Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

THE TEST AND CONFIDENCE INTERVAL FOR A CHANGE-POINT IN MEAN VECTOR OF MULTIVARIATE NORMAL DISTRIBUTION

BY WANG JING-LONG AND WANG JIN

China Normal University, Shanghai, China
and
Institute of Nuclear Power Operation, Wuhan, China

Let $X_1, \cdots, X_n$ be a sequence of independent random vectors, such that the first k of this sequence, $X_1, \cdots, X_k$ have a common multivariate normal distribution $N_p(\mu, \Sigma)$, and the last $n-k$ of this sequence, $X_{k+1}, \cdots, X_n$ have a common multivariate normal distribution $N_p(\mu^*, \Sigma)$, where mean vectors μ and μ^* are all unknown. The integer k, which is called the change-point, is also unknown. In this paper, the hypothesis to be test is $H_0 : \mu = \mu^*$ (i.e. no change) against $H_1 : \mu \neq \mu^*$. The maximum likelihood methods are used to test for a change-point in mean vector of multivariate normal distribution when the covariance matrix Σ is known, or unknown. In the case of the covariance matrix Σ known, the exact null distribution of the test statistic is found, the table of critical values is given, it is shown that the power of test is a increasing function of $||\mu^* - \mu||$, and the probability that the MLE $\hat{k}$ of change-point k is just equal to $k, P(\hat{k} = k)$ is a increasing function of $||\mu^* - \mu||$. In the case of the covariance matrix Σ unknown, the null distribution of test statistic is simulated, and the table of approximate critical values is given. In both cases the confidence interval for the change-point is discussed.

1. Introduction. In the exploration of some oil field of China,we take soil samples from 670m to 1019.875m beneath the earth, and get observations at intervals of 0.125m. Thus, we get 2800×7 values of seven factors:

GΓ	:	national Γ parameter
SON	:	time difference of sound wave
DEN	:	density
SND	:	compensate neutron
ILD	:	interaction
M4	:	four meter gradient
CAC	:	diameter of well

* This work was supported by the National Natural Science Foundation of China
AMS 1980 Subject Classifications: Primary 62H15, 62H12.
Key words and phrases: Change-point,confidence interval, multivariate normal distribution, test.

How should we divide the strata? This is a problem in cluster analysis of ordered samples. Oil exploration experts have applied standard methods to this problem, but the results are not satisfactory . Here we study this problem from another angle.

Let $X_1, \cdots, X_n$ be a sequence of independent random vectors, such that

$$X_i \sim N_p(\mu_i, \Sigma), \qquad i = 1, \cdots, n.$$

The hypothesis to be tested is

$$\begin{aligned}
&H_0 : \mu_1 = \cdots = \mu_n, \\
&H_1 : \text{for some integer} \quad k,\ 1 \le k \le n-1, \\
&\qquad \mu_1 = \cdots = \mu_k = \mu, \\
&\qquad \mu_{k+1} = \cdots \mu_n = \mu^*, \qquad \mu \ne \mu^*.
\end{aligned}$$

If the null hypothesis H_0 is rejected, we can divide the sequence $X_1, \cdots, X_n$ into two groups: $X_1, \cdots, X_k$ and $X_{k+1}, \cdots, X_n$. Thus one change-point means the sequence can be divided into two groups. Generally speaking, m change-points means that $m + 1$ groups are identified. So we can consider the problem of cluster analysis of ordered samples as the problem of testing for change-points. In this paper, we mainly deal with the problem of at most one change-point. As to the problem of more than one change-point, we can discuss it using the result of the problem of at most one change-point.

When $p = 1$, i.e. for the case of a sequence of normally distributed random variables, considerable attention has been devoted to change-point problems (see D. W. Hawkins (1977), K. J. Worsley (1986), X. R. Chen (1988)). When $p > 1$, i.e. for the case of a sequence of normally distributed random vectors, change-point problems have been addressed only recently. In 1989, D. L. Hawkins dealt with the change-point problem, when $p = 2$ and the covariance matrix Σ is known. P. R. Krishnaiah, B. Q. Miao and L. H. Zhao(1990) dealt with the change-point problem, when $p > 1$ and the number of change-points is unknown, using the local likelihood method. The results of both papers are large sample theory. In this paper, we consider small sample theory.

In Section 2, the likelihood ratio method is applied to the change-point problem when the covariance matrix Σ is known. In Section 3, the likelihood ratio method is applied to the change-point problem, when the covariance matrix Σ is unknown. In Section 4, the stepwise discrimination procedure for analysis of the oil exploration data is proposed.

2. Likelihood Ratio Method when Σ Is Known. When the covariance matrix Σ is known, we can set it equal to I without loss of generality.

2.1. Test Statistic.

For fixed k, the likelihood ratio

$$\lambda_k = \max_{\mu} \prod_{i=1}^{n} \left\{ \left(\frac{1}{\sqrt{2\pi}}\right)^p \cdot \exp\left[-\frac{(x_i-\mu)'(x_i-\mu)}{2}\right]\right\} \Big/ \left\{ \max_{\mu,\mu^*} \prod_{i=1}^{k} \left\{ \left(\frac{1}{\sqrt{2\pi}}\right)^p \exp\left[-\frac{(x_i-\mu)'(x_i-\mu)}{2}\right]\right\} \cdot \prod_{i=k+1}^{n} \left\{ \left(\frac{1}{\sqrt{2\pi}}\right)^p \exp\left[-\frac{(x_i-\mu^*)'(x_i-\mu^*)}{2}\right]\right\}\right\}.$$

It is easy to show that

$$E_k = -2 \cdot \log \lambda_k = T_k' \cdot T_k,$$

where

$$T_k = \sqrt{\frac{n}{k\cdot(n-k)}} \cdot \sum_{i=1}^{k} (X_i - \overline{X}),$$

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Thus, for unknown k, minus twice the log likelihood ratio, i.e., the likelihood ratio test statistic is equivalent to

$$U = \max_{1 \le k \le n-1} E_k.$$

We reject H_0 for large values of U.

2.2. The Null Distribution of U.

Because the process $\{T_1, \cdots, T_{n-1}\}$ is Markovian, the null distribution of U can be found by a straightforward generalization of the iterative method employed by D. W. Hawkins (1977). The null distribution function of U is

$$F(x) = P(U < x) = \int_{T_{n-1}' \cdot T_{n-1} < x} F_{n-1}(T_{n-1}, x) \cdot f(T_{n-1})\, dT_{n-1}.$$

where $f(x)$ is a density function of $N_p(0, I)$, and $F_k(T_k, x)$, $k = 2, \cdots, n-1$, have the recurrence formulas:

$$F_2(T_2, x) = P(T_1' \cdot T_1 < x \mid T_2) = \int_{T_1' \cdot T_1 < x} f(T_1 \mid T_2)\, dT_1,$$

$$F_{k+1}(T_{k+1}, x) = P(T_1' \cdot T_1 < x, \cdots, T_k' \cdot T_k < x \mid T_{k+1}) = \int_{T_k' \cdot T_k < x} F_k(T_k, x) \cdot f(T_k \mid T_{k+1})\, dT_k, \qquad k = 2, \cdots, n-2.$$

TABLE 1

n	α	$p = 2$	3	4	5	6	7
10	0.10	8.20	10.29	12.16	13.92	15.63	17.25
		(9.00)	(11.12)	(13.03)	(14.83)	(16.54)	(18.19)
	0.05	9.75	11.93	13.93	15.81	17.59	19.28
		(10.39)	(12.61)	(14.62)	(16.50)	(18.29)	(20.01)
	0.01	13.18	15.64	17.81	19.85	21.78	23.64
		(13.60)	(16.05)	(18.23)	(20.28)	(22.21)	(24.07)
15	0.10	8.80	10.92	12.86	14.69	16.42	18.05
		(9.88)	(12.07)	(14.05)	(15.90)	(17.66)	(19.36)
	0.05	10.37	12.64	14.67	16.58	18.38	20.09
		(11.27)	(13.56)	(15.62)	(17.55)	(19.38)	(21.14)
	0.01	13.84	16.34	18.58	20.86	22.61	24.49
		(14.49)	(16.99)	(19.21)	(21.30)	(23.26)	(25.16)
20	0.10	9.16	11.36	13.32	15.14	16.89	18.60
		(10.49)	(12.73)	(14.74)	(16.63)	(18.42)	(20.15)
	0.05	10.76	13.02	15.10	17.01	18.86	20.64
		(11.88)	(14.21)	(16.31)	(18.27)	(20.12)	(21.92)
	0.01	14.27	16.78	19.00	21.11	23.09	24.98
		(15.10)	(17.64)	(19.88)	(22.01)	(23.98)	(25.91)
25	0.10	9.46	11.66	13.64	15.50	17.26	18.94
		(10.96)	(13.23)	(15.27)	(17.18)	(19.00)	(20.75)
	0.05	11.00	13.36	15.45	17.39	19.23	20.98
		(12.35)	(14.71)	(16.83)	(18.82)	(20.69)	(22.50)
	0.01	14.58	17.08	19.38	21.49	23.49	25.40
		(15.57)	(18.14)	(20.40)	(22.54)	(24.53)	(26.49)
30	0.10	9.66	11.87	13.87	15.75	17.54	19.24
		(11.34)	(13.63)	(15.71)	(17.63)	(19.46)	(21.23)
	0.05	11.25	13.60	15.70	17.66	19.52	21.29
		(12.73)	(15.12)	(17.26)	(19.26)	(21.15)	(22.98)
	0.01	14.79	17.36	19.64	21.76	23.76	25.69
		(15.94)	(18.54)	(20.81)	(22.98)	(24.98)	(26.95)
35	0.10	9.81	12.03	14.05	15.94	17.75	19.46
		(11.66)	(13.98)	(16.06)	(18.01)	(19.85)	(21.63)
	0.05	11.43	13.78	15.88	17.85	19.73	21.53
		(13.04)	(15.46)	(17.61)	(19.63)	(21.53)	(23.37)
	0.01	14.95	17.56	19.84	21.95	23.99	26.34
		(16.26)	(18.88)	(21.16)	(23.35)	(25.35)	(27.34)
40	0.10	9.94	12.21	14.24	16.12	17.91	19.66
		(11.93)	(14.27)	(16.37)	(18.33)	(20.19)	(21.98)
	0.05	11.58	13.92	16.03	18.00	19.90	21.72
		(13.32)	(15.75)	(17.92)	(19.95)	(21.86)	(23.72)
	0.01	15.11	17.72	19.99	22.37	25.92	26.53
		(16.54)	(19.18)	(21.46)	(23.66)	(25.67)	(27.68)

With $f(x\,|\,T_{k+1})$ being the conditional density function given T_{k+1}, i.e., the density function of

$$N_p\big(\rho_{k,k+1}\cdot T_{k+1},\ (1-\rho^2_{k,k+1})\cdot I\big),$$

where

$$\rho_{k,k+1}=\left\{\frac{k\cdot(n-k-1)}{(k+1)\cdot(n-k)}\right\}^{1/2}, \qquad k=1,\cdots,n-2.$$

Using induction, it is easy to show that, for any orthogonal matrix Q,

$$F_k(T_k,x)=F_k(Q\cdot T_k,x), \qquad k=2,\cdots,n-1.$$

So the values of $F_k(T_k,x)$ remains constant at the hyperphere $T_k'\cdot T_k=E_k$, and we can write

$$F_k(T_k,x)=\widetilde{F}_k(E_k,x), \qquad k=2,\cdots,n-1.$$

The conditional distribution of $E_k=T_k'\cdot T_k$ given $T_{k+1},\ k=1,\cdots,n-2$, is

$$(1-\rho^2_{k,k+1})\cdot\chi^2(p,r_k),$$

where $\chi^2(p,r_k)$ is the noncentral χ^2 distribution on p degrees of freedom with the noncentrality parameter

$$r_k=\frac{\rho^2_{k,k+1}\cdot E_{k+1}}{1-\rho^2_{k,k+1}}.$$

Let $F_1(T_1,x)=\widetilde{F}_1(E_1,x)=1$. Then

$$\begin{aligned}\widetilde{F}_{k+1}(E_{k+1},x)=&\ E\big[\widetilde{F}_k(E_k,x)\cdot I_{(E_k<x)}|T_{k+1}\big]\\ =\sum_{i=o}^{\infty}&\exp\Big(-\frac{r_k}{2}\Big)\cdot\frac{(r_k/2)^i}{i\,!}\cdot\int_0^x\widetilde{F}_k(u,x)\\ &\cdot\frac{\exp\Big(-\dfrac{u}{2(1-\rho^2_{k,k+1})}\Big)\cdot u^{i+p/2-1}\cdot\Big(\dfrac{1}{2(1-\rho^2_{k,k+1})}\Big)^{i+p/2}}{\Gamma(i+p/2)}du,\end{aligned}$$

where I_A denotes the characteristic function of the set A. Therefore, the error in the recursive computation of $F(x)$ is independent of dimension p. Of course, the error is dependent of sample size n.

The critical values U_α of test statistic U have been computed iteratively and are listed in the Table 1.

Conservative tests may be made as follows. Since $E_1, \cdots, E_{n-1}$ are identically distributed as $\chi^2(p)$,

$$P(U > c) = P\left(\max_{1\leq k\leq n-1} E_k > c\right) \leq \sum_{k=1}^{n-1} P(E_k > c) = (n-1)\cdot P(E_1 > c).$$

Thus, a conservative level α critical value of test statistic may be based on the upper $\alpha/(n-1)$ fractile of $\chi^2(p)$. These approximations are listed in the Table 1 below the exact values. The difference between the exact value and approximate value is moderate.

2.3. Power of Test.

When the alternative hypothesis H_1 is true, the change-point is $k, 1 \leq k \leq n-1, X_1, \cdots, X_k$ iid $\sim N_p(\mu, \Sigma), X_{k+1}, \cdots, X_n$ iid $\sim N_p(\mu^*, \Sigma)$, $\delta = \mu^* - \mu \neq 0$. We set $\mu = 0, \mu^* = \delta$, without loss of generality. Because the process $\{T_1, \cdots, T_{n-1}\}$ is still Markovian

$$P(U < x) = \int_{T_k' \cdot T_k < x} P(E_1 < x, \cdots, E_{k-1} < x \mid T_k) \cdot P(E_{k+1} < x, \cdots, E_{n-1} < x \mid T_k) \cdot g(T_k)\, dT_k,$$

where $g(x)$ is the density function of

$$N_p\left(-\sqrt{\frac{k(n-k)}{n}} \cdot \delta,\ I_p\right).$$

No matter whether H_0 is true or H_1 is true, the sequence $\{T_1, \cdots, T_{k-1}\}$ has the same conditional distribution when T_k is given. The sequence $\{T_{k+1}, \cdots, T_{n-1}\}$ has this property as well. But, neither the sequence $\{T_1, \cdots, T_{\tau-1}\}$ nor $\{T_{\tau+1}, \cdots, T_{n-1}\}$ have this property when T_τ, $\tau \neq k$, is given. Hence

$$P(E_1 < x, \cdots, E_{k-1} < x \mid T_k) = F_k(T_k, x) = \widetilde{F}_k(E_k, x). \tag{2.1}$$

We note that when H_0 is true the sequence $\{T_{n-1}, \cdots, T_1\}$ obtained by taking the X_i in reverse order is a probabilistic replica of the given sequence $\{T_1, \cdots, T_{n-1}\}$. This implies that

$$P(E_{k+1} < x, \cdots, E_{n-1} < x \mid T_k) = F_{n-k}(T_k, x) = \widetilde{F}_{n-k}(E_k, x). \tag{2.2}$$

Hence,

$$P(U < x) = \int_{T_k' \cdot T_k < x} F_k(T_k, x) \cdot F_{n-k}(T_k, x) \cdot g(T_k)\, dT_k = E\left[\widetilde{F}_k(E_k, x) \cdot \widetilde{F}_{n-k}(E_k, x) \cdot I_{(E_k < x)}\right],$$

where

$$E_k \sim \chi^2\left(p,\ \Delta \cdot \frac{k(n-k)}{n}\right)$$
$$\Delta = \delta' \cdot \delta.$$

Therefore, $P(U > x)$, the power of test, can be calculated iteratively.

PROPOSITION 1. *The power of test, $P(U > x)$ is a increasing function of Δ.*

(Proof is given in Appendix 1)

2.4. MLE of Change-Point.

The MLE $\hat{k}$ of change-point k satisfies

$$E_{\hat{k}} = U = \max_{1 \le k \le n-1} E_k.$$

The distribution of $\hat{k}$ can be obtained. For $\tau = 1, \cdots, n-1$

$$\begin{aligned} P(\hat{k} = \tau) &= P\big(E_1 < E_\tau, \cdots, E_{n-1} < E_\tau\big) \\ &= \int P\big(E_1 \le E_\tau, \cdots, E_{\tau-1} \le E_\tau \,|\, T_\tau\big) \\ &\qquad \cdot P\big(E_{\tau+1} \le E_\tau, \cdots, E_{n-1} \le E_\tau \,|\, T_\tau\big) \cdot h(T_\tau)\, dT_\tau, \end{aligned}$$

where $h(x)$ is the density function of $N_p(a_{\tau,k} \cdot \delta,\ I)$

$$a_{\tau,k} = \begin{cases} -(n-k) \cdot \sqrt{\dfrac{\tau}{n(n-\tau)}} & \tau \le k; \\ -k \cdot \sqrt{\dfrac{n-\tau}{n\tau}} & \tau > k. \end{cases}$$

Especially, we pay considerable attention to the probability that the MLE $\hat{k}$ is just equal to change-point k,

$$\begin{aligned} P(\hat{k} = k) = &\int P\big(E_1 < E_k, \cdots, E_{k-1} < E_k \,|\, T_k\big) \\ &\quad \cdot P\big(E_{k+1} < E_k, \cdots, E_{n-1} < E_k \,|\, T_k\big) \cdot h(T_k)\, dT_k. \end{aligned}$$

Similar to the equations (2.1) and (2.2), we have

$$\begin{aligned} P(\hat{k} = k) &= \int F_k(T_k, E_k) \cdot F_{n-k}(T_k, E_k)\ \cdot\ h(T_k) dT_K \\ &= E\big[\widetilde{F}_k(E_k, E_k) \cdot \widetilde{F}_{n-k}(E_k, E_k)\big], \end{aligned}$$

where

$$E_k \sim \chi^2\left(p,\ \Delta \cdot \frac{k(n-k)}{n}\right),$$
$$\Delta = \delta' \cdot \delta.$$

PROPOSITION 2. *The probability that the MLE $\hat{k}$ is just equal to change-point k, $P(\hat{k} = k)$ is a increasing function of Δ.*

(The proof is given in Appendix 2)

2.5. Confidence Interval for the Change-Point.

The method of Cox and Spjϕtvoll (1982) can be modified and applied to constructing confidence interval for change-point k. The $1-\alpha$ confidence interval for the change-point k is the set of values τ for which we cannot H_0: different $k = \tau$ (i.e. τ is the change-point) against H_1: $k \neq \tau$ (i.e. τ isn't the change-point). We accept H_0 for small values of $M_\tau = \max\{U_\tau^-, U_\tau^+\}$, where U_τ^- and U_τ^+ are the equivalents of U evaluated for the subsequence $X_1, \cdots, X_\tau$ and $X_{\tau+1}, \cdots, X_n$ respectively. It is interesting to note that the distribution of M_τ is free of the nuisance parameters μ and μ^*. The exact $1-\alpha$ confidence interval for change-point k is

$$\widetilde{\mathcal{D}}_\alpha = \{\tau\colon\ M_\tau \leq M_\alpha(\tau)\},$$

where $M_\alpha(\tau)$ is a $(1-\alpha)$-fractile of M_τ,

$$P(M_\tau < M_\alpha(\tau)) = P(U_\tau^- < M_\alpha(\tau)) \cdot P(U_\tau^+ < M_\alpha(\tau)) = 1-\alpha.$$

So, the values of $M_\alpha(\tau)$ can be computed iteratively.

In order to simplify the problem, let

$$d_\alpha = \max_\tau\{M_\alpha(\tau)\}$$

The conservative $1-\alpha$ confidence interval for k is

$$\mathcal{D}_\alpha = \{\tau\colon M_\tau \leq d_\alpha\}$$

Obviously, $\mathcal{D}_\alpha \supset \widetilde{\mathcal{D}}_\alpha$. The values of d_α have been calculated and are listed in the Table 2.

PROPOSITION 3. *When there exists only one change-point in the sequence according to the test with level α, then $\hat{k} \in \mathcal{D}_\alpha$, which means that the MLE $\hat{k}$ of the change-point k is included in the confidence interval with level $1-\alpha$ for the change-point k.*

(The proof is given in Appendix 3)

TABLE 2

n	α	p=1	2	3	4	5	6	7
10	0.10	5.90	8.44	10.29	12.40	14.15	15.84	17.49
	0.05	7.24	9.90	12.09	14.08	15.94	17.73	19.44
	0.01	10.30	13.26	15.69	17.85	19.88	21.81	23.67
15	0.10	6.60	9.19	11.37	13.31	15.12	16.86	18.55
	0.05	7.92	10.72	12.96	15.00	16.92	18.76	20.51
	0.01	11.02	14.07	16.60	18.81	20.87	22.84	24.73
20	0.10	6.98	9.70	11.88	13.87	15.74	17.52	19.20
	0.05	8.41	11.21	13.55	15.63	17.57	19.59	21.16
	0.01	11.57	14.65	17.14	19.43	21.53	23.51	25.41
25	0.10	7.32	10.00	12.27	14.29	16.16	19.84	18.94
	0.05	8.72	11.59	13.90	15.99	17.96	19.84	21.65
	0.01	11.88	14.98	17.58	19.84	21.94	23.94	25.87
30	0.10	7.56	10.31	21.57	14.61	16.51	18.30	20.01
	0.05	8.93	11.84	14.21	16.35	18.33	20.19	21.98
	0.01	12.14	15.31	17.86	20.16	22.30	24.32	26.25
35	0.10	7.74	10.53	12.79	14.84	16.76	18.58	20.32
	0.05	9.13	12.04	14.47	16.61	18.60	20.49	22.29
	0.01	12.38	15.55	18.10	20.44	22.59	24.61	26.56
40	0.10	7.88	10.70	12.96	15.02	16.95	18.80	20.57
	0.05	9.32	12.25	14.66	16.80	18.81	20.71	22.54
	0.01	12.57	15.74	18.34	20.66	22.80	24.83	26.79

3. Likelihood Ratio Method when Σ is Unknown. For fixed k, the likelihood ratio λ_k is

$$\max_{\mu,\Sigma}\prod_{i=1}^{n}\Big\{\frac{1}{\sqrt{|\Sigma|}}\cdot\exp\Big[-\frac{(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}{2}\Big]\Big\}\Big/$$
$$\Big\{\max_{\mu,\mu^*,\Sigma}\prod_{i=1}^{k}\Big\{\frac{1}{\sqrt{|\Sigma|}}\exp\Big[-\frac{(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}{2}\Big]\Big\}$$
$$\cdot\prod_{i=k+1}^{n}\Big\{\frac{1}{\sqrt{|\Sigma|}}\exp\Big[-\frac{(x_i-\mu^*)'\Sigma^{-1}(x_i-\mu^*)}{2}\Big]\Big\}\Big\}.$$

It is easy to show that

$$\lambda_k^{-2/n}=\frac{|V|}{|V-T_k\cdot T_k'|}=\frac{1}{1-T_k'\cdot V^{-1}\cdot T_k},$$

where

$$V=\sum_{i=1}^{n}(X_i-\overline{X})\cdot(X_i-\overline{X})'.$$

Thus, for unknown k we can regard the following statistic as the likelihood ratio test statistic:

$$W = \max\{G_1, \cdots, G_{n-1}\},$$
$$G_k = T_k' \cdot V^{-1} \cdot T_k \qquad k = 1, \cdots, n-1.$$

We reject H_0 for large values of W.

The process $\{T_1, \cdots, T_{n-1}\}$ is Markovian. Howeven $\{T_1, \cdots, T_{n-1}\}$ and V are not independent and the process $\{T_1, \cdots, T_{n-1}\}$ is not Markovian when V is given. When Σ is unknown the critical values W_α of test statistic W can not be computed iteratively even if $p = 1$.

It is very difficult to get the exact critical values of W. The approximate critical values of W can be obtained using simulation. These approximate values of $W_\alpha, \alpha = 0.05$ are listed in the Table 3.

The approximate values can also be obtained using inequalities. Because $G_1, \cdots, G_{n-1}$ are identically distributed,

$$\begin{aligned} P(W > c) = & \; P\left(\max_{1 \le k \le n-1} G_k > c\right) \\ \le & \sum_{k=1}^{n-1} P(G_k > c) = (n-1) \cdot P(G_1 > c). \end{aligned}$$

Thus, an approximate level α critical value of test statistic may be based on the upper $\alpha/(n-1)$ fractile of G_1.

TABLE 3

n	α	$p=1$	2	3	4	5	6	7
30	0.05	0.2390	0.3455	0.4130	0.4480	0.5350	0.5375	0.6157
		(0.2997)	(0.3760)	(0.4350)	(0.4857)	(0.5320)	(0.5740)	(0.6126)
40	0.05	0.2075	0.2710	0.3253	0.3715	0.4135	0.4410	0.4990
		(0.2415)	(0.3023)	(0.3505)	(0.3921)	(0.4306)	(0.4659)	(0.4993)
50	0.05	0.1720	0.2155	0.2700	0.2997	0.3497	0.3650	0.4092
		(0.2029)	(0.2540)	(0.2948)	(0.3303)	(0.3628)	(0.3930)	(0.4214)
60	0.05	0.1470	0.1930	0.2260	0.2620	0.2815	0.3110	0.3445
		(0.1761)	(0.2202)	(0.2548)	(0.2857)	(0.3146)	(0.3405)	(0.3655)
80	0.05	0.0925	0.1255	0.1605	0.1856	0.2090	0.2137	0.2456
		(0.1396)	(0.1743)	(0.2018)	(0.2261)	(0.2486)	(0.2698)	(0.2895)
100	0.05	0.0633	0.0901	0.1065	0.1195	0.1380	0.1525	0.1743
		(0.1168)	(0.1452)	(0.1672)	(0.1874)	(0.2068)	(0.2240)	(0.2408)
150	0.05	0.0506	0.0658	0.0838	0.0950	0.1028	0.1135	0.1275
		(0.0840)	(0.1031)	(0.1184)	(0.1327)	(0.1467)	(0.1586)	(0.1703)

Since

$$G_k = \frac{G_k^*}{1 + G_k^*},$$

where $G_k^* = T_k' \cdot V_k^{-1} \cdot T_k$ with $V_k = V - T_k' \cdot T_k$, and G_k^* has Hotelling T^2 distribution with $n - p - 1$ degrees of freedom, we have

$$\frac{G_k}{1 - G_k} \cdot \frac{n-p-1}{p} = G_k^* \cdot \frac{n-p-1}{p} \sim F(p, n-p-1).$$

The approximate level α critical value is

$$W_\alpha \approx \frac{p \cdot F_{\alpha/(n-1)}(p, n-p-1)}{(n-p-1) + p \cdot F_{\alpha/(n-1)}(p, n-p-1)},$$

where $F_\alpha(m, n)$ denotes upper α fractile of $F(m, n)$ distribution. These approximations are listed in Table 1 below the values.

In order to get confidence interval when Σ is unknown, a method similar to that when Σ is known can also be applied. The approximate critical values with level $1 - \alpha = 0.05$ are listed in the Table 4.

TABLE 4

n	α	$p = 1$	2	3	4	5	6	7
30	0.05	0.5307	0.6427	0.7467	0.8088	0.8717	0.8737	9502
40	0.05	0.3948	0.5307	0.5987	0.6513	0.7237	0.7628	8438
50	0.05	0.3347	0.4607	0.5288	0.5789	0.6312	0.6727	7543
60	0.05	0.2827	0.3837	0.4513	0.4957	0.5708	0.5847	6527
80	0.05	0.2443	0.2987	0.3767	0.3927	0.4507	0.4787	5197
100	0.05	0.2007	0.2437	0.3002	0.3317	0.3747	0.3929	4303
150	0.05	0.1287	0.1782	0.1998	0.2353	0.2539	0.2767	3029

4. Stepwise Discrimination Procedure. The problem with more than one change-point is complex. Here the stepwise discrimination procedure will be used based on the result of the problem with at most one change-point.

Let $\{X_1, \cdots, X_N\}$ be the ordered sample. For the ordered subsample $\{X_1, \cdots, X_n\}$, $n = 1, \cdots, N$, the likelihood ratio statistic for the change-point is denoted by $W(n)$ (or $U(n)$, when the covariance matrix Σ is known). Suppose

$$N_1 = \min_n \{n : W(n) \geq W_\alpha(n)\},$$

where $W_\alpha(n)$ is the critical value of test statistic W with the level α. Thus, the sequence $\{X_1, \cdots, X_{\widehat{N}_1}\}$ is the first cluster, where $\widehat{N}_1$ is the MLE of the change-point for the sequence $\{X_1, \cdots, X_{N_1}\}$. Then, we consider the ordered sample $\{X_{\widehat{N}_1+1}, \cdots, X_N\}$ and repeat the above procedure. Thus, the sequence $\{X_{\widehat{N}_1+1}, \cdots, X_{\widehat{N}_2}\}$ is the second cluster, where $\widehat{N}_2$ is the MLE of change-point for the sequence

$$\{X_{\widehat{N}_1+1}, \cdots, X_{\widehat{N}_1+N_2}\},$$

where

$$N_2 = \min_n \{n :\ W(n) \geq W_\alpha(n)\}$$

and $W(n)$ is the likelihood ratio statistic of the sequence $\{X_{\widehat{N}_1+1},\cdots,X_{\widehat{N}_1+n}\}$. Similarly, the third, fourth, etc. clusters are obtained.

When the oil exploration experts applied the stanlanel methods to the problem of clustering an ordered sample, the results are not satisfatary. The reason is that when the ordered sample $\{X_1,\cdots,X_N\}$ is considered, the sequence $\{X_{\widehat{N}_m+1},\cdots,X_{\widehat{N}_{m+1}}\}$ is a cluster, but when the ordered subsample $\{X_s,\cdots,X_t\}, 1\le s<t\le N$, is considered, the sequence $\{X_{\widehat{N}_m+1},\cdots,X_{\widehat{N}_{m+1}}\}$ probably is not a cluster. They are insterested in this stepwise discrimination procedure, because using this procedure the above shortcoming can usually be overcome. So in this sense,this procedure is robust.

Appendix 1

As it is well known, noncentral $\chi^2(p,r)$ distribution has a monotone likelihood ratio in r. Because

$$F_1(T_1,x)=\widetilde{F}_1(E_1,x)=1,$$
$$\widetilde{F}_{i+1}(E_{i+1},x)=E\big[\widetilde{F}_i(E_i,x)\cdot I_{(E_i<x)}|T_{i+1}\big]$$

and the conditional distribution of $E_i=T_i'\cdot T_i$ given $T_{i+1}, i=1,\cdots,n-2$, is $(1-\rho_{i,i+1}^2)\cdot\chi^2(p,r_i)$ distribution on p degrees of freedom with the noncentrality parameter

$$r_i=\frac{\rho_{i,i+1}^2\cdot E_{i+1}}{1-\rho_{i,i+1}^2}.$$

By induction,we can prove that $\widetilde{F}_i(E_i,x),\ \ i=1,\cdots,n-2$, is a decreasing function of E_i.

Hence, because

$$E_k\sim\chi^2(p,\ r),\qquad r=\frac{k(n-k)}{n}\cdot\Delta$$

and

$$P(U<x)=E\big[\widetilde{F}_k(E_k,x)\cdot\widetilde{F}_{n-k}(E_k,x)\cdot I_{(E_k<x)}\big],$$

$P(U<x)$ is a decreasing function of Δ. This implies that the power of test, $P(U>x)$ is an increasing function of Δ.

Proposition 1 is proved.

Appendix 2

When T_k is given, the $T_1,\cdots,T_{k-1}$ have the conditional matrix normal distribution,

$$E(T_i\,|\,T_k)=\rho_{i,k}\cdot T_k,\qquad i<k,$$
$$\mathrm{COV}(T_i,T_j\,|\,T_k)=(\rho_{i,j}-\rho_{i,k}\cdot\rho_{j,k})\cdot I,\qquad i\le j<k,$$

where

$$\rho_{i,j} = \sqrt{\frac{i(n-j)}{j(n-i)}}, \qquad i \le j.$$

Obviously, for any orthogonal matrix Q, $T_1, \cdots, T_{k-1}$ and $Q \cdot T_1, \cdots$, $Q \cdot T_{k-1}$ have the same conditional distribution when T_k is given. Therefore, in calculating

$$\widetilde{F}_k(E_k, E_k) = P\big(T_1' \cdot T_1 < E_k, \cdots, T_{k-1}' \cdot T_{k-1} < E_k \,|\, T_k\big),$$

we can suppose that

$$T_k = (\sqrt{E_k}, 0, \cdots, 0),$$

without loss of generality.

Let

$$\widetilde{T}_i = T_i - \rho_{i,k} \cdot T_k, \qquad i < k$$

Then

$$\begin{aligned}\widetilde{F}_k(E_k, E_k) = P\big[&(\widetilde{T}_1 + \rho_{1,k} \cdot T_k)' \cdot (\widetilde{T}_1 + \rho_{1,K} \cdot T_k) < E_k,\\ &\cdots, (\widetilde{T}_{k-1} + \rho_{k-1,k} \cdot T_k)' \cdot (\widetilde{T}_{k-1} + \rho_{k-1,k} \cdot T_k) < E_k \,|\, T_k\big],\end{aligned}$$

where the $\widetilde{T}_1, \cdots, \widetilde{T}_{k-1}$ have the conditional matrix normal distribution given T_k,

$$\begin{aligned}&E(T_i|T_k) = 0, && i < k,\\ &\mathrm{COV}(T_i, T_j \,|\, T_k) = (\rho_{i,j} - \rho_{i,k} \cdot \rho_{j,k}) \cdot I, && i \le j < k,\end{aligned}$$

which is independent of T_k.

Let

$$\widetilde{T}_i = (\widetilde{T}_{i1}, \cdots, \widetilde{T}_{ip})', \qquad i = 1, \cdots, k-1.$$

Then

$$\begin{aligned}&(\widetilde{T}_i + \rho_{i,k} \cdot T_k)' \cdot (\widetilde{T}_i + \rho_{i,k} \cdot T_k)\\ &= (\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_k})^2 + \widetilde{T}_{i2}^2 + \cdots + \widetilde{T}_{ip}^2.\end{aligned}$$

When $E_{k1} < E_{k2}$, because $0 < \rho_{i,k} < 1$, $i = 1, \cdots, k$,

$$\begin{aligned}&(\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_{k1}})^2 + \widetilde{T}_{i2}^2 + \cdots + \widetilde{T}_{ip}^2 < E_{k1}\\ \Longrightarrow &(\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_{k1}})^2 < E_{k1}\\ \Longrightarrow &-(1 + \rho_{i,k}) \cdot \sqrt{E_{k1}} < \widetilde{T}_{i1} < (1 - \rho_{i,k}) \cdot \sqrt{E_{k1}}\\ \Longrightarrow &(\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_{k2}})^2 - E_{k2} - (\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_{k1}})^2 + E_{k1} < 0.\end{aligned}$$

Therefore

$$(\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_{k1}})^2 + \widetilde{T}_{i2}^2 + \cdots + \widetilde{T}_{ip}^2 < E_{k1}$$
$$\Longrightarrow (\widetilde{T}_{i1} + \rho_{i,k} \cdot \sqrt{E_{k2}})^2 + \widetilde{T}_{i2}^2 + \cdots + \widetilde{T}_{ip}^2 < E_{k2}.$$

This implies that

$$A_1 \subset A_2,$$

where

$$\begin{aligned} A_1 = \{(\widetilde{T}_1, \cdots, \widetilde{T}_{k-1})\colon & (\widetilde{T}_1 + \rho_{1,k} \cdot T_k)' \cdot (\widetilde{T}_1 + \rho_{1,k} \cdot T_k) < E_{k1}, \\ & \cdots, (\widetilde{T}_{k-1} + \rho_{k-1,k} \cdot T_k)' \cdot (\widetilde{T}_{k-1} + \rho_{k-1,k} \cdot T_k) < E_{k1}\}, \\ A_2 = \{(\widetilde{T}_1, \cdots, \widetilde{T}_{k-1})\colon & (\widetilde{T}_1 + \rho_{1,k} \cdot T_k)' \cdot (\widetilde{T}_1 + \rho_{1,k} \cdot T_k) < E_{k2}, \\ & \cdots, (\widetilde{T}_{k_1} + \rho_{k-1,k} \cdot T_k)' \cdot (\widetilde{T}_{k-1} + \rho_{k-1,k} \cdot T_k) < E_{k2}\}. \end{aligned}$$

Hence,

$$\begin{aligned} \widetilde{F}_k(E_{k1}, E_{k1}) &= P(A_1) \\ &\leq P(A_2) = \widetilde{F}_k(E_{k2}, E_{k2}). \end{aligned}$$

This implies that $\widetilde{F}_k(E_k, E_k)$ is a increasing function of E_k. So is $\widetilde{F}_{n-k}(E_k, E_k)$. Because $E_k \sim \chi^2(p, r)$ have monotone likelihood ratio in r,

$$r = \sqrt{\frac{k(n-k)}{n}} \cdot \Delta$$

$P(\hat{k} = k)$ is an increasing function of Δ.

Appendix 3

The fact that there exists only one change-point in the sequence according to the test with level α means that there is no change-point in both the sequence $\{X_1, \cdots, X_{\hat{k}}\}$ and the sequence $\{X_{\hat{k}+1}, \cdots, X_n\}$ according to the test with level α, i.e.,

$$\begin{aligned} U_{\hat{k}}^- &< U_\alpha(\hat{k}), \\ U_{\hat{k}}^+ &< U_\alpha(n - \hat{k}), \end{aligned}$$

where $U_\alpha(m)$ is the critical values of test with level α for samples of size m. Because $U_\alpha(\hat{k}), U_\alpha(n-\hat{k})$ and $M_\alpha(\hat{k})$ are the upper α fractile of the statistics $U_{\hat{k}}^-, U_{\hat{k}}^+$ and $M_{\hat{k}} = \max\{U_{\hat{k}}^-, U_{\hat{k}}^+\}$ respectively, $U_\alpha(\hat{k}) \leq M_\alpha(\hat{k})$, $U_\alpha(n-\hat{k}) \leq M_\alpha(\hat{k})$. Hence,

$$\begin{aligned} M_{\hat{k}} &= \max\{U_{\hat{k}}^-, U_{\hat{k}}^+\} \\ &\leq M_\alpha(\hat{k}) \leq d_\alpha \end{aligned}$$

Proposition 3 is proved.

Acknowledgement. This work was supported by the National Natural Science Foundation of China.

We wish to thank the editors and refrees for their valuable comments and corrections, which resulted in considerable improvement of our original paper.

REFERENCES

CHEN, X. R. (1988). Inference in a simple change-point model. *Chinese Sci.* Ser.A, 654–667.

COX, D. R., and SPJϕVOLL, E. (1982). On partitioning means into groups. *Scand. J. Statist.* **9**, 147–152.

HAWKINS, D. L. (1989). Estimating changes in a multi-parameter exponential family. *Commun. Statist. –Theory Meth.* **18**, 3595–3623.

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Am. Statist. Assdc.* **72**, 213–216.

KRISHNAINH, P. R., MIAO, B. Q. and ZHAO, L. C. (1990). Local likelihood method in the problems related to change points. *Chin. Ann. of Math.* **11**, 363–375.

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 MATHEMATICAL STATISTICS
EAST CHINA NORMAL UNIVERSITY
3663, ZHONGSHAN ROAD
SHANGHAI 200062
CHINA

ASSESSMENT CENTER
INSTITUTE OF NUCLEAR POWER OPERATION
WUHAN, HUBEI 430074
CHINA