Effect of bivariate data’s correlation on sequential tests of circular error probability

Yan Li\textsuperscript{a}, Yajun Mei\textsuperscript{b,\ast}

\textsuperscript{a} School of Finance and Statistics, East China Normal University, Shanghai 200241, China
\textsuperscript{b} H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

\begin{abstract}

The problem of evaluating a military or GPS/GSM system's precision quality is considered in this article, where one sequentially observes bivariate normal data \((X_i, Y_i)'s\) and wants to test hypotheses on the circular error probability (CEP) or the probability of nonconforming, i.e., the probabilities of the system hitting or missing a pre-specified disk target. In such a problem, we first consider a sequential probability ratio test (SPRT) developed under the erroneous assumption of the correlation coefficient \(\rho = 0\), and investigate its properties when the true \(\rho \neq 0\). It was shown that at least one of the Type I and Type II error probabilities would be larger than the required ones if the true \(\rho \neq 0\), and for the detailed effects, \(\exp(-2) \approx 0.1353\) turns out to be a critical value for the hypothesized probability of nonconforming. Moreover, we propose several sequential tests when the correlation coefficient \(\rho\) is unknown, and among these tests, the method of generalized sequential likelihood ratio test (GSLRT) in Bangdiwala (1982) seems to work well.

\end{abstract}

1. Introduction

Our research is motivated by the problem of evaluating a system's precision quality, which has important applications in the military science of ballistics, GPS (global positioning system) and GSM (global system for mobile communications). In a simple setting of such a problem, one observes two-dimensional vector data \((X_i, Y_i)'s\), and wants to investigate the probabilities of the system (bombs, missiles, bullets, GPS, GSM, etc.) hitting or missing a pre-specified disk target. Mathematically, we want to investigate the circular error probability \(CEP = P(X_i^2 + Y_i^2 \leq a^2)\), or equivalently, the probability of nonconforming, \(p = P(X_i^2 + Y_i^2 > a^2)\), for some pre-specified constant \(a > 0\), and without loss of generality, we assume \(a = 1\) hereafter. More formally, the problem of evaluating a system's precision quality can be formulated as testing the hypotheses

\[ H_0 : p = p_0 \text{ ("good quality") versus } H_1 : p = p_1 \text{ ("poor quality").} \]

where \(p_0 < p_1\) are pre-specified constants in \((0, 1)\), see Li et al. (2011).

In some real-world applications such as ballistics, it is often very expensive to collect data, and thus it is desired to use as few samples as possible to decide whether the precision quality of the system is good or poor. In the statistical literature, this type of problem is known as sequential hypothesis testing. For the usual statistical tests the sample size is fixed before the
data are taken, but for a sequential test the total sample size depends on the data and is thus a random variable. Sequential tests are often economical in the sense that we may reach a decision earlier via a sequential test than via a fixed sample size test on average. The sequential test was first investigated by Wald (1945), and since then it has undergone tremendous growth over the past seventy years and has been widely used in applications such as clinical trials and industrial quality control, see Siegmund (1985), Lai (2001) and Tartakovsky et al. (2014) for reviews. In the standard formulation of sequential hypothesis testing, we want to use as few samples as possible to decide which hypothesis in (1) is true, subject to the classical constraints on Type I and Type II error probabilities:

\[
P_{p=p_0} (\text{Reject } H_0) \leq \alpha \quad \text{and} \quad P_{p=p_1} (\text{Reject } H_1) \leq \beta,
\]

where \(\alpha\) and \(\beta\) are pre-specified bounds (usually \(0 < \alpha, \beta < 0.5\)).

For the sequential hypothesis testing problem in (1) and (2), two existing methods have been proposed in the context of the circular error probability (CEP), see Li et al. (2011). The first one is a naive approach that is based on the Bernoulli data \(Z_i = 1(X_i^2 + Y_i^2 > 1)\), where \(1(\cdot)\) is an indicator function. This is a robust nonparametric approach in the sense that it does not make any assumptions on the underlying distributions of two-dimensional vector data \((X_i, Y_i)\)'s, but unfortunately it is rarely used in practice as it needs larger sample sizes on average to make any decisions. The second method, proposed in Li et al. (2011), is a parametric sequential probability ratio test (SPRT) developed under the assumption that the observations \((X_i, Y_i)\)'s are independent and identically distributed (i.i.d.) according to a specific bivariate normal distribution \(N((0, 0), \sigma^2 I_{2 \times 2})\), where \(I_{2 \times 2}\) is a 2 \times 2 identity matrix. However, in a concrete application for the problem in (1) and (2) with \(p_0 = 0.1\) and \(p_1 = 0.4\), Li et al. (2011) found that their proposed parametric SPRT rejects the null hypothesis \(H_0 : p = 0.1\) after only taking \(n = 6\) observations even though all the 6 observations \((X_i, Y_i)\)'s fall inside the disk \(X_i^2 + Y_i^2 \leq 1\). In other words, it declares that the system’s precision quality is poor, although the empirical observed probability of nonconformance is \(\hat{p} = 0.6\). Such a disparity has a profound impact in applications to the producers and users of the system, as it can lead completely different conclusions on the precision quality of the system. Assumptions are important in statistical procedures, and thus it is interesting to investigate the effect of the specific \(N((0, 0), \sigma^2 I_{2 \times 2})\) model assumption on the sequential testing problem in (1) and (2).

Note that in the CEP context, it is standard to assume that the \((X_i, Y_i)\)'s are i.i.d. bivariate normal with \(E(X_i) = 0\) and \(E(Y_i) = 0\), as practitioners often adjust the bias of experiment outcomes so as to focus on the precision itself, see Fraser (1951), Fraser (1953), Solomon (1960), Harter (1960), Gillis (1991), Pyati (1993) and Shnidman (1995). Below we took a closer look at those 6 observed data \((X_i, Y_i)\)'s in Li et al. (2011). For this dataset, the point estimates of \(E(X_i), E(Y_i), \text{Var}(X_i), \text{Var}(Y_i)\) and \(\rho = \text{corr}(X_i, Y_i)\) are \(-0.148, 0.312, 0.326, 0.306\) and \(0.172\), respectively. Hence, the assumptions of equal variances (i.e., \(\text{Var}(X_i) = \text{Var}(Y_i)\)) and zero means (i.e., \(E(X_i) = E(Y_i) = 0\)) seem to be reasonable, as \(E(X_i)\) and \(E(Y_i)\) seem to be not significantly different from 0 given the observed variance values. However, the assumption of the correlation coefficient \(\rho = \text{corr}(X_1, Y_1) = 0\) seems questionable. We took a further step to use Fisher’s transformation to find that the 95% confidence interval on \(\rho\) is \([-0.65, 0.81]\), but we feel that such a wide confidence interval is due to the small sample size of 6, and does not necessarily prove that \(\rho = 0\). In addition, we should also mention that some weak correlations between \(X_i\)'s and \(Y_i\)'s for GPS data were also observed in Amiri-Simkooei (2009) based on the time series datasets from five permanent GPS stations—the correlation coefficients of the (north and east) coordinate components of GPS data are estimated as \(-0.06, -0.05, 0.07, 0.08\) and \(0.10\) for these five stations, respectively, and Amiri-Simkooei (2009) suspects that these weak correlations were intrinsic for GPS due to the simultaneous estimation of (north and east) coordinate components from the same dataset. Based on the above discussions, it is natural to investigate what happens to the sequential hypothesis testing problem in (1) and (2) under the bivariate normal model with zero means and equal variances when the correlation coefficient \(\rho\) between \(X_i\) and \(Y_i\) is small but not necessarily zero.

The primary objective of this paper is to understand the effect of misusing the correlation coefficient \(\rho = 0\) in the sequential hypothesis testing problem in (1) and (2). Intuitively, when the correlation coefficient \(\rho\) is not zero, but \(\rho = 0\) is erroneously assumed to develop a mis-specified SPRT, called mis-SPRT thereafter, for the problem in (1) and (2), the probabilities of Type I and Type II errors and the average sample numbers of the mis-SPRT will be different from what they should be under \(\rho = 0\). Here we will conduct a theoretical analysis to not only justify our intuition, but also gain a deeper understanding how the properties of the mis-SPRT vary as a function of \(\rho\) when \(\rho\) is small. Our results indicate that when the true correlation coefficient \(\rho \neq 0\), the stopping of the mis-SPRT is meaningless in the sense that the observed data still cannot provide enough evidence to decide which hypothesis in (1) is true subject to the Type I and Type II error probability constraints in (2).

A closely related objective of this paper is to develop a sequential test when the correlation coefficient \(\rho\) is unknown. While the sequential hypothesis testing problem with unknown nuisance parameters have been well studied in the literature, see, Wald (1947), to the best of our knowledge, no sequential tests have been developed in the context of CEP for two-dimensional random vector data. The main challenge is the computational complexity in the context of CEP when the correlation coefficient \(\rho \neq 0\). Here we demonstrate how to modify several existing sequential methods to the CEP context, and hopefully it will stimulate further research on sequential tests for CEP.

We should mention that the case of model or distributional mis-specification has been investigated in the statistical literature, though most, if not all, on the offline fixed-sample setting. For instance, Foutz and Srivastava (1977) considers the property of the fixed-sample likelihood ratio test when the (normal distribution) model is mis-specified. On the other hand, much research has been done on sequential tests of the bivariate normal model, or more generally, \(d\)-dimensional
exponential families, see, for example, Chan and Lai (2000) and also see the references in Siegmund (1985) or Lai (2001), they usually focus when the true model is indeed from either the null or the alternative hypothesis. To the best of our knowledge, no research has been done to investigate the sequential hypothesis testing problem in (1) and (2) under the bivariate normal model when the correlation coefficient $\rho$ is unknown and/or mis-specified as 0.

Note that here our formulation of evaluating a system’s precision quality is a hypothesis testing approach: we pre-specify the radius $a$, e.g., $a = 1$, and then test the hypotheses on $p = P(X_i^2 + Y_i^2 > a^2) = 1$ or $CEP = P(X_i^2 + Y_i^2 \leq 1)$. Another (possibly more popular) formulation is an estimation approach: one estimates the radius $a$ that satisfies, say, $0.5 = CEP = p = P(X_i^2 + Y_i^2 > a^2)$, and most, if not all, existing estimation methods assume that $(X_i, Y_i)$’s are bivariate normal with $E(X_i) = 0$ and $E(Y_i) = 0$, see Fraser (1951), Fraser (1953), Solomon (1960), Harter (1960), Gillis (1991), Pyati (1993) and Shnidman (1995). The effect of the correlation coefficient $\rho$ under the estimation approach was studied in Pyati (1993), which claimed that the point estimation of the radius $a$ “is not particularly sensitive to change in the correlation coefficient $\rho$”. However, no research has been done to investigate the effect of $\rho$ under our present sequential hypothesis testing approach, and this paper fills such a gap. Note that the estimation approach deals with the continuity of a function with respect to $\rho$, whereas the sequential hypothesis testing approach deals with more than the continuity property.

The remainder of this paper is organized as follows. In Section 2, we present the mis-SPRT that is developed under the mis-specified assumption of the correlation coefficient $\rho = 0$. In Section 3, we investigate how the performance of the mis-SPRT varies as a function of $\rho$, especially when $\rho \to 0$. In Section 4, we propose to apply several existing sequential methods to the CEP context when the true correlation coefficient $\rho$ is unknown, and the generalized sequential likelihood ratio test (GSLRT) proposed by Bangdiwala (1982) seems to work well. Section 5 provides some concluding remarks. The detailed technical proofs are provided in the Appendixes.

2. Preliminaries and mis-SPRT

In the hypothesis testing problem in (1) and (2), a sequential test design consists of an integer-valued random variable $N$ that indicates how many observations one should take, and a final decision $D \in \{0, 1\}$, where $(D = 0)$ means that one accepts $H_0 : p = p_0$ and $(D = 1)$ accepts $H_1 : p = p_1$. Note that the variable $N$ is a stopping time: the decision $\{N = n\}$ implies that we take $n$ observations when making a final decision, and such a decision is based only on the first $n$ observations.

As mentioned in the Introduction, in order to make a quick decision, Li et al. (2011) proposed to develop a sequential test for (1) and (2) under the assumption that the $(X_i, Y_i)$’s are i.i.d. with the specific bivariate normal distribution $N((0, 0), \sigma^2I_{2 \times 2})$. That is, the joint density function of $(X_i, Y_i)$’s is assumed to be

$$f_{\sigma^2}(x, y) = \frac{1}{2\pi \sigma^2} \exp \left\{ -\frac{x^2 + y^2}{2\sigma^2} \right\}, \quad (3)$$

where $\sigma > 0$ is the only unknown parameter that evaluates the precision quality of the system.

Below we will review the optimal sequential test under the model (3), as the ideas will help us better understand our discussions in the next sections when the model (3) is mis-specified. In this section and only in this section, we are working under the assumption of the bivariate normal model in (3). Under the model (3), it can be shown that the hypothesis testing problem in (1) becomes the problem of utilizing $(X_i, Y_i)$’s to test the unknown variance parameter $\sigma^2$ in (3):

$$H_0^* : \sigma^2 = \sigma_0^2 \quad \text{versus} \quad H_1^* : \sigma^2 = \sigma_1^2, \quad (4)$$

where

$$\sigma_j^* = \frac{1}{\sqrt{2|\ln(p_j)|}} \quad (5)$$

for $j = 0, 1$. Here $p_0$ and $p_1$ are the constants in (1), and $0 < \sigma_0^* < \sigma_1^* < \infty$, since $0 < p_0 < p_1 < 1$.

To see the equivalence of these two problem formulations in (1) and (4) under the model (3) assumption, it suffices to characterize the relationship between the probability of nonconforming, $p = P(X_i^2 + Y_i^2 > 1)$, and the $\sigma$ parameter of the model (3). Under the model assumption (3), it is clear that

$$p = \int_{\sigma^2(0,0,1)} \frac{1}{2\pi \sigma^2} \exp \left\{ -\frac{x^2 + y^2}{2\sigma^2} \right\} \, dx \, dy$$

$$= \int_{1}^{\infty} \int_{0}^{2\pi} \frac{1}{2\pi \sigma^2} \exp \left\{ -\frac{r^2}{2\sigma^2} \right\} \, r \, dr \, d\nu \quad \text{(let} x = r \cos \nu, \ y = r \sin \nu \text{)}$$

$$= \exp \left\{ -\frac{1}{2\sigma^2} \right\}. \quad (6)$$

Alternatively, a simpler argument to prove (6) is as follows: under the assumption (3), $p$ is the probability that a chi-squared variable $X^2 + Y^2$ with 2 degree of freedom, which is Exponential (2), exceeds $1/\sigma^2$. From the relation (6), solving $\sigma_0^*$ and $\sigma_1^*$ as a function of $p_0$ and $p_1$ yields (5), and thus the two hypothesis testing problems in (1) and (4) are equivalent.
Now the hypothesis testing problem (4) under the model assumption (3) is a special case of well-studied sequential hypothesis testing problems in the literature, and it is well-known (Wald, 1947) that the corresponding optimal sequential test is the SPRT defined as follows. Define the likelihood ratio statistic at time $n$ as

$$
\lambda_n^* = \prod_{i=1}^{n} f_{\rho_n^2}(X_i, Y_i) = \left( \frac{\sigma_0^2}{\sigma_1^2} \right)^n \exp \left\{ \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^{n} (X_i^2 + Y_i^2) \right\}.
$$

(7)

Then the SPRT stops at the first time

$$
N^* = \inf \left\{ n \geq 1 : \ln(\lambda_n^*) \notin [-B, A] \right\},
$$

(8)

where “inf” denotes that $N^* = \infty$ when no such $n$ exists. When $N^*$ in (8) stops, decision

$$
D^* = \begin{cases} 
0, & \text{if } \ln(\lambda_{N^*}) < -B \\
1, & \text{if } \ln(\lambda_{N^*}) > A.
\end{cases}
$$

(9)
is made, where $\{D^* = 0\}$ means that one accepts $H_0^*$, or equivalently, accepts $H_0$ in (1), and $\{D^* = 1\}$ accepts $H_1^*$ in (4) and $H_1$ in (1).

For the SPRT $(N^*, D^*)$ in (7)–(9), the thresholds $A$ and $B$ in (8) are chosen to satisfy the error probability constraints in (2) when $H_0$ and $H_1$ in (1) are replaced by $H_0^*$ and $H_1^*$ in (4). It is well-known from Wald (1947) and Siegmund (1985) that $A = (1 + o(1)) \ln((1 - \beta)/\alpha)$ and $B = (1 + o(1)) \ln((1 - \alpha)/\beta)$, as $\alpha + \beta \to 0$.

Note that the SPRT $(N^*, D^*)$ in (7)–(9) is developed under the specific bivariate normal model assumption (3) when the correlation coefficient $\rho$ is 0. However, as mentioned in the Introduction, it is possible that the true correlation coefficient $\rho \neq 0$ in practice. In such a scenario, the SPRT $(N^*, D^*)$ in (7)–(9) is still a well-defined sequential test design, but unfortunately its true performance will no longer be the same as what we thought for the case of $\rho = 0$, let alone the (asymptotic) optimality properties.

In the remainder of this paper, we assume that the true correlation coefficient $\rho \neq 0$ for the bivariate normal random vector $(X_i, Y_i)$. To avoid confusion, the SPRT design $(N^*, D^*)$ in (7)–(9) is referred to as the mis-SPRT, so as to emphasize the fact that it is developed under the mis-specified model (3) when $\rho$ is mis-specified as 0.

3. Properties of the mis-SPRT

In this section, we will investigate the properties of the mis-SPRT $(N^*, D^*)$ in (7)–(9) when the true correlation coefficient $\rho \neq 0$ is known. Later a sequential design will be developed in the next section when the true $\rho \neq 0$ is unknown.

Let us first state our true model to avoid any confusion. As mentioned in the Introduction, we assume that the observed $(X_i, Y_i)$’s are i.i.d. with a bivariate normal distribution

$$
N \left( (0, 0), \sigma^2 \left( \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right),
$$

(10)
i.e., the density function of $(X_i, Y_i)$’s is

$$
f_{\rho, \sigma^2}(x, y) = \frac{1}{2\pi \sigma^2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{x^2 + y^2 - 2\rho xy}{2\sigma^2(1 - \rho^2)} \right\},
$$

where $\rho \in (-1, 1)$ is the correlation coefficient between $X_i$’s and $Y_i$’s.

Next, as in (4), we can derive the equivalent formulation of (1) under the model (10) when the true correlation coefficient $\rho \neq 0$ is known. In this case, using the same logic as in (6), we have

$$
p = \mathbb{P}(X_i^2 + Y_i^2 > 1)
$$

$$
= \int_{x^2+y^2>1} \frac{1}{2\pi \sigma^2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{x^2 + y^2 - 2\rho xy}{2\sigma^2(1 - \rho^2)} \right\} dx dy
$$

$$
= \int_0^{2\pi} \int_1^{\infty} \frac{1}{2\pi \sigma^2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1 - \rho \sin(2\nu)}{2\sigma^2(1 - \rho^2)} \right\} r dr d\nu
$$

$$
= \frac{\sqrt{1 - \rho^2}}{2\pi} \int_0^{2\pi} \frac{1}{1 - \rho \sin(2\nu)} \exp \left\{ -\frac{1 - \rho \sin(2\nu)}{2\sigma^2(1 - \rho^2)} \right\} d\nu,
$$

(11)

where the last equality uses the fact that $\int_1^{\infty} \exp(-ar^2)r dr = \exp(-a)/(2a)$ for any $a > 0$. It is clear from (11) that the probability $p$ of nonconforming depends on both $\rho$ and $\sigma$, but unfortunately $p = p(\rho, \sigma)$ generally does not have a simple form as in (6) when $\rho = 0$. Indeed, the theoretical and computational challenges in the context of the CEP are mainly due to the messy function in (11).
Theoretically, for any $\rho \neq 0$, we can still use Eq. (11) to define $\sigma_0$ and $\sigma_1$ as those values satisfying $p_j = p(\rho, \sigma_j)$ for $j = 0$ and 1, where $p_0$ and $p_1$ are given in (1). Note that $\sigma_0$ and $\sigma_1$ should depend on the actual value of $\rho$, and here we suppress the index $\rho$ in the notation $\sigma_{0\rho}$ and $\sigma_{1\rho}$ for convenience, as we assume $\rho \neq 0$ is known in this section. It is well-known (Pyati, 1993; Shnidman, 1995) that there is no closed-form expression of $\sigma_j$ in term of $p_j$ as in (5), but we should be able to numerically solve $\sigma_0$ and $\sigma_1$ from (11) for $p = p_0$ and $p_1$ for a given $\rho \neq 0$.

Therefore, under the model assumption (10), the original hypothesis test problem in (1) reduces to the problem of using the data $(X_i, Y_i)$'s to test the hypotheses

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{versus} \quad H_1 : \sigma^2 = \sigma_1^2 \quad (0 < \sigma_0 < \sigma_1 < \infty).$$

(12)

It is evident that (4) is a special case of (12) when $\rho = 0$, but they are different when $\rho \neq 0$.

In the remainder of this section, we will investigate the performance of the mis-SPRT $(N^*, D^*)$ in (7)–(9) in the hypothesis testing problem (12) under the model (10) when $\rho \neq 0$ is known. Section 3.1 reports the theoretical results, whereas Section 3.2 presents the numerical simulation results.

### 3.1. Theoretical results

To present our theoretical results, we need to first introduce some notation. Recall that $p_0$ and $p_1$ are given in the null and alternative hypotheses in (1), and the true correlation coefficient $\rho \neq 0$ in the model (10) is assumed to be known in this section. For the value $\sigma_0$ in the null hypothesis of (12), we shall define two new constants, $\tilde{\rho}$ and $\tilde{\sigma}_1^2$, that satisfy

$$\begin{align*}
\tilde{\rho} = \frac{\rho - \sigma_0^2 (1 - \rho^2)}{\sigma_0^2 (1 - \rho^2)}; \\
\ln \left( \frac{|\ln(p_0)| - |\ln(p_1)|}{|\ln(p_0)| - |\ln(p_1)|} \right) = \frac{2\sigma_0^2 (1 - \rho^2)}{1 - \rho / \tilde{\rho}} \ln \left( \frac{\tilde{\sigma}_1^2 \sqrt{1 - \tilde{\rho}^2}}{\sigma_0^2 \sqrt{1 - \rho^2}} \right).
\end{align*}$$

(13)

The rationale of (13) and the existence of such a pair $(\tilde{\rho}, \tilde{\sigma}_1^2)$ are presented in the proof of Theorem 1 in the Appendix A. At the high-level description, $(\tilde{\rho}, \tilde{\sigma}_1^2)$ in (13) yields a new pseudo probability measure of $(X_i, Y_i)$'s:

$$\tilde{H}_1 : (X_i, Y_i) \sim N \left(0, 0, \tilde{\sigma}_1^2 \left(\frac{1}{\tilde{\rho}} \tilde{\rho} \frac{1}{\rho} \right)\right),$$

(14)

and the mis-SPRT design $(N^*, D^*)$ in (7)–(9) is the likelihood ratio test (up to a scalar) in the problem of testing $H_0 : \sigma^2 = \sigma_0^2$ in (12) against $\tilde{H}_1$ in (14).

With the notation of $(\tilde{\rho}, \tilde{\sigma}_1^2)$ in (13), the following theorem summarizes the true performance of the mis-SPRT design $(N^*, D^*)$ in (7)–(9) when the observations $(X_i, Y_i)$'s are actually from the model in (10) with the true $\rho \neq 0$. The proof of Theorem 1 is presented in Appendix A.

**Theorem 1.** When the true correlation coefficient $\rho \neq 0$, the Type I error probability of the mis-SPRT design $(N^*, D^*)$ under the null hypothesis $H_0 : p = p_0$ satisfies

$$\mathbb{P}_{H_0}(N^* \text{ rejects } H_0) \leq \exp \left\{ - \frac{A(1 - \rho / \tilde{\rho})}{2\sigma_0^2 (1 - \rho^2) \left| \ln(p_0) - |\ln(p_1)| \right|} \right\},$$

(15)

and the corresponding expected sample size under $H_0 : p = p_0$ is

$$E_{H_0}(N^*) = (1 + o(1)) |\ln(\beta)| / I_0,$$

(16)

as the error probability bounds $\alpha$ and $\beta$ in (2) go to 0 simultaneously, where $I_0$ is defined by

$$I_0 = \ln \left( \frac{|\ln(p_0)|}{|\ln(p_1)|} \right) - 2\left| \ln(p_0) - |\ln(p_1)| \right| \sigma_0^2.$$

(17)

Similar conclusions hold to the Type II error $\mathbb{P}_{H_1}(N^* \text{ rejects } H_1)$ and $E_{H_1}(N^*)$ under the alternative hypothesis $H_1 : p = p_1$, where the roles of $p_0$ and $p_1$ are interchanged and $(\beta, A, \tilde{\rho}, \tilde{\sigma}_1^2, \tilde{\sigma}_2^2)$ is replaced by $(\alpha, B, \tilde{\rho}, \tilde{\sigma}_2^2, \tilde{\sigma}_0^2)$ with $(\tilde{\rho}, \tilde{\sigma}_1^2)$ being defined similarly to (13).

While Theorem 1 gives the theoretical properties of the mis-SPRT design $(N^*, D^*)$ under the true $\rho \neq 0$, it is still difficult to understand the explicit effect of misusing $\rho = 0$, since the quantities $\tilde{\rho}$ and $\tilde{\sigma}_1^2$ in Theorem 1 depend on $\rho$ and they do not have explicit expressions. The main reason is that relation (11) characterizes $p = P(X_i^2 + Y_i^2 > 1) = p(\rho, \sigma)$ as a function of $\sigma^2$ and $\rho$, but the corresponding function is too complicated when $\rho \neq 0$.

To clearly see the effect of the correlation coefficient $\rho$ in Theorem 1, we take a further step by considering an asymptotic regime when $\rho \to 0$ while fixing $(p_0, p_1)$ and $(\alpha, \beta)$ in the original hypothesis testing problem in (1) and (2). The
This is summarized in Corollary 1 below, whose proof will be provided in Appendix C.

**Proposition 1.** As \( \rho \to 0 \), if

\[
p \geq \exp \left\{ - \frac{\rho^2(1 - \rho^2) - 2}{4\rho^2(1 - \rho^2)} + o(\rho^2) \right\},
\]

then the value of \( \sigma^2 \) in (11) satisfies

\[
\sigma^2 = \frac{1}{2|\ln(p)|} + \frac{2 + \ln(p)}{8|\ln(p)|} \rho^2 + o(\rho^2).
\]

It is useful to point out that the assumption (18) of Proposition 1 is not demanding for small \( \rho \), e.g., the right-hand side of (18) without the \( o(\rho^2) \) term reaches its maximum value 0.004947 at \( \rho = 0.5983 \), and as an illustration, has the value of \( 4.65 \times 10^{-12} \) when \( \rho = 0.2 \). As shown in the proof in Appendix B, this assumption is a necessary condition to solve a quadratic equation of \( \sigma^2 \) from the Taylor series expansion of (11).

Given Proposition 1 and relation (13), for fixed \((p_0, p_1)\), it is not difficult to solve \( \sigma^2 \) and \((\bar{\rho}, \bar{\sigma}_1^2)\) asymptotically as functions of \( \rho \) when \( \rho \to 0 \). Hence, the conclusions of Theorem 1 can be rewritten explicitly in terms of \( \rho \) as \( \rho \to 0 \). This is summarized in Corollary 1 below, whose proof will be provided in Appendix C.

**Corollary 1.** Assume \( 0 < p_0 < p_1 < 1 \) satisfies (18). Denote by \( \psi_0(\rho) \) the right-hand side of (15) in Theorem 1. Then as the correlation coefficient \( \rho \to 0 \),

\[
\psi_0(\rho) = \exp \left\{ -A \left\{ 1 - \frac{\xi_0(\xi_0 - 1)}{4(\xi_0 - \ln(\xi_0) - 1)} \rho^2 + o(\rho^2) \right\} \right\},
\]

where \( \xi_0 = |\ln(p_0)|/|\ln(p_1)| \). Meanwhile, if we denote by \( \omega_0(\rho) \) the term \( 1/l_0 \) in the right-hand side of (16) in Theorem 1, then we have

\[
\omega_0(\rho) = c + s \rho^2 + o(\rho^2),
\]

as \( \rho \to 0 \), where the constants \( c \) and \( s \) are given by

\[
c = \left[ \ln \left( \frac{|\ln(p_0)|}{|\ln(p_1)|} \right) + \frac{|\ln(p_1)|}{|\ln(p_0)|} - 1 \right]^{-1},
\]

and

\[
s = \frac{c^2[2 + \ln(p_0)][|\ln(p_0)| - |\ln(p_1)|]}{4|\ln(p_0)|}.
\]

Similarly, the properties of the mis-SPRT under \( H_1 \) can be defined by \( \psi_1(\rho) \) and \( \omega_1(\rho) \) which are the same as \( \psi_0(\rho) \) and \( \omega_0(\rho) \) except that the roles of \( p_0 \) and \( p_1 \) are switched and \((A, \beta)\) is replaced by \((B, \alpha)\).

It is useful to emphasize the importance of the functions \( \psi_0(\rho) \) and \( \omega_0(\rho) \) in Corollary 1. As a numerical illustration, when \( p_0 = 0.1 \) and \( p_1 = 0.2 \), these two functions become

\[
\psi_0(\rho) = \exp \left\{ -A + 0.829A\rho^2 + o(A\rho^2) \right\},
\]

where \( A \) is fixed in Corollary 1 and is chosen for the mis-SPRT to satisfy the error probability constraints in (2) under the mis-specified model (3) with \( \rho = 0 \), and

\[
\omega_0(\rho) = 17.508 - 6.980\rho^2 + o(\rho^2).
\]

When \( \rho = 0 \), the values of these two functions, \( \psi_0(\rho = 0) = \exp(-A) \) and \( \omega_0(\rho = 0) = c \), are exactly those values that characterize Type I error probability and ASN of the mis-SPRT \((N^l, D^r)\) in (7)–(9) under the null hypothesis \( H_0 : p = p_0 \) when \( \rho = 0 \) is true, see Siegmund (1985).

From Corollary 1, it is useful to discuss the robustness of mis-SPRT with respect to the correlation coefficient \( \rho \). On the one hand, the functions \( \psi_0(\rho), \psi_1(\rho), \omega_0(\rho) \) and \( \omega_1(\rho) \) are smooth functions of \( \rho \). This is consistent with our intuition that the performance of mis-SPRT should vary continuously as a function of \( \rho \). On the other hand, the coefficients of \( \rho^2 \) depend only on whether \( 2 + \ln(p_1) > 0 \) or not. In other words, when testing the hypotheses in (1), the critical value for both \( p_0 \) and \( p_1 \) is \( \exp(-2) \approx 0.1353 \). For example, when \( |\rho| \) is small, \( \psi_0(\rho) \), or equivalently, Type I error probability of mis-SPRT, would decrease as \( |\rho| \) increases when \( p_1 < \exp(-2) \) but will increase when \( p_1 > \exp(-2) \).
A more interesting scenario is when we further let $\alpha + \beta \rightarrow 0$ as in Theorem 1. Then $A = (1 + o(1)) \ln((1 - \beta)/\alpha) \rightarrow +\infty$ and thus the term $A \rho^2$ can be significant in $\psi_0(\rho)$, i.e., the ratio $\psi_0(\rho)/\psi_0(\rho = 0)$ can be significantly large even for small $\rho$. Meanwhile, the function $\omega_0(\rho)$ only depends on $(p_0, p_1)$ and does not depend on $(\alpha, \beta)$. That is, loosely speaking, the ASN of the mis-SPRT seem to be robust with respect to the correlation coefficient $\rho$, but the effect of a small $\rho \neq 0$ on the probabilities of Type I and Type II errors of the mis-SPRT can be severe when the constraints $\alpha$ and $\beta$ in (2) are also small.

The impact of Theorem 1 and Corollary 1 on the practitioners is as follows. When testing hypotheses in (1) and (2), suppose a test designer treats the correlation coefficient $\rho$ in (10) as 0, decides to use the mis-SPRT design $(N^*, D^*)$ in (7)–(9) due to its optimal properties, and then follows the standard sequential test literature to choose the thresholds $A$ and $B$ of the mis-SPRT. The designer may expect the ASNs of the mis-SPRT, $E_0(N^*)$ and $E_1(N^*)$, to be roughly $|\ln(\beta)|\omega_0(\rho = 0)$ and $|\ln(\alpha)|\omega_1(\rho = 0)$, respectively, and expect that the mis-SPRT can decide which hypothesis in (1) is true when stopping taking observations subject to the error probability constraints in (2). Unfortunately, if the true correlation coefficient $\rho$ is not 0, the actual performance of the mis-SPRT, especially the actual error probabilities, will deviate from one’s expectation. Table 1 summarizes the detailed deviation between the actual performance of the mis-SPRT design and the designer’s expectation, depending on the relationship between the null hypothesis $p_0$, the alternative hypothesis $p_1$, and the critical value $\exp[-2] \approx 0.1353$.

<table>
<thead>
<tr>
<th>Values of $p_0$ &amp; $p_1$</th>
<th>Is $P_{H_0}(N^* \text{ rejects } H_0) \leq \alpha$?</th>
<th>Is $P_{H_1}(N^* \text{ rejects } H_1) \leq \beta$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_0 &lt; p_1 &lt; \exp[-2]$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>$p_0 &gt; \exp[-2] &lt; p_1$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>$\exp[-2] &lt; p_0 &lt; p_1$</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1

The actual performance of the mis-SPRT design when $\rho \neq 0$.

It is clear from Table 1 that the mis-SPRT design $(N^*, D^*)$ does not satisfy the Type I and Type II error probability constraints in (2) when the true $\rho \neq 0$, no matter what combination of $p_0$ and $p_1$ ($> p_0$) is. It is also interesting to see that for the concrete example in Li et al. (2011), one faces the problem in (1) with $p_0 = 0.1$ and $p_1 = 0.4$, which belongs to the case of $p_0 < \exp(-2) < p_1$. In that case, when the true correlation coefficient $\rho \neq 0$, the mis-SPRT design stops earlier than what we expected, but unfortunately it also has larger probabilities on both Type I and Type II errors than what we expected. In other words, the earlier stopping of the mis-SPRT is meaningless: when the mis-SPRT stops taking observations, we still do not have enough evidence to decide which hypothesis in (1) is true, subject to the Type I and Type II error probability constraints in (2).

3.2. Numerical simulations

In this section, we run numerical simulation studies to validate our theoretical results. We consider the hypothesis testing problem in (1) under three different scenarios, depending on the relationship between the values $(p_0, p_1)$ and the critical value $\exp[-2] \approx 0.1353$:

(I) : $p_0 = 0.05$ and $p_1 = 0.1$  \quad (p_0 < p_1 < \exp[-2]) \quad (19)$

(II) : $p_0 = 0.1$ and $p_1 = 0.2$  \quad (p_0 < \exp[-2] < p_1) \quad (20)$

(III) : $p_0 = 0.2$ and $p_1 = 0.3$  \quad (\exp[-2] < p_0 < p_1) \quad (21)$

For the Type I and Type II error probability constraints (2), we set $\alpha = \beta = 0.05$. For each of these three scenarios in (19)–(21), we consider the corresponding mis-SPRT design $(N^*, D^*)$ in (7)–(9), and run numerical simulations to find the thresholds $A$ and $B$ of the mis-SPRT so as to satisfy the constraint in (2) by treating the correlation coefficient $\rho$ of $(X_i, Y_i)$’s as 0. Next, for each scenario, we simulate the true properties of the mis-SPRT design $(N^*, D^*)$ for $\rho \neq 0$ by generating $(X_i, Y_i)$’s from the model (10). Since $\rho$ is usually small in practice and the results are also symmetric to $\pm \rho$, we only consider the case of $\rho \in [0, 0.5]$ in our simulations. Figs. 1 and 2 plot the real error probabilities and ASNs of the mis-SPRT design as functions of $\rho \in [0, 0.5]$, and each curve in the plot corresponds to each of three scenarios. All numerical results are based on $10^6$ Monte Carlo runs.

It is clear from Fig. 1 that when the true $\rho \neq 0$, the Type I and Type II error probabilities of the mis-SPRT can be significantly larger than the pre-specified constraints $\alpha = \beta = 5\%$, and the situation becomes worse if $\alpha$ and $\beta$ are smaller. On the other hand, Fig. 2 shows that the ASN curves of the mis-SPRT in Fig. 2 seem to be relatively flat. All these numerical simulation results are consistent with our theoretical results in Table 1.

4. Our proposed sequential tests

In this section, we develop sequential tests for the hypothesis testing problem (1) subject to the Type I and Type II error constraints in (2), under the assumption that the observations $(X_i, Y_i)$’s are i.i.d. with the bivariate normal distribution in (10).
with an unknown correlation coefficient $\rho$. In practice $\rho$ is usually small, and thus it is reasonable to make an additional assumption that $|\rho| \leq \rho_0$ for some pre-specified constant $\rho_0$, say, $\rho_0 = 0.5$.

When $\rho \neq 0$ is unknown or not completely specified, the original hypothesis testing problem (1) can be rewritten as the problem of testing

$$H_0 : p(\rho, \sigma^2) = p_0 \quad \text{versus} \quad H_1 : p(\rho, \sigma^2) = p_1,$$

where $p(\rho, \sigma^2) = p = P(X_i^2 + Y_i^2 > 1)$ is defined in (11). In (22), $\rho$ can be regarded as an unknown nuisance parameter with $|\rho| \leq \rho_0$.

Mathematically, as in (12), one can solve $\sigma^2 = \sigma^2(p, \rho)$ from (11), and then the problem in (22) will be equivalent to the problem of testing composite hypotheses on $\sigma^2 = \sigma^2(p, \rho)$ with an unknown nuisance parameter $|\rho| \leq \rho_0$. While many efficient sequential tests such as Wald’s method of weight functions (Wald, 1947) have been developed for testing composite hypotheses, unfortunately they are generally computationally challenging in our context due to the messy function in (11).

More importantly, if we interpret our problem as testing composite hypotheses, then any point in the composite null hypothesis is uniquely related to a point in the composite alternative hypothesis (and vice versa) through a unique value of nuisance parameter $\rho$. Hence, it is more appropriate to view our problem as a functional hypothesis testing problem. Fortunately, such a problem has also been studied in other contexts and several methods have been developed, e.g., the generalized likelihood ratio test (GLRT) of Lai and Zhang (1994), and the generalized sequential likelihood ratio test (GSLRT) of Bangdiwala (1982). In addition, an intuitive method is the “plug-in” SPRT test that replaces the unknown true parameter $\rho$ in the SPRT by its maximum likelihood estimate (MLE). Below we demonstrate how to apply these methods in our context.

For the purpose of easy understanding, denote by $\theta = (\rho, \sigma^2)$. Then we are testing hypotheses on $p = p(\theta) = p(\rho, \sigma^2)$ in (22). It is straightforward to show that given the first $n$ i.i.d. two-dimensional vector observations, $\{(X_i, Y_i)\}_{i=1}^{n}$, the MLE
of \( \theta \) under the model assumption (10) is \( \hat{\theta}_n = (\hat{\rho}_n, \hat{\sigma}^2_n) \), where

\[
\hat{\rho}_n = 2 \sum_{i=1}^{n} (X_i Y_i) / \sum_{i=1}^{n} (X_i^2 + Y_i^2), \quad \hat{\sigma}^2_n = \frac{n}{n} \sum_{i=1}^{n} (X_i^2 + Y_i^2) / (2n).
\]  

(23)

Note that \( \hat{\rho}_n \) in (23) is consistent with the standard formula of correlation coefficient \( \sum (X_i - \mu_X)(Y_i - \mu_Y)/(n \sigma_X \sigma_Y) \) under the additional assumptions that \( \mu_X = \mu_Y = 0 \) and \( \sigma^2_X = \sigma^2_Y = \sigma^2 \).

In the remainder of this section, each of Sections 4.1–4.3 presents one of the three different sequential methods: the plug-in SPRT method, the GLRT method of Lai and Zhang (1994), and the GSLRT method of Bangdiwala (1982). In Section 4.4, we report numerical simulation results of these three methods. Our focus is on the stopping time of the sequential test for each method, as the decision whether to accept \( H_0 \) or \( H_1 \) is evident when stopping taking observations. The threshold(s) of each test will be chosen numerically to satisfy the Type I and Type II error probability constraints in (2), and the details will be explained in Section 4.4.

4.1. The plug-in SPRT

As in (12), when the correlation coefficient \( \rho \) were known, we can solve \( \sigma_{0\rho} \) and \( \sigma_{1\rho} \) from (11) with \( p = p_0 \) and \( p = p_1 \), where we add the subscript \( \rho \) to emphasize the dependence of \( \sigma \) on the unknown \( \rho \) for \( j = 0, 1 \) in this section. Then the original hypothesis testing problem in (1) becomes the problem of testing \( H_0 : \sigma^2 = \sigma^2_{0\rho} \) and \( H_0 : \sigma^2 = \sigma^2_{1\rho} \) under the bivariate normal model in (10). Hence, the corresponding optimal test is the SPRT that makes decisions based on the following log-likelihood ratio statistic

\[
\ln(\lambda_n) = \sum_{i=1}^{n} \ln [ f_{\rho, \sigma^2_{0\rho}}(X_i, Y_i) / f_{\rho, \sigma^2_{1\rho}}(X_i, Y_i) ]
\]

\[
= n \ln \left( \frac{\sigma_{0\rho}^2}{\sigma_{1\rho}^2} \right) + \frac{n}{2(1 - \rho^2)} \left( \frac{1}{\sigma_{0\rho}^2} - \frac{1}{\sigma_{1\rho}^2} \right) \sum_{i=1}^{n} (X^2_i + Y^2_i - 2\rho X_i Y_i).
\]

When the correlation coefficient \( \rho \) is unknown, it is natural to replace the true unknown \( \rho \) by its MLE \( \hat{\rho}_n \) in (23), and this yields the plug-in log-likelihood ratio statistic

\[
\ln(\lambda^\text{plug}_n) = n \ln \left( \frac{\sigma_{0\hat{\rho}_n}^2}{\sigma_{1\hat{\rho}_n}^2} \right) + \frac{n}{2(1 - \hat{\rho}_n^2)} \left( \frac{1}{\sigma_{0\hat{\rho}_n}^2} - \frac{1}{\sigma_{1\hat{\rho}_n}^2} \right) \sum_{i=1}^{n} (X^2_i + Y^2_i - 2\hat{\rho}_n X_i Y_i)
\]

\[
= n \ln \left( \frac{\sigma_{0\hat{\rho}_n}^2}{\sigma_{1\hat{\rho}_n}^2} \right) + \frac{n}{2} \left( \frac{1}{\sigma_{0\hat{\rho}_n}^2} - \frac{1}{\sigma_{1\hat{\rho}_n}^2} \right) \sum_{i=1}^{n} (X^2_i + Y^2_i).
\]

Here it is important to recall that \( \sigma_{j\rho} = \sigma(p_j, \rho = \hat{\rho}_n) \) is the solution of (11) when \( \rho = \hat{\rho}_n \) and \( p = p_j \) for \( j = 0, 1 \).

The plug-in SPRT design can then be defined by the stopping time

\[
N^\text{plug} = \inf \left\{ n \geq 1 : \ln(\lambda^\text{plug}_n) \notin [-B, A] \right\},
\]

for some bounds \( A > 0 \) and \( B > 0 \).

4.2. The GLRT of Lai and Zhang (1994)

The GLRT method of Lai and Zhang (1994) is a general sequential test that is asymptotically Bayes risk efficient, and its intuition is to use the likelihood functions to summarize the evidence how unlikely data are from \( H_0 \) or \( H_1 \), with a large value implying the rejection of one of these two hypotheses. One simple extension of the GLRT method to our context of CEP is to define the test statistic \( W_n^{\text{GLRT}} \) at time \( n \) as

\[
W_n^{\text{GLRT}} = \max_{j=0,1} \left( \sum_{i=1}^{n} \ln(f_{\theta_j}(X_i, Y_i)) - \sup_{\theta : p(\theta) = p_j} \sum_{i=1}^{n} \ln(f_0(X_i, Y_i)) \right),
\]

and \( \hat{\theta}_n \) is the MLE of \( \theta \) in (23).

The GLRT method is then defined by the stopping time

\[
N^{\text{GLRT}} = \inf \left\{ n \geq 1 : W_n^{\text{GLRT}} \geq C \right\},
\]

for some threshold \( C > 0 \).

We need to emphasize that we consider a very simplified version of the GLRT method of Lai and Zhang (1994), particularly with the constant threshold bound \( C \), to reduce the computational complexity in our context. Indeed, the computation of
the test statistic \( W_n^{\text{GLRT}} \) is straightforward but time consuming. Note that
\[
\sum_{i=1}^{n} \ln (f_{\hat{\theta}_n}(X_i, Y_i)) = \sum_{i=1}^{n} \ln (f_{\hat{\theta}_n}(X_i, Y_i)) = n \ln (2\pi \hat{\sigma}_{\nu}^2 \sqrt{1 - \hat{\rho}_n^2} - \sum_{i=1}^{n} (X_i^2 + Y_i^2 - 2\hat{\rho}_n X_i Y_i) /[2\hat{\sigma}_{\nu}^2 (1 - \hat{\rho}_n^2)],
\]
and
\[
\sup_{\theta : p(\theta) = p(\theta_n)} \sum_{i=1}^{n} \ln (f_i(X_i, Y_i)) = \sup_{|\rho| \leq \rho_0} \left\{ n \ln (2\pi \sigma_{\nu}^2 \sqrt{1 - \rho^2}) - \sum_{i=1}^{n} (X_i^2 + Y_i^2 - 2\rho X_i Y_i) /[2\sigma_{\nu}^2 (1 - \rho^2)] \right\},
\]
where \( \sigma_{\nu}^2 = \sigma^2(p, \rho) \) is the solution of (11) when \( p = p_j \) for \( j = 0, 1 \).

4.3. The GSLRT of Bangdiwala (1982)

In the context of testing hypotheses on \( p = p(\theta) = p(\rho, \sigma^2) \) in (22), the essential idea of the GSLRT method in Bangdiwala (1982) is to compare \( p(\hat{\theta}_n) \), the MLE of \( p \), with \( (p_0 + p_1)/2 \). The corresponding test statistic is based on the “normalized” version of \( p(\hat{\theta}_n) - (p_0 + p_1)/2 \) with respect to the variance of \( p(\hat{\theta}_n) \), which can be estimated from the observed Fisher information matrix based on the asymptotic properties of the MLE.

To be more concrete, the test statistic of the GSLRT method is defined as
\[
T_n = n(p_1 - p_0)(\hat{\gamma}_n^2)^{-1} \left[ p(\hat{\theta}_n) - (p_0 + p_1)/2 \right],
\]
where \( \hat{\gamma}_n^2 \) is the estimated variance of \( p(\hat{\theta}_n) \) defined by
\[
\hat{\gamma}_n^2 = \hat{p}(\hat{\theta}_n) \left[ \frac{1}{n} G(\hat{\theta}_n) \right]^{-1} \hat{p}(\hat{\theta}_n)^\top,
\]
with \( \cdot \) denoting transposition of a vector. Here
\[
\hat{p}(\hat{\theta}_n) = \begin{pmatrix} \frac{\partial p(\rho, \sigma^2)}{\partial \rho} & \frac{\partial p(\rho, \sigma^2)}{\partial \sigma^2} \end{pmatrix} \bigg| \rho = \hat{\rho}_n, \sigma^2 = \hat{\sigma}_n^2
\]
is the first-order derivative of \( p(\theta) \) in (11) with respect to \( \theta = (\rho, \sigma^2) \), and
\[
G(\hat{\theta}_n) = - \frac{\partial^2 \ln(\lambda_n)}{\partial \theta \partial \theta^\top} \bigg| \theta = \hat{\theta}_n
\]
is the observed Fisher information matrix.

Based on the test statistic \( T_n \) in (24), the GSLRT is defined by the stopping time
\[
N^{\text{GSLRT}} = \inf \{ n \geq n_0 : T_n \notin [-B, A] \}.
\]
Here \( n_0 \) is the initial sample size and it can be chosen as
\[
n_0 = (p_1 - p_0)^{-1 - \delta}
\]
for some \( \delta \in (0, 1) \).

In our context of two-dimensional observations \( (X_i, Y_i) \)’s with the bivariate normal distribution (10), the computation of \( T_n \) in (24) is tedious, especially the estimated variance \( \hat{\gamma}_n^2 \). The easier part is the observed Fisher information matrix, which is given by
\[
G(\hat{\theta}_n) = \begin{pmatrix} 1/\hat{\sigma}_n^4 & -\hat{\rho}_n/\hat{\sigma}_n^2 (1 - \hat{\rho}_n^2) \\ -\hat{\rho}_n/\hat{\sigma}_n^2 (1 - \hat{\rho}_n^2) & (1 + \hat{\rho}_n^2)/(1 - \hat{\rho}_n^2) \end{pmatrix},
\]
where the MLE \( \hat{\theta}_n = (\hat{\rho}_n, \hat{\sigma}_n^2) \) is given in (23). A more computationally expensive part is the computation of the first-order derivative \( \hat{p}(\hat{\theta}_n) \). By (11), we have
\[
\frac{\partial p(\rho, \sigma^2)}{\partial \rho} = \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ -\frac{1 - \rho \sin(2\nu)}{2\sigma^2(1 - \rho^2)} \right\} \frac{\rho}{\sin(2\nu)} - \frac{\rho(1 - \rho^2)}{\sin(2\nu)} \right\} \, d\nu,
\]
\[
\frac{\partial p(\rho, \sigma^2)}{\partial \sigma^2} = \frac{1}{4\pi \sigma^4 \sqrt{1 - \rho^2}} \int_0^{2\pi} \exp \left\{ -\frac{1 - \rho \sin(2\nu)}{2\sigma^2(1 - \rho^2)} \right\} \, d\nu.
\]
Then the actual values of \( \hat{p}(\hat{\theta}_n) \) can be computed through numerical integration via computer programming.
From Table 2, it is interesting to see that the plug-in SPRT has smaller ASNs under $\rho$ value and $A$ the GSLRT are $\rho = 0$ respectively. Meanwhile, the ASNsof the GSLRT tests are 50.26 and 40.15, which are slightly worse than the mis-SPRT when version to speed up computation. Among the three methods, the GSLRT seems to be the best choice in the sense of balancing properties need further investigation. Meanwhile, the GLRT has largest ASNs, possibly because we adopted a simplified performance, and the value $\alpha$ constraint in (2) with $\rho = \pm 0.5$.

The GLRT test

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Type I error</th>
<th>Type II error</th>
<th>$E_0(N)$</th>
<th>$E_1(N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0071 (0.0008)</td>
<td>0.0269 (0.0016)</td>
<td>83.17 (0.49)</td>
<td>70.95 (0.48)</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0070 (0.0008)</td>
<td>0.0251 (0.0016)</td>
<td>82.80 (0.50)</td>
<td>71.02 (0.48)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0081 (0.0009)</td>
<td>0.0390 (0.0017)</td>
<td>81.03 (0.50)</td>
<td>70.11 (0.48)</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0074 (0.0009)</td>
<td>0.0365 (0.0019)</td>
<td>74.42 (0.48)</td>
<td>68.72 (0.47)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0100 (0.0010)</td>
<td>0.0503 (0.0022)</td>
<td>56.35 (0.37)</td>
<td>66.49 (0.47)</td>
</tr>
</tbody>
</table>

The GSLRT test

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Type I error</th>
<th>Type II error</th>
<th>$E_0(N)$</th>
<th>$E_1(N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0495 (0.0022)</td>
<td>0.0461 (0.0021)</td>
<td>50.26 (0.37)</td>
<td>40.15 (0.30)</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0492 (0.0022)</td>
<td>0.0460 (0.0021)</td>
<td>50.38 (0.37)</td>
<td>40.18 (0.29)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0479 (0.0022)</td>
<td>0.0464 (0.0021)</td>
<td>50.33 (0.37)</td>
<td>40.03 (0.29)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0490 (0.0022)</td>
<td>0.0471 (0.0021)</td>
<td>50.55 (0.37)</td>
<td>39.91 (0.29)</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0490 (0.0022)</td>
<td>0.0481 (0.0021)</td>
<td>51.01 (0.38)</td>
<td>39.79 (0.29)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0483 (0.0021)</td>
<td>0.0498 (0.0022)</td>
<td>51.11 (0.38)</td>
<td>39.67 (0.29)</td>
</tr>
</tbody>
</table>

4.4. Numerical simulation studies

We now report our numerical simulation results for the sequential hypothesis testing problem (1) with $p_0 = 0.1$ and $p_1 = 0.2$, subject to the Type I and Type II error probability constraints in (2) with $\alpha = \beta = 0.05$. The data $(X_i, Y_i)$’s are assumed to be i.i.d. from the bivariate normal model (10) with unknown correlation coefficient $|\rho| \leq \rho_0 = 0.5$.

We will compare the properties of the three sequential tests in the previous three subsections: (1) The plug-in SPRT; (2) the GLRT of Lai and Zhang (1994); and (3) the GSLRT of Bangdiwala (1982). For each of these three sequential tests, we numerically search for the threshold(s) of the stopping time via brute force so as to satisfy the constraints in (2) with $\alpha = \beta = 0.05$ for any correlation coefficient value $|\rho| \leq \rho_0 = 0.5$.

The details of the numerical search of the thresholds are as follows. Assume the stopping time of a given sequential test involves two thresholds $A$ and $B$ (the situation is easier for the GLRT of Lai and Zhang (1994) which only involves one threshold $C$). First, we need to do two kinds of discretization: one is to discretize the interval $|\rho| \leq \rho_0 = 0.5$ by using points $\rho = 0, \pm 0.1, \pm 0.2, \ldots, \pm 0.5$, and the other is to discretize the domain $(0, \infty) \times (0, \infty)$ of $(A, B)$ with step size 0.01 until $A$ and $B$ are large, say $A = B = 20$. Next, for a given pair of thresholds, $(A, B)$, we simulate the error probabilities for each discrete point $\rho$ based on $10^4$ Monte Carlo runs, and then find the maximum error probabilities over the discretized $\rho$, which will be regarded as the Type I and type II error probabilities of the test with the thresholds $(A, B)$. Our numerical experiences suggest that the maximum error probabilities of the proposed three tests often occur at the boundary case of $\rho = \pm 0.5$, although we are unable to present a rigorous theoretical proof. Finally, we repeat the above process for each discretized pair $(A, B)$ until we find the smallest values of $A$ and $B$ whose Type I and type II error probabilities satisfy the constraint in (2) with $\alpha = \beta = 0.05$.

Using the above brute force algorithm, we found that the thresholds of the plug-in SPRT are $A = 3.75$ and $B = 3.40$, and the threshold of the GLRT of Lai and Zhang (1994) is $C = 5.30$. For the GSLRT of Bangdiwala (1982), it has an additional parameter $n_0$ in (26) for initial sample sizes. Our numerical experiences suggest that a larger value of $n_0$ yields a better performance, and the value $n_0 = 20$ seems to be a good choice in our simulation studies. When $n_0 = 20$, the thresholds of the GSLRT are $A = 1.58$ and $B = 5.62$.

Table 2 summarizes our numerical results of the three sequential tests, with their standard errors being in parentheses. From Table 2, it is interesting to see that the plug-in SPRT has smaller ASNs under $H_0$ when $|\rho| \geq 0.4$, but its theoretical properties need further investigation. Meanwhile, the GLRT has largest ASNs, possibly because we adopted a simplified version to speed up computation. Among the three methods, the GSLRT seems to be the best choice in the sense of balancing the tradeoff among asymptotic properties, computational time, and numerical performance on error probabilities and ASNs.

As suggested by a reviewer, it is also interesting to compare Table 2 with the solid line in Fig. 2 for the case of $p_0 = 0.1$ and $p_1 = 0.2$. In Fig. 2, when the correlation coefficient is indeed $\rho = 0$, the ASNs of the mis-SPRT test are 46.47 and 37.60, respectively. Meanwhile, the ASNs of the GSLRT tests are 50.26 and 40.15, which are slightly worse than the mis-SPRT when $\rho = 0$ is true. This is consistent with the fact that the mis-SPRT test is the optimal test when the true $\rho = 0$. Indeed, the mis-SPRT is not bad in the sense of having smaller ASNs than all three proposed sequential tests for any correlation coefficient value $\rho$, but unfortunately the error probabilities of the mis-SPRT are larger than the constraints $\alpha = \beta = 0.05$.
5. Concluding remarks

This paper is motivated from the application of evaluating a system’s precision quality, and investigates the effects of the correlation coefficient $\rho$ in the sequential hypothesis test problem in (1) and (2). Our theoretical analysis and numerical simulations show that when the true $\rho \neq 0$, but $\rho = 0$ is erroneously assumed, the performance of the mis-SPRT design is different from what they should be under $\rho = 0$. The specific effect of $\rho$ on the mis-SPRT design is derived asymptotically and is shown to be dependent on the relationship between $p_0$, $p_1$ and the critical value $\exp[-2] \approx 0.1353$. In particular, Type I and Type II error probabilities of the mis-SPRT can be significantly larger than the required constraints $\alpha$ and $\beta$ when the true $\rho \neq 0$.

For the case of unknown correlation coefficient $\rho$, we compare three existing methods that are computationally feasible in our contexts, and the GSLRT method of Bangdiwala (1982) seems to be the most promising. However, its drawback is that it requires a minimum sample size $n_0$ in (26) which can be quite large if $p_1 \to p_0$. It will be interesting to develop alternative efficient methods that can overcome this drawback and also hold nice properties both theoretically and numerically. Moreover, we focus on a specific bivariate normal model in (10) in this paper, and from the theoretical viewpoint, it will be interesting to develop efficient sequential tests in the problem in (1) and (2) under a more general bivariate normal model with all five parameters being unknown. Hopefully our paper can stimulate more research on evaluating a system’s precision quality.

Acknowledgments

Li’s research was supported in part by grants from the National Natural Science Foundation of China (11101156, 11271135, 11471119), the Program of Shanghai Subject Chief Scientist (14XD1401600), the 111 Project (B14019). Mei’s research was supported in part by grants from the National Natural Science Foundation of China (11101156, 11271135, 11471119), the Program of Shanghai Subject Chief Scientist (14XD1401600), the 111 Project (B14019). Mei’s research was supported in part by grants from the National Natural Science Foundation of China (11101156, 11271135, 11471119), the Program of Shanghai Subject Chief Scientist (14XD1401600), the 111 Project (B14019). Mei’s research was supported in part by grants from the National Natural Science Foundation of China (11101156, 11271135, 11471119), the Program of Shanghai Subject Chief Scientist (14XD1401600), the 111 Project (B14019).

Appendix A. Proof of Theorem 1

Proof. It is important to note that the mis-SPRT (N*, D*) in (7)–(9) is developed as a likelihood ratio test under the misspecified model (3) when $\rho$ is erroneously assumed to be 0, but it is no longer a likelihood ratio test under the model (10) with the true $\rho \neq 0$. To derive the properties of the mis-SPRT design under the null hypothesis $H_0$ of (12) when the true $\rho \neq 0$, the main mathematical tool is the change of measure technique, i.e., we need to find a pseudo hypothesis testing problem so that its likelihood ratio test is equivalent to the mis-SPRT up to a scalar constant.

For that purpose, we rewrite the complete form of the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ in (12) as

$$\tilde{H}_0 \sim \mathcal{N} \left( (X_i, Y_i) \right) \sim N \left( (0, 0), \sigma_0^2 \left( \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right)$$

where $\sigma_0$ satisfies Eq. (11) with $p = p_0$. For a given ($\tilde{\rho}$, $\tilde{\sigma}_1^2$), consider the problem of testing the null hypothesis $\tilde{H}_0$ against the alternative hypothesis in $\tilde{H}_1$ in (14). Then the corresponding log-likelihood ratio statistic is given by

$$\ln(\tilde{\lambda}_n) = \tilde{K} \sum_{i=1}^{n} (X_i^2 + Y_i^2 - \tilde{C}) + \left[ \frac{\tilde{\rho}}{\tilde{\sigma}_1^2 (1 - \tilde{\rho}^2)} - \frac{\rho}{\sigma_0^2 (1 - \rho^2)} \right] \sum_{i=1}^{n} X_i Y_i,$$

where

$$\tilde{K} = \frac{1}{2} \left[ \frac{1}{\sigma_0^2 (1 - \rho^2)} - \frac{1}{\tilde{\sigma}_1^2 (1 - \tilde{\rho}^2)} \right],$$

and

$$\tilde{C} = \ln \left( \frac{\tilde{\sigma}_1^2 (1 - \tilde{\rho}^2)}{\sigma_0^2 (1 - \rho^2)} \right) / \tilde{K}.$$

Meanwhile, for the mis-SPRT design ($N^*, D^*$) in (7)–(9), the decision statistic (in logarithm scale) is given by

$$\ln(\lambda_{n^*}) = K^* \sum_{i=1}^{n} (X_i^2 + Y_i^2 - C^*),$$

where

$$K^* = \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) > 0,$$
and
\[ C^* = \ln \left( \frac{\sigma^*_1^2}{\sigma^*_0^2} \right) / K^*. \]

A comparison between \( \ln(\hat{\lambda}_n) \) and \( \ln(\lambda^*_n) \) shows that
\[ \ln(\hat{\lambda}_n) = (\tilde{K} / K^*) \ln(\lambda^*_n) \]
as long as
\[ \frac{\tilde{\rho}}{\sigma^*_1(1 - \tilde{\rho}^2)} - \frac{\rho}{\sigma^*_0(1 - \rho^2)} = 0 \quad \text{and} \quad \tilde{C} = C^*. \quad (29) \]

Recall that \( \sigma^*_0 \) and \( \sigma^*_1 \) are defined in (5) as a function of \( p_0 \) and \( p_1 \), and thus they do not depend on the true \( \rho \neq 0 \) (though \( \sigma_0 \) does depend on \( \rho \)). Combining (5) with (29) yields (13). This explains the rationale of the definition of \( (\tilde{\rho}, \tilde{\sigma}^*_1) \) in (13).

Next, we need to show that there exists \( (\tilde{\rho}, \tilde{\sigma}^*_1) \) satisfying (13) when \( \rho \neq 0 \). Plugging the first equation of (13) into the second equation, we have a single equation as a function of \( \tilde{\rho} \),
\[ \frac{\ln(|\ln(p_0)|) - \ln(|\ln(p_1)|)}{|\ln(p_0)| - |\ln(p_1)|} = \frac{2\sigma^*_0(1 - \rho^2)}{1 - \rho/\tilde{\rho}} \ln \left( \frac{\tilde{\rho}\sqrt{1 - \rho^2}}{\rho\sqrt{1 - \tilde{\rho}^2}} \right). \quad (30) \]

Since the true \( \rho \neq 0 \), we can let \( \xi = \tilde{\rho}/\rho \in (0, 1/|\rho|) \) and then the right-hand side of (30) becomes
\[ 2\sigma^*_0(1 - \rho^2) \frac{\xi}{\xi - 1} \ln \left( \frac{\xi\sqrt{1 - \rho^2}}{\sqrt{1 - \rho^2\tilde{\rho}^2}} \right). \]

Clearly this is a continuous function of \( \xi \), as it has a finite limit of \( 2\sigma^*_0(1 - \rho^2) \) as \( \xi \to 1 \). Moreover, it converges to 0 and +\( \infty \) as \( \xi \) goes to 0 and 1/|\rho| respectively. Since the left-hand side of (30) is a positive constant that does not depend on \( \tilde{\rho} \), and the right-hand side is a continuous function of \( \tilde{\rho} \) ranging from 0 to +\( \infty \), there exists \( \tilde{\rho} \) satisfying (30). Hence, relation (13) has a solution \( (\tilde{\rho}, \tilde{\sigma}^*_1) \) when the true \( \rho \neq 0 \).

With the chosen \( (\tilde{\rho}, \tilde{\sigma}^*_1) \) satisfying (13), or equivalently, (29), we can derive the properties of the mis-SPRT design \( (N^*, D^*) \) in (7)-(9) under \( H_0 : p = p_0 \) in (1) when the true \( \rho \neq 0 \) by comparing it with the one-sided likelihood-ratio-based stopping time
\[ N^*_A = \inf\{ n \geq 1 : \ln(\lambda^*_n) > A \} \]
\[ = \inf\{ n \geq 1 : \ln(\hat{\lambda}_n) > A(\tilde{K} / K^*) \}. \]

Loosely speaking, under \( H_0 : p = p_0 \) in (1), the mis-SPRT with thresholds \( A \) and \( B \) has similar performance as its one-sided version \( N^*_A \) which turns out to be the likelihood ratio test with threshold \( A(\tilde{K} / K^*) \) when the true \( \rho \neq 0 \). To be more rigorous, for the Type I error probability,
\[ P_{H_0}(N^* \text{ rejects } H_0) \leq P_{H_0}(N^*_A < \infty) \]
\[ = P_{H_0}(N^*_A < \infty) \]
\[ = E_{\tilde{H}_1} \left\{ \exp\{-\ln(\hat{\lambda}_{N^*_A})\} ; N^*_A < \infty \right\} \]
\[ \leq \exp \left\{ -A(\tilde{K}/K^*) \right\}. \]

Here the first inequality holds because \( N^* \) stops whenever its one-sided version \( N^*_A \) stops, the second equation uses the fact that \( \tilde{H}_0 \) is just the null hypothesis \( H_0 \) in (12) written to match the notation \( \tilde{H}_1 \) in (14), the third equation is the standard change of measures argument, and the last inequality follows from the definition of \( N^*_A \).

As for the ASN of the mis-SPRT design \( (N^*, D^*) \), since \( \ln(\lambda^*_n) \) in (8) is still a random walk with increments \( \ln(\sigma^*_0^2 / \sigma^*_1^2) + \frac{1}{2} \left( \frac{1}{\sigma^*_0^2} - \frac{1}{\sigma^*_1^2} \right) (\mathcal{X}^*_n + \mathcal{Y}^*_n) \) when \( \rho \neq 0 \), it follows from the standard renewal theory that \( E_{H_0}(N^*) = (1 + o(1)) |\ln(\beta)| / I_0 \), where \( I_0 \) is the Kullback–Leibler information number under \( H_0 : p = p_0 \) defined by
\[ I_0 = E_{p_0} \left[ -\ln(\lambda^*_1) \right] = -\ln \left( \frac{\sigma^*_0^2}{\sigma^*_1^2} \right) - \left( \frac{1}{\sigma^*_0^2} - \frac{1}{\sigma^*_1^2} \right) \sigma^*_0^2. \]

Combining this with the definitions of \( \sigma^*_0^2 \) and \( \sigma^*_1^2 \) in (5) yields (17). Thus the theorem holds. \( \square \)
Appendix B. Proof of Proposition 1

Proof. To prove Proposition 1, it suffices to derive the asymptotic expression of (11) as $\rho \to 0$, and the key step is to consider the second-order Taylor series expansion of

$$\exp \left\{ \frac{\rho \sin(2\nu)}{2\sigma^2(1 - \rho^2)} \right\} = 1 + \frac{\sin(2\nu)}{2\sigma^2} \rho + \frac{1}{2} \left[ \frac{\sin(2\nu)}{2\sigma^2} \right]^2 \rho^2 + o(\rho^2)$$

$$= 1 + \frac{1}{2\sigma^2} + \frac{1}{8\sigma^4} + o(\rho^2) - \frac{1 - \rho \sin(2\nu)}{2\sigma^2} - \frac{1}{8\sigma^4} \rho^2$$

at $\rho = 0$. The second equation allows us to cancel the term $1/\{1 - \rho \sin(2\nu)\}$ in the integrand of (11) when appropriate. Plugging the above Taylor series expansions into (11) and simplifying the terms, we have

$$p = \frac{\sqrt{1 - \rho^2}}{2\pi} \exp \left\{ -\frac{1}{2\sigma^2(1 - \rho^2)} \right\} \times \left\{ \int_0^{2\pi} \frac{1}{1 - \rho \sin(2\nu)} \left[ 1 + \frac{1}{2\sigma^2} + \frac{1}{8\sigma^4} + o(\rho^2) \right] d\nu \right\}$$

$$- \frac{1}{2\sigma^2} \int_0^{2\pi} d\nu - \frac{1}{8\sigma^4} \int_0^{2\pi} [1 + \rho \sin(2\nu)] d\nu \right\}.$$ 

Now it is useful to note that

$$\int_0^{2\pi} \frac{1}{1 - \rho \sin(2\nu)} d\nu = \frac{2\pi}{\sqrt{1 - \rho^2}}.$$ 

This can be proved by using the obvious fact $1 = P(X_i^2 + Y_i^2 \geq 0)$:

$$1 = \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi \sigma^2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1 - \rho \sin(2\nu)}{2\sigma^2(1 - \rho^2)} r^2 \right\} r d\nu$$

$$= \frac{\sqrt{1 - \rho^2}}{2\pi} \int_0^{2\pi} \frac{1}{1 - \rho \sin(2\nu)} d\nu.$$ 

Combining the above results and integrating with respect to $\nu$, we have

$$p = \exp \left\{ -\frac{1}{2\sigma^2(1 - \rho^2)} \right\} \left[ 1 + \left( \frac{1}{2\sigma^2} + \frac{1}{8\sigma^4} \right) (1 - \sqrt{1 - \rho^2}) + o(\rho^2) \right].$$

or

$$\ln(p) = -\frac{1}{2\sigma^2(1 - \rho^2)} + \ln \left( 1 + \left( \frac{1}{2\sigma^2} + \frac{1}{8\sigma^4} \right) (1 - \sqrt{1 - \rho^2}) + o(\rho^2) \right).$$

Using the first-order Taylor series expansion of $\ln(1 + x) = x + o(x)$ and the fact that $1 - \sqrt{1 - \rho^2} = \rho^2/2 + o(\rho^2)$, we have

$$\ln(p) = -\frac{1}{2\sigma^2(1 - \rho^2)} + \left( \frac{1}{4\sigma^2} + \frac{1}{16\sigma^4} \right) \rho^2 + o(\rho^2).$$

Now we are ready to solve $\sigma^2$ as a function of $p$ and $\rho$, as the above relation can be rewritten as the form

$$b_0(\rho) + b_1(\rho)\sigma^2 + b_2(\rho, p)\sigma^4 = 0,$$

(31)

where

$$b_0(\rho) = -\rho^2(1 - \rho^2),$$

$$b_1(\rho) = 8 - 4\rho^2(1 - \rho^2),$$

$$b_2(\rho, p) = 16(1 - \rho^2)[\ln(p) - o(\rho^2)].$$

When $[b_1(\rho)]^2 - 4b_2(\rho, p)b_0(\rho) \geq 0$, which is exactly the assumption (18) of Proposition 1, relation (31) has two real roots

$$\sigma^2 = -b_1(\rho) \pm \sqrt{b_1^2(\rho) - 4b_2(\rho, p)b_0(\rho)}}{2b_2(\rho, p)}.$$

Note that the “+” root is unreasonable, as $\sigma^2$ would be 0 when $\rho = 0$. Thus plugging $b_0(\rho), b_1(\rho)$ and $b_2(\rho, p)$ into the “−” root, we have

$$\sigma^2 = \frac{4\rho^2 - 8 - 4\sqrt{4 + 4[\ln(p) - 1]\rho^2 + [5 - 8\ln(p)]\rho^4 + o(\rho^4) + o(\rho^2)}}{32(1 - \rho^2)[\ln(p) - o(\rho^2)]}.$$
Replacing $\rho \neq 0$, based on the proof of Theorem 1, the solution of $\tilde{\sigma}^2$ at Proposition 1 follows directly from the second-order Taylor series expansion of the right-hand side of the above equation at $\rho = 0$, completing the proof of Proposition 1. □

Appendix C. Proof of Corollary 1

Proof. First, we need to derive an explicit expression of $\sigma_0$ in the null hypothesis of (12) in terms of $\rho$ and $(p_0, p_1)$, as the true $\rho \to 0$. By Proposition 1, we have

$$\sigma_0^2 = \frac{1}{2|\ln(p_0)|} + \frac{2 + \ln(p_0)}{8|\ln(p_0)|} \rho^2 + o(\rho^2). \quad (32)$$

Next, for fixed $(p_0, p_1)$, we need to solve $(\tilde{\rho}, \tilde{\sigma}_1^2)$ from (13) as $\rho \to 0$. Combining two equations in (13) to cancel $\tilde{\sigma}_1^2$, we derive a single equation for $\tilde{\rho}$:

$$\frac{\ln(|\ln(p_0)|) - \ln(|\ln(p_1)|)}{2\sigma_0^2(1 - \rho^2) [|\ln(p_0)| - |\ln(p_1)|]} = \ln \left( \frac{\rho \sqrt{1 - \rho^2}}{\rho \sqrt{1 - \rho^2}} \right) / (1 - \rho / \tilde{\rho}). \quad (33)$$

Based on the proof of Theorem 1, the solution of $\tilde{\rho}$ exists and it is better to rewrite (33) in term of $\tilde{\xi} = \tilde{\rho} / \rho$. For any fixed $\rho \neq 0$, define a constant $C_\rho = \frac{\ln(|\ln(p_0)|) - \ln(|\ln(p_1)|)}{2\sigma_0^2(1 - \rho^2) [|\ln(p_0)| - |\ln(p_1)|]}$ and a function

$$h_\rho(\xi) = \frac{\xi}{\xi - 1} \left[ \ln(\xi) + \frac{1}{2} \ln \left( \frac{1 - \rho^2}{1 - \rho^2 \xi^2} \right) \right].$$

Then relation (33) becomes $C_\rho = h_\rho(\xi)$.

To solve $\tilde{\xi} = \tilde{\xi}(\rho)$ from this relation, we need to consider the Taylor series expansion of $h_\rho(\xi)$ for $\xi$ at some suitable $\tilde{\xi} = \tilde{\xi}_0$ as $\rho \to 0$. To find such $\tilde{\xi}_0$, it is natural to consider the relation $\lim_{\rho \to 0} C_\rho = \lim_{\rho \to 0} h_\rho(\xi)$. Note that as $\rho \to 0$, by (32), the constant $C_\rho$ converges to

$$\ln \left( \frac{|\ln(p_0)|}{|\ln(p_1)|} \right) / (1 - |\ln(p_1)| / |\ln(p_0)|),$$

whereas $h_\rho(\xi)$ converges to $\ln(\xi) / (1 - 1/\xi)$ for fixed $\xi$. Equating these two limits yields that

$$\tilde{\xi}_0 = \frac{|\ln(p_0)|}{|\ln(p_1)|}.$$ 

Now we consider the first-order Taylor series expansion of $h_\rho(\xi)$ at $\tilde{\xi}_0$, we have

$$C_\rho = h_\rho(\tilde{\xi}_0) \approx h_\rho(\tilde{\xi}_0) + h_\rho'(\tilde{\xi}_0)(\xi - \tilde{\xi}_0),$$

and thus

$$\tilde{\xi} \approx \tilde{\xi}_0 + \frac{[C_\rho - h_\rho(\tilde{\xi}_0)] / h_\rho'(\tilde{\xi}_0)}{2\sigma_0^2(1 - \rho^2) |\ln(p_1)|} - (\tilde{\xi}_0 - 1) \left[ \frac{\xi_0 \ln(\xi_0)}{2} + \frac{\xi_0}{2} \ln \left( \frac{1 - \rho^2}{1 - \rho^2 \xi_0^2} \right) \right].$$

Replacing $\sigma_0^2$ by (32) and through the second-order Taylor series expansion at $\rho = 0$, we have

$$\tilde{\xi} \approx \tilde{\xi}_0 + \frac{\xi_0 (\tilde{\xi}_0 - 1) \ln(\xi_0) [2 - \ln(p_0)] - 2\xi_0 (\tilde{\xi}_0 - 1) (\tilde{\xi}_0^2 - 1)}{4[\xi_0 - \ln(\xi_0) - 1]} \rho^2 + o(\rho^2). \quad (34)$$

Finally, we are ready to prove Corollary 1 by presenting the results of Theorem 1 in term of $\rho$ and $(p_0, p_1)$. The asymptotic expression of the $\psi_0(\rho)$ function follows by plugging $\tilde{\xi}$ in (34) and $\sigma_0$ in (32) into the right-hand side of (15) in Theorem 1 and then considering the corresponding Taylor series expansion of $\rho$. Combining $\sigma_0$ in (32), $l_0$ in (17), and Taylor series expansion, we can easily derive the asymptotic expression of $\omega_0(\rho) = 1/l_0$, and thus Corollary 1 holds. □
References


