Distributions of Angles, Beta Distributions, and their Extreme Value Distributions

May 22, 2021

Abstract

This report provides an overview of papers covering extreme value distributions, distributions of random angles, and applications of the beta distribution. Though the discussion of random angles includes scenarios where p grows with n, this report primarily focuses on the fixed p scenario, and includes both a proof and simulations supporting the main result for extreme values of angles in this scenario.

1 Motivation

High-dimensional statistical inference methods are developed to work with data with a large number p of explanatory variables and often a large number n of samples. It is a well known result that as the number of dimensions tends to infinity, almost all vectors are nearly orthogonal. This has implications for correlations in large datasets, since the correlation coefficient ρ for two random vectors is the cosine of the angle between those random vectors. Cai et al. [1] derive limiting distributions for random angles and their extreme values, and Frankl and Maehara [2] provide results on the beta distribution that have implications for the R^2 value given a response vector Y and any number p of explanatory variables $X_1, \ldots, X_p$. It helps to have knowledge of the basics of extreme value theory and the Fisher-Tippet theorem, which are both covered by Mikosch [3].

2 Overview of Papers

2.1 Regular Variation and Fisher-Tippet Theorem

Mikosch's "Regular Variation, Subexponentiality, and Their Applications in Probability Theory" defines regular variation and subsequently derives key theorems and lemmas related to extreme value theory.

Though no universal definition of heavy-tailed distributions can exist, but in the context of studying extrema of iid random samples, we often set the cutoff for 'light-tailed' at the exponential distribution. That is, a heavy-tailed distribution is ones with tails decay slower than that of the exponential distribution. Figure 1 shows an example of a light-tailed (Gamma) distribution and a heavy-tailed (Burr) distribution. Two classes of heavy-tailed distributions studied widely are regularly varying and subexponential distributions. Though Mikosch covers both, we just need the former to derive our key result on extreme value distributions.

Definition 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 2 ([3, p. 7]). *f is* ***slowly varying*** *(at infinity) if f is regularly varying with $\alpha = 0$*

Remark 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

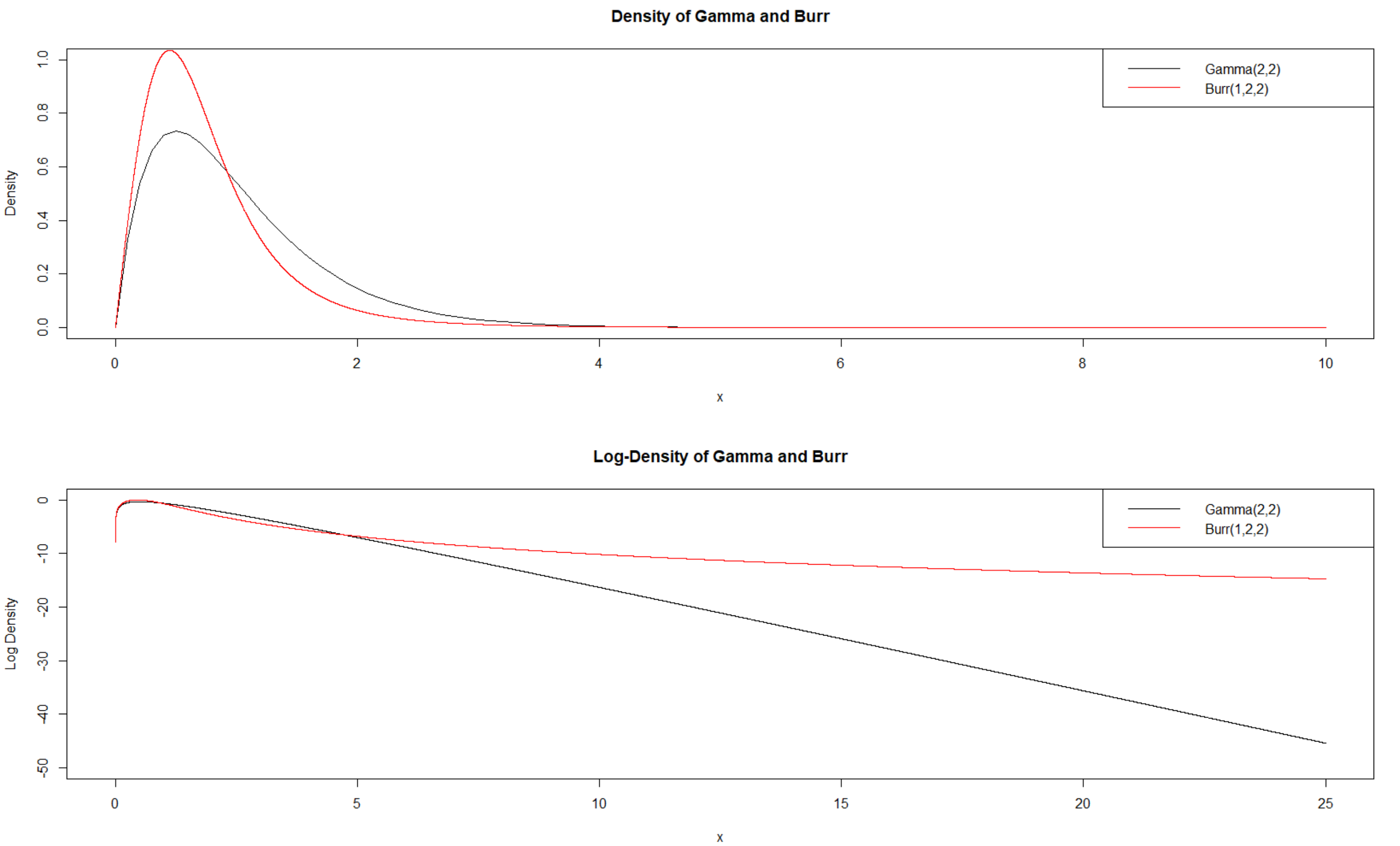

Figure 1: Tail behavior of Gamma(2,2) and Burr(1,2,2) Distributions.

We have now established our concept of regular variation. The next theorem and definition give us conditions for the conditions in Definition 1 to be uniform, and a notation of equivalence between two regularly varying functions.

Theorem 4 ([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 5 ([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.

Now that we have covered the important details of regularly varying functions, we can narrow our focus to random variables.

Definition 6 ([3, p. 12]). *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 7 ([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 8 ([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 9 ([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 $x \to \infty$ $P(X + Y > x) \sim P(X > x)$*

That is, if X, Y are regularly varying with index α then the event $\{X + Y > x\}$ is essentially due to the events of $\{X > x\}$ and $\{Y > x\}$ individually. However, if Y is not regularly varying but of smaller order than X then the event $\{X + Y > x\}$ is essentially due to the event $\{X > x\}$. These results immediately lead to our first result related to maximal values.

Corollary 10 ([3, p. 13]). *Let $X, X_1, ..., X_n$ 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 $M_n = \max_{i=1,...,n} X_i$, 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. We are now close to the Fisher-Tippet theorem, but we first need to make some requirements on the behaviour of the maxima of iid random samples. Specifically, we need to define max-stability.

Definition 11 ([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.*

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

Remark 12 ([3, p. 14]). *For a stable Y, we can find constants a_n for any n s.t. $S_n \stackrel{d}{=} a_n Y + b_n$*

Theorem 13 ([3, p. 14]). *A stable random variable X has characteristic function*

$$\phi_X(t) = E\exp\{iXt\} = \exp\{i\gamma t - c|t|^\alpha(1 - i\beta sign(t) z(t, \alpha)\}$$

With γ a constant, $c > 0, \alpha \in (0, 2], \beta \in [-1, 1]$,

$$z(t, \alpha) = \begin{cases} tan(\frac{\pi\alpha}{2}) & \alpha \neq 1 \\ \frac{-2}{\pi} ln|t| & \alpha = 1 \end{cases}$$

Definition 14 ([3, p. 14]). *α determines essential properties of X, so we may refer to X as $\boldsymbol{\alpha-}$ **stable**.*

2-stable distributions are Gaussian. For $\alpha < 2$, $\alpha-$stable distributions have infinite variance, and generally cannot be represented with elementary functions. The $1-$stable distribution (Cauchy) is one of few exceptions.

Definition 15 ([3, p. 15]). *A random variable X and its distribution F belong to the **domain of attraction** of the $\alpha-$stable distribution G_α if constants $a_n > 0, b_n \in \mathbb{R}$ s.t. the following holds:*

$$a_n^{-1}(S_n - b_n) \stackrel{d}{\to} G_\alpha \quad asn \to \infty$$

The domain of attraction gives us conditions on the distribution of X so that $a_n^{-1}(S_n - b_n)$ converges in distribution to an $\alpha-$stable random variable. Oftentimes, it is enough to know that X belongs to the domain of attraction of some non-specified $\alpha-$stable distribution, denoted $X \in DA(\alpha)$

Theorem 16 ([3, p. 15]). *A random varibale X and its distribution F belong to the DA of a Normal distribution iff*

$$\int_{|y|} y^2 dF(y) \text{ is slowly varying}$$

The central limit theorem provides us the desired results for $X \in DA(2)$, so we are more interested in the case where this integral is not slowly varying, and where the variance is infinite. In this case, the tail behavior of X is closely related to the tail behavior of its limiting $\alpha-$stable distribution. To tie this into extreme value distributions, we introduce max-stability.

Definition 17 ([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 $X, X_1, X_2, ...$ and appropriate constants $c_n > 0, d_n \in \mathbb{R}$

Remark 18 ([3, p. 17]). *If (X_n) is a sequence of iid max-stable random variables, then*

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

Theorem 19 ([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.*

Finally, we arrive at the Fisher-Tippet theorem.

Theorem 20 ([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:*

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

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

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

The proof of this theorem is very technical, and involves the convergence to types theorem and solving functional equations.

2.2 Random Angles and their Limiting Distributions

In "Distributions of Angles in Random Packing on Spheres" Cai et al. derive asymptotic properties of angles between unit vectors generated by random points distributed uniformly on the unit sphere in $\mathbb{R}^p$.

The limiting distributions of random angles are studied under two scenarios, for p fixed and p growing with n. Both results are summarized here, and both a complete proof and simulations are shown for the fixed p case in later sections. To start, we 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 $(i \neq j)$ in $\mathbb{R}^p$.

Theorem 21 ([1, p. 1839]). *With probability one, μ_n, the empirical distribution of the angles Θ_{ij}, converges weakly as $n \to \infty$ to density*

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

Remark 22 ([1, p. 1840]). *Θ_{ij}'s are identically distributed, so $h(\theta)$ is the pdf of any Θ_{ij}, with support $[0, \pi]$*

Let

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

Theorem 23 ([1, p. 1840]). *Both $n^{2/(p-1)}\Theta_{\min}$ and $n^{2/(p-1)}(\pi - \Theta_{\max})$ 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

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

Remark 24 ([1, p. 1840]). *As $n \uparrow$, $\Theta_{\min}$ is close to zero and $\Theta_{\max}$ is close to π*

The full proof of Theorem 23 is given in section 3 of this report.

We now have our limiting empirical distribution and the limiting extreme value distributions for random angles in $\mathbb{R}^p$ for fixed p. Before moving onto the growing p case, we have a useful theorem for the limiting distribution of the sum of the minimum and maximum random angles.

Theorem 25 ([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 23*

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

Remark 26 ([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.*

Remark 27 ([1, p. 1843]). *Since X, Y are iid, $X - Y$ is symmetric. Then, $\Theta_{\min} + \Theta_{\max}$ are symmetric around π.*

Now we consider the scenario where $n \to \infty$ and $\lim_{n\to\infty} p = \infty$. Here we use the normalized empirical distribution $\mu_{n,p}$.

Theorem 28 ([1, p. 1844]). *If $\lim_{n\to\infty} p = \infty$, then with probability one, $\mu_{n,p}$ converges weakly to $N(0, 1)$*

Since normalizing involves centering (subtracting $\frac{\pi}{2}$), Theorem 28 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. This is one way to show that most random vectors in high-dimensional Euclidean spaces are nearly orthogonal.

It's important to note that this theorem does not depend on the rate at which p goes to ∞. However, the limiting distribution for the minimal and maximal random angles depends on the rate at which p grows relative to n.

Theorem 29 ([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 $n \to \infty$:*

  $\Theta_{\min} \xrightarrow{p} 0$; $\Theta_{\max} \xrightarrow{p} \pi$

That is, in the sub-exponential case the minimum and maximum correlations go to 0, in the super-exponential case the minimum and maximum correlations go to -1 and 1 respectively, and in the exponential case the limit depends on the value . Notice that the fixed p case is similar to the super-exponential case, except in the super-exponential case we have a stronger mode of convergence.

2.3 R^2 and the Beta Distribution

In "Some Geometric Applications of the Beta Distribution", Frankl and Maehara cover various scenarios where the Beta distribution arises, one of which is the angle between a line and a 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 30 ([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.*

A full proof is given in section 3.

Observing that $cos^2(\theta)$ is $\rho^2 = R^2$, we have that for a response vector Y and (fixed) k iid explanatory random variables $X_1, \ldots, X_k$, the R^2 is a Beta-distributed random variable. In addition to having the limiting distributions for $\Theta_{\min}, \Theta_{\max}$ and so for $\rho_{\min}, \rho_{\max}$, if we find the extreme value distribution for iid Beta random variables, we can also derive the limiting distribution for $R^2_{\max}$.

From Mikosch [3], the beta distribution belongs to the Maximum Domain of Attraction (MDA) of the Weibull distribution

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

That is, 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 R^2.

3 Proof of Main Results

The proof of Theorem 23 requires covering some technical details and several lemmas first. The proof of Theorem 30 is quite short in comparison. Cai et al. cite other works for the proofs of lemmas, and those citations are given next to the lemma number. When other papers are cited in those proofs, the citation can be found inline.

3.1 Proof of Theorem 23

Recall that $X_1, X_2, \ldots$ are random points uniformly distributed on the unit sphere of $\mathbb{R}^p$, Θ_{ij} is the angle between $\vec{OX_i}$ and $\vec{OX_j}$, and $\rho_{ij} = \cos\Theta_{ij}$ for any $i \neq j$. The distribution of $(X_1, X_2, \ldots)$ is the same as the distribution of $(\frac{Y_1}{||Y_1||}, \frac{Y_1}{||Y_1||}, \ldots)$ for $Y_1, Y_2, ...$ iid $N_p(0, I_p)$. Thus

$$\rho_{ij} = \cos\Theta_{ij} = \frac{Y_i^T Y_j}{||Y_i|| \cdot ||Y_j||} \quad \forall i \neq j$$

Set

$$M_n = \max_{1\leq i<j\leq n} \rho_{ij} = \cos\Theta_{\min} \qquad \textbf{(1)}$$

Lemma 31 (Cai and Jiang (2012)[4, p. 15]). *Let* $p \geq 2$ *Then* ρ_{ij} *are pairwise independent and identically distributed with density function*

$$g(\rho) = \frac{\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p-1}{2})}(1-\rho^2)^{\frac{p-3}{2}},\ |\rho| < 1$$

Proof:

It is enough to prove that if $\{U, V, W\}$ are iid with an n-dimensional spherical distribution and $P(U = 0) = 0$, then $\rho_{U,V}$ and

$\rho_{U,W}$ are iid with density

$$\frac{\Gamma(\frac{p-1}{2})}{\sqrt{\pi}\Gamma(\frac{p-2}{2})}(1-\rho^2)^{\frac{p-4}{2}},\ |\rho|<1$$

where here the only difference from $g(\rho)$ is changing the degree of freedom from p to $p-1$.
Let $\{1\}$ be the span of 1. Then since $P(U=0)=0$, $Y := \frac{U}{||U||}$ is well defined. By definition, $OU \overset{P}{=} U$ for any orthogonal matrix O, so

$$OY = \frac{OU}{||OU||} \overset{P}{=} \frac{U}{||U||} = Y$$

That is, the probability measure generated by Y is an orthogonal-invariant measure on the unit sphere in $\mathbb{R}^p$. Y then has the uniform distribution on the unit sphere. Thus,

$$P(U \in \{1\}) = P(V \in \{1\}) = P(W \in \{1\}) = 0 \quad \text{where } \{1\} \text{ is the span of 1}$$

It follows from Jiang (2004) [5] and Muirhead (1982) [6] that $\rho_{U,V}$ and $\rho_{U,W}$ have the same density $f(\rho)$. Swapping p for $p-1$, we have our desired density $g(\rho)$
To show independence, we use that U, V, W are independent and so

$$E[g(\rho_{U,V})\dot{h}(\rho_{U,W})] = E\{E[g(\rho_{U,V})|U]\dot{E}[h(\rho_{U,W})|U]\}$$

Let $V = (V_1, \dots, V_p)^T \in \mathbb{R}^n$ and $\bar{V} = \frac{1}{n}\sum_{i=1}^n V_i$. For any numbers $u_1, \dots, u_n$ s.t. at least two are not identical, it follows from Muirhead (1982) [6] that

$$\rho_{u,V} = \frac{\sum_{i=1}^n (u_i - \bar{u})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^n (u_i - \bar{u})^2 \dot{\sum}_{i=1}^n (V_i - \bar{V})^2}}$$

has the density $f(\rho)$. So given U, the probability distribution of $\rho_{U,V}$ does not depend on the value of U. Thus,

$$E[h(\rho_{U,W})|U] = E[h(\rho_{U,W})] \text{ and } E[h(\rho_{U,V})|U] = E[h(\rho_{U,V})] \quad \square$$

Notice that $y = \cos x$ is strictly decreasing on $[0, \pi]$, hence $\Theta_{ij} = \cos^{-1}\rho_{ij}$.
Lemma 32 follows immediately from Lemma 31.

Lemma 32. *Let $p \geq 2$. Then Θ_{ij} are pairwise independent and identically distributed with density function*

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

If we replace Θ_{ij} with $\pi - \Theta_{ij}$, we get the same density.

Lemma 33 (Arratia et al. (1989)[7, p. 11])**.** *Let I be a finite set, and for each $\alpha \in I, X_\alpha$ be a Bernoulli random variable with $p_\alpha = P(X_\alpha = 1) = 1 - P(X_\alpha = 0) > 0$. Set $W = \sum_{\alpha \in I} X_\alpha$ and $\lambda = EW = \sum_{\alpha\in I} p_\alpha$. For each $\alpha \in I$, suppose we have chosen $B_\alpha \subset I$ with $\alpha \in B_\alpha$. Define*

$$b_1 = \sum_{\alpha\in I}\sum_{\beta\in B_\alpha} p_\alpha p_\beta \text{ and } b_2 = \sum_{\alpha\in I}\sum_{\alpha\neq\beta\in B_\alpha} P(X_\alpha = 1, X_\beta = 1)$$

For each $\alpha \in I$, assume X_α is independent of $\{X_\beta; \beta \in 1 - B_\alpha\}$. Then

$$|P(X_\alpha = 0 \ \ \forall\alpha \in I) - e^{-\lambda}| \leq b_1 + b_2$$

Proof
Define Z to be a Poisson random variable with $EZ = \lambda$, and h to be a function such that $||h|| = 1$. Let $\bar{h}(\cdot) = h(\cdot) - Eh(Z)$. Define linear operators S, T so that for some function f and for h

$$(Tf)(w) = wf(w) - \lambda f(w+1),\ w \geq 0$$

$$(Sh)(w+1) = -\lambda^{-1} P(Z = w)^{-1} E(h(Z); Z \leq w),\ w \geq 0$$

Define $f = S\bar{h}$, so that $Tf = \bar{h}$, and $E[Tf(W))] = E(h(W) - h(Z))$. Then we want to show that

$$|E\{h(w) - h(Z)\}| \leq (b_1 + b_2)||\Delta f|| + b_3'||f||$$

We first need bounds on $||f||$ and $||\Delta f||$, but we notice that if $Eh(Z) = 0$, then

$$(Sh)(w+1) = -\lambda^{-1} P(Z = w)^{-1} \text{cov}(h(z), (1(Z \leq w))$$

For fixed $k \geq 0$,

$$\text{cov}(h(z), (1(Z \leq w)) = P(Z \leq k \wedge w) - P(Z \leq k)P(Z \leq w)$$

Since $\frac{d}{d\lambda} P(Z \leq j) = -P(Z = j)$, we have

$$P(Z \leq j) = 1 - \int_0^\lambda e^{-v} v^j / j! dv = \int_\lambda^\infty e^{-v} v^j / j! dv$$

For $k = 0$, we have

$$(1 - e^{-\lambda})/\lambda = -f(1) > -f(2) > \cdots > 0$$

And so

$$||\Delta f|| \leq (1 - e^{-\lambda})/\lambda \text{ and } ||f|| \leq (1 - e^{-\lambda})/\lambda$$

Now we need to show that $|E\{h(w) - h(Z)\}| \leq (b_1 + b_2)||\Delta f|| + b_3'||f||$.
Let $V_\alpha := \sum_{\beta \in I - B_\alpha} X_\beta$ and W_α. We compute

$$E\{h(W) - h(Z)\} = E\{Wf(W) - \lambda f(W+1)\} = \sum_{\alpha \in I} E\{X_\alpha f(W) - p_\alpha f(W+1)\}$$

$$= \sum_{\alpha \in I} E\{p_\alpha f(W_\alpha + 1) - p_\alpha f(W_\alpha + 1)\} + \sum_{\alpha \in I} E\{X_\alpha f(W_\alpha) - p_\alpha f(W_\alpha + 1)\}$$

$$= \sum_{\alpha \in I} E\{p_\alpha X_\alpha [f(W_\alpha + 1) - f(W_\alpha + 2)]\} + \sum_{\alpha \in I} E\{(X - p_\alpha)[f(W_\alpha + 1) - f(V_\alpha + 1)]\} + \sum_{\alpha \in I} E\{(X_\alpha - p_\alpha) f(V_\alpha + 1)\}$$

$$\leq ||\Delta f|| \sum_{\alpha \in I} p_\alpha^2 + ||\Delta f|| (\sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} p_{\alpha\beta} + p_\alpha p_\beta) + b_3' ||f||$$

where

$$b_3' = \sum_{\alpha \in I} s_\alpha' \leq \sum_{\alpha \in I} s_\alpha = \sum_{\alpha \in I} E|E\{X_\alpha - p_\alpha| \sum_{\beta \in I - B_\alpha} X_\beta\}|$$

$$= (b_1 + b_2)||\Delta f|| + b_3'||f|| \leq b_1 + b_2)||\Delta f|| + b_3||f||$$

With our added condition that X_α is independent of $\{X_\beta;\ \beta \in I - B_\alpha\}$, $b_3 = 0$, and the final term simplifies to

$$(b_1 + b_2)||\Delta f||$$

And so

$$|P(W = 0) - e^{-\lambda}| \leq (b_1 + b_2)$$

and our desired result follows. □

The next lemma is just a special case of Lemma 33.

Lemma 34. *Let I be an index set and $\{B_\alpha, \alpha \in I\}$ be a set of subsets of I. Let also $\{\eta_\alpha, \alpha \in I\}$ be random variables. For a given $t \in \mathbb{R}$, set $\lambda - \sum_{\alpha \in I} P(\eta_\alpha > t)$. Then*

$$|P(\max_{\alpha \in I} \eta_\alpha \leq t) - e^{-\lambda}| \leq (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3)$$

where

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t) P(\eta_\beta > t), b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t), b_3 = \sum_{\alpha \in I} E|P(\eta_\alpha > t|\sigma(\eta_\beta, \beta \notin B_\alpha)) - P(\eta_\alpha > t)|$$

and $\sigma(\eta_\beta, \beta \notin B_\alpha)$ is the $\sigma-$algebra generated by $\{\eta_\beta, \beta \notin B_\alpha\}$. In particular, if η_α is independent of $\{\eta_\beta, \beta \notin B_\alpha\}$ for each α, $b_3 = 0$.

Proposition 35 ([1, p. 1854])**.** *Fix $p \geq 2$. Then $n^{4(p-1)}(1 - M_n)$ converges to distribution function*

$$F_1(x) = 1 - \exp\{-K_1 x^{\frac{p-1}{2}}\},\ x \geq 0 \qquad (2)$$

in distribution as $n \to \infty$, where

$$K_1 = \frac{2^{(p-5)/2}\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p+1}{2})}$$

Proof:
Set $t = t_n = 1 - xn^{-4/(p-1)}$ for $x \geq 0$. Then

$$t \to 1 \text{ and } t^2 = 1 - \frac{2x}{n^{-4/(p-1)}} + O(\frac{1}{n^{8(p-1)}}) \text{ as } n \to \infty$$

$$P(n^{4/(p-1)}(1 - M_n) < x) = P(M_n > t) = 1 - P(M_n \leq t)$$

Noting that $F_1(x)$ is continuous, to prove Proposition 32 it is then enough to show that

$$P(M_n \leq t) \to \exp\{-K_1 x(p-1)/2\} \text{ as } n \to \infty$$

Take $I = \{(i,j); 1 \leq i < j \leq n\}$. For $u = (i,j) \in I$, set $B_u = \{(k,l) \in I; \text{one of } k \text{ and } l = i \text{ or } j, \text{ but } (k, l \neq u)\}$
$\eta_u = \rho_{ij}$ and $A_u = A_{ij} = \{\rho_{ij} > t\}$
By Lemma 34,

$$|P(M_n \leq t) - e^{-\lambda_n}| \leq b_{1,n} + b_{2,n} \text{ where } \lambda_n = \frac{n(n-1)}{2} P(A_{12})$$

$$\text{and } b_{1,n} \leq 2n^3 P(A_{12})^2,\ b_{2,n} \leq 2n^3 P(A_{12}A_{13})$$

By Lemma 31, A_{12} and A_{13} are independent events with equal probability. Then whenever $n \geq 2$

$$b_{1,n} \vee b_{2,n} \leq 2n^3 P(A_{12})^2 \leq \frac{8n\lambda_n^2}{(n-1)^2} \leq \frac{32\lambda_n^2}{n}$$

Also by Lemma 31 we have

$$P(A_{12}) = \int_t^1 g(x)dx = \frac{\Gamma(\frac{p}{2})}{\sqrt{\pi}\Gamma(\frac{p-1}{2})} \int_t^1 (1 - x^2)^{\frac{p-3}{2}} dx$$

Let $m = \frac{p-3}{2} \geq -1/2$, and let $x = \sqrt{s}$ and so $dx = \frac{1}{2\sqrt{s}}ds$. Then

$$\int_t^1 (1-x^2)dx = \int_{t^2}^1 \frac{1}{2\sqrt{s}}(1-s)^m ds \;\sim\; \frac{1}{2}\int_{t^2}^1 (1-s)^m ds = \frac{1}{2m+2}(1-t^2)^{m+1}$$

Applying this, we have that as $n \to \infty$

$$\lambda_n \sim \frac{n^2\Gamma(\frac{p}{2}}{2\sqrt{\pi}\Gamma(\frac{p-1}{2})}\int_t^1 (1-x^2)^{\frac{p-3}{2}}dx \sim \frac{n^2\Gamma(\frac{p}{2})}{2\sqrt{\pi}(p-1)\Gamma(\frac{p-1}{2})}(1-t^2)^{\frac{p-1}{2}} = \frac{\Gamma(\frac{p}{2})}{4\sqrt{\pi}\Gamma(\frac{p+1}{2})}(n^{\frac{4}{p-1}}(1-t^2))^{\frac{p-1}{2}}$$

So as $n \to \infty$

$$n^{\frac{4}{p-1}}(1-t^2) = 2x + O(\frac{1}{n^{\frac{4}{p-1}}})$$

Therefore

$$\lambda_n \to \frac{2^{\frac{p-5}{2}}}{\sqrt{\pi}}\frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p+1}{2})}x^{\frac{p-1}{2}} = K_1 x^{\frac{p-1}{2}}$$

Then applying our bounds on $b_{1,n} \vee b_{2,n}$ and $|P(M_n \leq t) - e^{-\lambda_n}|$ we finally have

$$\lim_{n\to\infty} P(M_n \leq t) = -\exp\{-K_1 x^{\frac{p-1}{2}}\} \qquad \square$$

$M_n = \cos\Theta_{\min}$ by (1), so we use identity $1 - \cos h = 2\sin^2\frac{h}{2}$ for $h \in \mathbb{R}$ and get

$$n^{4/(p-1)}(1-M_n) = 2n^{4/(p-1)}\sin^2\frac{\Theta_{\min}}{2} \qquad (2)$$

By Proposition 35 and Slutsky's Theorem, $\sin\frac{\Theta_{\min}}{2} \to 0$ in probability as $n \to \infty$, which implies that $\Theta_{\min} \to 0$. From (2) and the fact that $\lim_{x\to 0}\frac{\sin x}{x} = 1$ we have

$$\frac{n^{4/(p-1)}(1-M_n)}{\frac{1}{2}n^{4/(p-1)}\Theta^2_{\min}} \xrightarrow{P} 1$$

Again using Proposition 35 and Slutskys Theorem, $\frac{1}{2}n^{4/(p-1)}\Theta^2_{\min}$ converges in distribution to $F_1(x)$. Additionally, for any $x > 0$

$$P(\frac{1}{2}n^{4/(p-1)}\Theta^2_{\min} \leq x) = P(n^{2/(p-1)}\Theta^2_{\min} \leq \frac{x^2}{2} \to 1 - \exp\{-K_1(x^2/2)^{(p-1)/2}\} = 1 - \exp\{-Kx^{p-1}\} \qquad (3)$$

where

$$K = 2^{(1-p)/2}K_1 = \frac{\Gamma(\frac{p}{2})}{4\sqrt{\pi}\Gamma(\frac{p+1}{2})}$$

Next, we must show that

$$n^{2(p-1)}(\pi - \Theta_{\max}) \text{ converges weakly to } F(x) \text{ as } n \to \infty$$

To show this, we only need to use a few properties of ρ_{ij}: That they are pairwise independent, that each has density $g(\rho)$ shown in Lemma 31, and that ρ_{ij} is independent of all $\rho_{k,l}$ where $\{k,l\} \cap \{i,j\} = \emptyset$.
Using Lemmas 31 and 32, these properties are equivalent to the following: Θij's are pairwise independent, Θij has the density $h(\theta)$ from Lemma 12, and that Θ_{ij} is independent of all $\Theta_{k,l}$ where $\{k,l\} \cap \{i,j\} = \emptyset$. Similarly, we have the same properties replacing Θ_{ij} with $\pi - \Theta ij$, where $\min(\pi - \Theta_{ij}) = \pi - \Theta_{\max}$.
Finally, recalling (3) we have

$$P(n^{2/(p-1)}(\pi - \Theta) \leq x) \to 1 - \exp\{-Kx^{p-1}\} \text{ as } n \to \infty \qquad \square$$

3.2 Proof of Theorem 30

Proof:

$N(\boldsymbol{O}, I)$ 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$(k/2, (n-k)/2)$ random variable.
Additionally, $sin^2(\theta) = \sum_{i=k+1}^{n} z_i^2 / \sum_{i=1}^{n} z_i^2$ is a Beta$((n-k)/2, k/2)$ random variable. □

4 Simulations

To keep this section brief, three scenarios are considered: p fixed at 2, 30, and at 1500 to represent a high-dimensional case (i.e. large omics data).

To be consistent with the paper by Cai et al., angles were computed by computing multivariate normal random variables of p dimension, and normalizing them to get points on the unit sphere in $\mathbb{R}^p$. All pairwise angles are then computed using the unit vectors generated from those points.

Figures 2-10 in the Appendix show the distributions of angles, their cosines, and their squared cosines as n grows for $p = 2, p = 30, p = 1500$, respectively. As is expected, for the fixed p scenario we actually don't have any change in the limiting distributions depending on the value of p itself. Additionally, also as expected, the rate of convergence does not appear to depend on p. With these figures we have a visual representation of the limiting distributions and tail behavior for the angles, correlations, and R^2 values themselves.

The proofs in section 3 of this report show the full derivation of the limiting distributions of the minima and maxima of random angles and of the R^2 values, and these simulations support the more general results on the asymptotic behavior of random angles for fixed p.

5 References

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.

[4] T. Tony Cai and Tiefeng Jiang. Phase transition in limiting distributions of coherence of high-dimensional random matrices. *Journal of Multivariate Analysis*, 107:24–39, 2012.

[5] Tiefeng Jiang. The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2), 2004.

[6] Robb John Muirhead. *Aspects of multivariate statistical theory.* John Wiley & Sons, 1982.

[7] R. Arratia, L. Goldstein, and L. Gordon. Two Moments Suffice for Poisson Approximations: The Chen-Stein Method. *The Annals of Probability*, 17(1):9 – 25, 1989.

6 Appendix

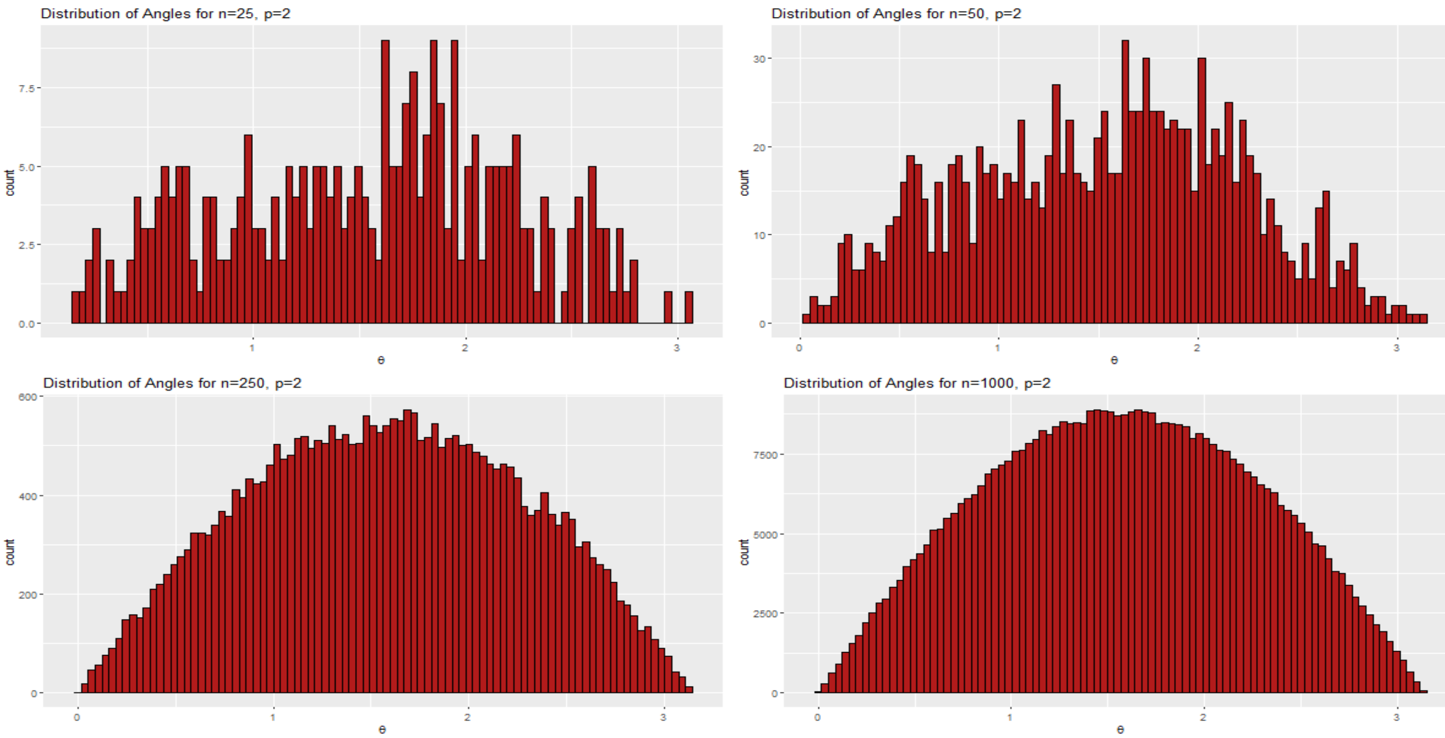

Figure 2: Distributions of Random Angles for $p = 2$.

Figure 3: Distributions of Cosines of Random Angles for $p = 2$.

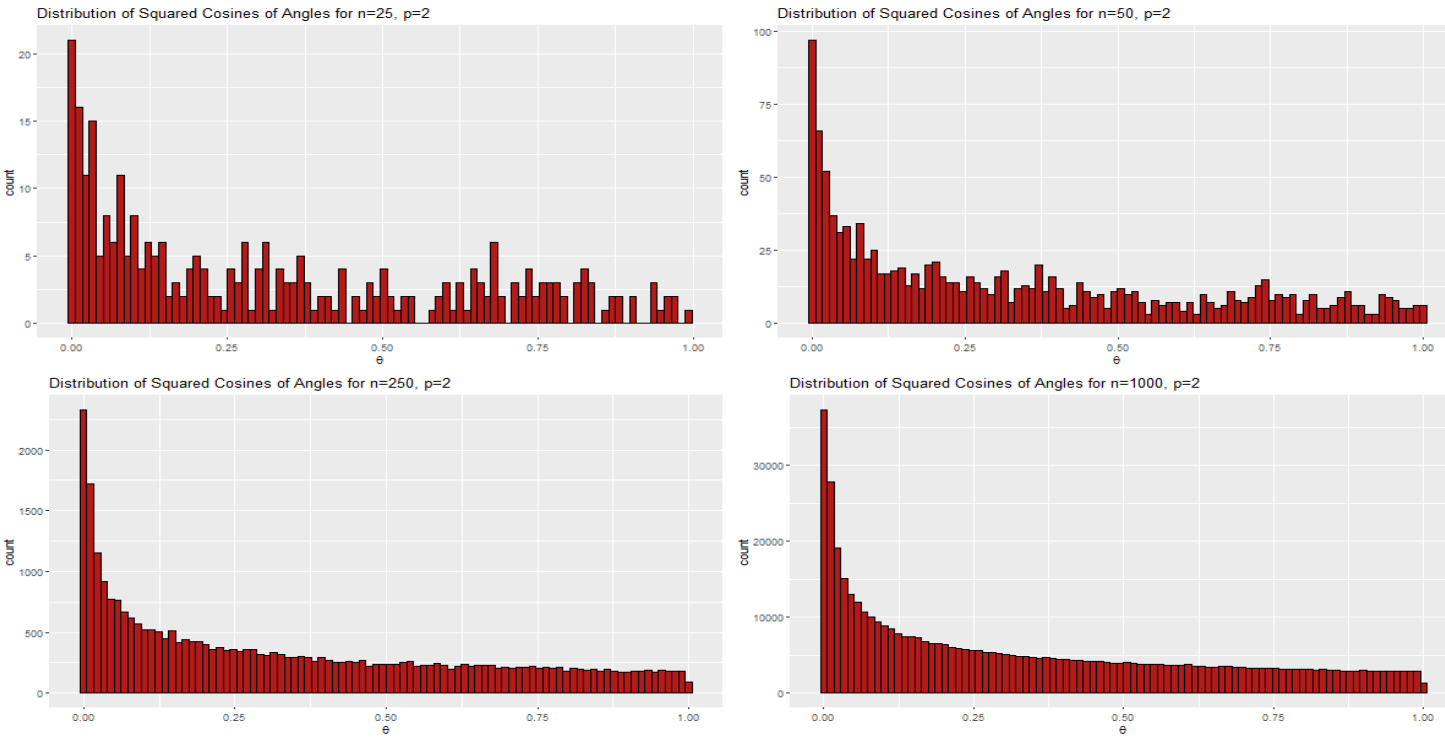

Figure 4: Distributions of Squared Cosines of Random Angles for $p = 2$.

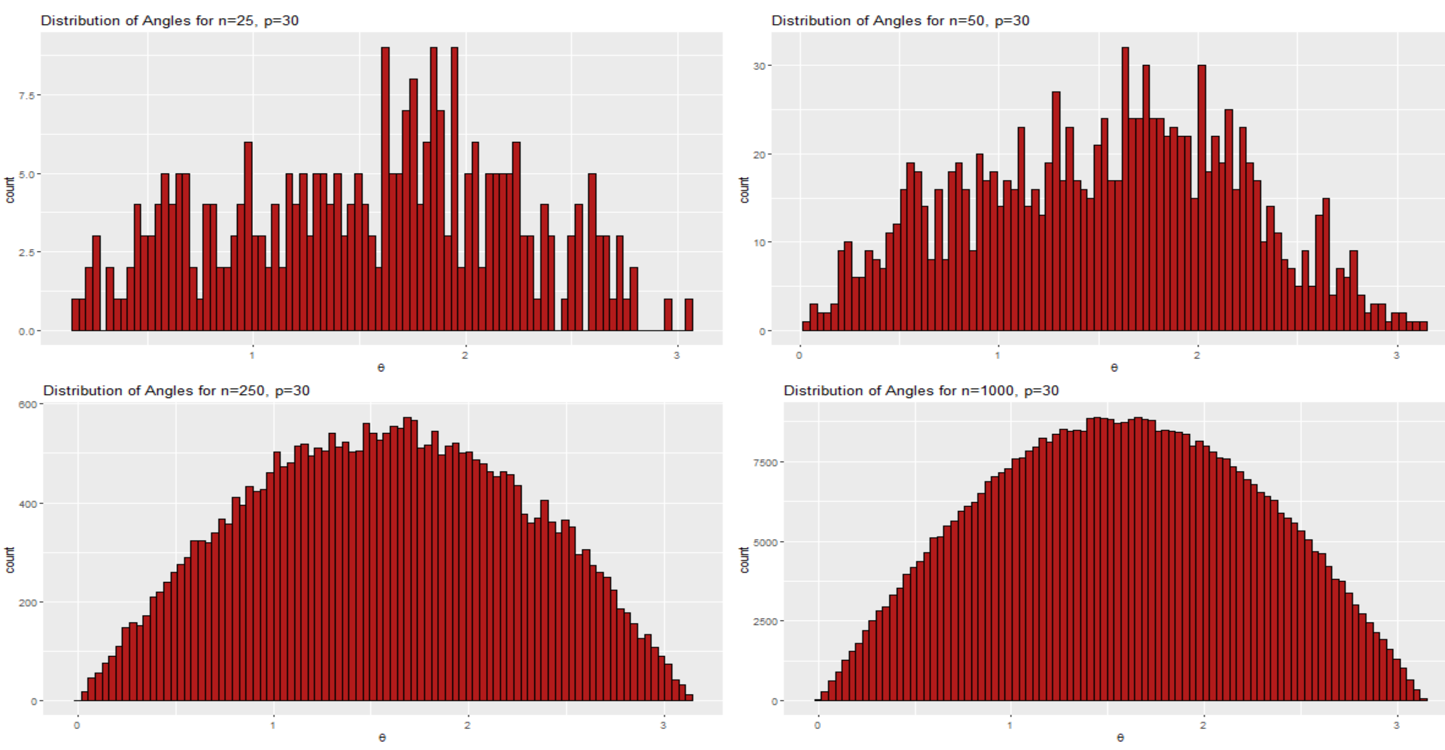

Figure 5: Distributions of Random Angles for $p = 30$.

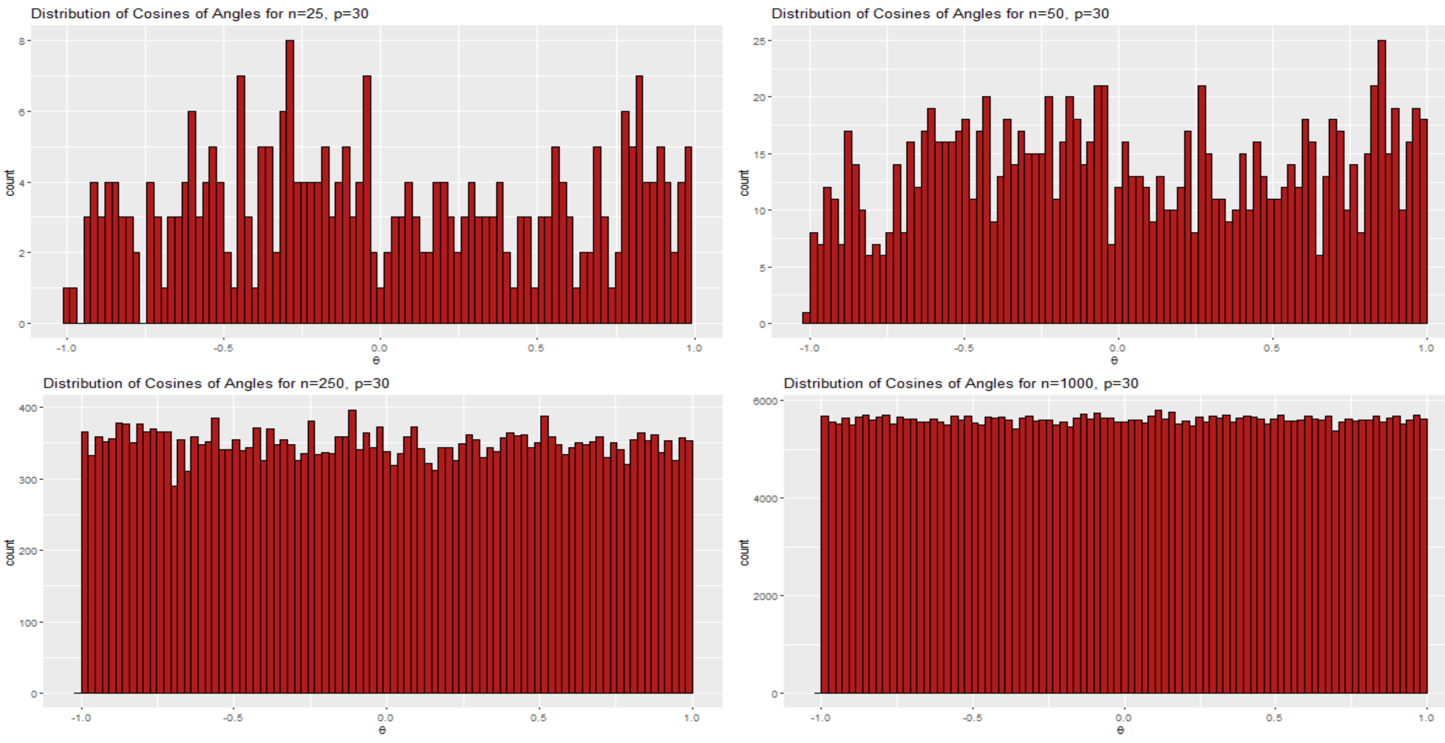

Figure 6: Distributions of Cosines of Random Angles for $p = 30$.

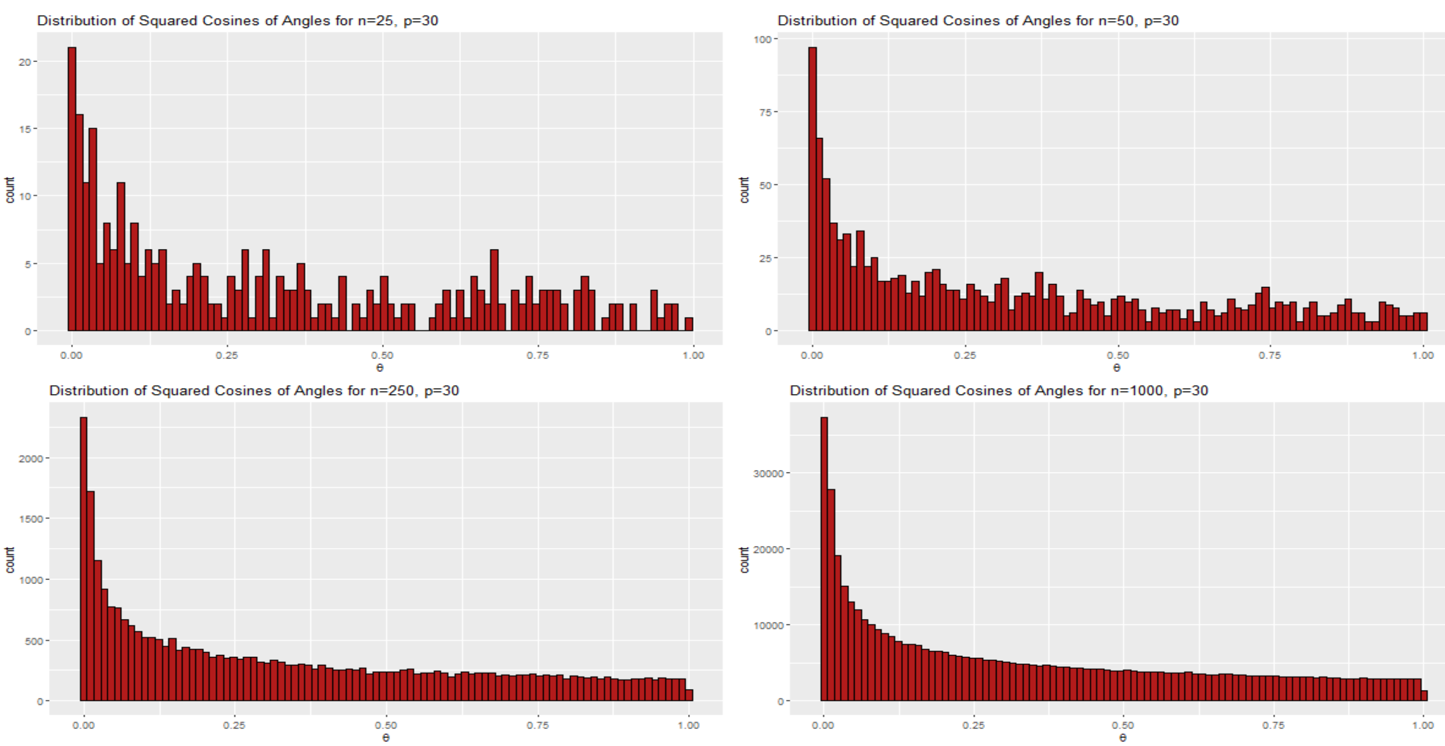

Figure 7: Distributions of Squared Cosines of Random Angles for $p = 30$.

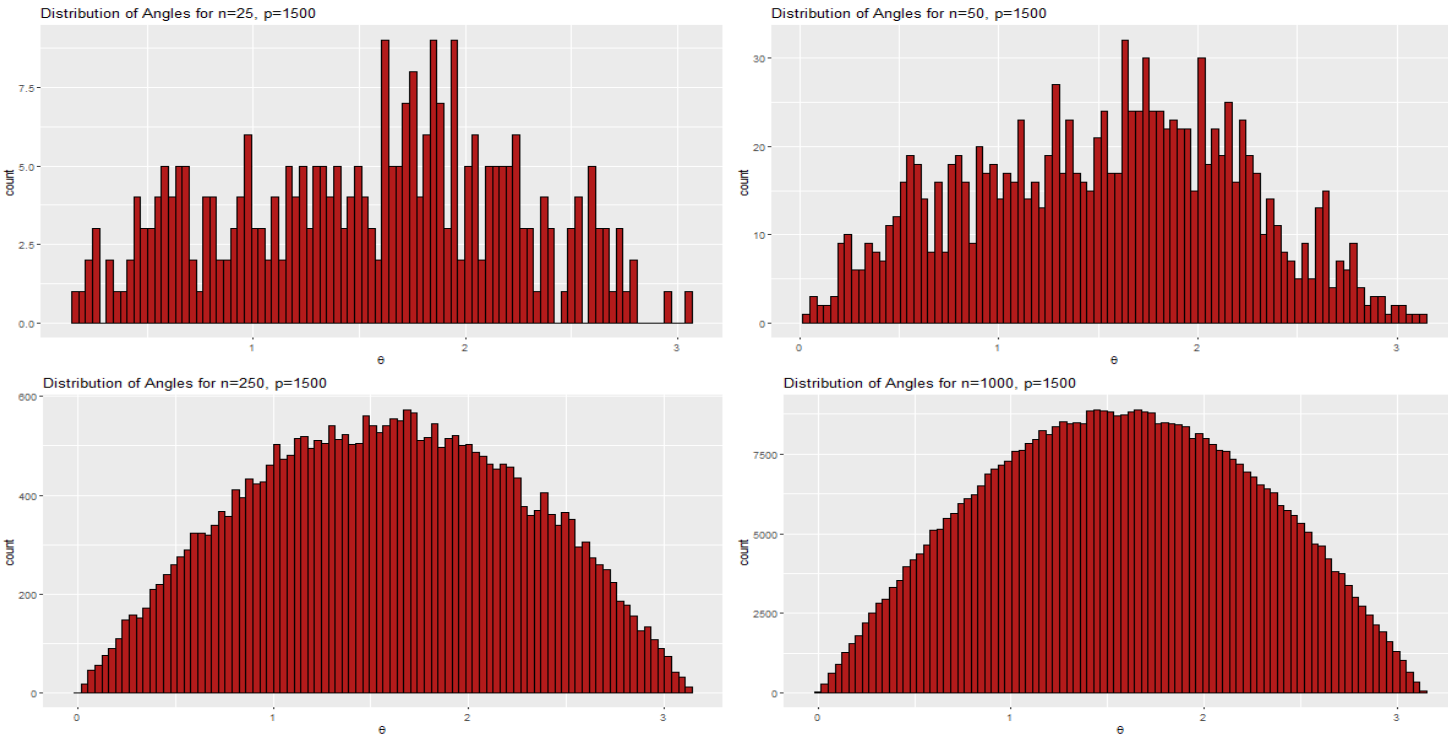

Figure 8: Distributions of Random Angles for $p = 1500$.

Figure 9: Distributions of Cosines of Random Angles for $p = 1500$.

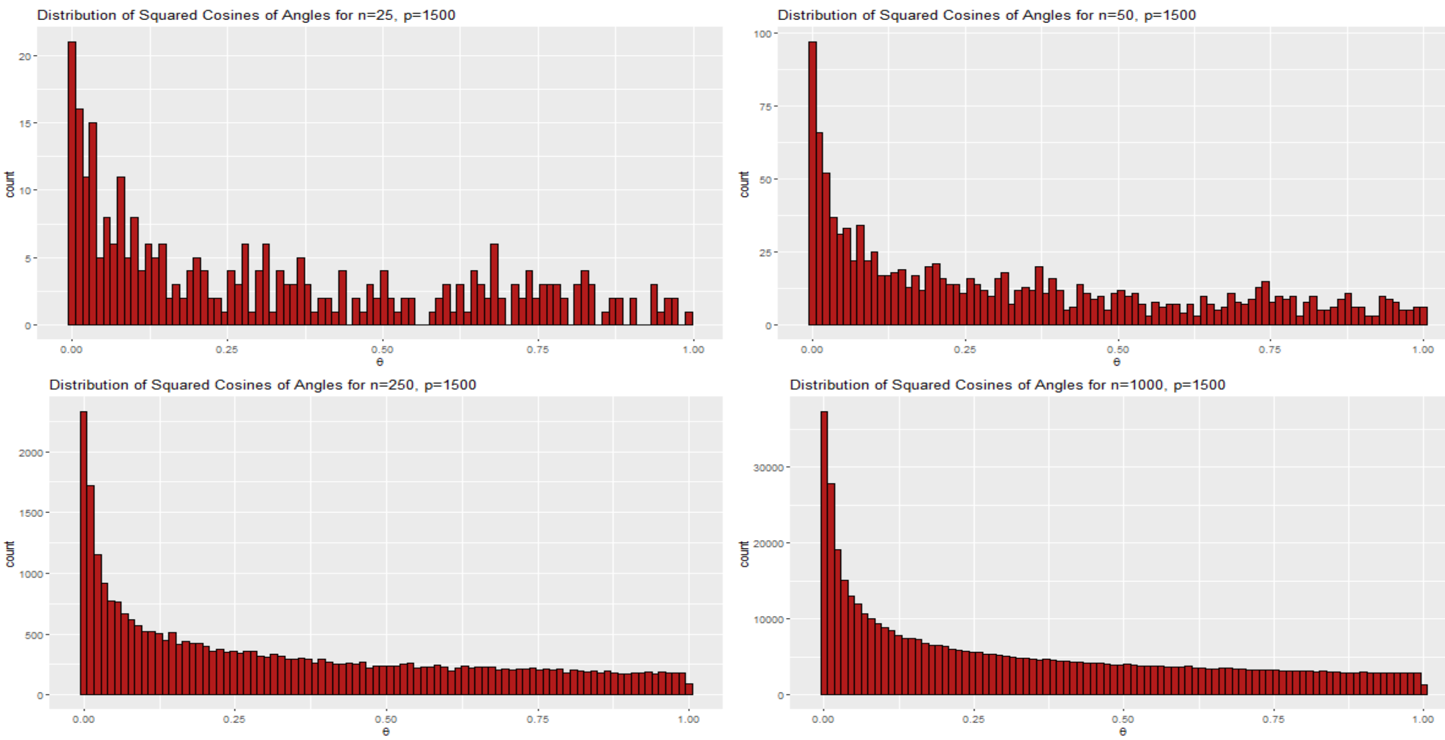

Figure 10: Distributions of Squared Cosines of Random Angles for $p = 1500$.