Sharp Anti-Concentration Inequalities for Extremum Statistics via Copulas

by Aryan Alavi

Based on the paper by Cattaneo, Masini, and Underwood

February 2025

Topics

- Anti-concentration Problem
- Copulas
- General Upper Bound
- General Lower Bound
- Upper Bound under Constraint
- Examples

Concentration Vs. Anti-concentration

- Concentration of measure -> quite well studied.
- Anti-concentration -> much less studied
- Formalized by Levy(1954) as the function:

$$L(Y, \varepsilon) := \sup_{x \in \mathbb{R}} \mathbb{P}(x \le Y \le x + \varepsilon)$$

Applications

- High-dimensional statistics

- Random matrix theory

- Geometric analysis

- Applied probability

Maximum statistic

- In this paper, concentration of maximum statistics of several random variables is studied.

$$Y := \max_{i\in[d]} X_i$$

- The trivial bound assuming density function f(x) for Y

$$L(Y,\varepsilon) \leq \varepsilon \sup_{x\in\mathbb{R}} f(x)$$

Previous Bounds

- Bounds based on variance (Bobkov and Chistyakov 2015)

- Bounds for log-concave distributions (Saumard and Wellner 2014)

- Multivariate Gaussian setting studied (Giessing 2023)

Notations

- cumulative distribution function of N(0, 1) Φ
- Lebesgue density function of N (0, 1) ϕ
- Minimum of two values $a \wedge b$
- Maximum of two values $a \vee b$

What is a copula?

- Two information are provided in a multivariate distribution:
    - univariate margins F1, . . . , Fd
    - copula C

- copula C captures the dependence among the d variables
    - irrespective of their marginal distributions

What is a copula?

- A d-variate copula C : [0, 1]^d → [0, 1] is the cdf of a random vector (U1, . . . , Ud) with Uniform(0, 1) margins:

$$C(\boldsymbol{u}) = \mathrm{P}[U_1 \leq u_1, \ldots, U_d \leq u_d]$$

$$\mathrm{P}[U_j \leq u_j] = u_j$$

Important properties

- If some component u_j is 0, then C(u) = 0
- C(1, . . . , 1, u_j , 1, . . . , 1) = u_j
- C is nondecreasing in each of the d variables
- C is a Lipschitz function

$$|C(\boldsymbol{u}) - C(\boldsymbol{v})| \leq |u_1 - v_1| + \cdots + |u_d - v_d|$$

Sklar theorem (2006)

- For any d-variate copula C and F1, . . . , Fd be univariate cdf’s, the following function is a d-variate cdf with marginals F1, . . . , Fd

$$F(\boldsymbol{x}) = C\big(F_1(x_1), \ldots, F_d(x_d)\big)$$

Sklar theorem (2006)

- For any d-variate distribution with cdf F and marginal cdfs F1, . . . , Fd, there exists a copula C holding:

$$F(\boldsymbol{x}) = C\big(F_1(x_1), \ldots, F_d(x_d)\big)$$

Copula diagon

- A function $\Delta : [0, 1] \to [0, 1]$ is a d-dimensional copula diagonal, if there exists a d-dimensional copula C such that:

$$\Delta(u) = C(u, \ldots, u)$$

Copula diagon

- A function $\Delta : [0, 1] \to [0, 1]$ is a d-dimensional copula diagonal if and only if it satisfies 3 properties:
  - $\Delta(1) = 1$
  - For all $0 <= u <= 1$: $\Delta(u) <= u$
  - For all $u, u' \in [0, 1]$ with $u \leq u'$: $0 \leq \Delta(u') - \Delta(u) \leq d(u' - u)$

Pointwise concentration function

- In this paper we study the pointwise concentration function of maximum statistic

$$\mathbb{P}\left(x < \max_{i \in [d]} X_i \leq x + \varepsilon\right) = C\big(F(x+\varepsilon), \ldots, F(x+\varepsilon)\big) - C\big(F(x), \ldots, F(x)\big)$$

- Can be written based on copula diagon

Pointwise concentration function

- We study the minimum and and maximum value of pointwise concentration function among all joint distribution satisfying all marginal cdfs equal to F

- $\mathcal{P}_d(F)$ is the family of the following joint distributions for some C

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

General bounds

- We study the minimum and and maximum value of pointwise concentration function among all joint distribution satisfying all marginal cdfs equal to F

- $\mathcal{P}_d(F)$ is the family of the following joint distributions for some C

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

General tight upper bound

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

General tight upper bound

- If $F(x + \varepsilon) - F(x) \leq F(x) / (d - 1)$ and $F(x) \in (0, 1)$, we obtain

$$\mathbb{P}_{\text{up}}\left(x < \max_{i\in[d]} X_i \leq x + \varepsilon\right) = d\left(F(x+\varepsilon) - F(x)\right)$$

- This precludes the possibility of obtaining anti-concentration bounds which hold uniformly in ε and depend sublinearly on d

Proof steps - uniform

- First, assume the marginal distributions follow U(0, 1)

- We simply study copula diagon

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

Proof steps - uniform

- We should prove:

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

Proof steps - general hold

- The inequality holds for every P simply because:

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

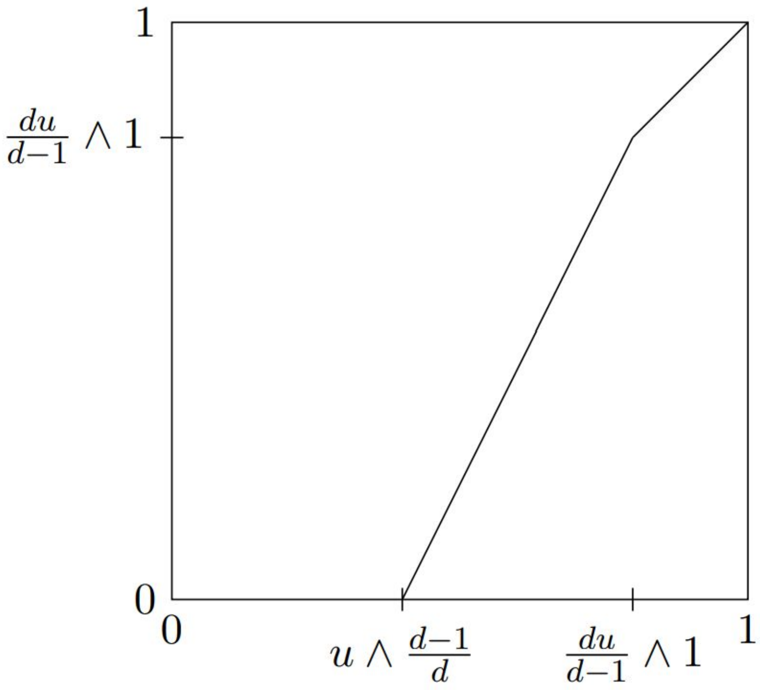
Proof steps - attainability

- We propose the following copula diagonal:

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

- By checking the three properties, it can be shown that it is a valid copula diagonal

Proof steps - attainability



Plot of the upper bound diagonal

Proof steps - general marginal

- By defining X = F^−1 (U), it will have cdf

$$\mathbb{P}(X \leq x) = \mathbb{P}\big(F^{-1}(U) \leq x\big) = \mathbb{P}\big(U \leq F(x)\big) = F(x)$$

- By defining u := F(x) and δ := F(x+ε)−F(x) , our goal will be obtained as bellow:

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

Applying to Gaussians

- Let d ∈ N, σ > 0 and ε ∈ [0, σ], the bound results there exists vector X with Xi ~ N (0, σ^2) such that

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

- It is far from the Nazarov's bound for jointly gaussian c:

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

Applying to uniform distributions

- In the case of jointly uniform distribution, we can simply obtain the following bound:

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

- On the other hand, the general bound can differ by factor approaching infinity by increasing d

$$\mathbb{P}_{\text{up}}\left(x < \max_{i\in[d]} X_i \leq x+\varepsilon\right) = d\varepsilon$$

General lower bound

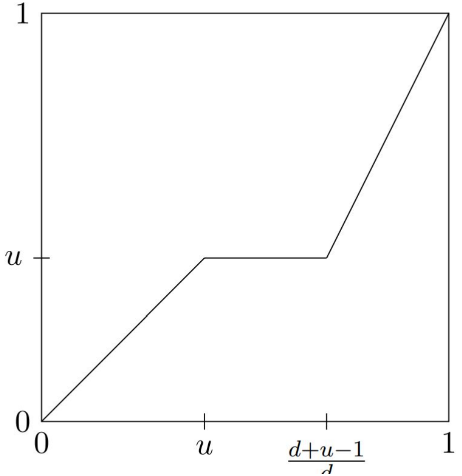
- The theorem also finds the lower bound for the concentration:

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

General lower bound - copula diagonal

- The minimum can be obtained by applying the copula diagonal:

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



Limiting the joint distribution

- General upper bound found for the concentration function in the previous theorem was not quite useful
- For sharper inequalities, we should make assumptions about the distribution (or the copula relating them)
- We study the case where the copula diagonal is a convex function on [0, 1]

- We study how sharp this bound is in special cases, and how constraining this condition is

Limiting the joint distribution

- We study the set $\mathcal{P}_d^{\mathrm{c}}(F)$
- This set consists of distributions P s.t.

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

- and Δ(x) = C(x, x, … x) is a convex function

Proof steps - uniform

- We start by studying the copula diagonal

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

Proof steps - uniform

- By properties of Δ:

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

Proof steps - attainability

- The upper bound can be attained by the following copula diagonal:

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

- It is a valid diagonal based on the three conditions, and also convex.
- It is in fact zero function until some point and linear afterwards.

Proof steps - general marginals

- By defining with u := F(x) and δ := F(x + ε):

$$\mathbb{P}\left(x < \max_{i\in[d]} X_i \leq x+\varepsilon\right) = \mathbb{P}\left(F(x) < \max_{i\in[d]} F(X_i) \leq F(x+\varepsilon)\right) = \Delta \circ F(x+\varepsilon) - \Delta \circ F(x)$$

$$\leq \left(F(x+\varepsilon) - F(x)\right)\left\{\frac{1}{1-F(x)} \wedge d\right\}.$$

Basic examples

- The convexity condition holds for independent components

$$\Delta_{\text{ind}}(u) = u^d$$

- It also holds for the d-dimensional Frechet–Hoeffding upper bound copula:

$$C^+(u_1, u_2, \ldots, u_d) = \min(u_1, u_2, \ldots, u_d)$$

$$\Delta_{\text{FHU}}(u) = u$$

Archimedean copula

- A class of copulas generated by decreasing ψ : [0, 1] → [0, ∞] such that ψ(0) = ∞ and ψ(1) = 0

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

Sufficient condition for archimedeans

- If $\psi'(x) < 0$ and is defined for all $x \in (0, 1)$, and the following function is non-increasing, then C has convex diagonal:

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

Proof

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

$$\Downarrow$$

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

$$\Downarrow$$

Δ' is non-decreasing on (0,1)

Example

- Clayton copulas with generator $\psi(x) = x \wedge (-r) - 1$ for $r > 0$

- Gumbel– Hougaard copulas with generator $\psi(x) = (-\log x) \wedge r$ for $r >= 1$

- Convex combination of copulas with convex copula

Jointly Gaussian distribution

- Suppose X is a multivariate Gaussian distribution with Xi ~ N (μ, σ^2)
- The copula diagon is a convex function
- By substituting in the theorem, the obtained result will be:

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

- It is quite similar to Nazarov's bound

Jointly Gaussian distribution

- By considering the multivariate Gaussian vector Y = (X1 − μ, . . . , Xd − μ, μ − X1, . . . , μ - Xd), the following result can be obtained:

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

Application - high dimensional statistical learning

- X is a R^d valued random vector constructed using samples taken from an underlying data set

- We assume that a coupling is available for X
- T = (T1, . . . , Td) and decreasing function p : [0, ∞) → [0, 1] exist s.t.

$$\mathbb{P}\left(\|X - T\|_\infty > \varepsilon\right) \leq p(\varepsilon)$$

Application - high dimensional statistical inference

- With significance level $\alpha \in (0, 1)$, the quantile for the maximum statistic of T is defined as

$$q_\alpha := \inf\{q \in \mathbb{R} : \mathbb{P}(\max_{i\in[d]} T_i \le q) \ge 1 - \alpha\}$$

- For maximum statistic of X

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

- Holds for every ε >= 0 -> the bound can be minimized over ε

Review

- Sharp upper and lower bound of pointwise concentration function of maximum statistic of same distribution random variables is calculated
  - Bounds not good enough

- Restricting the copula for better bounds -> convex diagonal:
  - Covers lots of cases
  - Quite sharper bounds
  - Studying example copulas
  - Possible Applications

Future studies

- Only studied distributions with common marginals
  - Quite important but restrictive case
  - Challenging problem

- Other applications
  - Other marginally Gaussian distributions with convex diagonal
- Other concentrations of multivariate distributions
  - Concentration in a rectangle

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

Thanks for your attention!