

Combining p-values via averaging

BY VLADIMIR VOVK

*Department of Computer Science
Royal Holloway, University of London
Egham, Surrey, UK
v.vovk@rhul.ac.uk*

5

RUODU WANG

*Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada
wang@uwaterloo.ca*

10

SUMMARY

This paper proposes general methods for the problem of multiple testing of a single hypothesis, with a standard goal of combining a number of p-values without making any assumptions about their dependence structure. An old result by Rüschendorf and, independently, Meng implies that the p-values can be combined by scaling up their arithmetic mean by a factor of 2 (and no smaller factor is sufficient in general). A similar result about the geometric mean (Mattner) replaces 2 by e . Based on more recent developments in mathematical finance, specifically, robust risk aggregation techniques, we extend these results to generalized means; in particular, we show that K p-values can be combined by scaling up their harmonic mean by a factor of $\ln K$ (asymptotically as $K \rightarrow \infty$). This leads to a generalized version of the Bonferroni–Holm procedure. We also explore methods using weighted averages of p-values. Finally, we discuss the efficiency of various methods of combining p-values and how to choose a suitable method in light of data and prior information.

15

20

25

Some key words: Hypothesis testing; Multiple hypothesis testing; Multiple testing of a single hypothesis; Robust risk aggregation

1. INTRODUCTION

Suppose we are testing the same hypothesis using $K \geq 2$ different statistical tests and obtain p-values p_1, \dots, p_K . How can we combine them into a single p-value?

30

Some early answers to this question include those of Tippett [41], Pearson [32] and Fisher [8]. These solutions assume that the p-values are independent (in practice, obtained from independent test statistics), whereas we would like to avoid any assumptions besides all p_1, \dots, p_K being bona fide p-values. Fisher’s method has been extended to dependent p-values in, e.g., [4, 20], but the combined p-values obtained in those papers are approximate; in this paper we are interested only in precise or conservative p-values. A summary of combination methods for p-values can be found in e.g., Oosterhoff [30] and Cousins [5]; see also the more specialized review Owen [31].

35

Without assuming any particular dependence structure among p-values, the simplest way of combining them is the Bonferroni method

$$F(p_1, \dots, p_K) := K \min(p_1, \dots, p_K) \quad (1)$$

(when $F(p_1, \dots, p_K)$ exceeds 1 it can be replaced by 1, but we usually ignore this trivial step). Albeit $F(p_1, \dots, p_K)$ is a p-value (see Section 2 for a precise definition of a *p-value*), it has been argued that in some cases it is overly conservative. Rüger [34] extends the Bonferroni method by showing that, for any fixed $k \in \{1, \dots, K\}$,

$$F(p_1, \dots, p_K) := \frac{K}{k} p_{(k)} \quad (2)$$

is a p-value, where $p_{(k)}$ is the k th smallest p-value among p_1, \dots, p_K ; see [29] for a simpler exposition. Hommel [15] develops this by showing that

$$F(p_1, \dots, p_K) := \left(1 + \frac{1}{2} + \dots + \frac{1}{K}\right) \min_{k=1, \dots, K} \frac{K}{k} p_{(k)} \quad (3)$$

is also a p-value. In the case of independent p_1, \dots, p_n , Simes [40] improves (3) by removing the first factor on the right-hand side of (3).

A natural way to combine K p-values is to average them, by using $\bar{p} := (p_1 + \dots + p_K)/K$ (at least when the tests have similar power). Unfortunately, \bar{p} is not necessarily a p-value. An old result by Rüschendorf [35, Theorem 1] shows that $2\bar{p}$ is a p-value; moreover, the factor of 2 cannot be improved in general. In the statistical literature this result was rediscovered by Meng [28, Theorem 1].

In this paper we turn to a general notion of the mean as axiomatized by Andrei Kolmogorov ([19]) and promoted in the context of combining independent p -values by Lipták [22]. We adapt various results of robust risk aggregation [6, 1, 7, 43, 18] to combining p -values by averaging them in Kolmogorov's wider sense. In particular, to obtain a p -value from given p -values p_1, \dots, p_K , it is sufficient to multiply their geometric mean by e , as noticed by [26], and to multiply their harmonic mean by $e \log K$ for $K > 2$. More generally, we consider the mean $M_{r,K}(p_1, \dots, p_K)$, sometimes referred to as generalized mean, defined by $\{(p_1^r + \dots + p_K^r)/K\}^{1/r}$ for $r \in [-\infty, \infty]$, and derive values of $a_{r,K}$ making $a_{r,K} M_{r,K}$ a merging function. In particular, our results cover the Bonferroni method (1), which corresponds to $M_{-\infty,K}(p_1, \dots, p_K) = K \min(p_1, \dots, p_K)$ (see, e.g., [13], (2.3.1)).

The median is also sometimes regarded as a kind of average. Rüger's (2), applied to $k := \lceil K/2 \rceil$, says that p-values can be combined by scaling up their median by a factor of 2 (exactly for even K and approximately for large odd K). Therefore, we have the same factor of 2 as in Rüschendorf's [35] result. (Taking $k = \lfloor (K+1)/2 \rfloor = \lceil K/2 \rceil$ is suggested in [25, Section 1.1].) More generally, the α quantile $p_{(\lceil \alpha K \rceil)}$ becomes a p-value if multiplied by $1/\alpha$.

It is often possible to automatically transform results about multiple testing of a single hypothesis into results about testing multiple hypotheses; the standard procedures are Marcus et al.'s ([24]) closed testing procedure and its modification by Hommel [16]. In particular, when applied to the Bonferroni method the closed testing procedure gives the well-known procedure due to Holm [14], which we will refer to as the Bonferroni–Holm procedure; see, e.g., [16, 17] for its further applications. The methods we develop in Section 3 can similarly be adapted to the testing of multiple hypotheses, as we briefly discuss in the Online Supplement. All proofs are also given in the Online Supplement.

Throughout we will use the following terminology. A function $F : [0, 1] \rightarrow [0, \infty)$ is said to be increasing (resp. decreasing) if $F(x_1) \leq F(x_2)$ (resp. $F(x_1) \geq F(x_2)$) whenever $x_1 \leq x_2$. A function $F : [0, 1]^K \rightarrow [0, \infty)$ is increasing (resp. decreasing) if it is increasing (resp. decreasing) in each of its arguments. A function is strictly increasing or strictly decreasing when these conditions hold with strict inequalities. 80

2. MERGING FUNCTIONS

A *p-variable* is a random variable P that satisfies

$$\mathbb{P}(P \leq \epsilon) \leq \epsilon, \quad \forall \epsilon \in (0, 1). \quad (4)$$

The values taken by a p-variable are *p-values* (allowed to be conservative). In Section 1 the expression “p-value” was loosely used to refer to p-variables as well. A *merging function* is an increasing Borel function $F : [0, 1]^K \rightarrow [0, \infty)$ such that $F(U_1, \dots, U_K)$ is a p-variable for any choices of random variables U_1, \dots, U_K (on the same probability space, which can be arbitrary) distributed uniformly on $[0, 1]$. Without loss of generality we can assume that U_1, \dots, U_K are defined on the same atomless probability space, which is fixed throughout the paper (cf. [9, Proposition A.27]). Let \mathcal{U} be the set of all uniformly distributed random variables (on our probability space). Using the notation \mathcal{U} , an increasing Borel function $F : [0, 1]^K \rightarrow [0, \infty)$ is a merging function if, for each $\epsilon \in (0, 1)$, 85

$$\sup \{ \mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \mid U_1, \dots, U_K \in \mathcal{U} \} \leq \epsilon \quad (5)$$

We say that a merging function F is *precise* if, for each $\epsilon \in (0, 1)$, 95

$$\sup \{ \mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \mid U_1, \dots, U_K \in \mathcal{U} \} = \epsilon. \quad (6)$$

Remark 1. The requirement that a merging function be Borel does not follow automatically from the requirement that it be increasing: see the remark after Theorem 4.4 in [11] (Theorem 4.4 itself says that every increasing function on $[0, 1]^K$ is Lebesgue measurable).

It may be practically relevant to notice that, for any merging function F , $F(P_1, \dots, P_K)$ is a p-variable whenever P_1, \dots, P_K are p-variables (on the same probability space). Indeed, for each $k \in \{1, \dots, K\}$ we can define a uniformly distributed (see, e.g., [36, Proposition 2.1]) random variable $U_k \leq P_k$ by 100

$$U_k(\omega) := \mathbb{P}(P_k < P_k(\omega)) + \Theta(\omega) \mathbb{P}(P_k = P_k(\omega)), \quad \omega \in \Omega,$$

where Θ is a random variable distributed uniformly on $[0, 1]$ and independent of P_1, \dots, P_K , and Ω is the underlying probability space extended (if required) to carry such a Θ ; we then have 105

$$\mathbb{P}(F(P_1, \dots, P_K) \leq \epsilon) \leq \mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \leq \epsilon, \quad \forall \epsilon \in (0, 1).$$

Therefore, combining p-values can be carried out in multiple layers (although it may make the final combined p-value overly conservative); we will discuss this further in Subsection 6.3.

3. COMBINING P-VALUES BY SYMMETRIC AVERAGING

In this section we present our methods of combining p-values via averaging. A general notion of averaging, axiomatized by Kolmogorov [19], is

$$M_{\phi,K}(p_1, \dots, p_K) := \psi \left(\frac{\phi(p_1) + \dots + \phi(p_K)}{K} \right), \quad (7)$$

where $\phi : [0, 1] \rightarrow [-\infty, \infty]$ is a continuous strictly monotonic function and ψ is its inverse (with the domain $\phi([0, 1])$). For example, arithmetic mean corresponds to the identity function $\phi(p) = p$, geometric mean corresponds to $\phi(p) = \ln p$, and harmonic mean corresponds to $\phi(p) = 1/p$.

The problem of finding precise p-value merging functions based on the averaging method in (7) is closely connected to *robust risk aggregation*, an active topic in mathematical finance (see Remark 2 below). The origin of this field lies in a problem posed by Kolmogorov (see, e.g., [23]) about bounds on the distribution function of the sum of random variables. To establish the main results in this paper, we will use many recent results in robust risk aggregation, in particular, many results in [6, 44, 1, 7, 43, 18].

Below, we first briefly discuss the connection between the two problems in Section 3.1, and then present our main results in Section 3.2.

3.1. Quantiles and robust risk aggregation

We start from a simple result (Lemma 1 below) that translates probability statements about merging functions into corresponding statements about quantiles. This result will allow us to freely use some recent results in the literature on robust risk aggregation. The proof of Lemma 1 is given on page 19 in Appendix A.

Define the left α -quantile of a random variable X for $\alpha \in (0, 1]$,

$$q_\alpha(X) := \sup\{x \in \mathbb{R} : \mathbb{P}(X \leq x) < \alpha\},$$

and the right α -quantile of X for $\alpha \in [0, 1)$,

$$q_\alpha^+(X) := \sup\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \leq \alpha\}.$$

Notice that $q_1(X)$ is the essential supremum of X and $q_0^+(X)$ is the essential infimum of X . For a function $F : [0, 1]^K \rightarrow [0, \infty)$ and $\alpha \in (0, 1)$, write

$$\underline{q}_\alpha(F) := \inf\{q_\alpha(F(U_1, \dots, U_K)) \mid U_1, \dots, U_K \in \mathcal{U}\}.$$

LEMMA 1. For an increasing Borel function $F : [0, 1]^K \rightarrow [0, \infty)$:

- (a) F is a merging function if and only if $\underline{q}_\epsilon(F) \geq \epsilon$ for all $\epsilon \in (0, 1)$;
- (b) F is a precise merging function if and only if $\underline{q}_\epsilon(F) = \epsilon$ for all $\epsilon \in (0, 1)$.

Remark 2. This remark discusses how the problem of combining p-values is related to robust risk aggregation in the field of mathematical finance. In quantitative risk management, the term *robust risk aggregation* refers to evaluating the value of a *risk measure* of an aggregation of risks X_1, \dots, X_K with specified marginal distributions and unspecified dependence structure. More specifically, if the risk measure is chosen as a quantile q_α , known as *Value-at-Risk* and very popular in finance, the quantities of interest are typically

$$\bar{q} := \sup\{q_\alpha(X_1 + \dots + X_n) \mid X_1 \sim F_1, \dots, X_n \sim F_n\}$$

and

$$\underline{q} := \inf \{q_\alpha(X_1 + \cdots + X_n) \mid X_1 \sim F_1, \dots, X_n \sim F_n\},$$

where F_1, \dots, F_n denote the prespecified marginal distributions of the risks. The motivation behind this problem is that, in practical applications of banking and insurance, the dependence structure among risks to aggregate is very difficult to accurately model, as compared with the corresponding marginal distributions. The interval $[\underline{q}, \bar{q}]$ thus represents all possible values of the aggregate risk measure given the marginal information. A more detailed introduction to this topic can be found in [27, Section 8.4.4] and [37, Chapter 4]. Via Lemma 1, the quantities \bar{q} and \underline{q} are obviously closely related to the problem of combining p-values. There are few explicit formulas for \bar{q} and \underline{q} but fortunately some do exist in the literature, and they become useful in our study of merging functions.

3.2. Main results

In this section, we present our main results on merging functions. The proofs of these results along with a few auxiliary results, can be found in Appendix A. We mostly focus on an important special case of (7), namely,

$$M_{r,K}(p_1, \dots, p_K) := \left(\frac{p_1^r + \cdots + p_K^r}{K} \right)^{1/r}, \quad (8)$$

where $r \in \mathbb{R} \setminus \{0\}$ and the following standard conventions are used: $0^c := \infty$ for $c < 0$, $0^c := 0$ for $c > 0$, $\infty + c := \infty$ for $c \in \mathbb{R} \cup \{\infty\}$, and $\infty^c := 0$ for $c < 0$. The case $r = 0$ (considered in [26]) is treated separately (as the limit as $r \rightarrow 0$):

$$M_{0,K}(p_1, \dots, p_K) := \exp \left(\frac{\ln p_1 + \cdots + \ln p_K}{K} \right) = \left(\prod_{k=1}^K p_k \right)^{1/K},$$

where, as usual, $\ln 0 := -\infty$, $-\infty + c := -\infty$ for $c \in \mathbb{R} \cup \{-\infty\}$, and $\exp(-\infty) := 0$. It is also natural to set

$$\begin{aligned} M_{\infty,K}(p_1, \dots, p_K) &:= \max(p_1, \dots, p_K), \\ M_{-\infty,K}(p_1, \dots, p_K) &:= \min(p_1, \dots, p_K). \end{aligned}$$

The most important special cases of $M_{r,K}$ are perhaps those corresponding to $r = -\infty$ (minimum), $r = -1$ (harmonic mean), $r = 0$ (geometric mean), $r = 1$ (arithmetic mean), and $r = \infty$ (maximum); the cases $r \in \{-1, 0, 1\}$ are known as Platonic means.

Our main aim is to identify merging functions of the form

$$a_{r,K} M_{r,K}(p_1, \dots, p_K), \quad r \in [-\infty, \infty], \quad K = 2, 3, \dots, \quad (9)$$

where $a_{r,K}$ is a constant, hopefully making the merging function (9) precise. In cases where an explicit formula of $a_{r,K}$ for (9) to be precise is not available, we obtain an asymptotically sharp bound in explicit form. The main results are summarized in Table 1, where a family F_K , $K = 2, 3, \dots$, of merging functions on $[0, 1]^K$ is called *asymptotically precise* if, for any $a \in (0, 1)$, the function aF_K is not a merging function for a large enough K ; in other words, this family of merging functions cannot be improved by a constant multiplier. It is well known [13, Theorem 16] that

$$M_{r_1,K} \leq M_{r_2,K} \text{ on } [0, 1]^K \text{ if } r_1 \leq r_2; \quad (10)$$

therefore, the constant $a_{r,K}$ making (9) a precise merging function should be generally decreasing in r .

Table 1: The main results of Section 3: examples of merging functions, all of them precise or asymptotically precise (except for the case $r = -1$ where the asymptotic formula is not a merging function for finite K)

range of r	merging function		special case	claimed in
$r = \infty$	$M_{r,K}$	precise	maximum	
$r \in [K-1, \infty)$	$K^{1/r} M_{r,K}$	precise		Prop. 3
$r \in [\frac{1}{K-1}, K-1]$	$(r+1)^{1/r} M_{r,K}$	precise	arithmetic	Prop. 2
$r \in (-1, \infty]$	$(r+1)^{1/r} M_{r,K}$	asymptotically precise		Prop. 1
$r = 0$	$e M_{r,K}$	asymptotically precise	geometric	Prop. 4
$r = -1$	$e(\ln K) M_{r,K}$ $(\ln K) M_{r,K}$	not precise (asymptotic formula)	harmonic	Prop. 6
$r \in (-\infty, -1)$	$\frac{r}{r+1} K^{1+1/r} M_{r,K}$	asymptotically precise		Prop. 5
$r = -\infty$	$K M_{r,K}$	precise	Bonferroni	

Two of our results, Theorems 1 and 2, are general in the sense of describing properties of Kolmogorov's general averaging function (7); Theorem 1 covers the case of integrable ϕ (and therefore, the case $r > -1$ in (8)), and Theorem 2 covers the case of non-integrable ϕ (and therefore, the case $r \leq -1$ in (8)). Propositions 1-6 give results shown in Table 1. We further show in Proposition 7 that, for the definition of a p-variable in (4), one can replace “for all ϵ ” by “for some ϵ ” in all averaging methods based on (8).

THEOREM 1. *Suppose a continuous strictly monotonic $\phi : [0, 1] \rightarrow [-\infty, \infty]$ is integrable, i.e., $\int_0^1 |\phi(u)| du < \infty$. Then, for any $K \in \{2, 3, \dots\}$ and any $\epsilon > 0$,*

$$\mathbb{P} \left(M_{\phi,K}(p_1, \dots, p_K) \leq \psi \left(\frac{1}{\epsilon} \int_0^\epsilon \phi(u) du \right) \right) \leq \epsilon. \quad (11)$$

As we stated it, Theorem 1 gives a critical region of size ϵ . An alternative statement is that $\Psi^{-1}(M_{\phi,K})$ is a merging function, where the strictly increasing function Ψ is defined by

$$\Psi(\epsilon) := \psi \left(\frac{1}{\epsilon} \int_0^\epsilon \phi(u) du \right), \quad \epsilon \in (0, 1). \quad (12)$$

In what follows, the expression $(r+1)^{1/r}$ is understood to be $e = \lim_{r \rightarrow 0} (r+1)^{1/r}$ when $r = 0$ and $1 = \lim_{r \rightarrow \infty} (r+1)^{1/r}$ when $r = \infty$.

Using Theorem 1 with $\phi(u) = u^r$, we see that $(r+1)^{1/r} M_{r,K}$ is a merging function for $r > -1$. Moreover, we can show the constant $(r+1)^{1/r}$ cannot be improved in general.

PROPOSITION 1. *Let $r \in (-1, \infty]$. Then $(r+1)^{1/r} M_{r,K}$, $K = 2, 3, \dots$, is a family of merging functions and it is asymptotically precise.*

In particular, Proposition 1 implies that for the geometric mean, it has to be multiplied by the constant factor e , which cannot be improved in general for large K . The next proposition characterizes the condition under which the merging function in Proposition 1 is precise.

PROPOSITION 2. For $r \in (-1, \infty)$ and $K \in \{2, 3, \dots\}$, the merging function $(r + 1)^{1/r} M_{r,K}$ is precise if and only if $r \in [\frac{1}{K-1}, K-1]$. 200

The most straightforward yet relevant example of Proposition 2, the arithmetic average multiplied by 2, namely,

$$M_{1,K}(p_1, \dots, p_K) := \frac{2}{K} \sum_{k=1}^K p_k,$$

is a precise merging function for all $K \geq 2$, a result obtained by [35]. As another special case of Proposition 2, the scaled quadratic average multiplied by $\sqrt{3}$, namely $\sqrt{3}M_{2,K}$, 205 is a merging function, and it is precise if and only if $K \geq 3$.

In the case $r \geq 1$, the merging function in Proposition 2 can be modified in an explicit way such that it remains precise even for $r > K-1$.

PROPOSITION 3. For $K \in \{2, 3, \dots\}$ and $r \in [1, \infty)$, the function $\min(r+1, K)^{1/r} M_{r,K}$ is a precise merging function. 210

Because of the importance of geometric mean as one of the Platonic means, the following result gives a precise (albeit somewhat implicit) expression for the corresponding precise merging function.

PROPOSITION 4. For each $K \in \{2, 3, \dots\}$, $a_K^G M_{0,K}$ is a precise merging function, where 215

$$a_K^G := \frac{1}{c_K} \exp(-(K-1)(1 - Kc_K))$$

and c_K is the unique solution to the equation

$$\ln(1/c - (K-1)) = K - K^2 c \quad (13)$$

over $c \in (0, 1/K)$. Moreover, $a_K^G \leq e$ and $a_K^G \rightarrow e$ as $K \rightarrow \infty$.

Table 2 reports several values of a_K^G/e calculated numerically and suggests that in practice there is no point in improving the factor e for $K \geq 5$.

Table 2: Numeric values of a_K^G/e for the geometric mean

K	a_K^G/e	K	a_K^G/e	K	a_K^G/e
2	0.7357589	5	0.9925858	10	0.9999545
3	0.9286392	6	0.9974005	15	0.9999997
4	0.9779033	7	0.9990669	20	1.0000000

The condition $r > -1$ in Proposition 1 ensures that the term (12) is finite, and also 220 that the condition $\int_0^1 |\phi(u)| du < \infty$ in Theorem 1 is satisfied. However, the condition rules out the harmonic mean (for which $r = -1$) and the minimum ($r = -\infty$). The next

simple corollary of another known result (Theorem 4.2 of [6]; see also Theorem 2.3 of [33]) will cover these cases as well.

225 **THEOREM 2.** *Suppose $\phi : [0, 1] \rightarrow [-\infty, \infty]$ is a strictly decreasing continuous function satisfying $\phi(0) = \infty$. Then, for any $\epsilon \in (0, 1)$ such that $\phi(\epsilon) \geq 0$,*

$$\mathbb{P}(M_{\phi,K}(p_1, \dots, p_K) \leq \epsilon) \leq \inf_{t \in (0, \phi(\epsilon)]} \frac{\int_{\phi(\epsilon)-t}^{\phi(\epsilon)+(K-1)t} \psi(u) du}{t}. \quad (14)$$

As $t \rightarrow 0$, the upper bound in (14) is not informative since, for $t \approx 0$,

$$\frac{\int_{\phi(\epsilon)-t}^{\phi(\epsilon)+(K-1)t} \psi(u) du}{t} \approx \frac{Kt\psi(\phi(\epsilon))}{t} = K\epsilon,$$

which is dominated by the Bonferroni bound. On the other hand, the upper bound is informative when $t = \phi(\epsilon)$ provided the integral is convergent. For example, we can
230 see that for $r < -1$, $\frac{r}{r+1} K^{1+1/r} M_{r,K}$ is a merging function. In what follows, the term $r/(r+1)$ should be understood as its limit 1 when $r = -\infty$.

PROPOSITION 5. *Let $r \in [-\infty, -1)$. Then $\frac{r}{r+1} K^{1+1/r} M_{r,K}$, $K = 2, 3, \dots$, is a family of merging functions and it is asymptotically precise.*

Proposition 5 includes the Bonferroni bound (1) as special case: for $r := -\infty$, we obtain
235 that $KM_{-\infty,K}$ is a merging function. On the other hand, Proposition 5 does not cover the case $r = -1$ of harmonic mean directly, but easily implies a bound leading to the merging function $e \ln KM_{-1,K}$, which turns out to be not so crude.

PROPOSITION 6. *Set $a_K^H := \frac{(y_K + K)^2}{(y_K + 1)K}$, $K > 2$, where y_K is the unique solution to the equation*

$$y^2 = K((y + 1) \ln(y + 1) - y), \quad y \in (0, \infty).$$

240 *Then $a_K^H M_{-1,K}$ is a precise merging function. Moreover, $a_K^H \leq e \ln K$ and $a_K^H / \ln K \rightarrow 1$ as $K \rightarrow \infty$.*

Even though $a_K^H / \ln K \rightarrow 1$, the rate of convergence is very slow, and $a_K^H / \ln K > 1$ for moderate values of K . In practice, it might be better to use the conservative merging function $(e \ln K) M_{-1,K}$. Table 3 reports several values of $a_K^H / \ln K$ calculated numerically.
245 For instance, for $K \geq 10$, one may use $(2 \ln K) M_{-1,K}$, and for $K \geq 50$, one may use $(1.7 \ln K) M_{-1,K}$.

Table 3: Numeric values of $a_K^H / \ln K$ for the harmonic mean

K	$a_K^H / \ln K$	K	$a_K^H / \ln K$	K	$a_K^H / \ln K$
3	2.499192	10	1.980287	100	1.619631
4	2.321831	20	1.828861	200	1.561359
5	2.214749	50	1.693497	400	1.514096

The main emphasis of this section has been on characterizing $a > 0$ such that $F := aM_{r,K}$ is a merging function, or a precise merging function. Recall that $F : [0, 1]^K \rightarrow [0, \infty)$ is a merging function if and only if (5) holds for all $\epsilon \in (0, 1)$, and that F is a

precise merging function if and only if (6) holds for all $\epsilon \in (0, 1)$. The next proposition shows that in both statements “for all” can be replaced by “for some” if $F = aM_{r,K}$. A practical implication is that even if an applied statistician is interested in the property of validity (4) only for specific values of ϵ (such as 0.01 or 0.05) and would like to use $aM_{r,K}$ as a merging function, she is still forced to ensure that (4) holds for all ϵ . 250

PROPOSITION 7. For any $a > 0$, $r \in [-\infty, \infty]$, and $K \in \{2, 3, \dots\}$: 255

- (a) $F := aM_{r,K}$ is a merging function if and only if (5) holds for some $\epsilon \in (0, 1)$;
- (b) $F := aM_{r,K}$ is a precise merging function if and only if (6) holds for some $\epsilon \in (0, 1)$.

4. COMBINING P-VALUES BY WEIGHTED AVERAGING

In this section we will briefly consider a more general notion of averaging:

$$M_{\phi, \mathbf{w}}(p_1, \dots, p_K) := \psi(w_1\phi(p_1) + \dots + w_K\phi(p_K))$$

in the notation of (7), where $\mathbf{w} = (w_1, \dots, w_K) \in \Delta_K$ is an element of the standard K -simplex 260

$$\Delta_K := \{(w_1, \dots, w_K) \in [0, 1]^K \mid w_1 + \dots + w_K = 1\}.$$

One might want to use a weighted average in a situation where some of p-values are based, e.g., on bigger experiments, and then we might want to take them with bigger weights. Intuitively, the weights reflect the prior importance of the p-values (see, e.g., [3, p. 5] for further details). 265

Much fewer mathematical results in the literature are available for asymmetric risk aggregation. For this reason, we will concentrate on the easier integrable case, namely, $r > -1$. Theorem 1 can be generalized as follows.

THEOREM 3. Suppose a continuous strictly monotonic $\phi : [0, 1] \rightarrow [-\infty, \infty]$ is integrable and $\mathbf{w} \in \Delta_K$. Then, for any $\epsilon > 0$, 270

$$\mathbb{P}\left(M_{\phi, \mathbf{w}}(p_1, \dots, p_K) \leq \psi\left(\frac{1}{\epsilon} \int_0^\epsilon \phi(u) du\right)\right) \leq \epsilon.$$

Similarly to (8), we set

$$M_{r, \mathbf{w}}(p_1, \dots, p_K) := (w_1 p_1^r + \dots + w_K p_K^r)^{1/r}$$

for $r \in \mathbb{R}$ and $\mathbf{w} = (w_1, \dots, w_K) \in \Delta_K$. We can see that Proposition 1 still holds when $M_{r,K}$ is replaced by $M_{r, \mathbf{w}}$, for any $r \in \mathbb{R}$ and $\mathbf{w} \in \Delta_K$. This is complemented by the following proposition, which is the weighted version of Proposition 2.

PROPOSITION 8. For $\mathbf{w} = (w_1, \dots, w_K) \in \Delta_K$ and $r \in (-1, \infty)$, the merging function $(r+1)^{1/r} M_{r, \mathbf{w}}$ is precise if and only if $w \leq 1/2$ and $r \in [\frac{w}{1-w}, \frac{1-w}{w}]$, where $w := \max_{k=1, \dots, K} w_k$. 275

Next we generalize Proposition 3 to non-uniform weights.

PROPOSITION 9. For $\mathbf{w} = (w_1, \dots, w_K) \in \Delta_K$ and $r \in [1, \infty)$, the function $\min(r+1, \frac{1}{w})^{1/r} M_{r, \mathbf{w}}$ is a precise merging function, where $w := \max_{k=1, \dots, K} w_k$. 280

An interesting special case of Proposition 9 is for $r = 1$ (weighted arithmetic mean). If $w \leq 1/2$, i.e., no single experiment outweighs the total of all the other experiments,

the optimal multiplier for the weighted average is 2, exactly as in the case of the arithmetic average. If $w > 1/2$, i.e., there is a single experiment that outweighs all the other experiments, our merging function is, assuming $w_1 = w$,

$$\frac{1}{w} M_{1,\mathbf{w}}(p_1, \dots, p_K) = p_1 + \sum_{k=2}^K \frac{w_k}{w} p_k.$$

It is obtained by adding weighted adjustments to the p-value obtained from the most important experiment.

5. EFFICIENCY OF MERGING FUNCTIONS

So far we have emphasized the *validity* of our methods of combining p-values: the combined p-value P is guaranteed to satisfy (4) under the null hypothesis $p_1, p_2, \dots, p_K \in \mathcal{U}$. In this section we will discuss their *efficiency*: under alternative hypotheses, we would like the combined p-value to be small.

We will be interested in asymptotic results as $K \rightarrow \infty$. Recall that our merging functions are

$$P_{r,K}(p_1, \dots, p_K) := a_{r,K} M_{r,K}(p_1, \dots, p_K),$$

where $r \in [-\infty, \infty]$, $P_{r,K}$ is the combined p-value, and

$$a_{r,K} := \begin{cases} (r+1)^{1/r} & \text{if } r \in (-1, \infty] \\ e \ln K & \text{if } r = -1 \\ \frac{r}{r+1} K^{1+1/r} & \text{if } r \in [-\infty, -1), \end{cases}$$

where for $r = 0$, $r = \infty$ and $r = -\infty$ one uses the limiting values of e , 1, and K , respectively. Notice that we do not truncate $P_{r,K}$ by replacing it with $\min(P_{r,K}, 1)$ and interpret large values of $P_{r,K}$, $P_{r,K} \gg 1$, as indicators of the weakness of the merging function: they show us how far it is from being useful. This will be discussed further after Proposition 10.

The rest of this section consists of two parts. In Subsection 5.1 we consider an infinite sequence of IID p-values p_1, p_2, \dots and analyze the performance of our merging functions on p_1, \dots, p_K as $K \rightarrow \infty$. Of course, we are not interested in independent p_1, p_2, \dots per se; after all, if we know the p-values to be independent, we should use much more efficient methods, such as Fisher's, that assume independence. The result in Subsection 5.1 will serve as a basic tool for the analysis of the general symmetric framework, considered in Subsection 5.2. We assume only the exchangeability of p_1, p_2, \dots , and according to de Finetti's theorem every exchangeable probability measure on $[0, 1]^\infty$ is a mixture of IID components (in particular, when we look at the realized sequence of p-values, we can only draw conclusions about the realized IID component).

5.1. IID p-values

In this subsection we assume that the p-values p_1, p_2, \dots are generated independently from the same probability measure Q on $[0, 1]$. If the support of Q is finite and does not include 0, the asymptotic performance of $P_{r,K}$ is poor for $r < -1$, but some of $r > -1$ may have reasonable performance: see Figure 1 for an example.

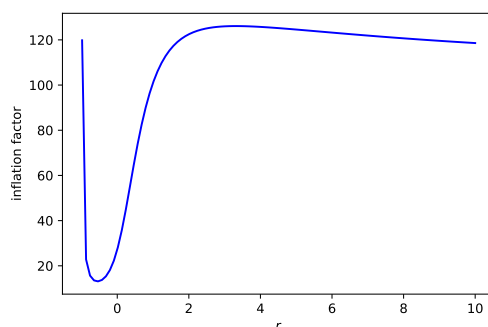


Fig. 1: The asymptotic inflation factor $\lim_{K \rightarrow \infty} P_{r,K}/p$ as function of r when each p -value p_k is either p or $100p$ (with equal probabilities), for some $p \in (0, 0.01)$. (The trivial factor is 100, and the ideal, but unattainable, factor is 1.)

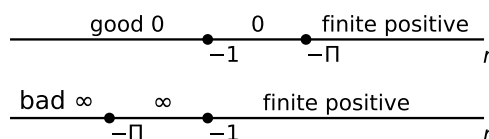


Fig. 2: The combined p -value for different r in the cases $\Pi < 1$ (top) and $\Pi > 1$ (bottom).

In general we impose integrability conditions on p_k . Set

$$\Pi = \Pi(Q) := \sup \left\{ m \in [0, \infty) \mid \mathbb{E}(p_1^{-m}) = \int p^{-m} Q(dp) < \infty \right\} \in [0, \infty].$$

By the Lyapunov inequality [12, Theorem 3.2.5], $\mathbb{E}(p_1^{-m}) < \infty$ for $m < \Pi$ and $\mathbb{E}(p_1^{-m}) = \infty$ for $m > \Pi$. The value of Π reflects the asymptotic power of the p -values; we will discuss this further at the end of this subsection.

The following proposition describes the performance of our merging functions for various values of r (except for the two critical values, -1 and $-\Pi$). To analyze the asymptotic behaviour of $P_{r,K}$ as $K \rightarrow \infty$, we will write $P_{r,K} \approx K^a$ for $a \in \mathbb{R}$ if

$$\liminf_{K \rightarrow \infty} \frac{P_{r,K}}{K^b} = 0 \quad \text{for } b > a \quad \text{and} \quad \lim_{K \rightarrow \infty} \frac{P_{r,K}}{K^b} = \infty \quad \text{for } b < a \quad \text{a.s.} \quad (15)$$

This means that $P_{r,K}$ is roughly of the order K^a , as least in the sense of \liminf . Also recall that for two sequences b_K and c_K , $K = 1, 2, \dots$, the symbol $b_K \sim c_K$ means $b_K/c_K \rightarrow 1$.

PROPOSITION 10. For $\Pi > 1$, we have:

(i) If $r > -1$, then

$$P_{r,K} \rightarrow (r+1)^{1/r} \mathbb{E}(p_1^r)^{1/r} \in (0, \infty) \quad \text{a.s.} \quad (16)$$

(ii) If $-\Pi < r < -1$, then

$$P_{r,K} \sim \frac{r}{r+1} K^{1+1/r} \mathbb{E}(p_1^r)^{1/r} \rightarrow \infty \quad \text{a.s.}$$

- (iii) If $-\infty < r < -\Pi$, then $P_{r,K} \approx K^{1-1/\Pi} \rightarrow \infty$.
- (iv) If $r = -\infty$, then $\lim_{K \rightarrow \infty} P_{r,K}/K^b = \infty$ in probability for $b < 1 - 1/\Pi$.

330 For $\Pi \in (0, 1)$, we have:

- (i) If $r > -\Pi$, then (16) holds.
- (ii) If $-1 < r < -\Pi$, then $P_{r,K} \approx K^{-1/r-1/\Pi} \rightarrow 0$.
- (iii) If $-\infty < r < -1$, then $P_{r,K} \approx K^{1-1/\Pi} \rightarrow 0$.
- (iv) If $r = -\infty$, then $\liminf_{K \rightarrow \infty} P_{r,K}/K^b = 0$ a.s. for $b > 1 - 1/\Pi$.

335 The results given in Proposition 10 are roughly summarized in Figure 2, where $r \in \{-\infty, \infty\}$ is also allowed. “Finite positive” means that $P_{r,K}$ converges to a finite positive number as $K \rightarrow \infty$, 0 means that it converges to 0 (at least in the sense of \liminf), “good 0” means that the rate of convergence to 0 is optimal (within the accuracy of (15)), ∞ means that it converges to ∞ , and “bad ∞ ” means the fastest rate of convergence to ∞ .

340 As we mentioned earlier, in the IID situation, the natural interpretation of Π is that it measures the asymptotic power of the p-values. When p_1 is uniformly distributed on $[0, 1]$ (as it is under the null hypothesis), $\Pi = 1$. If $\Pi > 1$, the p-values are asymptotically powerless for rejecting the null hypotheses, and when $\Pi < 1$, there is some power (e.g., the density function of p_1 is $f(p) := cp^{c-1}$ for $c \in (0, 1)$, as in e.g. [39]). The general
345 message of this section is that, in the IID case, using the Bonferroni merging function is a safe option, at least asymptotically, since it performs well in the interesting case $\Pi < 1$ (top of Figure 2); it might be also interesting to note that any merging function with $r < -1$ would achieve the same best rate of convergence to 0. The situation becomes more complicated in the case of dependent p-values, which will be considered in the next
350 subsection.

5.2. Exchangeable p-values

Now we relax our IID assumption to the more interesting case that the sequence p_1, p_2, \dots is exchangeable. This most general symmetry assumption includes cases of heavy dependence. By de Finetti’s theorem, every exchangeable distribution (of an infinite sequence) is a mixture $\int \tilde{Q}^\infty \nu(d\tilde{Q})$ of IID distributions \tilde{Q}^∞ ; see, e.g., [38, Theorem 1.49]. We can interpret ν as the limiting empirical distribution function of p_1, \dots, p_K .

The performance of our merging functions depends on the distribution of $\Pi(\tilde{Q})$. Applying Proposition 10 in combination with de Finetti’s theorem, we obtain the following two cases:

- 360 1. If $\nu(\Pi(\tilde{Q}) < 1) = 1$, the values $r \in [-\infty, -1]$ (including Bonferroni) perform very well (they are in the “good 0” area in Figure 2).
- 2. If $\nu(\Pi(\tilde{Q}) > 1) > 0$, the values $r \in [-\infty, -1]$ lead to an infinite expected combined p-value (or, with a positive probability, to a combined p-value of 1 when truncated) as $K \rightarrow \infty$.

365 Recall that the main motivation for our merging methods is the existence of unknown (possibly heavy) dependence. With dependence among p-values, the case $\nu(\Pi(\tilde{Q}) > 1) > 0$ is not unusual. For instance, in the extreme case of perfect dependence $p_1 = \dots = p_K$, if $\mathbb{P}(p_1 = 0) = 0$, then $\nu(\Pi(\tilde{Q}) = \infty) = 1$.

To analyze the influence of the strength of dependence on the performance of $r \in [-\infty, \infty]$, we consider the following simple example. Let

$$p_k = \Phi(X_k), \quad X_k = \rho Z + \sqrt{1 - \rho^2} Z_k - \mu, \quad k = 1, \dots, K, \quad (17)$$

where Φ is the standard normal distribution function, Z, Z_1, \dots, Z_K are IID standard normal random variables, and $\mu \geq 0$ and $\rho \in [0, 1]$ are constants. In other words, p_k is the p-value resulting from the k th standard one-sided z-test of the null hypothesis $\mu = 0$ against the alternative $\mu > 0$ using the statistic X_k from $N(-\mu, 1)$ with unknown μ , $k = 1, \dots, K$. Note that $\rho = 0$ corresponds to the case where p_1, \dots, p_K are independent, while $\rho = 1$ corresponds to the case where p_1, \dots, p_K are perfectly dependent.

Obviously, the model of (p_1, \dots, p_K) is exchangeable, and the marginal distribution Q of (p_1, \dots, p_K) does not depend on the correlation ρ . For $z \in \mathbb{R}$, let Q_z be the distribution of $\Phi(\sqrt{1 - \rho^2}W + \rho z - \mu)$, where W is a standard normal random variable. Clearly, conditional on $Z = z$, p_1, \dots, p_K are IID with distribution Q_z . Note that $\Pi(Q_z) = 1/(1 - \rho^2) > 1$ if $\rho > 0$ ($\Pi(Q_z)$ is computed in Lemma A.2 on page 27 in the Appendix). Therefore, in the presence of positive dependence, we are in Case 2 above ($\nu(\Pi(Q_z) > 1) = 1$); thus some choice of $r \geq -1$ may be optimal asymptotically as $K \rightarrow \infty$. For finite K , some simulation results of $P_{r,K}$ for the model (17) are reported in Subsection 6.4.

6. CHOOSING THE MERGING FUNCTION

In this section we discuss, for simplicity only in the symmetric case, how to choose a merging function.

6.1. A rule of thumb

First we state a crude rule of thumb for choosing r . Since any method based on the observed values of p_1, \dots, p_K would affect the validity of the method (see Subsection 6.3), we have to rely on prior or side information for a suitable choice of r . As a rule of thumb, if there is potentially substantial dependence among the p-values, then we should not use Bonferroni, and the harmonic mean might be a safer choice. If we are certain that the dependence is really strong, then the geometric and the arithmetic means might be an even better option. See Subsection 6.4 for a simulation study illustrating this point.

6.2. Practical issues

A fairly wide family of merging functions is provided in this paper. For a practitioner, given access to a variety of merging functions, it might be tempting to try many of them and then pick one of the merging functions that work well (perhaps the one yielding the smallest p-value). Of course, this is not a valid approach, and its inadmissibility has been discussed at length in recent literature (see, e.g., [45]). A search over various merging functions itself needs a multiple-testing correction, and the simplest procedure in the context of this paper would be to apply another merging function in our family to the outputs of the merging functions that we used in our search. This will be the topic of Subsection 6.3.

It might be also tempting to divide the available p-values into two parts, to find a suitable value of r from one part, and then compute the combined p-value using $a_{r,K}M_{r,K}$ applied to the other part. However, this is not a valid approach if dependence is present among the p-values. Under possibly heavy dependence, any peeking into the data is likely

to destroy validity (e.g., learning any of the p-values makes the conditional distribution of the remaining p-values non-uniform). From the vantage point of Section 5, the fact that we can never observe any IID components apart from the realized one prevents us from using interactive methods of adaptation such as the STAR method in [21].

6.3. Combining merging functions

In this subsection we will be interested in valid, necessarily very restrictive, ways of searching for a good merging function. Namely, we consider the following mechanism: for fixed $r_1, \dots, r_m \in [-\infty, \infty]$ and partition (A_1, \dots, A_m) of $[0, 1]^K$, we would like to use

$$F(p_1, \dots, p_K) := b \sum_{i=1}^m P_{r_i, K}(p_1, \dots, p_K) \mathbf{1}_{A_i}(p_1, \dots, p_K) \quad (18)$$

as a merging function, where b is a positive constant. In this case the observed p-values p_1, \dots, p_K determine which of the merging functions in our family $P_{r, K}$ is used, and the partition (A_1, \dots, A_m) serves as decision criterion. The value of b is chosen in such a way that F is a valid merging function.

A natural and simple choice of the decision criterion in (18) is to take the minimum of the merging functions $P_{r, K}$ to exploit their power. That is, by choosing A_i as the region in $[0, 1]^K$ where $P_{r_i, K}(p_1, \dots, p_K)$ is the smallest, we arrive at

$$F(p_1, \dots, p_K) := b \min_{i=1, \dots, m} P_{r_i, K}(p_1, \dots, p_K). \quad (19)$$

The constant b in (19) may be seen as the price to pay to exploit the power of different merging functions, and it is typically larger than 1. Clearly, choosing $b = m$ leads to a valid merging function in (19), and this is precisely applying the Bonferroni method on the combined p-values $P_{r_1, K}, \dots, P_{r_m, K}$. Simple examples show that in some cases a smaller value of b will also lead to a valid merging function: e.g., if all r_1, \dots, r_m are similar, then b can be chosen close to 1.

We consider in detail the most applicable cases where we search over only two values of r , one of them being $-\infty$ (the Bonferroni merging function). We would like to be competitive with the best of the two values, and so combine the two merging functions using the Bonferroni merging function. Namely, we consider the compound *Bonferroni-arithmetic* (BA) merging function

$$F_K^{\text{BA}}(p_1, \dots, p_K) := 2 \min(K \min(p_1, \dots, p_K), 2\bar{p}) \quad (20)$$

where \bar{p} is the arithmetic mean of p_1, \dots, p_K , and the compound *Bonferroni-geometric* (BG) merging function

$$F_K^{\text{BG}}(p_1, \dots, p_K) := 2 \min(K \min(p_1, \dots, p_K), e\tilde{p}) \quad (21)$$

where \tilde{p} is the geometric mean of p_1, \dots, p_K . Obviously, both methods are valid merging functions. Moreover, it turns out that these two merging functions are asymptotically precise. In other words, the price to pay for exploiting the power of both the Bonferroni method and the arithmetic/geometric average is precisely a factor of 2.

PROPOSITION 11. *Both families of merging functions F_K^{BA} and F_K^{BG} , $K = 2, 3, \dots$, in (20) and (21) are asymptotically precise.*

A crucial advantage of F_K^{BA} and F_K^{BG} over the merging functions in our family $a_{r, K} M_{r, K}$ is that both F_K^{BA} and F_K^{BG} improve greatly on the performance crudely shown in Figure

2: for F_K^{BA} and F_K^{BG} , in the top plot, we will get “finite positive” (the best value in that plot), and in the bottom plot, we will get “good 0” (also the best value in that plot). In particular, neither F_K^{BA} nor F_K^{BG} is dominated by $a_{r,K}M_{r,K}$ for any r .

6.4. A simulation study

We conduct some simulations for the correlated standard z-tests in model (17) to compare different merging methods. In Figure 3, we fix $\mu = 3$ (other values of μ give qualitatively similar results), and report the empirical performance for $K = 50, 400$ of merging methods with various values of r as well as the compound Bonferroni-geometric method in (21). The curves in Figure 3 are based on an average of 1,000 replications of $P_{r,K}$. For the best of visibility, we plot the range $r \in [-5, 0]$ for $\rho = 0.1, 0.5$ and $r \in [-5, 2]$ for $\rho = 0.9$. For values of r close to -1 , the asymptotically precise choice of the multiplier $a_{r,K}$ in Table 1 explodes as r approaches -1 . To avoid this explosion for finite K , we replace $a_{r,K}$ by $e \ln K$ if $a_{r,K} > e \ln K$ and $r \geq \ln K / (1 - \ln K)$, and this gives valid merging functions (cf. (A.6); there is still a kink around $r = -1$ since these choices are not precise). We make the following observations from Figure 3.

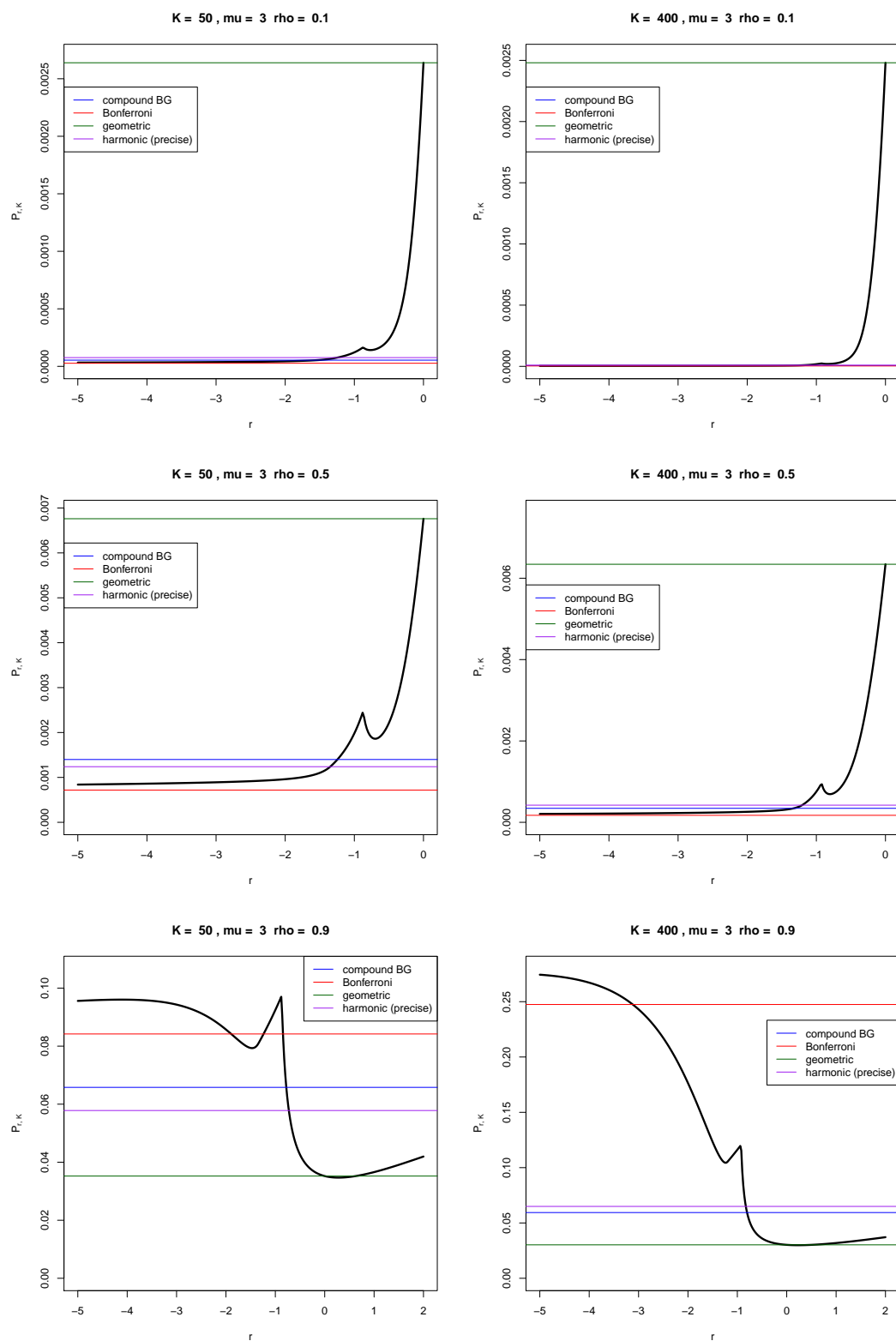
1. If the dependence is light or moderate ($\rho = 0.1, 0.5$), the Bonferroni method, as well as other methods based on $r < -1$, work quite well, and their power improves when K increases from 50 to 400.
2. In the case of very strong dependence ($\rho = 0.9$), the geometric averaging method and the arithmetic averaging method perform quite well. Notably, the performance of the Bonferroni method and other methods with $r < -1$ gets worse when K increases from 50 to 400.
3. The compound BG method generally performs quite well in all cases as it is often slightly worse (on average) than the best of the two base methods.

7. DIRECTIONS OF FURTHER RESEARCH

Perhaps the most important direction of further research is to find practically useful applications, in multiple testing of a single hypothesis or testing multiple hypotheses, for our methods of combining p-values. The Bonferroni method of combining a set of p-values works very well when experiments are almost independent, while it produces unsatisfactory results under heavy dependence (e.g., if all p-values are approximately equal). Our methods are designed to work for intermediate situations. Promising results have been obtained using the harmonic mean [46], but other merging functions proposed in this paper also deserve careful experimental study.

Our emphasis has been on finding valid methods of combining p-values, and we have just started exploration of their efficiency; the results of Section 5 are asymptotic and crude. A natural next step is to explore convergence of the combined p-values in distribution, as alluded after Proposition 10; such results would be still asymptotic but more precise (the price to pay will be the need to impose various conditions on the distribution of the p-values).

The main application of results about efficiency of various combination methods is choosing a suitable method. Subsection 6.3 is only a first step in this direction. Much wider families of potential merging functions deserve to be explored. In particular, finding the optimal value of b in merging functions of the form (18) is an interesting open problem.

Fig. 3: Empirical averages of $P_{r,K}$ for $K \in \{50, 400\}$

This paper concentrates on the symmetric case, where the merging function is a symmetric function of p_1, \dots, p_K . However, the weighted case (as in Section 4) is important in many applications because the quality of different p -values can be very different. It would be of much interest to study efficient ways of assigning weights to the p -values using prior or side information.

Acknowledgments

We thank the Editor, an Associate Editor and two referees for very constructive comments. We are grateful to Dave Cohen, Alessio Sancetta, Wouter Koolen, Lutz Mattner and Aaditya Ramdas for their advice and to Danqiao Guo for her substantial help on the simulation study. Our special thanks go to Paul Embrechts, who, in addition to illuminating discussions, has been instrumental in starting our collaboration.

For numeric calculations, we used Python and R. The first author's work was partially supported by the Cyprus Research Promotion Foundation and the European Union's Horizon 2020 Research and Innovation programme (671555), and the second author's by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590).

REFERENCES

- [1] Carole Bernard, Xiao Jiang, and Ruodu Wang. Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, 54:93–108, 2014.
- [2] Valeria Bignozzi, Tiantian Mao, Bin Wang, and Ruodu Wang. Diversification limit of quantiles under dependence uncertainty. *Extremes*, 19:143–170, 2016.
- [3] Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. *Introduction to Meta-Analysis*. Wiley, Chichester, 2009.
- [4] Morton B. Brown. A method for combining non-independent, one-sided tests of significance. *Biometrics*, 31:987–992, 1975.
- [5] Robert D. Cousins. Annotated bibliography of some papers on combining significances or p -values. Technical Report [arXiv:0705.2209](https://arxiv.org/abs/0705.2209) [physics.data-an], [arXiv.org](https://arxiv.org/) e-Print archive, December 2008.
- [6] Paul Embrechts and Giovanni Puccetti. Bounds for functions of dependent risks. *Finance and Stochastics*, 10:341–352, 2006.
- [7] Paul Embrechts, Bin Wang, and Ruodu Wang. Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19:763–790, 2015.
- [8] Ronald A. Fisher. Combining independent tests of significance. *American Statistician*, 2:30, 1948.
- [9] Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter, Berlin, third edition, 2011.
- [10] Jelle J. Goeman and Aldo Solari. Multiple testing for exploratory research. *Statistical Science*, 26:584–597, 2011.
- [11] Benjamin T. Graham and Geoffrey R. Grimmett. Influence and sharp-threshold theorems for monotonic measures. *Annals of Probability*, 34:1726–1745, 2006.
- [12] Allan Gut. *Probability: A Graduate Course*. Springer, New York, second edition, 2013.
- [13] G. H. Hardy, John E. Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, Cambridge, second edition, 1952.
- [14] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.
- [15] Gerhard Hommel. Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical Journal*, 25:423–430, 1983.
- [16] Gerhard Hommel. Multiple test procedures for arbitrary dependence structures. *Metrika*, 33:321–336, 1986.
- [17] Gerhard Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75:383–386, 1988.
- [18] Edwards Jakobsons, Xiaoying Han, and Ruodu Wang. General convex order on risk aggregation. *Scandinavian Actuarial Journal*, 2016:713–740, 2016.

- [19] Andrei N. Kolmogorov. Sur la notion de la moyenne. *Atti della Reale Accademia Nazionale dei Lincei. Classe di scienze fisiche, matematiche, e naturali. Rendiconti Serie VI*, 12(9):388–391, 1930.
- [20] James T. Kost and Michael P. McDermott. Combining dependent p-values. *Statistics and Probability Letters*, 60:183–190, 2002.
- [21] Lihua Lei, Aaditya Ramdas, and William Fithian. Star: A general interactive framework for FDR control under structural constraints. Technical Report [arXiv:1710.02776 \[stat.ME\]](#), [arXiv.org e-Print archive](#), October 2017.
- [22] Tamás Lipták. On the combination of independent tests. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 3:171–197, 1958.
- [23] G. D. Makarov. Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory of Probability and its Applications*, 26:803–806, 1981.
- [24] Ruth Marcus, Eric Peritz, and K. Ruben Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63:655–660, 1976.
- [25] Lutz Mattner. Combining individually valid and conditionally i.i.d. P-variables. Technical Report [arXiv:1008.5143 \[stat.ME\]](#), [arXiv.org e-Print archive](#), August 2011.
- [26] Lutz Mattner. Combining individually valid and arbitrarily dependent P-variables. In *Abstract Book of the Tenth German Probability and Statistics Days*, page p. 104, Mainz, Germany, March 2012. Institut für Mathematik, Johannes Gutenberg-Universität Mainz.
- [27] Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ, 2015.
- [28] Xiao-Li Meng. Posterior predictive p-values. *Annals of Statistics*, 22:1142–1160, 1993.
- [29] Dietrich Morgenstern. Berechnung des maximalen Signifikanzniveaus des Testes “Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen”. *Metrika*, 27:285–286, 1980.
- [30] Jacobus Oosterhoff. Combination of one-sided statistical tests. *MC Tracts*, 1976.
- [31] Art B. Owen. Karl Pearson’s meta-analysis revisited. *Annals of Statistics*, 37:3867–3892, 2009.
- [32] Karl Pearson. On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, 25:379–410, 1933.
- [33] Giovanni Puccetti and Ludger Rüschendorf. Sharp bounds for sums of dependent risks. *Journal of Applied Probability*, 50:42–53, 2013.
- [34] Bernhard Rüger. Das maximale Signifikanzniveau des Testes “Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen”. *Metrika*, 25:171–178, 1978.
- [35] Ludger Rüschendorf. Random variables with maximum sums. *Advances in Applied Probability*, 14:623–632, 1982.
- [36] Ludger Rüschendorf. On the distributional transform, Sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139:3921–3927, 2009.
- [37] Ludger Rüschendorf. *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg, 2013.
- [38] Mark J. Schervish. *Theory of Statistics*. Springer, New York, 1995.
- [39] Thomas Sellke, M. J. Bayarri, and James Berger. Calibration of p-values for testing precise null hypotheses. *American Statistician*, 55:62–71, 2001.
- [40] R. John Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73:751–754, 1986.
- [41] Leonard H. C. Tippett. *The Methods of Statistics: An Introduction Mainly for Experimentalists*. Williams and Norgate, London, 1931.
- [42] Bin Wang and Ruodu Wang. The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102:1344–1360, 2011.
- [43] Bin Wang and Ruodu Wang. Joint mixability. *Mathematics of Operations Research*, 41:808–826, 2016.
- [44] Ruodu Wang, Liang Peng, and Jingping Yang. Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, 17:395–417, 2013.
- [45] Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar. Moving to a world beyond “ $p < 0.05$ ”. *American Statistician*, 73(Supplement 1):1–19, 2019. Editorial, Special Issue on p-values.
- [46] Daniel J. Wilson. The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences*, 116:1195–1200, 2019.

Supplementary material for “Combining p-values via averaging”

BY VLADIMIR VOVK AND RUODU WANG

A. PROOFS OF MAIN RESULTS

600

This appendix contains proofs of our main results. All these proofs are based on known results in robust risk aggregation. For $a > 0$, let $\mathcal{U}(a)$ be the set of all random variables distributed uniformly over the interval $[0, a]$, $a \geq 0$; we can regard \mathcal{U} as an abbreviation for $\mathcal{U}(1)$.

Proof of Lemma 1

PART “IF” OF (A): Suppose $q_\epsilon(F) \geq \epsilon$ for all $\epsilon \in (0, 1)$. Consider arbitrary $U_1, \dots, U_K \in \mathcal{U}$. We have $q_\epsilon(F(U_1, \dots, U_K)) \geq \epsilon$ for all $\epsilon \in (0, 1)$. By the definition of left quantiles, $\mathbb{P}(F(U_1, \dots, U_K) < \epsilon) \leq \epsilon$. It follows that, for all $\delta \in (0, 1 - \epsilon)$,

$$\mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \leq \mathbb{P}(F(U_1, \dots, U_K) < \epsilon + \delta) \leq \epsilon + \delta,$$

which implies

$$\mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \leq \epsilon,$$

since δ is arbitrary. Therefore, F is a merging function.

PART “ONLY IF” OF (A): Suppose F is a merging function. Let $U_1, \dots, U_K \in \mathcal{U}$ and $\epsilon \in (0, 1)$. We have $\mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \leq \epsilon$. By the definition of right quantiles, $q_\epsilon^+(F(U_1, \dots, U_K)) \geq \epsilon$. It follows that, for all $\delta \in (0, \epsilon)$,

$$q_\epsilon(F(U_1, \dots, U_K)) \geq q_{\epsilon-\delta}^+(F(U_1, \dots, U_K)) \geq \epsilon - \delta,$$

which implies $q_\epsilon(F(U_1, \dots, U_K)) \geq \epsilon$ since δ is arbitrary.

PART “IF” OF (B): Suppose $q_\epsilon(F) = \epsilon$ for all $\epsilon \in (0, 1)$. By (a), F is a merging function. For all $\epsilon, \delta \in (0, 1)$, there exist $U_1, \dots, U_K \in \mathcal{U}$ such that $q_\epsilon(F(U_1, \dots, U_K)) \in [\epsilon, \epsilon + \delta)$, which implies $\mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon + \delta) \geq \epsilon$. Since δ is arbitrary, we have

$$\sup \{ \mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon) \mid U_1, \dots, U_K \in \mathcal{U} \} = \epsilon,$$

and thus F is precise.

PART “ONLY IF” OF (B): Suppose F is a precise merging function. Since F is a merging function, by (a) we have $q_\epsilon(M) \geq \epsilon$ for all $\epsilon \in (0, 1)$. Suppose, for the purpose of contradiction, that $q_\epsilon(M) > \epsilon$ for some $\epsilon \in (0, 1)$. Then, there exists $\delta \in (0, 1 - \epsilon)$ such that $q_\epsilon(F(U_1, \dots, U_K)) > \epsilon + \delta$ for all $U_1, \dots, U_K \in \mathcal{U}$. As a consequence, we have

$$\mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon + \delta/2) \leq \mathbb{P}(F(U_1, \dots, U_K) < \epsilon + \delta) \leq \epsilon.$$

Therefore,

$$\sup \{ \mathbb{P}(F(U_1, \dots, U_K) \leq \epsilon + \delta/2) \mid U_1, \dots, U_K \in \mathcal{U} \} < \epsilon + \delta/2,$$

contradicting F being precise.

In all proofs below, for statements that have a weighted version in Section 4, namely Theorem 1 and Propositions 2 and 3, we present a proof of the corresponding weighted version, which is stronger.

Proof of Theorem 3 (weighted version of Theorem 1)

Without loss of generality we can, and will, assume that ϕ is strictly increasing. Indeed, if ϕ is strictly decreasing, we can redefine $\phi := -\phi$ and $\psi(u) := \psi(-u)$ and notice that the statement of the theorem for new ϕ and ψ will imply the analogous statement for the original ϕ and ψ .

Define an accessory function $\Phi : (0, 1) \rightarrow [-\infty, \infty]$ by $\Phi(\epsilon) = \frac{1}{\epsilon} \int_0^\epsilon \phi(u) du$. Fix $\epsilon \in (0, 1)$. Since ϕ is integrable, $\Phi(\epsilon)$ is finite.

Known results from the literature on robust risk aggregation can be applied to random variables $X_k := \phi(U_k)$, where $U_k \in \mathcal{U}$; notice that the distribution function of X_k is ψ :

$$\mathbb{P}(X_k \leq x) = \mathbb{P}(\phi(U_k) \leq x) = \mathbb{P}(U_k \leq \psi(x)) = \psi(x).$$

635 Theorem 4.6 of [1] gives the following relation:

$$\underline{q}_\epsilon(M_{\phi, \mathbf{w}}) = \inf \left\{ q_1 \left(\psi \left(\sum_{k=1}^K w_k \phi(V_k) \right) \right) \middle| V_1, \dots, V_K \in \mathcal{U}(\epsilon) \right\}. \quad (\text{A.1})$$

Since

$$q_1(w_1 \phi(V_1) + \dots + w_K \phi(V_K)) \geq \mathbb{E}[w_1 \phi(V_1) + \dots + w_K \phi(V_K)] = \Phi(\epsilon)$$

for $V_1, \dots, V_K \in \mathcal{U}(\epsilon)$, we have $\underline{q}_\epsilon(M_{\phi, \mathbf{w}}) \geq \psi(\Phi(\epsilon))$.

Proof of Proposition 1

640 The case $r = \infty$ is trivial, and we focus on $r \in (-1, \infty)$. Let $\phi(u) = u^r$, which gives $\Phi(\epsilon) = \epsilon^r / (r + 1)$, in the notation of the previous proof. Evaluating the term (12) in (11), we obtain

$$\psi \left(\frac{1}{\epsilon} \int_0^\epsilon \phi(u) du \right) = \begin{cases} \epsilon/e & \text{if } r = 0 \\ (r + 1)^{-1/r} \epsilon & \text{otherwise.} \end{cases}$$

This shows that $(r + 1)^{1/r} M_{r, K}$ is indeed a merging function.

Using Corollary 3.4 of [7], we have

$$\lim_{K \rightarrow \infty} \frac{\phi(\underline{q}_\epsilon(M_{r, K}))}{\Phi(\epsilon)} = 1,$$

leading to

$$\lim_{K \rightarrow \infty} \underline{q}_\epsilon(M_{r, K}) = \epsilon(r + 1)^{-1/r}. \quad (\text{A.2})$$

It follows that, for $a < (r + 1)^{1/r}$,

$$\lim_{K \rightarrow \infty} \underline{q}_\epsilon(a M_{r, K}) < \epsilon$$

645 and so, by Lemma 1, $a M_{r, K}$ is not a merging function for K large enough.

Proof of Proposition 8 (weighted version of Proposition 2)

Let $M = (r + 1)^{1/r} M_{r, \mathbf{w}}$. Using (A.1) and Theorem 3, we have, for $\epsilon \in (0, 1)$:

$$(\underline{q}_\epsilon(M_{r, \mathbf{w}}))^r = \inf \left\{ q_1 \left(\sum_{k=1}^K w_k V_k^r \right) \middle| V_1, \dots, V_K \in \mathcal{U}(\epsilon) \right\} \geq \frac{\epsilon^r}{1 + r} \quad (\text{A.3})$$

if $r > 0$,

$$(\underline{q}_\epsilon(M_{r, \mathbf{w}}))^r = \sup \left\{ q_0^+ \left(\sum_{k=1}^K w_k V_k^r \right) \middle| V_1, \dots, V_K \in \mathcal{U}(\epsilon) \right\} \leq \frac{\epsilon^r}{1 + r} \quad (\text{A.4})$$

if $r < 0$, and

$$\underline{q}_\epsilon(M_{r, \mathbf{w}}) = \exp \left(\inf \left\{ q_1 \left(\sum_{k=1}^K w_k \ln V_k \right) \middle| V_1, \dots, V_K \in \mathcal{U}(\epsilon) \right\} \right) \geq \frac{\epsilon}{e} \quad (\text{A.5})$$

if $r = 0$. By Lemma 1, M is a precise merging function if and only if the inequality in (A.3)–(A.5) is an equality for all $\epsilon \in (0, 1)$. 650

Fix $\epsilon \in (0, 1)$ and $r \in (-1, \infty)$. For $k = 1, \dots, K$, let F_k be the distribution of $w_k V_k^r$ where $V_k \in \mathcal{U}(\epsilon)$. Using the terminology of [43], notice that the inequality in (A.3)–(A.5) is an equality if and only if (F_1, \dots, F_K) is jointly mixable due to a standard compactness argument (see [43, Proposition 2.3]). Therefore, we can first settle the cases $r = 0$ and $r < 0$, as in these cases the supports of F_1, \dots, F_K are unbounded on one side, and (F_1, \dots, F_K) is not jointly mixable (see [43, Remark 2.2]). 655

Next assume $r > 0$. Since F_1, \dots, F_K have monotone densities on their respective supports, by Theorem 3.2 of [43], (F_1, \dots, F_K) is jointly mixable if and only if the “mean condition”

$$w\epsilon^r \leq \frac{\epsilon^r}{1+r} \leq \epsilon^r - w\epsilon^r$$

is satisfied. This is equivalent to $w \leq \frac{1}{1+r} \leq 1 - w$ and, therefore, to the conjunction of $w \leq 1/2$ and $r \in [\frac{w}{1-w}, \frac{1-w}{w}]$. This completes the proof. 660

Proof of Proposition 9 (weighted version of Proposition 3)

Notice that, for each $k = 1, \dots, K$, the distribution of $w_k U_k^r$, where $U_k \in \mathcal{U}$, has a decreasing density on its support. Therefore, we can apply Corollary 4.7 of [18], which gives

$$\inf \{q_\epsilon(w_1 U_1^r + \dots + w_K U_K^r) \mid U_1, \dots, U_K \in \mathcal{U}\} = \max \left(w\epsilon^r, \frac{\epsilon^r}{1+r} \right).$$

Simple algebra leads to 665

$$\underline{q}_\epsilon(M_{r,K}) = \max \left(w, \frac{1}{1+r} \right)^{1/r} \epsilon,$$

and by Lemma 1, M is a precise merging function.

Proof of Proposition 4

First, we note that as $\lim_{r \rightarrow 0} (r+1)^{1/r} = e$, we know $a_K^G \leq e$ from Proposition 1. Moreover, by letting $r \rightarrow 0$ in (A.2), we know that $a_K^G \rightarrow e$.

Our next goal is to obtain the precise value of $q_\epsilon(M_{0,K})$. Set 670

$$\begin{aligned} b_K &:= \sup \{q_0^+(-(\ln U_1 + \dots + \ln U_K)) \mid U_1, \dots, U_K \in \mathcal{U}\} \\ &= \sup \{q_0^+(-(\ln V_1 + \dots + \ln V_K) + K \ln \epsilon) \mid V_1, \dots, V_K \in \mathcal{U}(\epsilon)\}. \end{aligned}$$

It is easy to see that

$$\begin{aligned} \underline{q}_\epsilon(M_{0,K}) &= \exp \left(\inf \left\{ q_1 \left(\frac{\ln V_1 + \dots + \ln V_K}{K} \right) \mid V_1, \dots, V_K \in \mathcal{U}(\epsilon) \right\} \right) \\ &= \exp \left(- \sup \left\{ q_0^+ \left(- \frac{\ln V_1 + \dots + \ln V_K}{K} \right) \mid V_1, \dots, V_K \in \mathcal{U}(\epsilon) \right\} \right) \\ &= \exp(-b_K/K + \ln \epsilon) \\ &= \epsilon \exp(-b_K/K). \end{aligned} \quad \text{675}$$

It is clear that $e^{b_K/K} M_{0,K}$ is a precise merging function, and $b_K/K \rightarrow 1$.

Since $-\ln U$ has the standard exponential distribution for $U \in \mathcal{U}$ and, therefore, a decreasing density on \mathbb{R} , we can apply Theorem 3.2 of [1] (essentially Theorem 3.5 of [42]) to arrive at 680

$$b_K = -(K-1) \ln(1 - (K-1)c_K) - \ln c_K,$$

where c_K is the unique solution to (13) (see [42, Corollary 4.1]). Using (13), we can write

$$b_K/K = -\ln c_K - (K-1)(1 - Kc_K).$$

Using $a_K^G = e^{b_K/K}$ one obtains the desired result.

Proof of Theorem 2

We will apply Theorem 4.2 of [6] in our situation where the function ϕ (and, therefore, ψ as well) in (7) is decreasing. Letting $X_k := \phi(p_k)$ and using the notation m_+ (used in Theorem 4.2 of [6]), we have, by the definition of m_+ ,

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^K X_k < s\right) &\geq m_+(s), \\ \mathbb{P}\left(\frac{1}{K} \sum_{k=1}^K \phi(p_k) < s/K\right) &\geq m_+(s), \\ \mathbb{P}(M_{\phi,K}(p_1, \dots, p_K) > \psi(s/K)) &\geq m_+(s), \\ \mathbb{P}(M_{\phi,K}(p_1, \dots, p_K) \leq \psi(s/K)) &\leq 1 - m_+(s).\end{aligned}$$

The lower bound on $m_+(s)$ given in Theorem 4.2 of [6] involves $1 - F(x)$, where F is the common distribution function of X_k , and in our current context we have:

$$1 - F(x) = \mathbb{P}(X_k > x) = \mathbb{P}(\phi(p_k) > x) = \mathbb{P}(p_k < \psi(x)) = \psi(x).$$

The last inequality and chain of equalities in combination with Theorem 4.2 of [6] give

$$\mathbb{P}(M_{\phi,K}(p_1, \dots, p_K) \leq \psi(s/K)) \leq K \inf_{r \in [0, s/K)} \frac{\int_r^{s-(K-1)r} \psi(x) dx}{s - Kr}.$$

Setting $\epsilon := \psi(s/K) \in [\psi(\infty), \psi(0)]$ (so that it is essential that $\psi(\infty) = 0$), we obtain, using $s = K\phi(\epsilon)$,

$$\mathbb{P}(M_{\phi,K}(p_1, \dots, p_K) \leq \epsilon) \leq K \inf_{r \in [0, \phi(\epsilon))} \frac{\int_r^{K\phi(\epsilon)-(K-1)r} \psi(x) dx}{(\phi(\epsilon) - r)K}.$$

Setting $t := \phi(\epsilon) - r$ and renaming x to u , this can be rewritten as (14).

Proof of Proposition 5

The case $r = -\infty$ is trivial, and we focus on $r \in (-\infty, -1)$. By Theorem 2 applied to $\phi(u) := u^r$, $r < -1$, we have

$$\mathbb{P}(M_{\phi,K}(p_1, \dots, p_K) \leq \epsilon) \leq \frac{\int_0^{K\phi(\epsilon)} \psi(u) du}{\phi(\epsilon)} = \frac{r}{r+1} K^{1+1/r} \epsilon.$$

This shows that $\frac{r}{r+1} K^{1+1/r} M_{r,K}$ is indeed a merging function.

We next show a simple property of a precise merging function via general averaging. Define the following constant:

$$b_{r,K} := \left(\frac{1}{K} \sup \{ q_0^+(U_1^r + \dots + U_K^r) \mid U_1, \dots, U_K \in \mathcal{U} \} \right)^{-1/r}.$$

It is clear that $b_{r,K} \geq 1$ for $r < 0$.

LEMMA A.1. *For $r < 0$, the function $b_{r,K} M_{r,K}$ is a precise merging function.*

Proof. By straightforward algebra and Theorem 4.6 of [1],

$$\begin{aligned}\underline{q}_\epsilon(b_{r,K} M_{r,K}) &= b_{r,K} \inf \{ q_1(M_{r,K}(V_1, \dots, V_K)) \mid V_1, \dots, V_K \in \mathcal{U}(\epsilon) \} \\ &= b_{r,K} \inf \{ \epsilon q_1(M_{r,K}(U_1, \dots, U_K)) \mid U_1, \dots, U_K \in \mathcal{U} \}\end{aligned}$$

$$\begin{aligned}
&= b_{r,K} \epsilon \left(\frac{1}{K} \sup \{ q_0^+(U_1^r + \cdots + U_K^r) \mid U_1, \dots, U_K \in \mathcal{U} \} \right)^{1/r} \\
&= \epsilon.
\end{aligned}$$

By Lemma 1, $b_{r,K} M_{r,K}$ is a precise merging function. □ 710

To construct precise merging functions, it remains to find values of $b_{r,K}$. Unfortunately, for $r < 0$ no analytical formula for $b_{r,K}$ is available. There is an asymptotic result available in [2], which leads to the following proposition.

PROPOSITION A.1. For $r \in (-\infty, -1)$,

$$\lim_{K \rightarrow \infty} \frac{b_{r,K}}{K^{1+1/r}} = \frac{r}{r+1}.$$

Proof. The quantity $\overline{\Delta}^{\mathcal{F}_d}$ in [2], defined as 715

$$\overline{\Delta}^{\mathcal{F}_d} := \lim_{\alpha \rightarrow 1} \frac{\sup \{ q_\alpha(U_1^r + \cdots + U_K^r) \mid U_1, \dots, U_K \in \mathcal{U}(\alpha) \}}{K(1-\alpha)^r},$$

satisfies

$$\overline{\Delta}^{\mathcal{F}_d} = \frac{1}{K} \sup \{ q_0^+(U_1^r + \cdots + U_K^r) \mid U_1, \dots, U_K \in \mathcal{U} \} = b_{r,K}^{-r}.$$

Using Proposition 3.5 of [2], we have, for $r < -1$, by substituting $\beta := -1/r$ in (3.25) of [2] and $\overline{\Delta}^{\mathcal{F}_d} = b_{r,K}^{-r}$,

$$\lim_{K \rightarrow \infty} \frac{b_{r,K}^{-r}}{K^{-r-1}} = \left(\frac{r}{r+1} \right)^{-r},$$

and this gives the desired result. □

The claim on the asymptotic precision in Proposition 5 immediately follows from Lemma A.1 and Proposition A.1. 720

Proof of Proposition 6

Let us first find the smallest value of the coefficient $\frac{r}{r+1} K^{1+1/r}$ in Proposition 5. Setting the derivative in r of the logarithm of this coefficient to 0, we obtain a linear equation whose solution is 725

$$r = \frac{\ln K}{1 - \ln K}. \quad (\text{A.6})$$

Plugging this into the coefficient gives $e \ln K$. Notice that r defined by (A.6) satisfies $r < -1$ and apply the inequality $M_{r,K} \leq M_{-1,K}$, a special case of (10). Hence, $e \ln K M_{-1,K}$ is a merging function, and $a_K^H \leq e \ln K$ follows from the fact that $a_K^H M_{-1,K}$ is a precise merging function, which we prove below.

Using the notation in Lemma A.1, we need to show $a_K^H = b_{-1,K}$, and we shall use Corollary 3.7 of [44]. Write 730

$$H(t) := \frac{K-1}{1-(K-1)t} + \frac{1}{t} = \frac{1}{t(1-(K-1)t)}, \quad t \in [0, 1/K].$$

By Corollary 3.7 of [44], we have

$$\begin{aligned}
b_{-1,K} &= \frac{1}{K} \sup \{ q_0^+(U_1^{-1} + \cdots + U_K^{-1}) \mid U_1, \dots, U_K \in \mathcal{U} \} \\
&= \frac{1}{K} H(x_K) = \frac{1}{K x_K (1 - (K-1)x_K)}
\end{aligned}$$

735 where x_K solves the equation

$$\int_x^{1/K} H(t) dt = \left(\frac{1}{K} - x \right) H(x), \quad x \in [0, 1/K].$$

Plugging in the expression for H and rearranging the above equation, we obtain

$$\begin{aligned} \frac{1 - Kx}{Kx(1 - (K-1)x)} &= \int_x^{1/K} \left(\frac{K-1}{1 - (K-1)t} + \frac{1}{t} \right) dt \\ &= \int_{(K-1)x}^{(K-1)/K} \frac{1}{1-y} dy + \int_x^{1/K} \frac{1}{t} dt \\ &= \ln(1 - (K-1)x) - \ln x. \end{aligned} \quad (\text{A.7})$$

740 The uniqueness of the solution x_K to (A.7) can be easily checked, and it is a special case of Lemma 3.1 of [18]. Writing $y = \frac{1-Kx}{x} > 0$, (A.7) reads as $\frac{y}{(y+1)K/(y+K)} = \ln(y+1)$. Rearranging the terms gives

$$y^2 = K((y+1)\ln(y+1) - y), \quad (\text{A.8})$$

which admits a unique solution, $y_K = \frac{1-Kx_K}{x_K}$. Therefore,

$$b_{-1,K} = \frac{1}{Kx_K(1 - (K-1)x_K)} = \frac{(y_K + K)^2}{(y_K + 1)K},$$

and hence $a_K^H = b_{-1,K}$.

745 Next we analyze the asymptotic behaviour of a_K^H as $K \rightarrow \infty$. Using $\ln(y+1) \geq y - y^2/2$ for $y \geq 0$, we can see that (A.8) implies the inequality

$$y^2 \geq \frac{K}{2}y^2 - \frac{K}{2}y^3,$$

which leads to $2 \geq K(1-y)$. Hence, we have $\liminf_{K \rightarrow \infty} y_K \geq 1$.

Notice that $(y+1)\ln(y+1) - y$ is a strictly increasing function of $y \in (0, \infty)$. Using $\liminf_{K \rightarrow \infty} y_K \geq 1$, we obtain that

$$\liminf_{K \rightarrow \infty} y_K^2 \geq K(2\ln 2 - 1).$$

750 Therefore, $\lim_{K \rightarrow \infty} y_K = \infty$. Applying logarithms to both sides of (A.8) and taking a limit in their ratio, we obtain

$$1 = \lim_{K \rightarrow \infty} \frac{2 \ln y_K}{\ln K + \ln(y_K + 1) + \ln \left(\ln(y_K + 1) - \frac{y_K}{y_K + 1} \right)} = \lim_{K \rightarrow \infty} \frac{2 \ln y_K}{\ln K + \ln y_K},$$

and hence $\ln y_K / \ln K \rightarrow 1$ as $K \rightarrow \infty$. Using (A.8) again, we have

$$1 = \lim_{K \rightarrow \infty} \frac{y_K^2}{K((y_K + 1)\ln(y_K + 1) - y_K)} = \lim_{K \rightarrow \infty} \frac{y_K^2}{K(y_K \ln y_K)} = \lim_{K \rightarrow \infty} \frac{y_K}{K \ln K}.$$

Therefore, we have

$$\lim_{K \rightarrow \infty} \frac{a_K^H}{\ln K} = \lim_{K \rightarrow \infty} \frac{(y_K + K)^2}{(y_K + 1)K \ln K} = \lim_{K \rightarrow \infty} \frac{(K \ln K)^2}{(K \ln K)K \ln K} = 1.$$

This completes the proof.

Proof of Proposition 7

Let us check that for $F := aM_{r,K}$ the following statements are equivalent:

- (a) F is a merging function;
- (b) (5) holds for some $\epsilon \in (0, 1)$;
- (c) $\underline{q}_\epsilon(F) \geq \epsilon$ for some $\epsilon \in (0, 1)$.

The implication (a) \Rightarrow (b) holds by definition. 760

To check (b) \Rightarrow (c), let us assume (b). Since $\mathbb{P}(X \leq \epsilon) \leq \epsilon$ implies $q_\epsilon^+(X) \geq \epsilon$ for any random variable X ,

$$\inf \{q_\epsilon^+(F(U_1, \dots, U_K)) \mid U_1, \dots, U_K \in \mathcal{U}\} \geq \epsilon.$$

Using Lemma 4.5 of [1],

$$\underline{q}_\epsilon(F) = \inf \{q_\epsilon^+(F(U_1, \dots, U_K)) \mid U_1, \dots, U_K \in \mathcal{U}\},$$

and hence (c) follows.

It remains to check (c) \Rightarrow (a). For any $\epsilon \in (0, 1)$, by straightforward algebra and Theorem 4.6 765 of [1],

$$\begin{aligned} \underline{q}_\epsilon(F) &= \inf \{q_1(F(V_1, \dots, V_K)) \mid V_1, \dots, V_K \in \mathcal{U}(\epsilon)\} \\ &= \epsilon \inf \{q_1(F(U_1, \dots, U_K)) \mid U_1, \dots, U_K \in \mathcal{U}\}. \end{aligned}$$

Therefore, to check $\underline{q}_\epsilon(F) \geq \epsilon$ for all $\epsilon \in (0, 1)$, one only needs to check the inequality for one $\epsilon \in (0, 1)$. By Lemma 1, F is a merging function. 770

This completes the proof of the first part of Proposition 7, and the second part can be proved similarly.

Proof of Proposition 10

First we consider the integrable case $r > -\Pi$. By the Marcinkiewicz strong law of large numbers [12, Theorem 6.7.1] (and Kolmogorov's strong law of large numbers, which is its special case), we 775 have

$$\frac{1}{K} \sum_{k=1}^K p_k^r \rightarrow \mathbb{E}(p_1^r) \quad \text{a.s.}$$

This implies parts (i) and (ii) of the case $\Pi > 1$ and part (i) of the case $\Pi < 1$. Since $\mathbb{P}(p_1 > 0) = 1$ unless $\Pi = 0$, this also gives the "finite positive" entries in Figure 2.

Now let us consider the non-integrable case $r < -\Pi$. In this case $-\Pi/r \in [0, 1)$ and we can apply Marcinkiewicz's strong law, assuming $\Pi > 0$ (the simple case $\Pi = 0$ should be considered 780 separately). We have $\mathbb{E}((p_1^r)^{-\Pi/r-\epsilon}) < \infty$ for any $\epsilon > 0$, and so, by Marcinkiewicz's law (applicable when $\epsilon < -\Pi/r$),

$$\frac{\sum_{k=1}^K p_k^r}{K^{-r/\Pi+\epsilon}} \rightarrow 0 \quad \text{a.s.}$$

for any $\epsilon > 0$. This implies

$$\frac{\left(\sum_{k=1}^K p_k^r / K\right)^{1/r}}{K^{-1/\Pi-1/r-\epsilon}} \rightarrow \infty \quad \text{a.s.} \tag{A.9}$$

for any $\epsilon > 0$. On the other hand, we have $\mathbb{E}((p_1^r)^{-\Pi/r+\epsilon}) = \infty$ for any $\epsilon > 0$, and so, by the other 785 part of Marcinkiewicz's law (applicable when $\epsilon < 1 + \Pi/r$),

$$\frac{\sum_{k=1}^K p_k^r}{K^{-r/\Pi-\epsilon}} \rightarrow 0 \quad \text{a.s.}$$

fails for any $\epsilon > 0$. Therefore,

$$\frac{\left(\sum_{k=1}^K p_k^r / K\right)^{1/r}}{K^{-1/\Pi-1/r+\epsilon}} \rightarrow \infty \quad \text{a.s.}$$

fails for any $\epsilon > 0$. This implies that

$$\liminf_{K \rightarrow \infty} \frac{\left(\sum_{k=1}^K p_k^r / K\right)^{1/r}}{K^{-1/\Pi-1/r+\epsilon}} = 0 \quad (\text{A.10})$$

with a positive probability for any $\epsilon > 0$. Kolmogorov's zero-one law [12, Theorem 1.5.1] allows us to replace "with a positive probability" by "a.s." Combining (A.9) and (A.10) gives part (iii) of the case $\Pi > 1$ and parts (ii) and (iii) of the case $\Pi < 1$.

It remains to consider the case $r = -\infty$ (Bonferroni). First suppose $\Pi > 1$. For any $b \in (0, 1 - 1/\Pi)$ and $B > 0$, take $c = (1/(1-b), \Pi)$, and note that $\mathbb{E}(p_1^{-c}) < \infty$. We have

$$\begin{aligned} \mathbb{P}(P_{r,K} \leq BK^b) &= \mathbb{P}(\exists k \in \{1, \dots, K\} : p_k \leq BK^{b-1}) \\ &\leq K \mathbb{P}(p_1 \leq BK^{b-1}) \\ &= K \mathbb{P}(p_1^{-c} \geq (BK^{b-1})^{-c}) \\ &\leq K \mathbb{E}(p_1^{-c}) / (BK^{b-1})^{-c} = \mathbb{E}(p_1^{-c}) B^c K^{1-(1-b)c} \rightarrow 0. \end{aligned}$$

This gives $P_{r,K}/K^b \rightarrow \infty$ in probability and, therefore, part (iv) of the case $\Pi > 1$. Finally, suppose $\Pi < 1$. Note that for any $r < -1$,

$$P_{-\infty,K} = K \min_{k \in \{1, \dots, K\}} p_k \leq K M_{r,K}(p_1, \dots, p_K) = \frac{r+1}{r} K^{-1/r} P_{r,K}.$$

For $b > 1 - 1/\Pi$, we can take $r < -1$ small enough such that $b + 1/r > 1 - 1/\Pi$, and obtain by using (iii),

$$P_{-\infty,K}/K^b \leq \frac{r+1}{r} P_{r,K}/K^{b+1/r} \rightarrow 0 \quad \text{a.s.}$$

Proof of Proposition 11

We first analyze the compound BA method. To show that the family F_K^{BA} is asymptotically precise, it suffices to show that for some $\epsilon \in (0, 1)$,

$$\limsup_{K \rightarrow \infty} \left(\sup \left\{ \mathbb{P}(F_K^{\text{BA}}(U_1, \dots, U_K) \leq \epsilon) \mid U_1, \dots, U_K \in \mathcal{U} \right\} \right) = \epsilon. \quad (\text{A.11})$$

To show (A.11), fix $\epsilon \in (0, 1)$ and an even number K . Set $\delta := \epsilon/(2K)$. Let A_1, \dots, A_K, B be disjoint events with $\mathbb{P}(A_j) = \delta$, $j = 1, \dots, K$, and $\mathbb{P}(B) = (K-2)\delta$. Let $V, V_1, \dots, V_K, W_1, \dots, W_K$ be independent random variables, between themselves and of A_1, \dots, A_K, B , such that, for $j = 1, \dots, K$, V_j is uniformly distributed on $[0, \delta]$, V is uniformly distributed on $[\delta, (K-1)\delta]$, and W_j is uniformly distributed on $[(K-1)\delta, 1]$.

Define the random variables

$$p_j := \mathbf{1}_B V + \mathbf{1}_{A_j} V_j + \mathbf{1}_{(B \cup A_j)^c} W_j$$

for an odd number $j \in \{1, \dots, K\}$, and

$$p_j := \mathbf{1}_B (K\delta - V) + \mathbf{1}_{A_j} V_j + \mathbf{1}_{(B \cup A_j)^c} W_j$$

for an even number $j \in \{1, \dots, K\}$. We can easily check that p_1, \dots, p_K are uniformly distributed on $[0, 1]$. Note that $\frac{2}{K} \sum_{k=1}^K p_k \mathbf{1}_B = K\delta \mathbf{1}_B$, and, by writing $A := \bigcup_{j=1}^n A_K$, $\min(p_1, \dots, p_K) \mathbf{1}_A \leq \delta \mathbf{1}_A$. Therefore,

$$\mathbb{P}(F_K^{\text{BA}}(p_1, \dots, p_K) \leq \epsilon) = \mathbb{P}(\min(K \min(p_1, \dots, p_K), 2\bar{p}) \leq K\delta)$$

$$\geq \mathbb{P}(A \cup B) = (2K - 2)\delta = \epsilon - \epsilon/K.$$

Hence, (A.11) holds, and the family F_K^{BA} is asymptotically precise.

The case of the compound BG method is similar, although an explicit construction is more complicated. We fix $\epsilon \in (0, 1/2)$ and let $\delta := \epsilon/K$. Note that, for a uniform random variable U on $[0, 1]$, we have

$$\mathbb{E}(\ln U \mid U \in [0, \epsilon]) = \ln \epsilon - 1.$$

For $K \geq 3$, $\ln \delta < \ln \epsilon - 1$, and there exists a number $t(K) > 0$ such that

$$\mathbb{E}(\ln U \mid U \in [\delta, \epsilon - t(K)]) = \ln \epsilon - 1.$$

Clearly, $t(K) \rightarrow 0$ as $K \rightarrow \infty$. Next, take an arbitrary $\eta > 0$. By [7, Corollary 3.4], there exists $K_0 \in \{1, 2, \dots\}$ such that for $K > K_0$, there exist uniform random variables U_1, \dots, U_K on $[\delta, \epsilon - t(K)]$ satisfying

$$q_{1-\eta}(\ln U_1 + \dots + \ln U_K) < \mathbb{E}(\ln U_1 + \dots + \ln U_K) = K(\ln \epsilon - 1).$$

In other words,

$$\mathbb{P}(e\tilde{U} \leq \epsilon) \geq 1 - \eta,$$

where \tilde{U} is the geometric mean of U_1, \dots, U_K .

Let A_1, \dots, A_K, B be disjoint events with $\mathbb{P}(A_j) = \delta$, $j = 1, \dots, K$, and $\mathbb{P}(B) = \epsilon - t(K)$. Further, let $V_1, \dots, V_K, W_1, \dots, W_K$ be random variables, independent between themselves and of U_1, \dots, U_K and A_1, \dots, A_K, B , and such that, for $j = 1, \dots, K$, V_j is uniformly distributed on $[0, \delta]$ and W_j is uniformly distributed on $[\epsilon - t(K), 1]$.

Define the following random variables, for $j = 1, \dots, K$:

$$p_j := \mathbf{1}_B U_j + \mathbf{1}_{A_j} V_j + \mathbf{1}_{(B \cup A_j)^c} W_j.$$

We can easily check that p_1, \dots, p_K are uniformly distributed on $[0, 1]$. Note that

$$\mathbb{P}(e\tilde{p} \leq \epsilon \mid B) = \mathbb{P}(e\tilde{U} \leq \epsilon \mid B) \geq 1 - \eta.$$

By setting $A := \bigcup_{j=1}^n A_K$, we have $\min(p_1, \dots, p_K) \mathbf{1}_A \leq \delta \mathbf{1}_A$. Therefore,

$$\begin{aligned} \mathbb{P}(F_K^{\text{BG}}(p_1, \dots, p_K) \leq 2\epsilon) &= \mathbb{P}(\min(K \min(p_1, \dots, p_K), e\tilde{p}) \leq \epsilon) \\ &\geq \mathbb{P}(A) + \mathbb{P}(B)(1 - \eta) = \epsilon + (\epsilon - t(K))(1 - \eta). \end{aligned}$$

Since η is arbitrary and $t(K) \rightarrow 0$ as $K \rightarrow \infty$, we know that

$$\lim_{K \rightarrow \infty} (\sup \{ \mathbb{P}(F_K^{\text{BG}}(U_1, \dots, U_K) \leq 2\epsilon) \mid U_1, \dots, U_K \in \mathcal{U} \}) = 2\epsilon,$$

and so the family F_K^{BG} is asymptotically precise.

A lemma on moments of p -values from z -tests

LEMMA A.2. Let Z be a standard normal random variable and $z \in \mathbb{R}$. For all $\sigma > 0$, $\mathbb{E}[(\Phi(\sigma Z + z))^r] < \infty$ if and only if $r > -1/\sigma^2$ or both $r = -1/\sigma^2$ and $z > 0$.

Proof. It suffices to investigate the case $r < 0$. In this case, only small values of $\Phi(\sigma Z + z)$ matter. To analyze whether

$$\mathbb{E}[(\Phi(\sigma Z + z))^r] = \int_{-\infty}^{\infty} (\Phi(\sigma x + z))^r \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \quad (\text{A.12})$$

Algorithm 1. Generalized Bonferroni–Holm procedure

Inputs: significance level $\epsilon > 0$; parameter $r < -1$ (or, w.l.o.g., (B.1));

sequence of p-values p_1, \dots, p_K ordered as $p_{(1)} \leq \dots \leq p_{(K)}$.

For $k = 1, \dots, K$

reject := **true**

$I := \{k\}$

 For $i = K, \dots, 1, 0$

 If $\frac{r}{r+1} |I|^{1+1/r} M_{r,|I|}(p_{(I)}) > \epsilon$

reject := **false**

$I := I \cup \{i\}$

 If **reject** = **true**

 reject H_k

is finite, it suffices to look at the limiting behavior of $\Phi(\sigma x + z)$ as $x \rightarrow -\infty$. L'Hôpital's rule gives $\Phi(y) \sim -\frac{1}{y} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ as $y \rightarrow -\infty$. Hence, we have, for some constant $C > 0$, as $x \rightarrow -\infty$,

$$\begin{aligned} (\Phi(\sigma x + z))^r e^{-x^2/2} &\sim \left(-\frac{1}{\sigma x + z} \frac{1}{\sqrt{2\pi}} \right)^r e^{-r(\sigma x + z)^2/2 - x^2/2} \\ &\sim C x^{-r} e^{-(r\sigma^2 + 1)x^2/2 + r\sigma z x}. \end{aligned}$$

Therefore, if $r\sigma^2 + 1 > 0$, the integral in (A.12) is finite, and if $r\sigma^2 + 1 < 0$, the integral in (A.12) is infinite. If $r\sigma^2 + 1 = 0$, then $z \leq 0$ leads to an infinite integral in (A.12), and $z > 0$ leads to a finite integral in (A.12). \square

B. APPLICATION TO TESTING MULTIPLE HYPOTHESES

In this section we apply the results of the previous section, concerning multiple testing of a single hypothesis, to testing multiple hypotheses. Namely, we will arrive at a generalization of the Bonferroni–Holm procedure [14]. Fix a parameter

$$r \leq \frac{\ln K}{1 - \ln K} \quad (\text{B.1})$$

(cf. (A.6)); the Bonferroni–Holm case is $r := -\infty$.

Suppose p_k is a p-value for testing a composite null hypothesis H_k (meaning that, for any $\epsilon \in (0, 1)$, $\mathbb{P}(p_k \leq \epsilon) \leq \epsilon$ under H_k). For $I \subseteq \{1, \dots, K\}$, let H_I be the hypothesis

$$H_I := (\cap_{k \in I} H_k) \cap (\cap_{k \in \{1, \dots, K\} \setminus I} H_k^c),$$

where H_k^c is the complement of H_k .

Fix a significance level ϵ . Let us reject H_I when

$$\frac{r}{r+1} |I|^{1+1/r} M_{r,|I|}(p_I) \leq \epsilon,$$

where p_I is the vector of p_k for $k \in I$; by Proposition 1, the probability of error will be at most ϵ . If we now reject H_k when all H_I with $I \supseteq \{k\}$ are rejected, the family-wise error rate (FWER) will be at most ϵ . This gives the procedure described as Algorithm 1, in which (k_1, \dots, k_K) stands for a fixed permutation of $\{1, \dots, K\}$ such that $p_{k_1} \leq \dots \leq p_{k_K}$.

An alternative representation of the generalized Bonferroni–Holm procedure given as Algorithm 1 is in terms of adjusting the p-values p_1, \dots, p_K to new p-values p_1^*, \dots, p_K^* that are valid in the sense of the FWER: we are guaranteed to have $\mathbb{P}(\min_{k \in I} p_k^* \leq \epsilon) \leq \epsilon$ for all $\epsilon \in (0, 1)$, where

Algorithm 2. Generalized Bonferroni–Holm procedure for adjusting p-values

Inputs: parameter $r < -1$ (or, w.l.o.g., (B.1));

sequence of p-values p_1, \dots, p_K ordered as $p_{(1)} \leq \dots \leq p_{(K)}$.

For $k = 1, \dots, K$

$p_k^* := 0$

$I := \{k\}$

For $i = K, \dots, 1, 0$

If $\frac{r}{r+1} |I|^{1+1/r} M_{r,|I|}(p_{(I)}) > p_k^*$

$p_k^* := \frac{r}{r+1} |I|^{1+1/r} M_{r,|I|}(p_{(I)})$

$I := I \cup \{i\}$

I is the set of the indices k of the true hypotheses H_k . The adjusted p-values can be defined as

$$p_k^* := \max_{k \in I \subseteq \{1, \dots, K\}} \frac{r}{r+1} |I|^{1+1/r} M_{r,|I|}(p_I)$$

and computed using Algorithm 2.

If we do not insist on controlling the FWER, we can still use our ways of combining p-values instead of Bonferroni's in more flexible procedures for testing multiple hypotheses, such as those described in [10].

865