TRADE-OFF BETWEEN VALIDITY AND EFFICIENCY OF MERGING P-VALUES UNDER ARBITRARY DEPENDENCE

Yuyu Chen[1], Peng Liu[2], Ken Seng Tan[3] and Ruodu Wang[1]

[1] *University of Waterloo,* [2] *University of Essex,* [3] *Nanyang Technological University*

Supplementary Material

This document supplements the paper entitled "Trade-off between validity and efficiency of merging p-values under arbitrary dependence". Section S1 contains simulation studies and a real data example. The numerical results show the advantages of the Simes, the averaging and the Cauchy combination methods against dependence uncertainty of p-values. Section S2 presents additional remarks. Section S3 includes the proofs of all the theorems and propositions in the paper. Section S4 contains additional tables for the price for validity.

S1 Simulations and a real data example

S1.1 Simulation studies

We conduct K one-sided z-tests of the null hypothesis: $\mu_i = 0$ against the alternative hypothesis $\mu_i > 0$, $i = 1, \ldots, K$, using the test statistic X_i and the p-value p_i from the ith test, $i = 1, \ldots, K$. The tests are formulated as the following:

$$p_i = \Phi(X_i), \quad X_i = \rho Z + \sqrt{1-\rho^2} Z_i - \mu_i, \quad i = 1, \ldots, K.$$

where Φ is the standard normal distribution function, $Z, Z_1, \ldots, Z_K$ are iid standard normal random variables, $\mu_i \geq 0$, $i = 1, \ldots, K$, and ρ is a parameter in $[0,1]$. Note that for $\rho = 0$, the p-variables are independent, and $\rho = 1$ corresponds to the case where p-variables are comonotonic.

Let $K \in \{50, 200\}$ and set the significance level $\varepsilon = 0.01$. To see how different dependence structures and signals affect the size and the power for various methods using both VAD and VSD thresholds, the rejection probabilities (RPs) are computed over $\rho \in [0,1]$ under the following four cases:

(i) (no signal) 100% of μ_i's are 0;

(ii) (needle in a haystack) 98% of μ_i's are 0 and 2% of μ_i's are 4;

(iii) (sparse signal) 90% of μ_i's are 0 and 10% of μ_i's are 3;

(iv) (dense signal) 100% of μ_i's are 2.

The RP corresponds to the size under case (i), and it corresponds to the power under (ii), (iii) and (iv). The RP is computed as the ratio between the number of the combined values which are less than the critical threshold and the number of simulations for some $\rho \in [0,1]$, that is,

$$\mathrm{RP} = \frac{\sum_{i=1}^{N} \mathbb{1}_{\{F_i < g(\varepsilon)\}}}{N},$$

where N is the number of simulations and is equal to 15000 in our study, F_i is the realized value of the combining function for the i-th simulation, $i = 1, \ldots, N$, and $g(\varepsilon)$ is the corresponding critical value. For $\rho \in [0,1]$, graphs of RPs for different combining methods are drawn using VAD thresholds and VSD thresholds. Some observations from Figures 1-4 are made below, and those on the averaging methods using $M_{r,K}$ are consistent with the observations in Vovk and Wang (2020a).

1. All VAD methods give sizes less than $\varepsilon = 0.01$ as expected. Using VAD thresholds, the Bonferroni, the harmonic averaging, the Cauchy combination and the Simes methods have good powers.

2. The Simes method using thresholds b_F or c_F reports the right size for all values of ρ. Sarkar (1998) showed the validity of the Simes method in the so-called MTP_2 class including multivariate normal distributions with nonnegative correlations (the setting of our simulation).

3. Using thresholds b_F or c_F, the harmonic averaging and Cauchy combination methods perform similarly with sizes possibly larger than 0.01 (see Theorems 2 and 3).

4. The geometric averaging method using b_F and the Bonferroni and negative-quartic methods using c_F do not yield correct sizes under model misspecification, and the sizes increase rapidly as the misspecification gets bigger.

5. Using b_F or c_F, the harmonic averaging, the Cauchy combination and the Simes methods have good performances on capturing the signals.

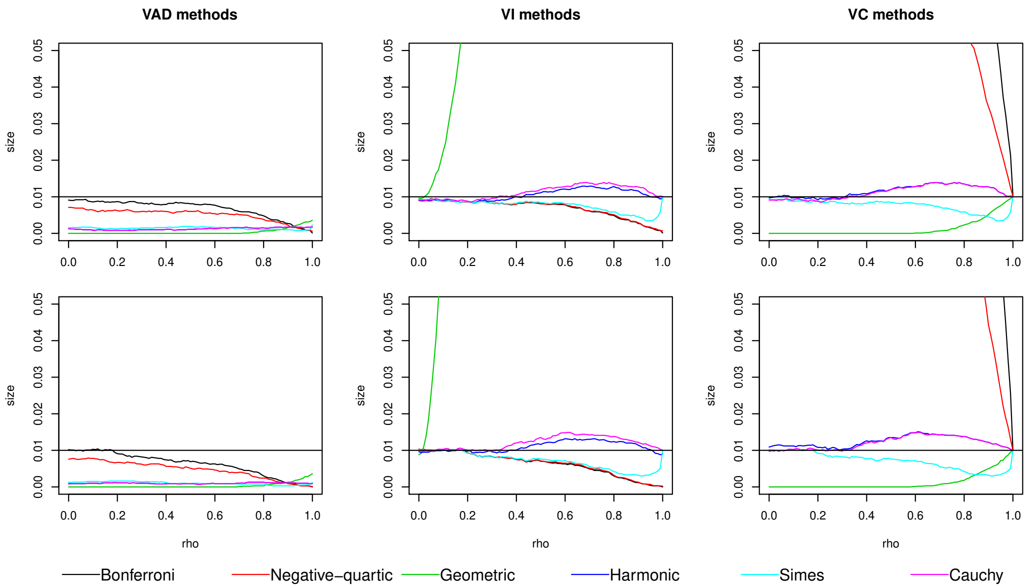

Figure 1: Case (i): size (top: $K = 50$, bottom: $K = 200$)

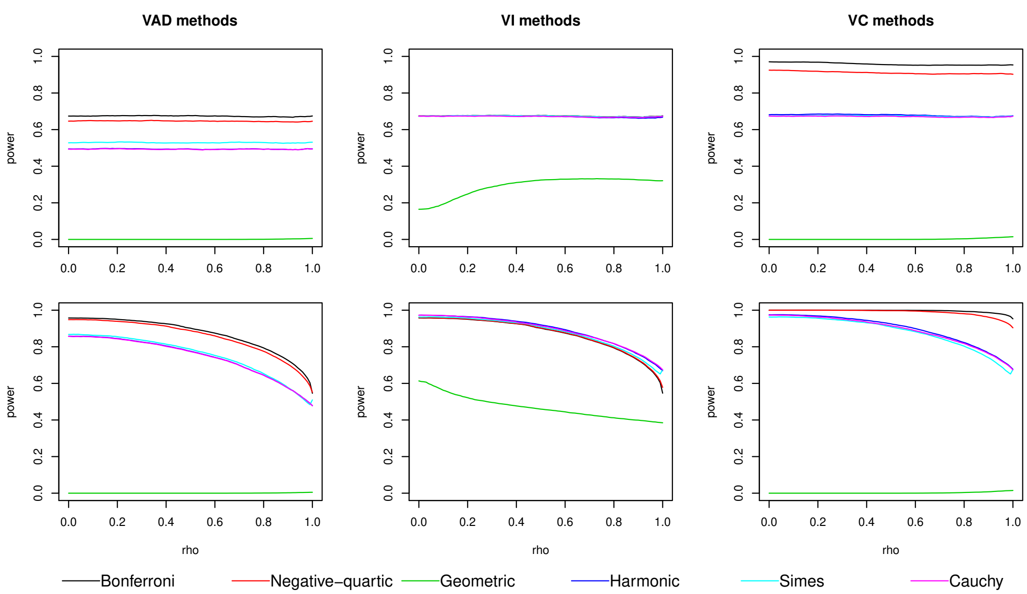

Figure 2: Case (ii): needle in a haystack (top: $K = 50$, bottom: $K = 200$)

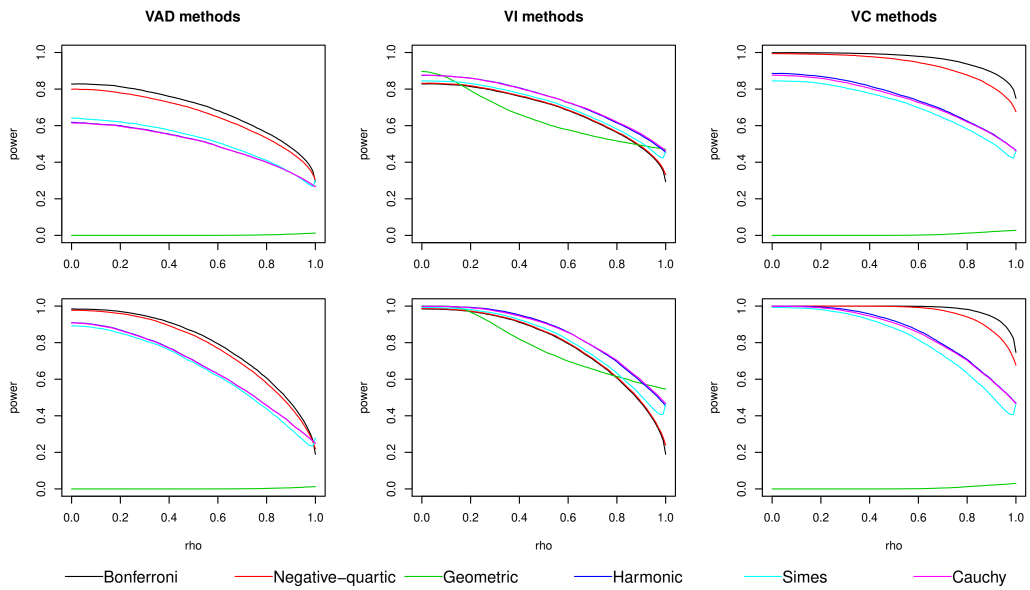

Figure 3: Case (iii): sparse signal (top: $K = 50$, bottom: $K = 200$)

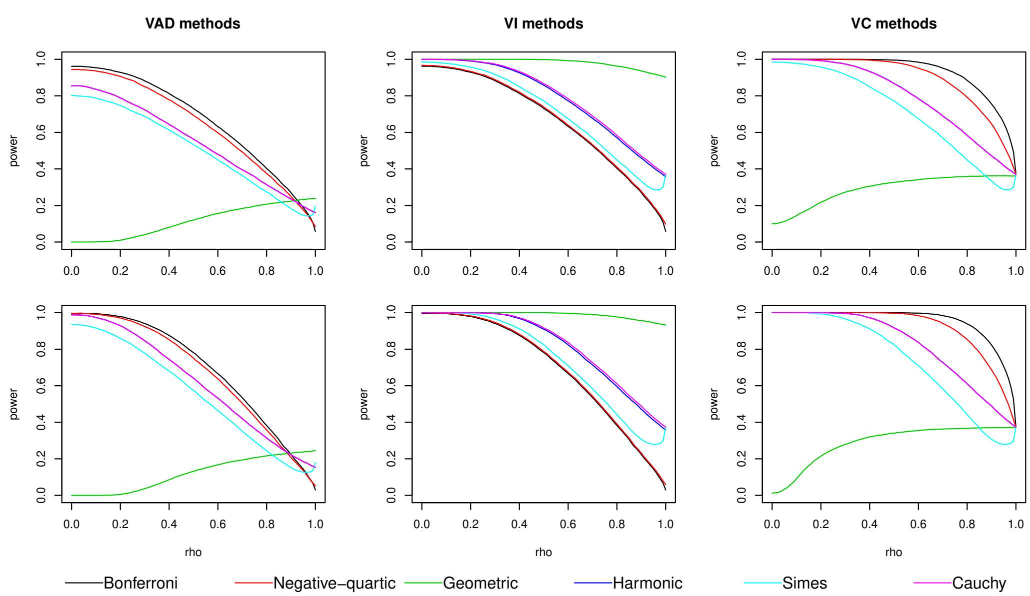

Figure 4: Case (iv): dense signal (top: $K = 50$, bottom: $K = 200$)

S1.2 Real data analysis

We apply several merging methods to a genomewide study to compare their performances. We use the dataset of p-values of Storey and Tibshirani (2003) which contains 3170 p-values computed based on the data from Hedenfalk et al. (2001) for testing whether genes are differentially expressed between BRCA1- and BRCA2-mutation-positive tumors. As mentioned in Section 2, $g^{-1} \circ F(P_1, \ldots, P_K)$ is a p-variable if the threshold g is strictly increasing, and it is the quantity we choose to compare combined p-values for different methods.

For each method, we calculate the combined p-value, and remove the smallest p-value from the dataset. Repeat this procedure until the resulting combined p-value loses significance. Using the Bonferroni combining function, this leads to the Bonferroni-Holm (BH) procedure (Holm (1979)); thus we mimic the BH procedure for other methods in a naive manner. The rough interpretation is to report the number of significant discoveries (this procedure generally does not control the family-wise error rate (FWER); to control FWER one needs to use a generalized BH procedure as in Vovk and Wang (2020a) or Goeman et al. (2019). This procedure can be seen as a lower confidence bound from a closed testing perspective). For a visual comparison of detection power, the combined p-values against the numbers of removed p-values are plotted in Figure 5, where we use both the VAD and the VI thresholds (comonotonicity is obviously unrealistic here). In the third panel of Figure 5, we present the number of omitted p-values in log-scale for better visualization.

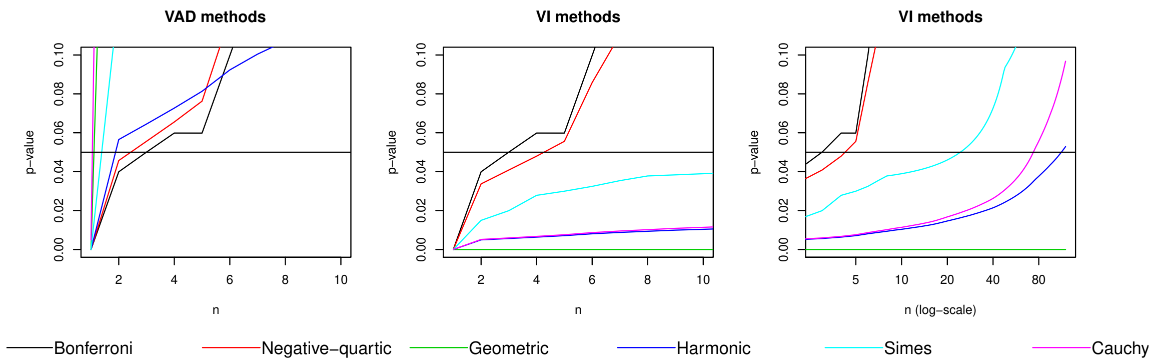

Figure 5: Combined p-value after removing n smallest p-values

All VAD methods lose significance at $\varepsilon = 0.05$ after omitting the first or the second smallest p-value (the smallest p-value is 0 and the second smallest is 1.26×10^{-5}). Using thresholds b_F for independence, the Bonferroni and the negative quartic methods behave similarly to their VAD versions (as their price for validity is close to 1). In contrast, the Simes, the Cauchy combination and the harmonic averaging methods lose significance at $\varepsilon = 0.05$ after removing around 20, 70 and 110 p-values respectively. The geometric averaging method (Fisher's) exceeds 0.05 only after removing around 400 p-values. However, this method relies heavily on the independence assumption, which is impossible to verify from just one set of p-values.

S2 Additional remarks

Remark 1 (Section 4)**.** The property of IC-balance should be seen as a necessary but not sufficient condition for a merging method to be insensitive to dependence between independence and comonotonicity. As shown by Sarkar (1998), the Simes method is valid for positive regression dependence, which is a large spectrum of dependence structures connecting independence and comonotonicity (larger than (4.1)); on the other hand, the Cauchy combination method using VI threshold is valid under a bivariate Gaussian assumption asymptotically but not precisely (Liu and Xie (2020)); see Theorem 2 below and the simulation studies in Section S1 of the supplementary material. Instead of arguing for the practical usefulness of IC-balance, we emphasize it as a necessary condition for insensitivity to dependence. The main aim of Theorem 1 is, via this necessary condition, to pin down the unique role of the Simes and the Cauchy combination methods among their respective generalized classes, thus justifying their advantages with respect to dependence.

Remark 2 (Section 5)**.** We note that the equivalence

$$\mathbb{P}\left(M_{\mathcal{C},K}(U_1,\ldots,U_K)<\varepsilon\right)\sim\mathbb{P}\left(M_{-1,K}(U_1,\ldots,U_K)<\varepsilon\right)$$

in (5.1) does not always hold under arbitrary dependence structures. Since the Cauchy distribution is symmetric, it is possible that $\mathbb{P}(\mathcal{C}^{-1}(U_1)+\cdots+\mathcal{C}^{-1}(U_K)=0)=1$ for some $U_1,\ldots,U_K\in\mathcal{U}$, implying $\mathbb{P}(M_{\mathcal{C},K}(U_1,\ldots,U_K)<1/2)=0$. Indeed, Theorem 4.2 of Puccetti et al. (2019) implies that there exist K standard Cauchy random variables whose sum is a constant c, for each $c \in [-K\log(K-1)/\pi, K\log(K-1)/\pi]$. On the other hand, $\mathbb{P}(M_{-1,K}(U_1,\dots,U_K) < \varepsilon) > 0$ for all $\varepsilon > 0$ and all $U_1,\dots,U_K \in \mathcal{U}$. Thus, $\mathbb{P}\left(M_{\mathcal{C},K}(U_1,\dots,U_K) < \varepsilon\right) \sim \mathbb{P}\left(M_{-1,K}(U_1,\dots,U_K) < \varepsilon\right)$ does not hold.

Remark 3 (Section 5)**.** The equivalence in Theorem 2 (ii) relies on the p-variables being uniform on $[0,1]$. For p-variables that are stochastically larger than uniform, the behaviour of the Cauchy combination method and that of the harmonic averaging method may diverge; nevertheless, by Theorem 2 (i), for a realized vector of p-values with at least one very small component, the two methods would produce similar values.

S3 Proofs of theorems and propositions

S3.1 Proof of Proposition 1

By definition, we have

$$a_F(\varepsilon) = \inf\{q_\varepsilon(F(U_1,\dots,U_K)) \mid U_1,\dots,U_K \in \mathcal{U}\},\ \varepsilon \in (0,1).$$

We shall show

$$a_F(\varepsilon) = \inf\{q_1(F(V_1,\dots,V_K)) \mid V_1,\dots,V_K \in \mathcal{U}_\varepsilon\},\ \varepsilon \in (0,1), \tag{S3.1}$$

where $\mathcal{U}_\varepsilon$ denotes the collection of all uniform random variables distributed on $[0,\varepsilon]$. Denote by $S = F(U_1,\dots,U_K)$ and $G_S^{-1}(t) = q_t(S),\ t \in (0,1]$. We can find $U_S \in \mathcal{U}$

such that $G_S^{-1}(U_S) = S$ a.s. (e.g., Lemma A.32 of Föllmer and Schied (2016)). Let $f_i(t) = \mathbb{P}\left(U_i \le t | U_S < \varepsilon\right),\ t \in [0,1]$. Then $f_i(U_i)$ conditionally on $U_S < \varepsilon$ is a uniform random variable on $[0,1]$ and $V_i^\varepsilon := \varepsilon f_i(U_i)$ conditionally on $U_S < \varepsilon$ is a uniform random variable on $[0,\varepsilon]$. We construct the following two random variables:

$$S_1 = S\mathbb{1}_{\{U_S<\varepsilon\}} + d\mathbb{1}_{\{U_S\ge\varepsilon\}},\ S_2 = F(V_1^\varepsilon,\dots,V_n^\varepsilon)\mathbb{1}_{\{U_S<\varepsilon\}} + d\mathbb{1}_{\{U_S\ge\varepsilon\}}, \tag{S3.2}$$

where $d > F(\varepsilon,\dots,\varepsilon)$. Noting the fact that $\varepsilon f_i(t) = \mathbb{P}(U_i \le t, U_S < \varepsilon) \le t,\ t \in [0,1]$ and F is increasing, we have $S_1 \ge S_2$. Hence $q_\varepsilon(S_1) \ge q_\varepsilon(S_2)$. Moreover, direct calculation shows $q_\varepsilon(S) = q_\varepsilon(S_1)$. Thus $q_\varepsilon(S) \ge q_\varepsilon(S_2)$. Let $\hat V_1,\dots,\hat V_n$ be uniform random variables on $[0,\varepsilon]$ such that $(\hat V_1,\dots,\hat V_n)$ has the joint distribution identical to the conditional distribution of $(V_1^\varepsilon,\dots,V_n^\varepsilon)$ on $U_S < \varepsilon$. Hence, for $x < d$,

$$\begin{aligned}\mathbb{P}(S_2 \le x) &= \mathbb{P}(F(V_1^\varepsilon,\dots,V_n^\varepsilon) \le x, U_S < \varepsilon)\\ &= \varepsilon\mathbb{P}(F(V_1^\varepsilon,\dots,V_n^\varepsilon) \le x | U_S < \varepsilon)\\ &= \varepsilon\mathbb{P}(F(\hat V_1,\dots,\hat V_n) \le x).\end{aligned}$$

This implies $q_\varepsilon(S_2) = q_1(F(\hat V_1,\dots,\hat V_n))$. Thus we have

$$a_F(\varepsilon) \ge \inf\{q_1(F(V_1,\dots,V_K)) \mid V_1,\dots,V_K \in \mathcal{U}_\varepsilon\}.$$

We next show "$\le$" in (S3.1). Take $V_1,\dots,V_n \in \mathcal{U}_\varepsilon$ and $U \in \mathcal{U}$ such that U is independent of $V_1,\dots,V_n$. Let $\hat U_i = V_i\mathbb{1}_{\{U<\varepsilon\}} + U\mathbb{1}_{\{U\ge\varepsilon\}},\ i = 1,2,\dots,n$. It is clear that $\hat U_i \in \mathcal{U},\ i = 1,2,\dots,n$ and $F(\hat U_1,\dots,\hat U_n) = F(V_1,\dots,V_n)\mathbb{1}_{\{U<\varepsilon\}} + F(U,\dots,U)\mathbb{1}_{\{U\ge\varepsilon\}}$.

Noting that F is increasing, we have $q_1(F(V_1,\ldots,V_n)) = q_\varepsilon(F(\hat{U}_1,\ldots,\hat{U}_n))$. This implies

$$a_F(\varepsilon) \le \inf\{q_1(F(V_1,\ldots,V_K)) \mid V_1,\ldots,V_K \in \mathcal{U}_\varepsilon\}.$$

Therefore, (S3.1) holds. By (S3.1) and the homogeneity of F we have that for $\varepsilon \in (0,1)$,

$$\begin{aligned} a_F(\varepsilon) &= \inf\{q_1(F(V_1,\ldots,V_K)) \mid V_1,\ldots,V_K \in \mathcal{U}_\varepsilon\} \\ &= \inf\{q_1(F(\varepsilon U_1,\ldots,\varepsilon U_K)) \mid U_1,\ldots,U_K \in \mathcal{U}\} \\ &= \varepsilon \inf\{q_1(F(U_1,\ldots,U_K)) \mid U_1,\ldots,U_K \in \mathcal{U}\}. \end{aligned}$$

This completes the proof. □

S3.2 Proof of Proposition 2

It is well known that the Bonferroni correction yields $a_F(\varepsilon) = \varepsilon/K$. Also, since the average of identical objects is itself, $c_F(\varepsilon) = \varepsilon$ for any averaging method, including the Bonferroni method. For iid standard uniform random variables $V_1,\ldots,V_K$, we have $\mathbb{P}(\min\{V_1,\ldots,V_K\} \le x) = 1-(1-x)^K$. Therefore, $b_F(\varepsilon) = 1-(1-\varepsilon)^{1/K}$ for $\varepsilon \in (0,1)$. □

S3.3 Proof of Proposition 3

(a) Suppose $r < 0$. We first fix K and find the asymptotic of b_r as $\varepsilon \downarrow 0$ satisfying

$$\mathbb{P}\left(\sum_{i=1}^{K} P_i^r \geq K\left(b_r(\varepsilon)\right)^r\right) = \varepsilon.$$

Observe that the random variables P_i^r, $i = 1, \ldots, K$, follow a common Pareto distribution with cdf $\mathbb{P}(P_i^r \leq x) = 1 - x^{1/r}$, $x \in (1, \infty)$, $i = 1, \ldots, K$. Note that the tail probability of the sum of iid Pareto random variables is asymptotically the same as that of the maximum of the iid Pareto random variables (e.g., Embrechts et al. (2013), Corollary 1.3.2). Hence

$$\lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}\left(\sum_{i=1}^{K} P_i^r \geq K\left(b_r(\varepsilon)\right)^r\right)}{\mathbb{P}\left(\max\{P_1^r, \ldots, P_K^r\} > K\left(b_r(\varepsilon)\right)^r\right)} = \lim_{\varepsilon \downarrow 0} \frac{\varepsilon}{1 - \left(1 - K^{\frac{1}{r}} b_r(\varepsilon)\right)^K} = 1.$$

This implies

$$b_r(\varepsilon) \sim \frac{1 - (1 - \varepsilon)^{\frac{1}{K}}}{K^{\frac{1}{r}}} \sim K^{-1-1/r}\varepsilon, \qquad \text{as } \varepsilon \downarrow 0.$$

The case $K \to \infty$ follows directly from the generalized central limit theorem (e.g., Theorem 1.8.1 of Samorodnitsky (2017)).

(b) If $r = 0$, in a similar way, we first have,

$$\mathbb{P}\left(2\sum_{i=1}^{K} \log\frac{1}{P_i} \geq 2K \log\frac{1}{b_r(\varepsilon)}\right) = \varepsilon.$$

The random variable $\log\frac{1}{P_i}$, $i = 1, \ldots, K$, follows exponential distribution with parameter 1. Thus $2\sum_{i=1}^{K} \log\frac{1}{P_i}$ follows a chi-square distribution with parameter

$2K$. We denote $q_\alpha(\chi^2_\nu)$ the α-quantile of the chi-square distribution with ν degrees of freedom. Hence

$$b_r(\varepsilon) = \exp\left(-\frac{1}{2K} q_{1-\varepsilon}\left(\chi^2_{2K}\right)\right).$$

(c) If $r > 0$, using the result of Wang (2005), we have for $0 \le x \le K^{-r}$,

$$\begin{aligned}\mathbb{P}\left(M_{r,K}(U_1, \ldots, U_K) \le x\right) &= \mathbb{P}\left(\sum_{i=1}^K U_i^r \le K x^r\right) \\ &= \lambda\left\{(x_1, \ldots, x_K) : \sum_{i=1}^K x_i^r \le K x^r,\ x_1, \ldots, x_K \ge 0\right\} \\ &= \frac{(\Gamma(1+1/p))^K}{\Gamma(1+K/p)} K^{K/r} x^K,\end{aligned}$$

where λ is the Lebesgue measure. This implies that if $\varepsilon \le \frac{(\Gamma(1+1/p))^K}{\Gamma(1+K/p)}$,

$$b_r(\varepsilon) = \frac{(\Gamma(1+K/p))^{1/K} \varepsilon^{1/K}}{K^{1/r}\Gamma(1+1/p)}. \tag{S3.3}$$

The asymptotic behaviour of $b_r(\varepsilon)$ for fixed $\varepsilon \in (0,1)$ as $K \to \infty$ can be obtained by the Central Limit Theorem. Note that the random variables P_i^r, $i = 1, \ldots, K$, follow a common Beta distribution with mean and variance given by, respectively,

$$\mu = (r+1)^{-1}, \text{ and } \sigma^2 = r^2(1+2r)^{-1}(1+r)^{-2}.$$

The Central Limit Theorem gives $(\sum_{i=1}^K P_i^r - K\mu)/\sqrt{K}\sigma \xrightarrow{\text{d}} \mathrm{N}(0,1)$. Hence

$$b_r(\varepsilon) \sim \left(\frac{\sigma}{\sqrt{K}} \Phi^{-1}(\varepsilon) + \mu\right)^{\frac{1}{r}}, \text{ as } K \to \infty,$$

where Φ^{-1} is the inverse of the standard normal distribution function. □

S3.4 Proof of Proposition 4

By symmetry of the standard Cauchy distribution,

$$\begin{aligned} a_F(\varepsilon) &= \mathcal{C}\left(\inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \mathcal{C}^{-1}(U_i)\right) \mid U_1,\dots,U_K \in \mathcal{U}\right\}\right) \\ &= \mathcal{C}\left(\frac{-1}{K}\sup\left\{q_{1-\varepsilon}\left(\sum_{i=1}^K \mathcal{C}^{-1}(U_i)\right) \mid U_1,\dots,U_K \in \mathcal{U}\right\}\right). \end{aligned}$$

Moreover, $\mathcal{C}^{-1}(U_i),\ i = 1,\dots,K$, follow the standard Cauchy distribution with decreasing density on $[\mathcal{C}^{-1}(1-\varepsilon),\infty]$ for $\varepsilon \in (0,1/2)$. The proposition follows directly from applying Corollary 3.7 of Wang et al. (2013). □

S3.5 Proof of Theorem 1

(i) IC-balance of $M_{\phi,K}$ for all $K \in \{2,3,\dots\}$ is equivalent to $\frac{1}{K}\sum_{i=1}^K \phi(V_i) \overset{\mathrm{d}}{=} \phi(U)$ for all $K \in \{2,3,\dots\}$, which is further equivalent to the fact that $\phi(U)$ follows a strictly 1-stable distribution. We know that strictly 1-stable distributions are Cauchy distributions (see, e.g., Theorem 14.15 of Sato (1999)). This proves the statement of part (i).

(ii) For the Simes function $S_{\alpha,K} = S_K$, $\alpha_i = i$ for $i \in \{1,\dots,K\}$ and $b_F(x) = c_F(x) = x$ for $x \in [0,1]$. Therefore, $S_{\alpha,K}$ is IC-balanced.

Below we show the opposite direction of the statement. For $n \in \{2,\dots,K\}$, let $V_{(1)},\dots,V_{(n)}$ be the order statistics for n independent standard uniform random variables $V_1,\dots,V_n$. Let $(X_1,\dots,X_{n-1}) = (V_{(1)}/V_{(n)},\dots,V_{(n-1)}/V_{(n)})$ which is identically distributed as the order statistics for $n-1$ independent standard uniform random variables, independent of $V_{(n)}$. Hence, for $x \in (0, 1/\alpha_n)$,

$$\begin{aligned}
&\mathbb{P}\left(S_{\alpha,n}(V_1,\ldots,V_n) > x\right) \\
&= \mathbb{P}\left(V_{(1)} > x\alpha_1, \ldots, V_{(n-1)} > x\alpha_{n-1}, V_{(n)} > x\alpha_n\right) \\
&= \mathbb{P}\left(X_1 > x\alpha_1/V_{(n)}, \ldots, X_{n-1} > x\alpha_{n-1}/V_{(n)}, V_{(n)} > x\alpha_1\right) \\
&= \int_{x\alpha_n}^{1} \mathbb{P}\left(X_1 > x\alpha_1/p, \ldots, X_{n-1} > x\alpha_{n-1}/p\right) np^{n-1}\,\mathrm{d}p \\
&= \int_{x\alpha_n}^{1} \mathbb{P}\left(S_{\alpha,n-1}(V_1,\ldots,V_{n-1}) > x/p\right) np^{n-1}\,\mathrm{d}p, \qquad \text{(S3.4)}
\end{aligned}$$

where for simplicity we use $S_{\alpha,n-1}$ for $S_{(\alpha_1,\ldots,\alpha_{n-1}),n-1}$. Note that

$$\mathbb{P}\left(S_{\alpha,1}(V_1) > x\right) = 1 - \alpha_1 x, \quad x \in (0, 1/\alpha_1). \qquad \text{(S3.5)}$$

Plugging (S3.5) in (S3.4), we obtain that $\mathbb{P}\left(S_{\alpha,2}(V_1, V_2) > x\right)$ is a polynomial function of x of degree less than or equal to 2. Recursively, using (S3.4) we are able to show that the function $\mathbb{P}\left(S_{\alpha,n}(V_1,\ldots,V_n) > x\right)$ for $x \in (0, 1/\alpha_n)$ is a polynomial of x of degree less than or equal to n for $n = 2, \ldots, K$. Hence, there exist K constants $\beta_0, \ldots, \beta_{K-1}$ such that

$$\mathbb{P}\left(S_{\alpha,K-1}(V_1,\ldots,V_{K-1}) > x\right) = \sum_{i=0}^{K-1} \beta_i x^i, \quad x \in (0, 1/\alpha_{K-1}).$$

Moreover, noting that $S_{\alpha,K}$ is IC-balanced, we have

$$\int_{x\alpha_K}^{1} \mathbb{P}\left(S_{\alpha,K-1}(V_1,\ldots,V_{K-1}) > x/p\right) Kp^{K-1}\,\mathrm{d}p = \mathbb{P}\left(S_{\alpha,K}(U,\ldots,U) > x\right) = 1 - x\alpha_K,$$

for $x \in (0, 1/\alpha_K)$. Therefore, we have

$$\int_{x\alpha_K}^1 \left(\sum_{i=0}^{K-1} \beta_i x^i p^{-i} \right) K p^{K-1} \, \mathrm{d}p = 1 - x\alpha_K,$$

which implies that for $x \in (0, 1/\alpha_K)$,

$$\sum_{i=0}^{K-1} \frac{K\beta_i}{K-i} x^i - \left(\sum_{i=0}^{K-1} \frac{K\beta_i}{K-i} \alpha_K^{K-i} \right) x^K = 1 - x\alpha_K.$$

Solving the above equation, we get $\beta_0 = 1$, $\beta_1 = -\frac{K-1}{K}\alpha_K$ and $\beta_2 = \cdots = \beta_{K-1} = 0$. Consequently,

$$\mathbb{P}\left(S_{\alpha,K-1}(V_1, \ldots, V_{K-1}) > x \right) = 1 - \frac{K-1}{K} \alpha_K x, \quad x \in (0, 1/\alpha_{K-1}).$$

Recursively, using (S3.4) we have

$$\mathbb{P}\left(S_{\alpha,n}(V_1, \ldots, V_n) > x \right) = 1 - \frac{n}{K} \alpha_K x, \quad x \in (0, 1/\alpha_n) \tag{S3.6}$$

for $n = 1, \ldots, K$, which gives, using (S3.5),

$$\alpha_K = K\alpha_1. \tag{S3.7}$$

Inserting (S3.6) into (S3.4), we obtain, for $x \in (0, 1/\alpha_n)$ and $n = 2, \ldots, K$,

$$\begin{aligned} 1 - \frac{n}{K}\alpha_K x &= \int_{x\alpha_n}^1 \left(1 - \frac{n-1}{K} \alpha_K x p^{-1} \right) n p^{n-1} \, \mathrm{d}p \\ &= 1 - \frac{n}{K}\alpha_K x + \left(\frac{n}{K} \alpha_K \alpha_n^{n-1} - \alpha_n^n \right) x^n. \end{aligned}$$

Consequently,

$$\alpha_n = \frac{n}{K}\alpha_K, \quad n = 2, \ldots, K,$$

which together with (S3.7) implies $\alpha_n = n\alpha_1, \ k = 1, \dots, K$. This gives the desired statement. □

In the following example, we shall employ several theorems from Sato (1999). To make our paper more self-contained, we display the useful part of these theorems as below.

Theorem 8.1 in Sato (1999): μ is an infinitely divisible distribution in $\mathbb{R}$ if and only if there exist $d \geq 0$, $\gamma \in \mathbb{R}$ and a measure ν on $\mathbb{R}$ satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}}(|x|^2 \wedge 1)\nu(\,\mathrm{d}x) < \infty$, such that the characteristic function of μ is

$$\hat{\mu}(z) = \exp\left(-\frac{1}{2}dz^2 + i\gamma z + \int_{\mathbb{R}}(e^{izx} - 1 - izx\mathbb{1}_{[-1,1]}(x))\nu(\,\mathrm{d}x)\right), \ z \in \mathbb{R}, \quad \text{(S3.8)}$$

where $\mathbb{1}_{[-1,1]}(\cdot)$ is the indicator function and $i^2 = -1$.

Theorem 27.16 in Sato (1999): Suppose μ satisfies (S3.8). If $d = 0$ and ν is discrete with total measure infinite, then μ is a continuous distribution.

Example 1 (IC-balanced generalized mean for a finite K)**.** We show that IC-balance of $M_{\phi,K}$ for a finite K does not imply $M_{\phi,K}$ that ϕ is the Cauchy quantile function (up to an affine transform). For this purpose, we construct a continuous distribution μ such that

$$\frac{1}{K}\sum_{i=1}^{K} X_i \overset{\mathrm{d}}{=} X, \quad \text{(S3.9)}$$

where X and $X_i, i = 1, \dots, K$ are iid random variables with distribution μ, but μ is not a Cauchy distribution. Define

$$\hat{\mu}(z) = \exp\left(\int_{\mathbb{R}} \left(e^{izx} - 1 - \mathbb{1}_{[-1,1]}(x)\right) \nu(\mathrm{d}x)\right), \ z \in \mathbb{R},$$

where ν is a symmetric measure on $\mathbb{R} \setminus \{0\}$ satisfying

$$\nu(\{K^n\}) = \nu(\{-K^n\}) = K^{-n}, \ n \in \mathbb{Z}, \ \text{and} \ \nu\left(\mathbb{R} \setminus \left(\{0\} \cup \bigcup_{n\in\mathbb{Z}} \{K^n, -K^n\}\right)\right) = 0.$$

It follows from Theorem 8.1 of Sato (1999) that $\hat{\mu}$ is the characterization function of some infinitely divisible distribution μ. Also noting that $\nu(\mathbb{R}\setminus\{0\}) = \infty$, by Theorem 27.16 of Sato (1999) we know that μ is a continuous distribution. By Theorem 14.7 of Sato (1999), $(\hat{\mu}(z))^b = \hat{\mu}(bz), \ z \in \mathbb{R}, b > 0$ holds if and only if

$$T_b\nu(B) = b\nu(B), \ \text{and} \ \int_{1<|x|\leq b} x\nu(\mathrm{d}x) = 0,$$

where $T_b\nu(B) = \nu(b^{-1}B)$ for all Borel sets $B \subset \mathbb{R}$. By symmetry of ν, $\int_{1<|x|\leq b} x\nu(\mathrm{d}x) = 0$ holds for any $b > 0$. However, $T_b\nu(B) = b\nu(B)$ holds only for $b \in \{K^n, n \in \mathbb{Z}\}$. Consequently, $(\hat{\mu}(z))^b = \hat{\mu}(bz), \ z \in \mathbb{R}$ if and only if $b \in \{K^n, n \in \mathbb{Z}\}$. This implies that μ is not a Cauchy distribution (strictly 1-stable distribution) but (S3.9) holds.

S3.6 Proof of Theorem 2

(i) Recall that

$$\mathcal{C}^{-1}(x) = \tan\left(-\frac{\pi}{2} + \pi x\right), \quad x \in (0,1);$$
$$\mathcal{C}(y) = \frac{1}{\pi}\arctan(y) + \frac{1}{2}, \quad y \in \mathbb{R}.$$

Note that $\mathcal{C}^{-1}(x) \sim -1/(\pi x)$ as $x \downarrow 0$ and $\mathcal{C}(y) \sim -1/(\pi y)$ as $y \to -\infty$. For any $\delta_1, \delta_2 \in (0, 1/K)$, there exists $0 < \varepsilon < 1$ and $m < 0$ such that for all $x \in (0, \varepsilon)$ and $y \in (-\infty, m)$,

$$-\frac{(1+\delta_1)}{\pi x} \leq \mathcal{C}^{-1}(x) \leq -\frac{(1-\delta_1)}{\pi x}; \tag{S3.10}$$

$$-\frac{(1-\delta_2)}{\pi y} \leq \mathcal{C}(y) \leq -\frac{(1+\delta_2)}{\pi y}. \tag{S3.11}$$

For $0 < c < 1$, there exists $0 < \varepsilon' < \varepsilon$ such that

$$\sup_{x \in [\varepsilon, c]} \left| \tan\left(-\frac{\pi}{2} + \pi x\right) + \frac{1}{\pi x} \right| \leq \frac{\delta_1}{\pi \varepsilon'}. \tag{S3.12}$$

Take $(p_1, \ldots, p_K)$ such that $p_{(1)} < \varepsilon'$ and $p_{(K)} \leq c < 1$. Let $l = \max\{i = 1, \ldots, K : p_{(i)} < \varepsilon\}$. As a consequence of (S3.10), we have

$$-\sum_{i=1}^{l} \frac{(1+\delta_1)}{\pi p_{(i)}} \leq \sum_{i=1}^{l} \tan\left(-\frac{\pi}{2} + \pi p_{(i)}\right) \leq -\sum_{i=1}^{l} \frac{(1-\delta_1)}{\pi p_{(i)}}.$$

For $j > l$, (S3.12) implies

$$\left| \tan\left(-\frac{\pi}{2} + \pi p_{(j)}\right) + \frac{1}{\pi p_{(j)}} \right| \leq \frac{\delta_1}{\pi \varepsilon'} \leq \frac{\delta_1}{\pi p_{(1)}}.$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^{K} \tan\left(-\frac{\pi}{2} + \pi p_i\right) &\leq -\sum_{i=1}^{l} \frac{(1-\delta_1)}{\pi p_{(i)}} - \sum_{i=l+1}^{K} \frac{1}{\pi p_{(i)}} + \frac{(K-l)\delta_1}{\pi p_{(1)}} \\
&\leq -\sum_{i=1}^{K} \frac{(1-K\delta_1)}{\pi p_{(i)}} \\
&= -\sum_{i=1}^{K} \frac{(1-K\delta_1)}{\pi p_i}.
\end{aligned}$$

Similarly, we can show

$$\sum_{i=1}^{K} \tan\left(-\frac{\pi}{2} + \pi p_i\right) \geq \sum_{i=1}^{K} -\frac{(1+K\delta_1)}{\pi p_i}.$$

Using (S3.11), for any $(p_1, \ldots, p_K)$ satisfying $p_{(1)} < \min(\varepsilon', \frac{K\delta_1 - 1}{K\pi m})$ and $p_{(K)} \leq c < 1$,

$$\frac{1-\delta_2}{1+K\delta_1} M_{-1,K}(p_1, \ldots, p_K) \leq M_{\mathcal{C},K}(p_1, \ldots, p_K) \leq \frac{1+\delta_2}{1-K\delta_1} M_{-1,K}(p_1, \ldots, p_K).$$

We establish the claim by letting $\delta_1, \delta_2 \downarrow 0$, and the above inequalities hold as long as $p_{(1)}$ is sufficiently small.

(ii) The statement

$$\mathbb{P}\left(M_{\mathcal{C},K}(U_1, \ldots, U_K) < \varepsilon\right) \sim \varepsilon \quad \text{as } \varepsilon \downarrow 0$$

follows directly from Theorem 1 of Liu and Xie (2020) by noting that standard Cauchy distribution is symmetric at 0. Below we show $\mathbb{P}\left(M_{-1,K}(U_1, \ldots, U_K) < \varepsilon\right) \sim \varepsilon$ as $\varepsilon \downarrow 0$, based on similar techniques as in Theorem 1 of Liu and Xie (2020). Observe that

$$\mathbb{P}\left(M_{-1,K}(U_1, \ldots, U_K) < \varepsilon\right) = \mathbb{P}\left(\frac{1}{K}\sum_{i=1}^{K} U_i^{-1} > 1/\varepsilon\right).$$

Condition (G) means that for any $1 \leq i < j \leq K$, $(\Phi^{-1}(U_i), \Phi^{-1}(U_j))$ is a bivariate normal random variable with $\mathrm{cov}(\Phi^{-1}(U_i), \Phi^{-1}(U_j)) = \sigma_{ij}$, where Φ is the standard normal distribution function and Φ^{-1} is its inverse. Clearly, $\sigma_{ij} = 1$ implies that $U_i = U_j$ a.s. In this case we can combine them in one and the corresponding coefficient becomes $2/K$. Thus, it suffices to prove the stronger statement

$$\mathbb{P}\left(\sum_{i=1}^{K} w_i U_i^{-1} > 1/\varepsilon\right) \sim \varepsilon, \text{ as } \varepsilon \downarrow 0, \tag{S3.13}$$

where $w_i > 0,\ i = 1, \ldots, K,\ \sum_{i=1}^{K} w_i = 1$ and $\sigma_{ij} < 1,\ i, j = 1, \ldots, K$. We choose some positive constant δ_ε depending on ε, such that $\delta_\varepsilon \to 0$ and $\delta_\varepsilon/\varepsilon \to \infty$ as $\varepsilon \downarrow 0$. Denote by $S = \sum_{i=1}^{K} w_i U_i^{-1}$, and define the following events: for $i \in \{1, \ldots, K\}$,

$$A_{i,\varepsilon} = \left\{U_i^{-1} > \frac{1+\delta_\varepsilon}{w_i \varepsilon}\right\}, \quad B_{i,\varepsilon} = \left\{U_i^{-1} \leq \frac{1+\delta_\varepsilon}{w_i \varepsilon},\ S > 1/\varepsilon\right\}.$$

Let $A_\varepsilon = \bigcup_{i=1}^{K} A_{i,\varepsilon}$ and $B_\varepsilon = \bigcap_{i=1}^{K} B_{i,\varepsilon}$ and thus we have

$$\mathbb{P}\left(S > 1/\varepsilon\right) = \mathbb{P}(A_\varepsilon) + \mathbb{P}(B_\varepsilon).$$

First we show $\mathbb{P}(B_\varepsilon) = o(\varepsilon)$. Note that $S > 1/\varepsilon$ implies that there exists $i \in \{1, \ldots, K\}$ such that $U_i^{-1} > \frac{1}{w_i K \varepsilon}$. Hence,

$$\begin{aligned}
\mathbb{P}\left(B_\varepsilon\right) &\leq \sum_{i=1}^{K} \mathbb{P}\left(\frac{1}{w_i K \varepsilon} < U_i^{-1} \leq \frac{1+\delta_\varepsilon}{w_i \varepsilon}, S > 1/\varepsilon\right) \\
&\leq \sum_{i=1}^{K} \mathbb{P}\left(\frac{1}{w_i K \varepsilon} < U_i^{-1} \leq \frac{1-\delta_\varepsilon}{w_i \varepsilon}, S > 1/\varepsilon\right) + \sum_{i=1}^{K} \mathbb{P}\left(\frac{1-\delta_\varepsilon}{w_i \varepsilon} < U_i^{-1} \leq \frac{1+\delta_\varepsilon}{w_i \varepsilon}\right) \\
&\leq \sum_{i=1}^{K} \mathbb{P}\left(\frac{1}{w_i K \varepsilon} < U_i^{-1} \leq \frac{1-\delta_\varepsilon}{w_i \varepsilon}, S > 1/\varepsilon\right) + \sum_{i=1}^{K} w_i \varepsilon \left(\frac{1}{1-\delta_\varepsilon} - \frac{1}{1+\delta_\varepsilon}\right) \\
&=: I_1 + I_2.
\end{aligned}$$

Noting that $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$, we have $I_2 = o(\varepsilon)$. We next focus on I_1. Observe

$$\begin{aligned} I_1 &\le \sum_{i=1}^K \mathbb{P}\left(\frac{1}{w_i K\varepsilon} < U_i^{-1} \le \frac{1-\delta_\varepsilon}{w_i\varepsilon}, \sum_{j\neq i}^K w_j U_j^{-1} > \delta_\varepsilon/\varepsilon \right) \\ &\le \sum_{i=1}^K \sum_{j\neq i}^K \mathbb{P}\left(\frac{1}{w_i K\varepsilon} < U_i^{-1} \le \frac{1-\delta_\varepsilon}{w_i\varepsilon}, U_j^{-1} > \frac{\delta_\varepsilon}{w_j K\varepsilon} \right). \end{aligned}$$

It remains to show for $1 \le i \neq j \le K$,

$$I_{i,j} := \mathbb{P}\left(\frac{1}{w_i K\varepsilon} < U_i^{-1} \le \frac{1-\delta_\varepsilon}{w_i\varepsilon}, U_j^{-1} > \frac{\delta_\varepsilon}{w_j K\varepsilon} \right) = o(\varepsilon).$$

Condition (G) implies that there exist $Z_{i,j}$ and $\delta_{i,j}$ such that

$$\Phi^{-1}(U_j) = \sigma_{ij}\Phi^{-1}(U_i) + \delta_{ij} Z_{ij}, \tag{S3.14}$$

where Z_{ij} is a standard normal random variable that is independent of U_i and $\sigma_{ij}^2 + \delta_{ij}^2 = 1$. If $\sigma_{ij} = -1$, we have $U_i = 1 - U_j$. This implies that $I_{i,j} = 0$ for $\varepsilon > 0$ sufficiently small. Next, assume $|\sigma_{ij}| < 1$, and write $\gamma_{ij} = \Phi^{-1}(w_i K\varepsilon)$ if $-1 < \sigma_{ij} \le 0$ and $\gamma_{ij} = \Phi^{-1}\left(\frac{w_i\varepsilon}{1-\delta_\varepsilon}\right)$ if $0 < \sigma_{ij} < 1$. We have

$$\begin{aligned} I_{i,j} &= \mathbb{P}\left(\frac{1}{w_i K\varepsilon} < U_i^{-1} \le \frac{1-\delta_\varepsilon}{w_i\varepsilon}, \sigma_{ij}\Phi^{-1}(U_i) + \delta_{ij} Z_{ij} < \Phi^{-1}\left(\frac{w_j K\varepsilon}{\delta_\varepsilon}\right) \right) \\ &\le \mathbb{P}\left(\frac{1}{w_i K\varepsilon} < U_i^{-1} \le \frac{1-\delta_\varepsilon}{w_i\varepsilon}, \delta_{ij} Z_{ij} < \Phi^{-1}\left(\frac{w_j K\varepsilon}{\delta_\varepsilon}\right) - \sigma_{ij}\gamma_{ij} \right) \\ &= \mathbb{P}\left(\frac{1}{w_i K\varepsilon} < U_i^{-1} \le \frac{1-\delta_\varepsilon}{w_i\varepsilon} \right) \mathbb{P}\left(\delta_{ij} Z_{ij} < \Phi^{-1}\left(\frac{w_j K\varepsilon}{\delta_\varepsilon}\right) - \sigma_{ij}\gamma_{ij} \right). \end{aligned}$$

Note that $\Phi^{-1}(\varepsilon) \sim -\sqrt{-2\ln\varepsilon}$, as $\varepsilon \downarrow 0$, which is a slowly varying function. Taking $\delta_\varepsilon = -1/\log\varepsilon$, we have

$$\Phi^{-1}\left(\frac{w_i\varepsilon}{1-\delta_\varepsilon}\right) \sim \Phi^{-1}(w_i K\varepsilon) \sim \Phi^{-1}\left(\frac{w_j K\varepsilon}{\delta_\varepsilon}\right) \quad \text{as } \varepsilon \downarrow 0.$$

This implies

$$\Phi^{-1}\left(\frac{w_j K\varepsilon}{\delta_\varepsilon}\right) - \sigma_{ij}\gamma_{ij} \to -\infty, \text{ as } \varepsilon \downarrow 0.$$

Hence $I_{i,j} = o(\varepsilon)$. Consequently, $I_1 = o(\varepsilon)$ and further $\mathbb{P}(B_\varepsilon) = o(\varepsilon)$. Next, we show $\mathbb{P}(A_\varepsilon) \sim \varepsilon$. By the Bonferroni inequality, we have,

$$\sum_{i=1}^{K} \mathbb{P}(A_{i,\varepsilon}) - \sum_{1\le i<j\le K} \mathbb{P}(A_{i,\varepsilon} \cap A_{j,\varepsilon}) \le \mathbb{P}(A_\varepsilon) \le \sum_{i=1}^{K} \mathbb{P}(A_{i,\varepsilon}).$$

Direct calculation gives

$$\sum_{i=1}^{K} \mathbb{P}(A_{i,\varepsilon}) = \sum_{k=1}^{K} \frac{w_i \varepsilon}{1+\delta_\varepsilon} \sim \varepsilon.$$

For any $1 \le i < j \le K$, since the Gaussian copula is tail independent (e.g., Example 7.38 of McNeil et al. (2015)), we have, writing $w = \max\{w_i, w_j\}$,

$$\begin{aligned}\mathbb{P}(A_{i,\varepsilon} \cap A_{j,\varepsilon}) &= \mathbb{P}\left(U_i^{-1} > \frac{1+\delta_\varepsilon}{w_i\varepsilon}, U_j^{-1} > \frac{1+\delta_\varepsilon}{w_j\varepsilon}\right) \\ &\le \mathbb{P}\left(U_i < \frac{w\varepsilon}{1+\delta_\varepsilon}, U_j < \frac{w\varepsilon}{1+\delta_\varepsilon}\right) = o(1)\mathbb{P}\left(U_1 < \frac{w\varepsilon}{1+\delta_\varepsilon}\right) = o(1)\varepsilon.\end{aligned}$$

Hence $\mathbb{P}(A_{i,\varepsilon} \cap A_{j,\varepsilon}) = o(\varepsilon)$. This implies $\mathbb{P}(A_\varepsilon) \sim \varepsilon$, and we establish (S3.13).

(iii) By Lemma A.1 of Vovk and Wang (2020a), we have

$$a_{\mathcal{H}}(\varepsilon) = \varepsilon\left(\sup\left\{q_0^+\left(\frac{1}{K}\sum_{i=1}^{K} P_i^{-1}\right) \mid P_1, \dots, P_K \in \mathcal{U}\right\}\right)^{-1}, \quad \varepsilon \in (0,1),$$

where $q_0^+(X) = \sup\{x \in \mathbb{R} \mid \mathbb{P}(X \le x) = 0\}$. Note that for any $\delta > 0$, there exists $0 < \varepsilon_\delta < 1$ such that for all $x \in (0, \varepsilon_\delta)$

$$-\frac{(1+\delta)}{x} < \tan\left(-\frac{\pi}{2} + x\right) < -\frac{(1-\delta)}{x}.$$

For $\delta > 0$, letting $0 < \varepsilon < \varepsilon_\delta/\pi$ and using Theorem 4.6 in Bernard et al. (2014), we have

$$\begin{aligned}
&\inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \mathcal{C}^{-1}(P_i)\right) \mid P_1,\dots,P_K \in \mathcal{U}\right\}\\
&= \inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \tan\left(\pi\left(P_i - \frac{1}{2}\right)\right)\right) \mid P_1,\dots,P_K \in \mathcal{U}\right\}\\
&= \inf\left\{q_1\left(\frac{1}{K}\sum_{i=1}^K \tan\left(\pi\left(\varepsilon P_i - \frac{1}{2}\right)\right)\right) \mid P_1,\dots,P_K \in \mathcal{U}\right\}\\
&\le \inf\left\{q_1\left(\frac{1}{K}\sum_{i=1}^K -\frac{1-\delta}{\varepsilon\pi P_i}\right) \mid P_1,\dots,P_K \in \mathcal{U}\right\}\\
&= -\frac{1-\delta}{\varepsilon\pi}\sup\left\{q_0^+\left(\frac{1}{K}\sum_{i=1}^K P_i^{-1}\right) \mid P_1,\dots,P_K \in \mathcal{U}\right\} = -\frac{1-\delta}{a_{\mathcal{H}}(\varepsilon)\pi}.
\end{aligned}$$

Similarly, we obtain, for $0 < \varepsilon < \varepsilon_\delta/\pi$,

$$\inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \mathcal{C}^{-1}(P_i)\right)\right\} \ge -\frac{1+\delta}{a_{\mathcal{H}}(\varepsilon)\pi}.$$

Consequently,

$$\inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \mathcal{C}^{-1}(P_i)\right)\right\} \sim -\frac{1}{a_{\mathcal{H}}(\varepsilon)\pi} \quad \text{as } \varepsilon \downarrow 0.$$

Plugging the above result in the formula for $a_{\mathcal{C}}$ in (3.6), and using $\mathcal{C}(y) \sim -1/(\pi y)$ as $y \to -\infty$, we have, as $\varepsilon \downarrow 0$,

$$\begin{aligned}
a_{\mathcal{C}}(\varepsilon) &= \mathcal{C}\left(\inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \mathcal{C}^{-1}(P_i)\right)\right\}\right)\\
&\sim -\frac{1}{\pi}\left(\inf\left\{q_\varepsilon\left(\frac{1}{K}\sum_{i=1}^K \mathcal{C}^{-1}(P_i)\right)\right\}\right)^{-1} \sim a_{\mathcal{H}}(\varepsilon).
\end{aligned}$$

This completes the proof.

(iv) By (i), it suffices to show that for $r \neq -1$

$$\frac{M_{-1,K}(p_1,\dots,p_K)}{M_{r,K}(p_1,\dots,p_K)} \not\to 1, \text{ as } \max_{i\in\{1,\dots,K\}} p_i \downarrow 0.$$

Take $p_1 = p^2$ and $p_i = x_i p$ with $x_i > 0$ and $p > 0$ for $i = 2,\dots,K$. By homogeneity of M_r, for $r \leq -1$,

$$\frac{M_{-1,K}(p_1,\dots,p_K)}{M_{r,K}(p_1,\dots,p_K)} = \frac{M_{-1,K}(p,x_2,\dots,x_K)}{M_{r,K}(p,x_2,\dots,x_K)}.$$

Hence

$$\lim_{p\downarrow 0} \frac{M_{-1,K}(p_1,\dots,p_K)}{M_{r,K}(p_1,\dots,p_K)} = K^{1/r+1} \neq 1, \quad r < -1.$$

This proves the claim of (iv) for $r < -1$. The case for $r > -1$ can be argued similarly. □

S3.7 Proof of Theorem 3

Take arbitrary $p_1,\dots,p_K \in (0,1]$, and let $j \in \{1,\dots,K\}$ be such that $\min_{k\in\{1,\dots,K\}} p_{(k)}/k = p_{(j)}/j$. Noting that

$$\sum_{i=1}^K \frac{1}{p_i} = \sum_{i=1}^K \frac{1}{p_{(i)}}, \text{ and } \frac{p_{(j)}}{j} \leq \frac{p_{(i)}}{i}, \ i = 1,\dots,K,$$

we have

$$\frac{S_K(p_1,\dots,p_K)}{M_{-1,K}(p_1,\dots,p_K)} = \frac{1}{j}p_{(j)}\left(\sum_{i=1}^K \frac{1}{p_i}\right) = \sum_{i=1}^K \frac{1}{j}p_{(j)}\frac{1}{p_{(i)}} \leq \sum_{i=1}^K \frac{1}{i}p_{(i)}\frac{1}{p_{(i)}} = \sum_{i=1}^K \frac{1}{i} = \ell_K.$$

Moreover,

$$\frac{S_K(p_1,\dots,p_K)}{M_{-1,K}(p_1,\dots,p_K)} = \frac{1}{j}p_{(j)}\left(\sum_{i=1}^K \frac{1}{p_{(i)}}\right) \geq \frac{1}{j}p_{(j)}\left(\sum_{i=1}^j \frac{1}{p_{(j)}} + \sum_{i=j+1}^K \frac{1}{p_{(i)}}\right) \geq 1.$$

Therefore, $M_{-1,K} \le S_K \le \ell_K M_{-1,K}$. The two special cases of equalities are straightforward to check. □

S3.8 Proof of Proposition 5

(i) Recall that $a_F(x) = a_F x$ for $x \in (0,1)$. By (i) of Proposition 3, we have $b_F(\delta) \sim \delta$ as $\delta \downarrow 0$. Hence $\lim_{\delta\downarrow 0} b_F(\delta)/a_F(\delta) = 1/a_F$. By Proposition 6 of Vovk and Wang (2020a), we have $a_F \sim 1/\log K$, as $K \to \infty$. Consequently,

$$\lim_{\delta\downarrow 0} \frac{b_F(\delta)}{a_F(\delta)} \sim \log K, \text{ as } K \to \infty.$$

Moreover, for the harmonic averaging method, $c_F(\varepsilon) = \varepsilon$. This implies $c_F(\varepsilon)/a_F(\varepsilon) = 1/a_F$. We establish the claim by the fact $a_F \sim 1/\log K$, as $K \to \infty$.

(ii) By Theorem 2, we have $a_{\mathcal{C}}(\delta) \sim a_{\mathcal{H}}(\delta)$ and $b_{\mathcal{C}}(\delta) \sim b_{\mathcal{H}}(\delta)$ as $\delta \downarrow 0$, which together with (i) leads to

$$\lim_{\delta\downarrow 0} \frac{b_{\mathcal{C}}(\delta)}{a_{\mathcal{C}}(\delta)} \sim \log K, \text{ as } K \to \infty.$$

The rest of the statement follows by noting that $c_{\mathcal{C}}(\delta) = b_{\mathcal{C}}(\delta)$.

(iii) For the Simes method, recall that $a_F(x) = x/\ell_K$ and $b_F(x) = c_F(x) = x$. The claim follows directly from the fact that $\ell_K = \sum_{k=1}^K \frac{1}{k} \sim \log K$, as $K \to \infty$. □

S4 Additional tables

In Tables 1 and 2 we report numerical results of prices for validity for $\varepsilon = 0.05$

Table 1: $b_F(\varepsilon)/a_F(\varepsilon)$ and $c_F(\varepsilon)/a_F(\varepsilon)$ for $\varepsilon = 0.05$ and $K \in \{50, 100, 200, 400\}$

	$K = 50$		$K = 100$		$K = 200$		$K = 400$	
	b_F/a_F	c_F/a_F	b_F/a_F	c_F/a_F	b_F/a_F	c_F/a_F	b_F/a_F	c_F/a_F
Bonferroni	1.025	50.000	1.026	100.000	1.026	200.000	1.026	400.000
Negative-quartic	1.367	25.071	1.367	42.164	1.368	70.911	1.368	119.257
Simes	4.499	4.499	5.187	5.187	5.878	5.878	6.570	6.570
Cauchy	6.623	6.623	7.463	7.463	8.274	8.274	9.055	9.055
Harmonic	6.793	6.625	7.650	7.459	8.485	8.273	9.306	9.072
Geometric	15.679	2.718	16.874	2.718	17.755	2.718	18.395	2.718

and 0.0001, respectively.

Table 2: $b_F(\varepsilon)/a_F(\varepsilon)$ and $c_F(\varepsilon)/a_F(\varepsilon)$ for $\varepsilon = 0.0001$ and $K \in \{50, 100, 200, 400\}$

	$K = 50$		$K = 100$		$K = 200$		$K = 400$	
	b_F/a_F	c_F/a_F	b_F/a_F	c_F/a_F	b_F/a_F	c_F/a_F	b_F/a_F	c_F/a_F
Bonferroni	1.000	50.000	1.000	100.000	1.000	200.000	1.000	400.000
Negative-quartic	1.333	25.071	1.333	42.164	1.333	70.911	1.333	119.257
Simes	4.499	4.499	5.187	5.187	5.878	5.878	6.570	6.570
Cauchy	6.625	6.625	7.465	7.465	8.274	8.274	9.055	9.055
Harmonic	6.625	6.625	7.459	7.459	8.272	8.272	9.071	9.071
Geometric	5416.222	2.718	6601.414	2.718	7523.231	2.718	8214.151	2.718

Bibliography

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, *54*, 93–108.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (2013). *Modelling Extremal Events for Insurance and Finance*. Springer Science & Business Media.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Walter de Gruyter, Berlin, Fourth Edition.

Goeman, J. J., Meijer, R. J., Krebs, T. J. and Solari, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika*, *106*(4), 841–856.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Raffeld, M., et al. (2001). Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine*, *344*(8), 539–548.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*, 65–70.

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*, 393–402.

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

Puccetti, G., Rigo, P., Wang, B. and Wang, R. (2019). Centers of probability measures without the mean. *Journal of Theoretical Probability*, *32*(3), 1482–1501.

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.

Samorodnitsky, G. (2017). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Routledge.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.

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. and Wang, R. (2020a). 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.

Wang, X. (2005). Volumes of generalized unit balls. *Mathematics Magazine*, *78*(5), 390–395.