A Note on the Finite-Exceedance Dependence Principle for Sum–Maximum Asymptotic Independence

Hongjian Shi

Abstract

Asymptotic independence between sum-type and maximum-type statistics arises repeatedly across many areas of probability and statistics, particularly in high-dimensional testing and random matrix theory. Motivated by the shared structure underlying existing proofs, which is obscured by their model-specific formulations, this note isolates a common intermediate principle. We introduce a *finite-exceedance dependence principle*: if the *finite-exceedance dependence error* vanishes at every fixed order and high-order exceedances are mildly controlled, then a finite-order Bonferroni argument yields asymptotic independence between the sum and the maximum. We further give an independent-proxy criterion for verifying this: for each finite set of exceedances, it suffices to construct an exactly independent proxy for the sum-type statistic whose approximation error is negligible in aggregate. Local deletion under dependency graphs and Gaussian regression and residualization are two examples. The framework gives modular derivations of existing sum–max independence results.

Keywords: sum–maximum asymptotic independence; high-dimensional testing; local dependence; Gaussian regression; Gaussian residualization.

1 Introduction

The joint asymptotic behavior of sums and maxima is a classical topic in statistics. The sum is typically driven by the aggregate contribution of many moderate observations, whereas the maximum is determined by a small number of extreme ones. It has long been known that, for independent observations, this separation of scales leads to asymptotic independence under suitable tail assumptions; see, e.g., Tiago de Oliveira (1962); Rosengard (1962); Rossberg (1965); Chow and Teugels (1979). Related results have also been developed for strongly mixing and weakly dependent stationary sequences and stationary Gaussian sequences (see, among others, Anderson and Turkman (1991, 1993, 1995); Hsing (1995a,b); Ho and Hsing (1996)), and, in related forms, for both discrete and continuous-time Gaussian processes (e.g., Ho and McCormick (1999); McCormick and Qi (2000)).

A closely related phenomenon has become increasingly important in high-dimensional statistics. Sum-type statistics are powerful against dense alternatives, while max-type statistics are powerful against sparse alternatives. Their asymptotic independence permits adaptive combination procedures. Examples include high-dimensional independence testing (Li and Xue, 2015), adaptive U-statistic tests (He, Xu, Wu, and Pan, 2021), one-sample mean problem (Chen, Feng, and Liang, 2024), finite-L_q rank blocks combined with maxima (Zhao, Wang, and Feng, 2026), and dependent Gaussian sum–max tests (Feng, Jiang, Li, and Liu, 2024). A point-process version of the same sum–maximum separation also appears in a recent work on random matrix theory (Heiny and Kleemann, 2023).

Although the final conclusions look similar and the proofs are often presented in problem-specific language, a recurring structure can be identified as follows.

(i) Write the max event as a union of rare exceedance events and apply Bonferroni inclusion–exclusion.

(ii) For a finite set of candidate extreme coordinates, construct a modified sum-type statistic that is *exactly independent* of those exceedance events.

(iii) Show that the modification is negligible on the scale of the sum-type statistic, both unconditionally and after regression on and summing over the rare exceedance configurations.

(iv) Use anti-concentration of the sum-type statistic to transfer the exact independence of the proxy back to the original statistic.

The purpose of this note is to isolate this argument at what we regard as a useful level of abstraction. We would like to characterize when and why the asymptotic factorization of every finite exceedance intersection occurs. The balanced formulation is therefore an *independent-proxy criterion* combined with a generic finite-exceedance Bonferroni-type lemma.

The core result has two layers. First, define the aggregated *finite-exceedance dependence error* as

$$D_{n,k}(y) = \sup_{x\in\mathbb{R}} \sum_{|I|=k} \left| \mathbb{P}\{T_n \le x, \mathscr{E}_{n,I}(y)\} - \mathbb{P}\{T_n \le x\}\mathbb{P}\{\mathscr{E}_{n,I}(y)\} \right|,$$

where T_n is a real-valued statistic of interest (in this paper, most often a sum-type statistic) and $\mathscr{E}_{n,I}(y)$ denotes the intersection of k exceedance events $\{X_{n,i} > u_n(y)\}$ indexed by $i \in I$. Theorem 2.2 then uses $D_{n,k}(y)$ to provide a bound for

$$\sup_{x\in\mathbb{R}} |\mathbb{P}\{T_n \le x, \mathscr{U}_n(y)\} - \mathbb{P}\{T_n \le x\}\mathbb{P}\{\mathscr{U}_n(y)\}|$$

where $\mathscr{U}_n(y) = \{\max_i X_{n,i} > u_n(y)\}$ is the exceedance event for the corresponding maximum. Thus, showing that $D_{n,k}(y) \to 0$ for each fixed k, together with a mild high-order binomial-moment condition, is sufficient for sum–max asymptotic independence. Second, Theorem 3.1 gives a reusable sufficient condition for $D_{n,k}(y) \to 0$ based on an exactly independent proxy. Therefore, the problem is reduced to constructing an appropriate independent proxy and verifying the conditions of Theorem 3.1.

When a sparse dependency graph can be formulated, the local-deletion argument in Section 4 is a natural way to construct an appropriate proxy. This mechanism is used in all examples in Section 5 except Section 5.2, where Gaussian regression and residualization provide an alternative construction.

2 Finite exceedances and the Bonferroni reduction

Let $d_n \to \infty$ as $n \to \infty$ and let $\mathscr{E}_{n,1}(y), \dots, \mathscr{E}_{n,d_n}(y)$ be exceedance events indexed by a fixed threshold parameter $y \in \mathbb{R}$. In applications, one typically has

$$\mathscr{E}_{n,i}(y) = \{X_{n,i} > u_n(y)\}, \tag{1}$$

where random variables $X_{n,1}, \dots, X_{n,d_n}$ are not necessarily i.i.d. The union of all exceedance events

$$\mathscr{U}_n(y) := \bigcup_{i=1}^{d_n} \mathscr{E}_{n,i}(y)$$

is exactly the event that the maximum $\max_{1\le i\le d_n} X_{n,i}$ exceeds the level $u_n(y)$ when the $\mathscr{E}_{n,i}(y)$ are exceedance events given in (1).

Let T_n be an arbitrary real-valued statistic and let

$$F_n(x) := \mathbb{P}(T_n \le x).$$

For $I \subseteq [\![d_n]\!] := \{1, \dots, d_n\}$, define

$$\mathscr{E}_{n,I}(y) = \bigcap_{i \in I} \mathscr{E}_{n,i}(y).$$

For $k \ge 1$, define the kth exceedance probability sum

$$H_{n,k}(y) = \sum_{|I|=k} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} \equiv \sum_{I \subseteq [\![d_n]\!], |I|=k} \mathbb{P}\{\mathscr{E}_{n,I}(y)\}.$$

Let $N_n(y) = \sum_{i=1}^{d_n} \mathbb{1}\{\mathscr{E}_{n,i}(y)\}$ denote the exceedance count. Then $H_{n,k}(y)$ is the k-th binomial moment of $N_n(y)$ (Sibuya, 1991)

$$H_{n,k}(y) = \mathbb{E}\binom{N_n(y)}{k} = \frac{1}{k!}\mathbb{E}\big[(N_n(y))_k\big],$$

where $(a)_k := a(a-1)\cdots(a-k+1)$ is the falling factorial.

The central quantity in this note is the following.

Definition 2.1 (Finite-exceedance dependence error)**.** For $k \ge 1$, define

$$D_{n,k}(y) = \sup_{x \in \mathbb{R}} \sum_{|I|=k} \Big| \mathbb{P}\{T_n \le x, \mathscr{E}_{n,I}(y)\} - \mathbb{P}\{T_n \le x\}\mathbb{P}\{\mathscr{E}_{n,I}(y)\} \Big|.$$

The following theorem establishes asymptotic independence under $D_{n,k}(y) \to 0$ for each fixed k, together with a mild condition on the higher-order binomial moment $H_{n,2r}(y)$. We defer discussion for proving $D_{n,k}(y)$ small to Theorem 3.1.

Theorem 2.2 (Finite-exceedance dependence principle)**.** *For every integer $r \ge 1$ and every $y \in \mathbb{R}$,*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}\{T_n \le x, \mathscr{U}_n(y)\} - \mathbb{P}\{T_n \le x\}\mathbb{P}\{\mathscr{U}_n(y)\}| \le \sum_{k=1}^{2r} D_{n,k}(y) + H_{n,2r}(y). \tag{2}$$

Consequently, if for every fixed $k \ge 1$,

$$D_{n,k}(y) \longrightarrow 0, \tag{3}$$

and

$$\lim_{r \to \infty} \limsup_{n \to \infty} H_{n,2r}(y) = 0, \tag{4}$$

then

$$\sup_{x \in \mathbb{R}} |\mathbb{P}\{T_n \le x, \mathscr{U}_n(y)\} - \mathbb{P}\{T_n \le x\}\mathbb{P}\{\mathscr{U}_n(y)\}| \longrightarrow 0. \tag{5}$$

The same conclusion holds with $\mathscr{U}_n(y)$ replaced by its complement.

Proof. For $k \geq 1$, define

$$J_{n,k}(x,y) = \sum_{\substack{I \subseteq [\![d_n]\!] \\ |I|=k}} \mathbb{P}\{T_n \leq x, \mathscr{E}_{n,I}(y)\}.$$

Bonferroni's inequalities imply

$$\sum_{k=1}^{2r} (-1)^{k-1} J_{n,k}(x,y) \leq \mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)\} \leq \sum_{k=1}^{2r-1} (-1)^{k-1} J_{n,k}(x,y), \tag{6}$$

and

$$\sum_{k=1}^{2r} (-1)^{k-1} H_{n,k}(y) \leq \mathbb{P}\{\mathscr{U}_n(y)\} \leq \sum_{k=1}^{2r-1} (-1)^{k-1} H_{n,k}(y). \tag{7}$$

The difference between the odd and even Bonferroni truncations in (7) is exactly $H_{n,2r}(y)$.

By definition of $D_{n,k}(y)$,

$$\left| J_{n,k}(x,y) - F_n(x) H_{n,k}(y) \right| \leq D_{n,k}(y) \qquad \text{uniformly in } x.$$

Combining this inequality with the upper bound in (6) and the lower bound in (7) gives

$$\mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)\} - \mathbb{P}\{T_n \leq x\}\mathbb{P}\{\mathscr{U}_n(y)\} \leq \sum_{k=1}^{2r-1} D_{n,k}(y) + H_{n,2r}(y). \tag{8}$$

Similarly, combining the lower bound in (6) and the upper bound in (7) yields

$$\mathbb{P}\{T_n \leq x\}\mathbb{P}\{\mathscr{U}_n(y)\} - \mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)\} \leq \sum_{k=1}^{2r} D_{n,k}(y) + H_{n,2r}(y). \tag{9}$$

Taking the supremum on both (8) and (9) over x proves (2).

For fixed r, assumption (3) makes the finite sum on the right-hand side vanish as $n \to \infty$. Taking $\limsup_n$ in (2) and then sending $r \to \infty$ proves (5). The statement for $\mathscr{U}_n(y)^c$ follows from

$$\mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)^c\} - \mathbb{P}\{T_n \leq x\}\mathbb{P}\{\mathscr{U}_n(y)^c\} = -\Big[\mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)\} - \mathbb{P}\{T_n \leq x\}\mathbb{P}\{\mathscr{U}_n(y)\}\Big].$$

□

Remark 2.3 (When a fixed Bonferroni order is sufficient—and when it is not)**.** When the goal is to establish asymptotic independence, Theorem 2.2 permits the order r to be fixed while $n \to \infty$, followed only afterwards by $r \to \infty$. This is often substantially simpler than choosing a truncation order that grows with dimension. However, it is worth emphasizing that the Bonferroni bound given in (2) is non-asymptotic. Choosing a growing order r may still be important for obtaining an explicit uniform convergence rate, but it is not necessary to the asymptotic independence itself.

Theorem 2.2 establishes asymptotic independence but does not, by itself, guarantee the existence of a joint limiting distribution. The following corollary fills this gap by providing a more self-contained approach to establishing both the marginal limit of the maximum and its asymptotic independence from the sum.

Corollary 2.4 (Poisson criterion). *Suppose that for every fixed $k \geq 1$ and $y \in \mathbb{R}$,*

$$H_{n,k}(y) \longrightarrow \{\lambda(y)\}^k \Big/ k! \tag{10}$$

for some $0 \leq \lambda(y) < \infty$, and that (3) holds. Then

$$\mathbb{P}\{\mathscr{U}_n(y)^c\} \longrightarrow e^{-\lambda(y)} \tag{11}$$

and

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)^c\} - F_n(x) e^{-\lambda(y)} \right| \longrightarrow 0. \tag{12}$$

If in addition $T_n \Rightarrow T$ with continuous distribution function F, then for every $x, y \in \mathbb{R}$,

$$\mathbb{P}\{T_n \leq x, \mathscr{U}_n(y)^c\} \longrightarrow F(x) e^{-\lambda(y)}. \tag{13}$$

Proof. Equation (10) implies (4). The Bonferroni argument applied only to $\mathscr{U}_n(y)$ yields $\mathbb{P}\{\mathscr{U}_n(y)^c\} \to e^{-\lambda(y)}$, while Theorem 2.2 gives the factorization (12). If $F_n \to F$ weakly and F is continuous, then the convergence of distribution functions is everywhere. This fact together with (12) concludes (13). □

3 An independent-proxy criterion

Theorem 2.2 isolates the exact quantity to be controlled. We now give a general mechanism for proving this vanishing. This together with Theorem 2.2 yields the main reusable criterion of the note.

Theorem 3.1 (Independent-proxy criterion). *Fix $k \geq 1$ and $y \in \mathbb{R}$. Assume that, for every $I \subseteq [\![d_n]\!]$ with $|I| = k$, a proxy statistic $T^\circ_{n,I}$, which may depend on I and need not have the same distribution for different I, can be constructed such that*

$$T^\circ_{n,I} \perp\!\!\!\perp \mathscr{E}_{n,I}(y).$$

For $\eta > 0$, define

$$A_{n,k}(\eta, y) = \sum_{|I|=k} \mathbb{P}\{|T_n - T^\circ_{n,I}| > \eta\} \, \mathbb{P}\{\mathscr{E}_{n,I}(y)\},$$
$$B_{n,k}(\eta, y) = \sum_{|I|=k} \mathbb{P}\{|T_n - T^\circ_{n,I}| > \eta, \mathscr{E}_{n,I}(y)\},$$

and the Lévy concentration function

$$\omega_n(\eta) = \sup_{x \in \mathbb{R}} \{F_n(x+\eta) - F_n(x-\eta)\}.$$

Then, for every $\eta > 0$,

$$D_{n,k}(y) \leq \underbrace{A_{n,k}(\eta, y)}_{\substack{\text{weighted unconditional} \\ \text{proxy error}}} + \underbrace{B_{n,k}(\eta, y)}_{\substack{\text{weighted conditional} \\ \text{proxy error}}} + \underbrace{\omega_n(2\eta) \times H_{n,k}(y)}_{\text{anti-concentration}}. \tag{14}$$

Consequently, if there exists $\eta_n \to 0$ such that

$$A_{n,k}(\eta_n, y) \to 0, \qquad B_{n,k}(\eta_n, y) \to 0, \qquad \omega_n(2\eta_n) \to 0, \qquad H_{n,k}(y) = O(1), \tag{15}$$

then $D_{n,k}(y) \to 0$.

Proof. Fix $I \subseteq [\![d_n]\!]$ with $|I| = k$. Abbreviate $\mathscr{E}_I = \mathscr{E}_{n,I}(y)$ and $T_I^\circ = T_{n,I}^\circ$. Let

$$\alpha_I(\eta) = \mathbb{P}(|T_n - T_I^\circ| > \eta), \qquad \beta_I(\eta) = \mathbb{P}(|T_n - T_I^\circ| > \eta, \mathscr{E}_I), \qquad p_I = \mathbb{P}(\mathscr{E}_I).$$

It holds that

$$T_n \le x, \quad |T_n - T_I^\circ| \le \eta \quad \Longrightarrow \quad T_I^\circ \le x + \eta.$$

Hence, using this triangle inequality,

$$\begin{aligned} \mathbb{P}(T_n \le x, \mathscr{E}_I) &\le \mathbb{P}(T_I^\circ \le x + \eta, \mathscr{E}_I) + \mathbb{P}(|T_n - T_I^\circ| > \eta, \mathscr{E}_I) \\ &= \mathbb{P}(T_I^\circ \le x + \eta)\, p_I + \beta_I(\eta), \end{aligned} \tag{16}$$

where the last step is by the exact independence $T_{n,I}^\circ \perp\!\!\!\perp \mathscr{E}_{n,I}(y)$. Using the triangle inequality once again gives

$$\begin{aligned} \mathbb{P}(T_I^\circ \le x + \eta) &\le \mathbb{P}(T_n \le x + 2\eta) + \mathbb{P}(|T_n - T_I^\circ| > \eta) \\ &= F_n(x + 2\eta) + \alpha_I(\eta). \end{aligned} \tag{17}$$

Combining (16) and (17) shows

$$\mathbb{P}(T_n \le x, \mathscr{E}_I) - p_I F_n(x) \le \omega_n(2\eta)\, p_I + \alpha_I(\eta)\, p_I + \beta_I(\eta). \tag{18}$$

Similarly, the implication

$$T_I^\circ \le x - \eta, \quad |T_n - T_I^\circ| \le \eta \quad \Longrightarrow \quad T_n \le x$$

yields

$$\mathbb{P}(T_n \le x, \mathscr{E}_I) \ge \mathbb{P}(T_I^\circ \le x - \eta)\, p_I - \beta_I(\eta),$$

and

$$\mathbb{P}(T_I^\circ \le x - \eta) \ge F_n(x - 2\eta) - \alpha_I(\eta).$$

Therefore

$$p_I F_n(x) - \mathbb{P}(T_n \le x, \mathscr{E}_I) \le \omega_n(2\eta)\, p_I + \alpha_I(\eta)\, p_I + \beta_I(\eta). \tag{19}$$

Combining (18) and (19), summing over $|I| = k$, and then taking the supremum over x proves (14). □

Remark 3.2 (The two approximation errors are genuinely different)**.** The term $A_{n,k}$ is the *weighted unconditional proxy error*: it weights the unconditional probability that the proxy approximation fails by the probability of the corresponding rare configuration. The term $B_{n,k}$ is the *weighted conditional proxy error*:

$$B_{n,k}(\eta, y) = \sum_{|I|=k} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta \mid \mathscr{E}_{n,I}(y)\}\, \mathbb{P}\{\mathscr{E}_{n,I}(y)\},$$

which weights the conditional proxy-failure probability given the rare configuration by the probability of that configuration. The latter is the quantity that a purely unconditional $o_{\mathbb{P}}(1)$ statement can miss. In many applications both terms are controlled by an extremely strong uniform tail estimate, but the distinction is useful in more delicate problems.

Corollary 3.3 (Uniform proxy criterion). *Define the following uniform bound for* $T_n - T^\circ_{n,I}$

$$\rho_{n,k}(\eta) := \sup_{|I|=k} \mathbb{P}(|T_n - T^\circ_{n,I}| > \eta). \tag{20}$$

Then, for every $\eta > 0$,

$$D_{n,k}(y) \le \rho_{n,k}(\eta) H_{n,k}(y) + \rho_{n,k}(\eta) \binom{d_n}{k} + \omega_n(2\eta) \times H_{n,k}(y) \tag{21}$$

If there exists $\eta_n \to 0$ *such that*

$$\rho_{n,k}(\eta_n) \binom{d_n}{k} \to 0, \qquad \omega_n(2\eta_n) \to 0, \qquad H_{n,k}(y) = O(1), \tag{22}$$

then $D_{n,k}(y) \to 0$.

Proof. In view of (14), to prove (21), it suffices to bound $A_{n,k}(\eta, y)$ and $B_{n,k}(\eta, y)$ in terms of $\rho_{n,k}(\eta)$ and $H_{n,k}(y)$:

$$\begin{aligned} A_{n,k}(\eta, y) &= \sum_{|I|=k} \mathbb{P}\{|T_n - T^\circ_{n,I}| > \eta\} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} \\ &\le \rho_{n,k}(\eta) \sum_{|I|=k} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} = \rho_{n,k}(\eta) H_{n,k}(y), \end{aligned} \tag{23}$$

and

$$\begin{aligned} B_{n,k}(\eta, y) &= \sum_{|I|=k} \mathbb{P}\{|T_n - T^\circ_{n,I}| > \eta, \mathscr{E}_{n,I}(y)\} \\ &\le \sum_{|I|=k} \mathbb{P}\{|T_n - T^\circ_{n,I}| > \eta\} \le \rho_{n,k}(\eta) \binom{d_n}{k}. \end{aligned} \tag{24}$$

Notice that the upper bound in (24) also bounds $A_{n,k}(\eta_n, y)$, so (22) implies that the right-hand side of (21) tends to 0 taking $\eta = \eta_n$. □

Although the upper bound in (24) also bounds $A_{n,k}(\eta_n, y)$, it is typically much cruder than (23). Therefore, one needs either a strong enough bound for $\rho_{n,k}$ to pay the cost of factor $\binom{d_n}{k}$, as in (22) and in the proofs of Sections 5.3 and 5.4, or a more delicate estimate for the weighted conditional proxy error.

Corollary 3.4 (L^2 proxy criterion). *Assuming* $T^\circ_{n,I} \perp\!\!\!\perp \mathscr{E}_{n,I}(y)$, *suppose for fixed* $k \ge 1$ *and* $y \in \mathbb{R}$ *that*

$$\sup_{|I|=k} \mathbb{E}\big(T_n - T^\circ_{n,I}\big)^2 \to 0, \tag{25}$$

$$\sum_{|I|=k} \mathbb{E}\Big[\big(T_n - T^\circ_{n,I}\big)^2 \mathbb{1}\{\mathscr{E}_{n,I}(y)\}\Big] \to 0, \tag{26}$$

$$H_{n,k}(y) = O(1). \tag{27}$$

The first condition controls the weighted unconditional proxy error, while the second controls the weighted conditional proxy error. If $T_n \Rightarrow T$ *with a continuous distribution, then* $D_{n,k}(y) \to 0$.

Proof. Let a_n^{ord} and a_n^{rare} denote the two quantities in (25) and (26) and let a_n be the maximum of them. Set $\eta_n = a_n^{1/4}$ when $a_n > 0$ and $\eta_n = n^{-1}$ when $a_n = 0$. Markov's inequality gives

$$A_{n,k}(\eta_n, y) \le H_{n,k}(y)\frac{a_n^{\mathrm{ord}}}{\eta_n^2} = O(a_n^{1/2})$$

and

$$B_{n,k}(\eta_n, y) \le \frac{a_n^{\mathrm{rare}}}{\eta_n^2} \le a_n^{1/2}.$$

If $T_n \Rightarrow T$ and T has a continuous distribution function F, then Pólya's theorem yields

$$\|F_n - F\|_\infty \longrightarrow 0.$$

Since every continuous distribution function is uniformly continuous, for every deterministic $\eta_n \to 0$,

$$\omega_n(2\eta_n) \le 2\|F_n - F\|_\infty + \sup_{x\in\mathbb{R}}\{F(x+2\eta_n) - F(x-2\eta_n)\} \longrightarrow 0. \tag{28}$$

Thus Theorem 3.1 applies. □

Proposition 3.5 (Multivariate $\boldsymbol{T}_n$ version)**.** *For fixed K, Theorem 3.1 extends verbatim to $\boldsymbol{T}_n \in \mathbb{R}^K$. For $\eta > 0$, define*

$$\begin{aligned}
A_{n,k}(\eta, y) &= \sum_{|I|=k} \mathbb{P}\{\|\boldsymbol{T}_n - \boldsymbol{T}_{n,I}^\circ\|_\infty > \eta\}\,\mathbb{P}\{\mathscr{E}_{n,I}(y)\},\\
B_{n,k}(\eta, y) &= \sum_{|I|=k} \mathbb{P}\{\|\boldsymbol{T}_n - \boldsymbol{T}_{n,I}^\circ\|_\infty > \eta, \mathscr{E}_{n,I}(y)\},\\
\omega_n(\eta) &= \sup_{\boldsymbol{x}\in\mathbb{R}^K}\{F_n(\boldsymbol{x}+\eta\mathbf{1}) - F_n(\boldsymbol{x}-\eta\mathbf{1})\},
\end{aligned}$$

where $F_n(\boldsymbol{x}) := \mathbb{P}(\boldsymbol{T}_n \le \boldsymbol{x})$. Then, for every $\eta > 0$,

$$\begin{aligned}
D_{n,k}(y) &= \sup_{\boldsymbol{x}\in\mathbb{R}^K} \sum_{|I|=k} \left|\mathbb{P}\{\boldsymbol{T}_n \le \boldsymbol{x}, \mathscr{E}_{n,I}(y)\} - \mathbb{P}\{\boldsymbol{T}_n \le \boldsymbol{x}\}\mathbb{P}\{\mathscr{E}_{n,I}(y)\}\right|\\
&\le A_{n,k}(\eta, y) + B_{n,k}(\eta, y) + \omega_n(2\eta) \times H_{n,k}(y).
\end{aligned}$$

Theorem 2.2 and Corollaries 3.3 and 3.4 also extend verbatim. Their proofs are unchanged. The multivariate version of Corollary 3.3 is used in Section 5.4.

4 Local deletion

The independent-proxy formulation contains local deletion as an immediate special case. Let

$$\{X_{n,v} : v \in \mathscr{V}_n\}$$

be a triangular array with a dependency graph $G_n = (\mathscr{V}_n, E_n)$. Recall that G_n is a dependency graph if whenever two disjoint vertex sets $A, B \subseteq \mathscr{V}_n$ have no edge between them, the families $\{X_{n,v} : v \in A\}$ and $\{X_{n,v} : v \in B\}$ are independent. Dependency graphs are classical tools for normal approximation of locally dependent sums; see Janson (1988); Baldi and Rinott (1989). Consider

$$T_n = \frac{1}{\sigma_n}\sum_{v\in\mathscr{V}_n}\{X_{n,v} - \mathbb{E}X_{n,v}\}, \qquad \text{where } \sigma_n > 0 \text{ is a normalizing factor.}$$

For $I \subseteq \mathcal{V}_n$, let $\Gamma_n[I]$ denote the closed graph neighborhood of I, i.e., the vertices in I together with every vertex adjacent to at least one vertex of I. Suppose $\mathcal{E}_{n,v}(y)$ is measurable with respect to $X_{n,v}$ and set $\mathcal{E}_{n,I}(y) = \bigcap_{v\in I} \mathcal{E}_{n,v}(y)$. Define the proxy statistic

$$T^{\circ}_{n,I} = \frac{1}{\sigma_n} \sum_{v \notin \Gamma_n[I]} \{X_{n,v} - \mathbb{E}X_{n,v}\} \tag{29}$$

by the "local deletion". There is no edge from I to $\mathcal{V}_n \setminus \Gamma_n[I]$, hence $T^{\circ}_{n,I} \perp\!\!\!\perp \mathcal{E}_{n,I}(y)$. Thus Theorem 3.1 gives the following.

Corollary 4.1 (Local-deletion construction)**.** *Fix k and y. Suppose that $T^{\circ}_{n,I} \perp\!\!\!\perp \mathcal{E}_{n,I}(y)$ and define*

$$R_{n,I} := T_n - T^{\circ}_{n,I} = \frac{1}{\sigma_n} \sum_{v \in \Gamma_n[I]} \{X_{n,v} - \mathbb{E}X_{n,v}\}. \tag{30}$$

If $H_{n,k}(y) = O(1)$, $T_n \Rightarrow T$ with continuous distribution, and there exists $\eta_n \to 0$ such that

$$\sum_{|I|=k} \mathbb{P}(|R_{n,I}| > \eta_n)\mathbb{P}\{\mathcal{E}_{n,I}(y)\} \to 0, \tag{31}$$

$$\sum_{|I|=k} \mathbb{P}\{|R_{n,I}| > \eta_n, \mathcal{E}_{n,I}(y)\} \to 0, \tag{32}$$

then $D_{n,k}(y) \to 0$. If these conditions hold for every fixed k and $\lim_{r\to\infty} \limsup_{n\to\infty} H_{n,2r}(y) = 0$ holds, then T_n and the max event $\mathcal{U}_n(y)$ are asymptotically independent.

Remark 4.2 (What a small dependency degree does and does not imply)**.** Let $N_n = |\mathcal{V}_n|$ and let D_n be the maximum graph degree. For fixed k, $|\Gamma_n[I]| \le k(D_n + 1)$. Thus $D_n = o(N_n)$ says that finitely many extreme coordinates contaminate only a vanishing fraction of the summands. This is a useful structural heuristic, but it is not by itself sufficient: the deleted neighborhood may carry a non-negligible fraction of the total variance or may become large precisely on the extreme event. Conditions (31)–(32) express the missing "local energy" requirement.

5 Examples and applications

5.1 The i.i.d. Gaussian sum and maximum

Let $Z, Z_1, \ldots, Z_n$ be i.i.d. $N(0,1)$ and define

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i, \qquad M_n = \max_{1\le i\le n} Z_i.$$

Set $\mathcal{E}_{n,i}(y) = \{Z_i > u_n(y)\}$ where $u_n(y)$ is chosen to be the Gaussian extreme-value threshold

$$u_n(y) = \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}} + \frac{y}{\sqrt{2\log n}}$$

satisfying

$$q_n(y) := \mathbb{P}(Z > u_n(y)) = \overline{\Phi}(u_n(y)) \sim \frac{e^{-y}}{n}, \qquad n\overline{\Phi}(u_n(y)) \to e^{-y},$$

with $\overline{\Phi}$ the standard normal survival function.

For $I \subseteq [\![n]\!]$, define the proxy statistic and its remainder

$$T^{\circ}_{n,I} = \frac{1}{\sqrt{n}} \sum_{i \notin I} Z_i, \qquad R_{n,I} := T_n - T^{\circ}_{n,I} = \frac{1}{\sqrt{n}} \sum_{i \in I} Z_i.$$

This is a special case of the local-deletion mechanism of Section 4 when the graph is edgeless and $\Gamma_n[I] = I$. Then $T^{\circ}_{n,I} \perp\!\!\!\perp \mathscr{E}_{n,I}(y)$.

Fix $k \geq 1$ and $y \in \mathbb{R}$. In this example, the proxy criterion can be verified particularly simply through conditions (25)–(27) of Corollary 3.4.

We first verify (25). Since for any $I \subseteq [\![n]\!]$ with $|I| = k$,

$$\mathbb{E} R^2_{n,I} = \frac{k}{n},$$

we have

$$\sup_{|I|=k} \mathbb{E} R^2_{n,I} = \frac{k}{n} \longrightarrow 0. \tag{33}$$

Now we verify (26). Using the Cauchy inequality $(\sum_{i \in I} Z_i)^2 \leq k \sum_{i \in I} Z_i^2$ and independence,

$$\begin{aligned} \mathbb{E}\big[R^2_{n,I} \mathbb{1}\{\mathscr{E}_{n,I}(y)\}\big] &\leq \frac{k}{n} \sum_{i \in I} \mathbb{E}\Big[Z_i^2 \mathbb{1}\{Z_i > u_n(y)\} \prod_{j \in I \setminus \{i\}} \mathbb{1}\{Z_j > u_n(y)\}\Big] \\ &= \frac{k^2}{n} \mathbb{E}\big[Z^2 \mathbb{1}\{Z > u_n(y)\}\big] q_n(y)^{k-1}, \end{aligned}$$

where $Z \sim N(0,1)$. Hence

$$\sum_{|I|=k} \mathbb{E}\big[R^2_{n,I} \mathbb{1}\{\mathscr{E}_{n,I}(y)\}\big] \leq \binom{n}{k} \frac{k^2}{n} \mathbb{E}\big[Z^2 \mathbb{1}\{Z > u_n(y)\}\big] q_n(y)^{k-1}. \tag{34}$$

For a standard Gaussian variable Z,

$$\mathbb{E}\big[Z^2 \mathbb{1}\{Z > u\}\big] = u\phi(u) + \overline{\Phi}(u) \sim u^2 \overline{\Phi}(u), \qquad u \to \infty.$$

Since $u_n(y)^2 \sim 2 \log n$ and $n q_n(y) \to e^{-y}$,

$$\mathbb{E}\big[Z^2 \mathbb{1}\{Z > u_n(y)\}\big] q_n(y)^{k-1} \sim u_n(y)^2 q_n(y)^k.$$

Thus the right-hand side of (34) is on the order of

$$\binom{n}{k} \frac{k^2}{n} (\log n) \left(\frac{e^{-y}}{n}\right)^k \leq \frac{\log n}{n} \frac{k^2 e^{-yk}}{k!}$$

and therefore

$$\sum_{|I|=k} \mathbb{E}\big[R^2_{n,I} \mathbb{1}\{\mathscr{E}_{n,I}(y)\}\big] \longrightarrow 0. \tag{35}$$

Last, (27) holds because for fixed k,

$$H_{n,k}(y) = \binom{n}{k} q_n(y)^k \longrightarrow \frac{e^{-ky}}{k!}. \tag{36}$$

Combining (33), (35), and (36), Corollary 3.4 gives

$$D_{n,k}(y) \longrightarrow 0 \qquad \text{for every fixed } k,$$

that is, the finite-order condition (3) in Theorem 2.2 holds. Since also for each fixed r,

$$H_{n,2r}(y) \longrightarrow \frac{e^{-2ry}}{(2r)!}$$

the high-order condition (4) in Theorem 2.2 follows by letting $r \to \infty$. The theorem therefore recovers the classical asymptotic independence between the normalized Gaussian sum T_n and maximum M_n. Corollary 2.4 further gives for every $x, y \in \mathbb{R}$,

$$\mathbb{P}\{T_n \le x, M_n \le u_n(y)\} \longrightarrow \Phi(x)\exp\{-e^{-y}\}.$$

This example isolates the mechanism in its simplest form: delete the finitely many coordinates associated with an extreme configuration, obtain exact independence for the deleted proxy, and use an L^2-negligibility calculation to transfer that independence back to the original sum statistic.

5.2 Dependent Gaussian sum and maximum

Let $Z_{n,1}, \dots, Z_{n,n}$ again be jointly Gaussian with $N(0,1)$ marginals. Unlike in Section 5.1, however, we now allow them to be *dependent*. The dependency-graph specialization introduced in Section 4 and used in Section 5.1 is no longer suitable, since nonzero correlations may occur between all pairs of variables. Nevertheless, Theorem 3.1 remains applicable when the proxy is constructed via Gaussian regression and residualization rather than local deletion.

Let $\mathbf{Z}_n = (Z_{n,1}, \dots, Z_{n,n})^\top \sim N(0, \Sigma_n)$, where $\Sigma_n = (\sigma_{ij})_{1 \le i,j \le n}$. Denoting $S_n = \sum_{i=1}^n Z_{n,i}$ and $\sigma_n^2 = \mathrm{Var}(S_n)$, we consider the sum and maximum

$$T_n = S_n/\sigma_n, \qquad M_n = \max_{1 \le i \le n} Z_{n,i}.$$

Set once again the classical Gaussian extreme-value threshold

$$u_n(y) = \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}} + \frac{y}{\sqrt{2\log n}}$$

and $\mathcal{E}_{n,i}(y) = \{Z_{n,i} > u_n(y)\}$. We show that, for every $x, y \in \mathbb{R}$,

$$\mathbb{P}\{T_n \le x,\ M_n \le u_n(y)\} \longrightarrow \Phi(x)\exp\{-e^{-y}\}; \tag{37}$$

that is, T_n and M_n are asymptotically independent, with exactly the same joint limit as in Section 5.1.

To facilitate our analysis, we need to impose certain assumptions on the data generation process. Suppose that

$$\Sigma_n \succeq 0, \qquad \sigma_{ii} = 1, \qquad \mathbf{1}_n^\top \Sigma_n \mathbf{1}_n > 0,$$

which is assumed only for standardization and, in particular, $\mathbf{1}_n^\top \Sigma_n \mathbf{1}_n > 0$ is needed only to define T_n; also suppose that, for fixed $\varrho \in (0,1)$ and $\xi > 0$,

$$\max_{i \ne j} |\sigma_{ij}| \le \varrho, \qquad \max_i \sum_{j=1}^n \sigma_{ij}^2 \le (\log n)^\xi. \tag{38}$$

No eigenvalue or condition-number assumption is imposed.

Nearly uniform proxy criterion. We follow the approach of Corollary 3.3. Under the present assumptions, however, establishing a uniform bound of the form

$$\rho_{n,k}(\eta) := \sup_{|I|=k} \mathbb{P}(|T_n - T_{n,I}^\circ| > \eta)$$

may be difficult. We therefore partition the sets I into two classes, termed *good* and *bad*. Their precise definitions will be given below; for now, the key point is that the good sets are well behaved and make the dominant contribution.

We write

$$\begin{aligned} H_{n,k}(y) &= \sum_{|I|=k} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} \\ &= \sum_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} + \sum_{\substack{|I|=k \\ I \text{ bad}}} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} =: H_{n,k}^{\text{good}}(y) + H_{n,k}^{\text{bad}}(y), \end{aligned}$$

and we define

$$\rho_{n,k}^{\text{good}}(\eta_n) := \sup_{|I|=k,\ I \text{ good}} \mathbb{P}(|T_n - T_{n,I}^\circ| > \eta_n).$$

Using the definitions of the two proxy errors $A_{n,k}(\eta_n, y)$ and $B_{n,k}(\eta_n, y)$ in Theorem 3.1, we have

$$\begin{aligned} A_{n,k}(\eta_n, y) &= \sum_{|I|=k} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n\} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} \\ &= \sum_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n\} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} + \sum_{\substack{|I|=k \\ I \text{ bad}}} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n\} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} \\ &\le \rho_{n,k}^{\text{good}}(\eta_n) \sum_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} + \sum_{\substack{|I|=k \\ I \text{ bad}}} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} \\ &\le \rho_{n,k}^{\text{good}}(\eta_n) H_{n,k}^{\text{good}}(y) + H_{n,k}^{\text{bad}}(y), \end{aligned}$$

and

$$\begin{aligned} B_{n,k}(\eta_n, y) &= \sum_{|I|=k} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n, \mathcal{E}_{n,I}(y)\} \\ &= \sum_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n, \mathcal{E}_{n,I}(y)\} + \sum_{\substack{|I|=k \\ I \text{ bad}}} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n, \mathcal{E}_{n,I}(y)\} \\ &\le \sum_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}\{|T_n - T_{n,I}^\circ| > \eta_n\} + \sum_{\substack{|I|=k \\ I \text{ bad}}} \mathbb{P}\{\mathcal{E}_{n,I}(y)\} \\ &\le \rho_{n,k}^{\text{good}}(\eta_n) \binom{d_n}{k} + H_{n,k}^{\text{bad}}(y). \end{aligned}$$

Therefore, if one can find proxy $T_{n,I}^\circ \perp\!\!\!\perp \mathcal{E}_{n,I}(y)$ and prove for some $\eta_n \to 0$ that

$$\rho_{n,k}^{\text{good}}(\eta_n) \binom{d_n}{k} \to 0, \qquad \omega_n(2\eta_n) \to 0, \qquad H_{n,k}^{\text{good}}(y) = O(1), \qquad H_{n,k}^{\text{bad}}(y) \to 0, \tag{39}$$

then Theorem 3.1 gives $D_{n,k}(y) \to 0$.

Gaussian regression and residualization. Fix $k \geq 1$ and $y \in \mathbb{R}$. For every I with $|I| = k$, define the sum proxy and its remainder by

$$T^{\circ}_{n,I} := T_n - \mathbb{E}(T_n \mid Z_{n,I}), \qquad R_{n,I} := T_n - T^{\circ}_{n,I}. \tag{40}$$

This decomposition remains valid even when $\Sigma_{n,I}$ is singular, since the Gaussian conditional expectation is well defined. Let $I = \{i_1, \ldots, i_k\}$ and denote

$$Z_{n,I} = (Z_{n,i_1}, \ldots, Z_{n,i_k})^{\top}, \qquad \Sigma_{n,I} = \mathrm{Var}(Z_{n,I}) = (\Sigma_n)_{I,I},$$

and define the regression coefficient vector

$$r_{n,I} := \mathrm{Cov}(Z_{n,I}, T_n) = \frac{(\Sigma_n \mathbf{1}_n)_I}{\sigma_n}.$$

The decomposition in (40) can be explicitly written as

$$T^{\circ}_{n,I} = T_n - r^{\top}_{n,I} \Sigma^{-1}_{n,I} Z_{n,I}, \qquad R_{n,I} = r^{\top}_{n,I} \Sigma^{-1}_{n,I} Z_{n,I} \tag{41}$$

when $\Sigma_{n,I}$ is nonsingular, with the inverse $\Sigma^{-1}_{n,I}$ replaced by Moore–Penrose pseudoinverse when $\Sigma_{n,I}$ is singular; the proxy $T^{\circ}_{n,I}$ is indeed the Gaussian residualization. Joint Gaussianity gives

$$T^{\circ}_{n,I} \perp\!\!\!\perp Z_{n,I}, \qquad \text{hence} \quad T^{\circ}_{n,I} \perp\!\!\!\perp \mathscr{E}_{n,I}(y).$$

Good and bad sets $I \subseteq [\![n]\!]$. We now bound the quantities associated with the good and bad I required by (39) for the proxy and remainder in the form of (41), starting with the definitions of good and bad I.

Fix $k \geq 1$ and $y \in \mathbb{R}$. Consider every $I \subseteq [\![n]\!]$ with $|I| = k$. Take

$$\delta_n = (\log n)^{-2}.$$

We call a pair $i \neq j$ *weakly correlated* if $|\sigma_{ij}| \leq \delta_n$ and *strongly correlated* otherwise. We call a set I *weakly correlated* if $|\sigma_{ij}| \leq \delta_n$ for all distinct $i, j \in I$.

We call a set I *good* if it satisfies *both* of the following conditions:

- *I itself is weakly correlated*: $|\sigma_{ij}| \leq \delta_n$ for all distinct $i, j \in I$, and
- *$r_{n,I}$ is uniformly weak*: $|r_{n,i}| \leq \delta_n$ for all $i \in I$, or alternatively, $\|r_{n,I}\|_\infty \leq \delta_n$.

Every *good* set I is *weakly correlated*, but the converse cannot be guaranteed. *Bad* sets I are of two types:

- *Type I bad (strong)*: there exists at least one pair of distinct $i, j \in I$ such that $|\sigma_{ij}| > \delta_n$.
- *Type II bad (weak)*: I is weakly correlated, but $\|r_{n,I}\|_\infty > \delta_n$, that is, there exists $i \in I$ such that

$$|r_{n,i}| = |(r_n)_i| = \left| \frac{\sum_{j=1}^n \sigma_{ij}}{\sigma_n} \right| = \left| \frac{\sum_{j=1}^n \sigma_{ij}}{\sqrt{\sum_{i,j} \sigma_{ij}}} \right| > \delta_n,$$

where

$$r_n := \mathrm{Cov}(Z_n, T_n) = \frac{\Sigma_n \mathbf{1}_n}{\sigma_n}.$$

Compute $\mathbb{P}\{\mathscr{E}_{n,I}(y)\}$ for all weakly correlated I. I is weakly correlated if $|\sigma_{ij}| \le \delta_n$ for all distinct $i, j \in I$. By the Gershgorin circle theorem (Gerschgorin, 1931), the following squeeze argument (squeeze principle) shows that $\mathbb{P}\{\mathscr{E}_{n,I}(y)\}$ behaves asymptotically as if $\Sigma_{n,I} = \mathbf{I}_k$.

Setting $a_{k,n} = (k-1)\delta_n$, the eigenvalues of $\Sigma_{n,I}$ lie between $1-a_{k,n}$ and $1+a_{k,n}$ and

$$(1-a_{k,n})\mathbf{I}_k \preceq \Sigma_{n,I} \preceq (1+a_{k,n})\mathbf{I}_k.$$

Consequently, for all sufficiently large n,

$$\frac{\mathbf{I}_k}{1+a_{k,n}} \preceq \Sigma_{n,I}^{-1} \preceq \frac{\mathbf{I}_k}{1-a_{k,n}}, \qquad (1-a_{k,n})^k \le \det \Sigma_{n,I} \le (1+a_{k,n})^k.$$

Applying these bounds directly to the Gaussian density for $Z_{n,I}$ and integrating over the region $(z,+\infty)^k$ for $z > 0$, we obtain

$$\begin{aligned}
\left(\frac{1-a_{k,n}}{1+a_{k,n}}\right)^{k/2}\left[\overline{\Phi}\left(\frac{z}{\sqrt{1-a_{k,n}}}\right)\right]^k \le\ & \mathbb{P}(Z_{n,i_1} > z, \dots, Z_{n,i_k} > z) \\
&= \frac{1}{(2\pi)^{k/2}\det(\Sigma_{n,I})^{1/2}} \int_z^{+\infty} \cdots \int_z^{+\infty} \exp\left(-\frac{1}{2}\boldsymbol{x}_I^\top \Sigma_{n,I}^{-1} \boldsymbol{x}_I\right) \mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_k} \\
&\le \left(\frac{1+a_{k,n}}{1-a_{k,n}}\right)^{k/2}\left[\overline{\Phi}\left(\frac{z}{\sqrt{1+a_{k,n}}}\right)\right]^k.
\end{aligned}$$

Taking $z = u_n(y)$, since

$$u_n(y)^2 = 2\log n - \log\log n - \log(4\pi) + 2y + O\left(\frac{(\log\log n)^2}{\log n}\right)$$

and $a_{k,n}u_n(y)^2 = O(k\delta_n \log n) = O\{k(\log n)^{-1}\} = o(1)$ for any fixed k, Mills' ratio yields

$$\begin{aligned}
\frac{\overline{\Phi}(u_n(y)/\sqrt{1\pm a_{k,n}})}{\overline{\Phi}(u_n(y))} &\sim \frac{\phi(u_n(y)/\sqrt{1\pm a_{k,n}})}{\phi(u_n(y))} \times \frac{u_n(y)}{u_n(y)/\sqrt{1\pm a_{k,n}}} \\
&= \sqrt{1\pm a_{k,n}} \exp\left\{\frac{z^2}{2}\left(1 - \frac{1}{1\pm a_{k,n}}\right)\right\} \longrightarrow 1.
\end{aligned}$$

Meanwhile,

$$q_n(y) := \overline{\Phi}(u_n(y)) \sim \frac{e^{-y}}{n}.$$

Therefore, the preceding bounds imply

$$\mathbb{P}\{\mathscr{E}_{n,I}(y)\} = \mathbb{P}(Z_{n,i_1} > u_n(y), \dots, Z_{n,i_k} > u_n(y)) = (1+o(1))q_n(y)^k = (1+o(1))\left(\frac{e^{-y}}{n}\right)^k. \tag{42}$$

All error bounds depend only on the fixed k and $\delta_n = (\log n)^{-2}$, and not on the choice of index set $I = \{i_1, \dots, i_k\}$. Hence, the convergence is uniformly over all *good* (in fact, over all *weakly correlated*) I, completing the proof.

Compute $H_{n,k}^{\text{good}}(y)$. Recall the definition

$$H_{n,k}^{\text{good}}(y) := \sum_{|I|=k:\ I \text{ good}} \mathbb{P}\{\mathscr{E}_{n,I}(y)\}$$

We will show below that the cardinality $\#\{|I|=k:\ I \text{ bad}\}$ is $o(n^k)$:

$$\sum_{|I|=k:\ I \text{ bad}} 1 = o(n^k) \tag{43}$$

Then (42) gives

$$H_{n,k}^{\text{good}}(y) = \sum_{|I|=k:\ I \text{ good}} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} = (1+o(1))\left(\frac{e^{-y}}{n}\right)^k \left\{\binom{n}{k} - o(n^k)\right\} = O(1). \tag{44}$$

Bound $\rho_{n,k}^{\text{good}}(\eta_n)$. Recall the definition of $\rho_{n,k}^{\text{good}}(\eta_n)$

$$\rho_{n,k}^{\text{good}}(\eta_n) = \sup_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}(|T_n - T_{n,I}^{\circ}| > \eta_n) = \sup_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}(|R_{n,I}| > \eta_n)$$

and the explicit form of our constructed remainder $R_{n,I}$

$$R_{n,I} = \boldsymbol{r}_{n,I}^{\top} \Sigma_{n,I}^{-1} \boldsymbol{Z}_{n,I}, \qquad \text{with} \quad \mathbb{E}R_{n,I}^2 = \boldsymbol{r}_{n,I}^{\top} \Sigma_{n,I}^{-1} \boldsymbol{r}_{n,I}.$$

Notice that the covariance matrix $\Sigma_{n,I}$ is always invertible for good (and thus weakly correlated) I, since its eigenvalues all lie between $1-a_{k,n}$ and $1+a_{k,n}$. Since $R_{n,I}$ is a linear transformation of a Gaussian vector

$$R_{n,I} \sim N\left(0, \boldsymbol{r}_{n,I}^{\top} \Sigma_{n,I}^{-1} \boldsymbol{r}_{n,I}\right).$$

For good I, we have $\lambda_{\min}(\Sigma_{n,I}) \geq 1-(k-1)\delta_n$ by its weak correlation within I, and

$$||\boldsymbol{r}_{n,I}||^2 \leq k\delta_n^2,$$

by its weak correlation with the standardized sum. Thus

$$\mathbb{E}R_{n,I}^2 \leq \frac{k\delta_n^2}{1-(k-1)\delta_n}.$$

The Gaussian tail bound (Chernoff bound) gives

$$\mathbb{P}(|R_{n,I}| > \eta_n) \leq 2\exp\left\{-\frac{\eta_n^2[1-(k-1)\delta_n]}{2k\delta_n^2}\right\}.$$

Taking $\eta_n = (\log n)^{-1/2}$ and recalling $\delta_n = (\log n)^{-2}$ yields

$$\sup_{\substack{|I|=k \\ I \text{ good}}} \mathbb{P}\left\{|R_{n,I}| > (\log n)^{-1/2}\right\} \leq 2\exp\left\{-\frac{(1+o(1))(\log n)^3}{2k}\right\}.$$

This convergence rate is fast enough to offset the combinatorial factor $\binom{d_n}{k}$.

Bound $H_{n,k}^{\text{bad}}(y)$. Recalling the two types of bad sets I:

- *Type I bad (strong)*: there exists at least one pair of distinct $i, j \in I$ such that $|\sigma_{ij}| > \delta_n$
- *Type II bad (weak)*: I is weakly correlated, however, there exists $i \in I$ such that

$$|r_{n,i}| = \left| \frac{\sum_{j=1}^n \sigma_{ij}}{\sigma_n} \right| > \delta_n.$$

We further write

$$H_{n,k}^{\text{bad}}(y) = H_{n,k}^{\text{bad (strong)}}(y) + H_{n,k}^{\text{bad (weak)}}(y) := \sum_{\substack{|I|=k \\ I \text{ bad (strong)}}} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} + \sum_{\substack{|I|=k \\ I \text{ bad (weak)}}} \mathbb{P}\{\mathscr{E}_{n,I}(y)\}.$$

Control $H_{n,k}^{\text{bad (strong)}}(y)$. When $k = 2$, to control $H_{n,k}^{\text{bad (strong)}}(y)$, applying the normal comparison inequality in Li and Shao (2002) yields the desired bound. However, this argument does not directly extend to $k \geq 3$. We instead consider the following connected-component argument, which applies to all $k \geq 2$. The heuristic is to partition the variables into components such that variables within each component may be strongly correlated, whereas dependence between different components is weak. Consequently, the component averages are nearly independent, so their weighted sum behaves approximately like a sum of independent random variables. This approach resembles classical dependency-graph arguments, except that distinct components here are only weakly correlated rather than independent.

Strong-correlation graph and connected components. We consider a graph G_n with vertex set $[\![n]\!]$, where two distinct vertices i and j are connected whenever $|\sigma_{ij}| > \delta_n$—they have "strong correlation". This graph has maximum degree at most

$$\#\{j \neq i : |\sigma_{ij}| > \delta_n\} \leq \frac{\max_i \sum_{j=1}^n \sigma_{ij}^2}{\delta_n^2} \leq (\log n)^{\xi+4} =: D_n, \tag{45}$$

where $\max_i \sum_{j=1}^n \sigma_{ij}^2 \leq (\log n)^{\xi}$ is in our assumption (38).

For every Type I *bad* set I with $|I| = k$, we must have $k \geq 2$ since there exists at least one pair of distinct $i, j \in I$ such that $|\sigma_{ij}| > \delta_n$. Its induced subgraph $G_n[I]$ contains at least one edge, because such *strongly correlated* pair exists.

Let $m \in [\![k-1]\!]$ denote the number of connected components of $G_n[I]$—here m is at most $k-1$ since $G_n[I]$ contains at least one edge. We show the number of such sets with a given m is at most $C_k n^m D_n^{k-m}$ for some constant C_k depending only on k: note that every such $G_n[I]$ contains a spanning forest consisting of m trees with each connected component containing a spanning tree. The counting is conducted by choosing one root per component and adding the remaining vertices along spanning trees. Each k-set is counted at least $k!$ times, so we divide the resulting count by $k!$:

$$\#\{I \subseteq [\![n]\!] : |I| = k,\ G_n[I] \text{ has } m \text{ components}\} \leq \frac{|\mathscr{F}_{k,m}|}{k!} n^m D_n^{k-m},$$

where $\mathscr{F}_{k,m}$ denotes the collection of labeled forests on index set $[\![k]\!]$ with exactly m connected components. Every such labeled forest has exactly $k-m$ edges. Thus, by Harary (1969, Chapter 15)

$$|\mathscr{F}_{k,m}| \leq \binom{\binom{k}{2}}{k-m} \leq 2^{\binom{k}{2}}.$$

Gaussian tail bound via weighted component averages. For any such set $I \subseteq [\![n]\!]$ with $|I| = k$ whose induced subgraph $G_n[I]$ has m components, write its components as $\mathscr{C}_1, \dots, \mathscr{C}_m$, with sizes $\varkappa_\ell = |\mathscr{C}_\ell|$, and define

$$Y_\ell = \frac{1}{\varkappa_\ell} \sum_{i \in \mathscr{C}_\ell} Z_{n,i}, \qquad w_\ell = \frac{\varkappa_\ell}{1 + (\varkappa_\ell - 1)\varrho}, \qquad \widehat{Y} = \sum_{\ell=1}^{m} w_\ell Y_\ell, \qquad w = \sum_{\ell=1}^{m} w_\ell.$$

Here w_ℓ is chosen as the reciprocal of the following upper bound on $\mathrm{Var}(Y_\ell)$:

$$\mathrm{Var}(Y_\ell) \le \frac{1}{\varkappa_\ell^2} \left(\sum_{i \in \mathscr{C}_\ell} \sigma_{ii} + \sum_{i,j \in \mathscr{C}_\ell : i \ne j} |\sigma_{ij}| \right) \le \frac{\varkappa_\ell + \varkappa_\ell(\varkappa_\ell - 1)\varrho}{\varkappa_\ell^2} = \frac{1 + (\varkappa_\ell - 1)\varrho}{\varkappa_\ell}.$$

Accordingly, w_ℓ can be viewed as the "effective number of independent variables" (a.k.a. the "degrees of freedom") that the component $\mathscr{C}_\ell$ contributes to the Gaussian tail exponent. For example,

- If $\varkappa_\ell = 1$, then $w_\ell = 1$; if $\varkappa_\ell \ge 2$, then $w_\ell \ge 2/(1+\varrho) > 1$, and is monotonically increasing in $\varkappa_\ell$.
- If $\varrho = 0$, then $w_\ell = \varkappa_j$; the variables in this component are i.i.d. Gaussian.
- If $\varrho \uparrow 1$, then $w_\ell \downarrow 1$; the variables in this component can be up to almost fully correlated.

For any Type I bad k-set I, since at least one component $\mathscr{C}_{\ell^*}$ contains two or more vertices, we have

$$w - m = \sum_{\ell=1}^{m} \frac{(\varkappa_\ell - 1)(1-\varrho)}{1 + (\varkappa_\ell - 1)\varrho} \ge \frac{(\varkappa_{\ell^*} - 1)(1-\varrho)}{1 + (\varkappa_{\ell^*} - 1)\varrho} \ge \frac{1-\varrho}{1+\varrho} =: \gamma > 0,$$

and the variance bounds $\mathrm{Var}(Y_\ell) \le w_\ell^{-1}$ and $|\mathrm{Cov}(Y_\ell, Y_h)| \le \delta_n$ for $\ell \ne h$ imply

$$\mathrm{Var}(\widehat{Y}) = \sum_{\ell} w_\ell^2 \mathrm{Var}(Y_\ell) + 2 \sum_{\ell < h} w_\ell w_h \mathrm{Cov}(Y_\ell, Y_h) \le w + 2\delta_n \sum_{\ell < h} w_\ell w_h \le w + \delta_n w^2.$$

On $\mathscr{E}_{n,I}(y)$ we have $Y_\ell > u_n(y)$ for all ℓ and thus $\widehat{Y} > w u_n(y)$; hence the Gaussian tail bound gives

$$\begin{aligned} \mathbb{P}(\mathscr{E}_{n,I}(y)) &\le \mathbb{P}(\widehat{Y} > w u_n(y)) \le \exp\left\{ -\frac{w^2 u_n(y)^2}{2\,\mathrm{Var}(\widehat{Y})} \right\} \\ &\le \exp\left\{ -\frac{w u_n(y)^2}{2(1 + \delta_n w)} \right\} \le C_{k,y} (\log n)^{w/2} n^{-w} \le C_{k,y} (\log n)^{k/2} n^{-m-\gamma}. \end{aligned}$$

where $m \le w \le k$ and recall that

$$u_n(y)^2 = 2\log n - \log\log n - \log(4\pi) + 2y + O\left(\frac{(\log\log n)^2}{\log n} \right).$$

Here $\delta_n w$ can be absorbed by some constant $C_{k,y}$ because $w u_n(y)^2 \times \delta_n w = O(k^2 \delta_n \log n) = O\{k^2 (\log n)^{-1}\} = o(1)$.

Combining component counts and weighted-sum tail bounds. Summing over $m \in \{1, \dots, k-1\}$ shows that the total contribution is

$$\begin{aligned} H_{n,k}^{\text{bad (strong)}}(y) &= \sum_{|I|=k:\ I \text{ bad (strong)}} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} \\ &\le C_{k,y} (\log n)^{k/2} \sum_{m=1}^{k-1} D_n^{k-m} n^{-\gamma} \le C_{k,y} (\log n)^{c_{k,\xi}} n^{-\gamma} = o(1). \end{aligned} \tag{46}$$

To conclude the part of controlling $H_{n,k}^{\text{bad (strong)}}(y)$, note also that the total number of Type I bad k-sets is at most

$$\sum_{m=1}^{k-1}\binom{\binom{k}{2}}{k-m}n^m D_n^{k-m} \le 2^{\binom{k}{2}}n^{k-1}D_n = o(n^k), \tag{47}$$

which will be used later.

Control $H_{n,k}^{\text{bad (weak)}}(y)$. Now it remains to handle the Type II bad sets I. By the spectral theorem, $\Sigma_n^2 \preceq \lambda_{\max}(\Sigma_n)\Sigma_n$ and together with our assumption $\max_i \sum_{j=1}^n \sigma_{ij}^2 \le (\log n)^\xi$ in (38),

$$\|r_n\|^2 \le \lambda_{\max}(\Sigma_n) \le \sqrt{\operatorname{tr}(\Sigma_n^2)} \le \sqrt{n}\,(\log n)^{\xi/2}.$$

Thus the exceptional set

$$K_n := \{i : |r_{n,i}| > \delta_n\} \quad \text{satisfies} \quad |K_n| \le \sqrt{n}\,(\log n)^{\xi/2+4} = o(n).$$

Then the number of Type II *bad* k-sets, i.e., that of k-sets containing at least one exceptional coordinate is at most

$$|K_n|\binom{n-1}{k-1} = o(n^k) \tag{48}$$

Notice that Type II bad sets I are still weakly correlated, so together with (42),

$$H_{n,k}^{\text{bad (weak)}}(y) = \sum_{|I|=k:\ I \text{ bad (weak)}} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} = (1+o(1))\left(\frac{e^{-y}}{n}\right)^k o(n^k) = o(1). \tag{49}$$

To conclude the part of controlling $H_{n,k}^{\text{bad (weak)}}(y)$, combining (48) and (47) proves (43).

Combine good and bad parts. We have verified all conditions in (39) except that $\omega_n(2\eta_n) \to 0$, which is obvious: T_n under our normalization is exactly $N(0,1)$, which implies

$$\omega_n(2\eta_n) \le \frac{4\eta_n}{\sqrt{2\pi}} \longrightarrow 0. \tag{50}$$

Therefore, Theorem 3.1 gives $D_{n,k}(y) \to 0$ for every fixed k. Putting (44), (46), (49) together implies

$$H_{n,k}(y) = \sum_{|I|=k} \mathbb{P}\{\mathscr{E}_{n,I}(y)\} = \left(\frac{e^{-y}}{n}\right)^k \binom{n}{k} + o(1) \longrightarrow \frac{e^{-ky}}{k!} \qquad (k \ge 1 \text{ fixed}). \tag{51}$$

Since (51) also implies the high-order binomial-moment condition (4), Corollary 2.4 yields (37). Thus the Gaussian sum and maximum are asymptotically independent under dependence structure characterized by (38).

McCormick and Qi (2000) studied the joint limit of such statistics using a maximum-proxy argument. Unlike them, the proof above directly applies Theorem 3.1 by constructing a proxy for the sum T_n. The assumptions are first two parts of Assumption 3 in Feng, Jiang, Li, and Liu (2024).

5.3 High-dimensional independence testing based on pairwise sample covariances

We next explain how the results of Li and Xue (2015) fit into our framework. Their setting concerns high-dimensional mutual-independence testing based on pairwise sample covariances or Spearman's ρ. We illustrate how our framework applies to the former; indeed, it can accommodate both.

Write the p_n data columns as $\boldsymbol{X}_1,\dots,\boldsymbol{X}_{p_n}$ and index the $d_n=\binom{p_n}{2}$ pairs in $\{(j,h):1\le j<h\le p_n\}$ in any fixed order (e.g., lexicographic order) by writing

$$\{(j,h):1\le j<h\le p_n\}=\{(j_\ell,h_\ell):1\le \ell\le d_n\}.$$

Their quadratic statistic is, after centering and normalization,

$$T_n=b_n\sum_{\ell=1}^{d_n}(V_\ell-1),\qquad \text{where}\quad V_\ell=\frac{(\boldsymbol{X}_{j_\ell}^\top \boldsymbol{X}_{h_\ell})^2}{n},$$

and b_n is some regime-dependent normalization factor defined in their Lemma 1. The max statistic is defined as $\max_\ell V_\ell$. Set $\mathscr{E}_{n,\ell}(y)=\{V_\ell>u_n(y)\}$ and $\mathscr{E}_{n,I}(y)=\bigcap_{\ell\in I}\mathscr{E}_{n,\ell}(y)$ for the threshold

$$u_n(y)=4\log p_n-\log\log p_n+y. \tag{52}$$

Suppose that the underlying data-generating process $\{X_{ij}:1\le i\le n,\ 1\le j\le p_n\}$ are i.i.d. with a fixed distribution satisfying

$$\mathbb{E}X_{11}=0,\qquad \mathbb{E}X_{11}^2=1,\qquad \mathbb{E}X_{11}^4>1,\qquad \mathbb{E}e^{t_0X_{11}^2}<\infty\quad\text{for some } t_0>0. \tag{53}$$

We further assume the following growth regime:

$$n,p_n\to\infty,\qquad p_n\ge n,\qquad \log p_n=o(n^{1/6}). \tag{54}$$

For $I\subseteq[\![d_n]\!]$ with $|I|=k$, Li and Xue define W_I to be the set of indices ℓ such that the pair (j_ℓ,h_ℓ) shares at least one coordinate with some pair indexed by I. The cardinality of W_I satisfies

$$|W_I|\le 2k\,p_n.$$

Because the original columns $\boldsymbol{X}_1,\dots,\boldsymbol{X}_{p_n}$ are independent under the null hypothesis, the variables $\{V_\ell:\ell\in I\}$ are independent of the collection $\{V_\ell:\ell\notin W_I\}$. Hence the proxy with its remainder

$$T_{n,I}^\circ=b_n\sum_{\ell\notin W_I}(V_\ell-1),\qquad R_{n,I}:=T_n-T_{n,I}^\circ=b_n\sum_{\ell\in W_I}(V_\ell-1) \tag{55}$$

satisfies $T_{n,I}^\circ\perp\!\!\!\perp\mathscr{E}_{n,I}(y)$. This is precisely the local-deletion mechanism of Section 4.

In this example, we will use Corollary 3.3 to verify the proxy criterion. We only need to verify the three conditions in (22), starting from the one about $\rho_{n,k}(\eta_n)$. Equation (23) of Li and Xue (2015) provides a uniform tail bound for the proxy remainder. Set

$$\eta_n=\min\{p_n^{-1/5},\sqrt{n/p_n}\}.$$

For each fixed k, their argument yields constants $C_k,c_k>0$ such that

$$\rho_{n,k}(\eta_n)=\sup_{|I|=k}\mathbb{P}\{|R_{n,I}|>\eta_n\}\le C_ke^{-c_kn^{1/3}}. \tag{56}$$

In fact, they state that the constants can be chosen universally over a certain range of k, which is needed for their growing-order Bonferroni argument; we discuss it in the end of this section.

Next we compute $H_{n,k}(y)$. We show that for each fixed $k \geq 1$ and $y \in \mathbb{R}$,

$$H_{n,k}(y) \longrightarrow \frac{\{\lambda(y)\}^k}{k!}, \qquad \lambda(y) = \frac{e^{-y/2}}{\sqrt{8\pi}}. \tag{57}$$

Compute $q_n(y)$ and $H_{n,1}(y)$. Recall $u_n(y) = 4\log p_n - \log\log p_n + y$ and define $q_n(y) = \mathbb{P}\{V_1 > u_n(y)\}$. We have

$$q_n(y) = \mathbb{P}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{i1}X_{i2} > \sqrt{u_n(y)}\right\} + \mathbb{P}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{i1}X_{i2} < -\sqrt{u_n(y)}\right\}.$$

By Cramér-type moderate deviation,

$$\mathbb{P}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{i1}X_{i2} > z\right\} = (1+o(1))\overline{\Phi}(z), \qquad \mathbb{P}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{i1}X_{i2} < -z\right\} = (1+o(1))\overline{\Phi}(z),$$

uniformly over $0 \leq z = o(n^{1/6})$. Taking $z = \sqrt{u_n(y)}$, Mills' ratio yields

$$2\overline{\Phi}\Big(\sqrt{u_n(y)}\Big) \sim \sqrt{\frac{2}{\pi}}\,\frac{e^{-u_n(y)/2}}{\sqrt{u_n(y)}} \sim \sqrt{\frac{2}{\pi}}\,\frac{p_n^{-2}(\log p_n)^{1/2}e^{-y/2}}{2(\log p_n)^{1/2}} = \frac{e^{-y/2}}{\sqrt{2\pi}}\,p_n^{-2},$$

and thus

$$H_{n,1}(y) = \sum_{\ell}\mathbb{P}\{\mathscr{E}_{n,\ell}(y)\} = \sum_{\ell}\mathbb{P}\{V_\ell > u_n(y)\} = \binom{p_n}{2} q_n(y) \longrightarrow \lambda(y) := \frac{e^{-y/2}}{\sqrt{8\pi}}. \tag{58}$$

Compute $H_{n,k}^{\mathrm{disjoint}}(y)$. We first compute the total contribution from all k-sets I for which $\{j_\ell, h_\ell : \ell \in I\}$ contains exactly $2k$ distinct coordinates. The number of such k-sets I is

$$\binom{p_n}{2k}\frac{(2k)!}{2^k k!} = \frac{(p_n)_{2k}}{2^k k!},$$

recalling $(a)_k = a(a-1)\cdots(a-k+1)$. For every such set I, we have $\mathbb{P}\{\mathscr{E}_{n,I}(y)\} = q_n(y)^k$ and thus

$$H_{n,k}^{\mathrm{disjoint}}(y) = \frac{(p_n)_{2k}}{2^k k!}\,q_n(y)^k = \frac{p_n^{2k}}{2^k k!}\left(\frac{e^{-y/2}}{\sqrt{2\pi}\,p_n^2}\right)^k + o(1) = \frac{1}{k!}\left(\frac{e^{-y/2}}{\sqrt{8\pi}}\right)^k + o(1). \tag{59}$$

Bound $H_{n,k}^{\mathrm{overlap}}(y)$ and compute $H_{n,k}(y)$. The number of k-sets I for which $\{j_\ell, h_\ell : \ell \in I\}$ has $2k-1$ or less distinct coordinates is $O_k(p_n^{2k-1})$.

Consider every such I with $|I| = k$. For any fixed $\boldsymbol{\varepsilon} = (\varepsilon_\ell)_{\ell\in I} \in \{-1,1\}^k$, define

$$W_{\boldsymbol{\varepsilon},i} = \sum_{\ell\in I}\varepsilon_\ell X_{ij_\ell}X_{ih_\ell}.$$

Denoting $S_\ell = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_{ij_\ell}X_{ih_\ell}$, notice that

$$\varepsilon_\ell S_\ell > \sqrt{u_n(y)} \quad \text{for all } \ell \in I \qquad \Longrightarrow \qquad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{\boldsymbol{\varepsilon},i} > k\sqrt{u_n(y)}.$$

We also have $\mathbb{E}W_{\varepsilon,i} = 0$ and $\mathrm{Var}(W_{\varepsilon,i}) = k$ since $\{j_\ell, h_\ell\} \neq \{j_{\ell'}, h_{\ell'}\}$ for $\ell \neq \ell'$, and thus by (53)

$$\log \mathbb{E} e^{t W_{\varepsilon,i}} = \frac{k t^2}{2} + O_k(|t|^3).$$

Summing over 2^k choices of signs and using the Chernoff bound gives

$$\begin{aligned}\mathbb{P}\{\mathscr{E}_{n,I}(y)\} = \mathbb{P}\left\{\bigcap_{\ell\in I}\left\{V_\ell > u_n(y)\right\}\right\} &= \sum_{\varepsilon\in\{-1,1\}^k} \mathbb{P}\left\{\bigcap_{\ell\in I}\left\{\varepsilon_\ell S_\ell > \sqrt{u_n(y)}\right\}\right\} \\ &\le 2^k \exp\left\{-\frac{k u_n(y)}{2} + O_k\left(\frac{u_n(y)^{3/2}}{\sqrt{n}}\right)\right\} \le C_{k,y}\, p_n^{-2k} (\log p_n)^{k/2}.\end{aligned}$$

This together with the fact that the number of such I is $O_k(p_n^{2k-1})$ yields

$$H_{n,k}^{\text{overlap}}(y) \le C_{k,y} \frac{(\log p_n)^{k/2}}{p_n} = o(1). \tag{60}$$

Combining (59) and (60) concludes

$$H_{n,k}(y) = \frac{(p_n)_{2k}}{2^k k!}\, q_n(y)^k + O_{k,y}\left(\frac{(\log p_n)^{k/2}}{p_n}\right).$$

The remaining ingredient of Corollary 3.3 is the Lévy concentration probability mass. Lemma 1 of Li and Xue (2015) gives a normal approximation of F_n whose Kolmogorov error is of order η_n. Therefore

$$\omega_n(2\eta_n) = \sup_{x\in\mathbb{R}}\{F_n(x+2\eta_n) - F_n(x-2\eta_n)\} = O(\eta_n), \tag{61}$$

because the standard normal distribution has a bounded density; this is stronger than necessary—the convergence $F_n \to \Phi$ alone suffices to show that $\omega_n(2\eta_n) = o(1)$.

Putting (56), (57), (61) into Corollary 3.3 gives the bound

$$\begin{aligned}D_{n,k}(y) &\le \rho_{n,k}(\eta_n) H_{n,k}(y) + \rho_{n,k}(\eta_n)\binom{d_n}{k} + \omega_n(2\eta_n) \times H_{n,k}(y) \\ &= o(1) \qquad \text{for every fixed } k.\end{aligned}$$

The asymptotic independence follows by noticing that (57) also implies the high-order condition (4) in Theorem 2.2. The joint limit law follows from Corollary 2.4 by plugging (57) into (10) and combining with the marginal limit law for T_n.

Most existing literature on high-dimensional independence testing based on max-type statistics (e.g., Jiang (2004); Zhou (2007); Han, Chen, and Liu (2017); Drton, Han, and Shi (2020)) establishes the limiting distributions of the maxima by deriving the limits of $H_{n,1}$, together with Poisson approximation via the Chen–Stein method (Arratia, Goldstein, and Gordon, 1989, 1990). Here, by directly calculating $H_{n,k}(y)$ for every fixed k, we also provide an alternative way to establish the limiting distribution of the maximum through Corollary 2.4.

The fixed-order Bonferroni argument above is intended only to establish the joint limit, whereas the growing-order Bonferroni argument in Li and Xue (2015) is still needed to establish the uniform convergence rate. Our framework also accommodates the latter.

5.4 High-dimensional independence testing: finite-L_q blocks of pairwise rank-based statistics

Theorem 2.2 of Zhao, Wang, and Feng (2026) gives a natural multivariate application of the local-deletion principle. They also aim to test mutual independence in high dimensions, but rely on pairwise rank statistics. We retain the notation from the previous section: write the p_n data columns as $X_1,\dots,X_{p_n}$ and index the $d_n = \binom{p_n}{2}$ pairs in any ordering by

$$(j_\ell, h_\ell), \qquad 1 \le \ell \le d_n.$$

Let U_ℓ denote a generic rank-based statistic, which is either

(V) a simple linear rank statistic (e.g., Spearman's ρ (Spearman, 1904)), or

(W) a non-degenerate rank-based U-statistic (e.g., Kendall's τ (Kendall, 1938)), or

(Q) a degenerate rank-based U-statistic (e.g., Hoeffding's D (Hoeffding, 1948), Blum–Kiefer–Rosenblatt's R (Blum et al., 1961), Bergsma–Dassios–Yanagimoto's τ^* (Bergsma and Dassios, 2014; Yanagimoto, 1970)),

computed from X_{j_ℓ} and X_{h_ℓ} in each case. Let $\widetilde{U}_\ell = U_\ell - \mathbb{E}(U_\ell)$; here and throughout, expectations and probabilities are taken under the null. Since U_ℓ is rank-based, we can assume without loss of generality that all the marginals in the underlying data-generating process $\{X_{ij} : 1 \le i \le n,\ 1 \le j \le p_n\}$ are Unif$[0,1]$, and, in particular, are i.i.d. Unif$[0,1]$ under the null.

They define, for each fixed even q,

$$S^U_{n;q} = \sum_{\ell=1}^{d_n} \{\widetilde{U}_\ell^q - \mathbb{E}(\widetilde{U}_\ell^q)\}, \qquad \sigma^U_{n;q} = \sqrt{\sum_{\ell=1}^{d_n} \mathrm{Var}(\widetilde{U}_\ell^q)}, \qquad T^U_{n;q} = \frac{S^U_{n;q}}{\sigma^U_{n;q}}.$$

For a fixed finite set $\mathcal{Q} = \{q_1,\dots,q_K\} \subseteq 2\mathbb{Z}_+$, the sum-type object is defined as the vector

$$\boldsymbol{T}^U_{n;\mathcal{Q}} = (T^U_{n;q_1},\dots,T^U_{n;q_K})^\top.$$

For types (V) and (W), define their max statistics as

$$M^U_n = \frac{\max_{1\le\ell\le d_n} \widetilde{U}_\ell^2}{\mathrm{Var}(\widetilde{U}_1)} \qquad \text{and set} \quad \mathcal{E}^U_{n,\ell}(y) = \left\{ \frac{\widetilde{U}_\ell^2}{\mathrm{Var}(\widetilde{U}_1)} > u_n(y) \right\}$$

and $\mathcal{E}^U_{n,I}(y) = \bigcap_{\ell\in I} \mathcal{E}^U_{n,\ell}(y)$ for the extreme-value threshold $u_n(y) = 4\log p_n - \log\log p_n + y$, as in (52). The growing regime is given by (54).

When U is of type (Q)—a degenerate rank-based U-statistic with a kernel h of order m that is rank-based, symmetric, and satisfies the following three properties:

(i) h is bounded.

(ii) h is mean-zero and degenerate under independent continuous margins, i.e., $h_1\big((x_{11},x_{12})\big) = 0$.

(iii) $h_2\big((x_{11},x_{12}),(x_{21},x_{22})\big)$ has uniformly bounded eigenfunctions, that is, it admits the expansion

$$h_2\big((x_{11},x_{12}),(x_{21},x_{22})\big) = \sum_{v=1}^{\infty} \lambda_v \phi_v\big((x_{11},x_{12})\big)\phi_v\big((x_{21},x_{22})\big),$$

where $\{\lambda_v\}$ and $\{\phi_v\}$ are the eigenvalues and eigenfunctions satisfying the integral equation

$$\mathbb{E}h_2\big((x_{11},x_{12}),(X_{21},X_{22})\big)\phi\big((X_{21},X_{22})\big) = \lambda\phi\big((x_{11},x_{12})\big) \quad \text{for all } (x_{11},x_{12}) \in \mathbb{R}^2,$$

with $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, $\Lambda := \sum_{v=1}^\infty \lambda_v \in (0,\infty)$, and $\sup_v \|\phi_v\|_\infty < \infty$.

Here we write $h_1\big((x_{11},x_{12})\big) := \mathbb{E}h\big((x_{11},x_{12}),(X_{21},X_{22}),\ldots,(X_{m1},X_{m2})\big)$ and $h_2\big((x_{11},x_{12}),(x_{21},x_{22})\big) := \mathbb{E}h\big((x_{11},x_{12}),(x_{21},x_{22}),(X_{31},X_{32})\ldots,(X_{m1},X_{m2})\big)$; recall expectations and probabilities are always taken under the null. The max statistic for type (Q) is accordingly defined as

$$M_n^U = \frac{n-1}{\lambda_1\binom{m}{2}} \max_{1\le \ell \le d_n} \widetilde{U}_\ell \qquad \text{and set} \quad \mathscr{E}_{n,\ell}^U(y) = \left\{ \frac{n-1}{\lambda_1\binom{m}{2}} \widetilde{U}_\ell > u_n^U(y) \right\}$$

and $\mathscr{E}_{n,I}^U(y) = \bigcap_{\ell\in I} \mathscr{E}_{n,\ell}^U(y)$ for the extreme-value threshold

$$u_n^U(y) = 4\log p_n + (\mu_1 - 2)\log\log p_n - \frac{\Lambda}{\lambda_1} + y,$$

and the growth regime is given by

$$n, p_n \to \infty, \qquad p_n \ge n, \qquad \log p_n = o(n^\theta), \tag{62}$$

where θ is characterized by Equation (4.1) in Drton, Han, and Shi (2020) and can be taken as $1/8-\delta$ for any arbitrarily small, prespecified constant $\delta > 0$ for the aforementioned three statistics D, R, and τ^*.

We also impose the remaining regularity assumptions for types (V), (W), and (Q); we refer the reader to the original paper for details.

We use the multivariate version of Corollary 3.3 to verify the proxy criterion. Fix $k \ge 1$ and $y \in \mathbb{R}$. For $I \subseteq [\![d_n]\!]$ with $|I| = k$, W_I is defined the same as in Li and Xue (2015) with $|W_I| \le 2k p_n$. For $q \in \mathscr{Q}$, define the proxy and the remainder by

$$\big(S_{n,I;q}^U\big)^\circ = \sum_{\ell \notin W_I} \{\widetilde{U}_\ell^q - \mathbb{E}(\widetilde{U}_\ell^q)\}, \qquad R_{n,I;q}^U = \sum_{\ell \in W_I} \{\widetilde{U}_\ell^q - \mathbb{E}(\widetilde{U}_\ell^q)\},$$

and form the proxy vector and the remainder vector

$$\big(\boldsymbol{T}_{n,I;\mathscr{Q}}^U\big)^\circ = \left(\frac{\big(S_{n,I;q_1}^U\big)^\circ}{\sigma_{n;q_1}^U}, \ldots, \frac{\big(S_{n,I;q_K}^U\big)^\circ}{\sigma_{n;q_K}^U} \right)^\top, \qquad \boldsymbol{R}_{n,I;\mathscr{Q}}^U = \left(\frac{R_{n,I;q_1}^U}{\sigma_{n;q_1}^U}, \ldots, \frac{R_{n,I;q_K}^U}{\sigma_{n;q_K}^U} \right)^\top.$$

Under the null of mutual independence, it holds that $\big(\boldsymbol{T}_{n,I;\mathscr{Q}}^U\big)^\circ \perp\!\!\!\perp \mathscr{E}_{n,I}^U(y)$.

We start with bounding $\rho_{n,k}^U$. Combining their assumption that all kernel/score functions considered are uniformly bounded, along with a rank-moment inequality, their equation (C.3) gives for every fixed even $\tau \ge 4$ and $\eta \in (0,1)$

$$\rho_{n,k}^U(\eta) = \sup_{|I|=k} \mathbb{P}\left\{ \|\boldsymbol{T}_{n;\mathscr{Q}}^U - \big(\boldsymbol{T}_{n,I;\mathscr{Q}}^U\big)^\circ\|_\infty > \eta \right\} \le C_{\mathscr{Q},\tau}^U \, \eta^{-\tau} k^{2\tau} p_n^{-\tau/2}. \tag{63}$$

The polynomial bound in (63) is not as tight as the exponential one in (56), though still strong enough to absorb the factor $\binom{d_n}{k}$. Indeed, choosing $\tau = 6k$ and $\eta_n = p_n^{-1/12}$ in (63) gives, for fixed k,

$$\rho_{n,k}^U(\eta_n)\binom{d_n}{k} \le C_{\mathscr{Q},k}^U \, p_n^{6k/12 - 3k + 2k} = O(p_n^{-k/2}) \longrightarrow 0.$$

Next we compute $H_{n,k}^U(y)$. We show that for each fixed $k \geq 1$ and $y \in \mathbb{R}$,

$$H_{n,k}^U(y) \longrightarrow \frac{\{\lambda^U(y)\}^k}{k!}, \qquad \lambda^U(y) := \begin{cases} \dfrac{e^{-y/2}}{\sqrt{8\pi}}, & U \text{ is of type (V) or (W)}, \\ \dfrac{2^{\mu_1/2-2}\varkappa}{\Gamma(\mu_1/2)} e^{-y/2}, & U \text{ is of type (Q)}, \end{cases} \tag{64}$$

where $\mu_1 := \#\{v : \lambda_v = \lambda_1\}$ and $\varkappa := \prod_{v>\mu_1}(1-\lambda_v/\lambda_1)^{-1/2}$.

Compute $q_n^U(y)$ and $H_{n,1}^U(y)$. Define $q_n^U(y) = \mathbb{P}\{\mathscr{E}_{n,1}^U(y)\}$. By Cramér-type moderate deviation arguments in Han, Chen, and Liu (2017) and Cramér-type moderate deviation bounds for degenerate U-statistics in Drton, Han, and Shi (2020), respectively, we obtain

$$q_n^U(y) \sim \begin{cases} \dfrac{e^{-y/2}}{\sqrt{2\pi}} p_n^{-2}, & U \text{ is of type (V) or (W)}, \\ \dfrac{2^{\mu_1/2-1}\varkappa}{\Gamma(\mu_1/2)} e^{-y/2} p_n^{-2}, & U \text{ is of type (Q)}. \end{cases} \tag{65}$$

Thus

$$H_{n,1}^U(y) = \sum_{\ell} \mathbb{P}\{\mathscr{E}_{n,\ell}^U(y)\} = \binom{p_n}{2} q_n^U(y) \longrightarrow \lambda^U(y).$$

Compute $(H_{n,k}^U)^{\text{disjoint}}(y)$. We first compute the total contribution from all k-sets I such that $\{j_\ell, h_\ell : \ell \in I\}$ have $2k$ distinct elements. The number of such k-sets I is

$$\binom{p_n}{2k} \frac{(2k)!}{2^k k!} = \frac{(p_n)_{2k}}{2^k k!},$$

recalling $(a)_k = a(a-1)\cdots(a-k+1)$. For every such set I, we have $\mathbb{P}\{\mathscr{E}_{n,I}^U(y)\} = q_n^U(y)^k$ and thus

$$(H_{n,k}^U)^{\text{disjoint}}(y) = \frac{(p_n)_{2k}}{2^k k!} q_n^U(y)^k = \frac{1}{k!}\lambda^U(y)^k + o(1). \tag{66}$$

Bound $(H_{n,k}^U)^{\text{overlap}}(y)$ and compute $H_{n,k}^U(y)$. We next consider the k-sets I for which $\{j_\ell, h_\ell : \ell \in I\}$ contains $v \leq 2k-1$ distinct coordinates. We consider the graph on vertex set $[\![p_n]\!]$ with edges given by the pairs (j_ℓ, h_ℓ), $\ell \in I$.

Rank-based statistics have many desirable properties concerning independence (see, e.g., Lemma 2.1 in Leung and Drton (2018)), and using those stated in Proposition 2.1 of Drton, Han, and Shi (2020) gives the following "*forest independence*":

$$\mathbb{P}\left(\bigcap_{(j_\ell,h_\ell)\in E(\mathscr{F})} \mathscr{E}_{n,\ell}^U(y) \right) = q_n^U(y)^{|E(\mathscr{F})|}.$$

This follows by successively removing leaves from a forest $\mathscr{F}$, conditioning on the remaining variables at each step. This "forest independence" says, as long as the underlying graph induced by I is acyclic, the joint probability of the exceedance events indexed by I factors into the product of their individual probabilities, even when the pairs (j_ℓ, h_ℓ), $\ell \in I$, share coordinates.

Now consider the graph associated with a k-set I. Let v and c denote the numbers of non-isolated vertices and nontrivial connected components, respectively. A spanning forest $\mathscr{F}_I$ of this graph has exactly $v-c$ edges. For each fixed v, we have at most

$$\binom{p_n}{v}\binom{\binom{v}{2}}{k} = O_k(p_n^v) \tag{67}$$

such graphs. Using the "forest independence" yields

$$\mathbb{P}\left(\bigcap_{\ell\in I}\mathscr{E}^U_{n,\ell}(y)\right) \le \mathbb{P}\left(\bigcap_{(j_\ell,h_\ell)\in E(\mathscr{F}_I)}\mathscr{E}^U_{n,\ell}(y)\right) = q_n^U(y)^{v-c} = O^U_{y;k}\left(p_n^{-2(v-c)}\right).$$

Furthermore, we have $v \ge 2c+1$ because every connected component has at least two vertices, and at least one of them has three or more—coordinates in $\{j_\ell, h_\ell : \ell \in I\}$ are not distinct. Therefore

$$\mathbb{P}\left(\bigcap_{\ell\in I}\mathscr{E}^U_{n,\ell}(y)\right) = O^U_{y;k}\left(p_n^{-v-1}\right). \tag{68}$$

Taking (67) and (68) together yields

$$(H^U_{n,k})^{\text{overlap}}(y) \le \sum_{v=3}^{2k-1} O_k(p_n^v)\, O^U_{y;k}(p_n^{-v-1}) = O^U_{y;k}(p_n^{-1}) = o(1). \tag{69}$$

Hence, combining (66) and (69), we finish the proof of (64).

We are left with controlling the anti-concentration term. Let $F_n^U(\boldsymbol{x}) := \mathbb{P}\{\boldsymbol{T}^U_{n;\mathscr{Q}} \le \boldsymbol{x}\}$, with $F^U_{n,j}(x)$ denoting its jth marginal, and let ω_n^U denote the corresponding multivariate concentration function. The marginal Gaussian approximations in Theorem 2.1 of Zhao, Wang, and Feng (2026) gives

$$\omega_n^U(2\eta_n) \le 2\sum_{j=1}^{K}\sup_{x\in\mathbb{R}}|F^U_{n,j}(x)-\Phi(x)| + \frac{4K\eta_n}{\sqrt{2\pi}} = o(1). \tag{70}$$

Combining (63), (64), and (70) gives, for every fixed k,

$$D^U_{n,k}(y) \le \rho^U_{n,k}(\eta_n)H^U_{n,k}(y) + \rho^U_{n,k}(\eta_n)\binom{d_n}{k} + \omega_n^U(2\eta_n)\times H^U_{n,k}(y) = o(1)$$

Since (64) implies (4), Theorem 2.2 yields asymptotic independence. Whenever the block $\boldsymbol{T}^U_{n;\mathscr{Q}}$ converges in distribution, the joint limit law follows from the multivariate version of Corollary 2.4, since (64) verifies (10).

6 Discussion and extensions

Theorem 2.2 and Theorem 3.1 separate a sum–max asymptotic-independence proof into a generic combinatorial layer and a model-specific approximation layer. The same philosophy should extend from maxima to extreme point processes. Instead of a single family of threshold events $\mathscr{E}_{n,i}(y)$, one may work with factorial moment measures for counts in finitely many disjoint extreme-value sets. An independent-proxy criterion of the same type should then yield asymptotic independence between the sum-type statistic and the entire limiting point process. Such a result would subsume sum–max asymptotic independence by the continuous mapping theorem and would connect naturally with the point-process approach of Heiny and Kleemann (2023). We leave this extension for future work.

Acknowledgment

The author acknowledges the use of ChatGPT (OpenAI) for language editing and manuscript refinement.

References

Anderson, C. W., & Turkman, K. F. (1991). The joint limiting distribution of sums and maxima of stationary sequences. *Journal of Applied Probability, 28*(1), 33–44. https://doi.org/10.2307/3214738

Anderson, C. W., & Turkman, K. F. (1993). Limiting joint distributions of sums and maxima in a statistical context. *Theory of Probability and Its Applications, 37*(2), 314–316. https://doi.org/10.1137/1137063

Anderson, C. W., & Turkman, K. F. (1995). Sums and maxima of stationary sequences with heavy tailed distributions. *Sankhyā: The Indian Journal of Statistics, Series A, 57*(1), 1–10. https://www.jstor.org/stable/25051025

Arratia, R., Goldstein, L., & Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *The Annals of Probability, 17*(1), 9–25. https://doi.org/10.1214/aop/1176991491

Arratia, R., Goldstein, L., & Gordon, L. (1990). Poisson approximation and the Chen–Stein method. *Statistical Science, 5*(4), 403–434. https://doi.org/10.1214/ss/1177012015

Baldi, P., & Rinott, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *The Annals of Probability, 17*(4), 1646–1650. https://doi.org/10.1214/aop/1176991178

Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli, 20*(2), 1006–1028. https://doi.org/10.3150/13-BEJ514

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics, 32*(2), 485–498. https://doi.org/10.1214/aoms/1177705055

Chen, D. C., Feng, L., & Liang, D. C. (2024). Asymptotic independence of the quadratic form and maximum of independent random variables with applications to high-dimensional tests. *Acta Mathematica Sinica, English Series, 40*(12), 3093–3126. https://doi.org/10.1007/s10114-024-2498-2

Chow, T. L., & Teugels, J. L. (1979). The sum and maximum of i.i.d. random variables. In P. Mandl and M. Hušková, editors, *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, pages 81–92.

Drton, M., Han, F., & Shi, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *The Annals of Statistics, 48*(6), 3206–3227. https://doi.org/10.1214/19-AOS1926

Feng, L., Jiang, T., Li, X., & Liu, B. (2024). Asymptotic independence of the sum and maximum of dependent random variables with applications to high-dimensional tests. *Statistica Sinica, 34*(3), 1745–1763. https://doi.org/10.5705/ss.202022.0354

Gerschgorin, S. (1931). Über die Abgrenzung der Eigenwerte einer Matrix. *Bulletin de l'Académie des Sciences de l'URSS, Classe des sciences mathématiques et naturelles, 6*, 749–754. https://www.mathnet.ru/eng/im5235

Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika, 104*(4), 813–828. https://doi.org/10.1093/biomet/asx050

Harary, F. (1969). *Graph theory*. Addison-Wesley, Reading, MA.

He, Y., Xu, G., Wu, C., & Pan, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *The Annals of Statistics, 49*(1), 154–181. https://doi.org/10.1214/20-AOS1951

Heiny, J., & Kleemann, C. (2023). Asymptotic independence of point process and Frobenius norm of a large sample covariance matrix. *arXiv preprint* arXiv:2302.13914. https://arxiv.org/abs/2302.13914

Ho, H.-C., & Hsing, T. (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of Applied Probability, 33*(1), 138–145. https://doi.org/10.2307/3215271

Ho, H.-C., & McCormick, W. P. (1999). Asymptotic distribution of sum and maximum for Gaussian processes. *Journal of Applied Probability, 36*(4), 1031–1044. https://doi.org/10.1239/jap/1032374753

Hoeffding, W. (1948). A non-parametric test of independence. *The Annals of Mathematical Statistics, 19*(4), 546–557. https://doi.org/10.1214/aoms/1177730150

Hsing, T. (1995a). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *The Annals of Probability, 23*(2), 938–947. https://doi.org/10.1214/aop/1176988296

Hsing, T. (1995b). On the asymptotic independence of the sum and rare values of weakly dependent stationary random variables. *Stochastic Processes and their Applications, 60*(1), 49–63. https://doi.org/10.1016/0304-4149(95)00054-2

Janson, S. (1988). Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *The Annals of Probability, 16*(1), 305–312. https://doi.org/10.1214/aop/1176991903

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability, 14*(2), 865–880. https://doi.org/10.1214/105051604000000143

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1–2), 81–93. https://doi.org/10.1093/biomet/30.1-2.81

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics, 46*(1), 280–307. https://doi.org/10.1214/17-AOS1550

Li, W. V., & Shao, Q.-M. (2002). A normal comparison inequality and its applications. *Probability Theory and Related Fields, 122*(4), 494–508. https://doi.org/10.1007/s004400100176

Li, D., & Xue, L. (2015). Joint limiting laws for high-dimensional independence tests. *arXiv preprint* arXiv:1512.08819. https://arxiv.org/abs/1512.08819

McCormick, W. P., & Qi, Y. (2000). Asymptotic distribution for the sum and maximum of Gaussian processes. *Journal of Applied Probability, 37*(4), 958–971. https://doi.org/10.1017/S0021900200018155

Rosengard, A. (1962). Étude des lois-limites jointes et marginales de la moyenne et des valeurs extrêmes d'un échantillon. *Publications de l'Institut de Statistique de l'Université de Paris, 11*, 3–53. https://hal.science/hal-04094510

Rossberg, H. J. (1965). Die asymptotische Unabhängigkeit der kleinsten und größten Werte einer Stichprobe vom Stichprobenmittel. *Mathematische Nachrichten, 28*(5–6), 305–318. https://doi.org/10.1002/mana.19650280505

Sibuya, M. (1991). Bonferroni-type inequalities; Chebyshev-type inequalities for the distributions on $[0, n]$. *The Annals of the Institute of Statistical Mathematics, 43*(2), 261–285. https://doi.org/10.1007/BF00118635

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101. https://doi.org/10.2307/1412159

Tiago de Oliveira, J. (1962). The asymptotic independence of the sample mean and extremes. *Revista da Faculdade de Ciências de Lisboa, Série II-A, 8*(2), 299–310.

Yanagimoto, T. (1970). On measures of association and a related problem. *Annals of the Institute of Statistical Mathematics, 22*(1), 57–63. https://doi.org/10.1007/BF02506323

Zhao, P., Wang, H., & Feng, L. (2026). Rank-based tests for mutual independence of high-dimensional random vectors via L_q norm. *arXiv preprint* arXiv:2605.25380. https://arxiv.org/abs/2605.25380

Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society, 359*(11), 5345–5363. https://doi.org/10.1090/S0002-9947-07-04192-X