Statistics & Probability Letters 49 (2000) 195–204

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Sequential change-point detection with likelihood ratios

Edit Gombay [1]

Department of Mathematical Sciences, University of Alberta, 632, Central Academic Building, Edmonton, Alberta T6G 2G1, Canada

Received July 1999; received in revised form September 1999

Abstract

We consider the problem of sequential change-point detection when the family of distributions is exponential, and distinguish between parameters of interest, and nuisance parameters. Likelihood ratios are used as test statistics, and their large sample approximations under the alternative hypothesis of change are given. Our formulae allow type II error approximations and they suggest different schemes for change detection and change-point estimation. © 2000 Elsevier Science B.V. All rights reserved

Keywords: Maximum likelihood ratio; Strong approximation; Sequential change-point detection; Stochastic process

1. Introduction

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with densities belonging to the exponential family of distributions, that is, for the density/probability function of X_i, $i = 1, 2, \ldots$, we have

$$\log f(x;\xi) = T(x)\xi' + S(x) - A(\xi), \quad \xi = (\theta, \eta),$$

$$T(x) = (T_1(x), \ldots, T_d(x), T_{d+1}(x), \ldots, T_{d+p}(x)) = (T^d(x), T^p(x)),$$

where $\xi \in \Omega$, $\theta \in \Omega_1 \subset \mathbb{R}^d$, $\eta \in \Omega_2 \subset \mathbb{R}^p$, $d \geqslant 1$, $p \geqslant 0$, and $\Omega = \Omega_1 \times \Omega_2$. (The transpose of a vector is denoted by ξ'.) In our discussion θ is the parameter of interest and η is a nuisance parameter. We are interested in testing for a change in θ from a known initial value θ^0 to some unknown value in the presence of nuisance parameter η. Our null hypothesis is

$$\mathrm{H}_0\colon \theta = \theta^0 \quad \text{for r.v.'s } X_1, X_2, \ldots$$

E-mail address: gombay@stat.ualberta.ca (E. Gombay)

[1] Research supported in part by the Natural Sciences and Engineering Council of Canada.

PII: S0167-7152(00)00048-1

and the alternative is

$$\begin{aligned} \mathrm{H}_a\colon\ \theta &= \theta^0 \quad \text{for r.v.'s } X_1, X_2, \ldots, X_\tau, \\ \theta_i &= \theta^a \quad \text{for r.v.'s } X_{\tau+1} X_{\tau+2}, \ldots \ . \end{aligned}$$

The change-point τ, and the parameters θ^a, η are unknown.

Testing for H_0 against H_a can be done, for example, by Shewhart charts, by Shiryayev–Roberts procedure, by cumulative sum tests, by exponentially weighted moving averages, depending on the model assumptions. See Shewhart (1931), Shiryaev (1963), Roberts (1966), Pollack and Siegmund (1985), Srivastava and Wu (1993), Gosh and Sen (1991), and references therein. Now, we propose the use of the generalized likelihood ratio. This allows testing for any parameter/parameter vector, not just the mean, and allowing the presence of nuisance parameters is also an improvement compared to most existing procedures. For the application of the likelihood ratio test to the change-point problem we refer to Chapter 1 in Csörgő and Horváth (1997).

The connection between sequential tests and sequential change detection procedures is very close. Our null hypothesis is the same as that of a sequential test for a value θ^0. Gombay (1996) derived almost sure approximations for the generalized likelihood ratio under H_0. The test statistic is

$$\Lambda_n = \frac{\sup_{\eta\in\Omega_2} \prod_{i=1}^n f(X_i; \theta^0, \eta)}{\sup_{\theta\in\Omega_1,\, \eta\in\Omega_2} \prod_{i=1}^n f(X_i; \theta, \eta)}, \quad n \geqslant 1.$$

Hence, we have an approximation which can be used to obtain critical values to control the level of significance. See Gombay (1998) for practical implementations and weighted versions. The family of distributions considered in this note is smaller than in Gombay (1996), so we state the conditions for our present case and state the result we want to use.

Let Γ_0 be a neighbourhood of (θ^0, η), and let Γ_a denote a neighbourhood of the $(d+p)$-dimensional interval spanned by endpoints (θ^0, η) and (θ^a, η) where η is the true value of the nuisance parameter.

To unify the conditions for the two theorems we state them for a general $\Gamma \subset \mathbb{R}^{d+p}$. ($\nabla$ denotes the vector of partial derivatives.)

(C1) $\nabla_\eta A$, $\nabla_\xi A$ are continuous and have unique inverses in Γ, which are Lipschitz continuous of order one in each variable.
(C2) The matrices $\nabla^2_{\eta^2} A$, and $\nabla^2_{\xi^2} A$ are positive definite, Lipschitz continuous of order one in each variable in Γ. Furthermore, their inverses exist.
(C3) $(\partial^3/\partial\xi_i\partial\xi_j\partial\xi_k)A(\xi)$, $i,j,k = 1,\ldots,d+p$, exist and are bounded in Γ.
(C4) $E_\xi |T_i(X)|^\gamma < \infty$, $1 \leqslant i \leqslant d+p$, with some $\gamma > 2$.

We also assume that the probability space is rich enough to support all stochastic processes defined below.

Theorem 1. *If* H_0 *and conditions* (C1)–(C4), *are satisfied for* $\Gamma = \Gamma_0$, *then there exist independent, Wiener processes* $W_1(x), \ldots, W_d(x)$, *such that with* $V_d(x) = (1/x)\sum_{1\leqslant j\leqslant d} W_j^2(x)$

(i) $\sup_{1\leqslant t<\infty} |-2\log\Lambda_{[nt]} - V_d(nt)| = \mathrm{O}(n^{-\alpha}(\log\log n)^{1/2})$, *a.s., where* $0 < \alpha \leqslant \frac{1}{2} - 1/\gamma$,
(ii) *for any* $0 < \varepsilon < 1$

$$\left| \sup_{1\leqslant k\leqslant n} (-2\log\Lambda_k)^{1/2} - \sup_{1\leqslant t\leqslant n} V^{1/2}(t) \right| = \mathrm{O_P}(\exp(-\log n)^{1-\varepsilon}).$$

The proof of Theorem 1 is more simple than those of the corresponding theorems in Gombay (1996), so it will be omitted.

Theorem 2. *Let* $\Gamma = \Gamma_a$. *Under conditions* (C1)–(C4), *if* $n > \tau$, *then*

$$\sup_{n\to\infty} \left| \frac{-2\log\Lambda_n - nQ_n}{n^{1/2} V_n^{1/2}} - \left(\frac{n-\tau}{n}\right)^{1/2} \frac{W(n-\tau)}{(n-\tau)^{1/2}} + \left(\frac{\tau V_n^*}{n V_n}\right)^{1/2} \mathrm{N}(0,1) \right| = \mathrm{o_P}(1),$$

where $W(\cdot)$ *is a Wiener process, the standard normal random variable* $\mathrm{N}(0,1)$ *is independent of* $W(\cdot)$,

$$Q_n = (\xi_n^a - (\theta^0, \eta_n^{0a}))\nabla^2_{\xi^2} A(\xi^*)(\xi_n^a - (\theta^0, \eta_n^{0a}))',$$

$$V_n = (\xi_n^a - (\theta^0, \eta_n^{0a}))\nabla^2_{\xi^2} A(\theta^a, \eta)(\xi_n^a - (\theta^0, \eta_n^{0a}))',$$

$$V_n^* = (\xi_n^a - (\theta^0, \eta_n^{0a}))\nabla^2_{\xi^2} A(\theta^0, \eta)(\xi_n^a - (\theta^0, \eta_n^{0a}))',$$

ξ_n^a, η_n^{0a} *are defined in* (3.1) *and* (3.4), *respectively, while for* ξ^* (3.6) *holds.*

The complexity of the notation of the different parameter values is present because of the crucial role these different points in the parameter space play in deriving these results, and it cannot be avoided.

By (C1), (C2), and by the definitions of ξ_n^a, (θ^0, η_n^{0a}), one can see that $V_n \neq 0$, $V_n^* \neq 0$, and the quadratic form Q_n in the mean is positive and bounded away from zero as $n \to \infty$. Theorem 2 generalizes some of the results in Gombay (1997), where $\tau = 0$ was considered.

In Theorem 2 nothing is assumed on the connection between τ and n. The result covers the case when τ does not depend on n as well as the case when τ is a function of n. For example, if $\tau = [n\lambda]$, for some $0 < \lambda < 1$, then in Theorem 2 we would have

$$\sup_{n\to\infty} \left| \frac{-2\log \Lambda_n - nQ_n}{n^{1/2} V_n^{1/2}} - \left(\lambda\left(1 + \frac{V_n^*}{V_n}\right)\right)^{1/2} \mathrm{N}(0,1) \right| = o_{\mathrm{P}}(1).$$

Furthermore, in this case Q_n, V_n, and V_n^* would be constants, not depending on n, only depending on λ.

Combining Theorems 1 and 2, we can see that the sequential generalized likelihood ratio-based statistic $-2\log \Lambda_n$ has the following large sample behaviour.

Until the change point it has approximately a constant mean value d (number of parameters of interest), that starts to increase at the change point by a value Q_n with each additional observation. The variance will also change at τ from the approximate value of $2d$ to the sum of two weighted quadratic forms.

The increments in the process $\{-2\log \Lambda_n\}$, when additional observations are taken after change, are independent in the leading term, as shown in (3.5). They are not identically distributed as the mean and variance are changing. See examples for demonstration. Hence cumulative-sum, and other change-in-the mean procedures may be adopted for $\{-2\log \Lambda_n\}$. In particular, if we prepare a cusum plot for it, the point, where the graph of observed values starts to increase, can be taken as our estimator for τ.

2. Examples

Example 1. Let $\{X_i\}$ be normal random variables, the mean, μ, the parameter of interest, the variance, σ^2, a nuisance parameter. The test statistic is

$$-2\log \Lambda_n = n\left\{-\log\left(\sum (X_i - \overline{X})^2\right) + \log\left(\sum (X_i - \mu^0)^2\right)\right\}.$$

As the initial parameter value is assumed to be known, without loss of generality, we may assume that $\mu^0 = 0$. We have to express the parameters in terms of the natural parameters of the exponential family. We have $d = p = 1$, $\Omega_1 = \mathbb{R}$, $\Omega_2 = (0, \infty)$, $T_1(x) = x$, $T_2(x) = x^2$, $\theta = \mu/\sigma^2$, $\eta = -1/2\sigma^2$, $A(\theta, \eta) = \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(-2\eta) - \frac{1}{4}\theta^2/\eta, \nabla_\xi A(\theta, \eta) = (-\theta/2\eta, -1/2\eta + \theta^2/4\eta^2)$. The matrix of the different quadratic forms, in terms of the original parameters, is

$$\nabla^2_{\xi^2} A = \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 4\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix}.$$

Solving Eqs. (3.1) and (3.4) we get for the components of ξ_n^a, in terms of the original parameters

$$\theta_n^a = \frac{\mu^a}{(\tau/n)(\mu^a)^2 + \sigma^2},$$

$$\eta_n^a = \left(-\frac{1}{2}\right) \frac{1}{((n-\tau)/n - ((n-\tau)/n)^2)(\mu^a)^2 + \sigma^2},$$

and

$$\eta_n^{0a} = \left(-\frac{1}{2}\right) \frac{1}{[(n-\tau)/n](\mu^a)^2 + \sigma^2}.$$

The vector of the quadratic forms is

$$(\xi_n^a - (\theta^0, \eta_n^{0a})) = \left(\frac{\mu^a}{(\tau/n)(\mu^a)^2 + \sigma^2}, \left(-\frac{1}{2}\right) \frac{1}{((n-\tau)/n - ((n-\tau)/n)^2)(\mu^a)^2 + \sigma^2} + \frac{1}{2} \frac{1}{[(n-\tau)/n](\mu^a)^2 + \sigma^2}\right).$$

Example 2. Let σ^2 be the parameter of interest, and μ the nuisance parameter in the normal observations. As the initial value is known, we assume w.l.g. that $\sigma^0 = 1$. Then the test statistic is

$$-2 \log \Lambda_n = n \log\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2\right) + n - \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

In terms of the natural parameters, $\theta = -1/2\sigma^2$, $\eta = \mu/\sigma^2$, $T_1(x) = x^2$, $T_2(x) = x$, $\nabla_\xi A(\theta, \eta) = (1/2\theta - \eta^2/4\theta^2, \eta/2\theta)$. The null hypothesis is that there is no change in the distribution along the sequence of observations, while the alternative hypothesis is

$$\begin{aligned} \mathrm{H}_a\colon\ \sigma^2 &= (\sigma^0)^2, \quad \mu \text{ unknown, for r.v.'s } X_1, X_2, \ldots, X_\tau, \\ \sigma^2 &= (\sigma^a)^2, \quad \mu \text{ unknown, for r.v.'s } X_{\tau+1}, X_{\tau+2}, \ldots, \end{aligned}$$

and in terms of the natural parameters θ and η it is

$$\begin{aligned} \mathrm{H}_a\colon\ \theta &= \theta^a, \quad \eta \text{ unknown, for r.v.'s } X_1, \ldots, X_\tau, \\ \theta &= \theta^a, \quad \eta^{0a} \text{ unknown, for r.v.'s } X_{\tau+1}, X_{\tau+2}, \ldots, \end{aligned}$$

with $\eta^{0a} = \mu/(\sigma^a)^2$. (This change in the η value leaves the proof of the theorem valid.) Solving equations we get for the components of ξ_n^a,

$$\theta_n^a = \left(-\frac{1}{2}\right) \frac{1}{(\tau/n)(\sigma^0)^2 + [(n-\tau)/n](\sigma^a)^2},$$

$$\eta_n^a = \frac{\mu}{(\tau/n)(\sigma^0)^2 + [(n-\tau)/n](\sigma^a)^2}.$$

For η_n^{0a} we have

$$\eta_n^{0a} = \frac{\mu}{(\sigma^0)^2}.$$

The vector of the quadratic forms in Theorem 2 is

$$\xi_n^a - (\theta^0, \eta_n^{0a}) = \left(-\frac{1}{2} \frac{1}{(\tau/n)(\sigma^0)^2 + [(n-\tau)/n](\sigma^a)^2} - \frac{-1}{2(\sigma^0)^2}, \frac{\mu}{(\tau/n)(\sigma^0)^2 + [(n-\tau)/n](\sigma^a)^2} - \frac{\mu}{(\sigma^0)^2}\right),$$

and its matrix is

$$\nabla^2_{\xi^2} A = \begin{pmatrix} 2\sigma^4 + 4\sigma^2\mu^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & \sigma^2 \end{pmatrix}.$$

Example 3. Now, we are looking for simultaneous change in both parameters of the normal distribution. Without loss of generality, we may assume $\theta^0 = (0, 1)$, and then the test statistic is

$$-2\log \Lambda_n = -n \log\left(\frac{\sum (X_i - \overline{X})^2}{n}\right) + \sum X_i^2 - n.$$

In this case $d = 2$, $p = 0$, as there is no nuisance parameter. $T_1(x) = x$, $T_2(x) = x^2$, $\theta_1 = \mu/\sigma^2$, $\theta_2 = -1/2\sigma^2$, $\xi = (\theta_1, \theta_2)$, $\nabla_\xi A(\theta_1, \theta_2) = (-\theta_1/2\theta_2, -1/2\theta_2 + \frac{1}{2}\sigma_1^2/\theta_2^2)$. Solving equations we get for the components of $\xi_n^a = (\theta_1, \theta_2)_n^a$

$$\theta_{1n}^a = \frac{[(n-\tau)/n]\mu^a}{\tau/n + [(n-\tau)/n]((\sigma^a)^2 + (\mu^a)^2) - ([(n-\tau)/n]\mu^a)^2},$$

$$\theta_{2n}^a = \left(-\frac{1}{2}\right) \frac{1}{\tau/n + [(n-\tau)/n]((\sigma^a)^2 + (\mu^a)^2) - ([(n-\tau)/n]\mu^a)^2}.$$

The vector in the quadratic form is

$$(\xi_n^a - (\theta^0, \eta_n^{0a})) = (\theta_{1n}^a, \theta_{2n}^a - \tfrac{1}{2}),$$

and the matrix $\nabla^2_{\xi^2} A$ is the same as in Example 1.

Example 4. Consider three independent sequences of normal observations $\{X_{1i}\}, \{X_{2i}\}$, and $\{X_{3i}\}$, all with variance one. Test that the three population means are equal, i.e.,

$$\mathrm{H}_0\colon \mu_1 = \mu_2 = \mu_3 = \mu, \quad \mu \text{ unknown, for r.v.'s } X_{1i}, X_{2i}, X_{3i}, \quad i = 1, 2, \ldots$$

against a change

$$\mathrm{H}_a\colon \mu_1 = \mu_2 = \mu_3 = \mu, \quad \mu \text{ unknown, for r.v.'s } X_{1i}, X_{2i}, X_{3i}, \quad i = 1, 2, \ldots, \tau,$$

$$\mu_j \neq \mu_k, \quad \text{for some } k \neq j, \text{ for r.v.'s } X_{1i}, X_{2i}, X_{3i}, \quad i = \tau+1, \tau+2, \ldots,$$

where τ and μ_j, $j = 1, 2, 3$, are unknown constants. This is like a sequential ANOVA test for the equality of three means, when they are initially equal. To apply our theorem we transform the parameters: $\theta_1 = \mu_1 + \mu_2 - 3\mu_3$, $\theta_2 = \mu_1 - \mu_2$, $\eta = \mu_3$, and we get $T_1(x_1, x_2, x_3) = (x_1 + x_2)/2$, $T_2(x_1, x_2, x_3) = (x_1 - x_2)/2$, $T_3(x_1, x_2, x_3) = x_1 + x_2 + x_3$. Our hypotheses become

$$\mathrm{H}_0\colon (\theta_1, \theta_2) = (0, 0), \quad \eta \text{ unknown, for r.v.'s } X_{1i}, X_{2i}, X_{3i}, \quad i = 1, 2, \ldots$$

and

$$\mathrm{H}_a\colon (\theta_1, \theta_2) = (0, 0), \quad \eta \text{ unknown, for r.v.'s } X_{1i}, X_{2i}, X_{3i}, \quad i = 1, 2 \ldots, \tau,$$

$$(\theta_1, \theta_2) \neq (0, 0), \quad \eta \text{ unknown, for r.v.'s } X_{1i}, X_{2i}, X_{3i}, \quad i = \tau+1, \tau+2, \ldots .$$

The test statistic is

$$-2\log \Lambda_n = \frac{(\sum_{j=1}^3 \sum_{i=1}^n X_{ji})^2}{3n} - \frac{(\sum_{i=1}^n X_{1i})^2}{n} - \frac{(\sum_{i=1}^n X_{2i})^2}{n} - \frac{(\sum_{i=1}^n X_{3i})^2}{n}.$$

$A(\theta_1,\theta_2,\eta) = \theta_1^2/4+\theta_2^2/4+\frac{3}{2}\eta^2+\theta_1\eta$, $\nabla_\xi A(\xi) = (\frac{1}{2}\theta_1+\eta, \frac{1}{2}\theta_2, 3\eta+\theta_1)$. Solving equations again, we get for ξ_n^a in terms of the original parameters $\xi_n^a = ([(n-\tau)/n](\mu_1+\mu_2) - 2[(n-\tau)/n]\mu_3, [(n-\tau)/n](\mu_1-\mu_2), (\tau/n)\mu + [(n-\tau)/n]\mu_3)$, and $\eta_n^{0a} = (\tau/n)\mu + [(n-\tau)/n](\mu_1+\mu_2+\mu_3)/3$. So the vector of the quadratic forms is

$$(\xi_n^a - (\theta^0, \eta_n^{0a})) = \left(\frac{n-\tau}{n}(\mu_1+\mu_2-2\mu_3), \frac{n-\tau}{n}(\mu_1-\mu_2), \frac{n-\tau}{n}\left(\mu_3 - \frac{\mu_1+\mu_2+\mu_3}{3}\right)\right),$$

and their matrix is

$$\nabla_{\xi^2}^2 A = \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \\ 1 & 0 & 3 \end{pmatrix}.$$

As this is a constant matrix, the three quadratic forms Q_n, V_n, and V_n^* are identical, and Theorem 2 gets a very simple form.

Example 5. Consider the problem of two population proportions that are unknown but equal initially. The interest is in detecting a change in this situation. More precisely, assume Y_{ij}, $j=1,2,\dots$, $i=1,2$, are independent sequences from populations with probability functions $f_{ij}(y) = \pi_{ij}^y(1-\pi_{ij})^{1-y}$, $y=0,1$, $0<\pi_{ij}<1$, $i=1,2$, $j=1,2,\dots$, and test

$$\begin{aligned} &\mathrm{H}_0\colon \pi_{1j} = \pi_{2j},\ \pi_{2j} = \pi_2, \quad j=1,2,\dots \\ &\mathrm{H}_a\colon \pi_{1j} = \pi_{2j},\ \pi_{2j} = \pi_2, \quad j=1,2,\dots,\tau, \\ &\qquad \pi_{1j} = \pi_1, \quad \pi_{2j} = \pi_2, \ \pi_1 \neq \pi_2, \quad j=\tau+1,\tau+2,\dots, \end{aligned}$$

where τ, π_1, π_2 are unknown. To reparametrize, we write

$$\log f(y_1,y_2,\pi_1,\pi_2) = y_1 \log\frac{\pi_1}{1-\pi_1} + y_2 \log\frac{\pi_2}{1-\pi_2} - (-\log[(1-\pi_1)(1-\pi_2)]),$$

and let $\theta = \log(\pi_1/(1-\pi_1)) - \log(\pi_2/(1-\pi_2))$, $\eta = \log(\pi_2/(1-\pi_2))$. We test

$$\begin{aligned} &\mathrm{H}_0\colon \theta = 0, \quad \eta \text{ unknown, for r.v.'s } Y_{ij}, \quad i=1,2,\ j=1,2,\dots, \\ &\mathrm{H}_a\colon \theta = 0, \quad \eta \text{ unknown, for r.v.'s } Y_{ij}, \quad i=1,2,\ j=1,\dots,\tau, \\ &\qquad \theta = \theta^a, \quad \eta \text{ unknown, for r.v.'s } Y_{ij}, \quad i=1,2,\ j=\tau+1,\tau+2,\dots, \end{aligned}$$

where τ, θ^a, η are unknown. Now, $T_1(y_1,y_2) = y$, and $T_2(y_1,y_2) = y_1+y_2$,

$$A(\theta,\eta) = -\log\left[\left(1 - \frac{e^{\theta+\eta}}{1+e^{\theta+\eta}}\right)\left(1 - \frac{e^\eta}{1+e^\eta}\right)\right].$$

The matrix of the quadratic forms in terms of the original parameters is

$$\nabla_{\xi^2}^2 A = \begin{pmatrix} \pi_1 - \pi_1^2 & \pi_1 - \pi_1^2 \\ \pi_1 + \pi_2 - \pi_1^2 - \pi_2^2 & \pi_1 - \pi_1^2 \end{pmatrix}.$$

Calculations of $\xi_n^a - (\theta^0, \eta_n^{0a})$ give for its first component

$$\log\frac{([(n-\tau)/n]\pi_1 + [(n-\tau)/n]\pi_2)(1 + [(n-\tau)/n]\pi_2 - (\tau/n)\pi_2)}{((\tau/n)\pi_2 - [(n-\tau)/n]\pi_2)(1 - [(n-\tau)/n]\pi_1 - [(n-\tau)/n]\pi_2)}$$

and for its second component

$$\log\frac{((\tau/n)\pi_2 - [(n-\tau)/n]\pi_2)(1 - (\tau/n)\pi_2 - [(n-\tau)/n]\pi_1)}{(1 - (\tau/n)\pi_2 + [(n-\tau)/n]\pi_2)((\tau/n)\pi_2 + [(n-\tau)/n]\pi_1)}.$$

3. Proofs

Proof of Theorem 2. We have

$$-2\log \Lambda_n = 2\sum_{i=1}^{n}\{T(X_i)\tilde{\xi}_n' - A(\tilde{\xi}_n)\} - 2\sum_{i=1}^{n}\{T^d(X_i)\theta^{0'} + T'(X_i)\hat{\eta}_n' - A(\theta^0, \hat{\eta}_n)\},$$

where $T(X_i) = (T^d(X_i), T^p(X_i))$,

$$\tilde{\xi}_n = \text{inv } \nabla_\xi A\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right),$$

$$\hat{\eta}_n = \text{inv } \nabla_\eta A\left(\theta^0, \frac{1}{n}\sum_{i=1}^{n} T^p(X_i)\right),$$

which are to be interpreted as

$$\nabla_\xi A(\tilde{\xi}_n) = \frac{1}{n}\sum_{i=1}^{n} T(X_i),$$

$$\nabla_\eta A(\theta^0, \hat{\eta}_n) = \frac{1}{n}\sum_{i=1}^{n} T^p(X_i),$$

respectively. When $n > \tau$

$$E\sum_{i=1}^{n} T(X_i) = \tau\nabla_\xi A(\theta^0, \eta) + (n-\tau)\nabla_\xi A(\theta^a, \eta).$$

Define $\xi = \xi_n^a$ as the solution of equations

$$\nabla_\xi A(\xi) = \frac{\tau}{n}\, ET(X_1) + \frac{n-\tau}{n}\, ET(X_{\tau+1}),$$

which gives

$$\xi_n^a = \text{inv } \nabla_\xi A\left(\frac{\tau}{n}\, ET(X_1) + \frac{n-\tau}{n}\, ET(X_{\tau+1})\right). \tag{3.1}$$

From the law of iterated logarithm we get that

$$\left\|ET(X_1) - \frac{1}{\tau}\sum_{i=1}^{\tau} T(X_i)\right\| = O_P\left(\left(\frac{\log\log\tau}{\tau}\right)^{1/2}\right),$$

so using (C1) we obtain

$$\|\xi_n^a - \tilde{\xi}_n\| = O_P(n^{-1}(\tau\log\log\tau)^{1/2} \vee n^{-1}((n-\tau)\log\log(n-\tau))^{1/2}). \tag{3.2}$$

We can write

$$\begin{aligned}
-2\log \Lambda_n &= \left\{2\sum_{i=1}^{n}[T(X_i)\tilde{\xi}_n' - A(\tilde{\xi}_n)] - 2\sum_{i=1}^{n}[T(X_i)\xi_n^{a'} - A(\xi_n^a)]\right\} \\
&\quad + \left\{2\sum_{i=1}^{n}[T(X_i)\xi_n^{a'} - A(\xi_n^a)] - 2\sum_{i=1}^{n}[T(X_i)(\theta^0, \hat{\eta}_n)' - A(\theta^0, \hat{\eta}_n)]\right\} \\
&= 2L_{1n} + 2L_{2n}.
\end{aligned}$$

By Taylor expansion and the definition of $\tilde{\xi}_n$

$$L_{1n} = \sum_{i=1}^{n} T(X_i)(\tilde{\xi}_n - \xi_n^a)' - n(A(\tilde{\xi}_n) - A(\xi_n^a))$$

$$= \sum_{i=1}^{n} T(X_i)(\tilde{\xi}_n - \xi_n^a)' - n\{\nabla_\xi A(\tilde{\xi}_n)(\tilde{\xi}_n - \xi_n^a)' + \frac{1}{2}(\tilde{\xi}_n - \xi_n^a)\nabla^2_{\xi^2} A(\xi_n^a)(\tilde{\xi}_n - \xi_n^a)'\} + R_n$$

$$= \frac{n}{2}(\tilde{\xi}_n - \xi_n^a)\nabla^2_{\xi^2} A(\xi_n^a)(\tilde{\xi}_n - \xi_n^a)' + R_n$$

where

$$R_n = \frac{n}{2}(\tilde{\xi}_n - \xi_n^a)(\nabla^2_{\xi^2} A(\tilde{\xi}_n) - \nabla^2_{\xi^2} A(\xi_n^a))(\tilde{\xi}_n - \xi_n^a)'$$

$$+ n\{\text{third order terms in } (\tilde{\xi}_n - \xi_n^a)_j,\ j = 1, \dots, d+p\}$$

$$= \mathrm{O_P}(n^{-2}(\tau \log\log\tau)^{3/2} \vee n^{-2}((n-\tau)\log\log(n-\tau))^{3/2})$$

$$= \mathrm{O_P}(\gamma_{\tau n}), \tag{3.3}$$

by conditions (C2), (C3), and (3.2).

Define $\eta = \eta_n^{0a}$ as the solution of equations

$$\nabla_\eta A(\theta^0, \eta) = \frac{\tau}{n}\, ET^p(X_1) + \frac{n-\tau}{n}\, ET^p(X_{\tau+1}),$$

that is,

$$\eta_n^{0a} = \text{inv } \nabla_\eta A\left(\theta^0, \frac{\tau}{n}\, ET^p(X_1) + \frac{n-\tau}{n}\, ET^p(X_{\tau+1})\right). \tag{3.4}$$

Taylor expansion gives

$$A(\xi_n^a) - A(\theta^0, \hat{\eta}_n) = A(\xi_n^a) - A(\theta^0, \eta_n^{0a}) + A(\theta^0, \eta_n^{0a}) - A(\theta^0, \hat{\eta}_n)$$

$$= A(\xi_n^a) - A(\theta^0, \eta_n^{0a}) + \nabla_\eta A(\theta^0, \hat{\eta}_n)(\eta_n^{0a} - \hat{\eta}_n)'$$

$$+ \tfrac{1}{2}(\eta_n^{0a} - \hat{\eta}_n)\nabla^2_{\eta^2} A(\theta^0, \eta_n^{0a})(\eta_n^{0a} - \hat{\eta}_n)'$$

$$+ \tfrac{1}{2}(\eta_n^{0a} - \hat{\eta}_n)(\nabla^2_{\eta^2} A(\theta^0, \hat{\eta}_n) - \nabla^2_{\eta^2} A(\theta^0, \eta^*))(\eta_n^{0a} - \hat{\eta}_n)',$$

where $\|\eta^* - \hat{\eta}_n\| \leqslant \|\eta_n^{0a} - \hat{\eta}_n\|$.

Using the law of iterated logarithm as in (3.2) before, (C2), and the definitions of η_n^{0a} and $\hat{\eta}_n$, we get that the last quadratic form above is $\mathrm{O_P}(\gamma_{\tau n})$. We have

$$L_{2n} = \sum_{i=1}^{n} [T(X_i)\xi_n^{a'} - A(\xi_n^a)] - \sum_{i=1}^{n} [T(X_i)(\theta^0, \eta_n^{0a})' - A(\theta^0, \eta_n^{0a})]$$

$$+ \sum_{i=1}^{n} [T(X_i)(\theta^0, \eta_n^{0a})' - A(\theta^0, \eta_n^{0a})] - \sum_{i=1}^{n} [T(X_i)(\theta^0, \hat{\eta}_n)' - A(\theta^0, \hat{\eta}_n)]$$

$$= \sum_{i=1}^{n} T(X_i)(\xi_n^a - (\theta^0, \eta_n^{0a}))' - n(A(\xi_n^a) - A(\theta^0, \eta_n^{0a})) + \sum_{i=1}^{n} T^p(X_i)(\eta_n^{0a} - \hat{\eta}_n)^{\mathrm{T}}$$

$$- n\left\{\nabla_\eta A(\theta^0, \hat{\eta}_n)(\eta_n^{0a} - \hat{\eta}_n)' + \tfrac{1}{2}(\eta_n^{0a} - \hat{\eta}_n)\nabla^2_{\eta^2} A(\theta^0, \eta_n^{0a})(\eta_n^{0a} - \hat{\eta}_n)'\right\} + \mathrm{O_P}(\gamma_{\tau n}).$$

Combining the expressions for L_{1n} and L_{2n}, we get when $n > \tau$

$$\begin{aligned}
-2\log\Lambda_n &= 2\sum_{i=1}^{n} T(X_i)(\xi_n^a - (\theta^0, \eta_n^{0a}))' - 2n(A(\xi_n^a) - A(\theta^0, \eta_n^{0a})) \\
&\quad + n(\tilde{\xi}_n - \xi_n^a)\nabla^2_{\xi^2} A(\xi_n^a)(\tilde{\xi}_n - \xi_n^a)' \\
&\quad - n(\eta_n^{0a} - \hat{\eta}_n)\nabla^2_{\eta^2} A(\theta^0, \eta_n^{0a})(\eta_n^{0a} - \hat{\eta}_n)' + \mathrm{O_P}(\gamma_{\tau n}) \\
&= 2\sum_{i=1}^{n} T(X_i)(\xi_n^a - (\theta^0, \eta_n^{0a}))' - 2n(A(\xi_n^a) - A(\theta^0, \eta_n^{0a})) \\
&\quad + \mathrm{O_P}\left(\frac{\tau}{n}\log\log\tau \vee \frac{n-\tau}{n}\log\log(n-\tau)\right),
\end{aligned} \tag{3.5}$$

as by (3.2), and by a similar statement for $\|\eta_n^{0a} - \hat{\eta}_n\|$ the above quadratic terms are $\mathrm{O_P}((\tau/n)\log\log\tau \vee [(n-\tau)/n]\log\log(n-\tau))$. The remaining sum in (3.5) is of independent components. Its expected value is two times

$$\begin{aligned}
&E\left\{\sum_{i=1}^{n} T(X_i)(\xi_n^a - (\theta^0, \eta_n^{0a}))' - (A(\xi_n^a) - A(\theta^0, \eta_n^{0a}))\right\} \\
&\quad = (\tau\nabla_{\xi} A(\theta^0, \eta) + (n-\tau)\nabla_{\xi} A(\theta^a, \eta))(\xi_n^a - (\theta^0, \eta_n^{0a}))' \\
&\qquad - n\nabla_{\xi} A(\xi_n^a)(\xi_n^a - (\theta^0, \eta_n^{0a}))' + \frac{n}{2}(\xi_n^a - (\theta^0, \eta_n^{0a}))\nabla^2_{\xi^2} A(\xi^*)(\xi_n^a - (\theta^0, \eta_n^{0a}))',
\end{aligned} \tag{3.6}$$

where ξ^* is a point between ξ_n^a and (θ^0, η_n^{0a}), in the $(d+p)$-dimensional sense. As the definition of ξ_n^a causes cancellation in (3.6), we get that apart from an error of $\mathrm{O}((\tau/n)\log\log\tau \vee [(n-\tau)/n]\log\log(n-\tau))$, the expected value of $-2\log\Lambda_n$ is

$$nQ_n = n(\xi_n^a - (\theta^0, \eta_n^{0a}))\nabla^2_{\xi^2} A(\xi^*)(\xi_n^a - (\theta^0, \eta_n^{0a}))' > 0.$$

Terms $U(X_i) = T(X_i)(\xi_{An} - (\theta^0, \eta_n^{0a}))'$, $i = 1, 2, \ldots$, are independent. We note that, if there is no change

$$\sum_{i=1}^{n} U(X_i) = 0,$$

as in that case

$$\xi_n^a = (\theta^0, \eta), \quad \eta_n^{0a} = \eta.$$

Let $Y_i \stackrel{\mathscr{D}}{=} U(X_{\tau+i})$, $i = 1, 2, \ldots$, be independent random variables, and $n > \tau$. By (C4) and Komlós et al. (1975, 1976), there exists a Wiener process, s.t.

$$\left|\sum_{i=1}^{k}(Y_i - EY_i)/V_n^{1/2} - W(k)\right| \stackrel{\text{a.s.}}{=} \mathrm{o}(k^{1/\mu}), \tag{3.7}$$

where

$$\begin{aligned}
V_n &= \mathrm{Var}(U(X_{\tau+1})) \\
&= (\xi_n^a - (\theta^0, \eta_n^{0a}))\nabla^2_{\xi^2} A(\theta^a, \eta)(\xi_n^a - (\theta^0, \eta_n^{0a}))'.
\end{aligned}$$

From this we get

$$\left|((n-\tau)V_n)^{-1/2}\sum_{i=\tau+1}^{n}(U(X_i)-EU(X_i))-(n-\tau)^{-1/2}W(n-\tau)\right|\overset{\text{a.s.}}{=}\mathrm{o}((n-\tau)^{-\alpha}),$$
$$0\leqslant\alpha\leqslant\tfrac{1}{2}-\frac{1}{\mu}. \tag{3.8}$$

By the central limit theorem

$$\frac{1}{\tau^{1/2}}\sum_{i=1}^{\tau}(U(X_i)-EU(X_i))/(V_n^*)^{1/2}$$

is approximately a standard normal random variable, where

$$V_n^*=\mathrm{Var}(U(X_1))$$
$$=(\xi_n^a-(\theta^0,\eta_n^{0a}))\nabla_{\xi^2}^2A(\theta^0,\eta)(\xi_n^a-(\theta^0,\eta_n^{0a}))'.$$

As the two sums $\sum_{i=1}^{\tau}U(X_i)$ and $\sum_{i=\tau+1}^{\tau}U(X_i)$ are independent, (3.5), (3.7) and (3.8) give the theorem.

References

Csörgő, M., Horváth, L., 1997. Limit Theorems in Change-point Analysis. Wiley, New York.
Gombay, E., 1996. The weighted sequential likelihood ratio. Canad. J. Statist. 24, 229–239.
Gombay, E., 1997. The likelihood ratio under noncontiguous alternatives. Canad. J. Statist. 25, 417–423.
Gombay, E., 1998. Sequential tests of composite hypotheses. manuscript.
Gosh, B.K., Sen, P.K., 1991. Handbook of Sequential Analysis. Marcel Dekker, New York.
Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent R.V.'s and the sample DF. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111–131.
Komlós, J., Major, P., Tusnády, G., 1976. An approximation of partial sums of independent R.V.'s and the sample DF. II. Z. Wahrscheinlichkeitstheorie verw. Gebiete 34, 33–58.
Roberts, S.W., 1966. A comparison of some control chart procedures. Technometrics 8, 411–430.
Shewhart, W.A., 1931. The Economic Control of Quality of a Manufactured Product. Van Nostrand, Princeton.
Shiryaev, A.N., 1963. On optimum methods in quickest detection problems. Theory Probab. Appl. 8, 26–51.
Srivastava, M.S., Wu, Y., 1993. Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. Ann. Statist. 21, 645–670.