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

# ASYMPTOTIC MINIMAXITY IN THE CHANGE-POINT PROBLEM

By Andrew L. Rukhin
*University of Maryland*

A lower bound on the limit of the minimax risk under the zero-one loss function is established in the classical setting of the change-point estimation problem. This bound is attained by the maximum likelihood estimator in the situation when the two probability distributions before and after the change point are completely known. The nature of this bound is related to the multiple decision problem and a variety of inequalities relating it to the information-type measures is deduced. Minimaxity of the maximum likelihood procedure is proved for normal observations with unknown means.

**1. Introduction and Summary.** In this paper the classical setting of the change-point estimation problem which was studied by many 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 parameter (Carlstein, 1988) or the distributions (Ritov, 1990) to depend on the sample size in some fashion.

In Section 2 a positive lower bound on the limit of the minimax risk under the zero-one loss function is established. Although the maximum likelihood estimator is not consistent (its deviation from the true parameter is positive with positive probability) it attains this bound when the two probability distributions before and after the change-point are completely known. The nature of this bound is related to the classification problem, and this relationship is used to derive various inequalities for the probability of the correct decision in terms of information-type characteristics.

When the observations are normal with unknown means it is shown in Section 3 that the maximum likelihood procedure is asymptotically minimax

Research supported by NSF grant #DMS-9124009.

AMS 1991 Subject Classification: Primary 62F12, Secondary 62E20.

Key words and phrases: Change-point problem, error probability, information-type measures, maximum likelihood procedure.

in the sense that its risk attains the mentioned bound. This result might be extendable to more general situations. Numerical results give evidence of the reasonable behaviour of this procedure for moderate sample sizes.

**2. Minimaxity Notion for the Change-Point Estimators.** Let $P$ and $Q$ be two different probability distributions with densities $p$ and $q$ and assume that the observed data

$$\mathbf{x} = (x_1, \cdots, x_\nu, x_{\nu+1}, \cdots, x_n) = (\mathbf{x}_1, \mathbf{x}_2)$$

consists of two independent parts, the first $\mathbf{x}_1$ being a random sample from distribution $P$, and the second random sample $\mathbf{x}_2$ coming from distribution $Q$. In other words $\nu$ is the change-point, the parameter of interest. It is known (cf. Hinkley, 1970) that there is no consistent estimator of $\nu$, so that this setting is usually modified for the purpose of studyng the asymptotic minimaxity.

In our approach asymptotic efficiency is defined by means of the error probability which does not tend to zero as sample size increases but which satisfies the following inequality analogous to the classical multiple decision problem.

LEMMA 1. *Let $\delta = \delta(\mathbf{x})$ be any estimator of $\nu$ and denote by $\hat{\delta}$ the maximum likelihood estimator. Then if as $n \to \infty$ both $n - \nu \to \infty$ and $\nu \to \infty$*

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n}\sum_{\nu=1}^{n} Pr\Big(\delta(\mathbf{x}) \neq \nu\Big) &\geq \lim_{\nu\to\infty, n-\nu\to\infty} Pr\Big(\hat{\delta}(\mathbf{x}) \neq \nu\Big) \\ &= 1 - \exp\Big\{-\sum_{k=1}^{\infty} k^{-1}\Big[P\big(\sum_{1}^{k} Z_j \leq 0\big) + Q\Big(\sum_{1}^{k} U_j < 0\Big)\Big]\Big\} \\ &= 1 - \exp\{-\rho(P,Q)\}.\end{aligned} \tag{2.1}$$

*Here $Z_j = \log p(x_j) - \log q(x_j)$ with $x_j$ having the distribution $P$ and $U_j = \log q(x_j) - \log p(x_j)$ with $x_j$ having the distribution $Q$ and it is assumed that $P(\sum_1^k Z_j = 0) = Q(\sum_1^k U_j = 0) = 0$ for any $k$.*

PROOF. The proof of inequality (2.1) is based on the following facts. The maximum likelihood estimator is the Bayes estimator against the uniform distribution for $\nu$ under the zero-one loss function. Therefore

$$\frac{1}{n}\sum_{i=1}^{n} Pr\Big(\delta(\mathbf{x}) = i\Big) \leq \frac{1}{n}\sum_{i=1}^{n} Pr\Big(\hat{\delta}(\mathbf{x}) = i\Big).$$

One has

$$Pr\Big(\hat{\delta}(\mathbf{x}) = i\Big) = Pr\Big(\sum_{1}^{k} Z_j > 0, k = 1, \ldots, i-1, \sum_{i}^{m} U_j > 0, m = i+1, \ldots, n\Big)$$

$$= P\Big(\sum_{1}^{k} Z_j > 0, k = 1, \ldots, i-1\Big) Q\Big(\sum_{1}^{k} U_j > 0, k = 1, \ldots, n-i\Big)$$

$$= p_{i-1} q_{n-i}.$$

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

$$p_i \to p = \exp\Big\{ -\sum_{k=1}^{\infty} k^{-1} P(\sum_{1}^{k} Z_j \leq 0)\Big\}$$

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

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

Lemma 1 shows that the maximum likelihood estimator is asymptotically minimax and also that the quantity $\rho(P,Q)$ defined by (2.1) provides a new "information-type" divergence between distributions $P$ and $Q$. Indeed, as is easy to see, $\rho(P,Q) = \rho(Q,P)$,

$$\rho(P,P) = \infty$$

and

$$\rho(P,Q) = 0$$

if $P$ and $Q$ are singular.

These and some other properties of $\rho(P,Q)$ are discussed by Lorden (1977) who showed the fundamental role of this quantity for sequential multiple hypotheses testing problems.

For example if $P$ and $Q$ are two normal distributions with the same, say, unit variance and means $\theta_1$ and $\theta_2$, then

$$\rho(P,Q) = \rho_o(\Delta) = 2\sum_{k=1}^{\infty} k^{-1} \Phi(-\Delta\sqrt{k}) \qquad (2.2)$$

with $\Delta = 0.5 \mid \theta_1 - \theta_2 \mid$ and $\Phi$ denoting the standard normal distribution function. Function (2.2) plays an important role in sequential analyis and renewal theory; its values are tabulated in Woodroofe, 1982, p. 33. For $\Delta \to 0$

$$\rho_o(\Delta) \sim -\log \Delta$$

and as $\Delta \to \infty$

$$\rho_o(\Delta) \sim 2\exp(-\Delta^2/2)/\sqrt{2\pi}\Delta.$$

Also notice that the term $r_k = [P(\sum_1^k Z_j \leq 0) + Q(\sum_1^k U_j < 0)]/2$ in the sum defining $\rho(P,Q)$ in (2.1) is formed by probabilities of large deviations for sums of i.i.d. random variables. Indeed $E^P Z_j = K(P,Q) > 0$ and $E^Q U_j = K(Q,P)$ are information numbers, so that

$$P(\sum_1^k Z_j \leq 0) = P\Big(\sum_1^k [Z_j - E^P Z_1] \leq -kK(P,Q)\Big)$$

and known inequalities for such probabilities can be used to estimate $r_k$ from above. Also $r_k$ can be interpreted as the error probability of the maximum likelihood procedure in the multiple decision problem for two probability distributions $P$ and $Q$ on the basis of random sample $x_1, \ldots, x_k$. More precisely this is the Bayes risk of the Bayes rule against the uniform prior and the zero-one loss function.

These facts lead to many useful inequalities. For example,

$$\rho(P,Q) \leq \sum_{k=1}^{\infty} \frac{r^k}{k} = -\log(1-r)$$

where

$$r = \inf_{0<s<1} \int p^s(x) q^{1-s}(x)\, d\mu(x) \geq (r_k)^{1/k}.$$

Also

$$\rho(P,Q) \geq \sum_{k=1}^{\infty} \frac{h^{k/2}}{k} = -\log(1-h^2),$$

where

$$h^{2k} = \Big[\int p^{1/2}(x) q^{1/2}(x)\, d\mu(x)\Big]^{2k} \leq r_k(1-r_k)$$

so that

$$1 - \exp\{-\rho(P,Q)\} \geq h^2. \tag{2.3}$$

In the normal example

$$h = \exp\{-\Delta^2/2\}$$

which shows that (2.3) provides a better bound for large $\Delta$ then for small $\Delta$.

If the likelihood ratio $q(x)/p(x)$ is bounded, say,

$$\ell_1 \leq q(x)/p(x) \leq \ell_2$$

then the Hoeffding's (1963) inequality implies that

$$r_k \leq \frac{1}{2}\left[\exp\left\{-\frac{2kK^2(P,Q)}{(\log \ell_2 - \log \ell_1)^2}\right\} + \exp\left\{-\frac{2kK^2(Q,P)}{(\log \ell_2 - \log \ell_1)^2}\right\}\right].$$

Also it follows from Rukhin (1993) that

$$2r_k \geq \frac{1 + \ell_1^k \ell_2^k - 2\ell_1^k}{\ell_2^k - \ell_1^k}.$$

For any two distributions $P$ and $Q$ with given means $\theta_1, \theta_2$ and variances $\sigma_1^2, \sigma_2^2$

$$r_k \leq \frac{1}{2(1 + k(\theta_1 - \theta_2)^2/(\sigma_1 + \sigma_2)^2)}$$

(see Chernoff, 1971).

**3. Asymptotic Efficiency of the Maximum Likelihood Procedure.** In this section we study the change-point estimation problem for a normal sample assuming a known and constant variance, which can then be taken to be unity. Thus assume that $\mathbf{x}_1 = (x_1, \ldots, x_\nu)$ is a normal random vector whose components are independent and have the normal distribution with mean $\theta_1$ and the unit variance. Also let $\mathbf{x}_2 = (x_{\nu+1}, \ldots, x_n)$ be another such vector whose components have mean $\theta_2$ . Here both $\theta_1$ and $\theta_2 = \theta_1 - 2\Delta$ are unknown, but for the sake of concreteness we assume in the following that $\theta_2 < \theta_1$. Also let $\nu/n = p_n$ be such that $p_n \rightarrow p$ with $0 < p < 1$.

Our goal is to investigate the asymptotic efficiency in the sense of Lemma 1 of maximum likelihood estimator $\delta_o$ which is known to have the form

$$\delta_o = \arg\max_{1 \leq k < n} \frac{[S_k - kS_n/n]^2}{k(1 - k/n)}$$

with $S_k = x_1 + \ldots + x_k$. James et al. (1987) studied the properties of tests about the change-point based on this statistic.

One has

$$Pr(\delta_o = \nu) = EPr\Big(\max_{1 \leq k < \nu} \frac{[S_k - kS_n/n]^2}{k(1 - k/n)} \leq s^2 \,\Big|\, \frac{S_\nu - \nu S_n/n}{\sqrt{\nu(1 - \nu/n)}} = s\Big)$$
$$\times\, Pr\Big(\max_{\nu+1 \leq k < n} \frac{[S_k - kS_n/n]^2}{k(1 - k/n)} \leq s^2 \,\Big|\, \frac{S_\nu - \nu S_n/n}{\sqrt{\nu(1 - \nu/n)}} = s\Big) = E\Pi_1\Pi_2. \tag{3.1}$$

The expected value here is taken with respect to the distribution of

$$[S_\nu - \nu S_n/n]/\sqrt{\nu(1-\nu/n)},$$

which is normal with mean $2\Delta\sqrt{\nu(1-p_n)}$ and the unit variance.

We look at the first conditional probability $\Pi_1$ in (3.1). It follows from Hinkley (1970) p. 11 that for bounded $m = \nu - k$ the distribution of

$$\frac{S_k - kS_n/n}{\sqrt{k(1-k/n)}} - \frac{S_\nu - \nu S_n/n}{\sqrt{\nu(1-\nu/n)}}$$

is approximately that of the sum $\sum_1^m Y_j$ where $Y_j, j = 1, \ldots, \nu - 1$, are independent normal random variables with mean $-\Delta/[n(1-p)p]^{1/2}$ and variance $1/[np(1-p)]$. Also the argument given by James et al. (1987) shows that for any positive $m$

$$Pr\left( \max_{1 \le k < \nu} \frac{[S_k - kS_n/n]^2}{k(1-k/n)} = \max_{\nu - m \le k < \nu} \frac{[S_k - kS_n/n]^2}{k(1-k/n)} \right) \to 1$$

and the same formula holds for conditional probability $\Pi_1$. Because of these facts

$$\Pi_1 \to \exp\{-\sum_{m=1}^{\infty} m^{-1}\Phi(-\Delta\sqrt{m})\} = \lim_{m\to\infty} Pr(V_1 + \ldots + V_k < 0, k = 1, \ldots, m)$$

with $V_1, V_2, \ldots$ being independent normal random variables with negative mean $-\Delta$ and unit variance.

The second conditional probability $\Pi_2$ in (3.1) has the same limit, which is easily seen by using a similar argument.

This leads to the following result.

THEOREM 1. *Maximum likelihood procedure $\delta_o$ is an asymptotically efficient estimator of the change-point parameter $\nu$ in a sequence of independent normal observations with the same known variance and unknown means $\theta_1$ and $\theta_2$ in the sense that for any fixed positive $\Delta = 0.5|\theta_1 - \theta_2|$*

$$\lim_{n\to\infty} Pr(\delta_o = \nu) = \exp\{-\rho(P,Q)\} = \exp\left\{ -2\sum_m m^{-1}\Phi(-\Delta\sqrt{m})\right\}.$$

Table 1: Probabilities of the correction decision for estimators $\delta_0$ and $\hat{\delta}$ when $n = 50$ and $\nu = 25$ and asymptotic efficiencies $\exp\{-\rho_0(\Delta)\}$ for various values of $\Delta$

| $\Delta$ | $Pr(\delta_o = \nu)$ | $Pr(\hat{\delta} = \nu)$ | $\exp\{-\rho_o\}$ |
|---|---|---|---|
| .05 | .014 | .022 | .005 |
| .10 | .021 | .035 | .018 |
| .15 | .032 | .054 | .038 |
| .20 | .053 | .074 | .063 |
| .30 | .108 | .132 | .127 |
| .40 | .178 | .203 | .201 |
| .50 | .259 | .280 | .280 |
| .60 | .330 | .360 | .360 |
| .70 | .419 | .438 | .438 |
| .80 | .495 | .511 | .511 |
| .90 | .564 | .579 | .579 |
| 1.00 | .630 | .640 | .641 |
| 2.00 | .950 | .952 | .978 |

Numerical results show that estimator $\delta_o$ behaves quite reasonably for moderate sample sizes. In fact Table 1 containing the probabilities of the correct decision for $\delta_o$ and maximum likelihood procedure

$$\hat{\delta} = \arg\max_k \left[ S_k - k\frac{\theta_1 + \theta_2}{2} \right],$$

which uses the exact values of $\theta_1$ and $\theta_2$, shows that when $n = 50, \nu = 25$ estimators $\delta_o$ and $\hat{\delta}$ exhibit similar behavior. Although $\hat{\delta}$ outperforms $\delta_o$ for all $\Delta$, when $\Delta$ is large this becomes less noticeable. For most values of $\Delta$ the probabilities of the correct decision for procedure $\hat{\delta}$ are almost equal to the limiting value $\exp\{-\rho_o\}$ determined by (2.2) which is also given in Table 1. For small values of $\Delta$ these probabilities even exceed $\exp\{-\rho_o\}$. These results are based on Monte Carlo simulations with 75,000 replicas of i.i.d. standard normal variables.

Hopefully these results can be extended to a more general situation of multivariate normal vectors with an unknown covariance matrix (see Srivastava and Worsley, 1986 for the hypothesis testing problem and James et al., 1992 for the confidence estimation problem) and to observations from an exponential family as in Worsley (1986).

## REFERENCES

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

CHERNOFF, H. (1971). A bound on the classification error for discriminating between populations with specified means and variances. In *Studi di probabilita, statistica e ricerca operativa in onore di Giuseppe Pompilj.* Oderisi-Gubbio.

COBB, G. W. (1978). The problem of the Nile: conditional solution to a change- point problem. *Biometrika* **62**, 243–251.

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

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

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 tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statistica Sinica* **2**, 69–90.

LORDEN, G. (1977). Nearly-optimal sequential tests for finitely many parameter values. *Ann. Statist.* **5**, 1–21.

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

RUKHIN, A. L. (1993). Lower bound on the error probability for families with bounded likelihood ratios. *Proc. Amer. Math. Soc.* **91**.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer, New York.

SRIVASTAVA, M. and WORSLEY, K. L. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *J. Amer. Statist. Assoc.* **81**, 99–204.

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

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

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MARYLAND
BALTIMORE COUNTY CAMPUS
BALTIMORE, MD 21228