Research Article

Degenerate-Generalized Likelihood Ratio Test for One-Sided Composite Hypotheses

Dongdong Xiang, Xiaolong Pu, Lei Wang, and Yan Li

School of Finance and Statistics, East China Normal University, Shanghai 200241, China

Correspondence should be addressed to Yan Li, yli@stat.ecnu.edu.cn

We propose the degenerate-generalized likelihood ratio test (DGLRT) for one-sided composite hypotheses in cases of independent and dependent observations. The theoretical results show that the DGLRT has controlled error probabilities and stops sampling with probability 1 under some regularity conditions. Moreover, its stopping boundaries are constants and can be easily determined using the provided searching algorithm. According to the simulation studies, the DGLRT has less overall expected sample sizes and less relative mean index (RMI) values in comparison with the sequential probability ratio test (SPRT) and double sequential probability ratio test (2-SPRT). To illustrate the application of it, a real manufacturing data are analyzed.

1. Introduction

Consider the following hypotheses test problem:

$$H_0 : \theta \leq \theta_0 \text{ versus } H_1 : \theta \geq \theta_1 \quad (\theta_0 < \theta_1) \tag{1.1}$$

with the error constraints

$$
\begin{aligned}
P_\theta\{\text{accept } H_1\} &\leq \alpha \quad \text{for } \theta \leq \theta_0 \\
P_\theta\{\text{accept } H_0\} &\leq \beta \quad \text{for } \theta \geq \theta_1.
\end{aligned}
\tag{1.2}
$$

Here, θ_0, $\theta_1 \in \Theta$, and Θ is the parameter space. Sequential tests for the problem (1.1) with independently and identically distributed (i.i.d.) observations have been widely studied. In cases of the one parameter exponential family with monotone likelihood ratio, the sequential probability ratio test (SPRT) proposed by Wald [1] provided an optimal solution to

the problem (1.1), in the sense of minimizing the expected sample sizes (ESSs) at $\theta = \theta_0$ and $\theta = \theta_1$, among all tests satisfying the constraints (1.2).

However, its ESSs at other parameter points are even larger than that of the test methods with fixed sample sizes. This led Weiss [2], Lai [3], and Lorden [4] to consider the problem (1.1) from the minimax perspective. Subsequently, Huffman [5] extended Lorden's [4] results to show that the 2-SPRT provides an asymptotically optimal solution to the minimax sequential test problem (1.1). Instead of the minimax approach, Wang et al. [6] proposed a test minimizing weighted ESS based on mixture likelihood ratio (MLR). Since the ESSs over $[\theta_0, \theta_1]$ are hard to control and are usually focused on applications, Wang et al. [6] paid much attention to investigate the performance of the ESS over $[\theta_0, \theta_1]$. Many tests for the problem (1.1) under independent observations are developed from other perspectives, including [7–11] and so forth.

It is true that in many practical cases the independence is justified, and hence these tests have been widely used. However, such tests may not be effective in cases when the observations are dependent, for example, Cauchy-class process for sea level (cf. [12]), fractional Gaussian noise with long-range dependence (cf. [13, 14]) and the power law type data in cyber-physical networking systems [15]. Especially for the power law data, the sequential tests for dependent observations are particularly desired. This need is not limited to these cases.

So far, many researchers studied sequential tests for various dependent scenarios. Phatarfod [16] extended the SPRT to test two simple hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ when observations constitute a Markov chain. Tartakovsky [17] showed that certain combinations of one-sided SPRT still own the asymptotical optimality in the ESS under fairly general conditions for a finite simple hypotheses. Novikov [18] proposed an optimal sequential test for a general problem of testing two simple hypotheses about the distribution of a discrete-time stochastic process. Niu and Varshney [19] proposed the optimal parametric SPRT with correlated data from a system design point of view. To our best knowledge, however, there are few references available for considering the problem (1.1) with dependent observations from the perspective of minimizing the ESS over $[\theta_0, \theta_1]$. Similar to Wang et al. [6], one can extend the MLR to the dependent case. However, unlike the i.i.d. case, the MLR under the dependent case may not be available because of the complexity of its computation. Besides, its test needs to divide $[\theta_0, \theta_1]$ into two disjoint parts by inserting a point. In i.i.d. cases, this point can be selected following Huffman's [5] suggestion. But, in the dependent case, this suggestion may not be effective. One also can use the generalized likelihood ratio (GLR) instead of the MLR. Unfortunately, as opposite to the MLR, the GLR does not preserve the martingale properties which allow one to choose two constant stopping boundaries in a way to control two types of error. Moreover, the computation of the GLR is hard to be obtained in cases when the maximum likelihood estimator should be searched. This usually happens in the dependent case.

In this paper, we propose a test method for both dependent and independent observations. It has the following features: (1) it has good performances over $[\theta_0, \theta_1]$ in the sense of less overall expected sample sizes; (2) its computation is reasonably simple; (3) its stopping boundaries can be determined conveniently. The rest of the paper is organized as follows. In Section 2, we describe the construction of the proposed test in details and present its basic theoretical properties. Based on these theoretical results, we provide a searching algorithm to compute stopping boundaries for our proposed test. In Section 3, we conduct some simulation studies to show the performance of the proposed test. Some concluding remarks are given in Section 4. Some technical details are provided in the appendix.

2. The Proposed Test

Let $x^i =: (x_1, x_2, \ldots, x_i)$, $i = 1, 2, \ldots$ and suppose that the conditional probability distribution of each $x_i | x^{i-1}$, $f(x_i | x^{i-1}, \theta)$ has an explicit form. Here, $x_1 | x^0 =: x_1$ and $f(x_1 | x^0, \theta) =: f(x_1, \theta)$. Thus, likelihood ratio can be defined as

$$R_n(\theta, \theta') = \prod_{i=1}^{n} \frac{f(x_i \mid x^{i-1}, \theta)}{f(x_i \mid x^{i-1}, \theta')}, \quad \theta, \theta' \in \Theta. \tag{2.1}$$

Lai [20] introduced this model to construct a sequential test for many simple hypotheses when the observations are dependent. It is very general and also includes the i.i.d. cases.

Example 2.1. Consider, for instance, a simple nonlinear time series model:

$$x_i = \theta x_{i-1}^2 + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1). \tag{2.2}$$

In this case, $R_n(\theta, \theta') = \prod_{i=1}^{n} \phi(x_i - \theta x_{i-1}^2) / \phi(x_i - \theta' x_{i-1}^2)$, $x_0 = 0$, and $\phi(\cdot)$ is the probability density function of the standard normal distribution.

To overcome the difficulty stated in Section 1, we propose a test statistic which minimizes the likelihood ratio with restriction to a finite parameter points in $[\theta_0, \theta_1]$. First, we insert k (≥ 3) points into $[\theta_0, \theta_1]$ uniformly, denoted as $\tilde{\theta}_i$ with $\tilde{\theta}_i = \theta_0 + (i-1)(\theta_1 - \theta_0)/(k-1)$, $i = 1, \ldots, k$. Next, we define the test statistic as $\max_{1 \leq i \leq k} R_n(\tilde{\theta}_i, \theta')$. It can be checked that this test statistic not only preserves the martingale properties, but also inherits the merit of the GLR. As long as k is not very large (e.g., $k > 100$), its computation will be very simple. Thus, it has all the three features stated in Section 1. Since this maximization is restricted to some finite points, we refer to it as degenerate-generalized likelihood ratio (DGLR).

Based on the DGLR, we define a stopping rule T for the problem (1.1) by

$$T = \inf \left\{ n \geq 1, \max_{1 \leq i \leq k} R_n(\tilde{\theta}_i, \theta_0) \geq A \text{ or } \max_{1 \leq i \leq k} R_n(\tilde{\theta}_i, \theta_1) \geq B \right\}, \tag{2.3}$$

with the terminal decision rule

$$\Delta = \begin{cases} \text{accept } H_1, & \max_{1 \leq i \leq k} R_T(\tilde{\theta}_i, \theta_0) \geq A, \\ \text{accept } H_0, & \max_{1 \leq i \leq k} R_T(\tilde{\theta}_i, \theta_1) \geq B, \\ \text{continue sampling}, & \text{else}, \end{cases} \tag{2.4}$$

where $0 < A, B < \infty$ are two stopping boundaries. Hereafter, the sequential test method with (2.3) and (2.4) is called the degenerate-generalized likelihood ratio test (DGLRT). It has some theoretical properties which are stated as follows. These theoretical properties provide a guide to the design of the DGLRT, whose proofs are provided in the appendix.

Let

$$\alpha'(\theta, A, B) = P_\theta \left\{ \max_{1 \leq i \leq k} R_T\left(\tilde{\theta}_i, \theta_0\right) \geq A \right\}, \quad \theta \in \Theta_0,$$

$$\beta'(\theta, A, B) = P_\theta \left\{ \max_{1 \leq i \leq k} R_T\left(\tilde{\theta}_i, \theta_1\right) \geq B \right\}, \quad \theta \in \Theta_1 \tag{2.5}$$

be the real error probabilities, where Θ_0 and Θ_1 represent the parameter subsets under H_0 and H_1, respectively.

Proposition 2.2. *Suppose*

$$\int \frac{f\left(x_i \mid x^{i-1}, \theta''\right)}{f\left(x_i \mid x^{i-1}, \theta'\right)} f\left(x_i \mid x^{i-1}, \theta\right) dx_i \leq 1, \tag{2.6}$$

for any positive integer n and every triple $\theta \leq \theta' \leq \theta''$. For the DGLRT defined by (2.3) and (2.4), one has $\alpha'(\theta, A, B) \leq k/A$ for all $\theta \in \Theta_0$ and $\beta'(\theta) \leq k/B$ for all $\theta \in \Theta_1$.

Remark 2.3. The assumption (2.6) given in Proposition 2.2 is not restrictive. This holds for the general one parameter exponential family and many others (cf. Robbins and Siegmund [21]).

Proposition 2.4. *Suppose that there exists a constant $\varepsilon > 0$ such that $E_{\theta''}[\log\{f(x_i|x^{i-1};\theta')\} - \log\{f(x_i|x^{i-1};\theta)\}] \geq \varepsilon$ for all i and every triple $\theta \leq \theta' \leq \theta''$. Under the assumptions stated in Proposition 2.2, one has $P_\theta\{T < \infty\} = 1$ for all $\theta \in \Theta$.*

Remark 2.5. For $\theta'' \geq \theta'$, we have

$$E_{\theta''}\left[\log\left\{f\left(x_i \mid x^{i-1};\theta'\right)\right\} - \log\left\{f\left(x_i \mid x^{i-1};\theta\right)\right\}\right]$$

$$= -E_{\theta''}\left[\log\left\{f\left(x_i \mid x^{i-1};\theta\right)\right\} - \log\left\{f\left(x_i \mid x^{i-1};\theta'\right)\right\}\right]$$

$$\geq -\log\left\{E_{\theta''}\left[\frac{f\left(x_i \mid x^{i-1};\theta\right)}{f\left(x_i \mid x^{i-1};\theta'\right)}\right]\right\} \tag{2.7}$$

$$\geq 0.$$

The last inequality follows from (2.6). $E_{\theta''}[\log\{f(x_i|x^{i-1};\theta')\} - \log\{f(x_i|x^{i-1};\theta)\}]$ is positive with probability 1 if $\theta \neq \theta'$. Heuristically, the requirement that the difference be greater than the constant $\varepsilon > 0$ for all i amounts to assuming that the sequence of data cumulatively adds information about all the $\theta'' \geq \theta'$, which is generally true in sequential studies.

From Proposition 2.2, we conclude that the DGLRT satisfies the error constraints (1.2) if $A = k/\alpha$ and $B = k/\beta$. From Proposition 2.4, it is easy to find that we absolutely stop sampling after finite observations. These results imply that the DGLRT can be useful in a sequential study for testing the problem (1.1).

In the DGLRT (2.3) and (2.4), the value of the parameter k should be large but finite. In practice, we suggest that $k = 10$ (cf. Section 3). Regarding A and B, we can compute them by simulation. Proposition 2.2 shows $A \leq k/\alpha$ and $B \leq k/\beta$. Thus, we can search (A, B) over

Table 1: The ESSs at $\theta = -0.8$ (0.1) 0 for $-\theta_0 = \theta_1 = 0.5$ and $\alpha = \beta = 0.01$.

θ	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0
$k = 3$	6.293	7.121	8.181	9.545	11.241	13.156	15.136	16.600	17.141
$k = 5$	6.355	7.173	8.254	9.557	11.172	13.073	14.958	16.396	16.974
$k = 10$	6.380	7.228	8.264	9.577	11.138	13.039	14.897	16.344	16.889
$k = 50$	6.394	7.231	8.265	9.578	11.113	13.070	14.872	16.321	16.862

$[1, k/\alpha] \times [1, k/\beta]$ with the real error probabilities being computed by simulations. One may consider a density grid searching on $[1, k/\alpha] \times [1, k/\beta]$. But this is a time consuming job. To reduce the computation, we introduce an efficient approach as follows. In the first step, we can use bisection searching to find A_1 ($\in [1, k/\alpha]$) such that $\alpha'(\theta_0, A_1, k/\beta) = \alpha$. Then, fix A_1 to find B_1 ($\in [1, k/\beta]$) such that $\beta'(\theta_1, A_1, B_1) = \beta$. Since $\alpha'(\theta_0, x, y)$ and $1 - \beta'(\theta_1, x, y)$ increase in x and decrease in y, we conclude that $(A, B) \in [1, A_1] \times [1, B_1]$. Hence, we repeat the above step over $[1, A_1] \times [1, B_1]$. In this way, we generate a sequence of pairs $(A_1, B_1), (A_2, B_2), \ldots$. Following the above program, we have

$$A_1 \geq A_2 \geq \cdots \geq 1, \qquad B_1 \geq B_2 \geq \cdots \geq 1. \tag{2.8}$$

It can be checked that these pairs converge to the exact stopping boundaries. In practice, we repeat the above process and stop at step l if $|\alpha'(\theta_0, A_l, B_l) - \alpha| \leq \text{tol}_1$ and $|\beta'(\theta_1, A_l, B_l) - \beta| \leq \text{tol}_2$. Here, $\text{tol}_1 = 2\%\alpha$ and $\text{tol}_2 = 2\%\beta$. Computation involved in finding A and B is not difficult partly due to the rapid developments in information technology. For example, in the nonlinear time series model (2.2), setting $-\theta_0 = \theta_1 = 0.25$, $\alpha = 0.01$, $\beta = 0.05$, and $k = 10$, it requires 15 minutes to obtain the stopping boundaries A and B for the DGLRT based on 100,000 simulations, using Intel-Core i7-2.80 GHz CPU. Since this is a one-time computation before testing, it is convenient to accomplish.

3. Numerical Studies

In this section, we present some simulation results regarding the numerical performance of the proposed DGLRT. In the DGLRT, the parameter k needs to be chosen. We first investigate the effect of k on the performance of the DGLRT according to i.i.d. observations from the normal distribution $N(\theta, 1)$. Setting $-\theta_0 = \theta_1 = 0.5$ and $\alpha = \beta = 0.01$, we compare the DGLRTs with $k = 3, 5, 10, 50$. The corresponding stopping boundaries (A, B) are $(69.3, 69.3)$, $(74.3, 74.3)$, $(75.7, 75.7)$, and $(76.7, 76.7)$, respectively. The ESSs at $\theta = -0.8$ (0.1) 0.8 (i.e., θ takes values from -0.8 to 0.8 with step 0.1) are computed based on 100,000 simulated data and are provided in Table 1.

Because of the symmetry, we only include results for $\theta \in [-0.8, 0]$. Table 1 shows that the ESSs under a larger k are smaller than those under a smaller k if $\theta \in (\theta_0, \theta_1)$. Meanwhile, it can be seen that a smaller k has a better performance outside (θ_0, θ_1). In order to assess the overall performance of the tests, we compute their relative mean index (RMI) values.

The RMI is introduced by Han and Tsung [22] for comparing the performance of several control charts. It is defined as

$$\text{RMI} = \frac{1}{N} \sum_{l=1}^{N} \frac{\text{ESS}(\theta_l) - \text{MESS}(\theta_l)}{\text{MESS}(\theta_l)}, \tag{3.1}$$

where N is the total numbers of parameter points (i.e., θ_l's) we considered, $\text{ESS}(\theta_l)$ denotes the ESS at θ_l, and $\text{MESS}(\theta_l)$ is the smallest one among all the three $\text{ESS}(\theta_l)$. So, $(\text{ESS}(\theta_l) - \text{MESS}(\theta_l))/\text{MESS}(\theta_l)$ can be considered as a relative difference of the given test, compared to the best test, at θ_l, and RMI is the average of all such difference values. By this index, a test with smaller RMI value is considered better in its overall performance. Since we focus on the performance over the parameter interval $[\theta_0, \theta_1]$, $\theta_l = -0.5 + 0.1(i - 1)$, $i = 1, \ldots, 10$ in this illustration. The resulting RMIs for the DGLRT under $k = 3, 5, 10, 50$ are 0.0116, 0.0042, 0.0017, and 0.0011, respectively, which shows that the DGLRT under a larger k is more efficient than the one under a smaller k. The improvement is minor when k is large enough. Considering the complexity of computation, we select $k = 10$ for practical purposes. From now on, the DGLRT is always the DGLRT under $k = 10$ unless otherwise stated.

Next, we investigate the performance of the DGLRT in controlling the ESSs over $[\theta_0, \theta_1]$. In the i.i.d. case, we know the 2-SPRT has a better performance in controlling the maximum ESS. For the ESSs over the neighborhoods of θ_0 and θ_1, the SPRT provides a closely approximation. Based on extensive simulations, we conclude that these features still preserve in the dependent case. Therefore, the SPRT and the 2-SPRT are compared with the DGLRT in this paper. The following three cases are considered.

Case 1. Observations collected from normal distributions with mean θ and variance 1. Set $-\theta_0 = \theta_1 = 0.5$ and $\alpha = \beta = 0.01$ for the test problem (1.1).

Case 2. Observations collected from exponential distributions with mean $1/\theta$. The problem (1.1) is set with $\theta_0 = 0.5$, $\theta_1 = 2$, and $\alpha = \beta = 0.01$.

Case 3. Consider the test problem (1.1) for the simple nonlinear time series model (2.2) with $\theta_0 = 0$, $\theta_1 = 1$ and $\alpha = \beta = 0.01$.

In each case, the inserted point for the 2-SPRT is searched over $[\theta_0, \theta_1]$. The stopping boundaries are also computed following the searching algorithm stated in Section 2. These stopping boundaries (A, B) are listed in the order of the SPRT, 2-SPRT, and DGLRT: Case 1: $(56.4, 56.4)$, $(37.4, 37.4)$, and $(75.7, 75.7)$; Case 2: $(63.8, 25.5)$, $(42.5, 23.5)$, and $(79.5, 39.5)$; and Case 3: $(14.5, 25.5)$, $(8.2, 26.8)$, and $(22.5, 36.5)$. Figures 1–3 display the ESS curves over $[\theta_0 - 0.5, \theta_1 + 0.5]$ under the three tests for Cases 1–3 with the dashed line for the SPRT, the dotted line for the 2-SPRT, and the solid line for the DGLRT. Figure 1 shows that the DGLRT is comparable to the 2-SPRT in the middle of the parameter range and performs as well as the SPRT in the two tails. It implies that the DGLRT controls both the maximum ESS and the ESSs under H_0 and H_1 very well. The same conclusions can also be obtained from Figures 2 and 3. The RMIs for the SPRT, 2-SPRT, and DGLRT under the three cases are also computed. The results are listed in Table 2. It can be seen that the RMI for the DGLRT is the smallest one among the three tests under all three cases. Thus, the DGLRT performs the best, compared with the SPRT and the 2-SPRT over $[\theta_0, \theta_1]$.
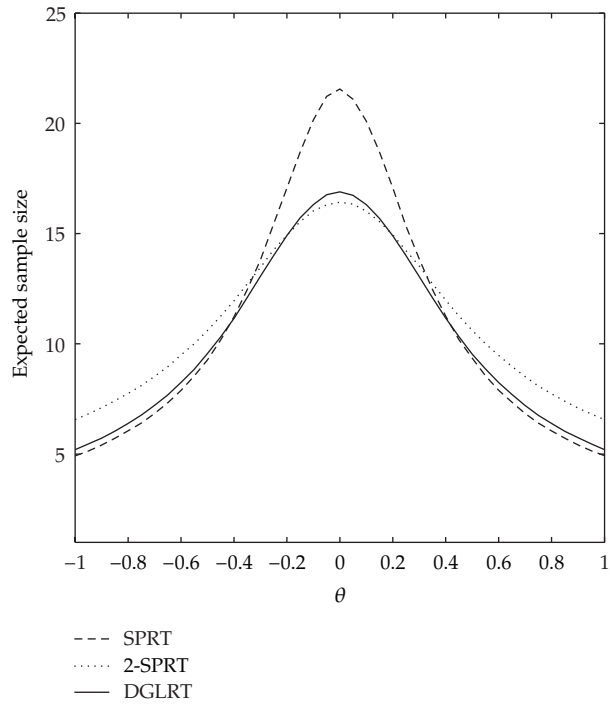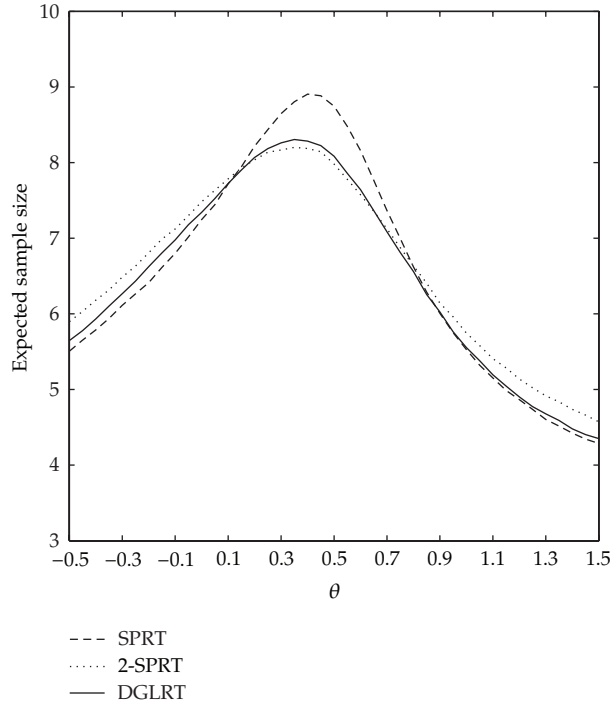
Figure 1: Comparison of ESS curves under the SPRT, the 2-SPRT, and the DGLRT for Case 1: $-\theta_0 = \theta_1 = 0.5$ for the normal distribution with mean θ and variance 1.
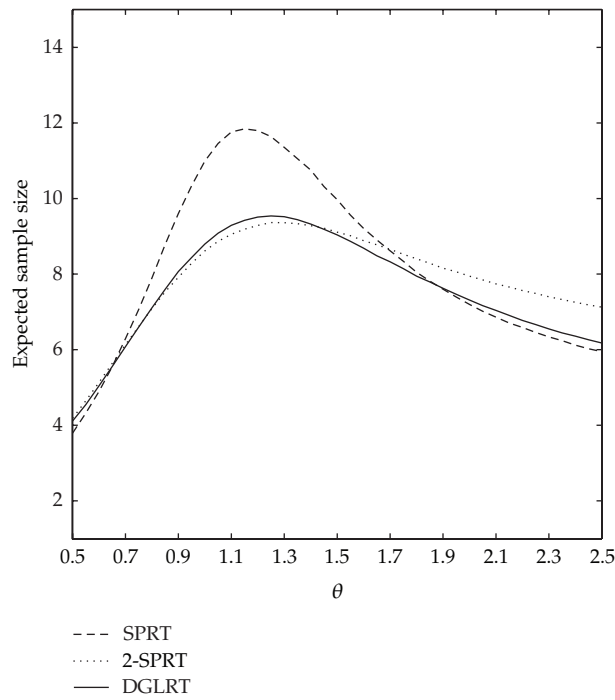
Figure 2: Comparison of ESS curves under the SPRT, the 2-SPRT, and the DGLRT for Case 2: $\theta_0 = 0.5$ and $\theta_1 = 2$ for the exponential distribution with mean $1/\theta$.

Table 2: The RMI for the SPRT, 2-SPRT, and DGLRT under Cases 1–3.

Case	The SPRT	The 2-SPRT	The DGLRT
1	0.1194	0.0402	0.0103
2	0.1148	0.0263	0.0135
3	0.0370	0.0105	0.0059

Figure 3: Comparison of ESS curves under the SPRT, the 2-SPRT, and the DGLRT for Case 3: $\theta_0 = 0$ and $\theta_1 = 1$ for the nonlinear time series (2.2).

To illustrate the DGLRT, we apply it to a real manufacturing data (cf. Chou et al. [23]). A customer specifies an average breaking strength of a strapping tape as 200 psi, and the standard deviation is 12 psi. The data are the breaking strength of different strapping tapes, so the random errors mainly stem from the measurement errors. Thus, the observations can be assumed to be independent. The Shapiro and Wilk [24] test shows that the data are taken from a normal distribution. Consider the test problem (1.1) with $\theta_0 = 200$ and $\theta_1 = 212$ and standardize the observations by using a transformation $X_i \rightarrow (X_i - 206)/12, i = 1, 2, \ldots$. Then the resulting test problem is equivalent to $H_0 : \theta \leq -0.5$ versus $H_1 : \theta \geq 0.5$. Under $\alpha = \beta = 0.01$, the corresponding stopping boundaries for the DGLRT are $(75.7, 75.7)$. Based on the first 20 real observations, we compute the test statistics of the DGLRT, which are displayed in Table 3. In Table 3, standardized X_i indicates $(X_i - 206)/12$. Table 3 shows that $\max_{1 \leq j \leq k} R_i(\widetilde{\theta}_j, \theta_1)$ increases in i rapidly, while $\max_{1 \leq j \leq k} R_i(\widetilde{\theta}_j, \theta_0)$ keeps constant for $i = 1, 2, \ldots, 20$ under the real data. Since $\max_{1 \leq j \leq k} R_i(\widetilde{\theta}_j, \theta_1)$ crosses its stopping boundary at the 11th observation, we should accept the null hypothesis according to the terminal decision rule (2.4).

Table 3: Implementation of the DGLRT with the first 20 observations of breaking strength of a strapping tape.

		Standardized	The DGLRT	
i	X_i	X_i	$\max_{1 \leq j \leq k} R_i(\theta_j, \theta_0)$	$\max_{1 \leq j \leq k} R_i(\theta_j, \theta_1)$
1	191	−1.250	1	3.490
2	193	−1.083	1	10.309
3	204	−0.167	1	12.182
4	215	0.750	1	5.755
5	182	−2.000	1	42.521
6	223	1.417	1	10.309
7	194	−1.000	1	28.022
8	202	−0.333	1	39.095
9	214	0.667	1	20.065
10	210	0.333	1	14.382
11	186	−1.667	1	**76.172**
12	211	0.417	1	50.199
13	202	−0.333	1	70.035
14	201	−0.417	1	106.272
15	191	−1.250	1	370.925
16	193	−1.083	1	1095.537
17	196	−0.833	1	2519.964
18	189	−1.417	1	10394.166
19	194	−1.000	1	28254.274
20	209	0.250	1	22004.450

4. Concluding Remarks

In this paper, we have proposed the DGLRT test in cases where the conditional density function has an explicit form. It has been shown that the properties of the DGLRT can guarantee bounding two error probabilities. To make our method be more applicable, we further discuss the selection of the parameter k and the searching algorithm for its stopping boundaries. From our numerical results, we conclude that the DGLRT has several merits: (1) in contrast to the SPRT, the DGLRT has much smaller ESS for θ in the middle of the parameter range and nearly has the same performance for θ outside the interval (θ_0, θ_1). It is not surprising that the 2-SPRT performs the best in minimizing the maximum ESS because it is designed to be optimal in the minimax sense. However, the relative difference of the maximum ESS between the DGLRT and the 2-SPRT is minor. Moreover, for θ outside (θ_0, θ_1), the ESSs of the DGLRT are much smaller than those of the 2-SPRT. That is to say, the DGLRT controls the maximum ESS and the ESSs under two hypotheses; (2) under the RMI criteria, the DGLRT performs more efficiently than the SPRT and the 2-SPRT over $[\theta_0, \theta_1]$; (3) its implementation is very simple.

While our focus in this paper is on methodological development, there are still some related questions unanswered yet. For instance, at this moment, we do not know how to determine the critical stopping boundaries for the DGLRT in an analytical way instead of the Monte Carlo method. Besides, our method controls the ESS in pointwise, so it can be used to construct control chart for detecting the small shifts. These questions will be addressed in our future research.

Appendix

Proof of Proposition 2.2. Let

$$
\begin{aligned}
T_1 &= \inf\left\{ n \geq 1, \ \max_{1 \leq i \leq k} R_n\left(\tilde{\theta}_i, \theta_0\right) \geq A \right\}, \\
T_2 &= \inf\left\{ n \geq 1, \ \max_{1 \leq i \leq k} R_n\left(\tilde{\theta}_i, \theta_1\right) \geq B \right\}.
\end{aligned}
\tag{A.1}
$$

So,

$$
\begin{aligned}
\alpha'(\theta, A, B) &= P_\theta\{\text{accept } H_1\} = P_\theta\left\{ T < \infty, \max_{1 \leq i \leq k} R_T\left(\tilde{\theta}_i, \theta_0\right) \geq A \right\} \\
&= P_\theta\left\{ T_1 \leq T_2, T < \infty, \max_{1 \leq i \leq k} R_T\left(\tilde{\theta}_i, \theta_0\right) \geq A \right\} \\
&\leq P_\theta\{T_1 < \infty\} \leq \int_{\{T_1 < \infty\}} \frac{1}{A} \max_{1 \leq i \leq k} R_{T_1}\left(\tilde{\theta}_i, \theta_0\right) dP_\theta \\
&\leq \sum_{i=1}^{k} \frac{1}{A} \int_{\{T_1 < \infty\}} R_{T_1}\left(\tilde{\theta}_i, \theta_0\right) dP_\theta \\
&\leq \frac{k}{A}.
\end{aligned}
\tag{A.2}
$$

The last inequality follows from (2.6). Till now, we prove that the result $\alpha'(\theta, A, B) \leq k/A$ for all $\theta \in \Theta_0$. The other result can also be proven in a similar way. $\qquad\square$

Proof of Proposition 2.4. Since we insert k (≥ 3) points in $[\theta_0, \theta_1]$, we can find a point θ_2 which belongs to (θ_0, θ_1). Thus, there exists a $\varepsilon > 0$ such that $E_\theta[\log\{f(x_i|x^{i-1}; \theta_2)\} - \log\{f(x_i|x^{i-1}; \theta_0)\}] \geq \varepsilon$. It implies that $E_\theta[R_n(\theta_2, \theta_0)] \to \infty$ for $\theta \geq \theta_2$. So,

$$
\lim_{n \to \infty} P_\theta\left\{ \max_{1 \leq i \leq k} R_n\left(\tilde{\theta}_i, \theta_0\right) \geq A \right\} \geq \lim_{n \to \infty} P_\theta\{R_n(\theta_2, \theta_0) \geq A\} = 1.
\tag{A.3}
$$

Thus, we have the result that $P_\theta\{T < \infty\} = 1$ for all $\theta \geq \theta_2$. In a similar way, we can obtain $P_\theta\{T < \infty\} = 1$ for all $\theta \leq \theta_2$. Combining the two results, we complete this proof. $\qquad\square$

Acknowledgments

References

[1] A. Wald, "Sequential tests of statistical hypotheses," *Annals of Mathematical Statistics*, vol. 16, pp. 117–186, 1945.

[2] L. Weiss, "On sequential tests which minimize the maximum expected sample size. ," *Journal of the American Statistical Association*, vol. 57, pp. 551–566, 1962.

[3] T. L. Lai, "Optimal stopping and sequential tests which minimize the maximum expected sample size," *The Annals of Statistics*, vol. 1, pp. 659–673, 1973.

[4] G. Lorden, "2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size," *The Annals of Statistics*, vol. 4, no. 2, pp. 281–291, 1976.

[5] M. D. Huffman, "An efficient approximate solution to the Kiefer-Weiss problem," *The Annals of Statistics*, vol. 11, no. 1, pp. 306–316, 1983.

[6] L. Wang, D. Xiang, X. Pu, and Y. Li, "Double sequential weighted probability ratio test for one-sided composite hypotheses," *Communication in Statistics—Theory and Method*. In press.

[7] T. L. Lai, "Nearly optimal sequential tests of composite hypotheses," *The Annals of Statistics*, vol. 16, no. 2, pp. 856–886, 1988.

[8] B. Darkhovsky, "Optimal sequential tests for testing two composite and multiple simple hypotheses," *Sequential Analysis*, vol. 30, no. 4, pp. 479–496, 2011.

[9] H. P. Chan and T. L. Lai, "Importance sampling for generalized likelihood ratio procedures in sequential analysis," *Sequential Analysis*, vol. 24, no. 3, pp. 259–278, 2005.

[10] Y. Li and X. Pu, "Method of sequential mesh on Koopman-Darmois distributions," *Science China A*, vol. 53, no. 4, pp. 917–926, 2010.

[11] Y. Li and X. Pu, "A method for designing three-hypothesis test problems and sequential schemes," *Communications in Statistics—Simulation and Computation*, vol. 39, no. 9, pp. 1690–1708, 2010.

[12] M. Li, "A class of negatively fractal dimensional Gaussian random functions," *Mathematical Problems in Engineering*, Article ID 291028, 18 pages, 2011.

[13] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 10.1155/2011/654284, 13 pages, 2011.

[14] M. Li and W. Zhao, "Variance bound of ACF estimation of one block of fGn with LRD," *Mathematical Problems in Engineering*, vol. 2010, Article ID 560429, 14 pages, 2010.

[15] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, 13 pages, 2012.

[16] R. M. Phatarfod, "Sequential analysis of dependent observations. I," *Biometrika*, vol. 52, pp. 157–165, 1965.

[17] A. Tartakovsky, "Asymptotically optimal sequential tests for nonhomogeneous processes," *Sequential Analysis*, vol. 17, no. 1, pp. 33–61, 1998.

[18] A. Novikov, "Optimal sequential tests for two simple hypotheses," *Sequential Analysis*, vol. 28, no. 2, pp. 188–217, 2009.

[19] R. Niu and P. K. Varshney, "Sampling schemes for sequential detection with dependent observations," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, part 2, pp. 1469–1481, 2010.

[20] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.

[21] H. Robbins and D. Siegmund, "A class of stopping rules for testing parametric hypotheses," in *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4, pp. 37–41, University of California Press, Berkeley, Calif, USA, 1973.

[22] D. Han and F. Tsung, "A reference-free Cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 368–386, 2006.

[23] Y.-M. Chou, R. L. Mason, and J. C. Young, "The SPRT control chart for standard deviation based on individual observations," *Quality Technology & Quantitative Management*, vol. 3, no. 3, pp. 335–345, 2006.

[24] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality: complete samples," *Biometrika*, vol. 52, pp. 591–611, 1965.