no point mass at $[\psi(\theta_1) - \psi(\theta_0)]/(\theta_1 - \theta_0)$.

Similarly

$$\lim_{m\to\infty} P\{\max_{\nu\leq n\leq \nu+m^a} \Lambda_n = \Lambda_\nu | C\} = P\{\tau'_+ = \infty\}, \tag{3.18}$$

if the distribution of $\max_{n\geq 1} S_n/n$ ($\min_{n\geq 1} S_n/n$) for $\theta_1 < \theta_0$ ($\theta_1 > \theta_0$) under P_{θ_1} has no point mass at $[\psi(\theta_1) - \psi(\theta_0)]/(\theta_1 - \theta_0)$. This completes the proof.

Corollary. *Under conditions of Theorem* 3.1, *for any two real numbers* $0 < \rho_1 < \rho_2 < 1$

$$\lim_{m\to\infty} \max_{m\rho_1\leq\nu\leq m\rho_2} P_{\nu,\theta_0,\theta_2}(\hat{\nu} = \nu) = \exp[-d(F_{\theta_0}, F_{\theta_1})].$$

Indeed the convergence in (3.2) is uniform in ν such that $m\rho_1 \leq \nu \leq m\rho_2$ which can be seen from Lemmas 4.1 and 4.5 of Hu (1991) and the fact that Chernoff type exponential inequality holds uniformly in exponential families.

Some numerical results for normal distributions are reported below. Let $\hat{\nu}$ be the maximum likelihood estimator of change-point ν in a sequence of independent normal observations with the same known variance, say, $\sigma^2 = 1$ and unknown means θ_1 and θ_2. Thus $\psi(\theta) = \theta^2/2, \mu(\theta) = \theta$ and $H(x) = x^2/2$. A simple calculation shows that in this case

$$\hat{\nu} = \arg \max_{1\leq k<m} \frac{[S_k - kS_m/m]^2}{k(1-k/m)}.$$

Let $\hat{\delta}$ be the maximum likelihood estimator of ν when means θ_1 and θ_2 are known, say, $\theta_2 > \theta_1$. Then

$$\hat{\delta} = \arg \max_{1\leq k<m} [S_k - k(\theta_1 + \theta_2)/2]$$

and (2.2) implies that

$$\lim_{n\to\infty} P(\hat{\delta} = \nu) = \exp\{-2\sum_{k=1}^{\infty} k^{-1}\Phi(-\Delta\sqrt{k})\},$$

where $\Delta = 0.5 \mid \theta_1 - \theta_2 \mid$. According to Theorem 3.1 $P(\hat{\nu} = \nu)$ has the same limit. Pettitt's estimator $\tilde{\nu}$ discussed in the end of Section 2 in this situation has the form

$$\tilde{\nu} = \arg \max_{1\leq k<m} \left[S_k - \frac{k}{m} S_m\right].$$

Figure 1 shows the probabilities of the correct decision for $\hat{\nu}$, $\hat{\delta}$ and $\tilde{\nu}$ when $\Delta = 1$ and $m = 40$ for $\nu = 1, \ldots, 20$. Notice that for all considered estimators these probabilities are symmetric about $\nu = 0.5m = 20$. These results are based on Monte Carlo simulations with 50,000 replicas of i.i.d. standard normal variables. This figure shows that the estimator $\hat{\nu}$ is doing just slightly worse than $\hat{\delta}$, which uses the knowledge about the exact values of θ_1 and θ_2, for ν not to far from $0.5m$. Surprisingly, for extreme values of ν the probabilities of the correct decision for $\hat{\delta}$ even exceed this bound which corresponds to $\exp\{-d_o\} = 0.641$ determined from (2.2) which is depicted in Figure 1 by the solid line. Estimator $\tilde{\nu}$ behaves well only for ν very close to $0.5m$ and rather poorly for other values of ν.

change-point

Figure 1. Probabilities of the correct decision for estimators $\hat{\nu}$ (dotted line '$\cdots$'), $\hat{\delta}$ (dash-dotted line '-·') and $\tilde{\nu}$ (dashed line '--') when $m = 40$ and $\Delta = 1$ for various values of $\nu = 1, \ldots, 20$.

It is worth noting that the situation with unknown σ is much more difficult. See James, James and Siegmund (1992) for some results concerning confidence "change in the mean estimation" for the multivariate normal distribution with unknown covariance matrix (which, however, remains the same for observations before and after change).

Acknowledgement

Research of the second author was supported by NSA Grant MD904-93-H-3015.

Appendix

We prove here that $\lim_{m\to\infty} P_{\nu,\theta_0,\theta_1}\{B_m\} = 1$. The other half of (3.3), being similar, is omitted. For $1 \le i \le \nu - m^a$ let $j = i + m^a$. A Taylor's expansion gives

$$jH(S_j/j) - iH(S_i/i) = (S_j - S_i)\hat{\theta}_i - (j-i)\psi(\hat{\theta}_i) + R_1, \tag{A.1}$$

where $\hat{\theta}_i = \theta(S_i/i)$, and $R_1 = j(S_j/j - S_i/i)^2 H''(\mu_i^*)/2$ with μ_i^* between S_j/j and S_i/i.

Similarly

$$(m-j)H\left(\frac{S_m - S_j}{m-j}\right) - (m-i)H\left(\frac{S_m - S_i}{m-i}\right) = -(S_j - S_i)\hat{\theta}_{m,i} + (j-i)\psi(\hat{\theta}_{m,i}) + R_2, \tag{A.2}$$

where $\hat{\theta}_{m,i} = \theta((S_m - S_i)/(m-i))$ and $R_2 = (m-j)[(S_m - S_j)/(m-j) - (S_m - S_i)/(m-i)]^2 H''(\mu_{m,i}^*)/2$ is the remainder in a Taylor's expansion.

From (A.1) and (A.2) it follows that

$$\Lambda_j - \Lambda_i = (S_j - S_i)(\hat{\theta}_i - \hat{\theta}_{m,i}) - (j-i)[\psi(\hat{\theta}_i) - \psi(\hat{\theta}_{m,i})] + R_1 + R_2. \tag{A.3}$$

In the following, K_i, $i = 1, 2, \ldots, 5$, and C denote constants. By Chernoff's inequality (see e.g. Chernoff (1952)) for $i > m^{2a}$ and any $\epsilon > 0$ we have

$$\limsup_{m\to\infty} \frac{1}{m^a} \log P\{|\hat{\theta}_i - \theta_0| > \epsilon\} \le -K_1(\epsilon). \tag{A.4}$$

For $i/m \to \iota$, let

$$\theta\Big(\frac{(1-\rho)\mu_1 + (\rho - \iota)\mu_0}{1-\iota}\Big) = \theta_\iota.$$

Then

$$\limsup_{m\to\infty} \frac{1}{m^a} \log P\{|\hat{\theta}_{m,i} - \theta_\iota| > \epsilon\} \le -K_2(\epsilon), \tag{A.5}$$

$$\limsup_{m\to\infty} \frac{1}{m^a} \log P\{R_i > C\} \le -K_3(C), \quad i = 1, 2. \tag{A.6}$$

From (A.3)-(A.6) for $i > m^{2a}$ it follows, except for a set of exponentially small probability (to be more precise, of order $\exp(-Km^a)$ for some positive constant K) that

$$\Lambda_j - \Lambda_i \sim (S_j - S_i)(\theta_0 - \theta_\iota) - (j-i)[\psi(\theta_0) - \psi(\theta_\iota)]. \tag{A.7}$$

By Chernoff's ineqality again, the right hand side of (A.7) is greater than

$$(1-\epsilon)m^a\{\mu_0(\theta_0 - \theta_\iota) - [\psi(\theta_0) - \psi(\theta_\iota)]\}$$

except for a set of exponentially small probability. Observing that $\mu_0(\theta_0 - \theta_\iota) - [\psi(\theta_0) - \psi(\theta_\iota)] > 0$, we have

$$\limsup_{m\to\infty} \frac{1}{m^a} \log P\{\Lambda_j - \Lambda_i \le 0\} \le K_5.$$

For $1 \le i \le m^{2a}$ by the strong law of large numbers, with probability one

$$\lim_{m\to\infty} \max_{1\le i\le m^{2a}} \frac{\Lambda_i}{m} = H[(1-\rho)\mu_1 + \rho\mu_0] < [(1-\rho)H(\mu_1) + \rho H(\mu_0)] = \lim_{m\to\infty} \frac{\Lambda_\nu}{m}.$$

Thus

$$\lim_{m\to\infty} P\{\hat\nu_0 \le m^{2a}\} = 0. \tag{A.8}$$

Let $A_i = \{\Lambda_i > \Lambda_j\}$. Clearly,

$$P\{B_m^c\} \le P\{\hat\nu_0 \le m^{2a}\} + P\{m^{2a} < \hat\nu_0 \le \nu - m^a\} \le P\{\hat\nu_0 \le m^{2a}\} + P\{\cup_{i=m^{2a}+1}^{\nu-m^a} A_i\}.$$

Now

$$P\{\cup_{i=m^{2a}+1}^{\nu-m^a} A_i\} \le \sum_{i=m^a+1}^{\nu-m^a} P(A_i) \le m\exp(-K_5 m^a) \to 0. \tag{A.9}$$

From (A.8) and (A.9) it follows that $P(B_m^c) \to 0$ which completes the proof.

Remark. It is not difficult to see that (3.3) still holds if B_m and B_m' are replaced by $\{\nu - (\log m)^\alpha \le \hat\nu_0\}$ and $\{\hat\nu_1 \le \nu + (\log m)^\alpha\}$ respectively for any $\alpha > 1$.

References

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

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist. **23**, 493-507.

Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a change-point problem. Biometrika **65**, 243-251.

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

Hu, I. (1991). Nonlinear renewal theory for conditional random walks. Ann. Probab. **19**, 401-422.

James, B., James, K. L. and Siegmund, D. (1987). Tests for a change-point. Biometrika **74**, 71-83.

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. Statistica Sinica **2**, 69-90.

Knopp, K. (1956). Infinite Sequences and Series. Dover, New York.

Krafft, O. and Puri, M. L. (1974). The asymptotic behavior of the minimax risk for multiple decision problems. Sankhyā **36**, 1-12.

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

Ritov, Y. (1990). Asymptotic efficient estimation of the change point with unknown distributions. Ann. Statist. **18**, 1829-1839.

Rukhin, A. L. (1994). Asymptotic minimaxity in the change-point problem. In Proceedings on Change-Point Problems (Edited by D. Siegmund, E. Carlstein and H. Mueller), Institute of Mathematical Statistics, Hayward, CA.

Rukhin, A. L. (1995). Asymptotic behavior of estimators of the change-point in a binomial probability. J. Appl. Statist. Sci. **2**.

Siegmund, D. (1985). Sequential Analysis: Tests and Confidence Intervals. Springer-Verlag, New York.

Woodroofe, M. (1982). Nonlinear Renewal Theory in Sequential Analysis. Regional Conferences Series in Applied Mathematics. SIAM, Philadelphia, PA.

Department of Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong.

Department of Mathematics, University of Maryland, Baltimore County Campus, Baltimore, MD 21228, U.S.A.

(Received November 1992; accepted May 1994)