

Model Introduced SPRT for Structural Change Detection of Time Series (II)

Yoshihide Koyama, Tetsuo Hattori

*Graduate School of Engineering, Kagawa University / 2217-20 Hayashi, Takamatsu City, Kagawa 761-0396, Japan
s11d451@stmail.eng.kagawa-u.ac.jp, hattori@eng.kagawa-u.ac.jp*

Katsunori Takeda

Canon IT Solutions Inc., Tosabori Dai Bldg. 2-2-4, Tosabori, Nishi-ku, Osaka-shi, Osaka 550-0001, Japan

Hiromichi Kawano

*NTT Advanced Technology/ Musashino-shi Nakamachi 19-18, Tokyo 180-0006, Japan
hiromichi.kawano@ntt-at.co.jp*

Abstract

In this paper, using the notion of a binary Channel Matrix as well known in Information Theory, we present an equivalent relation between the SPRT (Sequential Probability Ratio Test) and Bayes' Updating. Moreover, we show the relationship between the SPRT and NSPR (New Sequential Probability Ratio) where a Hidden Markov Model with Poisson process is introduced as structural change model. And we also provide the change point detection performance of SPRT and NSPR by experimental results.

Keywords: Change detection, SPRT, NSPR, Hidden Markov Model, Information Theory, Binary Channel, Bayes' Updating.

1. Introduction

There are three problems in dealing with ongoing time series data.¹⁻⁵ First, we have to make a prediction model from the early data. Second, we have to detect the structural change of the time series data, as quickly and correctly as possible, when the prediction model does not meet the observing data any more. Third, we have to remake the prediction model as soon as possible after the detected change point. For the second change detection problem, we previously proposed an application of SPRT (Sequential Probability Ratio Test) that has been mainly used in the field of quality control.⁵⁻⁸

In this paper, using the notion of a binary Channel as well known in Information Theory, we present an

equivalent relation between the SPRT and Bayes' Updating. Moreover, we also explain the relation between the SPRT and the New Sequential Probability Ratio of two conditional probabilities when a Hidden Markov Model with Poisson process is introduced as the never returning structural change model.⁹

2. SPRT and Bayes' Updating

2.1. SPRT

The Sequential Probability Ratio Test (SPRT) is used for testing a null hypothesis H_0 (e.g. the quality is under pre-specified limit 1%) against hypothesis H_1 (e.g. the quality is over pre-specified limit 1%). And it is defined as follows:

Let Z_1, Z_2, \dots, Z_i be respectively observed time series data at each stage of successive events, the probability ratio λ_i is computed as follows.

$$\lambda_i = \frac{P(Z_1 | H_1) \cdot P(Z_2 | H_1) \cdots P(Z_i | H_1)}{P(Z_1 | H_0) \cdot P(Z_2 | H_0) \cdots P(Z_i | H_0)} \quad (1)$$

where $P(Z | H_0)$ denotes the distribution of Z if H_0 is true, and similarly, $P(Z | H_1)$ denotes the distribution of Z if H_1 is true.

Two positive constants C_1 and C_2 ($C_1 < C_2$) are chosen. If $C_1 < \lambda_i < C_2$, the experiment is continued by taking an additional observation. If $C_2 < \lambda_i$, the process is terminated with the rejection of H_0 (acceptance of H_1). If $\lambda_i < C_1$, then terminate this process with the acceptance of H_0 .

$$C_1 = \frac{\beta}{1-\alpha}, \quad C_2 = \frac{1-\beta}{\alpha} \quad (2)$$

where α means type I error (reject a true null hypothesis), and β means type II error (accept a null hypothesis as true one when it is actually false).

2.2. Bayes' Updating based on binary channel

We regard that the observation and SPRT process for the aforementioned structural change point detection of time series data is essentially a kind of Bayes' Updating by a binary channel composed of a set of subjective probability as shown in Fig.1, where input event $X=(x_1, x_2)$ and output event $Y=(y_1, y_2)$.

Using the conditional probability, the transition matrix (or channel matrix) P_B is represented as shown in Eq.(3).

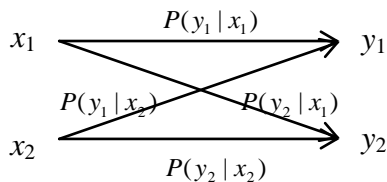


Fig.1. Binary channel.

$$P_B = \begin{bmatrix} P(y_1 | x_1) & P(y_2 | x_1) \\ P(y_1 | x_2) & P(y_2 | x_2) \end{bmatrix} \quad (3)$$

Let the prior probability distribution of input event X be $\{P_0(x_1), P_0(x_2)\}$ as initial values. Let the observed output sequence $\{y_{j_1}, y_{j_2}, \dots, y_{j_n}\}$ ($y_{j_k} = y_1$ or y_2 ($k=1, 2, \dots, n$)). Then, by the first observation of output, the posterior probability is given by Bayes' theorem.

$$p(x_1 | y_{j_1}) = \frac{P_0(x_1)P(y_{j_1} | x_1)}{P(y_{j_1})}$$

$$p(x_2 | y_{j_1}) = \frac{P_0(x_2)P(y_{j_1} | x_2)}{P(y_{j_1})}$$

$$\text{where } P(y_{j_1}) = \sum_i P_0(x_i)P(y_{j_1} | x_i) \quad (4)$$

According to the Bayes' Updating rule, we have the following updated prior probability $\{P_1(x_1), P_1(x_2)\}$.

$$P_1(x_1) = p(x_1 | y_{j_1}) = \frac{P_0(x_1)p(y_{j_1} | x_1)}{P(y_{j_1})}$$

$$P_1(x_2) = p(x_2 | y_{j_1}) = \frac{P_0(x_2)p(y_{j_1} | x_2)}{P(y_{j_1})} \quad (5)$$

Similarly, by the second observation, we have

$$P_2(x_1) = p(x_1 | y_{j_2}) = \frac{P_1(x_1)p(y_{j_2} | x_1)}{P(y_{j_2})}$$

$$P_2(x_2) = p(x_2 | y_{j_2}) = \frac{P_1(x_2)p(y_{j_2} | x_2)}{P(y_{j_2})}$$

$$\text{where, } P(y_{j_2}) = \sum_i P_1(x_i)P(y_{j_2} | x_i) \quad (6)$$

Thus, by the n th observation of output, we have

$$P_n(x_i) = p(x_i | y_{j_n}) = \frac{P_{n-1}(x_i)P(y_{j_n} | x_i)}{P(y_{j_n})}$$

$$= \frac{P_0(x_i) \prod_{k=1}^n P(y_{j_k} | x_i)}{\prod_{k=1}^n P(y_{j_k})} \quad (i=1,2) \quad (7)$$

Then the probability ratio λ_n is given as follows.

$$\lambda_n = \frac{P_n(x_2)}{P_n(x_1)} = \frac{P_0(x_2) \prod_{k=1}^n P(y_{j_k} | x_2)}{P_0(x_1) \prod_{k=1}^n P(y_{j_k} | x_1)} \quad (8)$$

If we assume that

$$P_0(x_1) = P_0(x_2) = \frac{1}{2}$$

Then, we obtain the same ratio as that of SPRT. Therefore, we consider that the Bayes' Updating based on binary channel is a kind of generalization of SPRT.

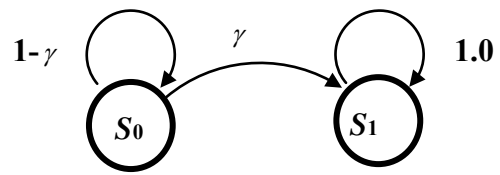
As for the conditional probability, we can generalize to introduce a kind of subjective probability. In addition, we can adopt not only the discrete probability, but also continuous density function.

3. Change Model and Conditional Probability Ratio

3.1. Hidden Markov Model

We present a Hidden Markov Model where the structural change is Poisson occurrence of average γ . In this model, once the change has occurred during the observing period, the structure does not go back to the previous one. The reason why we set such a model is that we focus on the detection of the first structural change in the sequential processing. The concept of the structural change model is shown in Fig.2-4.

Moreover, we introduce a more detailed model. Let R be the probability of the prediction failure that the data value goes out the tolerance zone when the structure is unchanged. Let Rc be the probability of the failure when the structure change has been occurred. We can consider that Rc is greater than R , i.e., $Rc > R$. The detailed model is illustrated as a probabilistic finite state automaton as shown in Fig.2-4.



S0 : State that the structure is unchanged.
S1 : State that the structural change has occurred.
 γ : Probability of the structural change occurrence.
(Poisson Process.)

Fig.2. Structural change model.

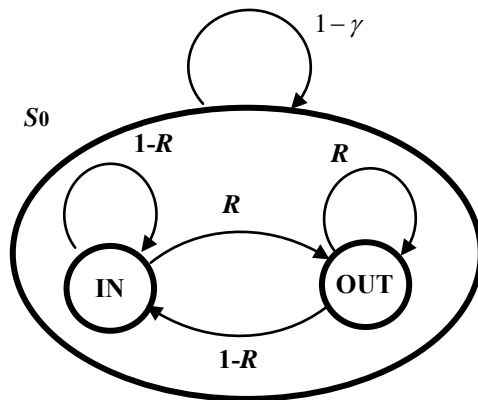


Fig.3. Internal model of the State S0.

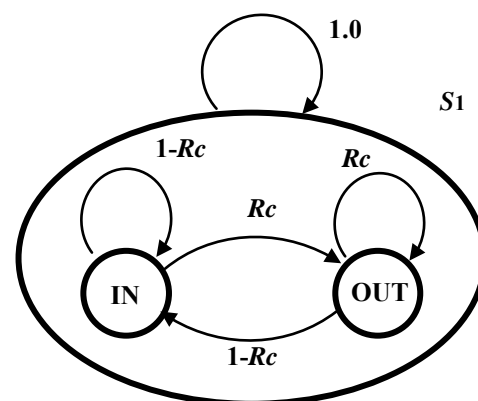


Fig.4. Internal model of the State S1.

3.2. Sequential Probability Ratio

Let $a_1 a_2, \dots, a_i, \dots, a_n$ $a_i \in \{\text{IN}, \text{OUT}\}$ be a string (or symbol sequence) obtained from the observed data.

Let θ_i and $\tilde{\theta}_i$ be the conditional probability that outputs the observed data (or above symbol sequence, $C_n = a_1 a_2 \dots a_n$ in the state S_0 and S_1 , respectively. That is, it means that $\theta_i \in \{R, 1-R\}$ and $\tilde{\theta}_i \in \{R_c, 1-R_c\}$, respectively.

And let $P(a_1 \dots a_n, H_0)$ and $P(a_1 \dots a_n, H_1)$ be the joint probability of the symbol sequence C_n happens with the event H_0 (the structural change is not occurred) and H_1 (the change is occurred), respectively.

Then, the following equations hold.

$$\begin{aligned} P(a_1 \dots a_n, H_0) &= P(C_n, H_0) \\ &= (1-\gamma)^n \theta_1 \dots \theta_n = (1-\gamma)^n \prod_{i=1}^n \theta_i \end{aligned} \quad (9)$$

$$\begin{aligned} P(a_1 \dots a_n, H_1) &= P(C_n, H_1) \\ &= \gamma \prod_{i=1}^n \tilde{\theta}_i + ((1-\gamma)\theta_1)(\gamma \prod_{i=2}^n \tilde{\theta}_i) \\ &\quad + ((1-\gamma)^2 \theta_1 \theta_2)(\gamma \prod_{i=3}^n \tilde{\theta}_i) + \dots \\ &= \sum_{k=1}^n ((1-\gamma)^{k-1} \cdot \prod_{j=0}^{k-1} \theta_j)(\gamma \prod_{i=k}^n \tilde{\theta}_i) \end{aligned} \quad (10)$$

$$\begin{aligned} P(H_0 | a_1 \dots a_n) &= P(H_0 | C_n) = \frac{P(H_0 | C_n) \cdot P(C_n)}{P(C_n)} \\ &= \frac{P(C_n, H_0)}{P(C_n)} = \frac{P(C_n, H_0)}{P(C_n, H_0) + P(C_n, H_1)} \end{aligned} \quad (11)$$

$$\begin{aligned} P(H_1 | a_1 \dots a_n) &= P(H_1 | C_n) = \frac{P(H_1 | C_n) \cdot P(C_n)}{P(C_n)} \\ &= \frac{P(C_n, H_1)}{P(C_n)} = \frac{P(C_n, H_1)}{P(C_n, H_0) + P(C_n, H_1)} \end{aligned} \quad (12)$$

Our New Sequential Probability Ratio (NSPR) of two conditional probabilities that we have previously

shown is represented as follows, using the aforementioned equations.¹⁰

$$\begin{aligned} \text{NSPR} &= \frac{P(H_1 | a_1 \dots a_n)}{P(H_0 | a_1 \dots a_n)} = \frac{P(H_1 | C_n)}{P(H_0 | C_n)} = \frac{P(C_n, H_1)}{P(C_n, H_0)} \\ &= \frac{\sum_{k=1}^n ((1-\gamma)^{k-1} \cdot \prod_{j=0}^{k-1} \theta_j)(\gamma \prod_{i=k}^n \tilde{\theta}_i)}{(1-\gamma)^n \prod_{i=1}^n \theta_i} \end{aligned} \quad (13)$$

On the contrary, there is another type of conditional probability ratio such as used in the SPRT. We call it "Semi Sequential Probability Ratio Test (SSPRT)".

That is,

$$\begin{aligned} \text{SSPRT} &= \frac{P(a_1 \dots a_n | H_1)}{P(a_1 \dots a_n | H_0)} = \frac{P(C_n | H_1)}{P(C_n | H_0)} \\ &= \frac{P(C_n, H_1) / P(H_1)}{P(C_n, H_0) / P(H_0)} \\ &= \left(\frac{P(H_0)}{P(H_1)} \right) \cdot \left(\frac{P(C_n, H_1)}{P(C_n, H_0)} \right) = \left(\frac{P(H_0)}{P(H_1)} \right) \cdot \text{NSPR} \end{aligned} \quad (14)$$

Let the notation $S_0^k \cdot S_1^l$ be an event sequence where, after the state S_0 continuously occurs k times, the state S_1 continuously occurs l times. Then, the $P(H_0)$ and $P(H_1)$ are represented as follows.

$$P(H_0) = P(S_0^n) = (1-\gamma)^n \quad (15)$$

$$\begin{aligned} P(H_1) &= P(S_1^n) + P(S_0 \cdot S_1^{n-1}) + P(S_0^2 \cdot S_1^{n-2}) \\ &\quad + P(S_0^k \cdot S_1^{n-k}) + \dots + P(S_0^{n-1} \cdot S_1) \\ &= \sum_{k=0}^{n-1} P(S_0^k \cdot S_1^{n-k}) = \sum_{k=0}^{n-1} (1-\gamma)^k \cdot \gamma \\ &= 1 - P(H_0) = 1 - (1-\gamma)^n \end{aligned} \quad (16)$$

Then we have the relation between the NSPR and Semi SPRT as follows.

$$\text{NSPR} = \left(\frac{P(H_1)}{P(H_0)} \right) \cdot \text{SSPRT} = \left(\frac{1}{(1-\gamma)^n} - 1 \right) \cdot \text{SSPRT} \quad (17)$$

When the NSPR is greater than 1.0, we regard that the structural change has been occurred before the present time.

In many cases, we can approximately assume the following probabilistic independence.

$$P(a_1 \dots a_n | H_1) = P(a_1 | H_1)P(a_2 | H_1) \dots P(a_n | H_1) \quad (18)$$

$$P(a_1 \dots a_n | H_0) = P(a_1 | H_0)P(a_2 | H_0) \dots P(a_n | H_0) \quad (19)$$

So, we have the following approximate relation.

$$\begin{aligned} \text{SSPRT} &= \frac{P(a_1 \dots a_n | H_1)}{P(a_1 \dots a_n | H_0)} = \frac{P(C_n | H_1)}{P(C_n | H_0)} \\ &= \frac{\prod_{i=1}^n P(a_i | H_1)}{\prod_{i=1}^n P(a_i | H_0)} = \lambda_n = \text{SPRT} \end{aligned} \quad (20)$$

In this case, from Eq.(20) and Eq.(17), we have

$$\text{NSPR} = \left(\frac{P(H_1)}{P(H_0)} \right) \cdot \text{SPRT} = \left(\frac{1}{(1-\gamma)^n} - 1 \right) \cdot \text{SPRT} \quad (21)$$

4. Experimentation for Remote Monitoring Data

As a simple automated change detection method of the scenes from a monitoring camera, we apply the SPRT method and the aforementioned NSPR one to the time series data of compressed JPEG file quantity (Kbyte).

Fig.5 shows the time series images when a person has moved at low speed, in front of a bookshelf.

Fig.6 shows the graph on time series data of JPEG file volumes, corresponding to Fig.5.

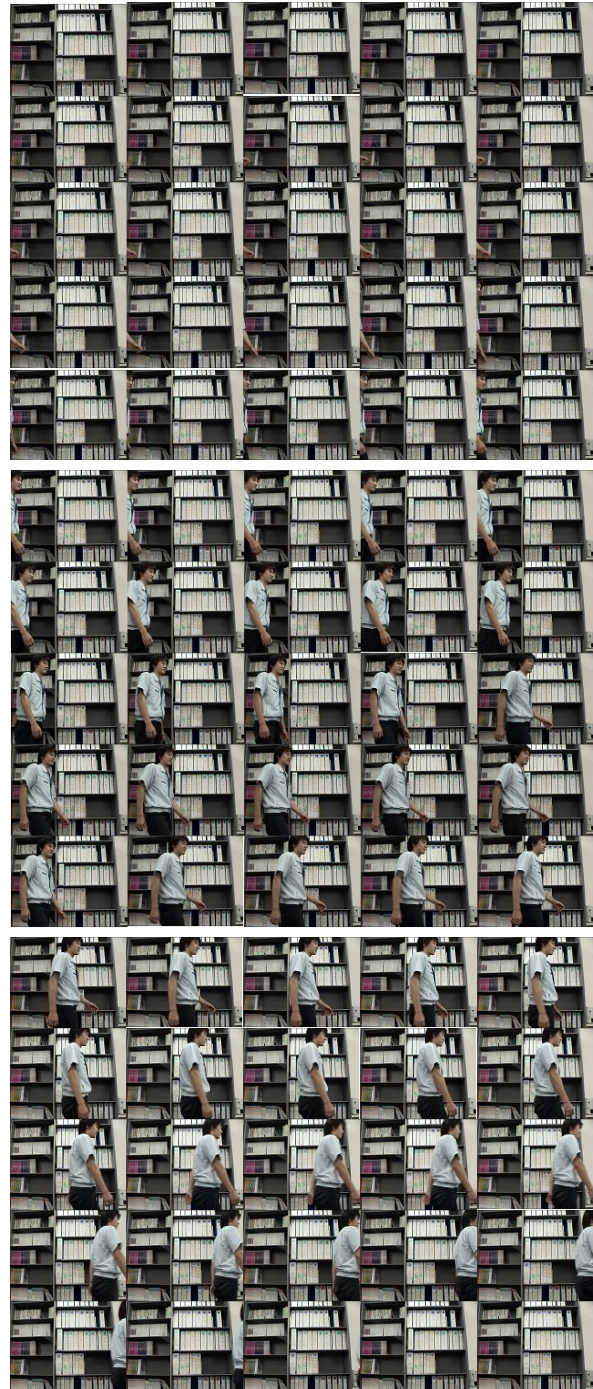


Fig.5. Example of images where someone passes by at low speed (Time is counted from upper left to lower right.)

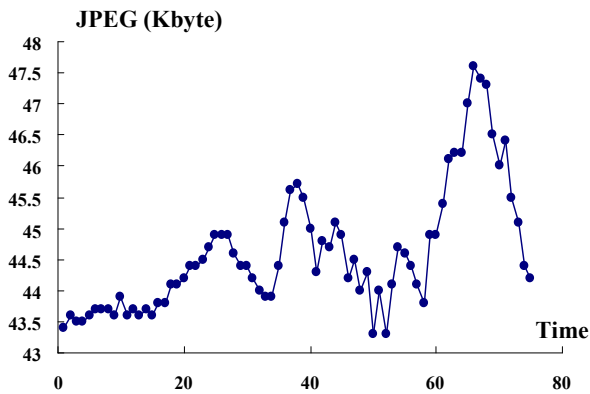


Fig.6. Total time series of image data, i.e., sequence of JPEG file volumes in Fig.5.

We assume that change point is the time point when some change happens, that is, when some person appears in the image data. From the No.8 image, we can see that the hand of a man comes into the image. Also from around the No.21 image, we can see that a part of someone's face gradually comes into the image sequence. The SPRT and NSPR can detect the change point correctly from the sequence of image data. We show experimental results of change detection using two tests (SPRT and Chow Test^{6,7,8}) and one ratio of NSPR.

Generally, several samples (i.e., some images) are used to construct the prediction model in the stage of learning and analysis, where samples are selected for concretely deciding prediction model (regression line) as shown in Fig.7.⁶

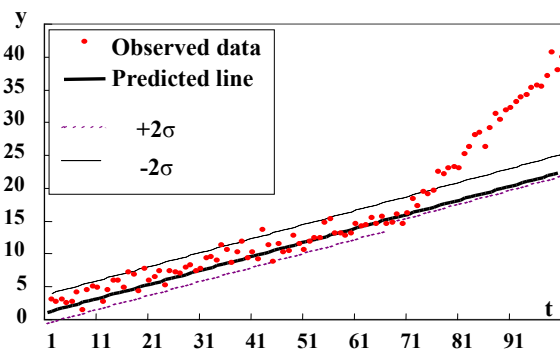


Fig.7. Example of regression line as a predicted model.⁶

In this experimentation, the prediction model is single regression line and the model is constructed by the first six images, i.e., No.1 through No.6 image. And the tolerable zone is the range of 2σ from the predicted value as shown in Fig.7.⁶

The results of change detection by the two Tests (SPRT and Chow Test) and NSPR are shown in the graph of Fig.8.

The SPRT detects change point at $t=21$, in the case of Condition 1 ($\theta_0=0.9, \theta_1=0.1$), and Condition 2 ($\theta_0=0.87, \theta_1=0.13$). The SPRT also detects the change point at $t=22$, under the Condition 3 ($\theta_0=0.8, \theta_1=0.2$).

The NSPR detects the change point at the following time points under the conditions, respectively.

- (a) ($\gamma = 0.001, \theta_i \in \{0.2, 0.8\}, \tilde{\theta}_i \in \{0.8, 0.2\}$): $t=24$,
- (b) ($\gamma = 0.001, \theta_i \in \{0.1, 0.9\}, \tilde{\theta}_i \in \{0.9, 0.1\}$): $t=22$,
- (c) ($\gamma = 0.005, \theta_i \in \{0.2, 0.8\}, \tilde{\theta}_i \in \{0.8, 0.2\}$): $t=23$,
- (d) ($\gamma = 0.005, \theta_i \in \{0.1, 0.9\}, \tilde{\theta}_i \in \{0.9, 0.1\}$): $t=21$.

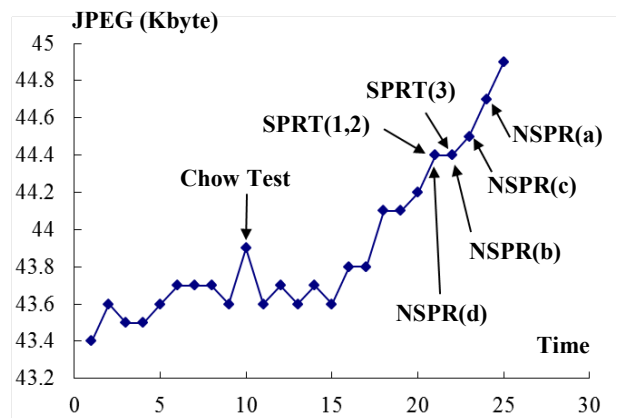


Fig.8. Change point detection (low speed passing-by movement in Fig.5) using Chow Test, SPRT, and NSPR. The "SPRT (1,2)" means the detection by SPRT under the Condition 1 ($\theta_0=0.9, \theta_1=0.1$) or Condition 2 ($\theta_0=0.87, \theta_1=0.13$). Similarly, the "SPRT(3)" means the case of the Condition 3 ($\theta_0=0.8, \theta_1=0.2$). The "NSPR (a) through (d)" mean the detections by NSPR under the following conditions, respectively.

- (a) ($\gamma = 0.001, \theta_i \in \{0.2, 0.8\}, \tilde{\theta}_i \in \{0.8, 0.2\}$),
- (b) ($\gamma = 0.001, \theta_i \in \{0.1, 0.9\}, \tilde{\theta}_i \in \{0.9, 0.1\}$),
- (c) ($\gamma = 0.005, \theta_i \in \{0.2, 0.8\}, \tilde{\theta}_i \in \{0.8, 0.2\}$),
- (d) ($\gamma = 0.005, \theta_i \in \{0.1, 0.9\}, \tilde{\theta}_i \in \{0.9, 0.1\}$).

From the experimental results for time series images of the low speed passing-by movement, we find that the SPRT and NSPR can detect the change quickly and correctly at the time when change happens. Moreover, the SPRT and NSPR method are very simple and take very low computational cost (i.e., required time and memory storage for computation). On the contrary, the Chow Test tends to make a mistake at the early stage before the change happens. In addition, the computational cost is considerably high.

Then, we consider that the performance of both SPRT and NSPR is very effective in comparison with Chow Test, even if the time series data of JPEG file volumes is fluctuated by lighting-influenced images and so on.

5. Conclusion

In this paper, we have presented the equivalent relation between the Bayes' Updating and SPRT (Sequential Probability Ratio Test), using the notion of binary channel. Then, we can find that the SPRT does not contain the irreversible model such as NSPR that we have previously proposed.

We also have presented the definition of the Semi Sequential Probability Ratio Test (SSPRT). And we have shown the approximate relation between NSPR and SPRT, because SSPRT approximately equals SPRT in many cases.

Moreover, based on the experimentations for the time series of JPEG file volumes of the slow movement image data, we have shown that the SPRT and NSPR have almost equivalent performance for the structural change detection, and that they are more effective than Chow Test for the change detection of time series data.

References

1. G.E.P.Box and G.M.Jenkins, *Time Series Analysis: Forecasting and Control*, (Prentice Hall, 1976).
2. Peter J. Brockwell and Richard A. Davis, *Introduction to Time Series and Forecasting*, (Springer; 2nd edition, 2003).
3. C.Han, P.K.Willet and D.A.Abraham, Some methods to evaluate the performance of Page's test as used to detect transient signals, *IEEE Trans. Signal processing*, **47**(8) (1999), pp.2112-2127.
4. S.D.Blostein, Quickest detection of a time-varying change in distribution, *IEEE Trans. Information Theory*, **37**(4) (1991), pp.1116-1122.
5. A. Wald, *Sequential Analysis*, (John Wiley & Sons, 1947).
6. Hiromichi Kawano, Tetsuo Hattori, Ken Nishimatsu (2008), Structural Change Point Detection Method of Time Series Using Sequential Probability Ratio Test – Comparison with Chow Test in the ability of early detection – (in Japanese), (*IEEJ Trans. EIS*, **128**(4) (2008), pp.583-592.
7. Chow, G. C. (1960), *Tests of Equality Between Sets of Coefficients in Two Linear Regressions*, (*Econometrica*, **28**(3), pp.591-605.
8. Katsunori Takeda, Tetsuo Hattori, Izumi Tetsuya, Hiromichi Kawano, Extended SPRT for Structural Change Detection of Time Series Based on Multiple Regression Mode, *International Journal of Artificial Life and Robotics*, (Springer, 2010) ISSN:1433-529, **15**(4), pp.417-420.
9. Tetsuo HATTORI, Hiromichi KAWANO (2011), Change Detection Method of Time Series as an Optimal Stopping Problem --Constructive Proof of Optimal Solution Theorem--, in *Proc. of ICBAKE 2011 (2011 International Conference on Biometrics and Kansei Engineering)*, (IEEE Computer Society, 2011) ISBN 978-0-7695-4512-7, pp.100-105.
10. Yoshihide Koyama, Tetsuo Hattori, Hiromichi Kawano, Model Introduced SPRT for Structural Change Detection of Time Series (I) -- Formulation --, *Journal of Robotics, Networks and Artificial Life*, (Atlantis Press, 2014) ISSN:2352-6386, **1**(1), pp.55-60.