Supplement to "Accurate and Efficient P-value Calculation via Gaussian Approximation: a Novel Monte-Carlo Method"

In Section 1, we provide the proofs of the main theorems and technical lemmas. In Section 2, we justify Condition (A.3) that is required for Theorem 2 and 3. In Section 3, we perform simulations to assess the accuracy of our main theorems. The formulae of asymptotical critical values and additional tables for the simulation results of uniform, gamma and Bernoulli distributions are given in Section 4.

1 Proofs of Main Results

For a smooth multivariable function $f(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}$, denote the partial derivatives by $\partial_j f = \frac{\partial f}{\partial x_j}(\mathbf{x}), \partial_k \partial_j f = \frac{\partial^2 f}{\partial x_k \partial x_j}(\mathbf{x})$ and $\partial_l \partial_k \partial_j f = \frac{\partial^3 f}{\partial x_l \partial x_k \partial x_j}(\mathbf{x})$. We omit the argument of partial derivative functions to simplify the notation. Write $a \preceq b$ if a is smaller than or equal to b up to some positive multiplicative constant independent of n and d. The constant c_0 in the definition of $T^*_{\mathrm{HC}}(\cdot)$ and $T^*_{\mathrm{BJ}}(\cdot)$ is not essential in the proof. For the brevity of exposition, hereafter we consider $c_0 = 1$ without loss of generality.

For all the three tests, we will first show the desired bound holds for the known variance case $\mathbf{r}^\sigma$ and then extend the results to incorporate the variance estimator s_y^2. The proof for the minimum p-value method involves a direct application of the results in Chernozhukov et al. (2014) and additional efforts for handling s_y^2. Lemma 2 plays an important role for the higher criticism test. The key component of the proof of Lemma 2 is to construct smooth functions satisfying some nice properties and is provided in Section 1.4. The proof strategy for the Berk–Jones test is analogous to the higher criticism test and we provide an outline in Section 1.5.

1.1 Proof of Theorem 1

Lemma 1. *Under the null hypothesis and Conditions (A.1) and (A.2), we have*

$$\sup_{t\in\mathbb{R}} \left| P\left(\max_{1\le i\le d} |r_i^\sigma| \le t\right) - P\left(\max_{1\le i\le d} |v_i| \le t\right) \right| \preceq \frac{B_n^{2/3} \log^{7/6}(dn)}{n^{1/6}}. \tag{1}$$

Proof. Let $\xi_{ki} = \tilde{x}_{ki}\varepsilon_k/\sigma$, then $r_i^\sigma = (\sum_{k=1}^n \xi_{ki})/\sqrt{n}$ under the null. It is easy to see that

$$\frac{1}{n}\sum_{k=1}^n E(\xi_{ki}^2) = \frac{1}{n}\sum_{k=1}^n \tilde{x}_{ki}^2 = 1. \tag{2}$$

Since $|\tilde{x}_{ki}| \le B_n$ and $E(\varepsilon_k^4)$ is bounded, it is obvious that for any $1 \le i \le d$,

$$\frac{1}{n}\sum_{k=1}^n E(\xi_{ki}^4) = \frac{E(\varepsilon_k^4)}{n\sigma^4}\sum_{k=1}^n (\tilde{x}_{ki})^4 \preceq \frac{1}{n}\sum_{k=1}^n (\tilde{x}_{ki})^2 \cdot B_n^2 = B_n^2 \tag{3}$$

and for any $1 \le k \le n$,

$$E[(\max_{1\le i\le d} |\xi_{ki}|/B_n^2)^4] \preceq E\varepsilon_k^4/(\sigma B_n)^4 \preceq 1. \tag{4}$$

Applying the second part of Proposition 2.1 in Chernozhukov et al. (2014) with (2),(3) and (4), we obtain (1).

□

Lemma 1 gives the bound of the Gaussian approximation error when the variance σ^2 is known. We next study the case with unknown variance. For any $t \in \mathbb{R}$,

$$\begin{aligned}
& P\left(\max_{1\le i\le d} |r_i| \le t\right) - P\left(\max_{1\le i\le d} |v_i| \le t\right) \\
= \; & \left[P\left(\max_{1\le i\le d} |r_i| \le t\right) - P\left(\max_{1\le i\le d} |r_i^\sigma| \le t + \delta_n\right)\right] \\
& + \left[P\left(\max_{1\le i\le d} |r_i^\sigma| \le t + \delta_n\right) - P\left(\max_{1\le i\le d} |v_i| \le t + \delta_n\right)\right] \\
& + \left[P\left(\max_{1\le i\le d} |v_i| \le t + \delta_n\right) - P\left(\max_{1\le i\le d} |v_i| \le t\right)\right] = I + II + III
\end{aligned}$$

First of all, applying Lemma 1 yields

$$II \preceq \frac{B_n^{2/3} \log^{7/6}(dn)}{n^{1/6}}, \tag{5}$$

and it follows from Theorem 3(i) in Chernozhukov et al. (2015) that

$$III \preceq \delta_n(\sqrt{2\log 2d}+1). \tag{6}$$

It remains to provide a bound for the first term. Note that $\max_{1\le i\le d}|r_i| = \frac{\sigma}{s_y}\max_{1\le i\le d}|r_i^\delta|$, then by Bonferroni inequality,

$$P\left(\max_{1\le i\le d}|r_i|\le t\right)\le P\left(\max_{1\le i\le d}|r_i^\sigma|\le t+\delta_n\right)+P\left(|\sigma/s_y-1|\cdot\max_{1\le i\le d}|r_i^\sigma|>\delta_n\right).$$

Therefore, we have

$$\begin{aligned} I &\le P\left(|\sigma/s_y-1|\cdot\max_{1\le i\le d}|r_i^\sigma|>\delta_n\right)\\ &\le P\left(|\sigma/s_y-1|>\delta_n/\sqrt{8\log dn}\right)+P\left(\max_{1\le i\le d}|r_i^\sigma|>\sqrt{8\log dn}\right)=I_1+I_2 \end{aligned}$$

Recall that $E(\varepsilon_k^4)\le C$, it follows from Chebyshevs inequality that

$$P(|s_y^2-\sigma^2|>c)\preceq\frac{E\varepsilon_k^4}{nc^2}\preceq\frac{1}{nc^2},$$

and then we have

$$P(|\sigma/s_y-1|>c)=P(|s_y^2-\sigma^2|>s_y(s_y+\sigma)\cdot c)\preceq\frac{1}{nc^2}. \tag{7}$$

Hence

$$I_1\preceq\frac{\log dn}{n\delta_n^2}. \tag{8}$$

Using Lemma 7 in Chernozhukov et al. (2015) yields

$$P\left(\max_{1\le i\le d}|v_i|-E[\max_{1\le i\le d}|v_i|]\ge c\right)\le 2e^{-c^2/2}.$$

Taking $c=\sqrt{2\log dn}$ and using the fact that $E[\max_{1\le i\le d}|v_i|]\le\sqrt{2\log 2d}$ (see, for example, Proposition 1.1.3 in Talagrand, 2003), we have

$$P\left(\max_{1\le i\le d}|v_i|\ge\sqrt{8\log dn}\right)\le\frac{2}{dn}. \tag{9}$$

It follows from Lemma 1 that

$$I_2\preceq\frac{B_n^{2/3}\log^{7/6}(dn)}{n^{1/6}}+\frac{1}{dn}. \tag{10}$$

By combining the bounds in (5),(6),(8) and (10) together and taking $\delta_n=n^{-1/3}$, we obtain

$$P\left(\max_{1\le i\le d}|r_i|\le t\right)-P\left(\max_{1\le i\le d}|v_i|\le t\right)\preceq\frac{B_n^{2/3}\log^{7/6}(dn)}{n^{1/6}}$$

for any $t\in\mathbb{R}$. A similar argument provides the bound in the other direction.

1.2 Proof of Theorem 2

It can be easily shown that $\psi_i(x)$ is strictly increasing for any $1 \le i \le d$. Recall that $T^*_{\mathrm{HC}}(\mathbf{v}) = \max_{1\le i\le \log d} \psi_i(v_{(i)})$, hence

$$\{T^*_{\mathrm{HC}}(\mathbf{v}) \le t\} = \bigcap_{1\le i\le \log d} \{v_{(i)} \le \psi_i^{-1}(t)\} = \{\max_{1\le i\le \log d}[v_{(i)} - \psi_i^{-1}(t)] \le 0\}. \tag{11}$$

We prepare some notations for the right-hand side of (11). Define $m_i(\mathbf{x}) : \mathbf{x} \mapsto |x_i|$ and $f_i(\mathbf{x}) : \mathbf{x} \mapsto x_{(i)}$, where $\mathbf{x} = (x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$ and $x_{(i)}$ is the i-th largest number, for any $i = 1, 2, \cdots, d$. For example, $f_1(\mathbf{x}) = \max_{1\le s\le d} x_s$. Further let $\mathbf{m}(\mathbf{x}) = (m_i(\mathbf{x}); 1 \le i \le d), \mathbf{f}(\mathbf{x}) = (f_i(\mathbf{x}); 1 \le i \le d)$ and $f_1^*(\mathbf{x}) = \max_{1\le s\le \log d} x_s$.

Let

$$H(\cdot) = f_1^*((\mathbf{f} \circ \mathbf{m})(\cdot) - \mathbf{b}), \tag{12}$$

where $\mathbf{b} \in \mathbb{R}^d$ is a constant vector. Specially, let $H_\psi(\cdot)$ denote $H(\cdot)$ with $\mathbf{b} = (\psi_1^{-1}(t), \cdots, \psi_d^{-1}(t))^T$. Then we have $H_\psi(\mathbf{v}) = \max_{1\le i\le \log d}[v_{(i)} - \psi_i^{-1}(t)]$, it remains to show

$$\sup_{t\in\mathbb{R}} |P\{H_\psi(\mathbf{r}) \le 0\} - P\{H_\psi(\mathbf{v}) \le 0\}| \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \tag{13}$$

The following key lemma indicates the desired bound holds in the situation of known variance and its proof is given in Section 1.3.

Lemma 2. *Under the null hypothesis and Conditions (A.1),(A.2) and (A.3), we have*

$$|P\{H(\mathbf{r}^\sigma) \le 0\} - P\{H(\mathbf{v}) \le 0\}| \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}$$

for any $\mathbf{b} \in \mathbb{R}^d$.

We next incorporate the variance estimator s_y^2. It is obvious that for any $1 \le i < j \le d$ and $x > 0$, $\psi_i(x) < \psi_j(x)$ and $\psi_i^{-1}(x) > \psi_j^{-1}(x)$. Let L be the largest integer not greater than $\log d$, Lemma 7 implies there exist a constant t_0 such that

$$\sqrt{10\log dn} \le \psi_L^{-1}(t_0) \le \psi_1^{-1}(t_0) \le \sqrt{20\log dn}. \tag{14}$$

For $t > t_0$,

$$\begin{aligned}
P\{H_\psi(\mathbf{r}) \ge 0\} &= P\left\{\max_{1\le i\le \log d}[r^\sigma_{(i)} - \frac{s_y}{\sigma}\psi_i^{-1}(t)] \ge 0\right\} \\
&\le P\left\{\max_{1\le i\le \log d}[r^\sigma_{(i)} - \frac{2}{\sqrt{5}}\psi_i^{-1}(t)] \ge 0\right\} + P\left\{\frac{s_y}{\sigma} \le \frac{2}{\sqrt{5}}\right\} \\
&\preceq P\left\{\max_{1\le i\le \log d}[v_{(i)} - \frac{2}{\sqrt{5}}\psi_i^{-1}(t)] \ge 0\right\} + \frac{1}{n} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}
\end{aligned}$$

where the last inequality follows from Lemma 2 and (7). Note that $\psi_L^{-1}(t) = \min_{1\le i\le \log d}\psi_i^{-1}(t)$, then we have

$$
\begin{aligned}
& |P\{H_\psi(\mathbf{r}) \ge 0\} - P\{H_\psi(\mathbf{v}) \ge 0\}| \\
\preceq\ & 2P\left\{\max_{1\le i\le \log d}[v_{(i)} - \frac{2}{\sqrt{5}}\psi_i^{-1}(t)] \ge 0\right\} + \frac{1}{n} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \\
\le\ & 2\log d P\left\{v_{(1)} \ge \frac{2}{\sqrt{5}}\min_{1\le i\le \log d}\psi_i^{-1}(t)\right\} + \frac{1}{n} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \\
\le\ & 2\log d P\left\{v_{(1)} \ge \sqrt{8\log dn}\right\} + \frac{1}{n} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \\
\le\ & \frac{4\log d}{dn} + \frac{1}{n} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}
\end{aligned}
$$

where the third inequality follows from (14) and the fourth from (9).

For $t \le t_0$,

$$
\begin{aligned}
P\{H_\psi(\mathbf{r}) \le 0\} &= P\left\{\max_{1\le i\le \log d}[r^\sigma_{(i)} - \frac{s_y}{\sigma}\psi_i^{-1}(t)] \le 0\right\} \\
&\le P\left\{\max_{1\le i\le \log d}[r^\sigma_{(i)} - \psi_i^{-1}(t)] \le \delta_n\sqrt{20\log dn}\right\} + P\left\{\frac{s_y}{\sigma} \ge 1+\delta_n\right\} \\
&\preceq P\left\{\max_{1\le i\le \log d}[v_{(i)} - \psi_i^{-1}(t)] \le \delta_n\sqrt{20\log dn}\right\} + \frac{1}{n\delta_n^2} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}
\end{aligned}
$$

where the first inequality follows from $\psi_1^{-1}(t) = \max_{1\le i\le \log d}\psi_i^{-1}(t)$ and (14), and the last from Lemma 2 and (7). Taking $\delta_n = n^{-1/3}$,

$$
\begin{aligned}
& P\{H_\psi(\mathbf{r}) \le 0\} - P\{H_\psi(\mathbf{v}) \le 0\} \\
\preceq\ & P\left\{0 \le \max_{1\le i\le \log d}[v_{(i)} - \psi_i^{-1}(t)] \le \delta_n\sqrt{20\log dn}\right\} + \frac{1}{n\delta_n^2} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \\
\le\ & \sum_{1\le i\le \log d} P\left\{0 \le v_{(i)} - \psi_i^{-1}(t) \le \delta_n\sqrt{20\log dn}\right\} + \frac{1}{n\delta_n^2} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \\
\le\ & \frac{(\log d)^2\sqrt{20\log dn}}{n^{1/3}} + \frac{1}{n^{1/3}} + \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}} \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}
\end{aligned}
$$

where the third inequality is due to the assumption that the density of v_i is bounded by $\log d$ up to some constant for $1 \le i \le \log d$. The bound in the other direction follows similarly.

Combining the results of the two cases, i.e., $t > t_0$ and $t \le t_0$, we have proved (13).

1.3 Proof of Lemma 2

The proof strategy relies on a Slepian-Stein method developed in Chernozhukov et al. (2013). Recall the definition of $H(\cdot)$ in (12). It is obvious that $H(\cdot)$ is continuous but not differentiable. Then $\tilde{H}(\cdot)$, defined in Lemma 5, is a smooth approximation to $H(\cdot)$.

Let $\tilde{\omega}_0(x) : \mathbb{R} \mapsto [0,1]$ be a non-increasing function in C^3 such that $\tilde{\omega}_0(x) = 1$ for $x \leq 0$ and $\tilde{\omega}_0(x) = 0$ for $x \geq 1$. Fix any $a \in \mathbb{R}$ and define $\tilde{\omega}(x) = \tilde{\omega}_0(\lambda(x-a))$, which is a smooth approximation to indicator function $I(x \leq a)$. Then

$$|\tilde{\omega}(x)| \leq 1, |\tilde{\omega}'(x)| \preceq \lambda, |\tilde{\omega}''(x)| \preceq \lambda^2, |\tilde{\omega}'''(x)| \preceq \lambda^3. \tag{15}$$

Define $\tilde{F}(\cdot) = (\tilde{\omega} \circ \tilde{H})(\cdot)$.

Step 1. Our goal of the first step is to show

$$\left| E[\tilde{F}(\mathbf{r}^\sigma) - \tilde{F}(\mathbf{v})] \right| \preceq \frac{\lambda\gamma^2 + \lambda^2\gamma + \lambda^3}{\sqrt{n}} B_n^3. \tag{16}$$

Let $\xi_{ki} = \tilde{x}_{ki}\varepsilon_k/\sigma, \eta_{ki} = \tilde{x}_{ki}e_k, \boldsymbol{\xi}_k = (\xi_{k1}, \xi_{k2}, \cdots, \xi_{kd})^T$ and $\boldsymbol{\eta}_k = (\eta_{k1}, \eta_{k2}, \cdots, \eta_{kd})^T$. Then $\mathbf{v} = (\sum_{k=1}^n \boldsymbol{\eta}_k)/\sqrt{n}$ and $\mathbf{r}^\sigma = (\sum_{k=1}^n \boldsymbol{\xi}_k)/\sqrt{n}$. Define the Slepian interpolant

$$\mathbf{Z}(s) = \sum_{k=1}^n \mathbf{Z}_k(s), s \in [0,1],$$

where

$$\mathbf{Z}_k(s) = \frac{1}{\sqrt{n}}(\sqrt{s}\boldsymbol{\xi}_k + \sqrt{1-s}\boldsymbol{\eta}_k).$$

Then we have

$$\tilde{F}(\mathbf{r}^\sigma) - \tilde{F}(\mathbf{v}) = \tilde{F}(\mathbf{Z}_k(1)) - \tilde{F}(\mathbf{Z}_k(0)) = \int_0^1 \frac{d\tilde{F}(\mathbf{Z}_k(s))}{ds} ds. \tag{17}$$

Denote the Stein leave-one-out term for $\mathbf{Z}(s)$ by

$$\mathbf{Z}^{(k)}(s) = \mathbf{Z}(s) - \mathbf{Z}_k(s).$$

Applying Taylor's theorem with integral remainder and (17), we obtain

$$E[\tilde{F}(\mathbf{r}^\sigma) - \tilde{F}(\mathbf{v})] = \sum_{i=1}^d \sum_{k=1}^n \int_0^1 E[\partial_i \tilde{F}(\mathbf{Z}(s)) Z'_{ki}(s)] ds = I + II + III,$$

where

$$I = \sum_{i=1}^{d} \sum_{k=1}^{n} \int_0^1 E[\partial_i \tilde{F}(\mathbf{Z}^{(k)}(s)) Z'_{ki}(s)] ds,$$

$$II = \sum_{i,j=1}^{d} \sum_{k=1}^{n} \int_0^1 E[\partial_j \partial_i \tilde{F}(\mathbf{Z}^{(k)}(s)) Z'_{ki}(s) Z_{kj}(s)] ds$$

$$III = \sum_{i,j,l=1}^{d} \sum_{k=1}^{n} \int_0^1 \int_0^1 (1-\tau) E[\partial_l \partial_j \partial_i \tilde{F}(\mathbf{Z}^{(k)}(s) + \tau \mathbf{Z}_k) Z'_{ki}(s) Z_{kj}(s) Z_{kl}(s)] d\tau ds.$$

Note that $\mathbf{Z}^{(k)}(s)$ is independent of $Z'_{ki}(s)$ and

$$Z'_{ki}(s) = \frac{1}{2\sqrt{n}} \left(\frac{\xi_{ki}}{\sqrt{s}} - \frac{\eta_{ki}}{\sqrt{1-s}} \right),$$

we have $EZ'_{ki}(s) = 0$ and hence $I = 0$.

Also, by the independence of $\mathbf{Z}^{(k)}(s)$ from $Z'_{ki}(s) Z_{kj}(s)$,

$$E[\partial_j \partial_i \tilde{F}(\mathbf{Z}^{(k)}(s)) Z'_{ki}(s) Z_{kj}(s)] = E[\partial_j \partial_i \tilde{F}(\mathbf{Z}^{(k)}(s))] \cdot E[Z'_{ki}(s) Z_{kj}(s)].$$

A direct calculation gives

$$E[Z'_{ki}(s) Z_{kj}(s)] = \frac{1}{2n} E \left(\xi_{ki}\xi_{kj} + \sqrt{\frac{1-s}{s}} \xi_{ki}\eta_{kj} - \sqrt{\frac{s}{1-s}} \eta_{ki}\xi_{kj} - \eta_{ki}\eta_{kj} \right) = 0,$$

the last equality is due to $E\xi_{ki}\xi_{kj} = E\eta_{ki}\eta_{kj}$ and the independence of ε_k from e_k. Therefore $II = 0$.

It remains to bound III. Recall that $\tilde{F} = \tilde{\omega} \circ \tilde{H}$, applying Lemma 3 with (15) and (27), we have

$$\sum_{i,j,l=1}^{d} |\partial_l \partial_j \partial_i \tilde{F}| \preceq \lambda\gamma^2 + \lambda^2\gamma + \lambda^3.$$

Thus

$$\begin{aligned} |III| &\preceq (\lambda\gamma^2 + \lambda^2\gamma + \lambda^3) \sum_{k=1}^{n} \int_0^1 \int_0^1 |1-\tau| \cdot E\left[\max_{1 \le i,j,l \le d} |Z'_{ki}(s) Z_{kj}(s) Z_{kl}(s)| \right] d\tau ds \\ &\preceq (\lambda\gamma^2 + \lambda^2\gamma + \lambda^3) \sum_{k=1}^{n} \int_0^1 E\left[\max_{1 \le i,j,l \le d} |Z'_{ki}(s) Z_{kj}(s) Z_{kl}(s)| \right] ds. \end{aligned}$$

Recall that $\tilde{x}_{ki}$'s are uniformly bounded by B_n, it is easy to see

$$\begin{aligned}\max_{1\le i,j,l\le d}|Z'_{ki}(s)Z_{kj}(s)Z_{kl}(s)| &\preceq n^{-3/2}\sqrt{s(1-s)}\max_{1\le i,j,l\le d}(|\xi_{ki}|+|\eta_{ki}|)(|\xi_{kj}|+|\eta_{kj}|)(|\xi_{kl}|+|\eta_{kl}|)\\ &\preceq n^{-3/2}B_n^3(|\varepsilon_k/\sigma|+|e_k|)^3.\end{aligned}$$

By Minkowski's inequality and recall $E(\varepsilon_k^4)$ is bounded,

$$E(|\varepsilon_k/\sigma|+|e_k|)^3 \le [(E|\varepsilon_k/\sigma|^3)^{1/3}+(E|e_k|^3)^{1/3}]^3 \preceq 1.$$

Hence

$$|III| \preceq \frac{\lambda\gamma^2+\lambda^2\gamma+\lambda^3}{\sqrt{n}}B_n^3.$$

We complete the proof of (16).

Step 2. Now we go back to deal with the original non-smooth functions. Recall the construction of $\tilde{\omega}(x)$, then

$$I(x\le a)\le\tilde{\omega}(x)\le I(x\le 1/\lambda+a). \tag{18}$$

Taking $a=4(\log d)^2/\gamma$, which is the upper bound in (26), we have

$$\begin{aligned}P\{H(\mathbf{r}^\sigma)\le 0\} &\le P\left(\tilde{H}(\mathbf{r}^\sigma)-4(\log d)^2/\gamma\le 0\right)\le E[(\tilde{\omega}\circ\tilde{H})(\mathbf{r}^\sigma)]\\ &\preceq E[(\tilde{\omega}\circ\tilde{H})(\mathbf{v})]+\frac{\lambda\gamma^2+\lambda^2\gamma+\lambda^3}{\sqrt{n}}B_n^3\\ &\le P\{H(\mathbf{v})\le 1/\lambda+a\}+\frac{\lambda\gamma^2+\lambda^2\gamma+\lambda^3}{\sqrt{n}}B_n^3,\end{aligned}$$

where the third inequality follows from (16) and the other inequalities follow from (18) and (26). Because the density of $v_{(i)}$ is bounded by $\log d$ for $1\le i\le\log d$, then

$$\begin{aligned}P\{H(\mathbf{v})\le 1/\lambda+a\}-P\{H(\mathbf{v})\le 0\} &= P\{0\le H(\mathbf{v})\le 1/\lambda+a\}\\ &\le \sum_{1\le i\le\log d}P\{0\le|v_i|-b_i\le 1/\lambda+a\}\preceq(\lambda^{-1}+a)(\log d)^2.\end{aligned}$$

Thus we have

$$P\{H(\mathbf{r}^\sigma)\le 0\}-P\{H(\mathbf{v})\le 0\}\preceq\frac{(\log d)^2}{\lambda}+\frac{(\log d)^4}{\gamma}+\frac{\lambda\gamma^2+\lambda^2\gamma+\lambda^3}{\sqrt{n}}B_n^6.$$

To minimize the right-hand side with respect to d and n, we take $\gamma=\lambda(\log d)^2$ and $\lambda=n^{1/8}/(B_n^{3/2}\sqrt{\log d})$. Therefore,

$$P\{H(\mathbf{r}^\sigma)\le 0\}-P\{H(\mathbf{v})\le 0\}\preceq\frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}.$$

A similar argument gives the bound in the other direction.

1.4 Technical Lemmas

It can be seen that $f_i(\cdot), m_i(\cdot), f_1^*(\cdot)$ and $H(\cdot)$ are continuous but not differentiable. We aim to find smooth functions to approximate them and use an over "$\sim$" to denote their corresponding smooth approximations. We begin with an useful lemma, which will be frequently used in the subsequent proof.

Lemma 3 (Bounds of derivatives)**.** *Let* $\boldsymbol{g}(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_q(\mathbf{x})) : \mathbb{R}^d \mapsto \mathbb{R}^q$ *be a map such that* $\sum_{j=1}^d |\partial_j g_i| \le c_1, \sum_{j,k=1}^d |\partial_k \partial_j g_i| \le c_2, \sum_{j,k,l=1}^d |\partial_l \partial_k \partial_j g_i| \le c_3$ *uniformly for any* $\mathbf{x} \in \mathbb{R}^d$ *and* $i = 1, 2, \cdots, q$, *where integers* $d, q \ge 1$ *and* c_1, c_2, c_3 *are some positive constants . Let* $h(\mathbf{y}) : \mathbb{R}^q \mapsto \mathbb{R}$ *be a thrice differentiable function such that* $\sum_{j=1}^q |\partial_j h| \le C_1, \sum_{j,k=1}^q |\partial_k \partial_j h| \le C_2, \sum_{j,k,l=1}^q |\partial_l \partial_k \partial_j h| \le C_3$ *for any* $\mathbf{y} \in \mathbb{R}^q$, *where* C_1, C_2 *and* C_3 *are some positive constants.*

Then we have

$$\sum_{j=1}^d |\partial_j (h \circ \mathbf{g})| \le c_1 C_1, \quad \sum_{j,k=1}^d |\partial_k \partial_j (h \circ \mathbf{g})| \le (c_1)^2 C_2 + c_2 C_1$$

and

$$\sum_{j,k,l=1}^d |\partial_l \partial_k \partial_j (h \circ \mathbf{g})| \le (c_1)^3 C_3 + 3 c_1 c_2 C_2 + c_3 C_1.$$

Proof. The results follow from repeated application of the chain rule. □

Now consider the max function $f_1(\mathbf{x})$ at first. Define the "smooth max function" as $\tilde{f}_1(\mathbf{x}) = \frac{1}{\gamma} \log\{\sum_{s=1}^d \exp(\gamma x_s)\}$, where $\gamma > 0$ is the smooth parameter which controls the level of approximation. An elementary calculation gives

$$0 \le \tilde{f}_1(\mathbf{x}) - f_1(\mathbf{x}) \le \frac{\log d}{\gamma}. \tag{19}$$

Through repeated application of the chain rule, it is easy to show that

$$\partial_j \tilde{f}_1 \ge 0, \quad \sum_{j=1}^d \partial_j \tilde{f}_1 = 1, \quad \sum_{k,j=1}^d |\partial_k \partial_j \tilde{f}_1| \le 2\gamma, \quad \sum_{l,k,j=1}^d |\partial_l \partial_k \partial_j \tilde{f}_1| \le 6\gamma^2. \tag{20}$$

for $1 \le j, k, l \le d$ (see also Lemma A.2 of Chernozhukov et al., 2013).

We next turn to construct a smooth function to approximate $f_i(\mathbf{x})$ for any $i = 1, 2, \cdots, d$. For any given i, define a set of indices sets $\mathscr{A}_i = \{\{s_1, s_2, \cdots, s_i\} : 1 \le s_1 < s_2 < \cdots < s_i \le d\}$. Then $|\mathscr{A}_i| = \binom{d}{i}$, where $|\mathscr{A}_i|$ denotes the cardinality of $\mathscr{A}_i$.

Observe that $f_i(\mathbf{x}) = \max_{A\in\mathscr{A}_i} \min_{s\in A} x_s = \max_{A\in\mathscr{A}_i}\{-\max_{s\in A}(-x_s)\}$, we can use the "smooth max function" recursively to construct a smooth function for $f_i(\mathbf{x})$. Specifically, define $g_{i,A}(\mathbf{x}) = -\frac{1}{\gamma}\log\{\sum_{s\in A}\exp(-\gamma x_s)\} + \frac{\log i}{\gamma}$ for any $A \in \mathscr{A}_i$, and

$$\tilde{f}_i(\mathbf{x}) = \frac{1}{\gamma}\log\left(\sum_{A\in\mathscr{A}_i}\exp\{\gamma g_{i,A}(\mathbf{x})\}\right). \tag{21}$$

Lemma 4. *(i)For every $1 \le i \le d$, we have*

$$0 \le \tilde{f}_i(\mathbf{x}) - f_i(\mathbf{x}) \le \frac{\log i + \log\binom{d}{i}}{\gamma}. \tag{22}$$

(ii) For every $1 \le i, j, k, l \le d$, we have

$$\partial_j \tilde{f}_i \ge 0, \quad \sum_{j=1}^{d}\partial_j\tilde{f}_i = 1, \quad \sum_{k,j=1}^{d}|\partial_k\partial_i\tilde{f}_i| \le 4\gamma, \quad \sum_{l,k,j=1}^{d}|\partial_l\partial_k\partial_j\tilde{f}_i| \le 24\gamma^2. \tag{23}$$

Proof. Note that $0 \le g_{i,A}(\mathbf{x}) - \min_{s\in A} x_s \le \frac{\log i}{\gamma}$ for any $A \in \mathscr{A}_i$, then (22) follows from (19). To show (23), we begin with some additional notation. Let $\mathscr{A}_i(j) = \{A \in \mathscr{A}_i : j \in A\}$ and $\tilde{g}_i(\mathbf{x}) = \sum_{A\in\mathscr{A}_i}\exp\{\gamma g_{i,A}(\mathbf{x})\}$, where $\mathscr{A}_i(j)$ is a collection of subsets of $\mathscr{A}_i$. Then $\tilde{f}_i(\mathbf{x}) = \frac{1}{\gamma}\log\tilde{g}_i(\mathbf{x})$. Hereafter, we suppress the argument $\mathbf{x}$ for all the functions, e.g., $g_{i,A} = g_{i,A}(\mathbf{x})$.

A direct calculation gives $\partial_j\tilde{f}_i = \frac{\partial_j\tilde{g}_i}{\gamma\tilde{g}_i} = \frac{1}{\tilde{g}_i}\sum_{A\in\mathscr{A}_i(j)}\exp(\gamma g_{i,A})\partial_j g_{i,A} \ge 0$. Recall that $\sum_{j\in A}\partial_j g_{i,A} = 1$ and the definition of $\tilde{g}_i$, it follows that

$$\sum_{j=1}^{d}\partial_j\tilde{f}_i = \frac{1}{\tilde{g}_i}\sum_{j=1}^{d}\sum_{A\in\mathscr{A}_i(j)}\exp(\gamma g_{i,A})\partial_j g_{i,A} = \frac{1}{\tilde{g}_i}\sum_{A\in\mathscr{A}_i}\sum_{j\in A}\exp(\gamma g_{i,A})\partial_j g_{i,A} = 1.$$

Because of (20), for any $A \in \mathscr{A}_i$,

$$\sum_{j=1}^{d}\partial_j g_{i,A} = 1, \quad \sum_{k,j=1}^{d}|\partial_k\partial_j g_{i,A}| \le 2\gamma, \quad \sum_{l,k,j=1}^{d}|\partial_l\partial_k\partial_j g_{i,A}| \le 6\gamma^2.$$

Then the bounds for the second and third derivatives in (23) directly follow from Lemma 3. □

Note that $|x_i| = \max\{x_i, -x_i\}$, we define the smooth function for $m_i(\mathbf{x})$ as

$$\tilde{m}_i(\mathbf{x}) = \frac{1}{\gamma}\log\{\exp(\gamma x_i) + \exp(-\gamma x_i)\}. \tag{24}$$

Further, define

$$\tilde{f}_1^*(\mathbf{x}) = \frac{1}{\gamma}\log\left(\sum_{1\le s\le \log d}\exp(\gamma x_s)\right). \tag{25}$$

Since $\tilde{m}_i(\mathbf{x})$ and $\tilde{f}_1^*(\mathbf{x})$ are special cases of the "smooth max function", properties (19) and (20) also hold for them with d replaced by 2 and $\log d$, respectively.

Lemma 5. *Let $\tilde{H}(\cdot) = \tilde{f}_1^*((\tilde{\mathbf{f}}\circ\tilde{\mathbf{m}})(\cdot)-\mathbf{b})$, where $\tilde{\mathbf{f}}, \tilde{\mathbf{m}}$, and $\tilde{f}_1^*$ are defined as (21),(24) and (25) respectively. We have*

$$0 \le \tilde{H}(\mathbf{x}) - H(\mathbf{x}) \le \frac{4(\log d)^2}{\gamma}. \tag{26}$$

and

$$\sum_{j=1}^{d}\partial_j\tilde{H} \preceq 1, \quad \sum_{k,j=1}^{d}|\partial_k\partial_j\tilde{H}| \preceq \gamma, \quad \sum_{l,k,j=1}^{d}|\partial_l\partial_k\partial_j\tilde{H}| \preceq \gamma^2. \tag{27}$$

Proof. By (22), for any $1 \le i \le \log d$, we have $0 \le \tilde{f}_i(\mathbf{x}) - f_i(\mathbf{x}) \le 2(\log d)^2/\gamma$. This, along with (19), gives

$$0 \le \tilde{H}(\mathbf{x}) - H(\mathbf{x}) \le \frac{\log 2 + \log d + 2(\log d)^2}{\gamma} \le \frac{4(\log d)^2}{\gamma}.$$

Through repeated application of Lemma 3 with (20) and (23), we obtain (27).

□

Lemma 6 (Mill's inequality)**.** *For any $x > 0$,*

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the p.d.f and c.d.f. of the standard normal distribution, respectively.

Lemma 7. *Let L be the largest integer not greater than $\log d$. We have*

$$\psi_L(\sqrt{10\log dn}) \le \psi_1(\sqrt{16\log dn}).$$

Proof. Denote $x_1 = \sqrt{16\log dn}$ and $x_L = \sqrt{10\log dn}$. Recall that $\pi(x) = 2[1-\Phi(x)]$. From Mill's inequality, it is easy to see $\pi(x_1) < \pi(x_L) < 1/(2d)$. Then it suffices to show

$$1 \le \left[\frac{\psi_1(x_1)}{\psi_L(x_L)}\right]^2 = \left[\frac{1/d-\pi(x_1)}{L/d-\pi(x_L)}\right]^2\cdot\frac{\pi(x_L)}{\pi(x_1)}\cdot\frac{1-\pi(x_L)}{1-\pi(x_1)}.$$

By $\pi(x_1) < \pi(x_L) < 1/(2d)$, we have

$$\left[\frac{1/d - \pi(x_1)}{L/d - \pi(x_L)}\right]^2 \geq \left[\frac{1/(2d)}{L/d}\right]^2 \geq \left[\frac{1}{2\log d}\right]^2.$$

and

$$\frac{1-\pi(x_L)}{1-\pi(x_1)} \geq \frac{d-1}{d}.$$

Applying Mill's inequality gives

$$\frac{\pi(x_L)}{\pi(x_1)} \geq \frac{\phi(x_L)/x_L}{2\phi(x_1)/x_1} = \frac{(dn)^3}{2}.$$

Recall that $n \geq 2$, therefore

$$\left[\frac{\psi_1(x_1)}{\psi_L(x_L)}\right]^2 \geq \left[\frac{1}{2\log d}\right]^2 \cdot \frac{d-1}{d} \cdot \frac{(dn)^3}{2} \geq \frac{d^2(d-1)}{(\log d)^2} \geq 1.$$

$\square$

1.5 Proof of Theorem 3

It is obvious that $\phi_i(x) \geq 0$ and $\phi_i(x)$ is strictly decreasing when $0 < x < \pi^{-1}(i/d)$ and strictly increasing when $x \geq \pi^{-1}(i/d)$, for any $1 \leq i \leq d$. Hence for any given $t \geq 0$, there exists positive constants $c_i^-(t) \leq c_i^+(t)$ such that

$$\{x : \phi_i(x) \leq t\} = \{c_i^-(t) \leq x \leq c_i^+(t)\}.$$

Recall that

$$T^*_{\mathrm{BJ}}(\mathbf{v}) = \max_{1\leq i\leq \log d} \phi_i(v_{(i)})$$

and then

$$\{T^*_{\mathrm{BJ}}(\mathbf{v}) \leq t\} = \bigcap_{1\leq i\leq \log d} \{c_i^-(t) \leq v_{(i)} \leq c_i^+(t)\} = \big\{\max\{J^+(\mathbf{v}), J^-(\mathbf{v})\} \leq 0\big\}$$

where $J^+(\mathbf{v}) = \max_{1\leq i\leq \log d}\{v_{(i)} - c_i^+(t)\}$ and $J^-(\mathbf{v}) = \max_{1\leq i\leq \log d}\{c_i^-(t) - v_{(i)}\}$. Let $J(\cdot) = \max\{J^+(\cdot), J^-(\cdot)\}$, then it remains to show

$$\sup_{t\in\mathbb{R}} |P\{J(\mathbf{r}) \leq 0\} - P\{J(\mathbf{v}) \leq 0\}| \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}$$

Note that $J^+(\cdot)$ can be considered as the function $H(\cdot)$ defined in (12) with $\mathbf{b} = (c_1^+(t), \cdots, c_d^+(t))^T$. Through a similar approach of constructing $\tilde{H}(\cdot)$, we can find a smooth approximation that satisfies (26) and (27) for $J^-(\cdot)$. The maximum function with two arguments in the outer layer of $J(\cdot)$ can be handled in a similar manner as $\tilde{m}_i(\mathbf{x})$ in (24). In the end, we can construct a smooth function $\tilde{J}(\cdot)$ that also satisfies (26) and (27) to approximate $J(\cdot)$.

Then using the same argument in the proof of Lemma 2, it can be shown that

$$|P\{J(\mathbf{r}^\sigma) \leq 0\} - P\{J(\mathbf{v}) \leq 0\}| \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}.$$

The remaining part for dealing with the estimate of variance s_y^2 is also similar to the proof for higher criticism in Section 1.2. Thus we omit the details of proof here.

2 Justification of Condition (A.3)

Let $\mathbf{\Sigma}$ denote the correlation matrix of the Gaussian vector $\mathbf{v}$. The following lemma shows that Condition (A.3) holds when $\mathbf{\Sigma} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix.

Lemma 8. *Let $v_1, v_2, \cdots, v_d$ be i.i.d. standard Gaussian variables and $v_{(i)}$ be the ith largest absolute value of them. Denote the density of $v_{(i)}$ by $g_i(x)$. For $1 \leq i \leq c_0 \log d$, where $c_0 > 0$ is some fixed constant, we have $g_i(x) \preceq \log d$.*

Proof. It is easy to see that

$$g_i(x) = \frac{d!}{(d-i)!(i-1)!}[2\Phi(x)-1]^{d-i}\{2[1-\Phi(x)]\}^{i-1}[2\phi(x)],$$

where $x \geq 0$. Next, we decompose the domain of the density functions into two parts and derive the upper bound for each part.

For $x > \sqrt{2\log d}$, we have

$$\begin{aligned} g_i(x) &\leq \frac{d!}{(d-i)!(i-1)!}\{2[1-\Phi(x)]\}^{i-1}[2\phi(x)] \leq \frac{d!}{(d-i)!(i-1)!} \cdot \frac{2^i[\phi(t)]^i}{t^{i-1}} \\ &\leq \frac{d \cdot (d-1)\cdots(d-i+1)}{(i-1)!} \cdot \frac{1}{d^i} \leq 1, \end{aligned}$$

where the second inequality follows from the left-hand side of Mill's inequality (Lemma 6).

For $0 \le x \le \sqrt{2\log d}$, we have

$$\begin{aligned} g_i(x) &= \frac{d!}{(d-i)!(i-1)!}[2\Phi(x)-1]^{d-i}\{2[1-\Phi(x)]\}^i \frac{\phi(x)}{1-\Phi(x)} \\ &\le \frac{d!}{(d-i)!(i-1)!}[2\Phi(x)-1]^{d-i}\{2[1-\Phi(x)]\}^i\{2\max(x,1)\} \\ &\le i\frac{d!}{(d-i)!i!}\left(1-\frac{i}{d}\right)^{d-i}\left(\frac{i}{d}\right)^i\{2\max(x,1)\} \\ &\le i\sqrt{\frac{e^{1/6}d}{2\pi(d-i)i}}\{2\max(x,1)\} \preceq \log d, \end{aligned}$$

where the first inequality follows from Mill's inequality, the second inequality from an elementary inequality that $(1-y)^{d-i}y^i \le \left(1-\frac{i}{d}\right)^{d-i}\left(\frac{i}{d}\right)^i$ for $0 \le y \le 1$, the third inequality from Stirling's approximation that $\sqrt{2\pi}n^{n+1/2}e^{-n} \le n! \le \sqrt{2\pi}n^{n+1/2}e^{-n}e^{1/(12n)}$ for any positive integer n, and the last inequality from $1 \le i \le c_0 \log d$ and $0 \le x \le \sqrt{2\log d}$. $\square$

As described in Remark 1 in the main text, for general correlation structure $\mathbf{\Sigma}$ and $2 \le i \le c_0 \log d$, we anticipate that the maximum of the density function of $v_{(i)}$ should be no large than that when $\mathbf{\Sigma} = \mathbf{I}$, up to some positive constant. This along with Lemma 8 implies that Condition (A.3) would hold for general correlation matrices $\mathbf{\Sigma}$.

We use simulations to examine some specific correlation matrices. Following the simulation setting in Section 4.1 of the main text, we consider $d = 5, 20, 50, 100, 300, 500$. For each value of d, three types of correlation matrices $\mathbf{\Sigma} = (\sigma_{ij})$ are studied: (i) Exponentially off-diagonal decay, $\sigma_{ij} = \rho^{|i-j|}$ for $1 \le i, j \le d$, where $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$; (ii) Exchangeable correlation, $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$ for $1 \le i \ne j \le d$, where $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$; (iii) Real genotype-based correlation structures, with $\mathbf{\Sigma}$ as the sample correlation matrix of SNPs in each of the genes used in Section 4.1 of the main text. We consider the order statistic $v_{(\lfloor \log d \rfloor)}$, where $\lfloor x \rfloor$ denotes the largest integer that is no larger than x. Under each correlation matrix $\mathbf{\Sigma}$, 10^4 Monte Carlo samples are generated to obtain the density function of the order statistic $v_{(\lfloor \log d \rfloor)}$.

Figure 1 demonstrates the simulation results. It can be seen that the density functions in the i.i.d. case have the largest maximum value among these correlation matrices. Therefore, Condition (A.3) is valid for the examples with these specific correlation matrices.

3 Accuracy of main theorems

We carry out simulations to access Theorem 1-3. We use p-p plots to compare the distributions of $T(\mathbf{r})$ and its Gaussian approximation $T(\mathbf{v})$ for the minimum p-value, higher criticism, and Berk-Jones tests, respectively.

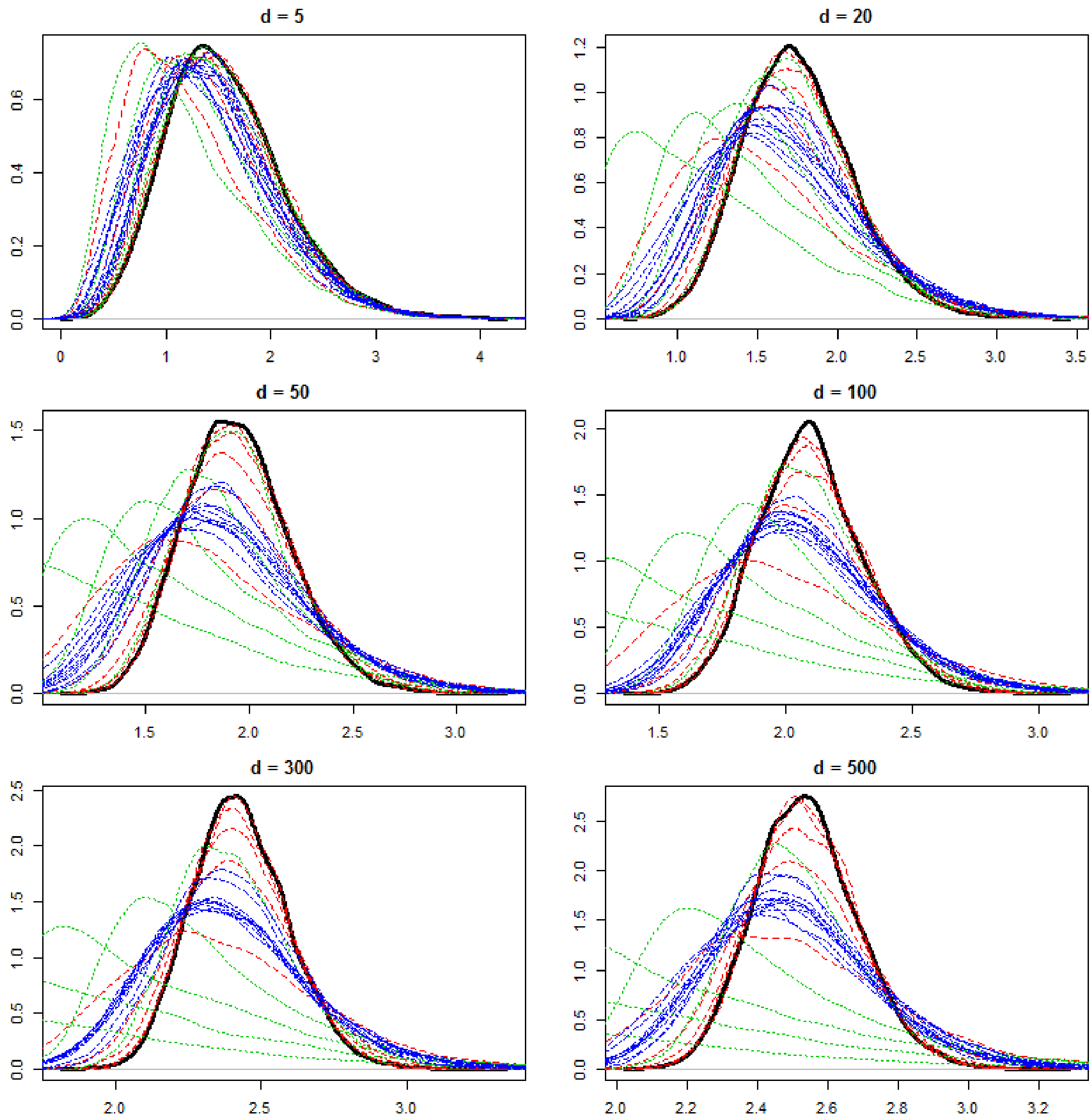

Figure 1: The density functions of $v_{(\lfloor \log d \rfloor)}$ under a variety of correlation matrices. The dimension of the correlation matrix d is demonstrated in the header of each subplot. In each subplot, the black curve corresponds to the i.i.d. case. The red, green and blue curves correspond to correlation structure (i),(ii) and (iii) described in the text, respectively.

Rewrite the regression model as $y_j = \mathbf{x}_j^T \boldsymbol{\beta} + \epsilon_j$, where y_j is the response and $\mathbf{x}_j$ is a $d \times 1$ covariates vector for the j-th subject, and $j = 1, 2, \cdots, n$. The covariates $\mathbf{x}_j$'s are generated independently from $N_d(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is chosen to have an AR(1) correlation structures (i.e., $\sigma_{ij} = \rho^{|i-j|}$ for $1 \le i, j \le d$) with $\rho = 0.9$ to mimic the correlation structure of SNP genotypes. The covariates $\mathbf{x}_j$'s are only simulated once and considered as non-stochastic. The responses y_j's are generated independently from $t(4)$ under the null hypothesis that $\boldsymbol{\beta} = 0$. Then we obtain the empirical distributions of $T(\mathbf{r})$ and $T(\mathbf{v})$ by simulating 5000 Monte Carlo samples.

We consider sample size $n = 1000$ and the number of covariates $d = 500, 1000, 2000$. The p-p plots are shown in Figure 2. It can be seen that the distributions of $T(\mathbf{r})$ and $T(\mathbf{v})$ are close to each other for all the three tests, as suggested by our Theorem 1–3. The Gaussian approximation is particularly accurate for the tail probabilities, which often correspond to p-values of interest in applications.

4 Supplementary Tables

The asymptotical critical values for the minimum p-value method, higher criticism and Berk–Jones tests are provided in Table 1. We compare the empirical sizes of the Gaussian approximation method, permutation and the method based on asymptotic distribution over a range of significance levels: $\alpha = 0.01, 0.001, 0.0001$. The simulation results of uniform, gamma and Bernoulli distributions are shown in Table 2–11. For the Bernoulli(p) distributions, $\mathbf{r}$ (i.e., Cochran-Armitage trend test statistics, see Section 2.3 in the main text) is used as the marginal test statistics when $p = 1/8$ and $p = 1/4$, and $\mathbf{r}^b$ is used when $p = 1/2$.

Table 1: Asymptotical critical values at significance level α, where $c_\alpha = \log\log(1-\alpha)^{-1}$.

MinP	$\sqrt{2\log d - \log\log d - \log\pi - 2c_\alpha}$
HC	$(4\log\log d + \log\log\log d - \log(4\pi) - 2c_\alpha)/\sqrt{8\log\log d}$
BJ	$\log\log d + (1/2)\log\log\log d - \log\sqrt{\pi/4} - c_\alpha$

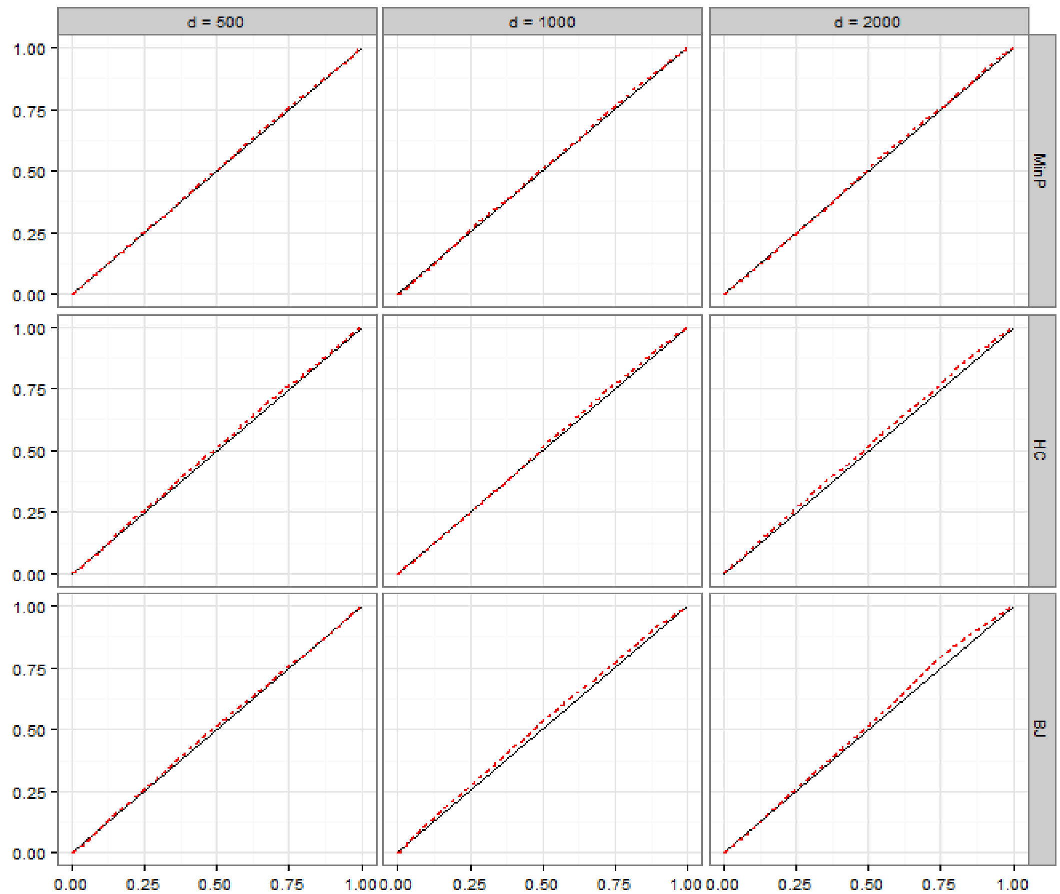

Figure 2: P-P plots comparing the distributions of $T(\mathbf{r})$ and its Gaussian approximation $T(\mathbf{v})$ for the minimum p-value, higher criticism, and Berk-Jones tests, respectively. The solid line in each plot is the 45° reference line. The sample size $n = 1000$. The columns correspond to dimension $d = 500, 1000, 2000$, respectively. The rows correspond to the three tests, respectively. The covariates $\mathbf{x}_j$'s are simulated from $N_d(0, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ has an AR(1) correlation structure with $\rho = 0.9$. The responses y_j's (or the error terms ϵ_j's) are drawn from $t(4)$ under the null.

Table 2: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 997$. The error term ε follows standardized Unif$(0, 1)$.

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.01	2.01	1.07	2.01	2.01	2.37	2.01	2.02	1.17
	3	3.01	3.02	1.48	3.01	3.02	3.45	3.01	3.02	1.68
	4	4.00	4.04	1.76	4.01	4.06	4.45	4.00	4.04	2.14
20	2	2.01	2.01	0.92	2.01	2.01	2.29	2.01	2.01	1.03
	3	3.03	3.01	1.25	3.03	3.02	3.35	3.02	3.00	1.41
	4	4.04	4.01	1.49	4.03	4.00	4.37	4.01	3.98	1.76
50	2	2.02	2.01	0.83	2.02	2.01	2.27	2.01	2.01	0.82
	3	3.04	3.01	1.14	3.04	3.01	3.33	3.04	3.02	1.14
	4	4.04	4.03	1.38	4.05	4.03	4.39	4.05	4.04	1.43
100	2	2.02	2.01	0.84	2.02	2.01	2.23	2.03	2.01	0.89
	3	3.04	3.03	1.16	3.05	3.03	3.30	3.04	3.01	1.25
	4	4.05	4.02	1.41	4.07	4.03	4.34	4.06	4.01	1.58
300	2	2.03	2.01	0.75	2.03	2.01	2.23	2.08	2.01	0.68
	3	3.04	3.02	1.06	3.05	3.02	3.29	3.10	3.03	0.97
	4	4.03	4.02	1.32	4.06	4.02	4.34	4.12	4.02	1.24
500	2	2.04	2.01	0.74	2.03	2.01	2.22	2.13	2.01	0.66
	3	3.06	3.02	1.04	3.06	3.03	3.28	3.18	3.04	0.94
	4	4.09	4.05	1.31	4.10	4.05	4.35	4.18	4.01	1.20

Table 3: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 1719$. The error term ε follows standardized Unif$(0, 1)$.

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.02	1.07	2.00	2.02	2.37	2.00	2.01	1.17
	3	3.00	3.01	1.48	3.00	3.01	3.43	3.00	3.03	1.67
	4	3.99	4.01	1.76	3.98	3.99	4.48	4.01	4.02	2.13
20	2	2.01	2.03	0.93	2.01	2.03	2.28	2.00	2.03	1.03
	3	3.02	3.04	1.25	3.02	3.03	3.33	3.00	3.04	1.42
	4	4.04	4.05	1.49	4.04	4.06	4.39	3.99	4.00	1.77
50	2	2.01	2.03	0.84	2.01	2.03	2.26	2.01	2.02	0.83
	3	3.02	3.05	1.14	3.02	3.05	3.30	3.02	3.03	1.16
	4	4.01	4.05	1.39	4.02	4.05	4.34	4.03	4.07	1.46
100	2	2.01	2.04	0.84	2.01	2.03	2.22	2.02	2.04	0.89
	3	3.03	3.05	1.16	3.03	3.05	3.27	3.03	3.06	1.25
	4	4.04	4.08	1.41	4.04	4.08	4.34	4.00	4.05	1.59
300	2	2.02	2.04	0.75	2.02	2.03	2.21	2.05	2.04	0.69
	3	3.03	3.05	1.06	3.03	3.04	3.26	3.06	3.07	0.97
	4	4.07	4.06	1.32	4.04	4.05	4.32	4.11	4.11	1.24
500	2	2.02	2.04	0.74	2.02	2.03	2.21	2.07	2.04	0.66
	3	3.04	3.05	1.04	3.04	3.05	3.26	3.10	3.06	0.93
	4	4.06	4.07	1.31	4.07	4.09	4.30	4.14	4.11	1.19

Table 4: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 997$. The error term ε follows standardized Gamma$(10, 1)$.

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.02	1.07	2.00	2.01	2.36	2.00	2.02	1.17
	3	2.99	3.01	1.48	2.98	3.01	3.39	3.01	3.02	1.68
	4	3.96	4.03	1.76	3.94	4.05	4.39	3.99	4.03	2.14
20	2	2.00	2.00	0.92	2.00	2.01	2.28	2.01	2.00	1.02
	3	3.00	3.01	1.24	2.99	3.01	3.30	3.02	3.01	1.41
	4	3.98	4.02	1.49	3.99	4.03	4.32	4.01	3.99	1.75
50	2	2.00	2.01	0.83	2.00	2.01	2.24	2.01	2.01	0.82
	3	2.98	3.01	1.13	2.97	3.01	3.23	3.03	3.01	1.14
	4	3.90	3.99	1.38	3.90	3.98	4.22	4.02	3.98	1.43
100	2	2.00	2.01	0.84	1.99	2.01	2.20	2.03	2.01	0.89
	3	2.97	3.02	1.15	2.97	3.02	3.20	3.05	3.02	1.25
	4	3.92	4.02	1.40	3.93	4.02	4.17	4.03	3.97	1.58
300	2	1.99	2.02	0.75	1.98	2.02	2.17	2.08	2.01	0.68
	3	2.94	3.03	1.05	2.93	3.03	3.14	3.11	3.03	0.97
	4	3.85	4.05	1.30	3.84	4.04	4.09	4.13	4.02	1.24
500	2	1.98	2.02	0.73	1.97	2.02	2.15	2.13	2.02	0.66
	3	2.93	3.02	1.03	2.92	3.02	3.12	3.17	3.03	0.93
	4	3.89	4.04	1.29	3.86	4.04	4.06	4.22	4.08	1.19

Table 5: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 1719$. The error term ε follows standardized Gamma$(10, 1)$.

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.02	1.07	2.00	2.02	2.36	2.00	2.02	1.17
	3	2.98	3.02	1.48	2.98	3.03	3.40	3.00	3.02	1.67
	4	3.93	4.01	1.76	3.91	4.00	4.41	4.01	4.02	2.13
20	2	2.00	2.03	0.93	2.00	2.02	2.27	2.01	2.03	1.03
	3	3.01	3.05	1.25	3.01	3.05	3.31	3.01	3.04	1.42
	4	3.97	4.06	1.49	3.98	4.06	4.32	4.02	4.06	1.77
50	2	2.00	2.03	0.84	2.00	2.03	2.24	2.01	2.02	0.83
	3	2.99	3.03	1.14	2.98	3.04	3.25	3.02	3.03	1.16
	4	3.96	4.05	1.38	3.95	4.05	4.27	3.99	4.04	1.46
100	2	2.00	2.04	0.84	1.99	2.04	2.20	2.02	2.04	0.89
	3	2.99	3.06	1.15	2.98	3.05	3.22	3.04	3.06	1.25
	4	3.96	4.11	1.41	3.95	4.10	4.21	4.02	4.09	1.59
300	2	1.99	2.03	0.75	1.99	2.03	2.18	2.05	2.04	0.69
	3	2.96	3.03	1.05	2.95	3.03	3.16	3.07	3.07	0.97
	4	3.93	4.04	1.31	3.91	4.05	4.17	4.12	4.07	1.24
500	2	1.99	2.04	0.73	1.98	2.03	2.16	2.07	2.04	0.66
	3	2.95	3.05	1.03	2.94	3.06	3.14	3.09	3.07	0.93
	4	3.89	4.07	1.29	3.87	4.04	4.12	4.11	4.11	1.19

Table 6: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 997$. The error term ε follows Bernoulli(1/2).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	1.99	1.05	2.00	2.00	2.34	2.00	2.00	1.16
	3	3.01	2.99	1.46	3.01	2.99	3.39	2.99	3.00	1.66
	4	4.00	4.02	1.74	3.99	4.00	4.42	3.98	4.00	2.11
20	2	2.01	2.00	0.93	2.01	2.00	2.29	2.00	2.00	1.03
	3	3.01	3.00	1.25	3.02	3.00	3.33	2.99	2.99	1.42
	4	4.04	4.01	1.50	4.04	4.01	4.39	3.93	3.91	1.76
50	2	2.01	2.00	0.83	2.01	2.00	2.26	2.00	2.00	0.82
	3	3.03	3.01	1.13	3.02	3.00	3.30	2.93	2.95	1.14
	4	4.05	4.00	1.38	4.05	4.00	4.33	3.72	3.73	1.43
100	2	2.02	2.00	0.84	2.01	2.00	2.22	1.98	2.00	0.88
	3	3.02	2.99	1.15	3.03	3.00	3.26	2.74	2.74	1.24
	4	4.04	4.00	1.41	4.04	4.01	4.30	3.56	3.40	1.56
300	2	2.02	2.00	0.74	2.03	2.00	2.20	1.79	1.75	0.67
	3	3.04	3.01	1.04	3.04	3.00	3.24	2.42	2.37	0.94
	4	4.04	3.98	1.30	4.04	3.98	4.27	2.72	2.60	1.18
500	2	2.03	2.00	0.73	2.03	2.00	2.20	1.58	1.63	0.63
	3	3.04	3.01	1.03	3.05	3.00	3.24	1.97	2.01	0.87
	4	4.07	4.00	1.30	4.07	3.99	4.29	2.18	2.27	1.09

Table 7: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 1719$. The error term ε follows Bernoulli(1/2).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.00	1.06	2.00	2.00	2.35	2.00	2.00	1.16
	3	3.01	3.00	1.46	3.01	3.00	3.41	3.00	3.00	1.66
	4	3.99	3.97	1.74	3.98	3.98	4.40	3.96	3.99	2.11
20	2	2.00	2.00	0.92	2.00	2.00	2.27	2.00	2.00	1.03
	3	3.01	3.01	1.24	3.01	3.00	3.32	2.99	2.99	1.42
	4	4.04	4.02	1.49	4.02	4.03	4.33	3.97	3.93	1.76
50	2	2.01	2.00	0.83	2.01	2.00	2.24	1.99	1.99	0.83
	3	3.02	3.01	1.13	3.01	3.00	3.27	2.92	2.91	1.16
	4	4.05	4.03	1.38	4.06	4.03	4.33	3.73	3.73	1.45
100	2	2.01	2.00	0.83	2.01	2.00	2.20	1.99	2.02	0.88
	3	3.02	3.01	1.14	3.02	3.01	3.24	2.81	2.87	1.24
	4	4.02	4.02	1.40	4.03	4.02	4.30	3.75	3.75	1.56
300	2	2.02	2.00	0.74	2.01	2.00	2.19	1.80	1.78	0.67
	3	3.02	2.99	1.04	3.02	2.99	3.21	2.47	2.46	0.94
	4	4.05	4.01	1.30	4.04	4.00	4.26	2.73	2.68	1.19
500	2	2.02	2.00	0.72	2.02	2.00	2.18	1.66	1.68	0.63
	3	3.03	3.01	1.02	3.04	3.01	3.23	2.18	2.20	0.88
	4	4.04	4.00	1.28	4.05	4.00	4.25	2.44	2.51	1.10

Table 8: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 997$. The error term ε follows Bernoulli(1/4).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.01	2.00	1.07	2.01	1.99	2.37	2.01	2.00	1.17
	3	3.00	2.99	1.48	2.99	3.00	3.41	3.01	3.00	1.68
	4	3.99	4.02	1.76	3.97	4.04	4.42	4.01	4.01	2.13
20	2	2.01	2.00	0.93	2.01	2.00	2.28	2.01	2.00	1.03
	3	3.01	3.00	1.25	3.01	3.00	3.31	3.02	3.00	1.42
	4	4.00	4.01	1.49	3.99	4.01	4.36	3.96	3.93	1.76
50	2	2.01	2.00	0.83	2.00	2.00	2.26	2.01	2.00	0.82
	3	3.00	3.01	1.13	2.99	3.00	3.28	2.99	2.96	1.14
	4	3.97	4.03	1.38	3.96	4.03	4.28	3.72	3.69	1.43
100	2	2.01	2.00	0.84	2.01	2.00	2.21	2.02	1.99	0.88
	3	3.00	3.00	1.16	2.99	2.99	3.22	2.93	2.88	1.24
	4	3.95	3.98	1.41	3.94	3.99	4.24	3.42	3.40	1.57
300	2	2.00	2.00	0.75	1.99	2.00	2.19	1.94	1.91	0.68
	3	2.98	2.99	1.05	2.97	2.99	3.20	2.39	2.38	0.96
	4	3.97	3.99	1.31	3.95	4.00	4.21	2.48	2.48	1.22
500	2	2.00	2.00	0.73	1.99	2.00	2.18	1.82	1.77	0.65
	3	2.97	3.00	1.04	2.96	3.00	3.17	2.08	2.06	0.91
	4	3.94	4.02	1.30	3.93	4.01	4.17	2.11	2.11	1.15

Table 9: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 1719$. The error term ε follows Bernoulli(1/4).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.00	1.07	2.00	1.99	2.37	2.00	2.00	1.17
	3	3.00	3.00	1.48	2.99	3.00	3.41	3.00	2.99	1.67
	4	3.96	3.97	1.76	3.94	3.97	4.42	4.03	3.99	2.13
20	2	2.00	2.00	0.93	2.00	2.00	2.27	2.00	2.00	1.03
	3	2.99	2.99	1.25	2.99	2.99	3.30	3.01	3.00	1.42
	4	4.00	3.99	1.49	3.99	3.98	4.33	4.02	3.99	1.76
50	2	2.00	2.00	0.84	2.00	2.00	2.25	2.01	2.01	0.83
	3	2.99	2.99	1.14	2.99	3.00	3.27	3.00	2.98	1.16
	4	3.96	3.98	1.38	3.95	3.98	4.29	3.90	3.89	1.46
100	2	2.01	2.00	0.84	2.00	2.00	2.21	2.01	2.00	0.89
	3	3.00	2.99	1.16	3.00	3.00	3.23	2.98	2.97	1.25
	4	3.97	3.98	1.41	3.97	3.99	4.25	3.78	3.74	1.58
300	2	2.00	2.00	0.75	2.00	2.00	2.19	2.02	1.98	0.68
	3	2.99	3.00	1.05	2.98	3.00	3.20	2.83	2.80	0.97
	4	3.99	4.00	1.31	3.98	4.01	4.21	3.16	3.15	1.23
500	2	2.00	2.00	0.73	2.00	2.00	2.18	1.93	1.94	0.65
	3	2.98	2.99	1.03	2.97	2.99	3.19	2.39	2.40	0.92
	4	3.98	3.98	1.30	3.95	3.98	4.20	2.48	2.48	1.16

Table 10: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 997$. The error term ε follows Bernoulli(1/8).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	1.99	1.99	1.07	1.99	1.99	2.34	2.00	2.00	1.17
	3	2.95	2.98	1.48	2.93	3.00	3.33	3.01	3.01	1.67
	4	3.88	4.02	1.75	3.83	4.03	4.27	3.99	3.98	2.13
20	2	2.00	2.00	0.92	1.99	2.00	2.26	2.01	2.00	1.03
	3	2.97	3.00	1.24	2.96	3.00	3.25	3.01	2.98	1.41
	4	3.89	3.99	1.49	3.86	3.98	4.21	4.02	3.98	1.75
50	2	1.98	2.00	0.83	1.97	2.01	2.21	2.01	2.00	0.82
	3	2.92	3.00	1.13	2.90	3.00	3.16	3.02	2.99	1.14
	4	3.79	4.03	1.37	3.77	4.03	4.08	3.95	3.91	1.43
100	2	1.97	1.99	0.84	1.96	2.00	2.15	2.03	2.00	0.88
	3	2.89	2.99	1.15	2.87	2.98	3.09	3.05	2.99	1.24
	4	3.77	4.00	1.39	3.75	4.01	4.01	4.00	3.92	1.58
300	2	1.94	2.00	0.74	1.91	2.00	2.10	2.06	2.06	0.68
	3	2.83	3.00	1.04	2.80	3.01	3.01	2.98	2.94	0.97
	4	3.69	4.02	1.29	3.66	4.01	3.89	3.47	3.46	1.24
500	2	1.92	2.00	0.72	1.89	2.01	2.06	2.05	1.97	0.65
	3	2.80	3.00	1.02	2.77	2.99	2.96	2.68	2.64	0.93
	4	3.65	4.01	1.27	3.64	4.01	3.83	2.83	2.82	1.18

Table 11: Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 1719$. The error term ε follows Bernoulli(1/8).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.00	1.07	2.00	2.00	2.36	2.00	2.00	1.17
	3	2.98	3.00	1.47	2.97	3.00	3.37	3.00	2.99	1.67
	4	3.93	4.01	1.76	3.89	4.01	4.36	4.00	4.01	2.13
20	2	2.00	2.00	0.92	2.00	2.00	2.26	2.00	2.00	1.03
	3	2.98	3.00	1.24	2.97	3.00	3.26	3.02	3.00	1.42
	4	3.93	3.99	1.49	3.93	4.00	4.25	3.96	3.95	1.76
50	2	1.99	2.00	0.84	1.98	2.00	2.22	2.01	2.00	0.83
	3	2.95	3.01	1.14	2.94	3.00	3.21	2.99	2.98	1.16
	4	3.88	4.01	1.38	3.86	4.01	4.15	3.80	3.78	1.46
100	2	1.98	2.00	0.84	1.97	2.00	2.17	2.01	2.03	0.89
	3	2.93	2.99	1.15	2.93	3.00	3.14	2.90	2.94	1.25
	4	3.84	3.97	1.40	3.84	3.97	4.06	3.39	3.39	1.58
300	2	1.96	2.00	0.75	1.95	2.00	2.13	1.96	1.94	0.68
	3	2.89	3.00	1.04	2.87	3.00	3.07	2.57	2.55	0.96
	4	3.81	4.02	1.29	3.77	4.02	4.00	2.72	2.71	1.22
500	2	1.95	2.00	0.73	1.93	2.00	2.11	1.84	1.84	0.64
	3	2.86	3.00	1.02	2.84	3.00	3.04	2.17	2.17	0.91
	4	3.76	3.99	1.28	3.73	4.00	3.94	2.22	2.22	1.15

References

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics 41*(6), 2786–2819.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Central limit theorems and bootstrap in high dimensions. *arXiv preprint arXiv:1412.3661*.

Chernozhukov, V., D. Chetverikov, and K. Kato (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields 162*(1-2), 47–70.

Talagrand, M. (2003). *Spin glasses: a challenge for mathematicians: cavity and mean field models*, Volume 46. Springer Science & Business Media.