Supplement to "Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures"

Yaowu Liu
Department of Biostatistics, Harvard School of Public Health
and
Jun Xie*
Department of Statistics, Purdue University
junxie@purdue.edu

In Section 1, we provide the proofs of technical lemmas, Corollary 1–2, and Theorem 1–3. In Section 2, we present additional simulation results about the accuracy of Cauchy approximation under multivariate t distributions and power comparison of different combination tests. In Section 3, we provide a toy example and some discussions to illustrate the finite-sample power of the Cauchy combination test.

1 Proof of main results

We introduce some notations. Let $\phi(x)$ and $\Phi(x)$ be the density function and the cumulative distribution function of the standard normal variable, respectively. Let $h(x) = \tan\{[2\Phi(|x|) - 3/2]\pi\}$ and $p(x) = 2\{1 - \Phi(|x|)\}$. The Cauchy combination test statistic is defined as $T(\mathbf{X}) = \sum_{i=1}^{d} \omega_i h(X_i)$, where $w_i \geq 0$ for any $1 \leq i \leq d$ and $\sum_{i=1}^{d} w_i = 1$.

1.1 Technical Lemmas

We first prove a few useful technical Lemmas. Note that d is fixed in Theorem 1. The notation $o(1)$ and $O(1)$ in the following proofs is with respect to t tending to $+\infty$.

*This work is supported by the National Institutes of Health Grant R21GM101504.

Lemma 1 (Bonferroni inequality)**.** *Let* $A = \cup_{i=1}^{d} A_i$. *For any* $k < [d/2]$, *we have*

$$\sum_{s=1}^{2k} (-1)^{s-1} B_s \le P(A) \le \sum_{s=1}^{2k-1} (-1)^{s-1} B_s$$

where $B_s = \sum_{1 \le i_1 < \cdots < i_s \le d} P(A_{i_1} \cap \cdots \cap A_{i_s})$.

Lemma 2 (Mill's ratio inequality)**.** *For any* $x > 0$,

$$\frac{x}{\phi(x)} \le \frac{1}{1 - \Phi(x)} \le \frac{x}{\phi(x)} \cdot \frac{1 + x^2}{x^2}.$$

Lemma 3 (Properties of function h)**.** *(i) For any* $|x| > \Phi^{-1}(3/4)$,

$$\frac{\cos[p(x)\pi]}{p(x)\pi} \le h(x) \le \frac{1}{p(x)\pi}. \tag{1}$$

(ii) For any constant $0 < |a| < 1$, *we have*

$$\lim_{x \to +\infty} \frac{h(x)}{x^2 h(ax)} > c_a > 0, \tag{2}$$

where c_a *is some constant only depend on* a.

(iii) Suppose W_0 *has a standard Cauchy distribution and* X_0 *has a standard normal distribution, then we have*

$$P\{W_0 > t\} = P\{h(X_0) > t\} = \frac{1}{t\pi} + O(1/t^3). \tag{3}$$

Proof. (i) Note that $p(x) < 1/2$ when $|x| > \Phi^{-1}(3/4)$ and $h(x) = \sin\{[1/2 - p(x)]\pi\} / \cos\{[1/2 - p(x)]\pi\} = \cos[p(x)\pi] / \sin[p(x)\pi]$, then (1) follows from the elementary inequalities that $z \cos z \le \sin z \le z$ for any $z \in (0, \pi/2)$.

(ii) Since $h(x) = h(-x)$, we only need to consider the case where $0 < a < 1$. To simplify the exposition, we write $f(x) \asymp g(x)$ if $\lim_{x \to +\infty} f(x)/g(x) = c_0$, where constant $c_0 > 0$ and $f(x)$ and $g(x)$ are two functions. Simple calculation gives that for $x > 0$,

$$h'(x) = \frac{2\pi\phi(x)}{\cos^2\{[1/2 - p(x)]\pi\}} > 0.$$

Through Mill's inequality and (1), we have $h(x) \asymp x e^{x^2/2}$ and $h'(x) \asymp x^2 e^{x^2/2}$. By the mean value theorem, $h(x) = h(ax) + h'(a_x x)(1 - a)x$ for some constant $a \le a_x \le 1$. Then, for $x > \Phi^{-1}(3/4)/a$, we have

$$\frac{h(x)}{x^2 h(ax)} \ge \frac{h'(a_x x)(1 - a)x}{x^2 h(ax)} \asymp \frac{(1 - a) a_x^2 x^2 e^{a_x^2 x^2/2}}{a x^2 e^{a^2 x^2/2}} \ge (1 - a)a.$$

(iii) Let $U_0 = p(X_0) \sim U[0,1]$. It follows from (1) that

$$P\left\{\frac{\cos[U_0\pi]}{U_0\pi} > t\right\} \le P\{h(X_0) > t\} \le P\left\{\frac{1}{U_0\pi} > t\right\} = \frac{1}{t\pi}.$$

By Taylor's series expansion, we have

$$P\left\{\frac{\cos[U_0\pi]}{U_0\pi} > t\right\} = P\left\{\frac{1-(\pi^2/2)U_0^2 + o_p(U_0^2)}{U_0\pi} > t\right\}$$

and

$$\begin{aligned} & P\left\{\frac{\pi}{2t}U_0^2 + U_0 - \frac{1}{t\pi} > 0\right\} = P\left\{U_0 < \frac{1}{\pi}(\sqrt{t^2+2} - t)\right\} \\ = & \frac{1}{\pi}(\sqrt{t^2+2} - t) = \frac{1}{t\pi} - \frac{1}{2\pi t^3} + o(\frac{1}{t^3}) = \frac{1}{t\pi} + O(\frac{1}{t^3}). \end{aligned}$$

Hence, we complete the proof.

□

1.2 Proof of Theorem 1

Since $h(x) = h(-x)$, then $h(X_i) = h(X_j)$ if $X_i = X_j$ or $X_i = -X_j$. Hence, we only need to consider the case where $|\sigma_{ij}| < 1$ for any $1 \le i < j \le d$. Without loss of generality, we assume $w_i > 0$ for any $1 \le i \le d$.

Let $A_{i,t} = \{h(X_i) > (1+\delta_t)t/\omega_i, T(\mathbf{X}) > t\}$ and $B_{i,t} = \{h(X_i) \le (1+\delta_t)t/\omega_i, T(\mathbf{X}) > t\}$, where constant δ_t only depends on t and satisfies that $\delta_t > 0$, $\delta_t \to 0$ and $\delta_t t \to +\infty$ as $t \to +\infty$. Further let $A_t = \bigcup_{i=1}^d A_{i,t}$, $B_t = \bigcap_{i=1}^d B_{i,t}$, then we can decompose the rejection region $\{T(\mathbf{X}) > t\}$ into the two disjoint sets A_t and B_t, and then

$$P\{T(\mathbf{X}) > t\} = P(A_t) + P(B_t).$$

Step 1. We first show that $P(B_t) = o(1/t)$. The event $\{T(\mathbf{X}) > t\}$ implies that there exists at least one i such that $h(X_i) > t/(\omega_i d)$. Then we have

$$\begin{aligned} P(B_t) &\le \sum_{i=1}^d P(B_{i,t} \cap \{h(X_i) > t/(\omega_i d)\}) = \sum_{i=1}^d P\left\{\frac{t}{\omega_i d} < h(X_i) \le \frac{(1+\delta_t)t}{\omega_i}, T(\mathbf{X}) > t\right\} \\ &\le \sum_{i=1}^d P\left\{\frac{t}{\omega_i d} < h(X_i) \le \frac{(1-\delta_t)t}{\omega_i}, T(\mathbf{X}) > t\right\} + \sum_{i=1}^d P\left\{\frac{(1-\delta_t)t}{\omega_i} < h(X_i) \le \frac{(1+\delta_t)t}{\omega_i}\right\} \\ &= I_1 + I_2 \end{aligned}$$

It is easy to see the $I_2 = o(1/t)$ by noting that $\delta_t \to 0$ and using Lemma 3(iii). Regarding I_1, we have

$$\begin{aligned} I_1 &\leq \sum_{i=1}^{d} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{(1-\delta_t)t}{\omega_i}, \sum_{j\neq i}^{d} \omega_j h(X_j) > \delta_t t\right\} \\ &\leq \sum_{i=1}^{d}\sum_{j\neq i}^{d} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{(1-\delta_t)t}{\omega_i}, h(X_j) > \frac{\delta_t t}{(d-1)\omega_j}\right\}. \end{aligned}$$

It remains to show for any $1 \leq i \neq j \leq d$,

$$I_{1,ij} = P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{(1-\delta_t)t}{\omega_i}, h(X_j) > \frac{\delta_t t}{(d-1)\omega_j}\right\} = o(1/t). \tag{4}$$

Because of the pairwise normality assumption of $\mathbf{X}$, we can write $X_j = \sigma_{ij}X_i + \gamma_{ij}Z_{ij}$, where $\sigma_{ij}^2 + \gamma_{ij}^2 = 1$ and Z_{ij} is independent of X_i and follows a standard normal distribution. If $\sigma_{ij} = 0$, then (4) directly follows from Lemma 3(iii). Note that we also have $|\sigma_{ij}| < 1$. Hence, we only need to consider the case where $0 < |\sigma_{ij}| < 1$. Let $h^{-1}(\cdot)$ be the inverse function of $h(x)$ when $x > 0$. The event in (4) implies that $|X_i| > h^{-1}(\frac{t}{\omega_i d}) \to +\infty$. Therefore, applying Lemma 3(ii), for sufficiently large t, we have

$$h(\sigma_{ij}X_i) \leq \frac{h(X_i)}{c_0 X_i^2} \leq \frac{(1-\delta_t)t}{c_0\omega_i[h^{-1}(\frac{t}{\omega_i d})]^2} < \frac{t}{c_0\omega_i[h^{-1}(\frac{t}{\omega_i d})]^2},$$

where $c_0 > 0$ is a constant only depending on σ_{ij}. Then by choosing δ_t such that $\delta_t[h^{-1}(\frac{t}{\omega_i d})]^2 \to +\infty$, we have $h(\sigma_{ij}X_i) \leq o(\frac{\delta_t t}{(d-1)\omega_j})$. Recall that $h(X_j) > \frac{\delta_t t}{(d-1)\omega_j}$ in (4) and $X_j = \sigma_{ij}X_i + \gamma_{ij}Z_{ij}$. These indicate that $|Z_{ij}| > \lambda_t$, where λ_t is some constant depending on t and tends to $+\infty$. Hence, by Lemma 3(iii),

$$I_{1,ij} \leq P\left\{\frac{t}{\omega_i d} < h(X_i), |Z_{ij}| > \lambda_t\right\} = P\left\{\frac{t}{\omega_i d} < h(X_i)\right\} P\left\{|Z_{ij}| > \lambda_t\right\} = o(1/t)$$

Step 2. In this step, we show that $P(A_t) = 1/(t\pi) + o(1/t)$. By Bonferroni inequality,

$$\sum_{i=1}^{d} P(A_{i,t}) - \sum_{1\leq i<j\leq d} P(A_{i,t}\cap A_{j,t}) \leq P(A_t) \leq \sum_{i=1}^{d} P(A_{i,t}).$$

Using a similar argument of proving $I_1 = o(1/t)$ in step 1, it is easy to see that $P(A_{i,t} \cap A_{j,t}) = o(1/t)$ for any $1 \leq i < j \leq d$. Further, observe that

$$P(A_{i,t}) = P\{h(X_i) > (1+\delta_t)t/\omega_i\} - P\{h(X_i) > (1+\delta_t)t/\omega_i, T(\mathbf{X}) \leq t\}$$

and

$$P\{h(X_i) > (1+\delta_t)t/\omega_i\} = \frac{\omega_i}{\pi(1+\delta_t)t} + o\left(\frac{1}{(1+\delta_t)t}\right) = \frac{\omega_i}{\pi t} + o(1/t),$$

it suffices to show

$$P\{h(X_i) > (1+\delta_t)t/\omega_i, T(\mathbf{X}) \le t\} = o(1/t) \tag{5}$$

for any $1 \le i \le d$. The event in (5) implies that there exists at least one $j \ne i$ such that $\omega_j h(X_j) \le -\delta_t t(d-1)$, then it can be easily seen that (5) also follows from a similar argument of proving $I_1 = o(1/t)$ in step 1. Therefore, we complete the proof.

1.3 Proof of Corollary 1

From Theorem 1, we have

$$\left| P\{T(\mathbf{X}) > t\} - \frac{1}{t\pi} \right| \le o(1/t),$$

for any given fixed weights $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_d)$ and $\sum_{i=1}^d \omega_i = 1$. Since $\boldsymbol{\omega}$ is independent of $\mathbf{X}$, it suffices to show that the upper bound $o(1/t)$ does not depend on the weights.

In the proof of Theorem 1, we provide the upper bound for multiple terms. First, we consider the term $I_{1,ij}$ in (4) and show that the upper bound for it does not depend on the weights. It is obvious that

$$I_{1,ij} \le P\left\{ \frac{t}{\omega_i d} < h(X_i) \le \frac{t}{\omega_i}, h(X_j) > \delta_t t \right\}.$$

Let $\varepsilon > 0$ be a sufficiently small constant. If $0 < \omega_i \le 1/t^{\varepsilon}$, then

$$I_{1,ij} \le P\left\{ \frac{t^{1+\varepsilon}}{d} < h(X_i) \right\} = \frac{d}{t^{1+\varepsilon}\pi} + o(1/t^{1+\varepsilon}).$$

If $1/t^{\varepsilon} < \omega_i < 1$, we have

$$\begin{aligned} I_{1,ij} &\le P\left\{t/d < h(X_i) \le t^{1+\varepsilon}, h(\sigma_{ij}X_i + \gamma_{ij}Z_{ij}) > \delta_t t\right\} \\ &\le P\left[t/d < h(X_i), |Z_{ij}| > \frac{1}{|\gamma_{ij}|}\left\{h^{-1}(\delta_t t) - |\sigma_{ij}| h^{-1}(t^{1+\varepsilon})\right\} \right]. \end{aligned}$$

As in the proof of Theorem 1, by choosing δ_t such that $\delta_t[h^{-1}(t^{1+\varepsilon})]^2 \to +\infty$, we have $h^{-1}(\delta_t t) - |\sigma_{ij}| h^{-1}(t^{1+\varepsilon}) \to +\infty$. Combing the two cases about ω_i together, we obtain an upper bound for $I_{1,ij}$, which does not depend on the weights and is $o(1/t)$.

For the other terms, it is easy to see or can be shown using a similar argument above for $I_{1,ij}$ that their upper bounds does not depend on the weights and is $o(1/t)$. Hence, we omit the details of the proof.

1.4 Proof of Theorem 2

To simplify the exposition, we introduce a new notation that $a \preceq b$ if a is smaller than or equal to b up to multiplying some positive constant independent of t and d. Similar to the proof of Theorem 1, we first decompose the rejection region $\{T(\mathbf{X}) > t\}$ into the two disjoint sets A_t and B_t.

Step 1. We show that $P(B_t) = o(1/t)$. The event $\{T(\mathbf{X}) > t\}$ implies that there exists at least one i such that $h(X_i) > t/(\omega_i d)$. Then we have

$$\begin{aligned}
P(B_t) &\leq \sum_{i=1}^{d} P(B_{i,t} \cap \{h(X_i) > t/(\omega_i d)\}) = \sum_{i=1}^{d} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{(1+\delta_t)t}{\omega_i}, T(\mathbf{X}) > t\right\} \\
&\leq \sum_{i=1}^{d} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{(1-\delta_t)t}{\omega_i}, T(\mathbf{X}) > t\right\} + \sum_{i=1}^{d} P\left\{\frac{(1-\delta_t)t}{\omega_i} < h(X_i) \leq \frac{(1+\delta_t)t}{\omega_i}\right\} \\
&\leq \sum_{i=1}^{d} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{(1-\delta_t)t}{\omega_i}, \sum_{j\neq i}^{d} \omega_j h(X_j) > \delta_t t\right\} \\
&\quad + \sum_{i=1}^{d} P\left\{\frac{(1-\delta_t)t}{\omega_i} < h(X_i) \leq \frac{(1+\delta_t)t}{\omega_i}\right\} \\
&= I_1 + I_2
\end{aligned}$$

It is easy to see the $I_2 = o(1/t)$ by noting that $\delta_t \to 0$ and using Lemma 3(iii). It remains to show $I_1 = o(1/t)$. Because the largest eigenvalue of $\mathbf{\Sigma} = (\sigma_{ij})$ is bounded by C_0, we have $\max_{1\leq i\leq d} \sum_{j=1}^{d} \sigma_{ij}^2 \leq C_0$. Let $0 < \sigma_0^2 < 1$ be a constant and $\mathcal{J}_i = \{j \neq i : \sigma_{ij}^2 \geq \sigma_0^2\}$. For any $1 \leq i \leq d$, the cardinality of $\mathcal{J}_i$ is less than or equal to C_0/σ_0^2. Let $\delta_t = (C_0/\sigma_0^2)t^{-\varepsilon}$, where constant $0 < \varepsilon < 1$. Then we have

$$\begin{aligned}
I_1 &\leq \sum_{i=1}^{d} \sum_{j\in\mathcal{J}_i} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{t}{\omega_i}, h(X_j) > \frac{t^{1-\varepsilon}}{\omega_j}\right\} \\
&\quad + \sum_{i=1}^{d} \sum_{j\in\mathcal{J}_i^c, j\neq i} P\left\{\frac{t}{\omega_i d} < h(X_i) \leq \frac{t}{\omega_i}, h(X_j) > \frac{t^{1-\varepsilon}}{d\omega_j}\right\} \\
&= I_{11} + I_{12}
\end{aligned}$$

Recall the definition of the notation "$\preceq$" introduced at the beginning of the proof. Let $d = t^\alpha$ for some $0 < \alpha < 1/2$. Note that $\min_{1\leq i\leq d} \omega_i \geq c_0/d$ and the cardinality of $\mathcal{J}_i$ is bounded by a constant C_0/σ_0^2 for any $1 \leq i \leq d$. Then we have

$$\begin{aligned}
I_{11} &= \sum_{i=1}^{d} \sum_{j\in\mathcal{J}_i} P\left\{t \preceq h(X_i) \preceq t^{1+\alpha}, t^{1+\alpha-\varepsilon} \preceq h(X_j)\right\} \\
&\preceq d \max_{1\leq i\leq d, j\in\mathcal{J}_i} P\left\{t \preceq h(X_i) \preceq t^{1+\alpha}, t^{1+\alpha-\varepsilon} \preceq h(X_j)\right\}
\end{aligned}$$

and

$$
\begin{aligned}
I_{12} &= \sum_{i=1}^{d} \sum_{j \in \mathcal{J}_i^c, j \neq i} P\left\{t \preceq h(X_i) \preceq t^{1+\alpha}, t^{1-\varepsilon} \preceq h(X_j)\right\} \\
&\preceq d^2 \max_{1 \leq i \leq d, j \in \mathcal{J}_i^c, j \neq i} P\left\{t \preceq h(X_i) \preceq t^{1+\alpha}, t^{1-\varepsilon} \preceq h(X_j)\right\}.
\end{aligned}
$$

It follows from Mill's inequality and (1) that $e^{x^2/2} \preceq h(x) \preceq |x| e^{x^2/2}$ for $|x| > \Phi^{-1}(3/4)$. Hence,

$$
\sqrt{2 \log t} \preceq h^{-1}(t) \preceq (1 + o(1)) \sqrt{2 \log t}, \tag{6}
$$

where $o(1)$ is positive.

Because of the pairwise normality assumption of $\mathbf{X}$, we can write $X_j = \sigma_{ij} X_i + \gamma_{ij} Z_{ij}$, where $\sigma_{ij}^2 + \gamma_{ij}^2 = 1$, $\sigma_{ij}\gamma_{ij} > 0$, and Z_{ij} is independent of X_i and follows a standard normal distribution. Combing this result with (6), we have

$$
\begin{aligned}
& P\left\{t \preceq h(X_i) \preceq t^{1+\alpha}, t^{b} \preceq h(X_j)\right\} \\
\leq\ & P\left\{t \preceq h(X_i), |X_i| \preceq h^{-1}(t^{1+\alpha}), h^{-1}(t^{b}) \preceq |X_j|\right\} \\
\leq\ & P\left\{t \preceq h(X_i), |X_i| \preceq (1 + o(1)) \sqrt{2(1+\alpha) \log t}, \sqrt{2b \log t} \preceq |\sigma_{ij} X_i + \gamma_{ij} Z_{ij}|\right\} \\
\leq\ & P\left\{t \preceq h(X_i), \frac{\sqrt{b} - \sqrt{(1 + o(1))(1+\alpha)\sigma_{ij}^2}}{|\gamma_{ij}|} \cdot \sqrt{2 \log t} \preceq |Z_{ij}|\right\} \\
=\ & P\left\{t \preceq h(X_i)\right\} P\left\{\frac{\sqrt{b} - \sqrt{(1 + o(1))(1+\alpha)\sigma_{ij}^2}}{\sqrt{1 - \sigma_{ij}^2}} \cdot \sqrt{2 \log t} \preceq |Z_{ij}|\right\},
\end{aligned}
$$

where $b > 0$ is some positive constant. Specifically, $b = 1 + \alpha$ for I_{11} and $b = 1 - \varepsilon$ for I_{12}.

If $b < (1 + o(1))(1 + \alpha)$, then the function $\left\{\sqrt{b} - \sqrt{(1 + o(1))(1+\alpha)\sigma_{ij}^2}\right\} / \sqrt{1 - \sigma_{ij}^2}$ is monotonically decreasing with respect to σ_{ij}^2 for $0 < \sigma_{ij}^2 < 1$. For both I_{11} and I_{12}, the values of b, i.e., $b = 1 + \alpha$ and $b = 1 - \varepsilon$, is less than $(1 + o(1))(1 + \alpha)$ by noting that $o(1)$ in (6) is positive. Therefore, the maximum in both I_{11} and I_{12} can be bounded by using the maximum of σ_{ij}^2.

In I_{11}, $\sigma_{ij}^2 \leq \sigma_{max}^2 < 1$ for any $1 \leq i \leq d, j \in \mathcal{J}_i$. In I_{12}, $\sigma_{ij}^2 \leq \sigma_0^2$ for any $1 \leq i \leq d, j \in \mathcal{J}_i^c, j \neq i$. Note that $P\{t \preceq h(X_i)\} \preceq 1/t$ by Lemma 3(iii). Hence, we have

$$
I_{11} \preceq t^{-1} \cdot t^{\alpha} \cdot P\left\{\frac{\sqrt{1+\alpha} - \sqrt{(1 + o(1))(1+\alpha)\sigma_{max}^2}}{\sqrt{1 - \sigma_{max}^2}} \cdot \sqrt{2 \log t} \preceq |Z_0|\right\}
$$

and

$$
I_{12} \preceq t^{-1} \cdot t^{2\alpha} \cdot P\left\{\frac{\sqrt{1-\varepsilon} - \sqrt{(1 + o(1))(1+\alpha)\sigma_0^2}}{\sqrt{1 - \sigma_0^2}} \cdot \sqrt{2 \log t} \preceq |Z_0|\right\},
$$

where Z_0 is a standard normal variable.

Through Mill's inequality, to prove that $I_{11} = o(1/t)$ and $I_{12} = o(1/t)$, it suffices to show that for some $0 < \alpha < 1/2$,

$$\frac{\sqrt{1+\alpha} - \sqrt{(1+o(1))(1+\alpha)\sigma_{max}^2}}{\sqrt{1-\sigma_{max}^2}} > \alpha \tag{7}$$

and

$$\frac{\sqrt{1-\varepsilon} - \sqrt{(1+o(1))(1+\alpha)\sigma_0^2}}{\sqrt{1-\sigma_0^2}} > 2\alpha. \tag{8}$$

It is obvious that the inequality (7) holds. By letting ε and σ_0^2 sufficient small, it can be easily seen that the inequality (8) also holds for any constant $0 < \alpha < 1/2$.

Step 2. We show that $P(A_t) = 1/(t\pi) + o(1/t)$. This can be done by using a similar argument in step 2 in the proof of Theorem 1 and a similar argument in step 1 in the proof of Theorem 2. Therefore, we omit the proof.

1.5 Proof of Corollary 2

The proof strategy is analogous to that of Theorem 1 or 2. Thus, we provide an outline here and omit the detail of the proof.

Let $\mathbf{X} = (X_1, \cdots, X_d)$ denote a vector of standard normal variables that has the same correlation matrix as $\tilde{\mathbf{X}}$ and satisfies the bivariate normality condition, where $\mathrm{var}(X_i) = 1$ for any $1 \le i \le d$. Note that in the proof of Theorem 1 and 2, we essentially did the following decomposition:

$$P\{T(\mathbf{X}) > t\} = \sum_{i=1}^{d} P\{h(X_i) > t/\omega_i\} + I,$$

where I contains multiple terms and we showed that $I = o(1/t)$.

Given the bivariate normality condition and the assumption that $\mathrm{var}(\tilde{X}_i) \le 1$, we have

$$P\{h(\tilde{X}_i) > t\} \le P\{h(X_i) > t\} \tag{9}$$

and

$$P\{h(\tilde{X}_i) > t, h(\tilde{X}_j) > s\} \le P\{h(X_i) > t, h(X_j) > s\}, \tag{10}$$

for any $1 \le i < j \le d$.

Then it follows from (9) that

$$\sum_{i=1}^{d} P\{h(X_i) > t/\omega_i\} \le \sum_{i=1}^{d} P\{h(\tilde{X}_i) > t/\omega_i\}.$$

The terms in I are all about the tail probabilities and we derived the upper bound for them. Through (9) and (10), it can be easily shown that $I = o(1/t)$ for $\tilde{X}_i$'s. Hence, we obtain

$$P\{T(\tilde{\mathbf{X}}) > t\} \leq P\{T(\mathbf{X}) > t\} + o(1/t).$$

1.6 Proof of Theorem 3

We first recall some notations and conditions for Theorem 3. Let $h(x) = \tan\{[2\Phi(|x|) - 3/2]\pi\}$ and $p(x) = 2\{1 - \Phi(|x|)\}$. The Cauchy combination test statistic is defined as $T(\mathbf{X}) = \sum_{i=1}^{d} \omega_i h(X_i)$, where $\min_{1\leq i\leq d} \omega_i \geq c_0/d$ for some constant $c_0 > 0$. We assume that the individual test statistics $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a banded correlation matrix, i.e., $\sigma_{ij} = 0$ for any $|i-j| > d_0$ for some positive constant $d_0 > 1$. Let $S = \{1 \leq i \leq d : \mu_i \neq 0\}$ denote the set of signals. Suppose that the number of signals $|S| = d^\gamma$, where $0 < \gamma < 1/2$ and $|S|$ is the cardinality of S. The non-zero μ_i's are assumed to have the same magnitude, i.e., $|\mu_i| = \mu_0 = \sqrt{2r \log d}$ for all $i \in S$, where $\sqrt{r} > 1 - \sqrt{\gamma}$.

Now we are ready to prove Theorem 3. Let $\mathbf{X} = \boldsymbol{\mu} + \mathbf{Z}$, where $\mathbf{Z} \sim N(0, \boldsymbol{\Sigma})$. We can decompose $T(\mathbf{X})$ into two parts:

$$T(\mathbf{X}) = \sum_{i\in S} \omega_i h(X_i) + \sum_{i\in S^c} \omega_i h(X_i).$$

Since $\boldsymbol{\Sigma}$ is a banded correlation matrix, it is easy to see that the second part in the decomposition is $O_p(1)$. Hence, to prove that $\lim_{d\to+\infty} P\{R_\alpha(\mathbf{X}) = 1\} = \lim_{d\to+\infty} P\{T(\mathbf{X}) > t_\alpha\} = 1$ for any $\alpha > 0$, it suffices to show that the first part in the decomposition converges to ∞ with probability 1.

It is obvious that

$$\sum_{i\in S} \omega_i h(X_i) \geq \frac{c_0}{d}\, h\left(\max_{i\in S} |X_i|\right) + \frac{d^\gamma - c_0}{d}\, h\left(\min_{i\in S} |X_i|\right).$$

Let $S_+ = \{i \in S, \mu_i > 0\}$ and assume that $|S_+| \geq |S|/2$ without loss of generality. Given the assumption that $\boldsymbol{\Sigma}$ is a banded correlation matrix, it follows from Lemma 6 of Cai et al. (2014) that $\max_{i\in S_+} Z_i \geq \sqrt{2\log|S_+|} + o_p(1)$. Then we have

$$\max_{i\in S} |X_i| \geq \max_{i\in S_+} |X_i| = \max_{i\in S_+} |Z_i + \mu_0| \geq \mu_0 + \max_{i\in S_+} Z_i = \mu_0 + \sqrt{2\log|S_+|} + o_p(1) \quad (11)$$

Hence,

$$\begin{aligned} h\left(\max_{i\in S} |X_i|\right) &\geq (2\pi)^{-1/2} \max_{i\in S} |X_i| \exp\left\{(\max_{i\in S} |X_i|)^2/2\right\} + o_p(1) \\ &\geq (2\pi)^{-1/2}(\sqrt{2\log|S_+|} + \mu_0) \exp\{\log|S_+| + \mu_0^2/2 + \mu_0\sqrt{2\log|S_+|}\} + o_p(1) \\ &\geq \exp\{\gamma \log d + \mu_0^2/2 + \mu_0\sqrt{2\gamma\log d}\} + o_p(1) = d^{(\sqrt{\gamma}+\sqrt{r})^2} + o_p(1), \end{aligned}$$

where the first inequality follows from the left-hand side of (1) and Mill's inequality and the second inequality follows from (11). Given the condition that $\sqrt{\gamma}+\sqrt{r}>1$, we obtain $(c_0/d)h(\max_{i\in S}|X_i|)\to+\infty$. Hence, it suffices to show that

$$d^{\gamma-1}h\left(\min_{i\in S}|X_i|\right)\geq o_p(1). \tag{12}$$

Let ε_d be constant such that $\varepsilon_d>0$ and $\varepsilon_d\to 0$ as $d\to+\infty$. First of all, we have

$$\begin{aligned} & P\{\min_{i\in S}|X_i|<\varepsilon_d\}\leq\sum_{i\in S}P\{|X_i|<\varepsilon_d\}=d^{\gamma}P\{|X_1|<\varepsilon_d\} \\ = \quad & d^{\gamma}\{\Phi(\mu_0+\varepsilon_d)-\Phi(\mu_0-\varepsilon_d)\}\leq 2\phi(\mu_0-\varepsilon_d)d^{\gamma}\varepsilon_d\leq d^{\gamma}\varepsilon_d. \end{aligned}$$

Following a similar argument in the proof of Lemma 3(i), it is easy to show that $h(x)\geq -1/[\{1-p(x)\}\pi]$ when $|x|<\Phi^{-1}(3/4)$. Therefore,

$$h(\varepsilon_d)\geq\frac{-1}{\{1-p(\varepsilon_d)\}\pi}=\frac{-1}{\{2\Phi(\varepsilon_d)-1\}\pi}=-\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\varepsilon_d}+o(1/\varepsilon_d),$$

where the last equality follows from $\Phi(\varepsilon_d)=1/2+\varepsilon_d/\sqrt{2\pi}+o(\varepsilon_d)$. Note that $0<\gamma<1/2$. By letting $\varepsilon_d=d^{\gamma_0-1}$ where $\gamma<\gamma_0<1/2$, we have

$$P\{\min_{i\in S}|X_i|<\varepsilon_d\}\leq d^{\gamma}\varepsilon_d=d^{\gamma+\gamma_0-1}=o(1)\quad\text{and}\quad d^{\gamma-1}h(\varepsilon_d)\geq-\sqrt{\frac{2}{\pi}}\cdot d^{\gamma-\gamma_0}+o(d^{\gamma-\gamma_0})=o(1).$$

Note that $h(x)$ is increasing when $x>0$, thus we prove (12).

2 Supplementary Figures

The simulation setting of Figure 1 is the same as that of Figure 1 in the main text, except that the individual test statistics $\mathbf{X}$ is generated from a multivariate t distribution with 4 degrees of freedom. The result demonstrates that the p-value calculation is also very accurate even if the normality assumption is violated.

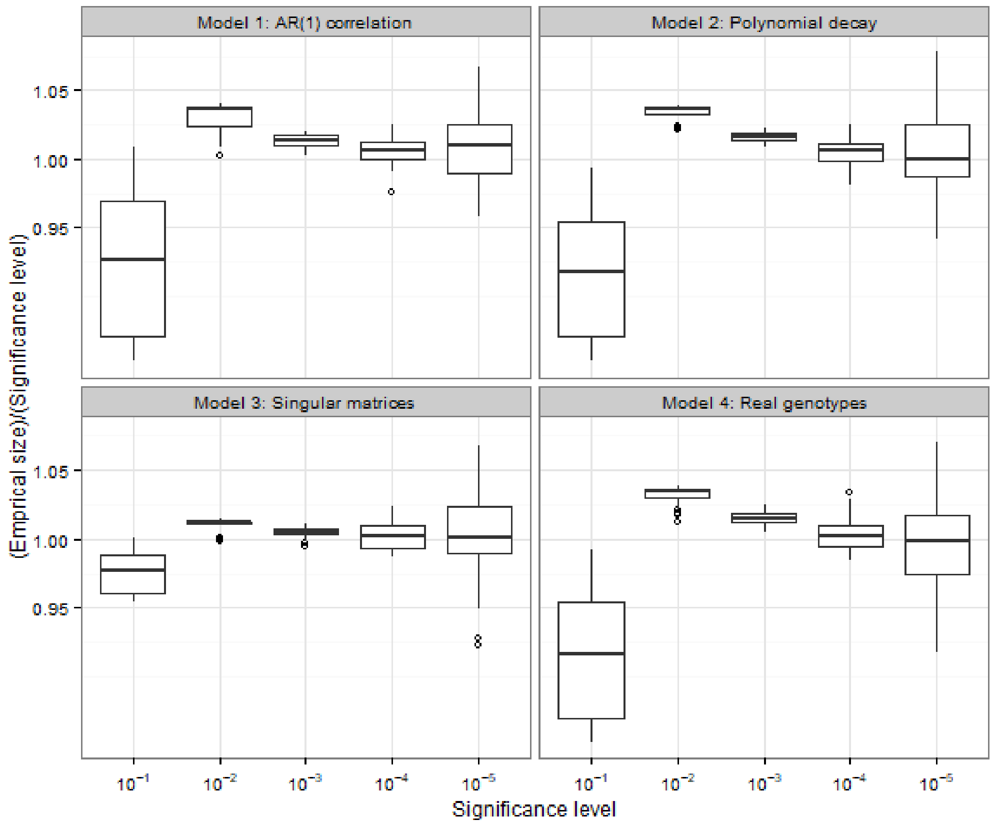

Figure 1: The ratio of empirical size to significance level under a variety of hypothetical and real-data-based correlation matrices. The simulation setting is the same as that of Figure 1 in the main text, except that $\mathbf{X} \sim t_4(0, \mathbf{\Sigma})$. The x-axis is the significance level at $\alpha = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$.

The simulation setting of Figure 2 is the same as that of Figure 3 in the main text, except that the critical values of CCT are calculated analytically through the Cauchy approximation. Figure 2 also demonstrates that CCT has more robust power across different correlation and sparsity levels, compared with the other three tests.

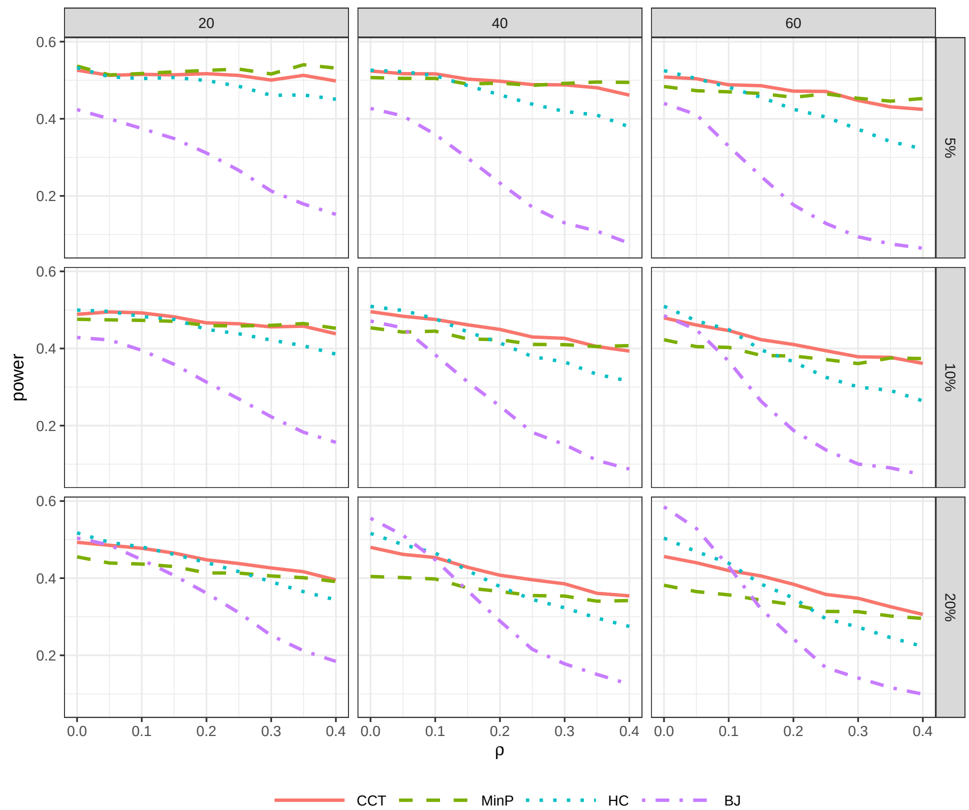

Figure 2: Power comparison of *CCT*, *MinP*, *HC* and *BJ*. The critical values of *CCT* are calculated analytically. The x-axis is the correlation strength ρ. The columns from left to right correspond to the dimension $d = 20, 40, 60$. The rows from top to bottom correspond to the signal percentage 5%, 10% and 20%.

3 Finite-sample power

As discussed in the main text, the Cauchy combination test essentially only uses a few smallest p-values to represent the overall significance. We illustrate this by a toy example provided in Table 1. In this example, there are seven p-values, where two of them are substantially smaller than the others. The Cauchy values for the smallest two p-values are much larger than those for the other p-values and dominate the summation. Although our combination test also relies on a sum of distributions, it essentially only uses a few smallest p-values to represent the overall significance and therefore would be very powerful in the presence of sparse signals.

Table 1: A toy example

P-values	0.45	0.35	0.25	0.15	0.05	5E-03	2E-03
Cauchy values	0.16	0.51	1.00	1.96	6.31	63.7	159

If there is one p-value very close to 1, the Cauchy combination test would tend to not reject the null hypothesis even if there are multiple other p-values that are moderately small. In comparison, the minimum p-value method (MinP) would reject the null hypothesis in this situation and might be more reasonable than the Cauchy combination test. However, the distribution of the p-value (under the null or alternative) is always stochastically larger than or equal to $U[0,1]$. Therefore, the p-value could be very close to 0 with a high probability (when from the alternative) but the chance to have a p-value close to 1 is always very small. Hence, the situation with p-values very close to 1 could but rarely happen and therefore would only lead to little power loss. Furthermore, from both the power theorem (i.e., Theorem 3 in main text) and the simulation studies (i.e., Figure 3 in the main text), we can see that the power of the Cauchy combination test is comparable with that of MinP. More importantly, the Cauchy combination test has fast p-value calculation and can be applied to analyze massive data but the MinP cannot.

References

Cai, T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372.