Optimal Robust Sequential Tests of Circular Nonconforming Probability

Qunzhi Xu and Yajun Mei

Department of Biostatistics

School of Global Public Health, New York University

Supplementary Material

S1 Proof of lemma 1

For the sake of being self-contained, let us first restate both the functional maximum likelihood estimator (MLE) problem and Lemma 1 in the main body of the paper.

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

$$L_p(f) = \prod_{i=1}^{n} f(x_i, y_i) \tag{S1.1}$$

subject to the constraints

$$f(u_i, v_i) \geq 0, \quad \sum_{i=1}^{\infty} f(u_i, v_i) = 1, \text{ and } \sum_{i=1}^{\infty} I(u_i^2 + v_i^2 > r^2) f(u_i, v_i) = p. \tag{S1.2}$$

Lemma 1 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 (S1.1)-(S1.2), the maximum value of likelihood function is given by*

$$L_p^* = \sup_f L_p(f) = \sup_f \prod_{i=1}^{n} f(x_i, y_i) = \big(\frac{1-p}{n-m}\big)^{n-m} \big(\frac{p}{m}\big)^m, \tag{S1.3}$$

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

We are now provide a rigorous proof to Lemma 1. In the functional MLE problem in (S1.1)-(S1.2), note that the likelihood function $L_p(f) = 0$ if at least one of $f(x_i, y_i) = 0$, and thus observed data (x_i, y_i)'s must be on the support of the optimal probability mass function f. Without loss of generality, we order the observed data (x_i, y_i) from largest to smallest based on the values of $x_i^2 + y_i^2$ (random ordering if there are ties), and assume $(u_i, v_i) = (x_i, y_i)$ for $i = 1 \cdots, n$. By (S1.1) and the definition of $m = \sum_{i=1}^{n} I(x_i^2 + y_i^2 > r^2)$, we have

$$\sum_{i=1}^{m} f(x_i, y_i) + \sum_{j=n+1}^{\infty} I(u_j^2 + v_j^2 > r^2) f(u_j, v_j) = p$$

$$\sum_{i=m+1}^{n} f(x_i, y_i) + \sum_{j=n+1}^{\infty} I(u_j^2 + v_j^2 \leq r^2) f(u_j, v_j) = 1 - p. \quad \text{(S1.4)}$$

Let $w_i = f(x_i, y_i)$ for $i = 1 \cdots, n$. Maximizing (S1.1) is equivalent to maximizing two sub-problems simultaneously:

$$\begin{aligned} &\max \prod_{i=1}^{m} w_i \quad \text{subject to} \quad \sum_{i=1}^{m} w_i \leq p, \quad \min_{1 \leq i \leq m} w_i \geq 0, \\ &\max \prod_{i=m+1}^{n} w_i \quad \text{subject to} \quad \sum_{i=m+1}^{n} w_i \leq 1 - p, \quad \min_{m+1 \leq i \leq n} w_i \geq 0. \end{aligned} \quad \text{(S1.5)}$$

We claim that the optimal solutions for (S1.5) are $w_i = p/m$ for $i = 1, \cdots, m$ and $w_i = (1-p)/(n-m)$ for $i = m+1, \cdots, n$.

To see this, let us focus on the case of $i = 1, \cdots, m$, and we first show that the optimized values w_i must be identical. Assume a pair w_i and w_j were different. Define $w_i' = w_j' = (w_i + w_j)/2$. Then $w_i' + w_j' = w_i + w_j$ but $w_i' w_j' - w_i w_j = (w_i - w_j)^2/4 > 0$. Thus if we replace w_i and w_j by $w_i' = w_j' = (w_i + w_j)/2$, we maintain the overall sum and increase the overall product. Hence, the product must be maximized when all w_i's are equal to w for $i = 1 \cdots, m$.

Next, for a given integer $m \in [0, n]$, the first optimization problem in (S1.5) becomes the problem of maximizing w^m subject to $0 \leq w \leq p/m$, which is attained at $w = p/m$. This shows that the optimized $w_i = p/m$ for $i = 1, \cdots, m$. Likewise, for the second optimization problem in (S1.5), the optimized $w_i = (1-p)/(n-m)$ for $i = m+1, \cdots, n$. Combining these results together yields (S1.3), completing the proof of Lemma 1.

S2 Proof of Theorem 1

For completeness, we restate the notation and results of Theorem 1 in the main body of the paper. For the bivariate data (X_i, Y_i) for $i = 1, \cdots, n$ with distribution f, when testing $H_0 : f = f_0 \in \Omega_0$ against $H_1 : f = f_1 \in \Omega_1$, the generalized likelihood ratio (GLR) test statistic is defined as

$$G_n = \frac{\sup_{f_1 \in \Omega_1} \prod_{i=1}^n f_1(X_i, Y_i)}{\sup_{f_0 \in \Omega_0} \prod_{i=1}^n f_0(X_i, Y_i)}. \tag{S2.1}$$

Meanwhile, define the binary quantizer variable

$$Z_i = \mathbf{1}\{X_i^2 + Y_i^2 > r^2\}, \quad i = 1, \ldots, n. \tag{S2.2}$$

and the corresponding the likelihood ratio test statistic is given by

$$L_n = \prod_{i=1}^n \Big(\frac{1-p_1}{1-p_0}\Big)^{1-Z_i} \Big(\frac{p_1}{p_0}\Big)^{Z_i}. \tag{S2.3}$$

Theorem 1 asserts that $G_n = L_n$. By Lemma 1, we have proved Theorem 1 under the discrete case in the main body of the paper. To be more specific, by Lemma 1, the GLR statistic G_n in (S2.1),

$$\begin{aligned} G_n &= \frac{L^*_{p_1}}{L^*_{p_0}} = \big(\frac{1-p_1}{n-m}\big)^{n-m} \big(\frac{p_1}{m}\big)^m \big(\frac{n-m}{1-p_0}\big)^{n-m} \big(\frac{m}{p_0}\big)^m \\ &= \big(\frac{1-p_1}{1-p_0}\big)^{n-m} \big(\frac{p_1}{p_0}\big)^m, \end{aligned}$$

which is the same as L_n in (S2.3), 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 (S2.2).

below we provide a detailed proof of Theorem 1 under the continuous case with pdf $f(u, v)$. The key idea is to add a small neighborhood of (x_i, y_i) and then apply Lemma 1 to the discrete probabilities of those small neighborhoods. While the areas of those small neighborhoods affect the maximum likelihood functions $L^*_{p_0}(f)$ and $L^*_{p_1}(f)$ themselves, they are canceled out in the ratios, and thus do not affect the GLR statistic G_n in (S2.1). To be more concrete, denote by $B_\epsilon(x_i, y_i)$ a small neighborhood of (x_i, y_i) and denoted its area by $\mu(B_\epsilon(x_i, y_i))$. By the definition of the pdf, we have

$$f(x_i, y_i) = \lim_{\epsilon \to 0} \frac{\mathbf{P}_f((u, v) \in B_\epsilon(x_i, y_i))}{\mu(B_\epsilon(x_i, y_i))}.$$

and thus

$$\mathcal{L}_p(f) = \prod_{i=1}^n f(x_i, y_i) = \lim_{\epsilon \to 0} \frac{\prod_{i=1}^n \mathbf{P}_f((u, v) \in B_\epsilon(x_i, y_i))}{\prod_{i=1}^n \mu(B_\epsilon(x_i, y_i))}.$$

The results follow directly by applying Lemma 1 to the corresponding discrete likelihood

$$\mathcal{L}_p(f_\epsilon) = \prod_{i=1}^n \mathbf{P}_f((u, v) \in B_\epsilon(x_i, y_i)).$$

since the term $\prod_{i=1}^n \mu(B_\epsilon(x_i, y_i))$ does not affect ratios. Of course, to be more rigorous in the application of Lemma 1, the neighborhood $B_\epsilon(x_i, y_i)$

should be chosen to satisfy (1) $u^2 + v^2 \leq r^2$ for any $(u, v) \in B_\epsilon(x_i, y_i)$ when $x_i^2 + y_i^2 \leq r^2$, and (2) $u^2 + v^2 > r^2$ for any $(u, v) \in B_\epsilon(x_i, y_i)$ when $x_i^2 + y_i^2 > r^2$. This completes the proof of Theorem 1.