Statistica Sinica Preprint No: SS-2024-0334	
Title	Optimal Robust Sequential Tests of Circular Nonconforming Probability
Manuscript ID	SS-2024-0334
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0334
Complete List of Authors	Qunzhi Xu and Yajun Mei
Corresponding Authors	Qunzhi Xu
E-mails	xuqunzhi66@gmail.com
Notice: Accepted author version.	

Statistica Sinica

Optimal Robust Sequential Tests of Circular Nonconforming Probability

Qunzhi Xu and Yajun Mei

Department of Biostatistics

School of Global Public Health, New York University

Abstract: We investigate the minimax nonparametric formulation of sequentially testing hypothesis on the circular nonconforming probability (CNP) that refers to the chance of the system missing a pre-specified 2D disk target. Such a problem occurs in the military science of ballistics, GPS, and GSM, etc., where we want to use as few as samples to assess the precision quality of the 2D system, but we do not make any parametric assumptions on the true underlying distributions for the observed raw data. We show that a Bernoulli sequential probability ratio test (SPRT) is optimal in the sense of minimizing the maximum expected sample sizes among all (fixed-sample or sequential) tests with the same or smaller Type I and Type II error probabilities. Since asymptotic theories in sequential analysis often assume small error probabilities, which are not always feasible in practice, we also propose algorithms to suitably design and implement Bernoulli SPRTs that are simple but useful for practitioners in real world applications.

Key words and phrases: Design of experiments, Kullback-Leibler divergence, se-

quential detection, sequential probability ratio test (SPRT).

1. Introduction

This research is motivated from the problem of testing a system's precision quality. It occurs in many important real-world applications such as the military science of ballistics, GPS (global positioning system) and GSM (global system for mobile communications). In the standard setting, one observes bivariate data (X_i, Y_i)'s that are independent and identically distributed (iid), and wants to make a quick but accurate inference about the circular nonconforming probability CNP $= \mathbf{P}(X_i^2 + Y_i^2 > r^2)$, or equivalently, about the circular conforming probability CCP $= 1 - \text{CNP} = \mathbf{P}(X_i^2 + Y_i^2 \leq r^2)$. These two probabilities correspond to the probabilities of the system (bombs, missiles, etc.) missing or hitting a disk target with the radius r.

There are several standard approaches from the statistical inference viewpoint. The first one is the point estimation approach where one wants to estimate the desired radius r that attains a given CNP, and some popular choices of CNP are 0.1 or 0.5. In particular, the desired r value with CNP $= \mathbf{P}(X_i^2 + Y_i^2 > r^2) = 0.5$ is termed as Circular Error Probable (CEP). Extensive research has been done on the point estimation approach under different model assumptions and contexts, see Fraser [8], Gillis [9], Harter [12], Liu, Duan and Yan [18], Pyati [22] and Shnidman [24].

The second one is the hypothesis testing approach where one tests the hypothesis on CNP for a pre-specified radius r. Research is rather limited along this direction, and for some exceptions, see Li, Pu and Xiang [17] and Li and Mei [16]. The hypothesis testing problem often occurs in ballistic applications, where new products such as bombs, missiles, bullets have to be tested to see whether it meets certain pre-specified requirements or not. In such applications, it is often very expensive to collect observations, and thus it is desired to use as few samples as possible to evaluate the system's precision quality. To speed up the decision making process, one standard statistical idea is to adopt sequential hypothesis testing, as the sequential test is economical in the sense that it may reach a decision earlier than via a fixed-sample-size test. The subfield of sequential hypothesis testing was originally developed by Wald [31] and has been extensively studied in the past seventy years, see classical books by Wald [32], Siegmund [25], Poor and Hadjiliadis [21], Tartakovsky, Nikiforov and Basseville [27]. Some nonparametric works in the sequential literature includes Bhattacharya and Frierson[3], Sen [23], Gordan and Pollak [10], Lhéritier and Cazals[14]. New real-world applications continue to inspire its modern developments, see, for example, Baron and Xu [1], Bartroff and Song [2], Chaudhuri and Fellouris

[5], De and Baron [6], He and Bartroff [13], Li et al. [15], Song and Fellouris [26], Xing and Fellouris [34]. Most of these work deal with asymptotic optimality when error probabilities go to 0, which are interesting from the theoretical viewpoint, but might not be feasible in practice. Here we feel that the CNP in ballistic applications provides another interesting motivation for new research directions in sequential hypothesis testing.

Statistically speaking, the problem of testing CNP for a pre-specified radius r can be stated as utilizing observed bivariate data (X_i, Y_i)'s to sequentially testing hypotheses on

$$\begin{aligned} &H_0 : \mathbf{P}(X_i^2 + Y_i^2 > r^2) = p_0 \;\; (\text{“good quality”}) \;\; \text{against} \\ &H_1 : \mathbf{P}(X_i^2 + Y_i^2 > r^2) = p_1 \;\; (\text{“bad quality”}), \end{aligned} \tag{1.1}$$

where $0 < p_0 < p_1 < 1$ are two pre-specified constants. We should emphasize that while the hypotheses H_0 and H_1 in (1.1) appear to be simple hypotheses, the true underlying distributions for the raw data (X_i, Y_i)'s are unknown and unspecified. In fact, one generally is not interested in finding the true distributions of (X_i, Y_i)'s in ballistic applications, since different manufacturing producers might build different systems that lead to different distributions for raw data (X_i, Y_i)'s. Moreover, a deep understanding of the problem in (1.1) would allow one to tackle the problem of testing composite hypotheses H_0: $p \leq p_0$ against $H_1 : p \geq p_0$ by using the monotone likelihood ratio properties, which will also be illustrated in our case study.

In this article, we tackle the minimax nonparametric framework of (1.1) where we do not make any parametric assumptions on the true underlying distributions for the observed data (X_i, Y_i)'s. Our contributions are twofold:

- From a theoretical standpoint, we prove a surprising result that the Bernoulli sequential probability ratio test (SPRT) is optimal in the sense of minimizing the maximum expected sample sizes subject to Type I and Type II error probabilities constraints, thereby providing fundamental detection limits. Indeed, one allegation often made in the CNP literature is that the Bernoulli SPRT is robust but might lose too much information when quantizing raw bivariate data into binary data, and it has been an open problem whether one can develop a test that is as robust as the Bernoulli SPRT but would have much smaller expected sample sizes, see Li and Mei [16]. Here our theoretical result provides a negative answer. It implies that one needs to make certain assumptions on the distribution of raw bivariate data if one wants to find a test that has a smaller sample size than the Bernoulli SPRT. In other words, one needs to balance the tradeoff between robustness (in term of underlying distributions of raw bivariate data) and efficiency (in term of expected sample sizes).

- From a practical viewpoint, we propose algorithms to suitably design and implement Bernoulli SPRTs that are easy to understand for practitioners in real world applications. There are a couple of challenges to take the full advantage of the optimal properties of the Bernoulli SPRT. First, due to the discrete nature of Bernoulli random variables, it is possible that no Bernoulli SPRTs can exactly attain pre-specified Type I and Type II error probability bounds. While the Bernoulli SPRTs would still be asymptotically optimal for small error probability bounds, it is unclear what are the optimal tests under the finite sample setting. Second, while one can use the Monte Carlo simulation or importance sampling to numerically find the Type I and Type II error probabilities of Bernoulli SPRTs for given boundary thresholds, it is generally computationally expensive to search boundary thresholds to attain the desire Type I and Type II error probabilities. This is because the error probabilities can be piece-wise discontinuous functions of boundary thresholds due to the discrete nature of Bernoulli random variables, and thus the bi-section search or similar would be inefficient. In the context of sequentially hypothesis testing in (1.1), we overcome these challenges by slightly adjusting (α, β) or (p_0, p_1) in a suitable way so that the corresponding Bernoulli SPRTs have explicit formulas on their finite-sample properties and also easy to understand and implement for practictioners.

The remainder of the paper is organized as follows. In Section 2, we state the minimax nonparametric formulation of sequential hypothesis testing problems in the CNP context. We present the Bernoulli SPRT in Section 3, and establish its optimality properties in Section 4, thereby providing fundamental lower bounds on the maximum expected sample sizes subject to the Type I and Type II error probabilities. Section 5 discusses how to design and implement Bernoulli SPRTs that are intuitively appealing to practitioners under the finite-sampling setting. A case study is discussed in Section 6 under two scenarios: one is non-symmetric design with standard error probability constraints, and the other is symmetric design with sample size considerations. Some concluding remarks are provided in Section 7. All technical proofs are included in Section 8.

2. Problem Formulation

In this section, we present a minimax nonparametric formulation of sequential hypothesis testing in the CNP contexts. Assume one observes iid bivariate data (X_i, Y_i)'s sequentially, i.e., one at a time, and we want to utilize as few observations as possible to test the hypotheses in (1.1). This can be naturally formulated as a sequential hypothesis testing problem (Wald [31]). A sequential test design consists of a stopping time N that indicates how many samples one should take, and a final decision $D \in \{0, 1\}$ that accepts the null hypothesis ($D = 0$) or accept the alternative hypothesis ($D = 1$). The stopping time N is an integer-valued random variable: $\{N = n\}$ implies that it takes n observations to make a final decision, and the decision upon $\{N = n\}$ depends only on the first n observations, i.e., we cannot use future observations to make current decisions.

Denote by $f_0^*(x, y)$ and $f_1^*(x, y)$ the unknown true probability density or mass functions of (X_i, Y_i)'s under H_0 and H_1 in (1.1), respectively. Also denote by $\mathbf{P}_{f_j^*}$ and $\mathbf{E}_{f_j^*}$ the probability measure and expectation of the bivariate data (X, Y)'s under hypothesis H_j for $j = 0, 1$. The hypotheses in (1.1) imply that $f_0^*(x, y)$ and $f_1^*(x, y)$ must satisfy

$$\mathbf{P}_{f_0^*}(X^2 + Y^2 > r^2) = p_0 \qquad \text{and} \qquad \mathbf{P}_{f_1^*}(X^2 + Y^2 > r^2) = p_1, \tag{2.2}$$

for a pre-specified $r > 0$. Denote by Ω_j the set of all possible probability density or mass functions, $f_j^*(x, y)$, that satisfies (2.2) with $j = 0$ or 1. Note that the set Ω_0 and Ω_1 will depend only on p_0 and p_1, respectively, unless we make additional model assumptions.

We follow the classical sequential hypothesis testing literature to consider the following mathematical problem:

Minimax Nonparametric Formulation: Given the sets Ω_0 and Ω_1 in (2.2), find a sequential test (T, D) that minimizes

$$\sup_{f_0^* \in \Omega_0} \mathbf{E}_{f_0^*}(T) \qquad \text{and} \qquad \sup_{f_1^* \in \Omega_1} \mathbf{E}_{f_1^*}(T), \tag{2.3}$$

subject to the constraints on Type I and Type II error probabilities:

$$\sup_{f_0^* \in \Omega_0} \mathbf{P}_{f_0^*}(D = 1) \leq \alpha \quad \text{and} \quad \sup_{f_1^* \in \Omega_1} \mathbf{P}_{f_1^*}(D = 0) \leq \beta, \tag{2.4}$$

where $\alpha, \beta \in (0, 1)$ are pre-specified constants.

The challenge of this minimax nonparametric formulation of sequential hypothesis testing problem is due to the nonparametric nature of the probability density or mass function sets, Ω_0 and Ω_1, as the true probability functions, $f_0^*(x, y)$ and $f_1^*(x, y)$, are unknown or unspecified. Have they been completely specified, the set Ω_0 and Ω_1 would contain only a single distribution, and then the problem would have been become the well-known classical sequential hypothesis testing problem of testing a simple null f_0^* against a simple alternative f_1^*. In addition, our minimax nonparametric formulation appears to be similar to the problem of sequentially testing a composite null hypothesis against a composite alternative hypothesis, the main difference is over the functional space in Ω_0 and Ω_1, which looks complicated but turns out to have an explicit solution due to the constraints in (2.2). To best of our knowledge, our minimax nonparametric formulation in (2.2)-(2.4) is new in the rich literature of sequential hypothesis testing.

3. The Bernoulli SPRT

In this section, we present the definition of the Bernoulli SPRT from the two different viewpoints: one is on the quantization of raw bivariate data, and the other is the generalized likelihood ratio (GLR) or maximum likelihood estimation (MLE) principle over the functional space. The former is intuitively appealing but often gives one a mis-impression that binary quantization of raw bivariate data would lead to significant information loss, and thus one might work on the wrong direction how to reduce such information loss. On the other hand, the GLR/MLE viewpoint is more sophisticated, but is useful to understand that there is no information loss from the statistical point of view. In addition, it also sheds light to prove the optimality of the Bernoulli SPRT in next section.

Let us begin with the quantization interpretation of the Bernoulli SPRT. Quantization techniques for sequential hypothesis testing problems were pioneered by Veeravalli, Basar and Poor [30] and have tremendous growths since then, see, for example, Veeravalli [29], Tartakovsky and Veeravalli [28], Mei [19], Nguyen, Wainwright and Jordan [20], Hadjiliadis, Zhang and Poor [11], Fellouris and Moustakides [7]. The main idea is to map raw bivariate data (X_i, Y_i)'s into binary data

$$Z_i = I\{X_i^2 + Y_i^2 > r^2\} = \begin{cases} 0, & \text{if } X_i^2 + Y_i^2 \leq r^2 \\ 1, & \text{if } X_i^2 + Y_i^2 > r^2 \end{cases}, \tag{3.5}$$

and then make statistical decisions based on quantized binary data Z_i's, instead of raw bivariate data (X_i, Y_i)'s. Under our settings, the minimax nonparametric formulation in (2.2)-(2.4) reduces to the problem of utilizing the iid Bernoulli data Z_i's to test the simple null hypothesis $H_0' : \mathbf{P}(Z_i = 1) = 1 - \mathbf{P}(Z_i = 0) = p_0$ against the simple alternative hypothesis $H_1' : \mathbf{P}(Z_i = 1) = 1 - \mathbf{P}(Z_i = 0) = p_1$. In this case, it is well-known that the optimal solution for this new simplified problem is the Bernoulli SPRT (Wald [31]). To be more specific, at time n, the likelihood ratio statistic of Z_i's is given by

$$L_n = \prod_{i=1}^{n} \left(\frac{1 - p_1}{1 - p_0}\right)^{1 - Z_i} \left(\frac{p_1}{p_0}\right)^{Z_i}. \tag{3.6}$$

For two constants $A > 1$ and $B > 1$, the Bernoulli SPRT is defined by the stopping time

$$N_{Bern}(A, B) = \inf\Big\{n \geq 1 : L_n \notin (B^{-1}, A)\Big\}, \tag{3.7}$$

and $N_{Bern}(A, B) = \infty$ if no such n exists. When stopping taking observations at time $N_{Bern}(A, B)$, the Bernoulli SPRT makes a final decision $D = 0$ that accepts the null hypothesis $H_0 : p = p_0$ if the lower bound of B^{-1} is crossed, and makes a final decision $D = 1$ that rejects the null hypothesis H_0 if the upper bound of A is crossed.

Next, let us provide a new interpretation of the Bernoulli SPRT from the GLR or MLE viewpoint. This sheds new light to understand our minimax nonparametric problem and provide ideas to prove the optimality of the Bernoulli SPRT in the next section. Under our minimax nonparametric formulation, at time n, the GLR statistic of (X_i, Y_i)'s is given by

$$G_n = \frac{\sup_{f_1 \in \Omega_1} \prod_{i=1}^{n} f_1(X_i, Y_i)}{\sup_{f_0 \in \Omega_0} \prod_{i=1}^{n} f_0(X_i, Y_i)}, \tag{3.8}$$

where the probability distribution sets Ω_0 and Ω_1 are defined in (2.2).

Our main results in this section can be summarized as follows:

Theorem 1. *The GLR statistic G_n in (3.8) satisfies*

$$G_n = L_n, \tag{3.9}$$

where L_n is the likelihood ratio statistic in (3.6). Thus the Bernoulli SPRT in (3.7) is actually a sequential GLR test under our minimax nonparametric formulation.

Since the Bernoulli SPRT is the GLR test under our minimax nonparametric formulation, it is not surprising that it does not lose much information, even if it is based on the quantized binary data Z_i's instead of raw bivariate data. This is the price we need to pay for the robustness properties with respect to the nonparametric nature of the probability density or mass function sets, Ω_0 and Ω_1. In other words, if we want to find a new test that improves the efficiency of the Bernoulli SPRT, we have to sacrifice the robustness properties, say, by making certain assumptions on Ω_0 and Ω_1.

4. Optimality of Bernoulli SPRT

In this section, we show that the Bernoulli SPRT is optimal under our minimax nonparametric formulation of the sequential hypothesis testing problem in (2.2)-(2.4). To the best of our knowledge, such optimality properties are new in the literature, and establish the best possible information bounds under non-parametric models.

The following theorem establishes the optimality properties of the Bernoulli SPRT under our minimax nonparametric formulation in (2.2)-(2.4). The result is identical to the optimality of the SPRT in the simplest parametric model of testing a simple null hypothesis against a simple alternative hypothesis, except that we are dealing with $\sup_{f_0^* \in \Omega_0}$ and $\sup_{f_1^* \in \Omega_1}$.

Theorem 2. *For a given Bernoulli SPRT $N_{Bern} = N_{Bern}(A, B)$ in (3.7), denote its Type I and Type II error probabilities by α^* and β^* respectively.*

Then for any sequential test (T, D_T), we have

$$\begin{aligned}\sup_{f_0^*\in\Omega_0} \mathbf{E}_{f_0^*}(T) &\geq \sup_{f_0^*\in\Omega_0} \mathbf{E}_{f_0^*}(N_{Bern}) \qquad \textit{and} \\ \sup_{f_1^*\in\Omega_1} \mathbf{E}_{f_1^*}(T) &\geq \sup_{f_1^*\in\Omega_1} \mathbf{E}_{f_1^*}(N_{Bern}),\end{aligned} \tag{4.10}$$

as long as

$$\sup_{f_0^*\in\Omega_0} \mathbf{P}_{f_0^*}(D_T = 1) \leq \alpha^* \quad \textit{and} \quad \sup_{f_1^*\in\Omega_1} \mathbf{P}_{f_1^*}(D_T = 0) \leq \beta^*. \tag{4.11}$$

By Theorem 2, the following corollary derives non-asymptotic lower bounds on the expected sample sizes under our minimax nonparametric formulation in (2.2)-(2.4).

Corollary 1. *Assume that the Type I and Type II error probabilities bounds α and β are attained exactly by the Bernoulli SPRT. Then for any sequential test (T, D_T) satisfying Type I and Type II error constraints in (2.4), we have*

$$\begin{aligned}\sup_{f_0^*\in\Omega_0} \mathbf{E}_{f_0^*}(T) &\geq \frac{1}{KL(p_0,p_1)}\{\alpha \log\frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta}\} \qquad \textit{and} \\ \sup_{f_1^*\in\Omega_1} \mathbf{E}_{f_1^*}(T) &\geq \frac{1}{KL(p_1,p_0)}\{(1-\beta)\log\frac{1-\beta}{\alpha} + \beta\log\frac{\beta}{1-\alpha}\}\end{aligned} \tag{4.12}$$

where

$$\begin{aligned}KL(p_0,p_1) &= (1-p_0)\log\frac{1-p_0}{1-p_1} + p_0\log\frac{p_0}{p_1} \qquad \textit{and} \\ KL(p_1,p_0) &= (1-p_1)\log\frac{1-p_1}{1-p_0} + p_1\log\frac{p_1}{p_0}.\end{aligned} \tag{4.13}$$

The proof of this corollary follows directly from the optimality and statistical properties of the Bernoulli SPRT (see Theorem 2.39 of Siegmund [25]), and thus omitted.

To better illustrate our theoretical results, it is informative to add several remarks:

1. In Theorem 2 and Corollary 1, a crucial assumption is that Type I and Type II error probabilities α and β are attainable by some Bernoulli SPRT. In this case, if one wants to find a test that has smaller expected sample sizes than the Bernoulli SPRT, one would need to sacrifice robustness, say, by making certain parametric assumptions on f_0^* and f_1^* in (2.2) to reduce the probability sets Ω_0 and Ω_1.

2. In the CNP contexts, one widely used parametric model is the bivariate normal models, in which the (X_i, Y_i)'s are assumed to be iid with bivariate normal distribution, see, for example, Fraser [8], Gillis [9], Harter [12], Pyati [22] and Shnidman [24]. This yields the Gaussian SPRT based on raw two-dimensional data (X_i, Y_i). Unfortunately, the Gaussian SPRT is not robust in the sense that its actual Type I and Type II error probabilities would become much larger than the bound α and β if the bivariate normal model assumption is violated, e.g., if the true model is the least favorable pdf in (8.33).

3. Based on the asymptotic properties, to satisfy error bounds α and β, one can approximately choose the thresholds of the Bernoulli SPRT as

$$A \approx \frac{1-\beta}{\alpha} \qquad \text{and} \qquad B \approx \frac{1-\alpha}{\beta}, \tag{4.14}$$

see equation (2.11) of Siegmund [25]. Under the finite sampling setting, the actual Type I and Type II error probabilities α^* and β^* could be very different from α and β due to the discrete nature of Bernoulli SPRT.

4. It is likely that Type I and Type II error probabilities bounds α and β cannot be attained exactly by any Bernoulli SPRTs, no matter how we tune the thresholds A and B based on the Monte Carlo simulations. This is not an issue under the asymptotic setting when $\alpha, \beta \to 0$, since the expected sample size are of order $\log \alpha^{-1}$ and $\log \beta^{-1}$ and thus the error bounds of (α, β) and $(\alpha/2, \beta/2)$ yield asymptotically equivalent expected sample size up to a negligible constant $\log(2)$. However, under the finite-sampling setting for error bounds α, β such as $5\% \sim 20\%$, the true Type I and Type II error probabilities of Bernoulli SPRTs could be much smaller than the pre-specified bounds α or β, which yields much larger expected sample size as compared

to other sequential tests such as Guassian SPRT that can attain the pre-specified bounds α or β. This might be the main reason why practitioners feel that the Bernoulli SPRT might lose information.

5. Algorithms for Design and Implementation

In this section, we discuss how to formulate the right sequential hypothesis testing problem in (1.1)-(2.4), and how to conveniently implement the Bernoulli SPRT in (3.6)-(3.7) from the practical viewpoint, so that the practitioners can have a simple but useful statistical procedure to use in real world applications. Note that in the rich literature of sequential hypothesis testing in statistics, theoretical researchers often pre-specify (p_0, p_1) in (1.1) and (α, β) in (2.4), and focus on the choices of the thresholds A and B of Bernoulli SPRT in (3.7).

Here we follow the practical setting in which some of the (α, β) or (p_0, p_1) values can be treated as initial working values and are adjustable, and the focus is how to implement the Bernoulli SPRT (3.7) as simple as possible. In addition, if one wants to find a Bernoulli SPRT that attains the pre-specified significant level α exactly, we illustrate this is doable if one is allowed to adjust the null hypothesis p_0 value a little bit. This might provide a new research direction for CNP and precision quality such as finding exotic radius r_0 that satisfies $\mathbf{P}(X^2 + Y^2 > r_0) = p_0$ for some non-standard null p_0 value. We will illustrate this through a case study in the next section.

For the purpose of practical implementation, there are two key facts about the Bernoulli SPRT in (3.6)-(3.7). The first one is that it is essentially characterized by the three parameters:

$$b = \frac{\log \frac{p_1}{p_0}}{\log \frac{1-p_0}{1-p_1}}, \qquad m_0 = \frac{\log A}{\log \frac{1-p_0}{1-p_1}}, \qquad m_1 = \frac{\log B}{\log \frac{1-p_0}{1-p_1}}. \tag{5.15}$$

To see this, note that it can be rewritten as

$$\begin{aligned} N^*_{Bern}(b, m_0, m_1) &= \inf\Big\{ n \geq 1 : \sum_{i=1}^{n} \Big[(1 - Z_i) - bZ_i \Big] \notin (-m_0, m_1) \Big\} \\ &= \inf\Big\{ n \geq 1 : S_n \notin (0, m_0 + m_1) \Big\}, \end{aligned} \tag{5.16}$$

where the new "score" statistic S_n has the following recursive form:

$$\begin{aligned} S_n &= m_0 + \sum_{i=1}^{n} \Big[1 - Z_i - bZ_i \Big] \\ &= S_{n-1} + 1 - Z_n - bZ_n, \end{aligned} \tag{5.17}$$

with the initial value $S_0 = m_0$.

The Bernoullis SPRT $N^*_{Bern}(b, m_0, m_1)$ in (5.16) can be implemented conveniently for practitioners as follows. Start with an initial integer score $S = m_0$, and add 1 to the score for each conforming observation (X_n, Y_n) with $X_n^2 + Y_n^2 \leq r^2$, and subtracting b for each nonconforming observation with $X_n^2 + Y_n^2 > r^2$. Then we will reject $H_0 : p \le p_0$ if the score falls to zero of less, and accept $H_0 : p \le p_0$ if the score reaches $m_0 + m_1$.

The second key fact is that when the parameters b, m_0, m_1 in (5.15) are integers or rational numbers, the exact formulas for the error probability and expected sample size of $N^*_{Bern}(b, m_0, m_1)$ in (5.16)-(5.17) have been derived in Burman [4] and Walker [33]. In particular, when b, m_0, m_1 are integers and when the Z_i's are iid Bernoulli(p), the power function of the Bernoulli SPRT $N^*_{Bern}(b, m_0, m_1)$ in (5.16) is given by

$$h(p, m_0, m_1) = \mathbf{P}_p(D = 1) = 1 - \frac{G_p(m_0)}{G_p(m_0 + m_1)}, \tag{5.18}$$

where

$$G_p(m) = \sum_{j=0}^{\lfloor \frac{m-1}{b+1} \rfloor} \frac{(-1)^j p^j}{(1-p)^{m-jb}} \binom{m - jb - 1}{j} \tag{5.19}$$
$$= \frac{1}{(1-p)^m} \left\{ 1 - (m - b - 1)p(1-p)^b + \binom{m - 2b - 1}{2} (p(1-p)^b)^2 - \cdots \right\},$$

and $\lfloor \frac{m-1}{b+1} \rfloor$ denotes the integral part of $\frac{m-1}{b+1}$. Since the exact formula becomes a little more complicated when b, m_0, m_1 in (5.15) are rational numbers, below we will focus on the case when b, m_0, m_1 in (5.15) are integers, so as to easily understand our main ideas.

To design and implement the Bernoulli SPRT in (3.6)-(3.7) conveniently from the practical viewpoint, our high-level idea is to choose b, m_0, m_1 in (5.15) to be integers based on the initial working values of (p_0, p_1) in (1.1) and (α, β) in (2.4). These would allow the practitioners to adopt its simpler yet equivalent form $N^*_{Bern}(b, m_0, m_1)$ in (5.16) in real-world applications while having its statistical properties in explicit forms without using computing-intensive Monte Carlo simulations.

Let us assume that we have the initial working values of (p_0, p_1) in (1.1) and (α, β) in (2.4), whereas the alternative hypothesis p_1 value can be changed. Below is the logical flow for finding a new set of $(p_0, p_1, \alpha, \beta)$ such that an easily implementable Bernoulli SPRT is optimal for the corresponding sequential hypothesis testing problem in (1.1)–(2.4), while satisfying the conditions in Theorem 2 and Corollary 1.

1. Compute

$$b^* = \left\lceil \frac{\log(p_1/p_0)}{\log((1-p_0)/(1-p_1))} \right\rceil, \tag{5.20}$$

where $\lceil x \rceil$ denote the smallest integer greater than or equal to x.

2. For the value b^* in (5.20), find $\widetilde{p}_1$ satisfying

$$\log \frac{\widetilde{p}_1}{p_0} = b^* \log \frac{1-p_0}{1-\widetilde{p}_1}. \tag{5.21}$$

3. By (5.15) and (5.21), we compute find the following integer-valued thresholds

$$\begin{aligned} m_0^* &= \left\lceil \frac{\log((1-\beta)/\alpha)}{\log((1-p_0)/(1-\widetilde{p}_1))} \right\rceil \qquad \text{and} \\ m_1^* &= \left\lceil \frac{\log((1-\alpha)/\beta)}{\log((1-p_0)/(1-\widetilde{p}_1))} \right\rceil. \end{aligned} \tag{5.22}$$

Algorithm 1 Formulating Right Problem with Bernoulli SPRT

Input: Initial values of (p_0, p_1) and (α, β).

1. Calculate $b = \left\lceil \frac{\log(p_1/p_0)}{\log\left(\frac{1-p_0}{1-p_1}\right)} \right\rceil$.

2. Update p_1 such that $\log(p_1/p_0) = b \log\left(\frac{1-p_0}{1-p_1}\right)$.

3. Find $m_0 = \left\lceil \frac{\log\left(\frac{1-\beta}{\alpha}\right)}{\log\left(\frac{1-p_0}{1-\widetilde{p}_1}\right)} \right\rceil, \quad m_1 = \left\lceil \frac{\log\left(\frac{1-\alpha}{\beta}\right)}{\log\left(\frac{1-p_0}{1-\widetilde{p}_1}\right)} \right\rceil$.

4. Update (α, β) or (p_0, p_1):

- Update $\alpha = h(p_0, m_0, m_1)$ and $\beta = 1 - h(p_1, m_0, m_1)$, or
- Update p_0 such that $h(p, m_0, m_1) = \alpha$ and p_1 such that $h(p, m_0, m_1) = 1 - \beta$.

Output: Updated values of $(p_0, p_1, \alpha, \beta)$.

4. The Type I error probability of the $N^*_{Bern}(b^*, m^*_0, m^*_1)$ in (5.16) is $\alpha^* = h(p_0, m^*_0, m^*_1)$ in (5.18) and Type II error probability at $\widetilde{p}_1$ is $\beta^* = 1 - h(\widetilde{p}_1, m^*_0, m^*_1)$. Meanwhile, if one want to keep (α, β) and adjust p_0 or $\widetilde{p}_1$, then

$$\begin{aligned} p^*_0 &= p \text{ such as } h(p, m^*_0, m^*_1) = \alpha \qquad \text{and} \\ p^*_1 &= p \text{ such as } h(p, m^*_0, m^*_1) = 1 - \beta. \end{aligned} \tag{5.23}$$

The above discussion can be summarized in Algorithm 1 and the implementation of the Bernoulli SPRT is summarized in Algorithm 2.

Algorithm 2 Bernoulli SPRT $N^*_{Bern}(b, m_0, m_1)$ in (5.16)

Input: b, m_0, m_1, r

Initialize: $S = m_0$

for each n with new data (X, Y) **do**

$Z = I(X^2 + Y^2 > r^2)$

$S \leftarrow S + 1 - 2Z$

if $S \leq 0$ **then return** $D = 0$ (accept H_0)

else if $S \geq m_0 + m_1$ **then return** $D = 1$ (reject H_0)

end if

end for

6. Case Study

In this section, we analyze a real data set in Li et al. [17] to demonstrate the usefulness of our theoretical results and practical algorithm in the previous sections. The dataset includes the recorded positions of the sampled bullets' falling points, and the first 6 values were: $(-0.82, -0.51)$, $(0.59, -0.06)$, $(-0.74, 0.39)$, $(0.24, 0.24)$, $(-0.33, 0.93)$, $(0.17, 0.88)$. In particular, the corresponding $\sqrt{X_i^2 + Y_i^2}$ values of these six points are $0.97, 0.59, 0.84, 0.34, 0.99$ and 0.90. Figure 1 plots these observations in the (X, Y) plane along with two radius: $r = 1$ and $r = 0.70$.

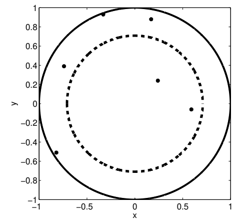

Figure 1: The six black dots represent the positions of the bullets' falling points. The solid and dotted circles correspond to the radius of $r = 1$ and $r = 0.70$, respectively.

Below we will investigate this dataset in two subsections: one is non-symmetric setting under the standard error probability constraints, and the other is the symmetric setting under the sample size considerations. Hopefully these allow us to illustrate how to formulate the right sequential hypothesis testing problem in (1.1)-(2.4) in the real-world applications that can take full advantage of the optimality properties of the Bernoulli SPRT.

6.1 Asymmetric Setting

Suppose that one practitioner focuses on the radius $r = 1$ and has a rough initial interest of testing hypotheses

$$H_0 : \mathbf{P}(X_i^2 + Y_i^2 > 1) = 0.1 \text{ against } H_1 : \mathbf{P}(X_i^2 + Y_i^2 > 1) = 0.4, \quad (6.24)$$

subject to the error probability constraints of $\alpha = 5\%$ and $\beta = 20\%$.

In this case, we have $p_0 = 0.1, p_1 = 0.4$, $\alpha = 0.05, \beta = 0.20$. Using to algorithm 1 to compute those values in (5.20)-(5.23), we have $b^* = \lceil 3.419 \rceil = 4, \widetilde{p}_1 \approx 0.3345, m_0 = \lceil 9.185 \rceil = 10$ and $m_1 = \lceil 5.162 \rceil = 6$. Now Given $b = 4$, the function in (5.19) becomes

$$\begin{aligned} G_p(m_0 = 10) &= \frac{1}{(1-p)^{10}}(1 - 5p(1-p)^4) \\ G_p(m_0 + m_1 = 16) &= \frac{1}{(1-p)^{16}}(1 - 11p(1-p)^4 + 21p^2(1-p)^8 - p^3(1-p)^{12}). \end{aligned}$$

and thus the power function in (5.18) becomes

$$h(p, m_0 = 10, m_1 = 6) \;=\; 1 - \frac{(1-p)^6(1 - 5p(1-p)^4)}{1 - 11p(1-p)^4 + 21p^2(1-p)^8 - p^3(1-p)^{12}}.$$

There are two different ways to use this power function. First, assume that we plan to keep $p_0 = 0.1$ and $\widetilde{p}_1 \approx 0.3345$, and want to adjust α and β so that the Bernoulli SPRT $N^*_{Bern}(b = 4, m_0 = 10, m_1 = 6)$ in (5.16) is the optimal. Plugging $p = 0.1$ and $p = 0.3345$ into the power function, we have $\alpha^* = h(p_0 = 0.1, m_0 = 10, m_1 = 6) = 0.0307$ and $\beta^* = 1 - h(\widetilde{p}_1 \approx 0.3345, m_0 = 10, m_1 = 6) = 0.1585$. This yield a new problem of testing

$$\begin{aligned} &H_0 : \mathbf{P}(X_i^2 + Y_i^2 > 1) = 0.1 \;\text{ against } H_1 : \mathbf{P}(X_i^2 + Y_i^2 > 1) = \widetilde{p}_1 \approx 0.3345, \\ &\text{subject to the constraints of } \alpha = 0.0307 \text{ and } \beta = 0.1585. \qquad (6.25) \end{aligned}$$

Second, if we want to keep $\alpha = 5\%$ and $\beta = 20\%$ and are willing to adjust p_0 or $\widetilde{p}_1$. By (5.23), we have $p_0^* = 0.1137$ and $p_1^* = 0.3150$. This leads to another new problem of testing

$$H_0 : \mathbf{P}(X_i^2 + Y_i^2 > 1) = 0.1137 \text{ against } H_1 : \mathbf{P}(X_i^2 + Y_i^2 > 1) = 0.3150,$$
$$\text{subject to the constraints of } \alpha = 0.05 \text{ and } \beta = 0.20. \tag{6.26}$$

What we want to declare is that the problem formulation in either (6.25) and (6.26) would be better than that in (6.24) for the practical purpose. The reason is that the Bernoulli SPRT $N^*_{Bern}(b = 4, m_0 = 10, m_1 = 6)$ in (5.16) is optimal under either (6.25) and (6.26) by Theorem 2 and is also easy to implement. Meanwhile, under the formula in (6.24), the Bernoulli SPRT is very difficult to implement in practical due to irrational value of b. Moreover, while asymptotically optimal as $\alpha, \beta \to 0$, it is unclear whether the Bernoulli SPRT is optimal under the finite-sample setting or not.

Let us now apply the Bernoullis SPRT $N^*_{Bern}(b = 4, m_0 = 10, m_1 = 6)$ in (5.16) to analyze those 6 data points in Figure 1. Since all six data points are inside the circle with radius of 1, the Bernoullis SPRT $N^*_{Bern}(b = 4, m_0 = 10, m_1 = 6)$ in (5.16) will stop at the 6th observation, and accept the null hypothesis H_0 in the problem of either (6.25) and (6.26).

6.2 Symmetric Setting

Suppose that another practitioner feels that $r = 0.70$ is the circular error probable in the sense of $\mathbf{P}(X^2 + Y^2 \geq 0.70^2) = 0.50$, and thus want to test composite hypotheses

$$H_0 : \mathbf{P}(X_i^2 + Y_i^2 > 0.70^2) \leq 0.50 \ \text{ against} \qquad (6.27)$$
$$H_1 : \mathbf{P}(X_i^2 + Y_i^2 > 0.70^2) > 0.50.$$

The practitioner might not be sure how to choose the error probabilities α and β here, but would have a rough idea of utilizing 6 or less observations to make a decision due to the budget or sampling constraints.

Let us illustrate how to apply our results to tackle this problem. As mentioned in Section II.3 in Siegmund [25], when testing composite hypotheses in (6.27), we can adopt the symmetric Bernoullis SPRT in the problem of

$$H_0 : \mathbf{P}(X_i^2 + Y_i^2 > 0.70^2) = p_0 \ \text{ against} \qquad (6.28)$$
$$H_1 : \mathbf{P}(X_i^2 + Y_i^2 > 0.70^2) = 1 - p_0,$$

subject to the error probability constraints of $\alpha = \beta$. Alternatively, (6.27) can be treated as the limit of (6.28) as $p_0 \to 0.5$.

In the problem of either (6.27) or (6.28), we have $p_1 = 1 - p_0$ and thus $b = 1$ for the Bernoullis SPRT $N^*_{Bern}(b, m_0, m_1)$ in (5.16). For the symmetric case, we would set $m_0 = m_1 = m$, and the power function in (5.18) with $b = 1$ has a simpler form of

$$h(p, m, m) = \mathbf{P}_p(D = 1) = \frac{p^m}{p^m + (1-p)^m}, \qquad (6.29)$$

see the first equation on page 15 of Siegmund [25]. In addition, for $b = 1$, the upper or lower integer bounds will be attained exactly without overshoot, and thus the expected sample sizes of the symmetric Bernoulli SPRT are the same as the lower bounds in (4.12) in Corollary 1, which becomes equality in this case.

Figure 2 plots error probability and expected sample size functions of the symmetric Bernoulli SPRT when p_0 various over $(0, 0.45)$ for $m = 1, 2, 3, 4$. First, for any non-integer bounds m^*, the Bernoulli SPRT $N^*_{Bern}(b = 1, m_0, m_1)$ in (5.16) with $m_0 = m_1 = m^*$ would be the same as that with $m_0 = m_1 = [m^*]$, the integer part of m^*. This means that the error probability and expected sample size functions are piece-wise constant function over the threshold m^*. Second, the plot shows that both functions are continuous, increasing functions of p_0. This is understandable, since the problem in (6.29) would become more difficult when p_0 increases from 0 to 0.5. Third, since this is symmetric design, it is not surprising to see that $\mathbf{P}_{p=0.5}(D = 1) = h(p = 0.5, m, m) = 0.5$ for any integer bound $m \geq 1$. Hence, the symmetric Bernoulli SPRT $N^*_{Bern}(b = 1, m, m)$ in (5.16) is useful for testing composite hypotheses in (6.27), the remaining question is how to choose m subject to the desired sampling size constraints. This can be answered through analyzing how to specify p_0 and $\alpha = \beta$ suitably when

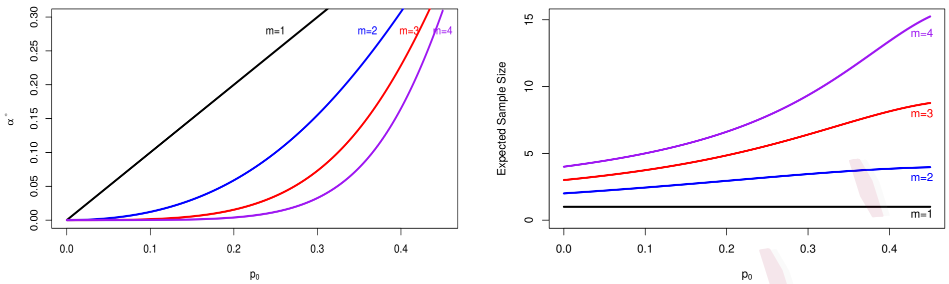

Figure 2: Performance of Symmetric Bernoulli SPRT $N^*_{Bern}(b = 1, m, m)$ in (5.16) with various p_0 and $m = 1, 2, 3, 4$. Left: error probabilities as a function of p_0. Right: expected sample size as a function of p_0.

testing hypotheses in (6.28).

As an illustration, assume that we want to set the error probability constraints in (2.3) as $\alpha = \beta = 15\%$. These relatively large α and β values are standard in ballistic or other applications as one wants to reach reasonable decision quickly to reduce the cost of experiments, also see Li et al. [17]. By Figure 2, in order to achieve $\alpha = \beta = 0.15$ for $m = 1, 2, 3, 4$, the corresponding critical p_0 values are $0.1500, 0.2959, 0.3594$ and 0.3933, respectively. The respective expected sample sizes would be $1.00, 3.42, 7.47$ and 13.12.

These information are useful to choose suitable design parameters in the experiments. For instance, if one would like to reach decisions based on at most 6 observations, then one might want to choose $m = 2$ and

$p_0 = 0.2959$. That is, with such a small sample size consideration, one should aim to test the problem in (6.28) with $p_0 = 0.2959$ subject to Type I and Type II error probabilities constraints of $\alpha = \beta = 0.15$. Moreover, in such a problem, the optimal test will be the symmetric Bernoulli SPRT $N^*_{Bern}(b = 1, m = 2, m = 2)$ in (5.16).

Let us go back the analysis of the original dataset of six observation in the context of testing the problem in (6.27) with $r = 0.70$ (instead of $r = 1$ in the previous subsection). In this new context, for these six data points, the observed binary data Z_i would be $1, 0, 1, 0, 1, 1$. Hence, the symmetric Bernoulli SPRT $N^*_{Bern}(b = 1, m = 2, m = 2)$ in (5.16) stops at time $n = 6$ and decides that there is enough evidence to reject $p = 0.2959$ at the significant level $\alpha = 15\%$ and declare the poor precision quality. Here the radius $r_a = 0.70$ can be replaced by any other values of $0.60 \leq r_a \leq 0.83$, and we would reach the same conclusion, as long as 4 out of 6 observations are outside of the radius.

It is interesting to note that for the same dataset, our conclusions in this subsection are different from those in the previous subsection. However, there are no contradictions here, as one is to accept the hypothesis $\mathbf{P}(X^2 + Y^2 > 1) \leq 0.1$ and the other is to reject the hypothesis $\mathbf{P}(X^2 + Y^2 > 0.70^2) \leq 0.5$. In practice, one might want to specify several hypothesis testing problems in (1.1) with respect to various radius r's. It would give us a complete picture of precision quality of the system, and it will be interesting to see how to fuse these decisions together to a single global decision.

7. Concluding Remarks

In this work, motivated by the application of evaluating a system's precision quality in the circular nonconforming probability (CNP) context, we have investigated the sequential hypothesis testing problem on the chance of the system missing a pre-specified disk target. Mathematically, it has been formulated as utilizing the observed bivariate data (X_i, Y_i)'s to sequentially test simple hypotheses on $p = \mathbf{P}(X_i^2 + Y_i^2 > r^2)$ when the true underlying distributions for the observed raw bivariate data (X_i, Y_i)'s are unknown. We derived several interesting results. Under the minimax nonparametric formulation, we showed that the intuitively appealing Bernoulli SPRT is actually the generalized likelihood ratio test and also is optimal. We also develop useful algorithms for the appropriate design and implementation of the Bernoulli SPRT so as to maintain both robustness and efficiency.

Several future directions can be pursued. First, instead of fully non-parametric approach, it will be interesting to develop semi-parametric or empirical Bayesian approaches that can utilize raw data better while conducting hypothesis testing. Second, in some applications, one might want to test multiple hypothesis testing problems simultaneously for various radius r, so as to better ensure the precision quality of the system. While we can combine multiple Bernoulli SPRTs into a single test, it will useful to investigate whether there are better approaches or not. Third, our hypothesis testing problem inspires new research direction to find suitable point estimation of the radius corresponding to exotic values of CNP. Hopefully our work can stimulate more non-asymptotic research on sequential hypothesis testing when the sample sizes are not so large.

8. Technical Proofs

This section includes the technical proofs of Theorem 1 and Theorem 2.

8.1 Proof of Theorem 1

Before we prove Theorem 1, let us investigate the following optimization problem for the functional MLE under the discrete case, which is involved in both numerator and denominator in (3.8).

Functional MLE Problem: Given $p \in (0, 1)$ and the n observed data vector (x_i, y_i)'s, find the real-valued probability mass function $f(u_i, v_i) =$

$\mathbf{P}(x = u_i, Y = v_i)$ for $i = 1, 2, \cdots$ that maximizes the likelihood function

$$L_p(f) = \prod_{i=1}^{n} f(x_i, y_i) \tag{8.30}$$

subject to the constraints

$$f(u_i, v_i) \geq 0, \quad \sum_{i=1}^{\infty} f(u_i, v_i) = 1, \text{ and } \sum_{i=1}^{\infty} I(u_i^2 + v_i^2 > r^2) f(u_i, v_i) = p. \tag{8.31}$$

The following lemma, whose proof will be presented in the supplementary material due to page limit, solves this functional MLE problem:

Lemma 1. *For a given $p \in (0, 1)$ and assume $m = \sum_{i=1}^{n} I(x_i^2 + y_i^2 > r^2) \in [0, n]$. For the functional MLE problem in (8.30)-(8.31), the maximum value of likelihood function is given by*

$$L_p^* = \sup_f L_p(f) = \sup_f \prod_{i=1}^{n} f(x_i, y_i) = (\frac{1-p}{n-m})^{n-m} (\frac{p}{m})^m, \tag{8.32}$$

where we adopt the classical notation $(\frac{b}{a})^a = 1$ whenever $a = 0$.

With Lemma 1, we are now ready to prove Theorem 1. Under the discrete case, the GLR statistic G_n in (3.8),

$$G_n = \frac{L_{p_1}^*}{L_{p_0}^*} = (\frac{1-p_1}{n-m})^{n-m} (\frac{p_1}{m})^m (\frac{n-m}{1-p_0})^{n-m} (\frac{m}{p_0})^m = (\frac{1-p_1}{1-p_0})^{n-m} (\frac{p_1}{p_0})^m,$$

which is the same as L_n in (3.6), since $m = \sum_{i=1}^{n} I(x_i^2 + y_i^2 > r^2) = \sum_{i=1}^{n} Z_i$ by the definition of Z_i in (3.5). The proof of Theorem 1 under continuous case is similar and will be presented in the supplementary material.

8.2 Proof of Theorem 2

To prove Theorem 2, the key idea is to show that the Bernoulli SPRT N_{Bern} is the limit of a sequence of SPRTs for testing a pair of least favorable pdfs. At high-level, Lemma 1 inspires us that least favorable pdfs consist of two uniform distributions: one is inside the unit disk, i.e., $X^2 + Y^2 \le r^2$, and the other is outside of the unit disk, i.e., $X^2+Y^2 > r^2$. Unfortunately, there is no uniform distribution over the region of $X^2 + Y^2 > r^2$. The good news is that we can approximate it by a sequence of uniform distributions over the region of $r^2 < X^2 + Y^2 \le M^2$ as $M \to \infty$.

To be more specific, for a given $p \in (0, 1)$ and a given $M > r$, consider the follwoing pdf of (X, Y) that is a mixture of three components:

$$f_{p,M}(x, y) = \begin{cases} \frac{1-p}{\pi r^2}, & \text{if } 0 \le x^2 + y^2 \le r^2; \\ \frac{p}{\pi(M^2-r^2)}, & \text{if } r^2 < x^2 + y^2 \le M^2; \\ 0, & \text{if } x^2 + y^2 > M^2. \end{cases} \quad (8.33)$$

Note that $f_{p,M}$ in (8.33) is essentially a combination of two uniform distributions: one is on $X^2 + Y^2 \le r^2$ and the other is on $r < X^2 + Y^2 \le M^2$. Clearly, $f_{p_0,M}(x, y)$ and $f_{p_1,M}(x, y)$ are well-defined pdfs of the raw data (X_i, Y_i)'s that satisfy (2.2) under the hypotheses H_0 and H_1 in (1.1). Hence, $f_{p_0,M}(x, y) \in \Omega_0$ and $f_{p_1,M}(x, y) \in \Omega_1$.

Next, for a given $M > r$, let us focus on the new parametric model

(8.33) for the raw data (X_i, Y_i)'s, and consider the problem of testing the simple null $f_{p_0,M}(x,y)$ against the simple alternative $f_{p_1,M}(x,y)$. At time n, the corresponding log-likelihood ratio statistic of (X_i, Y_i)'s is given by

$$\begin{aligned} L_{M,n} &= \prod_{i=1}^{n} \frac{f_{p_1,M}(X_i, Y_i)}{f_{p_0,M}(X_i, Y_i)} \\ &= \prod_{i=1}^{n} \Big(\frac{1-p_1}{1-p_0}\Big)^{I(X_i^2+Y_i^2 \le r^2)} \Big(\frac{p_1}{p_0}\Big)^{I(r^2 < X_i^2+Y_i^2 \le M^2)}, \end{aligned} \tag{8.34}$$

and define the corresponding SPRT as

$$N_M^* = \inf\Big\{n \ge 1 : L_{M,n} \notin (B^{-1}, A)\Big\}, \tag{8.35}$$

where the thresholds A and B are the same as those of the Bernoulli SPRT N_{Bern} in (3.7).

The key fact is that whenever $X^2 + Y^2 \le M^2$, the likelihood ratio statistic $L_{M,n}$ in (8.34) for the SPRT N_M^* is identical to the likelihood ratio statistic L_n in (3.6) for the Bernoulli SPRT N_{Bern}. This fact has two important implications. On the one hand, the SPRT N_M^* and the Bernoulli SPRT N_{Bern} has the same statistical properties under the hypothesis $f_{p_0,M}(x,y)$ or $f_{p_1,M}(x,y)$ in (8.33), since the data (X_i, Y_i) cannot be outside of radius M under these specified parameteric models. In particular, for the SPRT N_M^*, its Type I error probability is α^* under $f_{p_0,M}(x,y)$ and its Type II error is β^* under $f_{p_1,M}(x,y)$. On the other hand, a comparison between (3.6) and (8.34) shows that $L_{M,n} \to L_n$ as $M \to \infty$, and thus loosely speaking, the

Bernoulli SPRT N_{Bern} can be viewed as an asymptotic counterpart of the SPRTs N_M^* in (8.35) under any arbitrary distribution f^*. Below we will provide a rigorous proof on the optimality of the Bernoulli SPRT N_{Bern}.

To be more specific, let us focus on the first relation of (4.10) under an arbitrary $f_0^* \in \Omega_0$. For any sequential test with stopping time T satisfying the error probability constraints in (4.11), there are three steps when comparing $\mathbf{E}_{f_0^*}(T)$ and $\mathbf{E}_{f_0^*}(N_{Bern})$ through the optimal SPRT N_M^* in the problem of testing the simple null $f_{p_0,M}(x,y)$ against a simple alternative $f_{p_1,M}(x,y)$. First, let us compare T and N_M^* under $f_{p_0,M}(x,y)$, and we have

$$\sup_{f_0^* \in \Omega_0} \mathbf{E}_{f_0^*}(T) \geq \mathbf{E}_{f_{p_0,M}}(T) \geq \mathbf{E}_{f_{p_0,M}}(N_M^*), \tag{8.36}$$

where the result follows directly from the optimality of the SPRT N_M^*. Second, we will compare N_M^* with N_{Bern} under the specific parametric distribution $f_{p_0,M}(x,y)$ in (8.33). In this case, all data satisfy $X^2+Y^2 \leq M^2$, the optimal SPRT N_M^* is identical to the Bernoulli SPRT N_{Bern} under $f_{p_0,M}(x,y)$, and thus

$$\mathbf{E}_{f_{p_0,M}}(N_M^*) = \mathbf{E}_{f_{p_0,M}}(N_{Bern}). \tag{8.37}$$

Third, we will compare the performances of N_{Bern} under either $f_{p_0,M}(x,y)$ or under an arbitrary $f_0^* \in \Omega_0$. Note that the Bernoulli SPRT N_{Bern} is an equalizer procedure in the sense that $\mathbf{E}_{f_0^*}(N_{Bern})$ are constants over all possible pdf $f_0^* \in \Omega_0$ satisfying (2.2), and thus we have

$$\mathbf{E}_{f_{p_0,M}}(N_{Bern}) = \sup_{f_0^* \in \Omega_0} \mathbf{E}_{f_0^*}(N_{Bern}). \tag{8.38}$$

Combining (8.36), (8.37) and (8.38), we conclude that the Bernoulli SPRT N_{Bern} minimizes $\sup_{f_0^*} \mathbf{E}_{f_0^*}(T)$. Similarly, it also minimizes $\sup_{f_1^*} \mathbf{E}_{f_1^*}(T)$. This completes the proof of Theorem 2.

Supplementary Materials

This supplementary material provides detailed proof of Lemma 1 as well as the proof of Theorem 1 under continuous case.

Acknowledgements

The authors would like to thank the reviewers, Associate Editor, and co-Editors for their invaluable comments and encouragements. Y. Mei is partially supported by an NSF-DMS grant 2515158.

References

[1] Michael Baron and Rui Xu. Sequential testing for full credibility. *Variance*, 10(2):227–239, 2016.

[2] Jay Bartroff and Jinlin Song. Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates. *Sequential Analysis*, 39(1):65–91, 2021.

[3] PK Bhattacharya and Dargan Frierson Jr. A nonparametric control chart for detecting small disorders. *The Annals of Statistics*, pages 544–554, 1981.

[4] J. P. Burman. Sequential sampling formulae for a Binomial population. *Supplement to the Journal of the Royal Statistical Society*, 8(1):98–103, 1946.

[5] Anamitra Chaudhuri and Georgios Fellouris. Joint sequential detection and isolation for dependent data streams. *The Annals of Statistics*, 52(5):1899–1926, 2024.

[6] Shyamal K De and Michael Baron. Sequential tests controlling generalized familywise error rates. *Statistical Methodology*, 23:88–102, 2015.

[7] Georgios Fellouris and George V Moustakides. Decentralized sequential hypothesis testing using asynchronous communication. *IEEE Transactions on Information Theory*, 57(1):534–548, 2010.

[8] Donald Alexander Stuart Fraser. Generalized hit probabilities with a Gaussian target, II. *The Annals of Mathematical Statistics*, pages 288–294, 1953.

[9] J.T. Gillis. Computation of the circular error probability integral. *IEEE Transactions on Aerospace and Electronic Systems*, 27(6):906–910, 1991.

[10] Louis Gordon and Moshe Pollak. An efficient sequential nonparametric scheme for detecting a change of distribution. *The Annals of Statistics*, pages 763–804, 1994.

[11] Olympia Hadjiliadis, Hongzhong Zhang, and H Vincent Poor. One shot schemes for decentralized quickest change detection. *IEEE Transactions on Information Theory*, 55(7):3346–

3359, 2009.

[12] H Leon Harter. Circular error probabilities. *Journal of the American Statistical Association*, 55(292):723–731, 1960.

[13] Xinrui He and Jay Bartroff. Asymptotically optimal sequential FDR and pFDR control with (or without) prior information on the number of signals. *Journal of Statistical Planning and Inference*, 210:87–99, 2021.

[14] Alix Lhéritier and Frédéric Cazals. A sequential non-parametric multivariate two-sample test. *IEEE Transactions on Information Theory*, 64(5):3361–3370, 2018.

[15] Xiaoou Li, Yunxiao Chen, Xi Chen, Jingchen Liu, and Zhiliang Ying. Optimal stopping and worker selection in crowdsourcing: An adaptive sequential probability ratio test framework. *arXiv preprint arXiv:1708.08374*, 2017.

[16] Yan Li and Yajun Mei. Effect of bivariate data's correlation on sequential tests of circular error probability. *Journal of Statistical Planning and Inference*, 171:99–114, 2016.

[17] Yan Li, Xiaolong Pu, and Dongdong Xiang. Mixed variables-attributes test plans for single and double acceptance sampling under exponential distribution. *Mathematical Problems in Engineering*, 2011(1):575036, 2011.

[18] Bowen Liu, Xiaojun Duan, and Liang Yan. A novel Bayesian method for calculating circular error probability with systematic-biased prior information. *Mathematical Problems in Engineering*, 2018(1):5930109, 2018.

[19] Yajun Mei. Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks. *IEEE Transactions on Information Theory*, 54(5):2072–2089, 2008.

[20] XuanLong Nguyen, Martin J Wainwright, and Michael I Jordan. On optimal quantization rules for some problems in sequential decentralized detection. *IEEE Transactions on Information Theory*, 54(7):3285–3295, 2008.

[21] H Vincent Poor and Olympia Hadjiliadis. *Quickest Detection*. Cambridge University Press, 2008.

[22] Vittal P Pyati. Computation of the circular error probability (CEP) integral. *IEEE Transactions on Aerospace and Electronic Systems*, 29(3):1023–1024, 1993.

[23] Pranab Kumar Sen. Nonparametric methods in sequential analysis. *Handbook of sequential analysis*, pages 331–362, 1991.

[24] David A Shnidman. Efficient computation of the circular error probability (CEP) integral. *IEEE Transactions on Automatic Control*, 40(8):1472–1474, 1995.

[25] David Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer Science & Business Media, 2013.

[26] Yanglei Song and Georgios Fellouris. Sequential multiple testing with generalized error control. *The Annals of Statistics*, 47(3):1776–1803, 2019.

[27] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC press, 2014.

[28] Alexander G Tartakovsky and Venugopal V Veeravalli. Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis*, 27(4):441–475, 2008.

[29] Venugopal V Veeravalli. Sequential decision fusion: theory and applications. *Journal of the Franklin Institute*, 336(2):301–322, 1999.

[30] Venugopal V Veeravalli, Tamer Basar, and H Vincent Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Transactions on Information Theory*, 39(2):433–442, 1993.

[31] Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117 – 186, 1945.

[32] Abraham Wald. *Sequential Analysis.* J. Wiley & sons, Incorporated, 1947.

[33] A. M. Walker. Note on sequential sampling formulae for a Binomial population. *Journal of the Royal Statistical Society, Series B (Methodological)*, 12(2):301–307, 1950.

[34] Yiming Xing and Georgios Fellouris. Asymptotically optimal mutistage tests for non-iid data. *Statistica Sinica*, 34:2325–2346, 2024.

Department of Biostatistics, School of Global Public Health, New York University

E-mail: (xuqunzhi66@gmail.com) and (yajun.mei@nyu.edu)