Sub-uniformity of harmonic mean p-values

Yuyu Chen[*] Ruodu Wang[†] Yuming Wang[‡] Wenhao Zhu[§]

April 17, 2025

Abstract

We obtain several inequalities on the generalized means of dependent p-values. In particular, the weighted harmonic mean of p-values is strictly sub-uniform under several dependence assumptions of p-values, including independence, negative upper orthant dependence, the class of extremal mixture copulas, and some Clayton copulas. Sub-uniformity of the harmonic mean of p-values has an important implication in multiple hypothesis testing: It is statistically invalid (anti-conservative) to merge p-values using the harmonic mean unless a proper threshold or multiplier adjustment is used, and this applies across all significance levels. The required multiplier adjustment on the harmonic mean p-value grows sub-linearly to infinity as the number of p-values increases, and hence there does not exist a constant multiplier that works for any number of p-values, even under independence.

Keywords: Multiple hypothesis testing; merging function; p-value; stochastic order; negative dependence; Clayton copula.

1 Preamble: Connection between merging p-values and risk aggregation

This paper studies merging p-values via averaging in the context of testing a global hypothesis. We first explain the connection between two areas: merging p-values and risk aggregation, before a formal introduction of the research problems and our contributions.

The combined p-values of many merging methods take the following form

$$M_g(U_1,\dots,U_n) = g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(U_i)\right),$$

[*]Department of Economics, University of Melbourne, Australia. ✉ yuyu.chen@unimelb.edu.au

[†]Department of Statistics and Actuarial Science, University of Waterloo, Canada. ✉ wang@uwaterloo.ca

[‡]School of Mathematical Sciences, Peking University, China. ✉ wangyuming@pku.edu.cn

[§]School of Mathematical Sciences, Peking University, China. ✉ zwhao92@163.com

where the random variables $U_1, \dots, U_n$ represent possibly dependent p-values and $g : [0,1] \to [-\infty, \infty]$ is a strictly monotone and continuous function. For instance, the harmonic mean merging method (Wilson, 2019) corresponds to choosing g as the mapping $p \mapsto 1/p$ on $(0,1)$. As $M_g(U_1, \dots, U_n)$ may not be a valid p-value, i.e., $\mathbb{P}(M_g(U_1, \dots, U_n) \le p) \le p$ may not hold for some significance level $p \in (0,1)$, it is imperative to determine a valid threshold, denoted by t_p, such that $\mathbb{P}(M_g(U_1, \dots, U_n) \le t_p) \le p$.

If g is increasing, we can rewrite $\mathbb{P}(M_g(U_1, \dots, U_n) \le t_p)$ as $\mathbb{P}(\sum_{i=1}^n g(U_i) \le ng(t_p))$; the case with decreasing g is similar. Hence, the problem of deciding statistically valid thresholds t_p for the combined p-value can be converted into the computation of

$$\text{the quantile function of the test statistic } g(U_1) + \cdots + g(U_n) \text{ or bounds on it.} \tag{1}$$

Although bearing very different motivations, problem (1) has a prominent presence in Quantitative Risk Management (QRM), known as the problem of risk aggregation, which concerns quantitative assessments, such as a quantile or a risk measure, of the sum of random variables representing financial losses. We refer to McNeil et al. (2015, Chapter 8) for an introduction to risk aggregation. This connection allows us to utilize many existing results and techniques developed in QRM to study the problem of merging p-values. There are three connected streams of research on risk aggregation in QRM that are particularly relevant to merging p-values.

(a) The first one is Extreme Value Theory (EVT) in risk aggregation, because each $g(U_i)$ in (1) typically has a heavy-tailed, often even infinite-mean, distribution (as in the case of the harmonic mean merging method). For instance, Embrechts et al. (2009) obtained results on the asymptotic behaviour of the quantile of the aggregate risk via EVT, thus computing (1) in an asymptotic sense, and this helps us to understand the limiting behaviour of the combined p-value as the significance level goes to 0. We refer to Embrechts et al. (1997) for a comprehensive treatment of EVT in finance and insurance. The case of our main interest, the harmonic mean of uniform random variables, is discussed in Embrechts et al. (2002, Example 7) for $n = 2$.

(b) The second one is dependence modeling for risk aggregation with copulas. Copula models are very popular in the statistical and probabilistic modeling of aggregate risks, and they help to analyze merging p-values in (1) with certain classes of dependence structures. We refer to Frees and Valdez (1998), Denuit et al. (1999), Chavez-Demoulin et al. (2006), and Embrechts (2009) for using copulas in risk aggregation in various contexts of risk management. The textbook of McNeil et al. (2015) provides a general treatment.

(c) The third one is quantile bounds on dependence uncertainty in risk aggregation. In the context of merging p-values, this corresponds to the problem of averaging arbitrarily dependent p-values. For instance, results in Embrechts and Puccetti (2006), Wang et al. (2013), Embrechts et al. (2015), and Bignozzi et al. (2016) on dependence uncertainty, as well as the algorithm of Embrechts et al. (2013), are essential to the generalized mean of p-values under arbitrary dependence among p-values. We refer to Embrechts et al. (2014) for a summary.

The above selected references are subjective and may be slightly biased towards the work of Paul Embrechts, who has been a key figure in all streams of research above, not only with tremendous research contributions but also with remarkable mentorship for many younger scholars in the field.

This paper continues to enhance the connection between the two fields by studying merging p-values, especially the harmonic mean of p-values, through recent developments in risk modeling. Some of our results are either directly built on or inspired by the work of Embrechts.

2 Introduction

In multiple testing of a single hypothesis and testing multiple hypotheses, a decision maker often needs to combine several p-values into one p-value. Recently, Wilson (2019) proposed a statistical procedure based on the harmonic mean of p-values, which belongs to the larger class of generalized mean p-values studied by Vovk and Wang (2020). The class of generalized mean p-values also includes Fisher's combination method (Fisher, 1948) via the geometric mean, often applied under the assumption that p-values are independent. The harmonic mean p-value method of Wilson (2019) has some desirable properties such as being applicable under a wide range of dependence assumptions of p-values, and has received considerable attention in statistics and the natural sciences. Validity, admissibility, and threshold adjustments of the generalized mean methods for p-values with arbitrary dependence are studied further by Vovk et al. (2022) and Chen et al. (2023).

We say that a combined p-value, denoted by P, is conservative/valid (resp. anti-conservative/invalid) if $\mathbb{P}(P \le p) \le p$ for all $p \in (0,1)$ (resp. $\mathbb{P}(P \le p) > p$ for some $p \in (0,1)$); in practice one may only be interested in a specific value of p such as 0.01 or 0.05. Here, $\mathbb{P}$ is a fixed probability under the null hypothesis of interest. The harmonic mean of p-values, also called the harmonic mean p-value,[1] is known to be anti-conservative under some dependence structures, as noted by Wilson (2019). If the underlying dependence structure is arbitrary, a threshold correction of order $\log n$

[1] The harmonic mean p-value method usually refers to the statistical procedure proposed by Wilson (2019).

is needed, where n is the number of p-values to merge (Vovk and Wang, 2020). This correction generally leads to very conservative tests (i.e., the Type-I error rates are very small), and it may be reduced or even omitted under some specific dependence assumptions. In this paper, we study a stochastic order relation between a weighted generalized mean of standard uniform p-values and a standard uniform p-value under several dependence assumptions, and discuss its implications for the validity and threshold adjustment for harmonic mean p-values.

Let $\Delta_n = \{(w_1, \dots, w_n) \in [0,1]^n : w_1 + \dots + w_n = 1\}$ be the unit n-simplex. We always assume $n \geq 2$. For $r \in \mathbb{R} \setminus \{0\}$, $n \in \mathbb{N}$, and $\mathbf{w} = (w_1, \dots, w_n) \in \Delta_n$, the (weighted) r-mean function is defined as

$$M_r^{\mathbf{w}}(u_1, \dots, u_n) = (w_1 u_1^r + \dots + w_n u_n^r)^{1/r}, \qquad (u_1, \dots, u_n) \in (0, \infty)^n.$$

The 0-mean function is the weighted geometric mean, that is, $M_0^{\mathbf{w}}(u_1, \dots, u_n) = \prod_{i=1}^n u_i^{w_i}$, which is also the limit of $M_r^{\mathbf{w}}$ as $r \to 0$. The (-1)-mean function is referred to as the *harmonic mean.* If $w_1 = \dots = w_n = 1/n$, $M_r^{\mathbf{w}}$ is the symmetric r-mean function, denoted by M_r for simplicity, and it is defined on $\bigcup_{n \in \mathbb{N}} (0, \infty)^n$. Denote by $\Delta_n^+ = \Delta_n \cap (0,1)^n$. For some of our results, we will only consider $\mathbf{w} \in \Delta_n^+$ since if some components of $\mathbf{w}$ are zero, we can simply reduce the dimension of $M_r^{\mathbf{w}}$.

Throughout, $U_1, \dots, U_n$ are (standard) uniform random variables on $(0,1)$ that are possibly dependent, and they represent p-values to combine. The quantity $M_r^{\mathbf{w}}(U_1, \dots, U_n)$ is the weighted r-mean of p-values. For two random variables X and Y, we say X is less than Y in stochastic order, denoted by $X \preceq_{\mathrm{st}} Y$, if $\mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x)$ for all $x \in \mathbb{R}$. Moreover, we write $X \simeq_{\mathrm{st}} Y$ if X and Y have the same distribution. The main results in this paper concern the following inequality under several dependence assumptions of $U_1, \dots, U_n$

$$M_r^{\mathbf{w}}(U_1, \dots, U_n) \preceq_{\mathrm{st}} U_1, \tag{2}$$

where $r \leq -1$. Relation (2) is quite strong as it requires $\mathbb{P}(M_r^{\mathbf{w}}(U_1, \dots, U_n) \leq p) \geq p$ to hold for all threshold levels $p \in (0,1)$. Note that (2) cannot hold for $r > -1$ except for identical $U_1, \dots, U_n$ (see Proposition 1).

A non-negative random variable X is said to be *sub-uniform* if $X \preceq_{\mathrm{st}} U_1$. Moreover, X is *strictly sub-uniform* if

$$\mathbb{P}(X \leq p) > p \quad \text{for all } p \in (0,1). \tag{3}$$

Using a sub-uniform p-value is anti-conservative in hypothesis testing, since it has a larger type-I error rate than the nominal level. Therefore, if (2) holds true, then merging p-values using the harmonic mean, or any r-generalized mean function with $r \leq -1$, is anti-conservative across all significance levels in $(0, 1)$.

Remark 1 (Terminology). Although sub-uniformity is an important property for studying p-values, this term has been used with different meanings in the literature. Some of them are collected here. A non-negative random variable X is called super-uniform by Barber and Ramdas (2017) if $U_1 \preceq_{\mathrm{st}} X$ (anticipating that sub-uniformity should be defined by flipping the above inequality, their terminology is consistent with ours), but such a random variable is called sub-uniform by Ferreira and Zwinderman (2006). Moreover, Chen and Sarkar (2020) defined sub-uniformity in the strict sense (3). Rubin-Delanchy et al. (2019) called X sub-uniform if it is dominated by U_1 in convex order.[2]

When mentioning (strict) sub-uniformity later in this paper, we always refer to the corresponding property of $M_r^{\mathbf{w}}(U_1, \ldots, U_n)$ or $M_r(U_1, \ldots, U_n)$ with $r \leq -1$, which will be clear from the context.

The main objective of this paper is to study (2) given several dependence assumptions of $U_1, \ldots, U_n$, including negative upper orthant dependence (Proposition 4), the class of extremal mixture copulas (Theorem 1), and some Clayton copulas (Theorem 2). Some of our results are built on a recent study of Chen and Shneer (2024), where a stochastic ordering inequality on heavy-tailed random variables is established. As discrete p-values may arise in hypothesis testing (e.g., Vovk et al. (2005)), we also study sub-uniformity for discrete uniform random variables on $\{1/m, \ldots, m/m\}$ for $m \in \mathbb{N}$. This situation is quite different from the uniform case as (2) can never hold for $\mathbf{w} \in \Delta_n^+$ in the case of discrete uniform random variables unless they are identical. However, using the harmonic mean function to merge discrete uniform random variables can still be anti-conservative at some threshold levels (Theorem 3).

Most findings of this paper are negative results: Under many different assumptions of dependence among p-values, the harmonic mean p-value is anti-conservative and cannot be used without a proper adjustment, and the adjustment coefficient diverges even in the case of independence. To address these issues, while keeping the advantages of the harmonic mean p-value, it is recommendable to use the Simes method (Simes, 1986) or the Cauchy combination method (Liu and Xie, 2020), which are shown to be valid under various forms of dependence, and perform comparably to the harmonic mean p-value; see results in Chen et al. (2023) on comparing these three methods.

[2] A random variable X is said to be dominated by Y in convex order if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ holds for all convex functions f, provided that the expectations exist.

Other methods based on heavy-tailed transformation of p-values can also be used under different assumptions (Gui et al., 2023). An exception to the above negative results is the case of Clayton copulas treated in Theorem 2, where we obtain a positive result that the harmonic mean p-value can be made valid with a small threshold adjustment (a multiplicative factor of 1.131) under the assumption of Clayton copulas with parameter at least 1.

The rest of the paper is organized as follows. In Section 3, we first discuss the intuition behind (2) in the simplest case, and then present general properties related to sub-uniformity for dependent $U_1, \dots, U_n$. In Section 4, (2) is shown given the aforementioned dependence assumptions of $U_1, \dots, U_n$. In Section 5, the threshold of the harmonic mean p-value is studied for independent p-values, where we see that the adjustment increases at the rate of $\log n$ for a fixed probability level. Sub-uniformity for discrete uniform random variables is studied in Section 6. Numerical examples based on simple simulations are presented in Section 7, and Section 8 concludes the paper.

3 Some observations and general results on sub-uniformity

We set straight some first observations on the sub-uniformity inequality (2). First, the generalized mean is monotone in r; that is, given any $\mathbf{w} \in \Delta_n$, $M_r^{\mathbf{w}} \le M_s^{\mathbf{w}}$ for $r \le s$ (Theorem 16 of Hardy et al. (1934)). Hence, we have $M_r^{\mathbf{w}}(U_1, \dots, U_n) \le M_{-1}^{\mathbf{w}}(U_1, \dots, U_n)$ for all $r \le -1$. To get the sub-uniformity inequality (2) for all $r \le -1$, it suffices to show

$$\mathbb{P}(M_{-1}^{\mathbf{w}}(U_1, \dots, U_n) \le p) \ge p \text{ for all } p \in (0, 1). \tag{4}$$

This observation simplifies our journey by allowing us to focus on the case of the harmonic mean, which also happens to be the most popular method within the class of generalized mean methods for $r \in (-\infty, 0)$.

We begin with a simple proof for independent $U_1, \dots, U_n$ in the symmetric case. Although (4) in this case directly follows from Theorem 1 of Chen et al. (2024) or Theorem 1 of Chen and Shneer (2024), the proof below, different from Chen et al. (2024) and Chen and Shneer (2024), helps to understand (4) via a result well-known in the multiple testing literature, namely the Simes inequality (Simes, 1986). For $n \in \mathbb{N}$, write $[n] = \{1, \dots, n\}$. For $(u_1, \dots, u_n) \in (0, \infty)^n$, the Simes function is defined as $S(u_1, \dots, u_n) = \min_{i \in [n]}\{nu_{(i)}/i\}$ where $u_{(1)}, \dots, u_{(n)}$ are the order statistics of $u_1, \dots, u_n$, from the smallest to the largest. As shown by Simes (1986), $S(U_1, \dots, U_n)$ is uniformly distributed on $(0, 1)$ given independent $U_1, \dots, U_n$. Moreover, we have $M_{-1} \le S$; this inequality is in Theorem 3 of Chen et al. (2023) and was also discussed in Section 4 of the online appendix of

Wilson (2019), but one can also check it directly. Putting these two observations together, we have

$$M_{-1}(U_1, \dots, U_n) \le S(U_1, \dots, U_n) \simeq_{\mathrm{st}} U_1,$$

which implies that (2) holds for symmetric mean of independent $U_1, \dots, U_n$.

Before showing (2) holds under more general assumptions, we discuss several properties related to the target problem. We first explain that it is only meaningful to consider the sub-uniformity (2) for $r \le -1$. Indeed, as illustrated in Proposition 1 below, sub-uniformity can hold for some $r > -1$ only in the trivial case that $U_1, \dots, U_n$ are identical, and thus strict sub-uniformity can never hold for any $r > -1$.

Proposition 1. *The following statements are equivalent.*

(i) $M_r^{\mathbf{w}}(U_1, \dots, U_n) \preceq_{\mathrm{st}} U_1$ *for some* $r \in (-1, \infty)$ *and* $\mathbf{w} \in \Delta_n^+$*;*

(ii) $M_r^{\mathbf{w}}(U_1, \dots, U_n) \simeq_{\mathrm{st}} U_1$ *for some* $r \in (-1, \infty)$ *and* $\mathbf{w} \in \Delta_n^+$*;*

(iii) $U_1 = \cdots = U_n$ *a.s.;*

(iv) $M_r^{\mathbf{w}}(U_1, \dots, U_n) = U_1$ *a.s. for all* $r \in \mathbb{R}$ *and all* $\mathbf{w} \in \Delta_n$*;*

(v) $M_r^{\mathbf{w}}(U_1, \dots, U_n) = U_1$ *a.s. for some* $r \in (-1, \infty)$ *and* $\mathbf{w} \in \Delta_n^+$*.*

Proof. Note that the binary relation $X \preceq_{\mathrm{st}} Y$ is flipped under decreasing transformation on both X and Y. Hence, for $r < 0$, we can write $M_r^{\mathbf{w}}(U_1, \dots, U_n) \preceq_{\mathrm{st}} U_1$ as

$$\sum_{i=1}^{n} w_i U_i^r \succeq_{\mathrm{st}} U_1^r.$$

The case $r \ge 0$ can be argued similarly, and we will omit it from the discussion below.

<u>(i)⇒(ii)</u>: If $\sum_{i=1}^n w_i U_i^r$ and U_1^r are not identically distributed, then

$$\mathbb{E}\left[\sum_{i=1}^{n} w_i U_i^r\right] > \mathbb{E}\left[U_1^r\right],$$

which leads to a contradiction, since both expectations are equal to $1/(r+1)$. Hence, we have $\sum_{i=1}^n w_i U_i^r \simeq_{\mathrm{st}} U_1^r$.

<u>(ii)⇒(iii)</u>: Let $U = M_r^{\mathbf{w}}(U_1, \dots, U_n)$, which is uniformly distributed on $[0, 1]$. Note that

$$\sum_{i=1}^{n} w_i U_i^r = U^r. \tag{5}$$

Take $\epsilon \in (0, 1+r)$. From (5), we have

$$\mathbb{E}\left[\sum_{i=1}^{n} w_i U_i^r U^{-\epsilon}\right] = \mathbb{E}\left[U^{r-\epsilon}\right] = \frac{1}{1+r-\epsilon}. \tag{6}$$

For all $i \in [n]$, by Lemma 7.27 of McNeil et al. (2015) and the Fréchet-Hoeffding inequality, we have $\mathbb{E}[U_i^r U^{-\epsilon}] \leq \mathbb{E}[U^{r-\epsilon}]$, and the equality holds if and only if U_i^r and $U^{-\epsilon}$ are comonotonic. Together with (6), we obtain $\mathbb{E}[U_i^r U^{-\epsilon}] = \mathbb{E}[U^{r-\epsilon}]$ and thus U_i^r and $U^{-\epsilon}$ are comonotonic. Moreover, as U_i and U are identically distributed and $r < 0$, $U_i = U$ a.s. for all $i \in [n]$.

The remaining implications, (iii)$\Rightarrow$(iv) $\Rightarrow$(v) $\Rightarrow$(i), are straightforward by definition. $\square$

Joint distributions of standard uniform random variables are known as copulas; see Nelsen (2006) for an introduction to copulas. Next, we explain how to construct copulas for which (2) holds. We write $X \overset{\mathrm{d}}{\sim} F$ if a random variable or random vector X is distributed as F. Fix $r \leq -1$. In what follows, we say sub-uniformity holds for a copula C (or a random vector $(U_1, \dots, U_n)$), if $M_r^{\mathbf{w}}(U_1, \dots, U_n) \preceq_{\mathrm{st}} U_1$ holds for $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim} C$ and all $\mathbf{w} \in \Delta_n$; by saying that strict sub-uniformity holds for a copula C (or a random vector $(U_1, \dots, U_n)$), we mean that

$$\mathbb{P}(M_r^{\mathbf{w}}(U_1, \dots, U_n) \leq p) > p \text{ for all } p \in (0,1),$$

holds for $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim} C$ and all $\mathbf{w} \in \Delta_n^+$. Let $\mathcal{C}_k$ be the set of all k-dimensional copulas, $k \in \mathbb{N}$.

Proposition 2. *Let $r \leq -1$ and $\mathcal{C} \subseteq \mathcal{C}_n$. If sub-uniformity holds for each copula in $\mathcal{C}$, then it holds for any copula in the convex hull of $\mathcal{C}$.*

Proof. Note that for all $p \in (0,1)$ and any $\mathbf{w} \in \Delta_n$, $\mathbb{P}(M_r^{\mathbf{w}}(U_1, \dots, U_n) \leq p)$ is linear in the distribution of $(U_1, \dots, U_n)$. Since sub-uniformity holds for every element in $\mathcal{C}$, it also holds for every element from the convex hull of $\mathcal{C}$. $\square$

The following proposition shows that sub-uniformity can be passed from smaller groups to a larger group of p-values in two different ways. In what follows, for vectors $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$, their dot product is $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i$, and we denote by $\|\mathbf{x}\| = \sum_{i=1}^n |x_i|$ and $\mathbf{x}^{-1} = (x_1^{-1}, \cdots, x_n^{-1})$. Moreover, $\bigwedge \mathbf{x} = \min\{x_1, \dots, x_n\}$.

Proposition 3. *Let $r \leq -1$, and $C_i \in \mathcal{C}_{k_i}$ with $k_i \in \mathbb{N}$ and $i \in [n]$.*

(i) If sub-uniformity holds for C_i for all $i \in [n]$, then it holds for $C(\mathbf{u}_1, \dots, \mathbf{u}_n) = \prod_{i=1}^n C_i(\mathbf{u}_i)$, $\mathbf{u}_i \in \mathbb{R}^{k_i}$, $i \in [n]$.

(ii) Suppose that $C_i(\mathbf{u}_i) = \bigwedge \mathbf{u}_i$, $i \in [n]$. If sub-uniformity holds for $C^* \in \mathcal{C}_n$, then it holds for $C(\mathbf{u}_1, \dots, \mathbf{u}_n) = C^*(C_1(\mathbf{u}_1), \dots, C_n(\mathbf{u}_n))$, $\mathbf{u}_i \in \mathbb{R}^{k_i}$, $i \in [n]$.

Proof. Suppose that $(\mathbf{U}_1, \dots, \mathbf{U}_n) \overset{\mathrm{d}}{\sim} C$ such that $\mathbf{U}_i \overset{\mathrm{d}}{\sim} C_i$, $i \in [n]$.

(i) Let $\mathbf{w}_i \in [0,1]^{k_i}$, $i \in [n]$, such that $(\mathbf{w}_1, \dots, \mathbf{w}_n) \in \Delta_{\sum_{i=1}^n k_i}$. Let $U, V_1, \dots, V_n$ be iid standard uniform random variables. Since sub-uniformity holds for C_i, we have $\mathbf{w}_i \cdot \mathbf{U}_i^{-1} \succeq_{\mathrm{st}} \|\mathbf{w}_i\| V_i^{-1}$ for all $i \in [n]$. As $C(\mathbf{u}_1, \dots, \mathbf{u}_n) = \prod_{i=1}^n C_i(\mathbf{u}_i)$, $\mathbf{U}_1, \dots, \mathbf{U}_n$ are independent of each other. As stochastic order is preserved under convolution (e.g., Theorem 1.A.3 in Shaked and Shanthikumar (2007)), $\sum_{i=1}^n \mathbf{w}_i \cdot \mathbf{U}_i^{-1} \succeq_{st} \sum_{i=1}^n \|\mathbf{w}_i\| V_i^{-1}$. Moreover, by Theorem 1 in Chen et al. (2024), $\sum_{i=1}^n \|\mathbf{w}_i\| V_i^{-1} \succeq_{st} U^{-1}$. Consequently, we have the desired result as follows

$$\sum_{i=1}^n \mathbf{w}_i \cdot \mathbf{U}_i^{-1} \succeq_{st} \sum_{i=1}^n \|\mathbf{w}_i\| V_i^{-1} \succeq_{st} U^{-1}.$$

(ii) Since $C_i(\mathbf{u}_i) = \bigwedge \mathbf{u}_i$, the components of $\mathbf{U}_i$, $i \in [n]$, are perfectly positively dependent (i.e., they are almost surely equal as they follow the same distribution). The desired result is obvious as sub-uniformity holds for C^*. □

By Proposition 3 (i), if sub-uniformity holds for independent subgroups of standard uniform random variables, it also holds for the whole group. Proposition 3 (ii) says that, for a group of standard uniform random variables that consists of n subgroups of perfectly positively dependent components, if sub-uniformity holds for n components each of which comes from one distinct subgroup, then it also holds for the whole group.

In the rest of this paper, we will show that sub-uniformity holds for several dependence models including independence and some forms of positive and negative dependence. Together with Propositions 2 and 3, by mixing these dependence models or connecting them via groups, we can obtain a much wider range of dependence models for which sub-uniformity holds.

4 Sub-uniformity for dependent p-values

In this section, we study sub-uniformity for standard uniform random variables that are negatively or positively dependent in specific forms. In particular, we show that sub-uniformity holds for negative upper orthant dependence, extremal mixture copulas, and some Clayton copulas.

4.1 Negative upper orthant dependence

A random vector $\mathbf{X} = (X_1, \dots, X_n)$ is *negatively upper orthant dependent* (NUOD) if for all $x_1, \dots, x_n \in \mathbb{R}$, it holds that

$$\mathbb{P}(X_1 > x_1, \dots, X_n > x_n) \leq \prod_{i=1}^{n} \mathbb{P}(X_i > x_i). \tag{7}$$

It is said to be *negatively lower orthant dependent* (NLOD) if for all $x_1, \dots, x_n \in \mathbb{R}$,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \leq \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i). \tag{8}$$

In general, the two notions of negative dependence are not equivalent except when $n = 2$. If $\mathbf{X}$ is both NUOD and NLOD, it is said to be *negatively orthant dependent*, as introduced by Block et al. (1982). Negative orthant dependence includes multivariate normal distributions with non-positive correlations as special cases; see, e.g., Chi et al. (2024) for examples of negative dependence in multiple hypothesis testing.

Intuitively, (7) (resp. (8)) means that that $X_1, \dots, X_n$ are less likely to be large (resp. small) simultaneously compared to $X_1', \dots, X_n'$, which are iid copies of $X_1, \dots, X_n$, hence is a notion of negative dependence. Negative orthant dependence is closely related to other notions of negative dependence. It is implied by negative association (Joag-Dev and Proschan, 1983), negative regression dependence (Block et al., 1985), and weak negative association (Chen et al., 2024); see, e.g., Chen et al. (2024) for a discussion of these implication relations.

The next result gives sub-uniformity for NUOD standard uniform random variables. It is based on Theorem 1 of Chen and Shneer (2024), where a stochastic order relation for heavy-tailed random variables is shown. As such, this result is essentially known in a different context, but we present it here for its concise statement and clear interpretation.

Proposition 4. *Let $r \leq -1$. If $(U_1, \dots, U_n)$ is NUOD, then $M_r^{\mathbf{w}}(U_1, \dots, U_n)$ is strictly sub-uniform for all $\mathbf{w} \in \Delta_n^+$.*

Proof. It suffices to show

$$\mathbb{P}\left(\sum_{i=1}^{n} w_i U_i^{-1} > t\right) > \frac{1}{t} \quad \text{holds for all } t > 1 \text{ and all } \mathbf{w} \in \Delta_n^+.$$

Note that $U_i^{-1} - 1$, $i \in [n]$, follows a Pareto distribution with distribution function $\mathbb{P}(U_i^{-1} - 1 \leq x) = 1 - 1/(x + 1)$, $x > 0$. As $(U_1, \dots, U_n)$ is NUOD, $(U_1^{-1}, \dots, U_n^{-1})$ is NLOD. By Theorem 1 of

Chen and Shneer (2024),

$$\mathbb{P}\left(\sum_{i=1}^{n} w_i(U_i^{-1}-1)>x\right)>\mathbb{P}\left(U_1^{-1}-1>x\right)=\frac{1}{x+1}\ \text{ holds for all } x>0 \text{ and all } \mathbf{w}\in\Delta_n^+.$$

Hence, we have the desired result. □

Certainly, the result in Proposition 4 also holds for $(U_1,\dots,U_n)$ satisfying any of the stronger notions of negative dependence than negative upper orthant dependence.

4.2 Extremal mixture copulas

Next, we apply Proposition 4 to show that sub-uniformity holds for extremal mixture copula (McNeil et al., 2020). We say that $(U_1,\dots,U_n)$ follows an *extremal copula* $C\in\mathcal{C}_n$ with an index set $J\subseteq[n]$, if

$$U_j \overset{\mathrm{d}}{=} \mathbb{1}_{\{j\in J\}}U+\mathbb{1}_{\{j\in J^c\}}(1-U),\quad \text{for}\ \ j\in[n],$$

where U is a standard uniform random variable. For $n\geq 2$, there are 2^{n-1} different extremal copulas. Let $\mathbf{s}_i$ be a vector consisting of the digits of a n-digit binary number which represents the decimal number $i-1$, for each $i\in[2^{n-1}]$. For instance, if $n=3$, we have

$$\mathbf{s}_1=(0,0,0),\quad \mathbf{s}_2=(0,0,1),\quad \mathbf{s}_3=(0,1,0),\quad \mathbf{s}_4=(0,1,1).$$

For $n\in\mathbb{N}$, let J_i be the index set of zeros in $\mathbf{s}_i$, for each $i\in[2^{n-1}]$. Denote by $C^{(i)}$ the extremal copula with index set J_i. Note that $C^{(1)}$ is the comonotonicity copula. A copula C is an *extremal mixture copula* with a vector $(a_1,\dots,a_{2^{n-1}})\in\Delta_{2^{n-1}}$ if $C=\sum_{i=1}^{2^{n-1}}a_iC^{(i)}$. A random vector following an extremal mixture copula is not necessarily NUOD.

Theorem 1. *Let $r\leq -1$. If $(U_1,\dots,U_n)$ follows an extremal mixture copula with $(a_1,\dots,a_{2^{n-1}})\in\Delta_{2^{n-1}}$ such that $a_1<1$, then $M_r^{\mathbf{w}}(U_1,\dots,U_n)$ is strictly sub-uniform for all $\mathbf{w}\in\Delta_n^+$.*

Proof. Let $(V_1,\dots,V_n)$ follow an extremal copula $C^{(k)}\in\mathcal{C}_n$ with an index set $J_k\subseteq[n]$, $k\in[2^{n-1}]$. For $\mathbf{w}=(w_1,\dots,w_n)\in\Delta_n^+$ and $p\in(0,1)$,

$$\mathbb{P}(M_{-1}^{\mathbf{w}}(V_1,\dots,V_n)\leq p)=\mathbb{P}\left(\sum_{i\in J_k}w_iV^{-1}+\sum_{j\in J_k^c}w_j(1-V)^{-1}\geq p^{-1}\right)=\mathbb{P}(M_{-1}^{\eta}(V,1-V)\leq p),$$

where $\eta=(\sum_{i\in J_k}w_i,\sum_{j\in J_k^c}w_j)\in\Delta_2$ and V is a standard uniform random variable. Since $C^{(1)}$ is the comonotonicity copula, the above probability equals to p if $k=1$. It is straightforward to check that $(V, 1-V)$ is NUOD. By Proposition 4, $M_{-1}^{\eta}(V, 1-V)$ is strictly sub-uniform if $k \geq 2$. Hence, strict sub-uniformity holds for extremal copula $C^{(k)}$, $k = 2, \ldots, 2^{n-1}$. As extremal mixture copula is a weighted mixture of extremal copulas, by Proposition 2, sub-uniformity holds for any extremal mixture copula. It is clear that sub-uniformity is strict if $a_1 \neq 1$. $\square$

Remark 2. By McNeil et al. (2020, Theorem 1), every Kendall's rank correlation matrix (Kendall matrix, for short) can be attained by some extremal mixture copula. Together with Theorem 1, this implies that for any given Kendall matrix $\mathcal{T}$ and $r \leq -1$, there always exists a copula with the specified Kendall matrix $\mathcal{T}$ such that sub-uniformity holds, and strict sub-uniformity holds if $\mathcal{T}$ is not the matrix of ones.

4.3 Clayton copula and positive dependence

In this section, we study sub-uniformity for standard uniform random variables with a specific positive dependence structure, modelled by Clayton copulas, which have been used to model p-values in, e.g., Dickhaus and Gierl (2013), Bodnar and Dickhaus (2014), and Neumann et al. (2019). The Clayton copula C with parameter $t > 0$, denoted by Clayton(t), is given by

$$C(u_1, \ldots, u_n) = \left(u_1^{-t} + \cdots + u_n^{-t} - (n-1)\right)^{-1/t}, \quad (u_1, \ldots, u_n) \in (0,1)^n.$$

The Clayton copula with $t > 0$ represents a type of positive dependence, with $t \to \infty$ yielding comonotonicity and $t \downarrow 0$ yielding independence. For a random vector following a Clayton copula, Kendall's tau of any pair of its components is equal to $t/(t+2)$.

Clayton copulas arise naturally in the following context. Suppose that $X_1, \ldots, X_n$ are iid exponential random variables with parameter λ, and Y is a Gamma random variable with parameter $(1/t, 1)$ independent of $X_1, \ldots, X_n$; the exponential distribution with parameter $\lambda > 0$ is given by $F(x) = 1 - \exp(-x/\lambda)$, $x \geq 0$ and the Gamma distribution with parameter $(k, \theta) \in \mathbb{R}_+^2$ is given by $F(x) = (\Gamma(k)\theta^k)^{-1} \int_0^x y^{k-1} \exp(-y/\theta) \mathrm{d}y$, $x \geq 0$. The random vector $(T_1, \ldots, T_n) = (X_1/Y, \ldots, X_n/Y)$ is usually used to model the lifetimes of n objects in a system; see, e.g., Lindley and Singpurwalla (1986). The joint distribution of $(T_1, \ldots, T_n)$ is known as a multivariate Pareto distribution of type II with marginal distribution $G(x) = 1 - (\lambda/(\lambda + x))^{1/t}$ for $x \geq 0$. Let

$$(U_1, \ldots, U_n) = (1 - G(T_1), \ldots, 1 - G(T_n)) = (1 - G(X_1/Y), \ldots, 1 - G(X_n/Y)). \tag{9}$$

Each of $U_1, \ldots, U_n$ follows a standard uniform distribution and $(U_1, \ldots, U_n)$ has a Clayton(t) copula; see, e.g., Sarabia et al. (2016).

The next result gives sub-uniformity of the r-mean of p-values following a corresponding Clayton copula, as well as a positive result on a type-I error rate bound for the r-mean of p-values.

Theorem 2. *Let $t \geq 1$, $\mathbf{w} \in \Delta_n^+$ and $(U_1, \ldots, U_n) \overset{\mathrm{d}}{\sim} \mathrm{Clayton}(t)$. If $r \geq t$, then $M_{-r}^{\mathbf{w}}(U_1, \ldots, U_n)$ is strictly sub-uniform. Moreover, for $p \in (0,1)$ and $s \in [1, t]$, we have*

$$\mathbb{P}\left(M_{-s}^{\mathbf{w}}(U_1, \ldots, U_n) \leq p\right) \leq G_{1/t}\left(\frac{1}{p^{-t} - 1}\right),$$

where $G_{1/t}$ is the cdf of a Gamma distribution with parameter $(1/t, 1)$. In particular,

$$\mathbb{P}\left(M_{-1}^{\mathbf{w}}(U_1, \ldots, U_n) \leq p\right) \leq \sup_{b \geq 1} G_{1/b}\left(\frac{1}{p^{-b} - 1}\right). \tag{10}$$

Proof. We first show the case when $t = r$. Note that for $\mathbf{w} \in \Delta_n^+$, simple algebra leads to

$$\mathbb{P}(M_{-t}^{\mathbf{w}}(U_1, \ldots, U_n) \leq p) = \mathbb{P}\left(\frac{\sum_{i=1}^n w_i X_i}{Y} \geq \lambda p^{-t} - \lambda\right), \tag{11}$$

where $X_1, \ldots, X_n$ and Y are as in (9). Therefore, to show sub-uniformity, i.e., $\mathbb{P}(M_{-t}^{\mathbf{w}}(U_1, \ldots, U_n) \leq p) > p$, it suffices to show $\mathbb{P}(\sum_{i=1}^n w_i X_i / Y \geq x) > \mathbb{P}(X_1/Y \geq x)$ for all $x > 0$. We have

$$\mathbb{P}\left(\frac{\sum_{i=1}^n w_i X_i}{Y} \geq x\right) = \mathbb{E}\left[\mathbb{P}\left(Y \leq \frac{\sum_{i=1}^n w_i X_i}{x} \mid \sum_{i=1}^n w_i X_i\right)\right] = \mathbb{E}\left[\phi\left(\sum_{i=1}^n w_i X_i\right)\right],$$

where $\phi(y) = \mathbb{P}(Y \leq y/x)$ for $y > 0$. Assume $x = 1$ without losing generality. Taking the second derivative of ϕ, we get

$$\phi''(y) = \frac{y^{1/t-2} \exp(-y)(1/t - 1 - y)}{\Gamma(1/t)}.$$

As $t = r \geq 1$, we have $\phi'' < 0$ and ϕ is strictly concave. Therefore,

$$\mathbb{P}\left(\frac{\sum_{i=1}^n w_i X_i}{Y} \geq x\right) = \mathbb{E}\left[\phi\left(\sum_{i=1}^n w_i X_i\right)\right] > \mathbb{E}\left[\phi\left(X_1\right)\right] = \mathbb{P}\left(\frac{X_1}{Y} \geq x\right).$$

Hence, we obtain the desired stochastic dominance for $t = r$. The statement for $t \in [1, r)$ is due to the fact that $M_{-r} \leq M_{-t}$ (Hardy et al., 1934, Theorem 16).

To show the last inequality, note that Jensen's inequality gives

$$\mathbb{E}\left[\phi\left(\sum_{i=1}^n w_i X_i\right)\right] \leq \phi\left(\sum_{i=1}^n w_i \mathbb{E}[X_i]\right) = \phi(\lambda) = \mathbb{P}(Y \leq \lambda/x) = G_{1/t}(\lambda/x).$$

Hence, using (11), we get

$$\mathbb{P}(M_{-t}(U_1, \dots, U_n) \le p) \le G_{1/t}(1/(p^{-t} - 1)),$$

and the desired upper bound follows from noting again that $M_{-s} \ge M_{-t}$ for $s \le t$. $\square$

Applying Theorem 2 with the special case $r = t = s = 1$, we get that, if $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim}$ Clayton(1), then

$$p < \mathbb{P}\left(M_{-1}(U_1, \dots, U_n) \le p\right) \le 1 - e^{1-1/(1-p)} \le \frac{1}{1-p} - 1 = \frac{p}{1-p}.$$

As a consequence, although being anti-conservative without correction, the harmonic mean p-value becomes valid with the simple threshold $t_p = p/(1+p)$, i.e., $\mathbb{P}\left(M_{-1}(U_1, \dots, U_n) \le t_p\right) \le p$ holds for any $n \in \mathbb{N}$. The needed correction is minor since p and $p/(1+p)$ are very close for small p.

The null p-values following Clayton(1) is a strong assumption, and it can be relaxed by using (10). Suppose that $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim}$ Clayton(t) with some unknown $t \ge 1$. Define a constant

$$\kappa = \sup_{p \in (0,0.1],\ b \ge 1} \frac{1}{p} G_{1/b}\left(\frac{1}{p^{-b} - 1}\right).$$

By (10), we have $\mathbb{P}\left(M_{-1}^{\mathbf{w}}(U_1, \dots, U_n) \le p\right) \le \kappa p$ for all $p \in (0, 0.1]$. Numerical calculation gives $\kappa \approx 1.1304$; the maximum in computing κ is approximately attained at $p = 0.1$ and $b = 2.0853$. Therefore, if $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim}$ Clayton(t) with some $t \ge 1$, we can use a threshold $u_p = p/1.131$ for the harmonic mean p-value such that $\mathbb{P}(M_{-1}^{\mathbf{w}}(U_1, \dots, U_n) \le u_p) \le p$ for all $p \in (0, 0.1]$. This correction is valid for all n. This shows that positive dependence makes the harmonic mean p-value well behaved; in sharp contrast, the needed correction grows to infinity as $n \to \infty$ in case of independence; see Section 5.

The reason why the harmonic mean p-value behaves well for the Clayton copula is perhaps because U_i^{-1}, $i \in [n]$, is extremely heavy-tailed (i.e., U_i^{-1} does not have a finite mean). It is well known in EVT that for iid extremely heavy-tailed random variables $U_1^{-1}, \dots, U_n^{-1}$, a large value of $U_1^{-1} + \dots + U_n^{-1}$ (thus a small value of $M_{-1}(U_1, \dots, U_n)$) is most likely caused by one extremely large U_i^{-1} rather than by several moderately large U_i^{-1} (see Embrechts et al. (1997)). In the case of strong positive dependence, the possibility of getting one extremely large U_i^{-1} may be reduced (e.g., Alink et al. (2004)).

In the case of the symmetric mean function M_r, the distribution of $M_{-r}(U_1, \dots, U_n)$ with $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim}$ Clayton(r) has an analytical formula provided below.

Proposition 5. *Let $r \geq 1$ and $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim} \mathrm{Clayton}(r)$. For $p \in (0,1)$, we have*

$$\mathbb{P}\left(M_{-r}(U_1, \dots, U_n) \leq p\right) = 1 - B_{n,r}\left(\frac{np^{-r} - n}{np^{-r} - n + 1}\right),$$

where $B_{n,r}$ is a Beta cdf given by

$$B_{n,r}(x) = \frac{\Gamma\left(n + 1/r\right)}{\Gamma(n)\Gamma\left(1/r\right)} \int_0^x t^{n-1}(1-t)^{1/r-1}\mathrm{d}t, \quad x \in (0,1).$$

Proof. By (11), we have

$$\mathbb{P}\left(M_{-r}(U_1, \dots, U_n) \leq p\right) = \mathbb{P}\left(\sum_{i=1}^n T_i \geq \lambda(np^{-r} - n)\right).$$

Since λ is the scale parameter of the exponential distribution, the above probability is indifferent to λ. We assume $\lambda = 1$ for simplicity. Given $Y = y$, the conditional distribution of $\sum_{i=1}^n X_i/Y$ is a Gamma distribution with parameter $(n, 1/y)$. Since Y is also a Gamma distribution, $\sum_{i=1}^n T_i$ follows a compound Gamma distribution. Using (1.2) of Dubey (1970), we have the desired equality. $\square$

Besides the above Clayton copulas, we provide below two other positive dependence structures for which sub-uniformity holds.

Example 1. Let $X, X_i \overset{\mathrm{d}}{\sim} \mathrm{U}(0,1)$, $Y, Y_i \overset{\mathrm{d}}{\sim} \mathrm{U}(\beta, 1)$, and $Z, Z_i \overset{\mathrm{d}}{\sim} \mathrm{U}(0, \beta)$, $i \in [n]$, be independent, where $\beta \in (0,1)$. Assume that $(U_1, \dots, U_n)$ is modelled by one of the two cases below, that is,

$$U_i = \mathbb{1}_{\{X \leq \beta\}} Z_i + \mathbb{1}_{\{X > \beta\}} Y, \quad i \in [n] \qquad \text{or} \qquad U_i = \mathbb{1}_{\{X_i \leq \beta\}} Z + \mathbb{1}_{\{X_i > \beta\}} Y_i, \quad i \in [n]. \tag{12}$$

Clearly, $U_i \overset{\mathrm{d}}{\sim} \mathrm{U}(0,1)$ for all $i \in [n]$. Moreover, $(U_1, \dots, U_n)$ defined by (12) is positively dependent as for $i \neq j$, with $\mathrm{corr}(U_i, U_j) = 1 - \beta^3$ if $(U_1, \dots, U_n)$ is the first case and $\mathrm{corr}(U_i, U_j) = \beta^4$ if $(U_1, \dots, U_n)$ is the second case. If $(U_1, \dots, U_n)$ is modelled by either case in (12), it is known that $S(U_1, \dots, U_n) \simeq_{\mathrm{st}} U_1$ where $S(u_1, \dots, u_n) = \min\left\{nu_{(i)}/i\right\}$, $(u_1, \dots, u_n) \in (0, \infty)^n$ (see Example 1 in Samuel-Cahn (1996) and Proposition 3.4 in Xiong and Hu (2022)). Let $r \leq -1$. By Theorem 16 in Hardy et al. (1934) and Theorem 3 in Chen et al. (2023), $M_r \leq M_{-1} \leq S$. Hence, $M_r(U_1, \dots, U_n)$ is sub-uniform if $(U_1, \dots, U_n)$ is modelled by (12).

Remark 3. As the problem of merging p-values is closely related to the problem of risk aggregation (see Section 1), our results on (2) under various conditions have a direct interpretation in risk management. For all $i \in [n]$, let $X_i = U_i^r$ with $r \leq -1$. Then X_i follows a Pareto distribution with infinite mean, which is widely used in modeling extremely heavy-tailed financial losses. For

$(w_1, \ldots, w_n) \in \Delta_n$ and $x > 0$, sub-uniformity of $U_1, \ldots, U_n$ implies $\mathbb{P}(\sum_{i=1}^n w_i X_i > x) \geq \mathbb{P}(X_1 > x)$. Therefore, for identically distributed (but possibly dependent) infinite-mean Pareto losses, the average loss is strictly riskier than any individual loss; that is, diversification is harmful. Putting this in the language of risk measures, the Value-at-Risk, which is a quantile of a loss position, is strictly superadditive at any level for these losses. This issue was extensively discussed in Chen et al. (2024), and here we use more general dependence models. See also Nešlehová et al. (2006) for the diversification of infinite-mean models in the context of operational risk.

5 Threshold under independence

As we have seen above, the harmonic mean p-value is anti-conservative under a wide range of dependence assumptions. It is then worth studying its threshold by which the type-I error rate can be properly controlled below the significance level. However, explicit expressions of the probability distributions of the harmonic mean p-value are generally not available, even when p-values are independent. In this section, we focus on the independence case and use the generalized central limit theorem to derive the asymptotic threshold of the harmonic mean p-value as the number of p-values goes to infinity. The asymptotic threshold is potentially useful in genomewide study where there are a large number of p-values (e.g., Storey and Tibshirani (2003)), although the independence assumption may not be verifiable in such a context.

For $\alpha \in (0, 1]$, $q_\alpha(X)$ is the left α-quantile of a random variable X, defined as

$$q_\alpha(X) = \inf\{x \in \mathbb{R} \mid \mathbb{P}(X \leq x) \geq \alpha\}.$$

We also use $F^{-1}(\alpha)$ for $q_\alpha(X)$ if X follows a distribution F. Let $U_1, \ldots, U_n$ be independent. For $p \in (0, 1)$, denote by $a_{n,p}$ the threshold of the symmetric harmonic mean of p-values, that is, $a_{n,p} = q_p(M_{-1}(U_1, \ldots, U_n))$. It is clear that

$$\mathbb{P}(M_{-1}(U_1, \ldots, U_n) < a_{n,p}) \leq p.$$

Let S_1 be a distribution function with characteristic function given by

$$\int_{\mathbb{R}} \exp(i\theta x) \mathrm{d}S_1(x) = \exp\left(-|\theta|(1 + i\frac{2}{\pi}\operatorname{sgn}(\theta)\log|\theta|)\right) \quad \text{for} \quad \theta \in \mathbb{R},$$

where $\operatorname{sgn}(\cdot)$ is the sign function. The distribution S_1 is a stable distribution with tail parameter 1 (see Samorodnitsky and Taqqu (1994)). The following proposition gives an asymptotic approximation of $a_{n,p}$ for large n. Using $a_{n,p}$ is equivalent to the asymptotically exact test of the harmonic mean p-value method of Wilson (2019).

Proposition 6. *For $p \in (0,1)$, let $a_{n,p}$ be the threshold of M_{-1}. Then*

$$a_{n,p} \sim \left(\frac{\pi}{2} S_1^{-1}(1-p) + \log\left(\frac{n\pi}{2}\right) + 1 - \gamma\right)^{-1} \to 0 \quad \textit{as } n \to \infty, \tag{13}$$

where γ is the the Euler–Mascheroni constant.[3]

Proof. Note that the random variables $U_1^{-1}, \dots, U_n^{-1}$ follow a Pareto distribution with distribution function $\mathbb{P}(U^{-1} \le x) = 1 - x^{-1}$, $x \in [1, \infty)$. By the generalized central limit theorem (see Theorem 1.8.1 in Samorodnitsky and Taqqu (1994)), sum of iid Pareto random variables $U_1^{-1}, \dots, U_n^{-1}$ behaves like a stable distribution with tail parameter 1 for large n. Let $Z \overset{\mathrm{d}}{\sim} S_1$. Hence for $p \in (0,1)$,

$$\lim_{n\to\infty} \frac{\mathbb{P}\left(M_{-1}(U_1,\dots,U_n) \le p\right)}{\mathbb{P}\left(Z \ge c_n^{-1}(np^{-1} - d_n)\right)} = \lim_{n\to\infty} \frac{\mathbb{P}\left(c_n^{-1}(\sum_{i=1}^n U_i^{-1} - d_n) \ge c_n^{-1}(np^{-1} - d_n)\right)}{\mathbb{P}\left(Z \ge c_n^{-1}(np^{-1} - d_n)\right)} = 1,$$

where $d_n = \dfrac{\pi n^2}{2} \displaystyle\int_1^\infty \sin\left(\frac{2x}{n\pi}\right) x^{-2} \mathrm{d}x$ and $c_n = n\pi/2$. This implies

$$a_{n,p} = q_p(M_{-1}(U_1,\dots,U_n)) \sim \left(\frac{\pi}{2} S_1^{-1}(1-p) + \frac{n\pi}{2} \int_1^\infty \sin\left(\frac{2x}{n\pi}\right) x^{-2} \mathrm{d}x\right)^{-1} := b_{n,p} \quad \text{as } n \to \infty.$$

By Taylor's expansion and properties of the cosine integral, we get

$$\begin{aligned}
\frac{n\pi}{2} \int_1^\infty \sin\left(\frac{2x}{n\pi}\right) x^{-2} \mathrm{d}x &= \frac{n\pi}{2} \sin\left(\frac{2}{n\pi}\right) + \int_{\frac{2}{n\pi}}^\infty \frac{\cos(y)}{y} \mathrm{d}y \\
&= 1 + \sum_{i=1}^\infty \frac{(-1)^i (2n^{-1}\pi^{-1})^{2i}}{(2i+1)!} - \left(\gamma + \log\left(\frac{2}{n\pi}\right) - \int_0^{\frac{2}{n\pi}} \frac{1-\cos(y)}{y} \mathrm{d}y\right) \\
&\sim \log\left(\frac{n\pi}{2}\right) + 1 - \gamma.
\end{aligned}$$

Hence, as $n \to \infty$,

$$a_{n,p} \sim b_{n,p} \sim \left(\frac{\pi}{2} S_1^{-1}(1-p) + \log\left(\frac{n\pi}{2}\right) + 1 - \gamma\right)^{-1}.$$

We have the desired result. □

Proposition 6 means that as more independent p-values are merged by M_{-1}, a smaller threshold needs to be used. In other words, there does not exist a constant multiplier which makes the harmonic mean p-value valid, consistent with Wilson (2019). By (13), the multiplier κ_n such that

[3] The Euler–Mascheroni constant γ is approximately 0.57721.

$\mathbb{P}(M_{-1}(U_1, \dots, U_n) < p) \le \kappa_n p$ grows at a rate of $\log n$ as n goes to infinity. This is in sharp contrast to the dependence structure modelled by the Clayton copulas in Theorem 2, where the correction does not go to infinity as n increases.

Chen et al. (2023) showed that the harmonic mean p-value method is closely related to two commonly used merging methods, the Cauchy combination (Liu and Xie, 2020) and the Simes methods (Simes, 1986). The Cauchy combination method uses M_g (see Section 1) to combine p-values, where g is chosen to be the quantile function of the standard Cauchy distribution, i.e., $g(p) = \tan(\pi(p - 0.5))$ for $p \in (0, 1)$. The Simes method merges p-values via the Simes function S defined in Section 3. In contrast to the harmonic mean p-value method, the Simes and the Cauchy combination methods always produce valid merged p-values for any number of independent p-values. Hence, no correction is required for the merged p-values of the Simes and the Cauchy combination methods. This holds not only for independent p-values, but also for p-values modelled by a wide range of dependence structures. For instance, the Cauchy combination method is valid for the dependence structures considered by Pillai and Meng (2016). The Simes method is also conservative if the test statistics follow a multivariate normal distribution with nonnegative correlations (Sarkar, 1998). In this case, however, both the harmonic mean p-value and the Cauchy combination methods seem to be anti-conservative, based on numerical experiments; see Section 7 for the harmonic mean p-value and the simulation results in Chen et al. (2023) for the Cauchy combination method. See also Rustamov and Klosowski (2020) and Fang et al. (2023) for comparisons between the harmonic mean and the Cauchy combination methods.

6 Discrete uniform random variables

In this section, instead of considering standard uniform random variables, we study discrete uniform random variables $U_1^m, \dots, U_n^m$ on a finite set $\{1/m, \dots, m/m\}$ of m equidistant points. This setting concerns discrete p-values, which may be obtained from, for instance, binomial test and conformal p-scores; see Vovk et al. (2005) and the more recent Bates et al. (2023).

We first note that for discretely distributed $U_1^m, \dots, U_n^m$, one cannot expect

$$M_r^{\mathbf{w}}(U_1^m, \dots, U_n^m) \preceq_{\mathrm{st}} U_1^m$$

to hold for any $r \in \mathbb{R}$ and $\mathbf{w} \in \Delta_n^+$ unless $U_1^m, \dots, U_n^m$ are identical. The reason is that

$$\mathbb{P}\left(M_r^{\mathbf{w}}(U_1^m, \dots, U_n^m) \le 1/m\right) = \mathbb{P}\left(U_1^m = \dots = U_n^m = 1/m\right),$$

which is less than $1/m$ unless the events $U_i^m = 1/m$ for $i \in [n]$ occur together almost surely. Applying similar arguments on $\mathbb{P}\left(M_r^{\mathbf{w}}(U_1^m, \dots, U_n^m) \le k/m\right)$, $k \in [m]$, leads to $U_i^m = k/m$ for $i \in [n]$ also occur together almost surely for all $k \in [m]$. Hence, $M_r^{\mathbf{w}}(U_1^m, \dots, U_n^m) \preceq_{st} U_1$ implies that $U_1^m, \dots, U_n^m$ are identical. This argument is similar to Proposition 1.

In the context of hypothesis testing, we are more interested in whether the following inequality holds,

$$\mathbb{P}\left(M_r^{\mathbf{w}}(U_1^m, \dots, U_n^m) \le p\right) > p \quad \text{for some pre-specified } p \in (0,1), \tag{14}$$

where $r \le -1$. Based on previous discussions on sub-uniformity for standard uniform random variables, we may expect that (14) holds for $(U_1^m, \dots, U_n^m)$ with large m if it has a copula for which sub-uniformity holds. The intuition is that if m is very large, the distribution of each $U_1^m, \dots, U_n^m$ is close to the uniform distribution on $(0,1)$. If $U_1^m, \dots, U_n^m$ are NUOD, we show below that (14) holds asymptotically as m goes to infinity in the case of symmetric mean function. Following a similar line of thought, the corresponding result also holds if $(U_1^m, \dots, U_n^m)$ has certain Clayton copula or an extremal mixture copula, as in Section 4.

Theorem 3. *Let $r \le -1$, $p \in (0,1)$, and $U_1^m, \dots, U_n^m$ be NUOD discrete uniform random variables on $\{1/m, \dots, m/m\}$, $m \ge 2$. There exists a sequence $\{p_m : m \ge 2\}$ such that*

$$\mathbb{P}\left(M_r(U_1^m, \dots, U_n^m) \le p\right) \ge p_m \xrightarrow[m\to\infty]{} p.$$

Moreover, if $m > n^{-1/r} p^{-1}$, we can take

$$p_m = p - \frac{p^{1-r}}{m}\left(\left(np^r - (n-1)\left(\frac{m+1}{m}\right)^r\right)^{1/r} - \frac{1}{m}\right)^{r-1},$$

and $p_m = 0$ otherwise.

Proof. By Theorem 3 of Lin et al. (2024) and its proof, random variables $U_1^m, \dots, U_n^m$ are NUOD if and only if there exist NUOD standard uniform random variables $V_1^m, \dots, V_n^m$ such that

$$U_i^m = \sum_{j=1}^m \frac{j}{m} \mathbb{1}_{\{(j-1)/m < V_i^m \le j/m\}}, \quad i \in [n].$$

For $p \in (0,1)$, let $R(p) = \mathbb{P}\left(M_r(U_1^m, \dots, U_n^m) \le p\right)$, $r \le -1$. Define the following events

$$A = \left\{\frac{1}{n}\sum_{i=1}^n \left(V_i^m + \frac{1}{m}\right)^r < p^r\right\}, \quad B = \left\{\frac{1}{n}\sum_{i=1}^n (V_i^m)^r \ge p^r\right\},$$

$$\text{and} \quad C = \left\{ \frac{1}{n} \sum_{i=1}^{n} \left((V_i^m)^r + \frac{r}{m} (V_i^m)^{r-1} \right) < p^r \right\}.$$

Note that $U_i^m \leq V_i^m + 1/m$ for all $i \in [n]$. We have

$$\begin{aligned}
R(p) &= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} (U_i^m)^r \geq p^r \right) \\
&\geq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} \left(V_i^m + \frac{1}{m} \right)^r \geq p^r \right) \\
&\geq \mathbb{P}\left(\left\{ \frac{1}{n} \sum_{i=1}^{n} \left(V_i^m + \frac{1}{m} \right)^r \geq p^r \right\} \cap B \right) \\
&= \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B \cap C). \qquad (15)
\end{aligned}$$

The last equality is due to $A \subseteq C$ as for fixed $x \in (0,1)$, $(x+\epsilon)^r \geq x^r + \epsilon r x^{r-1}$, $\epsilon \in (0,1)$. Note that

$$\begin{aligned}
A &= \bigcap_{i=1}^{n} \left\{ \left(V_i^m + \frac{1}{m} \right)^r < np^r - \sum_{j \in [n]/i} \left(V_j^m + \frac{1}{m} \right)^r \right\} \\
&\subseteq \bigcap_{i=1}^{n} \left\{ \left(V_i^m + \frac{1}{m} \right)^r < np^r - (n-1) \left(\frac{m+1}{m} \right)^r \right\} \\
&= \bigcap_{i=1}^{n} \left\{ V_i^m > \left(np^r - (n-1) \left(\frac{m+1}{m} \right)^r \right)^{1/r} - \frac{1}{m} \right\} = \bigcap_{i=1}^{n} \left\{ V_i^m > z_p \right\}, \qquad (16)
\end{aligned}$$

where

$$z_p = \left(np^r - (n-1) \left(\frac{m+1}{m} \right)^r \right)^{1/r} - \frac{1}{m}.$$

Note that z_p is positive for large m and negative for small m. If $m \leq n^{-1/r} p^{-1}$, we let $p_m = 0$ to get the trivial bound $R(p) \geq 0$.

We next focus on the case that $m > n^{-1/r} p^{-1}$. Since $p > n^{-1/r} m^{-1} \geq (n^{-1}(m^{-r} + (n-1)(m+1)^r m^{-r}))^{1/r} > m^{-1}$, we can verify that $0 < z_p < 1$. Let

$$D = \bigcap_{i=1}^{n} \{V_i^m > z_p\}.$$

By (15) and (16), we have

$$
\begin{aligned}
R(p) &\geq \mathbb{P}(B) - \mathbb{P}(A \cap B \cap C) \\
&= \mathbb{P}(B) - \mathbb{P}(A \cap B \cap C \cap D) \\
&\geq \mathbb{P}(B) - \mathbb{P}(B \cap C \cap D) \\
&= \mathbb{P}\left(B\right) - \mathbb{P}\left(\left\{p^r \leq \frac{1}{n}\sum_{i=1}^{n}\left(V_i^m\right)^r < p^r - \frac{1}{n}\sum_{i=1}^{n}\frac{r}{m}\left(V_i^m\right)^{r-1}\right\} \cap D\right) \\
&\geq \mathbb{P}\left(B\right) - \mathbb{P}\left(\left\{p^r \leq \frac{1}{n}\sum_{i=1}^{n}\left(V_i^m\right)^r < p^r - \frac{r z_p^{r-1}}{m}\right\} \cap D\right) \\
&\geq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left(V_i^m\right)^r \geq p^r\right) - \mathbb{P}\left(p^r \leq \frac{1}{n}\sum_{i=1}^{n}\left(V_i^m\right)^r < p^r - \frac{r z_p^{r-1}}{m}\right) \\
&= \mathbb{P}\left(M_r(V_1^m, \ldots, V_n^m) \leq \left(p^r - \frac{r z_p^{r-1}}{m}\right)^{1/r}\right).
\end{aligned}
$$

By Proposition 4,

$$
\begin{aligned}
R(p) &\geq \left(p^r - \frac{r z_p^{r-1}}{m}\right)^{1/r} \\
&\geq p - \frac{z_p^{r-1}}{m}p^{1-r} \\
&= p - \frac{p^{1-r}}{m}\left(\left(np^r - (n-1)\left(\frac{m+1}{m}\right)^r\right)^{1/r} - \frac{1}{m}\right)^{r-1} = p_m.
\end{aligned}
$$

It is straightforward to verify that p_m goes to p as m goes to ∞. $\square$

7 Numerical examples

Throughout this section, let $R_n(p) = \mathbb{P}(M_{-1}(U_1, \ldots, U_n) \leq p)$ for $p \in (0,1)$, where $U_1, \ldots, U_n$ are standard uniform random variables or discrete uniform random variables on $\{1/m, \ldots, m/m\}$ with $m \geq 2$. We first provide a few small numerical examples to illustrate sub-uniformity for dependent $U_1, \ldots, U_n$. The first example is for standard uniform random variables, which follow the copula generated by an equicorrelated Gaussian distribution with $\rho \in [0,1]$. Let Φ be the standard normal distribution function, and $Z, Z_1, \ldots, Z_n$ be independent identically distributed standard normal random variables. Write

$$
U_i = \Phi(X_i), \text{ where } X_i = \rho Z + \sqrt{1-\rho^2} Z_i, \ i \in [n] \ .
$$

Fix $p = 0.1$. In Figure 1, we display $R_n(p)$ for $n = 5, 10, 15, 20$, and $\rho \in [0, 1]$. We observe that sub-uniformity holds for all $\rho \in [0, 1]$, and that as n increases, $R_n(p)$ gets larger. These results show that sub-uniformity may also hold for the class of equicorrelated Gaussian copulas with positive correlations, but the results in this paper can only cover the case of Gaussian copulas with non-positive correlations due to negative upper orthant dependence, and the corresponding sub-uniformity statement for a general positive ρ is not known in the literature. If the significance level p is extremely small, it is expected that the inflation issue will be less severe, as justified by Theorem 2 of Chen et al. (2023).

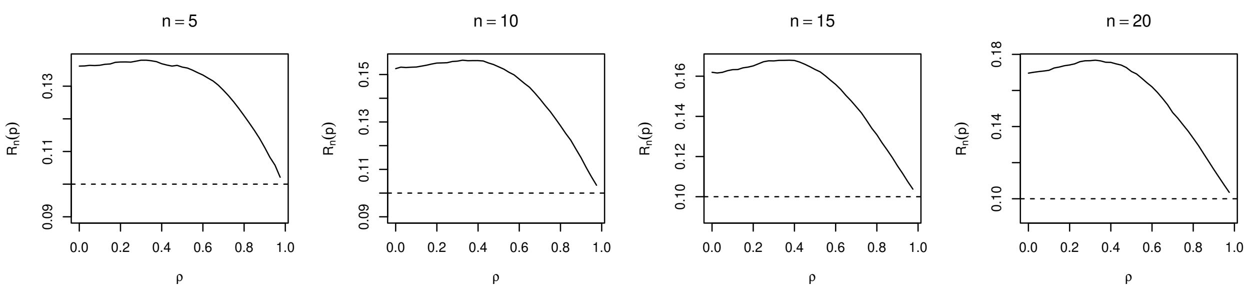

Figure 1: Equicorrelated Gaussian copula: $R_n(p)$ for $n = 5, 10, 15, 20$, and $\rho \in [0, 1]$ with $p = 0.1$.

In the second example, we consider the case of the Clayton copula. Assume that $p = 0.1$ and $(U_1, \dots, U_n) \overset{\mathrm{d}}{\sim} \mathrm{Clayton}(t)$ where $t \in (0, 1.5)$. From Figure 2, we can see that $R_n(p)$ decreases as t increases (i.e., dependence gets stronger). Moreover, for $t \geq 1$, $R_n(p)$ is only slightly larger than p, consistent with Theorem 2. On the other hand, as t approaches from 1 to 0, $R_n(p)$ will increase rapidly as the dependence structure becomes closer to independence.

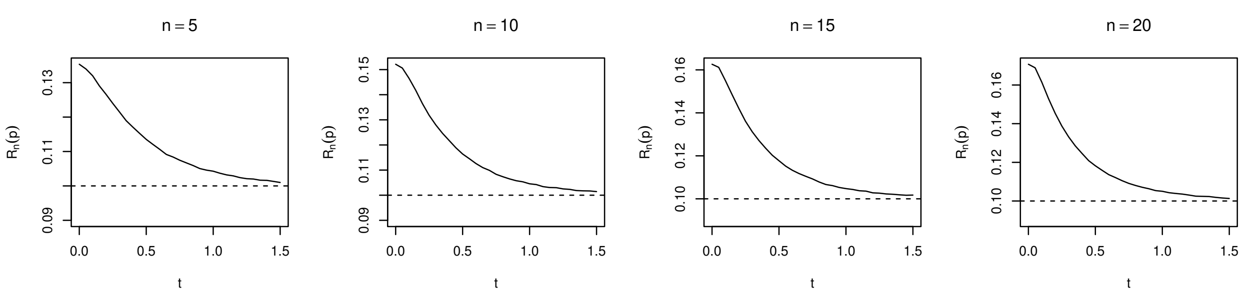

Figure 2: Clayton copula: $R_n(p)$ for $n = 5, 10, 15, 20$, and $t \in [0, 1.5]$ with $p = 0.1$.

As sub-uniformity holds for some Gaussian copulas and Clayton copulas, a natural question is whether it also holds for the more general classes of elliptical copulas or Archimedean copulas. A theoretical answer to this question is not available with the current techniques and we present below some simulation studies. We first compute $R_n(0.1)$ by assuming that $(U_1, \dots, U_n)$ follows an equicorrelated t-copula with degrees of freedom 4 and correlation coefficient $\rho \in [0, 1]$. Figure 3 suggests that sub-uniformity may hold for t-copulas with positive correlation coefficients and $R_n(p)$ decreases as the positive dependence becomes stronger. Next, we compute $R_n(p)$ for the Gumbel

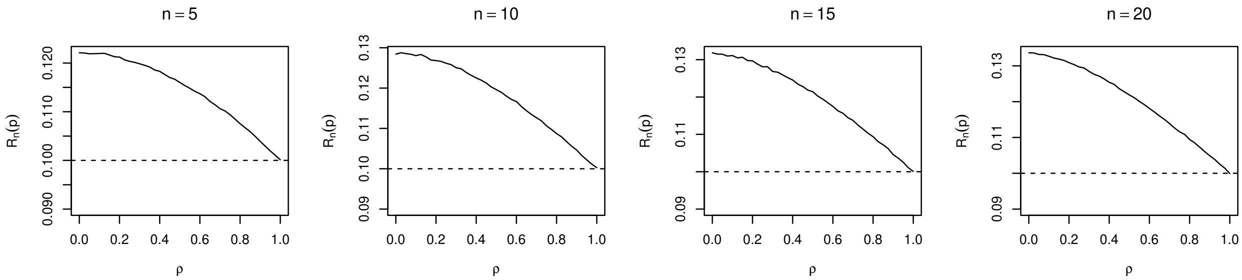

Figure 3: Equicorrelated t-copula: $R_n(p)$ for $n = 5, 10, 15, 20$, and $\rho \in [0, 1]$ with $p = 0.1$.

copula, which falls in the class of Archimedean copulas. The Gumbel copula with parameter $\theta \in (1, \infty)$ is defined as

$$C^G(u_1, \dots, u_n) = \exp\left(-\left((-\log u_1)^\theta + \cdots + (-\log u_2)^\theta\right)^{1/\theta}\right), \quad (u_1, \dots, u_n) \in (0, 1)^n.$$

If $\theta = 1$, C^G is the independence copula. As θ goes to ∞, C^G approaches the comonotonicity copula. Figure 4 plots $R_n(0.1)$ against $\theta \in (1, 10)$. The observation is similar to the case of the Clayton copula: $R_n(p)$ is large when the Gumbel copula is close to the independence copula and is small when the Gumbel copula is close to the comonotonicity copula. Based on these numerical results, we conjecture that sub-uniformity holds for more general elliptical copulas and Archimedean copulas than the ones considered in this paper. Nevertheless, we should keep in mind that sub-uniformity requires $R_n(p) \geq p$ for all $p \in (0, 1)$, and not only for a specific p as in our figures.

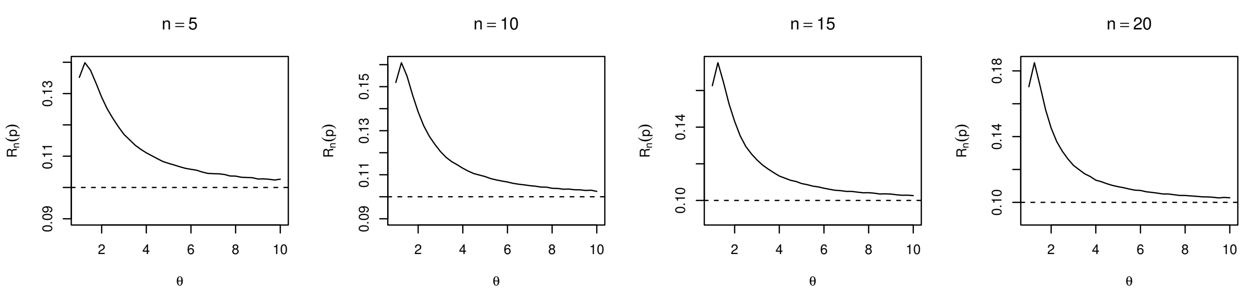

Figure 4: Gumbel copula: $R_n(p)$ for $n = 5, 10, 15, 20$, and $\theta \in [1, 10]$ with $p = 0.1$.

The fifth example presents the case of independent discrete uniform random variables on $\{1/m, \dots, m/m\}$, $m \geq 2$. Figure 5 gives $R_n(p)$ for 10 discrete uniform random variables with different discretization m. We can see that as m increases, $R_n(p)$ for discrete uniform random variables becomes closer to that for uniform random variables. Moreover, if m is large, (14) holds for a wide range of p in $(0,1)$ except for extremely small ones.

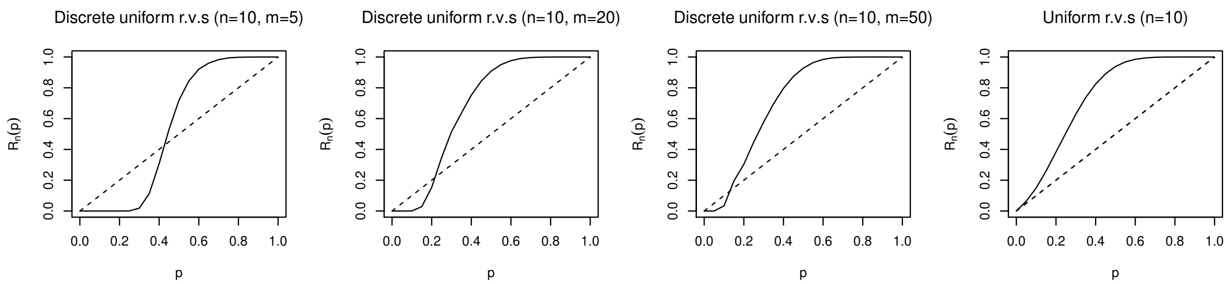

Figure 5: $R_n(p)$ for discrete p-values ($m = 5, 20, 50$) and uniform p-values, with $p \in (0,1)$.

In Figure 6, we numerically compute the threshold $a_{n,p}$ of the harmonic mean p-value for independent p-values and its asymptotic form (13). The thresholds are computed at significance levels 0.01, 0.05, and 0.1, up to 5000 p-values. The results suggest that the asymptotic threshold (13) can be a very good approximation of $a_{n,p}$ for large numbers of p-values. As hinted by the plot for $p = 0.01$, the simulated results are not stable for small significance levels, e.g., 0.005, and are thus omitted here.

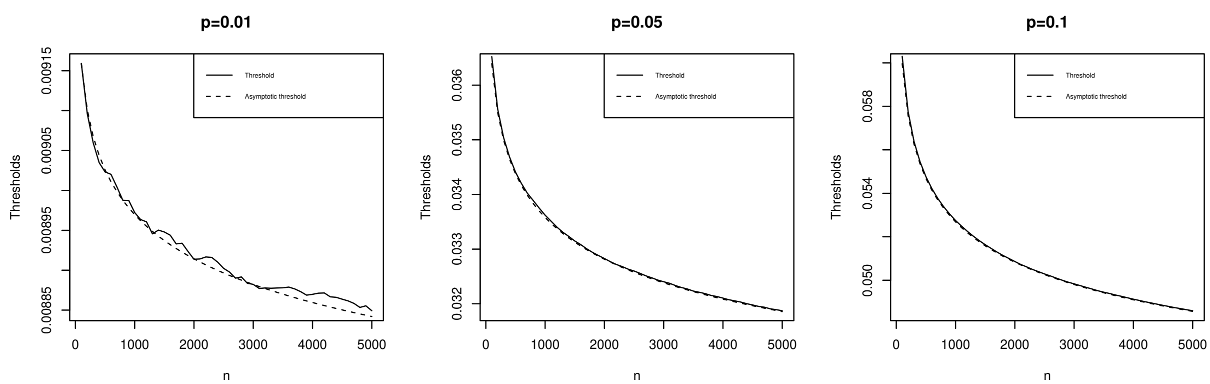

Figure 6: The asymptotic threshold (13) and $a_{n,p}$ of the harmonic mean p-value for independent p-values.

8 Conclusion

Sub-uniformity of generalized means of standard uniform random variables $U_1, \ldots, U_n$ is studied under several dependence assumptions. In particular, sub-uniformity is shown to hold in three cases: (i) negative upper orthant dependence (Proposition 4); (ii) the class of extremal mixture copulas (Theorem 1); (iii) some Clayton copulas (Theorem 2). These dependence structures can be used to construct a wide range of dependence structures for which sub-uniformity holds, as suggested by Propositions 2 and 3. Based on some numerical results, we conjecture that sub-uniformity also holds for Gaussian copulas with positive correlation coefficients. An important implication of sub-uniformity in multiple hypothesis testing is that merging p-values by any r-generalized mean function with $r \leq -1$ is anti-conservative across all significance levels in $(0, 1)$. Although sub-uniformity cannot hold for discrete uniform random variables, using an r-generalized mean function with $r \leq -1$ can still be anti-conservative if the number of discretizations is large (Theorem 3). An asymptotic threshold of the harmonic mean p-value for independent p-values is derived in Proposition 6. As the number of p-values increases, since the asymptotic threshold goes to 0, the harmonic mean p-value will be more anti-conservative if no adjustment is applied.

For the purpose of multiple testing under dependence, due to the anti-conservativeness results found in this paper, we recommend using the Simes method or the Cauchy combination, which are valid under independence and some other dependence assumptions, as well as their variants, over the harmonic mean p-value. The main advantage of the Simes and the Cauchy combination methods is that no correction is needed to make their merged p-values valid for a wide range of dependence assumptions and different numbers of p-values, thus making these methods robust in some sense. On the other hand, correction of the harmonic mean p-value is necessary and may vary according to different dependence structures and numbers of p-values. Theorem 2 also gives a small threshold correction for the harmonic mean p-value under Clayton copulas, suggesting that the harmonic mean p-value may behave quite well under some forms of positive dependence.

We close the paper by noting that, although sub-uniformity can hold under a wide range of dependence structures of $U_1, \ldots, U_n$, there always exists some dependence structure under which sub-uniformity does not hold. For instance, since comonotonicity (i.e., $U_1 = \cdots = U_n$ almost surely) does not maximize the distribution function of the sum of random variables (see Wang et al. (2013) for bounds on the distribution function of the sum), it is always possible to construct a dependence structure of $U_1, \ldots, U_n$ such that $\mathbb{P}(w_1 U_1^r + \cdots + w_n U_n^r \leq t) > \mathbb{P}(U_1^r \leq t)$ for some threshold $t \in \mathbb{R}$ of interest. Therefore, conditions on dependence structures that lead to sub-uniformity or super-uniformity, other than the ones studied in this paper, require further research. This may be achieved by research on either the statistical problem of merging p-values, the risk management problem of quantifying risk aggregation, or their interplay.

References

Alink, S., Löwe, M., and Wüthrich, M. V. (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics*, 35(1):77–95.

Barber, R. F. and Ramdas, A. (2017). The p-filter: Multilayer false discovery rate control for grouped hypotheses. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 79(4):1247–1268.

Bates, S., Candès, E., Lei, L., Romano, Y., and Sesia, M. (2023). Testing for outliers with conformal p-values. *Annals of Statistics*, 51(1):149–178.

Bignozzi, V., Mao, T., Wang, B., and Wang, R. (2016). Diversification limit of quantiles under dependence uncertainty. *Extremes*, 19(2):143–170.

Block, H. W., Savits, T. H., and Shaked, M. (1982). Some concepts of negative dependence. *Annals of Probability*, 10(3):765–772.

Block, H. W., Savits, T. H., and Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters*, 3(2):81–86.

Bodnar, T. and Dickhaus, T. (2014). False discovery rate control under Archimedean copula. *Electronic Journal of Statistics*, 8(2):2207–2241.

Chavez-Demoulin, V., Embrechts, P., and Nešlehová, J. (2006). Quantitative models for operational risk: Extremes, dependence and aggregation. *Journal of Banking and Finance*, 30(10):2635–2658.

Chen, X. and Sarkar, S. K. (2020). On Benjamini–Hochberg procedure applied to mid p-values. *Journal of Statistical Planning and Inference*, 205:34–45.

Chen, Y., Embrechts, P., and Wang, R. (2024). An unexpected stochastic dominance: Pareto distributions, dependence, and diversification. *Operations Research*, forthcoming. doi: 10.1287/opre.2022.0505.

Chen, Y., Liu, P., Tan, K. S., and Wang, R. (2023). Trade-off between validity and efficiency of merging p-values under arbitrary dependence. *Statistica Sinica*, 33(2):851–872.

Chen, Y. and Shneer, S. (2024). Risk aggregation and stochastic dominance for a class of heavy-tailed distributions. *arXiv:2408.15033*.

Chi, Z., Ramdas, A., and Wang, R. (2024). Multiple testing under negative dependence. *Bernoulli*, 31(2):1230–1255.

Denuit, M., Genest, C., and Marceau, É. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25:85–104.

Dickhaus, T. and Gierl, J. (2013). Simultaneous test procedures in terms of p-value copulae. *Proceedings on the 2nd Annual International Conference on Computational Mathematics, Computational Geometry & Statistics (CMCGS 2013)*, pp. 75–80. Global Science and Technology Forum (GSTF).

Dubey, S. D. (1970). Compound Gamma, Beta and F distributions. *Metrika*, 16(1):27–31.

Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg.

Embrechts, P., Lambrigger, D., and Wüthrich, M. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, 12(2):107-127.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (Eds: Dempster), pp. 176–223, Cambridge University Press.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of dependent risks. *Finance and Stochastics*, 10:341–352.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, 37(8):2750–2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, 2(1):25-48.

Embrechts, P., Wang, B., and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19:763–790.

Fang, Y., Chang, C., Park, Y., and Tseng, G. C. (2023). Heavy-tailed distribution for combining dependent p-values with asymptotic robustness. *Statistica Sinica*, 33:1115-1142.

Ferreira, J. and Zwinderman, A. (2006). On the Benjamini–Hochberg method. *Annals of Statistics*, 34(4):1827–1849.

Fisher, R. A. (1948). Combining independent tests of significance. *American Statistician*, 2(30).

Frees, E. W., and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

Gui, L., Jiang, Y., and Wang, J. (2023). Aggregating dependent signals with heavy-tailed combination tests. *arXiv*: 2310.20460.

Hardy, G. H., Littlewood, J. E., and Pólya, G (1934). *Inequalities*. Cambridge University Press.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Annals of Statistics*, 11(1):286–295.

Lin, L., Wang, R., Zhang, R., and Zhao, C. (2024). The checkerboard copula and dependence concepts. *arXiv*: 2404.15023.

Lindley, D. V. and Singpurwalla, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *Journal of Applied Probability*, 23(2):418–431.

Liu, Y. and Xie, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529):393–402.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton University Press.

McNeil, A. J., Neslehova, J. G., and Smith, A. D. (2020). On attainability of Kendall's tau matrices and concordance signatures. *Journal of Multivariate Analysis*, 191:105033.

Nelsen, R. (2006). *An Introduction to Copulas*. Springer, New York, Second Edition.

Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, 1(1):3–25.

Neumann, A., Bodnar, T., Pfeifer, D., and Dickhaus, T. (2019). Multivariate multiple test procedures based on nonparametric copula estimation. *Biometrical Journal*, 61(1):40–61.

Pillai, N. S. and Meng, X.-L. (2016). An unexpected encounter with Cauchy and Lévy. *Annals of Statistics*, 44 (5):2089–2097

Rubin-Delanchy, P., Heard, N. A., and Lawson, D. J. (2019). Meta-analysis of mid-p-values: Some new results based on the convex order. *Journal of the American Statistical Association*, 114(527):1105–1112.

Rustamov, R. M. and Klosowski, J. T. (2020). Kernel mean embedding based hypothesis tests for comparing spatial point patterns. *Spatial Statistics*, 38:100459.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Routledge.

Samuel-Cahn, E. (1996). Is the Simes improved Bonferroni procedure conservative? *Biometrika*, 83(4):928–933.

Sarabia, J. M., Gómez-Déniz, E., Prieto, F., and Jordá, V. (2016). Risk aggregation in multivariate dependent pareto distributions. *Insurance: Mathematics and Economics*, 71:154–163.

Sarkar, S. K. (1998). Some probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture. *Annals of Statistics*, 26(2):494–504.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.

Vovk, V., Gammerman, A. and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York.

Vovk, V., Wang, B., and Wang, R. (2022). Admissible ways of merging p-values under arbitrary dependence. *Annals of Statistics*, 50(1):351–375.

Vovk, V. and Wang, R. (2020). Combining p-values via averaging. *Biometrika*, 107(4):791–808.

Wang, R., Peng, L., and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417.

Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences*, 116(4):1195–1200.

Xiong, P. and Hu, T. (2022). On Samuel's p-value model and the Simes test under dependence. *Statistics and Probability Letters*, 187:109509.