Empirical and Extreme Value Distributions of Random Angles, Pearson Correlation, and the Beta Distribution

Steve Broll

May 6, 2021

Motivation

The cosine of the angle $\cos(\theta)$ between two vectors is just the Pearson correlation coefficient ρ [1]. By studying the distributions of the maximums or minimums of angles, and of the cosines of those angles, we can find the asymptotic behavior of the maximum and minimum correlation coefficients as $n \to \infty$.

Additionally, if we study the squared cosine of these angles $\cos^2(\theta) = \rho^2 = R^2$, we are studying Beta-distributed random-variables [2] and so we can use the tail properties of the Beta distribution.

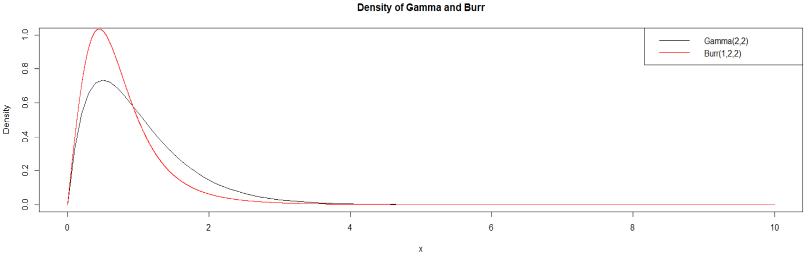
The material covered in this presentation is relevant in high dimensional statistics and machine learning applications.

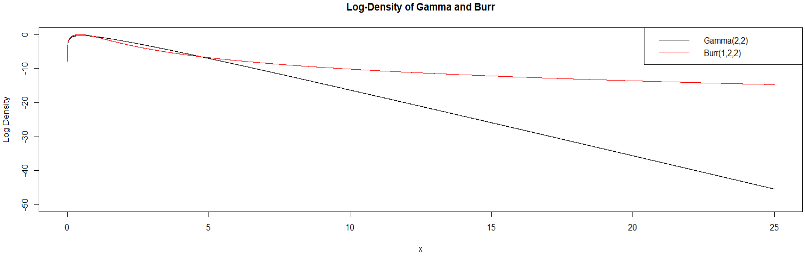
Introduction to Extreme Value Distributions (EVDs)

Mikosch's Regular Variation, Subexponentiality, and Their Applications in Probability Theory [3] provides an intuitive introduction to Extreme Value Distributions and their properties by first defining Regular Variation and Stable Distributions.

Heavy Tails

Introduction to Extreme Value Distributions





There is no universal definition for 'heavy-tailed' distributions, but we expect there to be some kind of strange behavior in samples caused by sample extrema. According to Mikosch [3], the exponential distribution is usually considered as the borderline between heavy- and light-tailed distributions.

Mikosch covers two classes of heavy-tailed distributions: Regularly Varying and Subexponential Distributions. In the following slides we will focus on Regularly Varying distributions.

Definition 1.1 ([3, p. 7])

*A positive measurable function f is **regularly varying** (at infinity) with index $\alpha \in \mathbb{R}$ if*

- It is defined on some neighborhood of infinity $[x_0, \infty)$

- $\lim_{x \to \infty} \frac{f(tx)}{f(x)} = t^{\alpha} \qquad \forall t > 0$

Definition 1.2 ([3, p. 7])

*f is **slowly varying** (at infinity) if f is regularly varying with $\alpha = 0$*

Remark 1.3 ([3, p. 7])

Every regularly varying f with index α has representation

$$f(x) = x^{\alpha} L(x)$$

for some slowly varying function L

Example 1.4 ([3, p. 8])

Slowly varying: Positive constants, logarithms and iterated logarithms
Regularly varying: $x^{\alpha}, (x \log(1 + x))^{\alpha}, x^{\alpha} \log(\log(e + x))$
Not regularly varying: $2 + \sin x, \exp\{ln(1 + x)\}$

Theorem 1.5 ([3, p. 9])

If f is regularly varying with index α, then for $0 < a \leq b < \infty$ the convergence of

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha}$$

is uniform (in t) on $[a, b]$ for $\alpha = 0$, $(0, b]$ if $\alpha > 0$, and $[a, \infty)$ if $\alpha < 0$

Definition 1.6 ([3, p. 11])

For any positive functions f and g, $f(x) \sim g(x)$ as $x \to \infty$ if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$$

We say f and g have the same tail behavior.

Definition 1.7

A non-negative random variable X and its distribution are ***regularly varying*** *with index $\alpha \geq 0$ if $\bar{F}_x$ is regularly varying with index $-\alpha$*
where $\bar{F}_x = 1 - F_x$ is the right distribution tail.

Lemma 1.8 (Convolution closure of regularly varying distributions [3, p. 12])

Let X, Y be independent, non-negative, regularly varying random variables with index $\alpha \geq 0$. Then $X + Y$ is regularly varying with index α. Additionally, as $x \to \infty$ $P(X + Y > x) \sim P(X > x) + P(Y > x)$

Remark 1.9 ([3, p. 12])

If X, Y are non-negative and regularly varying with index α_X, α_Y and $\alpha_X < \alpha_Y$, then $X + Y$ is regularly varying with index α_X

Remark 1.10 ([3, p. 12])

If X, Y are non-negative random variables s.t. $P(Y > x) = o(P(X > x))$ and X is regularly varying with index α, then as $x \to \infty$ $P(X + Y > x) \sim P(X > x)$

The previous slides immediately lead to our first maximal value result:

Corollary 1.11 ([3, p. 13])

Let $X, X_1, \ldots, X_n$ be iid non-negative regularly varying random variables and $S_n = X_1 + \cdots + X_n$. Then as $x \to \infty$,
$P(S_n > x) = P(X_1 + \cdots + X_n > x) \sim nP(X > x)$
And if we write $M_n = \max_{i=1,\ldots,n} X_i$, then
$P(S_n > x) \sim nP(X > x) \sim P(M_n > x)$

That is, for large enough x, $\{S_n > x\}$ is essentially due to $\{M_n > x\}$ and M_n is regularly varying with the same index as X.

Definition 1.12 ([3, p. 14])

*A random variable Y and its distribution are **stable** if for iid copies Y_1, Y_2 of Y, and all choices of non-negative constants c_1, c_2, $a, b \in \mathbb{R}$ s.t.*

$$c_1 Y_1 + c_2 Y_2 \overset{d}{=} aY + b$$

Remark 1.13 ([3, p. 14])

For a stable Y, we can find constants a_n for any n s.t. $S_n \overset{d}{=} a_n Y + b_n$

Definition 1.14 ([3, p. 15])

A non-degenerate random variable X and its distribution are ***max*** $-$ ***stable*** *if they satisfy the relation*

$$M_n \stackrel{d}{=} c_n X + d_n \quad \forall n \geq 2$$

for iid $X, X_1, X_2, \ldots$ and appropriate constants $c_n > 0, d_n \in \mathbb{R}$

Remark 1.15 ([3, p. 15])

If (X_n) is a sequence of iid max-stable random variables, then

$$c_n^{-1}(M_n - d_n) \stackrel{d}{=} X$$

From the previous slide, max-stable distributions are limit distributions for maxima of iid random variables. In particular,

Theorem 1.16 ([3, p. 17])

The class of max-stable distributions is equivalent to the class of all possible non-degenerate limit distributions for normalized maxima of iid random variables.

The next slide introduces the main result we have been working towards, and the basis of classical extreme value theory.

Theorem 1.17 (Fisher-Tippet theorem [3, p. 17])

For a sequence of iid random variables (X_n) *and their maximum* M_n*, if there exist constants* $c_n > 0, d_n \in \mathbb{R}$*, and non-degenerate distribution* H *s.t.* $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$*, then* H *belongs to one of the following:*

$$\textit{Fréchet}: \Phi_{\alpha>0}(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-1}\}, & x > 0 \end{cases}$$

$$\textit{Weibull}: \varphi_{\alpha>0}(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1 & x > 0 \end{cases}$$

$$\textit{Gumbel}: \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}$$

Distributions of Angles in Random Packing on Spheres

Cai, Fan, and Jiang [1] derive both the limiting empirical distributions of random angles and the limiting distributions of extreme angles.

To be more precise, these random angles are the pairwise angles of n uniformly distributed random unit vectors in $\mathbb{R}^p$ in two scenarios:

- $n \to \infty$, p fixed
- $n \to \infty$, p growing with n

The focus here will be on the limiting distributions of the extreme angles.

First, consider the scenario where $n \to \infty$ but $p \geq 2$ is fixed.
Let Θ_{ij} be the angle between the ith and jth random unit vector $(i \neq j)$.

Theorem 2.1 ([1, p. 1839])

With probability one, μ_n, the empirical distribution of the angles Θ_{ij}, converges weakly as $n \to \infty$ to density

$$h(\theta) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p+1}{2})} (\sin\theta)^{p-2}, \quad \theta \in [0, \pi]$$

Remark 2.2 ([1, p. 1840])

Θ_{ij}'s are identically distributed, so $h(\theta)$ is the pdf of any Θ_{ij}, with support $[0, \pi]$

We denote

$$\Theta_{\text{min}} = \min\{\Theta_{ij};\ 1 \le i < j \le n\}, \Theta_{\text{max}} = \max\{\Theta_{ij};\ 1 \le i < j \le n\}$$

Theorem 2.3 ([1, p. 1840])

Both $n^{2/(p-1)}\Theta_{\text{min}}$ *and* $n^{2/(p-1)}(\pi - \Theta_{\text{max}})$ *converge weakly as* $n \to \infty$ *to*

$$F(x) = \begin{cases} 1 - \exp\{-Kx^{p-1}\}, & x \ge 0; \\ 0, & x < 0 \end{cases}$$

Where

$$K = \frac{1}{4\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p+1}{2})}$$

Distribution of Extreme Angles (fixed p)

Remark 2.4 ([1, p. 1840])

The previous theorem gives us that, as $n\uparrow$, Θ_{min} *is close to zero and* Θ_{max} *is close to* π

Theorem 2.5 ([1, p. 1842])

$n^{2/(p-1)}(\Theta_{\max} + \Theta_{\min} - \pi)$ converges weakly to the distribution of $X - Y$, where X, Y are iid with cdf $F(x)$ from Theorem 2.3

$$F(x) = \begin{cases} 1 - \exp\{-Kx^{p-1}\}, & x \geq 0; \\ 0, & x < 0 \end{cases}$$

Remark 2.6 ([1, p. 1843])

Though $\Theta_{\min}$ and $\pi - \Theta_{\max}$ have identical distributions, $n^{2/(p-1)}\Theta_{\min}$ and $n^{2/(p-1)}(\pi - \Theta_{\max})$ are asymptotically independent and don't vanish as $n \to \infty$, so their difference is non-degenerate.

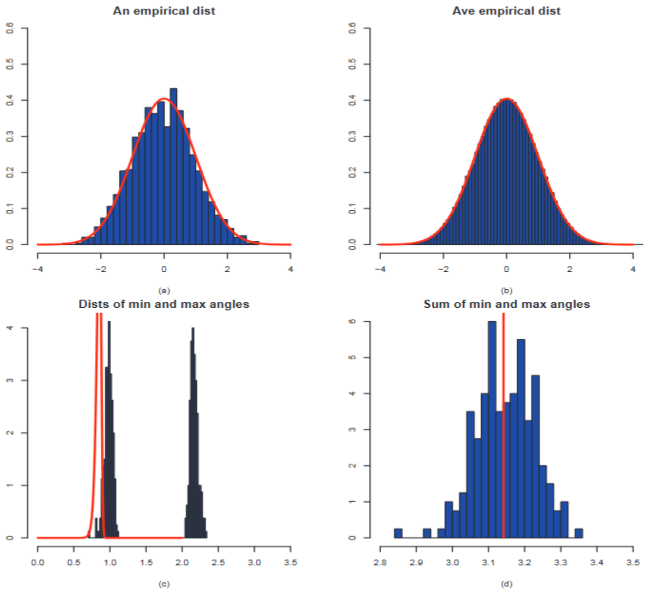
Distribution of Sum Of Max and Min Angles (fixed p)

Remark 2.7 ([1, p. 1843])

Since X, Y are iid, $X - Y$ is symmetric. Then, $\Theta_{\text{min}} + \Theta_{\text{max}}$ are symmetric around π.

Simulations in Cai et al. [1] show this symmetry around π. On the next slide is one example, with $p = 30$ and $n = 50$

Distributions of Angles in Random Packing on Spheres



Next, consider the scenario where $n \to \infty$ and $\lim_{n\to\infty} p = \infty$. It is appropriate here to use the normalized empirical distribution, denoted $\mu_{n,p}$.

Theorem 2.8 ([1, p. 1844])

If $\lim_{n\to\infty} p = \infty$, then with probability one, $\mu_{n,p}$ converges weakly to $N(0,1)$

Remark 2.9 ([1, p. 1840])

This theorem does not depend on the rate at which p goes to ∞

Since normalizing involves centering (subtracting $\frac{\pi}{2}$), Theorem 2.8 tells us that most of the random angles go to $\frac{\pi}{2}$ rapidly. Remembering that the Pearson correlation coefficient is the cosine of the angle, this tells us that most of the correlations rapidly go to 0! That is, most random vectors in high-dimensional Euclidean spaces are nearly orthogonal.

The convergence of $\Theta_{\min}$ and $\Theta_{\max}$ has 3 cases when p grows with n:

Theorem 2.10 ([1, p. 1845-1846])

- *Sub-Exponential Case* $\frac{\log n}{p} \to 0$ *as* $n \to \infty$*:*
  Both $\Theta_{\min}$ *and* $\Theta_{\max}$ *converge in probability to* $\pi/2$
- *Exponential Case* $\frac{\log n}{p} \to \beta \in (0, \infty)$ *as* $n \to \infty$*:*
  $\Theta_{\min} \xrightarrow{p} \cos^{-1}\sqrt{1 - \exp\{-4\beta\}}$; $\Theta_{\max} \xrightarrow{p} \pi - \cos^{-1}\sqrt{1 - \exp\{-4\beta\}}$
- *Super-Exponential Case* $\frac{\log n}{p} \to \infty$ *as* $n \to \infty$*:*
  $\Theta_{\min} \xrightarrow{p} 0$; $\Theta_{\max} \xrightarrow{p} \pi$

The fixed p case is similar to the Super-Exponential case but with weak convergence instead of convergence in probability.

Some Geometric Applications of the Beta Distribution

Frankl and Maehara [2] cover several properties and applications of the beta distribution.

In particular, we are interested in tying our extremal distributions of random angles and of Pearson correlations coefficients ρ to ρ^2, or R^2.

The result on the next slide doesn't just tell us the distribution of ρ^2 for any pair of random vectors, but for any random vector along with any random k-space.

Let L be a fixed 1-space (line) in $\mathbb{R}^n$, and let H be a random k-space in $\mathbb{R}^n$. That is, for $v_1, ..., v_k$ independent random points ($k < n$) in $\mathbb{R}^n$ with mean $\boldsymbol{O}$ (origin) and covariance I (identity), let H be the k-dimensional linear subspace spanned by $\boldsymbol{O}v_i$. Let θ be the angle between L and H.

Theorem 3.1 ([2, p. 464])

The random variables $\cos^2(\theta)$ *and* $\sin^2(\theta)$ *have the beta distributions Beta*$(k/2, (n-k)/2)$ *and Beta*$(n-k, (n-k)/2)$*, respectively.*

$N(\boldsymbol{O}, I)$ is 'isotropic', so we may assume L is also random. First take a random k-space H, then take a random point v and determine the random line $L = \boldsymbol{O}v$, where $v = (z_1, ..., z_n)$.
Since z_i's are iid $N(0, 1)$, $\sum_{i=1}^{k} z_i^2$ and $\sum_{i=k+1}^{n} z_i^2$ are independent chi-square random variables with k and $n - k$ degrees of freedom, respectively.
Then $\cos^2(\theta) = \sum_{i=1}^{k} z_i^2 / \sum_{i=1}^{n} z_i^2$ is a Beta random variable. □ $(*)$

(*) If X, Y are independent chi-square random variables with degrees of freedom a,b, then $X/(X + Y)$ is distributed as Beta$(a/2, b/2)$

To tie this beta distribution result back to EVDs, we return to Mikosch [3]. The beta distribution is regularly varying, and belongs to the 'Maximum Domain of Attraction' of the Weibull distribution

$$\varphi_{\alpha>0}(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1 & x > 0 \end{cases}$$

This means that for maximum M_n right endpoint x_F, and constants c_n, which can be chosen as x_F minus the $(1 - n^{-1})$ quantile of F, we have that

$$c_n^{-1}(M_n - x_F) \xrightarrow{d} \varphi_{\alpha}$$

And so we have a closed form limiting distribution for properly normalized maxima of $\cos^2$ of angles between random lines and random k-spaces.

Recap

- Fisher-Tippet theorem [3] tells us that Extrema are asymptotically distributed as either a Fréchet, Weibull, or Gumbel Distribution
- When $n \to \infty$ and p fixed, most pairwise angles are close to $\pi/2$ (and so most pairwise ρ's are close to 0) but the minimum and maximum pairwise angles are close to 0 and π, respectively
- When p also goes to ∞ normalized empirical distribution $\mu_{n,p}$ converges weakly to $N(0,1)$, most pairwise angles go to $\pi/2$ rapidly (and so most pairwise ρ's rapidly go to 0). Additionally, the minimum and maximum pairwise angles converge in probability, but to different values depending on the rate at which $p \to \infty$
- The $\cos^2$ of the angle, or R^2, between a random line and a random k-space is a Beta random variable
- The beta distribution has a regularly varying right tail, and normalized maxima of iid Beta random variables are asymptotically Weibull

References

[1] Tony Cai, Jianqing Fan, and Tiefeng Jiang. Distributions of angles in random packing on spheres. *Journal of Machine Learning Research*, 14(21):1837–1864, 2013.

[2] Peter Frankl and Hiroshi Maehara. Some geometric applications of the beta distribution. *Annals of the Institute of Statistical Mathematics*, 42(3):463–474, 1990.

[3] T. Mikosch. *Regular variation, subexponentiality and their applications in probability theory*. Report Eurandom. Eurandom, 1999.