Sharp Anti-Concentration Inequalities for Extremum Statistics via Copulas

Matias D. Cattaneo[1] Ricardo P. Masini[2*] William G. Underwood[3]

February 10, 2025

Abstract

We derive sharp upper and lower bounds for the pointwise concentration function of the maximum statistic of d identically distributed random variables with arbitrary dependence. Our first main result places no restrictions either on the common marginal law of the samples or on the copula describing their joint distribution. We show that, in general, strictly sublinear dependence of the concentration function on the dimension d is not possible. We then introduce a new class of copulas, namely those with a convex diagonal section, and demonstrate that restricting to this class yields a sharper upper bound on the concentration function. This allows us to establish several new dimension-independent and poly-logarithmic-in-d anti-concentration inequalities for a variety of marginal distributions under mild dependence assumptions. Our theory agrees with the best-known results in certain special cases. Applications to high-dimensional statistical inference are presented, including a specific example pertaining to Gaussian mixture approximation for factor models. Our main results lead to superior distributional guarantees under certain conditions on the mixture components.

Keywords: Anti-concentration; copulas; high-dimensional probability; concentration; extreme value theory; order statistics.

MSC: Primary 60E15; Secondary 62H05, 62G32.

[1]Department of Operations Research and Financial Engineering, Princeton University.
[2]Department of Statistics, University of California, Davis.
[3]Statistical Laboratory, University of Cambridge.
[*]Corresponding author: rmasini@ucdavis.edu

Contents

1 Introduction

Concentration of measure has been extensively studied throughout the probability and statistics literature. Anti-concentration phenomena, on the other hand, appear much less frequently and are generally not so well understood (Vershynin and Rudelson, 2007). While it is impossible to pin down the date when anti-concentration precisely became a topic of interest, its systematic study is commonly attributed to Lévy (1954), who defined the concentration function of a real-valued random variable Y as $L(Y, \varepsilon) := \sup_{x\in\mathbb{R}} \mathbb{P}(x \le Y \le x + \varepsilon)$ for $\varepsilon \ge 0$. The early focus was almost exclusively on the asymptotic behavior of the concentration function as $\varepsilon \to 0$, motivated by applications to quantitative central limit theorems. The last two decades have seen a revival of interest in anti-concentration, fueled by advances in high-dimensional and nonparametric statistics (Bakshi et al., 2020; Chernozhukov et al., 2013, 2014a; Koike, 2021; Kuchibhotla et al., 2021), random matrix theory (Litvak et al., 2017; Nie, 2022), geometric analysis (Livshyts, 2014, 2021; Paouris, 2012; Paouris and Valettas, 2018) and applied probability (Aizenman et al., 2009; Belloni et al., 2024; Chernozhukov et al., 2015; Fox et al., 2021; Götze et al., 2019; Krishnapur, 2016; Meka et al., 2015; Rudelson and Vershynin, 2015). Recently, attention has shifted to finding sharp non-asymptotic upper bounds for the concentration function in terms of ε and properties of the law of Y. One particular example of interest is the maximum statistic $Y := \max_{i\in[d]} X_i$, with $X_1, \dots, X_d$ real-valued random variables.

When the distribution of Y admits a density $f(x)$ with respect to the Lebesgue measure, a simple upper bound for the concentration function is obtained by observing that $L(Y, \varepsilon) \le \varepsilon \sup_{x\in\mathbb{R}} f(x)$. This technique was applied by Chernozhukov et al. (2015, Theorem 3) to $Y := \max_{i\in[d]} X_i$ with $(X_1, \dots, X_d)$ a zero-mean multivariate Gaussian random vector with a non-singular covariance matrix. Their proof leveraged the fact that conditioning on components preserves joint Gaussianity, and the resulting anti-concentration inequality was used to establish a conditional multiplier central limit theorem in a high-dimensional regime. A related approach is to provide bounds for the concentration function in terms of the variance of Y; Bobkov and Chistyakov (2015) used this method to establish matching upper and lower bounds (up to a constant factor) under a log-concavity assumption. Unfortunately, if X_i are log-concave random variables, then there is no guarantee in general that $\max_{i\in[d]} X_i$ is similarly log-concave (refer to Saumard and Wellner (2014) for a comprehensive review of log-concavity properties). Furthermore, lower bounds on the variance of $\max_{i\in[d]} X_i$ are typically not easy to obtain unless the joint distribution of $(X_1, \dots, X_d)$ is specified. In the multivariate Gaussian setting, Giessing (2023) recently established such bounds in terms of the dimension or metric entropy of the joint distribution. Another approach builds upon the seminal paper of Nazarov (2003), establishing anti-concentration inequalities for the maximum statistic by leveraging properties of the multivariate Gaussian distribution and tools from convex geometry (Chernozhukov et al., 2017a,b).

Our goal is to study the anti-concentration behavior of maximum statistics by providing upper and lower bounds for the pointwise concentration function

$$\mathbb{P}\Big(x < \max_{i\in[d]} X_i \le x + \varepsilon\Big). \tag{1}$$

In contrast to many prior results, we refrain from taking a supremum over $x \in \mathbb{R}$, with our main results focusing on pointwise (rather than uniform) anti-concentration phenomena. In principle, this can lead to sharper inequalities when restricting to x lying in a subset of $\mathbb{R}$ (see Section 4 for an illustration). Further, we seek to impose minimal assumptions on the dependence structure of the random vector $(X_1, \dots, X_d)$, as determined by its associated copula (see Durante and Sempi, 2016, for a contemporary review). We assume throughout that the variables $X_1, \ldots, X_d$ share a common marginal distribution.

Our first main result, given as Theorem 1 in Section 2, gives upper and lower bounds for the pointwise concentration function of $\max_{i\in[d]} X_i$, as defined in (1). Crucially, this theorem makes no assumptions at all on the copula describing the dependence structure of $(X_1, \ldots, X_d)$. As such, it is applicable even in cases where the joint distribution is intractable or unspecified. Moreover, we construct copulas which exactly attain our upper and lower bounds, respectively; therefore Theorem 1 is not improvable unless extra conditions are imposed on the copula. When considering marginally Gaussian random variables (Example 1), we show that the worst-case concentration function (i.e., the maximum overall possible copulas) is substantially larger (as a function of the dimension d) than when assuming joint Gaussianity (see Chernozhukov et al., 2017b, Theorem 1). It is therefore essential in applications, particularly in high-dimensional regimes, to consider properties of the copula associated with $(X_1, \ldots, X_d)$ as well as their marginal law. The proof of Theorem 1, presented in Appendix 5, relies only on basic properties of copulas and their diagonal sections. We remark that a similar copula-based approach was taken by Frank et al. (1987), who obtained optimal upper and lower bounds for the distribution function of the sum (and other combinations) of several random variables, under an arbitrary dependence structure.

In Section 3 we obtain a more refined result as Theorem 2 by restricting the class of copulas under consideration. Specifically, we impose a convexity condition on the diagonal section of the copula; this assumption is novel, to the best of the authors' knowledge, and leads to a class of copulas that could be useful in other applications. We present an explicit copula for which our concentration function upper bound is tight, demonstrating its optimality. The resulting anti-concentration inequality for the maximum statistic is typically substantially stronger than that obtained using Theorem 1; when applied to a joint distribution with Gaussian margins, we improve several well-known results in the literature where previously a multivariate Gaussian law was assumed (cf. Chernozhukov et al., 2015, 2017b). Moreover, we demonstrate the applicability of Theorem 2 to several popular families of copulas, and discuss the resulting concentration bounds for a variety of marginal distributions.

Section 4 presents an application of our main results in high-dimensional statistical inference, highlighting the importance of sharp anti-concentration bounds in distributional analysis. We give an explicit example in the context of Gaussian mixture approximations for high-dimensional factor models. In particular, we demonstrate that our main results lead to superior anti-concentration inequalities, and therefore better guarantees on the quality of the distributional approximation, whenever the Gaussian mixture components exhibit a wide range of variances.

In Section 6 we give concluding remarks; all proofs and further details are collected in Appendix 5.

1.1 Notation

We use $\mathbb{N} := \{1, 2, \ldots\}$ for the natural numbers, and for $d \in \mathbb{N}$ we define $[d] := \{1, \ldots, d\}$. The multivariate normal distribution with mean vector μ and covariance matrix Σ is denoted by $\mathcal{N}(\mu, \Sigma)$, and the cumulative distribution function (CDF) and Lebesgue density function of $\mathcal{N}(0, 1)$ are written as Φ and ϕ respectively. The uniform distribution on $[0, 1]$ is denoted by $\mathcal{U}$. For $a, b \in \mathbb{R}$, $a \wedge b$ and $a \vee b$ are their minimum and maximum, respectively. For a function F of a single real variable, we use $F^-(x)$ for its left limit at x if it exists. For two functions f and g, we write $f \circ g(x) = f\big(g(x)\big)$ for their composition whenever it is well-defined. We use log for the natural logarithm.

1.2 Preliminary results

Suppose that $X_1, \dots, X_d$ are real-valued identically distributed random variables with a common distribution function $F : \mathbb{R} \to [0,1]$. By a well-known theorem due to Sklar (Nelsen, 2006, Theorem 2.10.9), the joint distribution of $(X_1, \dots, X_d)$ can be decomposed into a copula and a marginal distribution as $\mathbb{P}(X_1 \le x_1, \dots, X_d \le x_d) = C\big(F(x_1), \dots, F(x_d)\big)$, where C is a d-dimensional copula (Nelsen, 2006, Definition 2.10.6). Considering $x_1 = \dots = x_d$, we obtain

$$\mathbb{P}\left(x < \max_{i\in[d]} X_i \le x + \varepsilon\right) = C\big(F(x+\varepsilon), \dots, F(x+\varepsilon)\big) - C\big(F(x), \dots, F(x)\big). \tag{2}$$

As such, the distribution of $\max_{i\in[d]} X_i$ depends on the copula associated with $(X_1, \dots, X_d)$ only through its diagonal section, as formalized in Definition 1.

Definition 1. *Let $d \in \mathbb{N}$. A function $\Delta : [0,1] \to [0,1]$ is a d-*dimensional copula diagonal *if there exists a d-dimensional copula $C : [0,1]^d \to [0,1]$ with $\Delta(u) = C(u, \dots, u)$ for all $u \in [0,1]$.*

Lemma 1 below gives a simple characterization of d-dimensional copula diagonals. In Fernández-Sánchez and Úbeda-Flores (2018), an explicit copula C is constructed with a specified diagonal Δ; for our purposes, any such copula suffices by (2). See Cuculescu and Theodorescu (2001) and Jaworski (2009) for further background on copulas and their diagonals.

Lemma 1 (Theorem 1, Fernández-Sánchez and Úbeda-Flores, 2018)**.** *A function $\Delta : [0,1] \to [0,1]$ is a d-dimensional copula diagonal if and only if it satisfies: (i) $\Delta(1) = 1$; (ii) $\Delta(u) \le u$ for all $u \in [0,1]$; and (iii) $0 \le \Delta(u') - \Delta(u) \le d(u' - u)$ for all $u, u' \in [0,1]$ with $u \le u'$.*

2 Anti-concentration inequalities for arbitrary copulas

We derive sharp upper and lower bounds on the pointwise concentration function of the maximum statistic of identically distributed (not necessarily independent) random variables, imposing no further assumptions on either their common marginal law or the copula describing their joint distribution. The relevant set of distributions is specified in Definition 2.

Definition 2. *Let $d \in \mathbb{N}$ and $F : \mathbb{R} \to [0,1]$ be a CDF. Write $\mathcal{P}_d(F)$ for the set of distributions $\mathbb{P}$ on $\mathbb{R}^d$, which have joint CDFs of the form*

$$\mathbb{P}\big(X_1 \le x_1, \dots, X_d \le x_d\big) = C\big(F(x_1), \dots, F(x_d)\big),$$

for some d-dimensional copula C.

Our first main result is given in Theorem 1.

Theorem 1. *Let $d \in \mathbb{N}$ and $F : \mathbb{R} \to [0,1]$ be a CDF. For each $x \in \mathbb{R}$ and $\varepsilon \ge 0$,*

$$\max_{\mathbb{P}\in\mathcal{P}_d(F)} \mathbb{P}\left(x < \max_{i\in[d]} X_i \le x + \varepsilon\right) = \big\{d\big(F(x+\varepsilon) - F(x)\big)\big\} \wedge F(x+\varepsilon), \tag{3}$$

$$\min_{\mathbb{P}\in\mathcal{P}_d(F)} \mathbb{P}\left(x < \max_{i\in[d]} X_i \le x + \varepsilon\right) = 0 \vee \big\{1 - F(x) - d\big(1 - F(x+\varepsilon)\big)\big\}. \tag{4}$$

Equation (3) in Theorem 1 gives a tight upper bound on the probability of the maximum statistic falling in $(x, x+\varepsilon]$. Further, (3) shows that if $F(x) \in (0,1)$ and $F(x+\varepsilon) - F(x) \le \frac{F(x)}{d-1}$, then there exists a joint distribution $\mathbb{P}_{\text{up}}$ such that the maximum statistic exhibits strong local concentration near x. That is, $\mathbb{P}_{\text{up}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = d\big(F(x+\varepsilon) - F(x)\big)$, which increases linearly with the dimension d; see (7) in Section 2.1. If also F admits a Lebesgue density f on $(x, x+\varepsilon]$ which is bounded above by M and below by m, then $\varepsilon \le \frac{F(x)}{M(d-1)}$ implies $dm\varepsilon \le \mathbb{P}_{\text{up}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) \le dM\varepsilon$. This is a form of "curse of dimensionality," precluding the possibility of obtaining anti-concentration bounds which hold uniformly in ε and depend strictly sublinearly on d. It is worth noting that this local concentration phenomenon can occur at any point $x \in \mathbb{R}$ satisfying $F(x) > 0$; contrast this with the independent setting, in which concentration is restricted to regions where $F(x)$ is close to 1.

For a lower bound on the concentration probability, (4) establishes a joint distribution $\mathbb{P}_{\text{lo}}$ which achieves perfect anti-concentration whenever $F(x+\varepsilon) - F(x) \le \frac{d-1}{d}\big(1 - F(x)\big)$, in the sense that $\mathbb{P}_{\text{lo}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = 0$. If F admits a Lebesgue density f on $(x, x+\varepsilon]$ which is bounded above by M, then $\varepsilon \le \frac{d-1}{Md}\big(1 - F(x)\big)$ suffices to ensure that $\mathbb{P}_{\text{lo}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = 0$; see (8) in Section 2.1.

In Example 1 we apply the result from (3) with marginally Gaussian random variables.

Example 1 (Normal distribution)**.** *Let $d \in \mathbb{N}$, $\sigma > 0$ and $\varepsilon \in [0, \sigma]$. By (3), there exists $(X_1, \ldots, X_d)$ with $X_i \sim \mathcal{N}(0, \sigma^2)$ for $i \in [d]$ such that*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(x < \max_{1\le i\le d} X_i \le x+\varepsilon\right) \ge \frac{d\varepsilon}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right) \wedge \Phi\left(\frac{\varepsilon}{\sigma}\right) \ge \frac{d\varepsilon}{\sigma}\frac{e^{-1/2}}{\sqrt{2\pi}} \wedge \frac{1}{2} \ge \frac{d\varepsilon}{5\sigma} \wedge \frac{1}{2}.$$

Compare Example 1 with Nazarov's inequality (Nazarov, 2003; see also Chernozhukov et al., 2017b, for a detailed proof) which, under the assumption of joint Gaussianity, obtains a bound of

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(x < \max_{1\le i\le d} X_i \le x+\varepsilon\right) \le \frac{\varepsilon}{\sigma}\Big(\sqrt{2\log d} + 2\Big). \tag{5}$$

This slow-growing dependence on d is crucial in high-dimensional statistical applications, where the dimension may be much larger than the sample size. Example 1 shows that marginal Gaussianity of each X_i alone is insufficient for obtaining such a bound. Therefore, in the upcoming Section 3 we present a restricted class of copulas for which sharper anti-concentration inequalities hold than those given in Theorem 1. We thus recover a form of Nazarov's inequality as a special case.

2.1 Overview of proof strategy

The proof of Theorem 1 is presented in Appendix 5 and proceeds as follows. Firstly, we consider the special case where the common distribution of the variables is the standard uniform distribution, and write U_i instead of X_i for clarity. In this setting, the joint law of $(U_1, \ldots, U_d)$ is a d-dimensional copula $C : [0,1]^d \to [0,1]$. With $\Delta(u) := C(u, \ldots, u)$ defined to be the diagonal section of C, we have for any $u \in [0,1]$ and $\delta \in [0, 1-u]$ that

$$\mathbb{P}\left(u < \max_{i\in[d]} U_i \le u+\delta\right) = \Delta(u+\delta) - \Delta(u). \tag{6}$$

Establishing (3) and (4) thus reduces to finding Δ_{up} and Δ_{lo} which maximize and minimize the right-hand side of (6) respectively over Δ, subject to the constraints enforced in Lemma 1. The resulting copula diagonals are described in (7) and (8), and are plotted in Figure 1. Fix $d \in \mathbb{N}$ and

$u \in [0,1]$, and for $t \in [0,1]$ define

$$\Delta_{\mathrm{up}}(t) := d \cdot \left\{ t - \left(u \wedge \frac{d-1}{d} \right) \right\} \cdot \mathbb{I}\left\{ u \wedge \frac{d-1}{d} < t \leq \frac{du}{d-1} \right\} + t \cdot \mathbb{I}\left\{ \frac{du}{d-1} \wedge 1 < t \right\}, \tag{7}$$

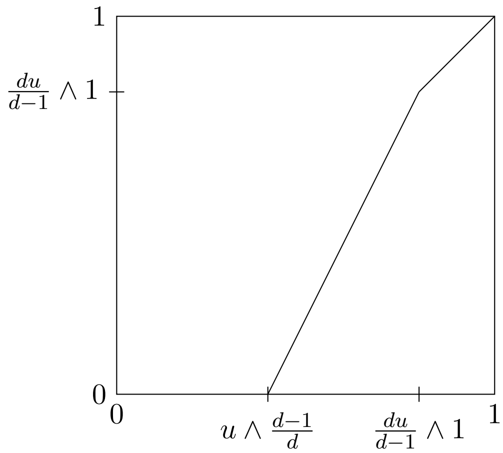
$$\Delta_{\mathrm{lo}}(t) := t \cdot \mathbb{I}\{t \leq u\} + u \cdot \mathbb{I}\left\{ u < t \leq \frac{d+u-1}{d} \right\} + (1-d+d\cdot t) \cdot \mathbb{I}\left\{ \frac{d+u-1}{d} < t \right\}. \tag{8}$$

For the upper bound (7, Figure 1a), we maximize the increment of Δ over $(u, u+\delta]$ to obtain Δ_{up}; for the lower bound ((8), Figure 1b), we minimize it, yielding Δ_{lo}. Therefore,

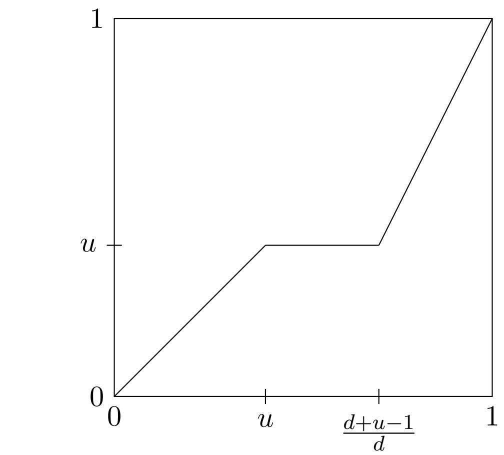
$$\mathbb{P}_{\mathrm{up}}\Big(u < \max_{i \in [d]} U_i \leq u + \delta \Big) = (d\delta) \wedge (u+\delta),$$
$$\mathbb{P}_{\mathrm{lo}}\Big(u < \max_{i \in [d]} U_i \leq u + \delta \Big) = 0 \vee \big(1 - u - d(1-u-\delta)\big).$$

The generalization to an arbitrary distribution function F then proceeds by a quantile transform, taking $u = F(x)$ and $\delta = F(x+\varepsilon) - F(x)$, and finally setting $X_i = F^{-1}(U_i)$.

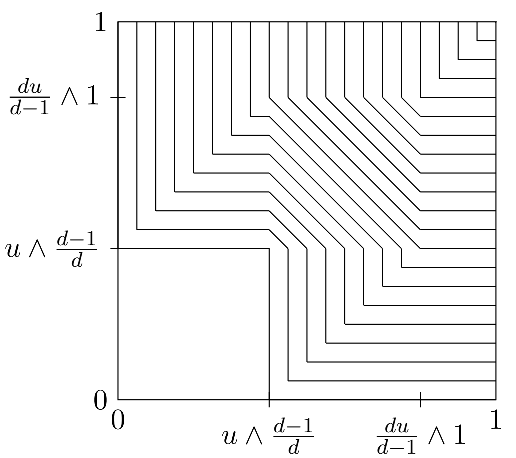
Analogous results to those established in Theorem 1 can be derived with the maximum statistic replaced by the minimum statistic by considering the variables $-X_i$, which have common CDF given by $G(x) = 1 - F^{-}(-x)$. Likewise, if F is symmetric about zero in the sense that $F(x) = 1 - F^{-}(-x)$ for all $x \in \mathbb{R}$, then similar results can be established for the maximum absolute value statistic (see Example 9) by observing that $\max_{i\in[d]} |X_i| = \max_{i\in[d]}(X_i \vee -X_i)$ and applying Theorem 1 to the $2d$-dimensional vector $(X_1, \ldots, X_d, -X_1, \ldots, -X_d)$.



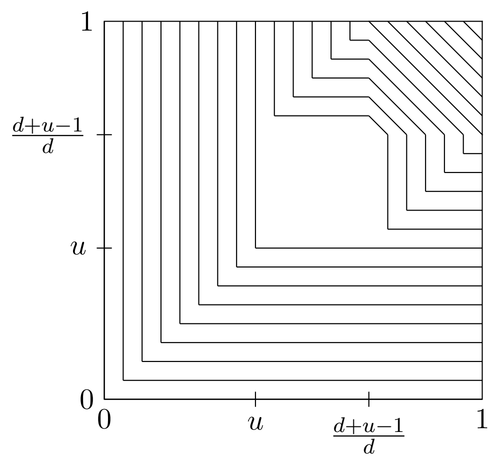
(a) The copula diagonal Δ_{up} defined in (7).



(b) The copula diagonal Δ_{lo} defined in (8).



(c) Extension of Δ_{up} to a copula with $d = 2$.



(d) Extension of Δ_{lo} to a copula with $d = 2$.

Figure 1: Top: the two d-dimensional copula diagonals (7) and (8) constructed to prove (3) and (4) respectively in Theorem 1. For the upper bound (a), the increment over $(u, u + \delta]$ is maximized, while for the lower bound (b) it is minimized. Bottom: contour plots for possible two-dimensional ($d = 2$) copulas (a) and (b) whose diagonals are given by Δ_{up} and Δ_{lo} respectively. We use the extension due to Fernández-Sánchez and Úbeda-Flores (2018, proof of Theorem 1), though this is not unique in general. Recall that every copula satisfies $C(0, \ldots, 0) = 0$ and $C(1, \ldots, 1) = 1$.

2.2 Comparisons with other well-known copulas

We provide some comparisons of the concentration properties established in Theorem 1 with those induced by other well-known copulas. For simplicity, we restrict to the case that $X_i \sim \mathcal{U}$ are uniformly distributed; extensions to arbitrary common laws proceed using a straightforward quantile transform, as in the proof of Theorem 1.

Example 2 (Independence copula)**.** *If $X_i \sim \mathcal{U}$ are independent for $i \in [d]$, then we have that*

$\mathbb{P}_{\mathrm{ind}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = (x+\varepsilon)^d - x^d$. *Taking* $x \in (0,1)$ *and* $\varepsilon \le \frac{x}{d-1} \wedge (1-x)$,

$$\mathbb{P}_{\mathrm{ind}}\left(x < \max_{i\in[d]} X_i \le x+\varepsilon\right) \le d\varepsilon(x+\varepsilon)^{d-1} \le d\varepsilon x^{d-1}\Big(1+\frac{1}{d-1}\Big)^{d-1} \le ed\varepsilon x^{d-1}.$$

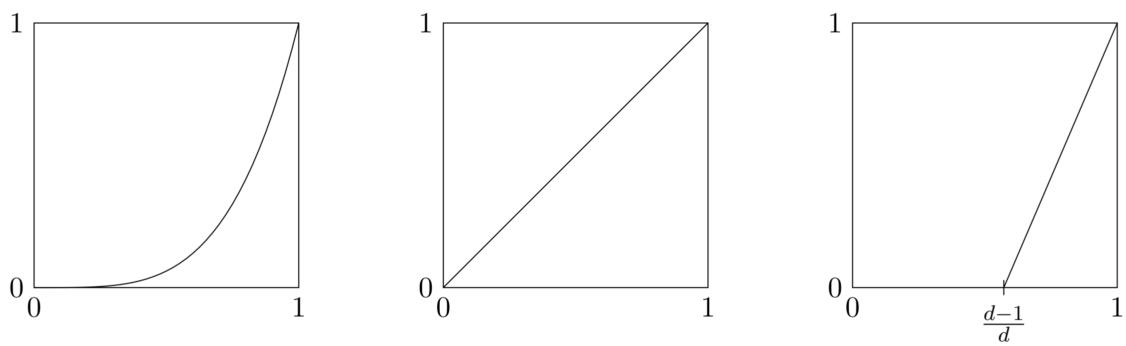
In contrast, the law $\mathbb{P}_{\mathrm{up}}$ *attaining the maximum in* (3) *satisfies* $\mathbb{P}_{\mathrm{up}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = d\varepsilon$; *its local concentration probability is greater by a factor of at least* $x^{1-d}/e \to \infty$ *as* $d \to \infty$.

If instead one takes $\varepsilon \in (1/d, 1)$ *and* $x = 1-\varepsilon$, *then for the independence copula one obtains* $\mathbb{P}_{\mathrm{ind}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = 1-(1-\varepsilon)^d$ *while* (3) *gives* $\mathbb{P}_{\mathrm{up}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = 1$. *Although both exhibit concentration of the maximum statistic at* $x=1$ *as expected, the independence copula does not attain exact concentration.*

Regarding lower bounds, if $x \in (0,1)$ *and* $\varepsilon \in (0, 1-x]$, *then the independence copula gives* $\mathbb{P}_{\mathrm{ind}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) \ge d\varepsilon x^{d-1} > 0$. *In contrast, whenever* $\varepsilon \le \frac{d-1}{d}(1-x)$, *the law* $\mathbb{P}_{\mathrm{lo}}$ *achieving the minimum in* (4) *satisfies* $\mathbb{P}_{\mathrm{lo}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = 0$.

Example 3 (Fréchet–Hoeffding upper bound)**.** *Write* $\mathbb{P}_{\mathrm{FHU}}$ *for the joint law of* $X_1 = \cdots = X_d \sim \mathcal{U}$. *For* $x \in [0,1)$ *and* $\varepsilon \in [0, 1-x]$, *we have* $\mathbb{P}_{\mathrm{FHU}}\big(x < \max_{i\in[d]} \le x+\varepsilon\big) = \varepsilon$. *Since* $0 \vee \big(1-x-d(1-x-\varepsilon)\big) \le \varepsilon \le (d\varepsilon) \wedge (x+\varepsilon)$, *we deduce that the Fréchet–Hoeffding upper bound copula interpolates between the upper bound* (3) *and the lower bound* (4) *of Theorem 1.*

Example 4 (Fréchet–Hoeffding lower bound)**.** *Let* $\mathbb{P}_{\mathrm{FHL}}$ *be any joint distribution of* $X_i \sim \mathcal{U}$ *for* $i \in [d]$ *with copula diagonal* $\Delta(x) = 0 \vee (d\cdot x - d + 1)$. *If* $x \ge \frac{d-1}{d}$ *and* $\varepsilon \in [0, 1-x]$, *then* $\mathbb{P}_{\mathrm{FHL}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = d\varepsilon$, *matching* (3). *Conversely, if* $x \in \big[0, \frac{d-1}{d}\big]$ *and* $\varepsilon \in \big[0, \frac{d-1}{d} - x\big]$, *then* $\mathbb{P}_{\mathrm{FHL}}\big(x < \max_{i\in[d]} X_i \le x+\varepsilon\big) = 0$, *agreeing with* (4).



(a) Independence (b) Upper Fréchet–Hoeffding (c) Lower Fréchet–Hoeffding

Figure 2: The diagonal sections of three well-known d-dimensional copulas.

3 Anti-concentration inequalities for diagonally convex copulas

The upper bound presented as (3) in Theorem 1 demonstrates that, without imposing further conditions on the dependence structure (the copula) of the random vector $(X_1, \ldots, X_d)$, it is impossible to obtain anti-concentration results which hold uniformly over $\varepsilon \ge 0$ and exhibit a strictly sublinear dependence on the dimension (see Example 1). As such, in order to obtain sharper upper bounds on the concentration probability, it is necessary to restrict the class of admissible copulas. For example, as discussed in Section 2, in the setting where $(X_1, \ldots, X_d)$ follows a multivariate

Gaussian law, Nazarov's inequality can be applied to the maximum statistic and produces a bound (5) with a poly-logarithmic dependence on the dimension.

Nonetheless, in this section, we propose a method which avoids the assumption of multivariate (joint) Gaussianity, replacing it with a mild nonparametric convexity condition on the copula describing the dependence structure (see Definition 3). We also allow for an arbitrary common marginal distribution; as such, we encompass a substantially wider range of joint distributions than those covered by Nazarov's inequality. See the upcoming Examples 8, 9, 10, 11 and 12 for a selection of novel anti-concentration results which can be derived using our results.

Definition 3. *Let $d \in \mathbb{N}$ and $F : \mathbb{R} \to [0,1]$ be a CDF. Write $\mathcal{P}_d^c(F)$ for the set of distributions $\mathbb{P}$ on $\mathbb{R}^d$ that have joint CDFs of the form*

$$\mathbb{P}\big(X_1 \le x_1, \ldots, X_d \le x_d\big) = C\big(F(x_1), \ldots, F(x_d)\big),$$

where C is a d-dimensional copula for which $\Delta : [0,1] \to [0,1]$ defined by $\Delta(x) = C(x, \ldots, x)$ is a convex function. We say that $\mathbb{P}$ and C are diagonally convex.

Theorem 2. *Let $d \in \mathbb{N}$ and $F : \mathbb{R} \to [0,1]$ be a CDF. For each $x \in \mathbb{R}$ and $\varepsilon \ge 0$,*

$$\max_{\mathbb{P} \in \mathcal{P}_d^c(F)} \mathbb{P}\left(x < \max_{i \in [d]} X_i \le x + \varepsilon\right) = \big(F(x+\varepsilon) - F(x)\big)\left\{\frac{1}{1 - F(x)} \wedge d\right\}.$$

The upper bound given in Theorem 2 holds uniformly over all diagonally convex copulas and, moreover, imposes no conditions on the common marginal law F. The proof of Theorem 2 is presented in Appendix 5 and relies only on convexity arguments.

3.1 Examples of diagonally convex copulas

Before applying Theorem 2 with some explicit marginal laws, we first verify that several popular copula families satisfy the convex diagonal section condition given in Definition 3. Naturally, the two copulas (3) and (4) constructed in Theorem 1 do *not* generally satisfy this assumption, as evidenced by the plots of their diagonal sections presented in Figure 1. We verify in Example 5 that diagonal convexity does hold for the independence copula, the Fréchet–Hoeffding upper bound copula, and any copula with diagonal section matching the Fréchet–Hoeffding lower bound; see Figure 2.

Example 5 (Diagonally convex copulas)**.** *The d-dimensional independence copula has diagonal section $\Delta_{\mathrm{ind}}(u) = u^d$ and is diagonally convex. Similarly, the d-dimensional Fréchet–Hoeffding upper bound copula has diagonal $\Delta_{\mathrm{FHU}}(u) = u$ and is diagonally convex. Any copula with diagonal section matching the d-dimensional Fréchet–Hoeffding lower bound satisfies $\Delta_{\mathrm{FHL}}(u) = 0 \vee (du - d + 1)$ and is diagonally convex.*

Next, Lemma 2 demonstrates that every multivariate Gaussian copula is diagonally convex. The proof of this result is given in Appendix 5, and depends on a precise characterization of the Lebesgue density associated with the maximum statistic of a multivariate Gaussian distribution (Chernozhukov et al., 2015, Lemmas 5 and 6).

Lemma 2. *Let $\mu \in \mathbb{R}^d$ and suppose $\Sigma \in \mathbb{R}^{d \times d}$ is a symmetric, positive semi-definite matrix. Then $\mathcal{N}(\mu, \Sigma)$ has a diagonally convex copula.*

We now give a general condition under which every member of a family of Archimedean copulas possesses a convex diagonal section. A d-dimensional copula C is said to be *Archimedean* (Nelsen, 2006, Theorem 4.6.2) if there is a continuous strictly decreasing function $\psi : [0,1] \to [0,\infty]$ with $\psi(0) = \infty$ and $\psi(1) = 0$ satisfying

$$C(x_1, \dots, x_d) = \psi^{-1}\left(\sum_{i=1}^{d} \psi(x_i)\right) \tag{9}$$

for all $(x_1, \dots, x_d) \in [0,1]^d$. The function ψ is known as the *generator* of C. Since our focus is on high-dimensional phenomena, we consider only strict Archimedean generators with completely monotone inverse functions; such generators yield valid copulas through (9) for every $d \in \mathbb{N}$.

Lemma 3. *Let C be a d-dimensional Archimedean copula with generator ψ which is differentiable on $(0,1)$ with $\psi'(x) < 0$ for all $x \in (0,1)$. Assume that $\Psi : (0,\infty) \to \mathbb{R}$, defined by*

$$\Psi(x) := \frac{d \cdot \psi' \circ \psi^{-1}(x)}{\psi' \circ \psi^{-1}(d \cdot x)} = \frac{\big(\psi^{-1}\big)'(d \cdot x)}{\big(\psi^{-1}\big)'(x)},$$

is non-increasing. Then C is diagonally convex.

We verify in the next example that several popular families of Archimedean copulas (Nelsen, 2006, Examples 4.23–4.25) satisfy the conditions of Lemma 3, and hence are diagonally convex. The details are contained in Appendix 5.

Example 6 (Archimedean copulas)**.** *The Clayton copulas are Archimedean with generator $\psi(x) = x^{-\theta} - 1$ for $\theta > 0$. Likewise, the Frank copulas have generator $\psi(x) = \log \frac{e^{-\theta}-1}{e^{-\theta x}-1}$ for $\theta > 0$, and the Gumbel–Hougaard copulas have generator $\psi(x) = (-\log x)^{\theta}$ for $\theta \geq 1$. All of these are diagonally convex.*

As a final example of a method for constructing diagonally convex copulas, we consider a model based on mixtures of copulas. This approach has applications in dependence-based clustering (Arakelian and Karlis, 2014).

Example 7 (Mixture copula)**.** *Take $K, d \in \mathbb{N}$ and suppose $p_1, \dots, p_K \geq 0$ are such that $\sum_{k=1}^{K} p_k = 1$. For each $k \in [K]$, let C_k be a d-dimensional diagonally convex copula. Then for $(x_1, \dots, x_d) \in [0,1]^d$, the mixture copula $(x_1, \dots, x_d) \mapsto \frac{1}{K} \sum_{k=1}^{K} p_k C_k(x_1, \dots, x_d)$ is diagonally convex.*

3.2 Examples with specific marginal distributions

Having established the existence of several copulas with convex diagonal sections, we now demonstrate the application of Theorem 2 with a selection of different common marginal laws. In Example 8 we consider the normal distribution; see Appendix 5 for details.

Example 8 (Marginal Gaussian distribution)**.** *Let $d \in \mathbb{N}$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Take $(X_1, \dots, X_d)$ a random vector with $X_i \sim \mathcal{N}\big(\mu, \sigma^2\big)$ for each $i \in [d]$, and suppose it has a diagonally convex copula. Then for each $x \in \mathbb{R}$ and $\varepsilon \geq 0$,*

$$\mathbb{P}\left(x < \max_{i \in [d]} X_i \leq x + \varepsilon\right) \leq \frac{\varepsilon}{\sigma}\left(\sqrt{2 \log d} + 1\right).$$

We do not require $(X_1, \dots, X_d)$ to be jointly Gaussian in Example 8; any copula with a convex diagonal section will suffice. In particular, a square root-logarithmic dependence on the dimension holds regardless of the structural form of the copula; for example, any of the copulas described in

Examples 5, 6 and 7 are permitted. Therefore, this example offers a version of Nazarov's inequality (Chernozhukov et al., 2017b, Theorem 1) for non-Gaussian joint distributions.

In particular, combining Example 8 with Lemma 2 allows us to deduce the following result for the maximum of identically distributed and jointly Gaussian random variables (and also for their maximum absolute deviation from the mean). We recover a form of Nazarov's inequality (Chernozhukov et al., 2017b, Theorem 1) with an improved constant.

Example 9 (Jointly Gaussian distribution)**.** *Let $d \in \mathbb{N}$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Suppose $(X_1, \dots, X_d)$ is multivariate Gaussian, with $X_i \sim \mathcal{N}(\mu, \sigma^2)$ for each $i \in [d]$. Then for any $x \in \mathbb{R}$ and $\varepsilon \geq 0$,*

$$\mathbb{P}\Big(x < \max_{i \in [d]} X_i \leq x + \varepsilon\Big) \leq \frac{\varepsilon}{\sigma}\Big(\sqrt{2 \log d} + 1\Big),$$

$$\mathbb{P}\Big(x < \max_{i \in [d]} |X_i - \mu| \leq x + \varepsilon\Big) \leq \frac{\varepsilon}{\sigma}\Big(\sqrt{2 \log 2d} + 1\Big).$$

More generally, in cases where F admits a decreasing Lebesgue density f, the dimension-dependence of an anti-concentration bound derived using Theorem 2 is determined by the quantity $H(x) := h(x) \wedge \{d \cdot f(x)\}$, where $h(x) := f(x)/\big(1 - F(x)\big)$ is the hazard function (or inverse Mills ratio) associated with F. Typically, if h is an increasing function (sometimes referred to as an "increasing failure rate" condition), then the maximum value of $H(x)$ is attained at a point x^* with $h(x^*) = d \cdot f(x^*)$, or equivalently with $x^* = F^{-1}(1 - 1/d)$, yielding a uniform upper bound of $H(x) \leq d \cdot f\big(F^{-1}(1 - 1/d)\big)$. If the hazard function h is instead decreasing (known as a "decreasing failure rate" condition), then generally, a dimension-independent bound is obtained.

In Example 10 we apply Theorem 2 to a family of Weibull distributions.

Example 10 (Weibull distribution)**.** *Let $d \in \mathbb{N}$, $\alpha \geq 1$ and $\lambda > 0$. Suppose $(X_1, \dots, X_d)$ is a random vector with a diagonally convex copula, and $\mathbb{P}(X_i \leq x) = 1 - \exp\big(-(x/\lambda)^\alpha\big)$ for $x \geq 0$ and $i \in [d]$. Then for each $x \geq 0$ and $\varepsilon \geq 0$,*

$$\mathbb{P}\Big(x < \max_{i \in [d]} X_i \leq x + \varepsilon\Big) \leq \frac{\varepsilon \alpha}{\lambda}\big(\log d + 1\big)^{\frac{\alpha - 1}{\alpha}}.$$

The dimension dependence in Example 10 is poly-logarithmic, with the exponent depending on the value of the shape parameter α. With $\alpha = 1$, we recover the exponential distribution, and the bound reduces to ε/λ. This dimension-independent result arises because the hazard function is constant. For $\alpha > 1$, the hazard function is increasing, yielding a dimension-dependent bound. When $\alpha = 2$, we recover a Rayleigh distribution and the dimension dependence scales as $\sqrt{\log d}$; the same as for the Gaussian distribution (Example 8).

Next, we consider a family of reverse Gumbel distributions.

Example 11 (Gumbel distribution)**.** *Let $d \in \mathbb{N}$ and $\lambda > 0$. Suppose $(X_1, \dots, X_d)$ is a random vector with a diagonally convex copula and $\mathbb{P}(X_i \leq x) = 1 - \exp\big(-e^{x/\lambda}\big)$ for $x \in \mathbb{R}$ and $i \in [d]$. For any $x \in \mathbb{R}$ and $\varepsilon \geq 0$,*

$$\mathbb{P}\Big(x < \max_{i \in [d]} X_i \leq x + \varepsilon\Big) \leq \frac{\varepsilon}{\lambda}\big(1 + \log d\big).$$

For such reverse Gumbel distributions, the hazard function is increasing, giving a dimension-dependent bound, here on the order of $\log d$.

In the final example, we consider a family of Pareto distributions.

Example 12 (Pareto distribution)**.** *Let $d \in \mathbb{N}$, $\alpha > 0$ and $\lambda > 0$. Suppose $(X_1, \dots, X_d)$ is a random vector with a diagonally convex copula and $\mathbb{P}(X_i \le x) = 1 - (\lambda/x)^\alpha$ for $x \ge \lambda$. For $x \ge 0$ and $\varepsilon \ge 0$,*

$$\mathbb{P}\left(x < \max_{i \in [d]} X_i \le x + \varepsilon\right) \le \frac{\alpha\varepsilon}{\lambda}.$$

Since Pareto distributions have decreasing hazard functions, the resulting bound in Example 12 is dimension-independent.

4 Application to high-dimensional statistical inference

We illustrate the applicability of our results with an example of a statistical inference using a potentially high-dimensional test statistic. Let X be an $\mathbb{R}^d$-valued random vector constructed using samples taken from an underlying data set. For example, X might represent (an appropriate transformation of) the fitted coefficients of a parametric model or a discretized version of a nonparametric estimator. Since weak convergence of the law of X routinely fails in high-dimensional settings, we suppose instead that a coupling (strong approximation) for X is available (see, for example, Chernozhukov et al., 2013, 2014a,b; Cattaneo et al., 2022, 2024; Cattaneo and Yu, 2025, and references therein). That is, there exists an $\mathbb{R}^d$-valued random vector $T = (T_1, \dots, T_d)$, on the same probability space as $X = (X_1, \dots, X_d)$, which satisfies

$$\mathbb{P}\big(\|X - T\|_\infty > \varepsilon\big) \le p(\varepsilon)$$

for some decreasing function $p : [0, \infty) \to [0, 1]$, where $\|x\|_\infty := \max_{1 \le i \le d} |x_i|$. It is typically the case that either one knows the law of T explicitly, or that one can draw samples from it. Inference proceeds by choosing a significance level $\alpha \in (0, 1)$ and computing a quantile $q_\alpha := \inf\big\{q \in \mathbb{R} : \mathbb{P}(\max_{1 \le i \le d} T_i \le q) \ge 1 - \alpha\big\}$. It is straightforward to verify that for all $\varepsilon \ge 0$,

$$\left|\mathbb{P}\Big(\max_{1 \le i \le d} X_i > q_\alpha\Big) - \alpha\right| \le p(\varepsilon) + \left\{\mathbb{P}\Big(q_\alpha - \varepsilon < \max_{1 \le i \le d} T_i \le q_\alpha\Big) \vee \mathbb{P}\Big(q_\alpha < \max_{1 \le i \le d} T_i \le q_\alpha + \varepsilon\Big)\right\};$$

this bound can then be minimized over $\varepsilon \ge 0$. In this setting, it suffices to control the anti-concentration terms in a neighborhood of the quantile q_α; our Theorem 2 provides sharper bounds than those derived using the Lévy concentration function $L(\max_{1 \le i \le d} T_i, \varepsilon)$.

As such, the validity of the resulting test based on X relies on the availability of both (i) a tight coupling inequality for $\|T - X\|_\infty$, and (ii) a sharp anti-concentration bound for $\max_{1 \le i \le d} T_i$.

Our main anti-concentration results given in Section 2 and Section 3 can be applied whenever the entries $T_1, \dots, T_d$ are identically distributed. While this is, in principle, a restrictive assumption, it is often satisfied in practice for the purpose of maximizing statistical power against alternative hypotheses. For example, consider the canonical setting in which the law of X is approximated using a multivariate Gaussian random vector T. It is usual to standardize the entries of X (and therefore also the corresponding entries of T) so that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\big[X_i^2\big] = 1$; it follows immediately that each T_i is distributed marginally as $\mathcal{N}(0, 1)$.

Our results are applicable to a substantially broader class of inference procedures than existing approaches, which typically require joint Gaussianity of the entries of T. Specifically, under Theorem 2 we allow for an arbitrary common marginal distribution as well as a wide range of joint distributions, so long as the underlying copula is diagonally convex (Definition 3). Moreover, the resulting anti-concentration inequality is agnostic to the precise form of the copula, allowing for settings where the joint distribution of T is unknown or difficult to estimate.

4.1 Gaussian mixture approximations for factor models

For a specific example in a non-Gaussian regime, we consider the Gaussian mixture approximation for martingale factor models discussed in Cattaneo et al. (2022, Section 2.5). In this setting, a size-n sample of d-dimensional observations $\big(X^{(1)}, \ldots, X^{(n)}\big)$ is taken from an underlying random process and is assumed to form a zero-mean martingale. For example, $X^{(1)}, \ldots, X^{(n)}$ may denote multivariate quantities derived from a time series estimation procedure. We impose further structure by means of a factor model; take $m \in \mathbb{N}$ and suppose that

$$X^{(i)} = Lg^{(i)} + \varepsilon^{(i)}$$

for $i \in [n]$, where L takes values in $\mathbb{R}^{d\times m}$, $g^{(i)}$ in $\mathbb{R}^m$, and $\varepsilon^{(i)}$ in $\mathbb{R}^d$. We interpret $g^{(i)}$ as a latent factor variable and L as a random factor loading, with independent disturbances $\varepsilon^{(i)}$. We assume that $\varepsilon^{(i)}$ is zero-mean and finite-variance for each $i \in [n]$, and that $\big(\varepsilon^{(1)}, \ldots, \varepsilon^{(n)}\big)$ is independent of L and $\big(g^{(1)}, \ldots, g^{(n)}\big)$. Suppose that $\mathbb{E}\big[g^{(i)} \mid L, g^{(1)}, \ldots, g^{(i-1)}\big] = 0$ for each $i \in [n]$.

Corollary 2.2 in Cattaneo et al. (2022) provides sufficient conditions for a coupling between $\sum_{i=1}^n X^{(i)}$ and an $\mathbb{R}^d$-valued random vector T with conditional distribution $T \mid L \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma := \sum_{i=1}^n \Big(L \operatorname{Var}\big[g^{(i)} \mid L\big] L^\mathsf{T} + \operatorname{Var}\big[\varepsilon^{(i)}\big] \Big).$$

We now impose some further conditions on the law of T; in particular, we ensure that the copula associated with T is the independence copula and that the entries T_j share a common marginal distribution. Therefore, suppose that the Lebesgue density function $f : \mathbb{R}^d \to [0, \infty)$ of T satisfies

$$f(x_1, \ldots, x_d) = \prod_{j=1}^d \left(\sum_{k=1}^K \frac{p_k}{\sigma_k} \phi\left(\frac{x_j}{\sigma_k}\right) \right) = \sum_{k_1=1}^K \cdots \sum_{k_d=1}^K \left(\prod_{j=1}^d p_{k_j} \right) \left(\prod_{j=1}^d \frac{1}{\sigma_{k_j}} \phi\left(\frac{x_j}{\sigma_{k_j}}\right) \right), \tag{10}$$

where $p_k \in [0,1]$ with $\sum_{k=1}^K p_k = 1$ and $\sigma_k > 0$, for $k \in [K]$. The variables $T_1, \ldots, T_d$ are seen to be independent and identically distributed by the factorization given in the first equality in (10). Further, T follows a multivariate Gaussian mixture distribution with K^d components by the second equality in (10); in general, T is not a Gaussian random vector. The copula associated with T is, therefore, the independence copula, which is diagonally convex by Example 5. In Example 13, we present an anti-concentration bound based on Theorem 2 for the maximum statistic of such a distribution, using the result from Example 8 and the structure of the multivariate mixture model.

Example 13 (Gaussian mixture with independent entries). *Let $d, K \in \mathbb{N}$ and suppose $(T_1, \ldots, T_d)$ has Lebesgue density as given in (10), where $p_k \in (0,1]$ with $\sum_{k=1}^K p_k = 1$ and $0 < \sigma_k \leq 1$, for $k \in [K]$. We assume that $\sigma_1 = 1$, and write $\sigma := \min_{1\leq k\leq K} \sigma_k$. Then for any $x \in \mathbb{R}$ and $\varepsilon \geq 0$,*

$$\mathbb{P}\left(x < \max_{j\in[d]} T_j \leq x + \varepsilon \right) \leq \left\{ \frac{\varepsilon}{p_1} \left(\sqrt{2\log d} + 2 \sum_{k=1}^K \frac{p_k}{\sigma_k} \right) \right\} \wedge \left\{ \frac{\varepsilon}{\sigma} \Big(\sqrt{2\log d} + 1 \Big) \right\}. \tag{11}$$

By applying Nazarov's inequality (5) instead of our Theorem 2, one can show that (11) admits an upper bound of $\frac{\varepsilon}{\sigma}\big(\sqrt{2\log d} + 2\big)$. As such, our Example 13 offers a sharper anti-concentration inequality, especially when the minimum variance σ is small. For instance, consider the setting where $\sigma_k = 1$ for $k \in [K-1]$ and $\sigma_K = \sigma \leq 1$, and suppose for simplicity that $p_k = 1/K$ for all $k \in [K]$. Then Example 13 gives

$$\mathbb{P}\left(x < \max_{j\in[d]} T_j \leq x + \varepsilon \right) \leq K\varepsilon \left(\sqrt{2\log d} + \frac{2(K-1)}{K} + \frac{2}{K\sigma} \right).$$

Crucially, the dependence on the dimension d (which may be large in high-dimensional regimes) and the minimum variance σ^2 (which could be small in the presence of approximately degenerate Gaussian mixture components) is additive; contrast this with the corresponding bound from Nazarov's inequality, which contains the multiplicative term $\frac{\varepsilon}{\sigma}\sqrt{2\log d}$.

Such a separation between the high-dimensional and the degenerate variance terms is to be expected in the resulting anti-concentration inequality for the following reason. Either (i) d is large: then since T_i are i.i.d., the maximum statistic will be realized far from the origin with high probability; here, the marginal density is dominated by the component with the largest (unit) variance; or (ii) d is small; then the maximum statistic may realize near zero, where the component with the smallest variance dominates the density. This approach is similar to that taken by Lopes et al. (2020) in the context of bootstrap approximations under variance decay.

5 Proofs

We give full proofs for all of our results, along with detailed calculations associated with examples given in the main paper.

5.1 Theorem 1

Before proving Theorem 1, we first establish the result in the special case where the common law is the standard uniform distribution, in Lemma 4. For clarity, we write $U_i \sim \mathcal{U}$ for each $i \in [d]$.

Lemma 4. *Let $d \in \mathbb{N}$. For $u \in [0,1]$ and $\delta \in [0, 1-u]$,*

$$\max_{\mathbb{P}\in\mathcal{P}_d(\mathcal{U})} \mathbb{P}\left(u < \max_{i\in[d]} U_i \le u+\delta\right) = (d\delta) \wedge (u+\delta), \tag{12}$$

$$\min_{\mathbb{P}\in\mathcal{P}_d(\mathcal{U})} \mathbb{P}\left(u < \max_{i\in[d]} U_i \le u+\delta\right) = 0 \vee \big\{1-u-d(1-u-\delta)\big\}. \tag{13}$$

Proof of Lemma 4. Firstly, it is clear by a union bound that for any $\mathbb{P} \in \mathcal{P}_d(\mathcal{U})$,

$$\mathbb{P}\Big(u < \max_{i\in[d]} U_i \le u+\delta\Big) \le \mathbb{P}\Big(\{U_1 \le u+\delta\} \cap \bigcup_{i\in[d]} \{u < U_i \le u+\delta\}\Big) \le (d\delta)\wedge(u+\delta),$$

$$\begin{aligned}\mathbb{P}\Big(u < \max_{1\le i\le d} U_i \le u+\delta\Big) &\ge 0 \vee \Big\{1 - \mathbb{P}\Big(\max_{1\le i\le d} U_i \le u\Big) - \mathbb{P}\Big(\max_{1\le i\le d} U_i > u+\delta\Big)\Big\} \\ &\ge 0 \vee \big\{1-u-d(1-u-\delta)\big\}.\end{aligned}$$

It remains to show these bounds can be attained. Let $\Delta : [0,1] \to [0,1]$ be a d-dimensional copula diagonal, and $C : [0,1]^d \to [0,1]$ be a copula satisfying $\Delta(t) = C(t,\ldots,t)$ for all $t \in [0,1]$ (see Lemma 1). By Sklar's theorem (Nelsen, 2006, Theorem 2.10.9), C is the joint distribution function of a random vector $(U_1,\ldots,U_d)$ where $U_i \sim \mathcal{U}$ for each $i \in [d]$. Further, if $(U_1,\ldots,U_d) \sim C$ then

$$\mathbb{P}\left(u < \max_{i\in[d]} U_i \le u+\delta\right) = C(u+\delta,\ldots,u+\delta) - C(u,\ldots,u) = \Delta(u+\delta) - \Delta(u).$$

For (12), consider taking $\Delta = \Delta_{\mathrm{up}}$ as defined by (7). Clearly $\Delta_{\mathrm{up}}(1) = 1$. If $t \le \frac{du}{d-1}$ then $d(t-u) - t \le 0$, so $\Delta_{\mathrm{up}}(t) \le t$ for all $t \in [0,1]$. Further, Δ_{up} is piecewise linear with the gradient of each piece bounded below by zero and above by d, so $0 \le \Delta_{\mathrm{up}}(t') - \Delta_{\mathrm{up}}(t) \le d(t'-t)$ whenever $0 \le t \le t' \le 1$. Thus Δ_{up} satisfies the conditions of Lemma 1 and is a d-dimensional copula diagonal.

If $u \le \frac{d-1}{d}$, then $\Delta_{\mathrm{up}}(u) = 0$, while $u > \frac{d-1}{d}$ implies $\Delta_{\mathrm{up}}(u) = du - d + 1$. If $u + \delta \le \frac{du}{d-1}$ (which occurs if and only if $d\delta \le u + \delta$), then $u \le \frac{d-1}{d}$ gives $\Delta_{\mathrm{up}}(u + \delta) = d\delta$, while $u > \frac{d-1}{d}$ implies $\Delta_{\mathrm{up}}(u + \delta) = du + d\delta - d + 1$. Conversely if $u + \delta > \frac{du}{d-1}$, then $u \le \frac{d-1}{d}$ and $\Delta_{\mathrm{up}}(u + \delta) = u + \delta$.

For (13), consider taking $\Delta = \Delta_{\mathrm{lo}}$ as defined by (8). Clearly $\Delta_{\mathrm{lo}}(1) = 1$, and $1 - d(1-t) - t \le 0$, so $\Delta_{\mathrm{lo}}(t) \le t$ for all $t \in [0,1]$. Further, Δ_{lo} is piecewise linear with each gradient bounded below by zero and above by d, so $0 \le \Delta_{\mathrm{lo}}(t') - \Delta_{\mathrm{lo}}(t) \le d(t' - t)$ when $0 \le t \le t' \le 1$. Thus Δ_{lo} satisfies the conditions of Lemma 1 and is a valid d-dimensional copula diagonal.

Now observe that $\Delta_{\mathrm{lo}}(u) = u$. If $u+\delta \le \frac{d+u-1}{d}$ (which occurs if and only if $1-u-d(1-u-\delta) \le 0$), then $\Delta_{\mathrm{lo}}(u + \delta) = u$. Conversely if $u + \delta > \frac{d+u-1}{d}$, then $\Delta_{\mathrm{lo}}(u + \delta) = 1 - d(1 - u - \delta)$. □

Proof of Theorem 1. Let $u := F(x) \in [0, 1]$ and $\delta := F(x+\varepsilon)-F(x) \in [0, 1-u]$. Define $F^{-1} : [0, 1] \to \mathbb{R}$ as the quantile function satisfying $F^{-1}(t) \le s$ if and only if $t \le F(s)$ for all $s \in \mathbb{R}$ and $t \in [0, 1]$. Note that if $U \sim \mathcal{U}$ and $X = F^{-1}(U)$, then $\mathbb{P}(X \le x) = \mathbb{P}\big(F^{-1}(U) \le x\big) = \mathbb{P}\big(U \le F(x)\big) = F(x)$, so $X \sim F$. To prove (3), take $(U_1, \ldots, U_d)$ as in (12) of Lemma 4 and set $X_i = F^{-1}(U_i)$ for $i \in [d]$ so

$$\begin{aligned}\mathbb{P}\left(x < \max_{i\in[d]} X_i \le x + \varepsilon\right) &= \mathbb{P}\left(x < \max_{i\in[d]} F^{-1}(U_i) \le x + \varepsilon\right) = \mathbb{P}\left(u < \max_{i\in[d]} U_i \le u + \delta\right) \\ &= (d\delta) \wedge (u + \delta) = \big\{d\big(F(x + \varepsilon) - F(x)\big)\big\} \wedge F(x + \varepsilon).\end{aligned}$$

To show (4), take $(U_1, \ldots, U_d)$ as in (13) of Lemma 4 and $X_i = F^{-1}(U_i)$ for $i \in [d]$ so that likewise

$$\begin{aligned}\mathbb{P}\left(x < \max_{i\in[d]} X_i \le x + \varepsilon\right) &= \mathbb{P}\left(u < \max_{i\in[d]} U_i \le u + \delta\right) = 0 \vee \big(1 - u - d(1 - u - \delta)\big) \\ &= 0 \vee \big\{1 - F(x) - d\big(1 - F(x + \varepsilon)\big)\big\}.\end{aligned}$$

□

5.2 Theorem 2

Proof of Theorem 2. Let Δ be the copula diagonal associated with the law of $(X_1, \ldots, X_d)$, and suppose it is a convex function on $[0, 1]$. Since $\Delta(1) = 1$, for any $u \in [0, 1]$ and $\delta \in [0, 1 - u]$,

$$\Delta(u + \delta) = \Delta\left(\frac{1 - u - \delta}{1 - u} \cdot u + \frac{\delta}{1 - u} \cdot 1\right) \le \frac{1 - u - \delta}{1 - u}\Delta(u) + \frac{\delta}{1 - u}.$$

Combining this with the facts that $\Delta(u + \delta) - \Delta(u) \le d\delta$ and $\Delta(u) \ge 0$ (see Lemma 1) gives

$$\Delta(u + \delta) - \Delta(u) \le \left\{\frac{\delta}{1 - u}\big(1 - \Delta(u)\big)\right\} \wedge (d\delta) \le \delta\left(\frac{1}{1 - u} \wedge d\right).$$

A quantile transform (see the proof of Theorem 1) with $u := F(x)$ and $\delta := F(x + \varepsilon) - F(x)$ yields

$$\begin{aligned}\mathbb{P}\left(x < \max_{i\in[d]} X_i \le x + \varepsilon\right) &= \mathbb{P}\left(F(x) < \max_{i\in[d]} F(X_i) \le F(x + \varepsilon)\right) = \Delta \circ F(x + \varepsilon) - \Delta \circ F(x) \\ &\le \big(F(x + \varepsilon) - F(x)\big)\left\{\frac{1}{1 - F(x)} \wedge d\right\}.\end{aligned}$$

We now show that the maximum is attained. For $d \in \mathbb{N}$ and $u \in [0, 1]$, define

$$\Delta_u(t) := \left\{t - \left(u \wedge \frac{d-1}{d}\right)\right\}\left(\frac{1}{1 - u} \wedge d\right) \cdot \mathbb{I}\left\{u \wedge \frac{d-1}{d} < t\right\},$$

which is a valid d-dimensional copula diagonal by Lemma 1. Note that for $\delta \in [0, 1 - u]$, we have $\Delta_u(u + \delta) - \Delta_u(u) = \delta\big(\frac{1}{1-u} \wedge d\big)$. Apply the quantile transform as above to conclude. □

5.3 Lemma 2

Proof of Lemma 2. If $d \geq 2$ and $\Sigma_{11} = 0$ then, by Sklar's theorem (Nelsen, 2006, Theorem 2.10.9), given a copula $C' : [0,1]^{d-1} \to [0,1]$ for $(X_2, \ldots, X_d)$, construct a copula for $(X_1, \ldots, X_d)$ as $C(x_1, \ldots, x_d) = x_1 \wedge C'(x_2, \ldots, x_d)$. The diagonal of C is $C(x, \ldots, x) = C'(x, \ldots, x)$. If $d = 1$ and $\Sigma_{11} = 0$ then take $C(x) = x$. We thus assume without loss of generality that $\Sigma_{ii} > 0$ for all $i \in [d]$.

The copula of $(X_1, \ldots, X_d)$ is the same as that for $(Z_1, \ldots, Z_d)$, where $Z_i = (X_i - \mu_i)/\sigma_i$ and $\sigma_i^2 = \Sigma_{ii}$ for each $i \in [d]$. Therefore, without loss of generality, set $\mu = 0$ and $\Sigma_{ii} = 1$ for each $i \in [d]$.

If $\Sigma_{ij} = 1$ where $1 \leq i < j \leq d$ then $\max_{i \in [d]} X_i = \max_{i \in [d] \setminus \{j\}} X_i$ almost surely so, without loss of generality, we take $\Sigma_{ij} < 1$ whenever $i \neq j$. Lemmas 5 and 6 in Chernozhukov et al. (2015) show $\max_{i \in [d]} X_i$ has Lebesgue density $f(x) = \phi(x) g(x)$ where $g : \mathbb{R} \to \mathbb{R}$ is increasing. For $x \in (0,1)$, note $\Delta(x) = \mathbb{P}\big(\max_{i \in [d]} X_i \leq \Phi^{-1}(x)\big)$, so Δ is differentiable on $(0,1)$ with increasing derivative

$$\Delta'(x) = \frac{f \circ \Phi^{-1}(x)}{\phi \circ \Phi^{-1}(x)} = g \circ \Phi^{-1}(x).$$

Thus Δ is convex on $(0,1)$, and also on $[0,1]$ as $0 = \Delta(0) \leq \Delta(x) \leq \Delta(1) = 1$ for all $x \in (0,1)$. □

5.4 Lemma 3

Proof of Lemma 3. The diagonal section of the Archimedean copula C is

$$\Delta(x) = \psi^{-1}\big(d \cdot \psi(x)\big).$$

Since ψ is differentiable with non-zero derivative on $(0,1)$, the inverse function theorem gives

$$\Delta'(x) = \frac{d \cdot \psi'(x)}{\psi' \circ \psi^{-1}\big(d \cdot \psi(x)\big)} = d \cdot \Psi \circ \psi(x).$$

Since ψ is strictly decreasing and Ψ is non-increasing, it follows that Δ' is non-decreasing on $(0,1)$. As $0 = \Delta(0) \leq \Delta(x) \leq \Delta(1) = 1$ for all $x \in (0,1)$, we conclude that Δ is convex on $[0,1]$. □

5.5 Examples

Details for Example 6. We apply Lemma 3 to each family. Firstly, the Clayton copula has inverse generator $\psi^{-1}(x) = (1+x)^{-1/\theta}$ for $\theta > 0$ (Nelsen, 2006, Example 4.23). With $x \in (0, \infty)$,

$$\big(\psi^{-1}\big)'(x) = -\frac{1}{\theta}(1+x)^{-1/\theta - 1} < 0, \qquad \Psi(x) = \left(\frac{1+x}{1+d \cdot x}\right)^{1/\theta+1}$$

and is decreasing, since $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1+x}{1+d\cdot x} = \frac{1-d}{(1+d\cdot x)^2} \leq 0$.

Next, the Frank copula has inverse generator $\psi^{-1}(x) = -\frac{1}{\theta}\log\big(1 - \big(1 - e^{-\theta}\big)e^{-x}\big)$ for $\theta > 0$ (Nelsen, 2006, Example 4.24). With $x \in (0, \infty)$ and setting $a = \big(1 - e^{-\theta}\big) \in (0,1)$,

$$\big(\psi^{-1}\big)'(x) = -\frac{1}{\theta}\frac{ae^{-x}}{1 - ae^{-x}} < 0, \qquad \Psi(x) = \frac{e^x - a}{e^{dx} - a},$$
$$\Psi'(x) = \frac{(1-d)e^{(d+1)x} - ae^x + ade^{dx}}{\big(e^{dx} - a\big)^2} \leq 0,$$

since by convexity of the exponential function, $e^{dx} \leq \frac{d-1}{d}e^{(d+1)x} + \frac{1}{d}e^x$. Thus Ψ is non-increasing.

Finally, the Gumbel–Hougaard copula has inverse generator $\psi^{-1}(x) = \exp\bigl(-x^{1/\theta}\bigr)$ for $\theta \geq 1$ (Nelsen, 2006, Example 4.25). With $x \in (0, \infty)$,

$$\bigl(\psi^{-1}\bigr)'(x) = -\frac{1}{\theta} x^{1/\theta - 1} \exp\bigl(-x^{1/\theta}\bigr) < 0, \qquad\qquad \Psi(x) = d^{1/\theta - 1} \exp\bigl(x^{1/\theta}(1 - d^{1/\theta})\bigr)$$

and is decreasing, since $d^{1/\theta} \geq 1$. $\square$

Details for Example 7. As each C_k is diagonally convex, $x \mapsto \frac{1}{K}\sum_{k=1}^{K} p_k C_k(x, \ldots, x)$ is convex. The mixture copula is easily verified to be a copula using Nelsen (2006, Definition 2.10.6). $\square$

Details for Example 8. Take $\mu = 0$ and $\sigma = 1$; consider $\mu + \sigma X_i$ for the general case. By Theorem 2,

$$\mathbb{P}\left(x < \max_{i \in [d]} X_i \leq x + \varepsilon\right) \leq \bigl(\Phi(x + \varepsilon) - \Phi(x)\bigr)\left\{\frac{1}{1 - \Phi(x)} \wedge d\right\}.$$

For $x \leq 0$ we have $\Phi(x) \leq 1/2$ and $\Phi(x + \varepsilon) - \Phi(x) \leq \varepsilon\phi(0) = \frac{1}{\sqrt{2\pi}}$, so

$$\bigl(\Phi(x + \varepsilon) - \Phi(x)\bigr)\left\{\frac{1}{1 - \Phi(x)} \wedge d\right\} \leq \frac{\varepsilon}{\sqrt{2\pi}}(2 \wedge d) = \varepsilon\sqrt{\frac{2}{\pi}}.$$

If $x \geq 0$ then $\Phi(x + \varepsilon) - \Phi(x) \leq \varepsilon\phi(x)$ and $\frac{\phi(x)}{1 - \Phi(x)} \leq \frac{2}{\sqrt{x^2 + 4} - x} \leq x + 1$ by Birnbaum (1942), so

$$\bigl(\Phi(x + \varepsilon) - \Phi(x)\bigr)\left\{\frac{1}{1 - \Phi(x)} \wedge d\right\} \leq \varepsilon\phi(x)\left\{\frac{x + 1}{\phi(x)} \wedge d\right\} = \varepsilon\bigl\{(x + 1) \wedge \bigl(d \cdot \phi(x)\bigr)\bigr\}.$$

Setting $x = \sqrt{2 \log d} \geq 0$ gives $x + 1 = \sqrt{2 \log d} + 1$ and $d \cdot \phi(x) = \frac{1}{\sqrt{2\pi}}$. Since $x \mapsto x + 1$ is continuous and strictly increasing, while $x \mapsto d \cdot \phi(x)$ is continuous and strictly decreasing on $[0, \infty)$, we have

$$\begin{aligned}
\sup_{x \geq 0}\bigl\{(x + 1) \wedge \bigl(d \cdot \phi(x)\bigr)\bigr\} &= \inf_{x \geq 0}\bigl\{(x + 1) \vee \bigl(d \cdot \phi(x)\bigr)\bigr\} \\
&\leq \left(\sqrt{2 \log d} + 1\right) \vee \frac{1}{\sqrt{2\pi}} = \sqrt{2 \log d} + 1.
\end{aligned} \tag{14}$$

The result follows as $\sqrt{2 \log d} + 1 \geq \sqrt{\frac{2}{\pi}}$. $\square$

Details for Example 9. The first result follows by Lemma 2 and Example 8. For the second, consider the $2d$-dimensional random vector $Y = (X_1 - \mu, \ldots, X_d - \mu, \mu - X_1, \ldots, \mu - X_d)$, which has a multivariate Gaussian distribution with $Y_i \sim \mathcal{N}(0, \sigma^2)$ for each $i \in [2d]$. By the first inequality,

$$\mathbb{P}\left(x < \max_{i \in [d]} |X_i - \mu| \leq x + \varepsilon\right) = \mathbb{P}\left(x < \max_{i \in [2d]} Y_i \leq x + \varepsilon\right) \leq \frac{\varepsilon}{\sigma}\left(\sqrt{2 \log 2d} + 1\right). \qquad \square$$

Details for Example 10. We assume that the scale parameter is $\lambda = 1$; the general result for $\lambda > 0$ then follows by considering λX_i for $i \in [d]$. The CDF and Lebesgue density of X_i are therefore

$$F(x) = 1 - \exp\bigl(-x^{\alpha}\bigr), \qquad\qquad f(x) = \alpha x^{\alpha - 1} \exp\bigl(-x^{\alpha}\bigr).$$

With $x_* = \bigl(\frac{\alpha - 1}{\alpha}\bigr)^{1/\alpha}$, f increases on $[0, x_*]$ and decreases on $[x_*, \infty)$. By Theorem 2, if $x \leq x_*$,

$$\mathbb{P}\left(x < \max_{i \in [d]} X_i \leq x + \varepsilon\right) \leq \frac{\varepsilon f(x_*)}{1 - F(x_*)} = \varepsilon\alpha x_*^{\alpha - 1} = \varepsilon\alpha\left(\frac{\alpha - 1}{\alpha}\right)^{\frac{\alpha - 1}{\alpha}} \leq \varepsilon\alpha.$$

Conversely, if $x > x_*$ then

$$\mathbb{P}\Big(x < \max_{i\in[d]} X_i \le x + \varepsilon\Big) \le \varepsilon f(x)\Big\{\frac{1}{1-F(x)} \wedge d\Big\} = \varepsilon\alpha x^{\alpha-1}\Big\{1 \wedge \Big(d\exp\big(-x^\alpha\big)\Big)\Big\}.$$

Take $x = (\log d + 1)^{1/\alpha} \ge x_*$ so that $x^{\alpha-1} = (\log d + 1)^{\frac{\alpha-1}{\alpha}}$ and $\exp\big(-x^\alpha\big) \le \frac{1}{d}$. As $x \mapsto x^{\alpha-1}$ is increasing and $x \mapsto x^{\alpha-1}\exp\big(-x^\alpha\big)$ is decreasing to zero on $[x_*, \infty)$, we have

$$\sup_{x\ge x_*}\Big\{x^{\alpha-1} \wedge \Big(dx^{\alpha-1}\exp\big(-x^\alpha\big)\Big)\Big\} \le \inf_{x\ge x_*}\Big\{x^{\alpha-1} \vee \Big(dx^{\alpha-1}\exp\big(-x^\alpha\big)\Big)\Big\} \le (\log d + 1)^{\frac{\alpha-1}{\alpha}}. \qquad \square$$

Details for Example 11. We assume that the scale parameter is $\lambda = 1$; the general result for $\lambda > 0$ follows by considering λX_i for $i \in [d]$. The CDF, Lebesgue density, and hazard function of X_i are

$$F(x) = 1 - \exp\big(-e^x\big), \qquad f(x) = \exp\big(x - e^x\big), \qquad h(x) = e^x.$$

Note that $f(x)$ is increasing on $(-\infty, 0]$, so for $x \le 0$, by Theorem 2, there exists $(X_1, \ldots, X_d)$ with

$$\mathbb{P}\Big(x < \max_{i\in[d]} X_i \le x + \varepsilon\Big) \le \frac{\varepsilon f(0)}{1 - F(0)} = \varepsilon.$$

Further, $f(x)$ is decreasing on $[0, \infty)$ while $h(x)$ is increasing on $[0, \infty)$. Note that for $d \ge 2$, we have $h(x) = d \cdot f(x)$ if and only if $x = \log\log d$, and if $d = 1$ then $d \cdot f(x) \le 1$. So for $x \ge 0$,

$$\mathbb{P}\Big(x < \max_{i\in[d]} X_i \le x + \varepsilon\Big) \le \varepsilon f(x)\Big\{\frac{1}{1-F(x)} \wedge d\Big\} \le \varepsilon\big\{h(x) \wedge \big(d \cdot f(x)\big)\big\} \le \varepsilon\big(1 + \log d\big). \qquad \square$$

Details for Example 12. We assume that the scale parameter is $\lambda = 1$; the general result follows by considering λX_i for $i \in [d]$. The CDF, Lebesgue density, and hazard function of X_i are therefore

$$F(x) = 1 - x^{-\alpha}, \qquad f(x) = \alpha x^{-\alpha-1}, \qquad h(x) = \alpha/x,$$

for $x \ge 1$. Since f and h are both decreasing, and by Theorem 2,

$$\mathbb{P}\Big(x < \max_{i\in[d]} X_i \le x + \varepsilon\Big) \le \varepsilon\big\{h(x) \wedge \big(d \cdot f(x)\big)\big\} \le \varepsilon\big\{h(1) \wedge \big(d \cdot f(1)\big)\big\} = \alpha\varepsilon. \qquad \square$$

Details for Example 13. We begin by showing that the first term on the right-hand side is an upper bound for the left-hand side. Note that the common CDF and density function of T_j, for $j \in [d]$, are

$$F(x) = \sum_{k=1}^K p_k \Phi\Big(\frac{x}{\sigma_k}\Big), \qquad f(x) = \sum_{k=1}^K \frac{p_k}{\sigma_k}\phi\Big(\frac{x}{\sigma_k}\Big),$$

respectively. As $(T_1, \ldots, T_d)$ are i.i.d., and recalling that the independence copula is diagonally convex by Example 5, we have by Theorem 2 that

$$\begin{aligned}
\mathbb{P}\Big(x < \max_{j\in[d]} T_j \le x + \varepsilon\Big) &\le \big(F(x+\varepsilon) - F(x)\big)\Big\{\frac{1}{1-F(x)} \wedge d\Big\} \\
&= \sum_{k=1}^K p_k\Big(\Phi(x/\sigma_k + \varepsilon/\sigma_k) - \Phi(x/\sigma_k)\Big)\left\{\frac{1}{1 - \sum_{k=1}^K p_k\Phi(x/\sigma_k)} \wedge d\right\} \\
&\le \sum_{k=1}^K p_k\Big(\Phi(x/\sigma_k + \varepsilon/\sigma_k) - \Phi(x/\sigma_k)\Big)\left\{\frac{1}{p_1\big(1 - \Phi(x)\big)} \wedge d\right\}.
\end{aligned}$$

Consider first the case that $x \leq 0$; then $\Phi(x+\varepsilon) - \Phi(x) \leq \varepsilon\phi(0)$ and $\Phi(x) \leq 1/2$, so

$$\mathbb{P}\Big(x < \max_{j\in[d]} T_j \leq x+\varepsilon\Big) \leq \frac{2\varepsilon\phi(0)}{p_1}\sum_{k=1}^{K}\frac{p_k}{\sigma_k}.$$

Next, if $x > 0$ then $\Phi(x+\varepsilon) - \Phi(x) \leq \varepsilon\phi(x)$ and $\frac{\phi(x)}{1-\Phi(x)} \leq x+1$ by Birnbaum (1942), so

$$\begin{aligned}\mathbb{P}\Big(x < \max_{j\in[d]} T_j \leq x+\varepsilon\Big) &\leq \varepsilon\sum_{k=1}^{K} p_k \frac{\phi(x/\sigma_k)}{\sigma_k\phi(x)}\bigg\{\frac{\phi(x)}{p_1\big(1-\Phi(x)\big)} \wedge \big(d\cdot\phi(x)\big)\bigg\} \\ &\leq \varepsilon\sum_{k=1}^{K} p_k \frac{\phi(x/\sigma_k)}{\sigma_k\phi(x)}\bigg\{\frac{x+1}{p_1} \wedge \big(d\cdot\phi(x)\big)\bigg\}.\end{aligned}$$

If $x \geq 1$ then since $0 < \sigma_k \leq 1$ for each $k \in [K]$, it follows that $2\log(1/\sigma_k) \leq 1/\sigma_k^2 - 1$ and so we have $x \geq \sqrt{\frac{2\log(1/\sigma_k)}{1/\sigma_k^2-1}}$. It follows that $\log(1/\sigma_k) \leq x^2(1/\sigma_k^2-1)/2$, so $e^{-x^2/(2\sigma_k^2)} \leq \sigma_k e^{-x^2/2}$ and hence $\phi(x/\sigma_k) \leq \sigma_k\phi(x)$. In this case, by (14),

$$\mathbb{P}\Big(x < \max_{j\in[d]} T_j \leq x+\varepsilon\Big) \leq \frac{\varepsilon}{p_1}\big\{(x+1)\wedge\big(d\cdot\phi(x)\big)\big\} \leq \frac{\varepsilon}{p_1}\Big(\sqrt{2\log d}+1\Big).$$

Alternatively, if $0 < x < 1$, then since $\sigma_k \leq 1$ for each $k \in [K]$, we have $\phi(x/\sigma_k) \leq \phi(x)$ and so

$$\mathbb{P}\Big(x < \max_{j\in[d]} T_j \leq x+\varepsilon\Big) \leq \frac{2\varepsilon}{p_1}\sum_{k=1}^{K}\frac{p_k}{\sigma_k}.$$

Combining these cases, we deduce that for all $x \in \mathbb{R}$,

$$\mathbb{P}\Big(x < \max_{j\in[d]} T_j \leq x+\varepsilon\Big) \leq \bigg\{\frac{\varepsilon}{p_1}\Big(\sqrt{2\log d}+1\Big)\bigg\} \vee \bigg\{\frac{2\varepsilon}{p_1}\sum_{k=1}^{K}\frac{p_k}{\sigma_k}\bigg\} \leq \frac{\varepsilon}{p_1}\bigg(\sqrt{2\log d}+2\sum_{k=1}^{K}\frac{p_k}{\sigma_k}\bigg).$$

We now address the second term on the right-hand side. By the second equality in (10), $(T_1,\ldots,T_d)$ follow a Gaussian mixture distribution with K^d components, so consider the following representation:

$$(T_1,\ldots,T_d) = \sum_{k_1=1}^{K}\cdots\sum_{k_d=1}^{K}\mathbb{I}\big\{Z=(k_1,\ldots,k_d)\big\}\big(Y_{k_1},\ldots,Y_{k_d}\big),$$

where the latent group assignment Z takes values in $[K]^d$ with $\mathbb{P}\big(Z=(k_1,\ldots,k_d)\big) = \prod_{j=1}^{d} p_{k_j}$, and with $Y_{k_j} \sim \mathcal{N}\big(0,\sigma_{k_j}\big)$ independently for $j \in [d]$ and independently of Z. Therefore by conditioning on Z and applying the result of Example 8,

$$\begin{aligned}
\mathbb{P}\Big(x < \max_{j\in[d]} T_j \le x+\varepsilon\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(x < \max_{j\in[d]} T_j \le x+\varepsilon \;\Big|\; Z\Big)\Big] \\
&= \sum_{k_1=1}^{K}\cdots\sum_{k_d=1}^{K}\Big(\prod_{j=1}^{d} p_{k_j}\Big)\mathbb{P}\Big(x < \max_{j\in[d]} T_j \le x+\varepsilon \;\Big|\; Z=(k_1,\ldots,k_d)\Big) \\
&= \sum_{k_1=1}^{K}\cdots\sum_{k_d=1}^{K}\Big(\prod_{j=1}^{d} p_{k_j}\Big)\mathbb{P}\Big(x < \max_{j\in[d]} Y_{k_j} \le x+\varepsilon\Big) \\
&\le \sum_{k_1=1}^{K}\cdots\sum_{k_d=1}^{K}\Big(\prod_{j=1}^{d} p_{k_j}\Big)\frac{\varepsilon}{\min_{1\le j\le d}\sigma_{k_j}}\Big(\sqrt{2\log d}+1\Big) \\
&\le \sum_{k_1=1}^{K}\cdots\sum_{k_d=1}^{K}\Big(\prod_{j=1}^{d} p_{k_j}\Big)\frac{\varepsilon}{\sigma}\Big(\sqrt{2\log d}+1\Big) \\
&= \frac{\varepsilon\big(\sqrt{2\log d}+1\big)}{\sigma}\prod_{j=1}^{d}\Big(\sum_{k=1}^{K} p_k\Big) = \frac{\varepsilon}{\sigma}\Big(\sqrt{2\log d}+1\Big).
\end{aligned}$$

We remark that the inequality $\min_{1\le j\le d}\sigma_{k_j} \ge \sigma$ is essentially optimal in the regime where the dimension d is much larger than the number of original components K, since they differ in only $(K-1)^d$ of the K^d possible values for $(k_1,\ldots,k_d)$. □

6 Conclusion

We presented sharp upper and lower bounds for the pointwise concentration function of the maximum (or minimum) statistic of d identically distributed random variables, under no further assumptions on their dependence structure (copula). When further restricted to copulas with convex diagonal sections, we demonstrated an improved (and similarly optimal) upper bound on the aforementioned concentration function. We verified this condition for a range of popular copulas, and applied our results to several different marginal distributions. Among other contributions, we recover a version of Nazarov's inequality with substantially relaxed assumptions and derive similar results for non-Gaussian laws. We presented an application for high-dimensional statistical inference, giving an explicit example pertaining to the Gaussian mixture approximation for factor models.

There are some potential directions for future research. Firstly, our main results apply only when the marginal distributions of each entry in $(X_1,\ldots,X_d)$ agree. This is a somewhat restrictive assumption, precluding applications in settings where the random vector of interest is not standardized entry-wise. For instance, in Example 9 we are currently unable to accommodate the setting of $(X_1,\ldots,X_d) \sim \mathcal{N}(\mu,\Sigma)$ with $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ an arbitrary positive semi-definite matrix, which is handled by Nazarov's inequality (Nazarov, 2003; Chernozhukov et al., 2017b) whenever $\min_{1\le i\le d}\Sigma_{ii} > 0$. It is desirable to know whether anti-concentration inequalities can be derived in regimes where the marginal laws are Gaussian but the copula is non-Gaussian, for example, following our Example 8. The main challenge in establishing such generalizations is to formulate a natural extension of the "diagonally convex" property introduced in Definition 3, along with a corresponding result analogous to our Theorem 2. A secondary task would then be to verify the new condition for a selection of popular multivariate copulas; initial investigation suggests that this is unlikely to be as straightforward as in Section 3.

A closely related problem is that of providing bounds for the probability that $(X_1, \ldots, X_d)$ lies near the perimeter of a (high-dimensional) rectangle. That is, to control

$$\mathbb{P}\left(\bigcap_{i=1}^{d}\{X_i \leq x_i + \varepsilon_i\}\right) - \mathbb{P}\left(\bigcap_{i=1}^{d}\{X_i \leq x_i\}\right) \tag{15}$$

where $x_i \in \mathbb{R}$ and $\varepsilon_i > 0$ for $i \in [d]$. Setting $Y_i := (X_i - x_i)/\varepsilon_i$ for $i \in [d]$, (15) is equal to $\mathbb{P}\big(0 < \max_{i \in [d]} Y_i \leq 1\big)$. Thus, in order to establish bounds for (15), it would be sufficient to generalize our main results to the setting where the marginal distributions of $X_1, \ldots, X_d$ are not necessarily equal.

7 Acknowledgments

We thank Boris Hanin and Grigoris Paouris for their comments. The authors gratefully acknowledge financial support from the National Science Foundation through grant DMS-2210561.

References

Aizenman, M., Germinet, F., Klein, A., and Warzel, S. (2009). On Bernoulli decompositions for random variables, concentration bounds, and spectral localization. *Probability Theory and Related Fields*, 143:219–238.

Arakelian, V. and Karlis, D. (2014). Clustering dependencies via mixtures of copulas. *Communications in Statistics–Simulation and Computation*, 43(7):1644–1661.

Bakshi, A., Diakonikolas, I., Hopkins, S. B., Kane, D., Karmalkar, S., and Kothari, P. K. (2020). Outlier-robust clustering of Gaussians and other non-spherical mixtures. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 149–159. IEEE.

Belloni, A., Fang, E. X., and Shen, S. (2024). Anti-concentration inequalities for the difference of maxima of Gaussian random vectors. arXiv:2408.13348.

Birnbaum, Z. W. (1942). An inequality for Mills' ratio. *Annals of Mathematical Statistics*, 13:245–246.

Bobkov, S. G. and Chistyakov, G. P. (2015). On concentration functions of random variables. *Journal of Theoretical Probability*, 28:976–988.

Cattaneo, M. D., Feng, Y., and Underwood, W. G. (2024). Uniform inference for kernel density estimators with dyadic data. *Journal of the American Statistical Association*, 119(548):2695–2708.

Cattaneo, M. D., Masini, R. P., and Underwood, W. G. (2022). Yurinskii's coupling for martingales. arXiv:2210.00362.

Cattaneo, M. D. and Yu, R. R. (2025). Strong approximations for empirical processes indexed by Lipschitz functions. *Annals of Statistics*.

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

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, 42(5):1787–1818.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014b). Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, 42(4):1564–1597.

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

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017a). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, 45(4):2309–2352.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017b). Detailed proof of Nazarov's inequality. arXiv:1711.10696.

Cuculescu, I. and Theodorescu, R. (2001). Copulas: Diagonals, tracks. *Revue Roumaine de Mathématiques Pures et Appliquées*, 46(6):731–742.

Durante, F. and Sempi, C. (2016). *Principles of Copula Theory*. Chapman and Hall/CRC Press.

Fernández-Sánchez, J. and Úbeda-Flores, M. (2018). Constructions of copulas with given diagonal (and opposite diagonal) sections and some generalizations. *Dependence Modeling*, 6(1):139–155.

Fox, J., Kwan, M., and Sauermann, L. (2021). Combinatorial anti-concentration inequalities, with applications. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 171, pages 227–248. Cambridge University Press.

Frank, M. J., Nelsen, R. B., and Schweizer, B. (1987). Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields*, 74(2):199–211.

Giessing, A. (2023). Anti-concentration of suprema of Gaussian processes and Gaussian order statistics. arXiv:2310.12119.

Götze, F., Naumov, A., Spokoiny, V. G., and Ulyanov, V. V. (2019). Large ball probabilities, Gaussian comparison and anti-concentration. *Bernoulli*, 25(4A):2538–2563.

Jaworski, P. (2009). On copulas and their diagonals. *Information Sciences*, 179(17):2863–2871.

Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Japanese Journal of Statistics and Data Science*, 4:257–297.

Krishnapur, M. (2016). Anti-concentration inequalities. Lecture notes.

Kuchibhotla, A. K., Mukherjee, S., and Banerjee, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli*, 27(1):192 – 217.

Lévy, P. (1954). *Théorie de l'Addition Des Variables Aléatoires*. Gauthier-Villars, Paris.

Litvak, A. E., Lytova, A., Tikhomirov, K., Tomczak-Jaegermann, N., and Youssef, P. (2017). Adjacency matrices of random digraphs: singularity and anti-concentration. *Journal of Mathematical Analysis and Applications*, 445(2):1447–1491.

Livshyts, G. V. (2014). Maximal surface area of a convex set in $\mathbb{R}^n$ with respect to log concave rotation invariant measures. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2011–2013*, pages 355–383. Springer.

Livshyts, G. V. (2021). Some remarks about the maximal perimeter of convex sets with respect to probability measures. *Communications in Contemporary Mathematics*, 23(05).

Lopes, M. E., Lin, Z., and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Annals of Statistics*, 48(2):1214–1229.

Meka, R., Nguyen, O., and Vu, V. H. (2015). Anti-concentration for polynomials of independent random variables. *Theory of Computing*, 12:1–17.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis: Israel Seminar, 2001–2002*, pages 169–187. Springer.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer Series in Statistics. Springer, New York.

Nie, Z. (2022). Matrix anti-concentration inequalities with applications. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 568–581.

Paouris, G. (2012). Small ball probability estimates for log-concave measures. *Transactions of the American Mathematical Society*, 364(1):287–308.

Paouris, G. and Valettas, P. (2018). A Gaussian small deviation inequality for convex functions. *Annals of Probability*, 46(3):1441–1454.

Rudelson, M. and Vershynin, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617.

Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8:45–114.

Vershynin, R. and Rudelson, M. (2007). Anti-concentration inequalities. In *Phenomena in High Dimensions, Third Annual Conference, Samos, Greece.*