Submitted to the Annals of Applied Probability

HIGH DIMENSIONAL SPARSE COVARIANCE ESTIMATION: ACCURATE THRESHOLDS FOR THE MAXIMAL DIAGONAL ENTRY AND FOR THE LARGEST CORRELATION COEFFICIENT

By Aharon Birnbaum* and Boaz Nadler†

The Hebrew University and Weizmann Institute of Science†*

The maxima of many independent, or weakly dependent, random variables, and their corresponding thresholds for given right tail probabilities are classical and well studied problems. In this paper we focus on two specific cases of interest related to estimation and hypothesis testing of high dimensional sparse covariance matrices. These are the distribution of the maximal diagonal entry of a sample covariance matrix and the largest off-diagonal correlation coefficient, both under the assumption of an identity population covariance. In both cases, as sample size and dimension tend to infinity, upon centering and scaling, there is asymptotic convergence to a Gumbel distribution. We show, however, that this convergence is slow and that finite sample distributions may be quite far from these asymptotic ones. Applying a perturbation approach, we identify the leading error terms, and derive more accurate distributions and corresponding thresholds. For non-Gaussian data, these depend explicitly on higher order moments via appropriate Edgeworth expansions. As a side result, we also derive sharp bounds for the left and right tail probabilities of a single χ^2_n random variable, which may be of independent interest.

1. Introduction. Many contemporary applications require analysis of high dimensional data, often with a relatively small number of samples, see [15] for a recent review. Prototypical examples include hypothesis testing, inference on, and estimation of a $p \times p$ population covariance matrix Σ, or of its leading eigenvectors, given a sample covariance matrix S computed from only n samples [4, 17]. Another example is screening for pairs of highly correlated variables, see [11]. In the high-dimension small-sample setting (known as the "large p, small n" scenario), considerable work has been devoted to various models of sparsity, to the development of thresholding schemes, and derivation of corresponding minimax rates of convergence, see for example [1, 5, 8, 18, 20, 25] and references therein. There are also works on estimation of a sparse inverse covariance matrix, though we shall not consider those in the present paper.

Assuming sparsity of either the population covariance matrix or of its leading eigenvectors, the two main thresholding schemes that have been proposed are: i) variable selection by thresholding the diagonal of the sample covariance matrix, and ii) screening highly dependent pairs of variables by thresholding the off-diagonal entries of sample covariance or correlation matrices. Similar schemes have been suggested for various hypothesis testing problems and for detection of sparse signals in noise.

On the theoretical front, for sparse covariance estimation, assuming that the underlying distribution satisfies appropriate tail conditions, several of the works cited above suggested thresholds of the form $C\sqrt{(\ln p)/n}$. Using relatively crude tail bounds, it can be proved that under various sparsity models such thresholds attain asymptotic consistency and in certain

AMS 2000 subject classifications: Primary 60F05; secondary 62E17

cases even achieve the minimax rates as $n, p \to \infty$. With respect to the largest off-diagonal entry of a sample correlation matrix, several works studied its limiting distribution as $p, n \to \infty$, assuming independence of all p coordinates, with an identity population covariance matrix, see [13, 6, 23, 35]. Beyond their theoretical interest, these results can be used to identify highly correlated pairs in high dimensional data and to reject hypotheses of variable independence. Indeed, such thresholding schemes have been proposed for detection of sparse signals in several practical applications. For example, Noh and Solo [28] proposed a thresholding scheme to detect signals in fMRI data, while Johnson and Potter [14] use a similar approach for outlier detection in a passive microwave sensing application. For a discussion of the importance of accurate thresholds in fMRI studies, see for example [24].

Given the above asymptotic results, an interesting practical question is their accuracy and relevance for typical applications, where the dimension and sample size are of course finite, with the latter parameter potentially rather small ($n = O(10 - 100)$, with $p = O(n)$ or even $p \gg n$). In this paper we focus on the hypothesis testing aspect of these problems: the determination of accurate thresholds for given false alarm rates, and in particular their dependence on the finite sample size and dimension, as well as on the underlying distributions. To set the corresponding thresholds for detection of sparse structures in data, we consider the null hypothesis that the observed data $X_i = (X_{i1}, \ldots, X_{ip})$ contains no structure at all, with all its p coordinates being independent and having the same distribution. Under this assumption, the diagonal entries of the sample covariance matrix S are all i.i.d. random variables. Similarly, the off-diagonal entries of the correlation matrix R are also all identically distributed, though they are weakly dependent. Hence, determination of appropriate thresholds for given type-I error probabilities amounts to the study of the maxima of many independent, or weakly dependent, random variables.

As is well known in extreme value theory, the convergence of the maxima to the limiting distribution may be very slow. In this paper we show that this is also the case for our two random variables of interest, the maxima of the diagonal of S and the largest pairwise correlation coefficient, albeit in a *non-trivial* way. The main difference between the classical theory of extreme values [19] and our setting, is that in the former the distribution of the p random variables is *fixed* and $p \to \infty$, whereas our setting involves *two* parameters p and n with the distribution of the underlying variables depending on the sample size n. To study the maxima of p such random variables, thus requires a careful analysis of various terms involving both of these parameters. In particular, we first point out that in the "large p – small n" setting, the standard approach of analyzing the relevant distributions in the joint limit as both $p, n \to \infty$ may give quite inaccurate results. The reason is that the leading asymptotic error terms, as both $p, n \to \infty$, typically of order $O(\ln \ln p / \ln p)$, are not the leading cause of error for finite and small values of n. The key point of our analysis, is that in the non-asymptotic "large p– small n" setting, the main source of inaccuracy of the limiting extreme value distributions is in *different* terms, of order $O((\ln p)^{3/2}/\sqrt{n})$ or $O((\ln p)^2/n)$, depending on the variable of interest. Since these terms are asymptotically negligible compared to $\ln \ln p / \ln p$, by studying the asymptotic limit $p, n \to \infty$ with dimension growing at most polynomially with sample size, these error terms are not considered, even though in practice they can be $O(1)$. By explicitly taking these higher order terms into account, we derive modified distributions and corresponding thresholds, which are far more accurate for practical sample size and dimensionality.

In our analysis, we consider both Gaussian and non-Gaussian distributions. In the Gaussian

case, we perform a delicate analysis of the known χ^2 distribution for the diagonal of the covariance matrix and of the distribution of Pearson's correlation coefficient for independent bi-variate Gaussians. In the course of this analysis we also derive sharp bounds for the left and right tail probabilities of a χ^2_n random variable, which may be of independent interest.

In the Gaussian case, we identify the correction terms as the leading order terms in the Edgeworth expansion of these distributions. Hence, in the non-Gaussian case, we study the corresponding Edgeworth expansions of the relevant distributions. The resulting modified thresholds thus depend explicitly on the higher order moments of the underlying distributions, and highlight the importance of Edgeworth expansions in high dimensional settings.

From a statistical perspective, our results allow determination of quite accurate non-asymptotic thresholds for a variety of hypothesis testing problems, as outlined above. In the context of high dimensional sparse linear regression [9], they allow to set appropriate thresholds for screening which predictor variables are highly correlated with a response variable.

However, our approach may have a broader applicability, as similar settings involving two parameters n, p with $p \gg 1$, occur frequently in many other high dimensional problems. In the context of tail inequalities for the maxima of p variables, it is known that behavior may change from double exponential to exponential, see for example [34]. In cases where accurate distributional results are needed, our perturbation technique, considering the leading order terms as a function of the finite values of both p and n may thus be applicable. A notable example is detection of significant bi-clusters or ANOVA-fit submatrices in high dimensional rectangular matrices [33]. Our analysis has also implications to estimation of sparse eigenvectors in principal component analysis [5], but these will be discussed in a separate publication.

The paper is organized as follows. In section 2 we present the problem formulation, a review of previous work and our main results. The outline of the proofs appears in sections 3 and 4, with more technical details postponed to the appendix. Section 5 contains simulations illustrating the accuracy of our modified distributions and thresholds.

2. Problem Setup and Main Results. For $i = 1, \ldots, n+1$, let $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^T$, be $(n+1)$ i.i.d. p-dimensional column vectors, with an unknown population covariance matrix Σ. The (unbiased) $p \times p$ sample covariance matrix S is given by

$$S = \frac{1}{n} \sum_{i=1}^{n+1} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{1}$$

where $\bar{\mathbf{x}}$ is the sample mean, whereas the (i, j)-th entry of the sample correlation matrix R is

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}}. \tag{2}$$

As mentioned in the introduction, in several modern applications, thresholding the matrices S or R are common tasks for covariance estimation, detection of sparse structures and for testing hypotheses of variable independence. The focus in this paper is on the determination of non-asymptotic *accurate thresholds* for these tasks for given false alarm probabilities, which depend both on the finite sample size n and dimension p, and on the underlying distribution.

Given the assumption that the structure to be discovered is sparse, in our analysis of the corresponding thresholds we consider the "null hypothesis" of no structure, i.e., we assume that all X_{ij} are i.i.d. from some underlying distribution $p(x)$ with a sufficient number of finite

moments. Without loss of generality, we assume that $\mathbb{E}[X_{ij}] = 0$, and $\mathbb{E}[X_{ij}^2] = 1$, so the population covariance matrix is the identity matrix, $\Sigma = I$. For future use, we denote the higher order moments by $\mu_k = \mathbb{E}[X_{ij}^k]$.

2.1. *Diagonal Thresholding.* Let us first consider the maxima of the diagonal entries of the sample covariance matrix, which we denote by Y_{np},

$$Y_{np} = \max_{1 \le i \le p} S_{ii}. \tag{3}$$

Under the null hypothesis of no structure in the data, the diagonal entries S_{ii} are all independent. Furthermore, as $n \to \infty$ and under suitable moment conditions on $p(x)$, from the CLT, each S_{ii} converges in distribution to a Gaussian random variable. Due to the independence of all S_{ii}, the theory of extreme values then states that as $p \to \infty$ the maxima, after proper centering and scaling, converges to a Gumbel distribution [19]. When the underlying data X_i are multivariate Gaussian $N(0,1)$, each S_{ii} follows a χ_n^2/n distribution. Approximating each diagonal entry by a $N(1, 2/n)$ random variable, standard results on the maxima of p independent Gaussians imply that the threshold $z(p,\alpha)$ which satisfies

$$\Pr\left[Y_{np} > 1 + \sqrt{\frac{2}{n}} z(p,\alpha)\right] = \alpha \tag{4}$$

is asymptotically given by

$$z^2(p,\alpha) = 2\ln p - \ln\ln p - \ln 4\pi - 2\ln|\ln(1-\alpha)| + O\left(\frac{\ln\ln p}{\ln p}\right). \tag{5}$$

As can be verified numerically, Eqs. (4) and (5) are quite accurate for the maxima of $p \gg 1$ i.i.d. Gaussian r.v.'s. The key difference in our setting is that the distribution of each of the p r.v. S_{ii} is only *approximately* Gaussian and depends on a second parameter n. Due to the slow convergence of the χ_n^2/n distribution to a $N(1, 2/n)$ Gaussian distribution, Eq. (5) may thus be a poor approximation to the required threshold. That is, the scaling $\sqrt{\frac{2}{n}} z(p,\alpha)$ in Eq. (4) may not be sufficiently accurate, with the required threshold having a non-negligible dependence on n, $z = z(n,p,\alpha)$. To illustrate this point, and motivate our work, consider the plots in Fig. 1. In the three panels from left to right, we compare the empirical density of Y_{np} to the Gaussian threshold of Eq. (5), for $(n,p) = (100,1000)$, $(1000,500)$ and $(1000,100)$, respectively. Note that in the left panel with $n \ll p$, the distribution of Y_{np} is very far from the limiting Gumbel distribution corresponding to maxima of purely Gaussian r.v.'s. Even in the other panels, where $n = 2p$ or $n = 10p$, the fit is not very accurate.

Our first result elucidates on the reason for this discrepancy. We show that for a given right tail probability α, the Gaussian approximation involves neglecting a higher order term $O(z^3/\sqrt{n})$ in the relevant equation for setting the threshold $z(n,p,\alpha)$. Since to leading order $z \sim (\ln p)^{1/2}$, asymptotically as $p, n \to \infty$ with $\frac{p}{n} = c$ for example, this term not only vanishes but is also significantly smaller than the next order correction term not present in Eq.(5), of $O(\ln\ln p/\ln p)$. However, for this term to be negligible for finite p, n, we need $z^{3/2}/\sqrt{n} \ll 1$. Fig. 2(a) shows the slow decay of this term as $n \to \infty$ with $p = c \cdot n$, for various values of c. Even if $n = 5p$ ($c = 0.2$), for this higher order term to be smaller than < 0.01, requires a sample size $n = O(10^6)$. This example illustrates that problems involving asymptotics with

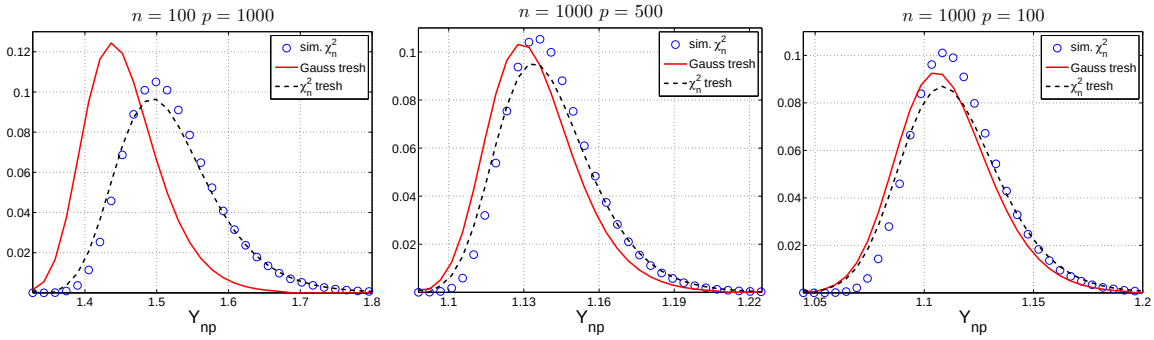


FIG 1. *Empirical density of Y_{np} (blue circles) compared to the asymptotic Gumbel distribution corresponding to the maxima of p Gaussians (solid red), and to our suggested correction, Eq. (6) (dashed black curve).*

two or more small or large parameters $(p, n$ in our case), need to be studied with extreme care. This is a well known issue in the applied mathematics literature, see [3].

Based on this observation, we suggest a modified threshold that takes this $O(z^3/\sqrt{n})$ term into account. In the Gaussian case this modified threshold can be computed explicitly using the known distribution of a χ^2_n random variable, as follows:

THEOREM 2.1. *Let $\{S_{ii}\}_{i=1}^p$ be p i.i.d. random variables $S_{ii} \sim \chi^2_n/n$, and let Y_{np} denote their maxima. Let $t = 1 + \sqrt{\frac{2}{n}} z_H$, where $z_H = z_H(n, p, \alpha)$ is given by*

$$z_H(n, p, \alpha) = z(p, \alpha)\left(1 + \frac{1}{3}\sqrt{\frac{2}{n}}\frac{z(p, \alpha)^3}{1 + z(p, \alpha)^2}\right) \tag{6}$$

and $z(p, \alpha)$ is the Gaussian threshold from Eq.(5). Then, for parameter values (n, p) such that $(\ln p)^{3/2} \ll \sqrt{n} \ll \frac{(\ln p)^{5/2}}{\ln \ln p}$,

$$\Pr\left[Y_{np} < t\right] = (1 - \alpha)\left(1 + O\left(\frac{\ln(1-\alpha)}{\ln p}\right) + O\left(\frac{|\ln(1-\alpha)|^2}{p}\right)\right). \tag{7}$$

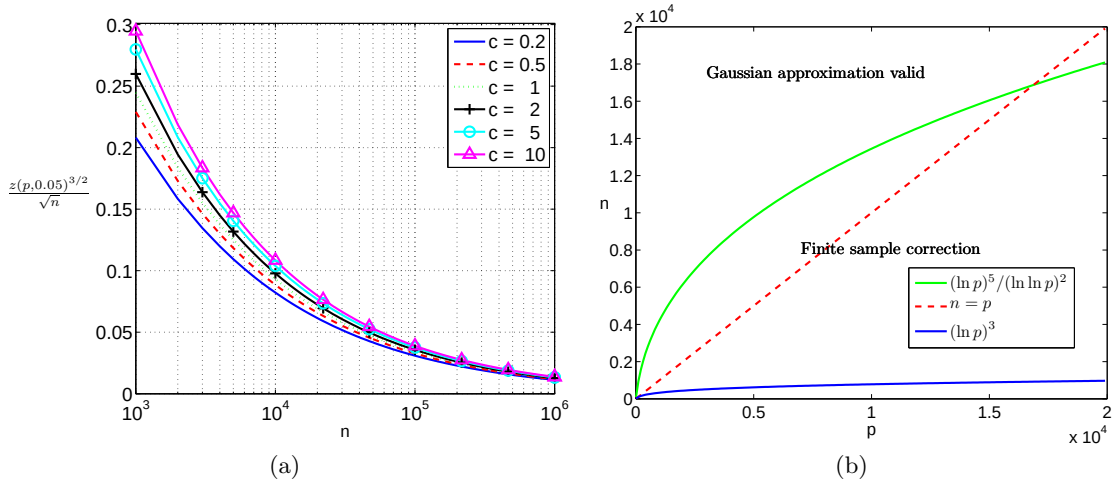
As shown in Figure 1, the modified expression of Eq.(6) yields a much better fit to the empirical density of Y_{np} for several values of (n, p). In particular, the fit is very accurate in the right tail, the most relevant region for calculation of the threshold z_H. The broad range of values of (n, p) where Eq. (6) is the leading correction term is illustrated in Figure 2(b).

To clarify the origin of the correction term in Eq. (6) we describe the first steps in the calculation of z_H. Since S_{ii} are i.i.d. we look for a threshold $t = t(\alpha)$ s.t.

$$\Pr[Y_{np} < t] = (1 - \Pr[S_{11} > t])^p = 1 - \alpha.$$

To this end, we should plug into this equation some expression for $\Pr[S_{11} > t]$. Since $S_{11} \sim \chi^2_n/n$ we use the following approximation (taken from lemma 3.1 below)

$$\Pr[S_{11} > 1 + \epsilon] \approx \frac{e^{-\frac{n}{2}(\epsilon - \ln(1+\epsilon))}}{\sqrt{\pi n}(\epsilon + \frac{2}{n})}.$$



(a) (b)

FIG 2. *(a) The ratio* $\frac{z(p,\alpha)^{3/2}}{\sqrt{n}}$ *as a function of sample size* n *with* $\alpha = 0.05$ *and* $p = c \cdot n$ *for various values of* c*. (b) Following the conditions of Theorem 2.1, the top and bottom solid curves are* $(\ln p)^5/(\ln\ln p)^2$ *and* $(\ln p)^3$*, respectively whereas the diagonal dashed curve is* $n = p$*. For sample size* n *significantly above the top curve (which in particular implies* $n \gg p$ *for* $p \le 20000$*), the Gaussian approximation (Eq. (5)) is reasonably accurate with all correction terms being negligible. Between the two curves finite* n *corrections are non-negligible. Below the bottom curve,* p *is exponential in* n*, and a different asymptotic approximation is needed.*

Replacing $\ln(1+\epsilon)$ by its Taylor expansion, and making a change of variables $\epsilon = \sqrt{\frac{2}{n}}z$ gives

$$\Pr\left[S_{11} > 1 + \sqrt{\tfrac{2}{n}}z\right] \approx \frac{e^{-z^2/2}}{\sqrt{2\pi}z} \cdot \frac{e^{-\frac{n}{2}\sum_{k=3}^{\infty}(-\sqrt{\frac{2}{n}}z)^k/k)}}{1+\sqrt{\frac{2}{n}}\frac{1}{z}}. \tag{8}$$

Let us compare this expression with the tail behavior for a Gaussian r.v., whereby for $z \gg 1$

$$\Pr\left[N(0,1) > z\right] = \frac{e^{-z^2/2}}{\sqrt{2\pi}z}\left(1 + O\left(\frac{1}{z}\right)\right).$$

Thus approximating $\frac{\chi^2_n}{n}$ by $N(1, \frac{2}{n})$, implicitly implies replacing the second term in Eq.(8) by unity. In particular for this approximation to be accurate, the next order term must be small $\sqrt{\frac{2}{n}}z(p,\alpha)^3 \ll 1$. However, as discussed above (and illustrated in Fig. 2(a)), this term is in fact $O(1)$ for practical finite values of (n,p) which explains the poor accuracy of Eq. (5) as a threshold for the maxima of many χ^2_n r.v.'s.

In summary, even though as $n \to \infty$ each diagonal entry S_{ii} converges to a Gaussian distribution, for finite values of n, p the Gaussian approximation may not be sufficiently accurate and a more careful analysis is required. When the observed data $X_{ij} \sim N(0,1)$, the leading correction term follows from an explicit analysis of the tail of a $\frac{\chi^2_n}{n}$ r.v., and in fact involves its Edgeworth expansion as $n \to \infty$. In the general case, there is no explicit expression for the distribution of the sample variance. In analogy to the χ^2_n/n case, we propose a modified threshold that takes into account the first term in the Edgeworth expansion of the sample variance. This gives the following proposition:

PROPOSITION 2.1. *Let S be the sample covariance matrix of an $(n+1) \times p$ matrix X, whose entries X_{ij} are all i.i.d. from some density $p(x)$ with zero mean, unit variance, and finite 8-th moments. Further assume that* $\limsup_{\|t\|\to\infty} \left|\mathbb{E}\left[e^{i(t_1 X_{11} + t_2 X_{11}^2)}\right]\right| < 1$. *Let*

$$z_E(n,p,\alpha) = z(p,\alpha)\left(1 + \frac{\kappa}{6\sigma^3}\frac{1}{\sqrt{n}}\frac{z(p,\alpha)^3}{1+z(p,\alpha)^2}\right) \tag{9}$$

where $z(p,\alpha)$ is the Gaussian threshold from Eq.(5), and

$$\sigma^2 = \mu_4 - 1 \tag{10}$$

$$\kappa = \mu_6 - 3\mu_4 + 6\mu_3^2 + 2. \tag{11}$$

Then, for parameter values (n,p) such that $(\ln p)^{3/2} \ll \sqrt{n} \ll \frac{(\ln p)^{5/2}}{\ln \ln p}$,

$$\Pr\left[Y_{np} < 1 + \frac{\sigma}{\sqrt{n}} z_E(n,p,\alpha)\right] \approx 1 - \alpha. \tag{12}$$

Note that for a Gaussian distribution, $\sigma^2 = 2, \kappa = 8$, and we recover Eq. (6) of Theorem 2.1. Figure 3 compares the empirical density of Y_{np} for several underlying distributions, to the limiting Gumbel density and to the density corresponding to Eq.(12). While our proposed threshold Eq.(9) is more accurate than the Gaussian threshold, its accuracy varies for different distributions. In contrast to the Gaussian case where Eq. (7) contained explicit error bounds, the errors involved in Eq. (12) are related to the accuracy of Edgeworth expansions. Deriving sharp (non-uniform and location dependent) bounds on the error of Edgeworth expansions is an interesting research topic beyond the scope of this article.

2.2. *Largest Correlation Coefficient.* Next, we consider the largest correlation coefficient, namely the maximal off-diagonal entry, in absolute value, of the sample correlation matrix,

$$L_n = \max_{i<j} |R_{ij}|. \tag{13}$$

The random variable L_n was suggested as a statistic for testing independence of p variates of a population, see [23, 26]. Related random variables, such as the maxima of individual rows of the correlation matrix R were recently suggested for screening interesting variables in large-scale correlation studies, see [11, 32]. Similarly, in the context of ultrahigh dimensional regression, screening variables based on their correlation with a response was studied by Fan and Lv [9]. Finally, the distribution of L_n plays a role in compressed sensing, since L_n is the coherence of the design matrix X, see [6].

The limiting distribution of L_n, in the joint limit $n, p \to \infty$ has been studied in several works. In [13], Jiang showed that if $n/p \to \gamma \in (0,\infty)$ and $E[X^r] < \infty$ for some $r > 30$, then

$$nL_n^2 - 4\ln p + \ln\ln p \to \exp\left(-e^{-y/2}/\sqrt{8\pi}\right). \tag{14}$$

Since then, several works showed that Eq.(14) continues to hold both with weaker moment conditions, as well as when the dimension is allowed to increase polynomially with sample

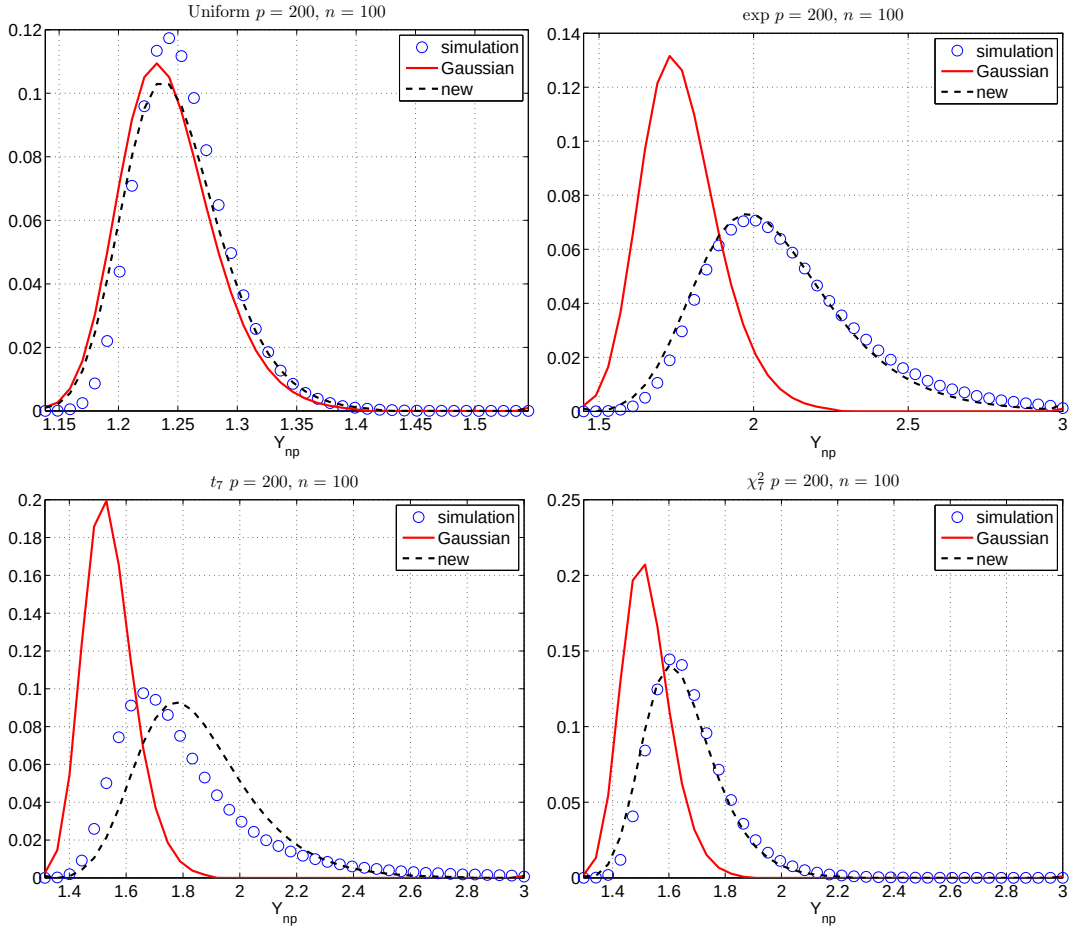


FIG 3. *Comparison of the empirical density of Y_{np} (blue circles) to the asymptotic Gumbel density of the maxima of p i.i.d. Gaussians, Eq. (5) (solid red), and to our suggested correction, Eq. (6) (dashed black curve).*

size, see [6, 22] and additional references therein. When dimension increases exponentially with sample size, there is a phase transition in the limiting distribution, see [7].

As in the case of the maxima of the diagonal entries of S, a key question is the accuracy of Eq.(14) for finite p, n, and in particular when $n \ll p$. Moreover, the parameter p has a different role here, as we now consider the maxima of $p(p-1)/2$ weakly dependent random variables, instead of only p variables as in the previous section. That is, even a modest value of p leads to the maxima of many random variables. In general, as already mentioned above, the convergence to limiting extreme value distributions is known to be slow. Indeed, in [23], the authors showed that the convergence rate in Eq. (14) is very slow, of $O(\ln\ln n/\ln n)$. Then, assuming that $p, n \to \infty$ with $c_1 n^\beta < p < c_2 n^\beta$ for some $\beta > 0$, and assuming some appropriate regularity conditions on the underlying distribution, Liu et. al. derived the following improved approximation (Thm. 1.2 in [23]), with a universal correction term independent of the underlying distribution,

$$\Pr\left[nL_n^2 - 4\ln p + \ln\ln p < \ y\right] \approx \exp\left(-\frac{p(p-1)}{2}\Pr\left[\chi_1^2 > \ 4\ln p - \ln\ln p + y\right]\right). \tag{15}$$

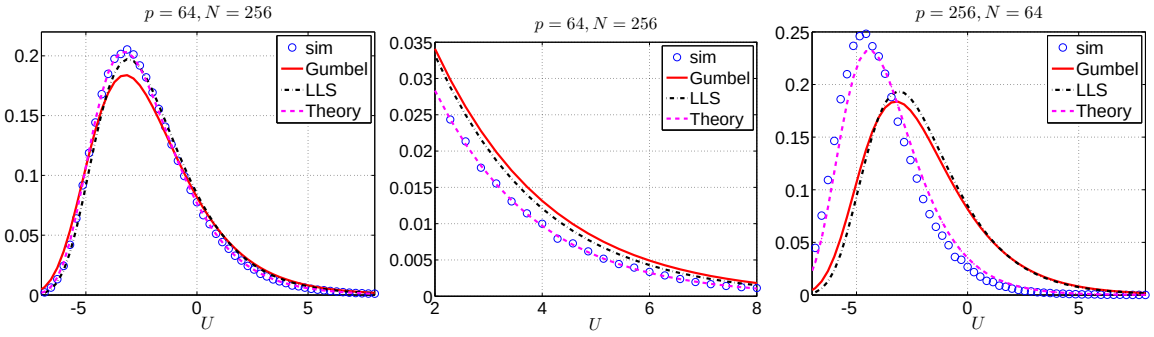


FIG 4. *Empirical density of* $U = nL_n^2 - 4\ln p + \ln\ln p$ *(blue circles) in comparison to the asymptotic Gumbel density, Eq. (14), (solid red curve), the correction by Liu, Lin and Shao (LLS) [23], Eq. (15), (dashed-dot black), and to our suggested correction, Eq. (17) (dashed purple curve). The left panel is for* $(p, n) = (64, 256)$*, the center panel is a zoom into the right tail region, whereas the right panel is for* $(p, n) = (256, 64)$*.*

Liu et. al. further showed that asymptotically, Eq. (15) has a smaller error, $O\left((\ln n)^{5/2}/\sqrt{n}\right)$.

In this paper we are interested in accurate approximations to the right tail probabilities of L_n. To motivate our work, consider the plots in Figure 4, which compare the empirical density of $nL_n^2 - 4\ln p + \ln\ln p$ to the limiting distribution (14) and to its correction (15) as suggested by [23], both for $(n, p) = (256, 64)$ as well as for $(n, p) = (64, 256)$, with underlying $N(0, 1)$ observations. As seen from these plots, neither (14) nor (15) provide accurate approximations to the required distributions, even though the latter is slightly better.

As we show below, and similar to the analysis of the random variable Y_{np}, the source for this non-negligible error is in a term of asymptotically smaller order $O((\ln p)^2/n)$, which was not considered in these previous works. In the Gaussian case, this term can be computed explicitly, using the known distribution of a single Pearson's correlation coefficient for independent bivariate Gaussian observations. Taking this term into account yields the following result:

THEOREM 2.2. *Let* R *be the correlation matrix of an* $(n+1) \times p$ *matrix* X*, whose entries* X_{ij} *are all i.i.d.* $N(0, 1)$*. Let*

$$w(y) = 2\ln(p(p-1)) - \ln\ln(p(p-1)) + \ln 2 + y. \tag{16}$$

Then, as $p, n \to \infty$*, with* $p/n \to c$

$$\Pr\left[L_n^2 < \frac{w(y)}{n-2}\right] = \exp\left(-\frac{e^{-y/2}}{\sqrt{8\pi}} \cdot A(w(y), n, p)\right)\left[1 + O\left(\frac{1}{n}, \frac{1}{\sqrt{n}(\ln p)^{3/2}}\right)e^{-y/2}\right] + O\left(e^{-y}\right) \tag{17}$$

where

$$A(w, n, p) = e^{-w^2/4(n-2)}\left(1 - \frac{n-2}{n}\frac{1}{w}\right)\sqrt{\frac{2\ln(p(p-1))}{w}}. \tag{18}$$

As illustrated in Fig. 4, Eq. (17) provides a much better fit to the empirical density of L_n than the asymptotic Gumbel distribution of Eq. (14), in particular at the right tail, which is the most relevant part for threshold calculation.

It is instructive to compare the difference between Eq. (17) and the limiting expression (14). Note that the latter follows from the former under the approximation $A(w,n,p) \approx 1$. For the limiting expression to be accurate, a necessary condition is thus that $w^2/4n \ll 1$. Since for large p, to leading order $w = 4\ln p(1+o(1))$, for this term to be negligible in practice, say $w^2/4n = 0.1$, the required condition is $n \geq 40(\ln p)^2$. Even at a moderately small dimension of $p = 10$, for the asymptotic distribution to be accurate requires $n \gtrsim 200$ samples. Our analysis thus illustrates that even with Gaussian observations, for practical values of (n,p) the limiting formula (14) may be quite far from the empirical one for the largest correlation coefficient, and as far as testing is concerned, may lead to rather inaccurate thresholds.

Let us provide a different point of view on the expression $A(w,n,p)$. As $n \to \infty$, each individual correlation coefficient R_{ij} converges in distribution to a Gaussian $N(0, 1/\sqrt{n})$ random variable. The term $\exp(-w^2/4(n-2))$ appearing in $A(w,n,p)$ is nothing but the leading order correction term in the Edgeworth expansion of the sample correlation coefficient, corresponding to independent bi-variate Gaussian observations. When the underlying random variables X_{ij} are non-Gaussian, an explicit expression for the density of the sample correlation coefficient is in general unknown. In analogy to the Gaussian case, we thus propose to approximate the probability $\Pr[|R_{ij}| > t]$ by its leading Edgeworth expansion.

PROPOSITION 2.2. *Let R be the correlation matrix of an $(n+1) \times p$ matrix X, whose entries X_{ij} are all i.i.d. from some density $p(x)$ with zero mean, unit variance, finite 8-th moments. Then,*

$$\Pr\left[nL_n^2 < w\right] \approx \exp\left(-\frac{p(p-1)}{2}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{w}}e^{-w/2}\left(1+\tfrac{1}{n}E(w)\right)\right) \tag{19}$$

where

$$E(w) = \frac{\alpha_3^2}{72}w(w^2-10w+15) + \frac{\alpha_4}{24}w(w-3) \tag{20}$$

and the coefficients α_3 and α_4 are given by

$$\alpha_3 = \mu_3^2 \tag{21}$$

$$\alpha_4 = -6 + (\mu_4 - 3)^2. \tag{22}$$

As discussed above, $E(w)/n$ is the leading term in the Edgeworth expansion of a single correlation coefficient. It depends on both the third and fourth moments of the underlying distribution. In fact, for a non-symmetric distribution with $\mu_3 \neq 0$, the correction involves not only a $O((\ln p)^2/n)$ term, but also a term $O((\ln p)^3/n)$, which asymptotically is even larger.

The top left panel of Fig. 5 shows the empirical density of L_n with Gaussian, uniform or χ_7^2 distributions, for $(n,p) = (128,128)$. Each of the remaining panels compares the right tail region to Eq. (19). At these parameter values, the Edgeworth expansion is quite accurate for the Gaussian and uniform distribution, but not for the χ_7^2 distribution which is asymmetric around its mean and has slower tail decay.

This example illustrates that the two-term Edgeworth expansion has limited applicability, in particular for $p \gg n$. The reason is that the Edgeworth expansion may not be accurate for values of the correlation coefficient several standard deviations away from zero, which are the relevant ones when $p \gg n$. Furthermore, the Edgeworth expansion may even give negative

densities for small sample sizes. While beyond the scope of this paper, one option to remedy this may be to apply some monotonic transformation that prevents negative densities without losing the asymptotic accuracy of the Edgeworth expansion, see for example [30].

Finally, the following proposition provides approximate expressions for the threshold of the largest correlation coefficient. Its proof, based on algebraic manipulations of Eqs. (17) and (19), is similar to that made in deriving the thresholds for the maximal diagonal entry of the sample variance, and is hence omitted.

PROPOSITION 2.3. *Let R be the correlation matrix of an $(n+1)\times p$ matrix X, whose entries X_{ij} are all i.i.d. from some density $p(x)$ with zero mean, unit variance, finite 8-th moments. Let*

$$w(n,p,\alpha) = w_0(n,p,\alpha)(1+\delta) \tag{23}$$

where

$$w_0(n,p,\alpha) = 2\ln(p(p-1)) - \ln\ln(p(p-1)) - \ln 4\pi - 2\ln|\ln(1-\alpha)| \tag{24}$$

and

$$\delta = \begin{cases} -\frac{w_0^2}{2(n-2)}\left(1 + w_0 + \frac{w_0^2}{n-2}\right)^{-1} & X_{ij}\sim N(0,1) \\ \frac{2E(w)}{n-1}\left(1 + w_0 - \frac{E(w)}{n-1}\right)^{-1} & \textit{otherwise} \end{cases}. \tag{25}$$

Then for parameter values (n,p) such that $(\ln p)^2 \ll n \ll \frac{(\ln p)^3}{\ln\ln p}$

$$\Pr\left[(n-2)L_n^2 < w(n,p,\alpha)\right] \approx 1-\alpha \tag{26}$$

For the case of Gaussian observations, similar to the analysis in Theorem 2.1, the error in Eq. (26) can be bounded explicitly. This error is $O(\ln(1-\alpha)\frac{\ln\ln p}{\ln p}) + O(\ln^2(1-\alpha))$.

3. Largest Diagonal Entry of the Sample Covariance Matrix. In this section we prove Theorem 2.1 and Proposition 2.1. Recall that under the null hypothesis of no structure the variables X_{ij} are assumed to be all i.i.d. Hence their sample variances S_{ii} are also i.i.d. and the exact equation for the threshold t as a function of the false alarm rate α is

$$1-\alpha = \Pr[Y_{np} < t] = (1 - \Pr[S_{11} > t])^p\,. \tag{27}$$

To simplify notation we denote $A_t = \Pr[S_{11} > t]$. Note that in fact A_t depends also on n since the distribution of S_{ii} depends on n. Taking logarithms on both sides of Eq. (27) gives

$$\ln(1-\alpha) = p\ln(1-A_t). \tag{28}$$

Since our interest is in right tail probabilities of the maxima Y_{np} where p is also large, we may assume that $A_t \ll 1$ and use the Taylor approximation $\ln(1-A_t) = A_t + O(A_t^2)$. Thus,

$$\ln(1-\alpha) = -p(A_t + O(A_t^2)) \approx -pA_t. \tag{29}$$

An approximate threshold $t = t(\alpha)$ can be found by inverting Eq.(29), namely $t(\alpha) = A^{-1}(\ln(1-\alpha)\frac{1}{p})$. The proof proceeds as follows: First we derive an explicit expression for A_t as function of n and t. Next, we plug this expression into Eq. (29) and solve for t, carefully analyzing the different error terms for finite p, n. We finish with an analysis of the error incurred by the approximations performed in the various steps of the derivation.

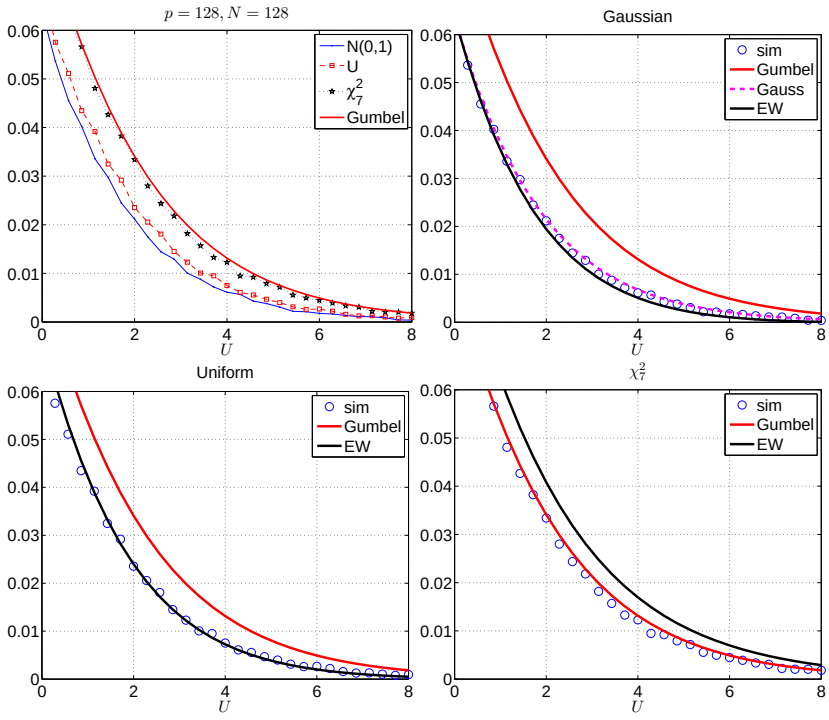


FIG 5. *Empirical density of* $U = nL_n^2 - 4\ln p + \ln\ln p$ *for various underlying distributions (Gaussian, uniform, and* χ^2_7*), and comparison to the corresponding Edgeworth expansion (EW).*

3.1. *Gaussian case.* In the multivariate Gaussian case, where $X_i = (X_{i1}, \ldots, X_{ip})^T \sim N(0, I_p)$, each S_{ii} follows a χ^2_n/n distribution. The following lemma gives non asymptotic bounds for the left and right tails of the χ^2_n distribution. This lemma may be of independent interest, as χ^2_n random variables appear in many statistical applications. In our case it provides an approximate yet accurate explicit expression for $A_t = \Pr[S_{11} > t]$.

LEMMA 3.1. *Let* W_n *be a random variable with a* χ^2_n *distribution. The following bounds hold for all* $n \geq 2$ *and for all* $\epsilon > 0$,

$$\Pr\left[\frac{W_n}{n} > 1+\epsilon\right] \leq \frac{1}{\sqrt{\pi n}}\frac{1}{\epsilon+\frac{2}{n}}\exp\left(-\tfrac{n}{2}(\epsilon - \ln(1+\epsilon))\right) \tag{30}$$

$$\Pr\left[\frac{W_n}{n} < 1-\epsilon\right] \leq \frac{1}{\sqrt{\pi n}}\frac{1-e^{-\frac{n}{2}\left(\epsilon-\frac{2}{n}\right)}}{\epsilon-\frac{2}{n}}\exp\left(\tfrac{n}{2}(\epsilon + \ln(1-\epsilon))\right) \tag{31}$$

In addition, if $\epsilon \geq \sqrt{\frac{2}{n}}z$ *for some* $z > 1$ *then the following lower bound also holds*

$$\Pr\left[\frac{W_n}{n} > 1+\epsilon\right] \geq \frac{1}{\sqrt{\pi n}}\frac{1}{\epsilon+\frac{2}{n}}\exp\left(-\tfrac{n}{2}(\epsilon - \ln(1+\epsilon))\right)\left(1 - \frac{1}{z^2} - \frac{1}{n}\right). \tag{32}$$

Remark 1: For $\epsilon < \frac{2}{n}$ it might seem that Eq. (31) gives a negative bound. However, this is not the case since the numerator and the denominator in the term $\frac{1-e^{-\frac{n}{2}\left(\epsilon-\frac{2}{n}\right)}}{\epsilon-\frac{2}{n}}$ have equal

sign for all ϵ. Additionally, there is no singularity at $\epsilon = \frac{2}{n}$ since as $\epsilon \to \frac{2}{n}$ both the numerator and the denominator vanish and the limit is well defined.

Remark 2: Several works derived χ^2_n bounds via various approximations to the exponential terms $\exp(-\frac{n}{2}(\epsilon - \ln(1+\epsilon))$ and $\exp\left(\frac{n}{2}(\epsilon + \ln(1-\epsilon))\right)$. For example, [17] and [21] proved that

$$\Pr\left[\frac{W_n}{n} > 1+\epsilon\right] \leq \exp\left(-\tfrac{n}{4}(\sqrt{1+2\epsilon}-1)^2\right) \tag{33}$$

$$\Pr\left[\frac{W_n}{n} < 1-\epsilon\right] \leq \exp\left(-\frac{n\epsilon^2}{4}\right). \tag{34}$$

Such bounds were mostly used to analyze minimax rates or prove consistency results. They are not sharp enough to determine accurate thresholds, which is the goal of this paper.

Remark 3: Lemma 3.1 slightly improves on the results in [16] as follows. Let $\epsilon = s\sqrt{2/n}$, and $\sigma_n = \sqrt{2n}$. For $0 < s < \sqrt{\frac{n}{8}}$ combining Eq. (47) of [16],

$$\ln(1+\epsilon) - \epsilon \leq \frac{-\epsilon^2/2}{1+2\epsilon/3} \qquad 0 \leq \epsilon \leq \frac{1}{2}$$

with our Eq. (30) gives that

$$\Pr\left[W_n - n > s\sigma_n\right] \leq \frac{1}{\sqrt{2\pi}} \frac{1}{s+\sqrt{\frac{2}{n}}} \exp\left(-\frac{s^2/2}{1+4s/3\sigma_n}\right) \tag{35}$$

which is slightly sharper than Eq. (43) in [16] as it has a smaller pre-exponential factor. Furthermore, from Eq.(35) it follows that Eq. (44) in [16], namely that for $0 < s < n^{1/6}$

$$\Pr\left[W_n - n > s\sigma_n\right] \leq \frac{1}{s} e^{-s^2/2} \tag{36}$$

holds for any $n \geq 2$, and not only for $n \geq 16$, as stated in [16].

We now return to our goal of deriving an explicit expression for the threshold t. We note that Eq.(30) and (32) imply that as $n \to \infty$ with $t = 1+\epsilon$, and $\epsilon = \sqrt{\frac{2}{n}}z$ with $z \geq z_0 > 1$,

$$A_t = \Pr\left[\frac{W_n}{n} > 1+\epsilon\right] = \frac{\exp\left(-\frac{n}{2}(\epsilon - \ln(1+\epsilon))\right)}{\sqrt{\pi n}(\epsilon + \frac{2}{n})}\left(1 + O\left(\frac{1}{z_0^2}, \frac{1}{n}\right)\right). \tag{37}$$

Plugging Eq.(37) into Eq.(29) gives

$$-p\frac{\exp\left(-\frac{n}{2}(\epsilon - \ln(1+\epsilon))\right)}{\sqrt{\pi n}(\epsilon + \frac{2}{n})}\left(1 + O\left(\frac{1}{z_0^2}, \frac{1}{n}\right)\right) = \ln(1-\alpha). \tag{38}$$

As described in section A.2 of the appendix, algebraic manipulations of Eq.(38) yield the following equation for $z = \sqrt{\frac{n}{2}}\epsilon$

$$z^2 - \frac{2}{3}\sqrt{\frac{2}{n}}z^3 + \ln z^2 + O\left(\frac{z^4}{n}, \frac{1}{\sqrt{n}z}, \frac{1}{z_0^2}, \frac{1}{n}\right) = 2\ln p - \ln 2\pi - 2\ln|\ln(1-\alpha)|. \tag{39}$$

Eq. (39) is an approximate transcendental equation for the required z. We look for the asymptotic solution for $z(\alpha, p, n)$, under the assumption that $n, p \gg 1$. It is common in extreme value theory to take only the first three terms in an asymptotic expansion. These are given by:

LEMMA 3.2. *As $p, n \to \infty$, with $(\ln p)^{3/2} \ll \sqrt{n}$, the first three terms in the asymptotic solution of Eq. (39) are*

$$z_\chi^2(\alpha, p) = 2\ln p - \ln\ln p - \ln 4\pi - 2\ln|\ln(1-\alpha)| + o(1) \tag{40}$$

Remark: Note that the condition $(\ln p)^{3/2} \ll \sqrt{n}$ holds for $p = O(n)$ as well as for $p = O(n^\beta)$ for any finite $\beta > 0$. However the condition does not hold if $p = O(e^{nc})$.

Lemma 3.2 shows that the first terms in the asymptotic expansion for the threshold z for the χ^2_n case are identical to those of Eq.(5), which corresponds to the Gaussian case. As shown in Figure 1 this might be quite inaccurate for finite values of p, n which hints that the $o(1)$ terms in Eq. (40) may be non-negligible. To elucidate the source of this inaccuracy we plug the value of z_χ back into the original equation (39) and obtain that the residual is:

$$\frac{2}{3}\sqrt{\frac{2}{n}}z_\chi^3 + O\left(\frac{1}{z_0^2}, \frac{1}{n}\right). \tag{41}$$

Note that for a fixed $0 < \alpha < 1$ (and bounded away from 1), as $p \to \infty$, $z_\chi \sim \sqrt{2\ln p}$ and $z_0 = \sqrt{\ln p}$. Therefore, the assumption $\sqrt{n} \ll \frac{(\ln p)^{5/2}}{\ln\ln p}$ implies that $\frac{z^3}{\sqrt{n}} = O\left(\frac{(\ln p)^{3/2}}{\sqrt{n}}\right) \gg O\left(\frac{1}{\ln p}, \frac{1}{n}\right)$. Hence the first term in Eq. (41) is the leading residual error. Moreover, it is significantly larger than the errors incurred in the derivation of Eq.(39). Additionally, while $\frac{(\ln p)^{3/2}}{\sqrt{n}}$ is negligible in the limit $n, p \to \infty$ its convergence to zero is very slow. As shown in Figure 2(a), for the condition $\frac{(\ln p)^{3/2}}{\sqrt{n}} < 0.1$ to hold requires $n \geq 10^5$ for a wide range of values of p/n. Hence we should not neglect this term for practical finite values of n, p.

To derive a more accurate threshold for finite values of n and p we return once more to Eq. (39). This time, we view the term $\frac{2}{3}\sqrt{\frac{2}{n}}z^3$ as a perturbation and look for a solution of the form $z = z_\chi(1+\gamma)$ where z_χ is the asymptotic solution from Eq. (40) and $\gamma = \gamma(n,p) = o(1)$. Plugging this expansion into Eq. (39) gives the following leading order equation for γ

$$2\gamma z_\chi^2 + 2\ln(1+\gamma) - \frac{2}{3}\sqrt{\frac{2}{n}}z_\chi^3 + \ln\left(1 - \frac{\ln\ln p + \ln 4\pi + 2\ln|\ln(1-\alpha)|}{2\ln p}\right) = 0.$$

Since $\gamma = o(1)$ we approximate $\ln(1+\gamma) = \gamma + O(\gamma^2)$ to get

$$2\gamma(1+z_\chi^2) = \frac{2}{3}\sqrt{\frac{2}{n}}z_\chi^3 - \ln\left(1 - \frac{\ln\ln p + \ln 4\pi + 2\ln|\ln(1-\alpha)|}{2\ln p}\right)$$

The first term in the r.h.s. is of order $O\left(\frac{(\ln p)^{3/2}}{\sqrt{n}}\right)$ while the second term is $O\left(\frac{\ln\ln p}{\ln p}\right)$. Hence, if $\sqrt{n} \ll \frac{\ln(p)^{5/2}}{\ln\ln p}$, the leading term in the r.h.s. is the first one and the leading solution for γ is

$$\gamma = \frac{1}{3}\sqrt{\frac{2}{n}}\frac{z_\chi^3}{1+z_\chi^2}. \tag{42}$$

This value of γ gives the threshold z_H of Eq. (6). If we plug z_H back into Eq.(39) we get that the leading error term is now $O\left(\frac{1}{z_0^2}\right) = O\left(\frac{1}{(\ln p)^2}\right)$ which also appears in the original equation.

To finish the proof of Theorem 2.1 we analyze the errors incurred by using the threshold z_H. This analysis is done by going "backward" in the derivation of z_H and collecting the error terms in the various steps. Since $\epsilon_H = \sqrt{\frac{2}{n}} z_H$ is the solution of Eq.(38) and $z_H > \sqrt{\ln p}$,

$$-p\frac{\exp\left(-\frac{n}{2}(\epsilon_H - \ln(1+\epsilon_H))\right)}{\sqrt{\pi n}(\epsilon_H + \frac{2}{n})}\left(1 + O\left(\frac{1}{\ln p}, \frac{1}{n}\right)\right) = \ln(1-\alpha). \tag{43}$$

Recall that by our assumption $\ln p \ll n$. Thus, the error term in the last equation is $O(\frac{1}{\ln p})$. By Eq.(37) we have that

$$A_{t_H}\left(1 + O\left(\frac{1}{\ln p}\right)\right) = \frac{\exp\left(-\frac{n}{2}(\epsilon_H - \ln(1+\epsilon_H))\right)}{\sqrt{\pi n}(\epsilon_H + \frac{2}{n})}. \tag{44}$$

Thus from the last two equations we have that

$$-pA_{t_H} = \ln(1-\alpha)\left(1 + O\left(\frac{1}{\ln p}\right)\right) \tag{45}$$

We use once more the Taylor approximation $p\ln(1 - A_{t_H}) = -pA_{t_H} + O(pA_{t_H}^2)$ to get

$$p\ln(1 - A_{t_H}) = \ln(1-\alpha)\left(1 + O\left(\frac{1}{\ln p}\right)\right) + O(pA_{t_H}^2). \tag{46}$$

Exponentiating this equation and approximating $e^x = 1 + x + O(x^2)$ gives

$$\Pr\left[Y_{np} < 1 + \sqrt{\frac{2}{n}} z_H\right] = (1 - A_{t_H})^p = (1-\alpha)\left(1 + O\left(\frac{\ln(1-\alpha)}{\ln p}\right) + O(pA_{t_H}^2)\right)$$

Using the approximation $pA \approx |\ln(1-\alpha)|$ (see Eq. (45)) gives Eq. (7).

□

3.2. *Non-Gaussian case.* To prove Proposition 2.1, which considers non-Gaussian observations, we first derive an explicit approximate expression for $\Pr[S_{ii} > t]$, that depends on the higher order moments of the underlying distribution $p(x)$ of the observations.

LEMMA 3.3. *Let $X_1, \ldots, X_{n+1}$ be $n+1$ i.i.d. scalar random variables with some density $p(x)$ that has zero mean, unit variance, and finite 8-th moments. Further assume that $\limsup_{\|t\|\to\infty} \left|\mathbb{E}\left[e^{i(t_1 X_{11} + t_2 X_{11}^2)}\right]\right| < 1$. Then, as $n \to \infty$, the Edgeworth expansion of the distribution of the unbiased sample variance estimator $S = \frac{1}{n}\sum(x_i - \bar{x})^2$ is*

$$\Pr\left[S > 1 + \frac{\sigma}{\sqrt{n}} z\right] = 1 - \Phi(z) + \frac{\phi(z)}{\sqrt{n}}\frac{\kappa}{6\sigma^3}(z^2 - 1) + O\left(\frac{1}{n}\right) \tag{47}$$

where $\Phi(z)$ is the c.d.f. of a $N(0,1)$ Gaussian r.v., $\phi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$ is its density, and

$$\sigma^2 = \mu_4 - 1 \tag{48}$$

$$\kappa = \mu_6 - 3\mu_4 + 6\mu_3^2 + 2. \tag{49}$$

Lemma 3.3 is proven in appendix A.4. Plugging Eq.(47) into Eq.(29) and approximating $\Phi(z) \approx 1 - \frac{\phi(z)}{z}$ yields the following approximate equation for the threshold $z(n,p,\alpha)$

$$-p\frac{e^{-z^2/2}}{\sqrt{2\pi}z}\left(1+\frac{\kappa}{6\sigma^3}\frac{z}{\sqrt{n}}(z^2-1)\right) = \ln(1-\alpha). \tag{50}$$

Taking logarithms and approximating $\ln(1+cz(z^2-1)) \approx cz^3$ gives the following equation:

$$z^2 - \frac{\kappa}{6\sigma^3}\frac{2}{\sqrt{n}}z^3 + \ln z^2 = 2\ln p - \ln 2\pi - 2\ln|\ln(1-\alpha)|. \tag{51}$$

The only difference between Eq. (51) and Eq.(39) for the Gaussian case, is in the coefficient of the $\frac{z^3}{\sqrt{n}}$ term. Solving Eq. (51) in the same way as we solved Eq. (39) gives Eq. (9).

4. Largest Correlation Coefficient. In this section we prove Theorem 2.2 and Proposition 2.2, regarding the distribution of the largest off-diagonal correlation coefficient. The first step is to derive an approximate relation between $\Pr[\max_{i<j}|R_{ij}| < t]$ and the much simpler event $\Pr[|R_{12}| < t]$. In contrast to the largest diagonal entry of S, where under the null hypothesis all p variables S_{ii} are independent, here the situation is a bit more complicated as the entries of R are weakly dependent.

One option to derive such a result is to employ the powerful Chen-Stein method, see [23]. Here, however, we show that for right tail probabilities, one may obtain similar results by simpler and more direct moment bounding methods. To this end, let $p_n(t)$ denote the density of a single correlation coefficient, computed from n samples, and let

$$A = A(n,t) = \Pr\left[|R_{ij}| > t\right] = \int_{-1}^{-t} p_n(r)dt + \int_{t}^{1} p_n(r)dt. \tag{52}$$

A key quantity that captures the dependence between some of these correlation coefficients is

$$A_2 = \Pr\left[|R_{ij}| > t \cap |R_{ik}| > t\right].$$

This quantity is related to a similar measure of dependence between correlation coefficients, recently analyzed in [11, 32], and also appears in the error bounds of the Chen-Stein method [23]. In terms of these quantities, we have the following claim.

Claim 4.1. *Let $s = p(p-1)/2$ be the total number of distinct correlation coefficients. Under the null hypothesis that all p variables are independent, the following inequalities between $\Pr[L_n < t]$ and $A(n,t)$ hold for any p, n, t,*

$$1 - sA \le \Pr[L_n < t] \le 1 - sA + \frac{1}{2}(sA)^2 - \frac{1}{2}sA^2 + s(p-2)(A_2 - A^2). \tag{53}$$

Note that for a Gaussian distribution, $A_2 = A^2$. While for general underlying distributions $A_2 \neq A^2$, as $n \to \infty$ any two pairs of correlation coefficients each converges to a Gaussian distribution and they become asymptotically independent. Hence, as $n \to \infty$ for any fixed t we have that $A_2 \to A^2$. Some bounds on a quantity similar to A_2 appear in [23]. A more detailed study of the rate of this convergence is beyond the scope of this paper.

Next, using the Taylor expansion $\exp(-sA) = 1 - sA + O\left((sA)^2\right)$, we may approximate Eq. (53) as

$$\Pr\left[\max_{i<j}|R_{ij}| < t\right] + O((sA)^2) = \exp(-sA) = \exp\left(-\frac{p(p-1)}{2}\Pr\left[|R_{ij}| > t\right]\right). \tag{54}$$

Eq. (54) is nothing but the Poisson approximation arising from the Chen-Stein method. Claim 4.1 gives somewhat different (typically larger but in some cases smaller) error bounds on the quality of this approximation, compared to those obtained by the Chen-Stein method. As we shall see below, the key to accurate thresholds is an accurate expression for $A = \Pr[|R_{ij}| > t$, rather than controlling the error in Eq. (54) above.

4.1. *Gaussian Case.* Given the analysis above, we now derive an approximate expression for $\Pr[|R_{ij}| > t]$. We first consider the Gaussian case, in which the distribution of a single Pearson's correlation coefficient is known explicitly (see [27], p. 147, Eq. 5)

$$p_n(r) = C_n(1-r^2)^{n/2-2} \tag{55}$$

with

$$C_n = \frac{\Gamma(\frac{n-1}{2})}{\pi^{1/2}\,\Gamma(\frac{n-2}{2})}.$$

Performing integration by parts

$$\begin{aligned}\Pr\left[|R_{ij}| > t\right] &= 2\int_t^1 p(r)dr = \frac{2C_n}{n-2}\int_t^1 \frac{(-(1-r^2)^{(n-2)/2})'}{r}dr \\ &= \frac{2C_n}{n-2}\frac{(1-t^2)^{(n-2)/2}}{t}\left[1 - \frac{1}{n}\frac{(1-t^2)}{t^2} + O\left(\frac{1}{n^2}\frac{1}{t^3}\right)\right] \end{aligned} \tag{56}$$

Next, for $t = \sqrt{w/(n-2)}$ we have that

$$(1-t^2)^{(n-2)/2} = e^{-w/2-w^2/4(n-2)}(1+o(1/n)). \tag{57}$$

Furthermore, from the asymptotics of the Gamma function,

$$C_n = \frac{1}{\sqrt{\pi}}\sqrt{\frac{n-2}{2}}\left(1+O(1/n)\right). \tag{58}$$

Combining Eqs. (56), (57) and (58) gives

$$\Pr\left[|R_{ij}| > \sqrt{\frac{w}{n-2}}\right] = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{w}}e^{-w/2-w^2/4(n-2)}\left(1 - \frac{n-2}{n}\frac{1}{w} + O\left(\frac{1}{n}\right) + O\left(\frac{1}{\sqrt{n}w^{3/2}}\right)\right). \tag{59}$$

Plugging Eq. (59) into the r.h.s. of Eq. (54) and choosing w as in Theorem 2.2 gives the first term in the r.h.s. of Eq.(17). The source of the additional error term in Eq. (17) is in the error term of Eq. (54). Plugging the definition of w into this error term gives:

$$(sA)^2 = \left(\frac{p(p-1)}{2}\right)^2\left(\frac{1}{p(p-1)}\frac{\sqrt{\ln(p(p-1))}}{\sqrt{2}}\frac{1}{\sqrt{w}}e^{-\frac{y}{2}} + o(1)\right)^2 = O\left(e^{-y}\right). \tag{60}$$

4.2. *Non-Gaussian case.* To prove Proposition 2.2, we first derive an approximate expression for $\Pr[|R_{12}| > t]$ when the $(n+1)\times p$ variables X_{ij} are i.i.d. but not necessarily Gaussian.

LEMMA 4.2. *Let X_1, X_2 be two i.i.d. random variables with the same distribution as a random variable X. Assume that $\mathbb{E}[X] = 0, \mathbb{E}[X^2] = 1$ and that X has finite 8-th moments. Then, as $n \to \infty$, the Edgeworth expansion of the sample correlation coefficient R_{12} from $n+1$ samples is*

(61)
$$\Pr\left[|R_{12}| > \frac{x}{\sqrt{n}}\right] = 2(1-\Phi(x)) + \frac{2\phi(x)}{n}\left(\frac{\alpha_3^2}{72}x(x^4-10x^2+15) + \frac{\alpha_4}{24}x(x^2-3)\right) + o(1/n)$$

where $\Phi(z)$ and $\phi(z)$ are the distribution and density functions of a $N(0,1)$ r.v., and

$$\alpha_3 = \mu_3^2 \tag{62}$$

$$\alpha_4 = -6 + (\mu_4 - 3)^2. \tag{63}$$

The proof of the lemma appears in appendix A.6. Combining the lemma with the relation (54) proves Proposition 2.2.

5. Simulation Results. We study the accuracy of our modified threshold for the sample variance in a series of simulations with several values of p, n and α. For Gaussian observations, we compare the accuracy of the threshold of the limiting Gumbel distribution, $t_G = 1 + \sqrt{\frac{2}{n}}z(p,\alpha)$, and our proposed threshold, $t_H = 1+\sqrt{\frac{2}{n}}z_H(n,p,\alpha)$, where $z(p,\alpha)$ and $z_H(n,p,\alpha)$ are given by (5) and (6), to empirical results based on 10^6 simulations. As Table 1 shows, the threshold t_H is much more accurate than the asymptotic Gumbel threshold t_G.

Similarly, for non-Gaussian distributions, we compare the accuracy of the asymptotic Gumbel threshold $t_G = 1 + \frac{\sigma}{\sqrt{n}}z(p,\alpha)$, with σ given by Eq. (10), to the proposed Edgeworth-based threshold $t_E = 1 + \frac{\sigma}{\sqrt{n}}z_E(n,p,\alpha)$, where z_E is given in Eq. (9). Table 2 shows that indeed the Edgeworth threshold is more accurate. However, its accuracy is not as good as that of the t_H threshold for the Gaussian case, due to higher order error terms in the Edgeworth expansion, that depend on the specific distribution.

	$n=100,\ p=1000$		$n=1000,\ p=500$		$n=1000,\ p=100$	
α	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_H]$	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_H]$	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_H]$
5%	27.9%	4.7%	8.5%	4.4%	6.5%	4.3%
2%	14.9%	1.8%	3.4%	1.7%	2.5%	1.6%
1%	9.0%	0.9%	1.6%	0.8%	1.1%	0.8%
0.1%	1.47%	0.09%	0.11%	0.07%	0.07%	0.07%

TABLE 1

For each value of α in the left column, we compare the accuracy of the Gaussian and χ^2_n high order thresholds.

Acknowledgments. We thank Alfred Hero, Bala Rajaratnam, Tiefeng Jiang, Tony Cai and Haruhiko Ogasawara for interesting discussions regarding various aspects of this work.

APPENDIX A: PROOFS

A.1. Bounds on χ^2_n Tail Probabilities. To prove lemma 3.1 we recall that the density of a χ^2_n random variable is

$$f_n(x) = C_n \cdot e^{h_n(x)}$$

	Uniform		Exponential	
α	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_E]$	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_E]$
5%	7.1%	4.8%	62.8%	8.7%
2%	3.0%	1.8%	50.2%	4.5%
1%	1.5%	0.9%	42.1%	2.8%
0.1%	0.18%	0.08%	23.23%	0.68%
	t_7		χ^2_7	
α	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_E]$	$\Pr[Y_{np} > t_G]$	$\Pr[Y_{np} > t_E]$
5%	60.4%	8.4%	42.5%	6.9%
2%	49.7%	5.6%	29.6%	3.2%
1%	43.0%	4.3%	22.4%	1.8%
0.1%	27.63%	2.09%	9.07%	0.32%

TABLE 2

For each value of α in the left column we compare the accuracy of the Gaussian and Edgeworth thresholds for different distributions with $n = 100$ and $p = 200$.

where $C_n = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}$ and $h_n(x) = -\frac{x}{2} + (\frac{n}{2} - 1)\ln x$. For simplicity from now on we omit the subscript n and write $h(x) = h_n(x)$.We wish to bound integrals of the form

$$C_n \int_a^b e^{h(x)} dx$$

where the endpoints (a, b) depend on whether our interest is in left tail or right tail probabilities. Note that $h'(x) = \frac{1}{2}(\frac{n-2}{x} - 1)$ and that for all $n \geq 2$

$$\frac{d^2h}{dx^2} = -(\tfrac{n}{2} - 1)\tfrac{1}{x^2} \leq 0.$$

Hence, for any $x_0 > 0$

$$h(x) \leq g(x) = h(x_0) + \left.\frac{dh}{dx}\right|_{x_0} (x - x_0).$$

To simplify our notation we denote $h_0 = h(x_0)$, $h'_0 = h'(x_0)$ and $h''_0 = h''(x_0)$. For any a, b

$$C_n \int_a^b e^{g(x)} dx = C_n e^{h_0} \int_a^b e^{h'_0(x-x_0)} dx = \frac{C_n e^{h_0}}{h'_0} e^{h'_0(x-x_0)}\Big|_a^b. \tag{64}$$

In our proof we will use the above equation with $x_0 = n(1 \pm \epsilon)$. Plugging this value into the different terms in the right hand side of (64) gives that

$$e^{h_0} = e^{-\frac{n(1\pm\epsilon)}{2} + \left(\frac{n}{2}-1\right)\ln(1\pm\epsilon) + \left(\frac{n}{2}-1\right)\ln n} = \frac{n^{\frac{n}{2}}}{n(1\pm\epsilon)} \cdot e^{-\frac{n}{2}(1\pm\epsilon - \ln(1\pm\epsilon))}$$

and

$$\frac{e^{h_0}}{h'_0} = \frac{n^{\frac{n}{2}}}{n(1\pm\epsilon)} \cdot \frac{e^{-\frac{n}{2}(1\pm\epsilon-\ln(1\pm\epsilon))}}{\frac{1}{2}\left(\frac{n-2}{n(1\pm\epsilon)} - 1\right)} = \frac{n^{\frac{n}{2}}}{\frac{n}{2}} \cdot \frac{e^{-\frac{n}{2}(1\pm\epsilon-\ln(1\pm\epsilon))}}{\left(\mp\epsilon - \frac{2}{n}\right)}.$$

Recall that for any $x \in \mathbb{R}$, $\ln\Gamma(x) \geq (x-\frac{1}{2})\ln x - x + \ln\sqrt{2\pi}$. Thus,

$$\begin{aligned}\frac{C_n e^{h_0}}{h_0'} &= \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\frac{n}{2}\Gamma(\frac{n}{2})}\cdot\frac{e^{-\frac{n}{2}(1\pm\epsilon-\ln(1\pm\epsilon))}}{\left(\mp\epsilon-\frac{2}{n}\right)} \\ &\leq \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}-1}}{\sqrt{2\pi}\left(\frac{n}{2}\right)^{\frac{n}{2}-\frac{1}{2}}e^{-\frac{n}{2}}}\cdot\frac{e^{-\frac{n}{2}(1\pm\epsilon-\ln(1\pm\epsilon))}}{\left(\mp\epsilon-\frac{2}{n}\right)} = \frac{e^{-\frac{n}{2}(\pm\epsilon-\ln(1\pm\epsilon))}}{\sqrt{\pi n}\left(\mp\epsilon-\frac{2}{n}\right)}. \qquad (65)\end{aligned}$$

Combining equations (64) and (65) gives that

$$C_n\int_a^b e^{g(x)}dx \leq \frac{e^{-\frac{n}{2}(\pm\epsilon-\ln(1\pm\epsilon))}}{\sqrt{\pi n}(\mp\epsilon-\frac{2}{n})}\cdot e^{h'(n(1\pm\epsilon))(x-n(1\pm\epsilon))}\Big|_a^b. \qquad (66)$$

With these preparations, Eq. (30) directly follows from Eq. (66) with $x_0 = n(1+\epsilon)$,

$$\Pr\left[\frac{W_n}{n} > 1+\epsilon\right] \leq C_n\int_{n(1+\epsilon)}^{\infty} e^{g(x)}dx \leq \frac{e^{-\frac{n}{2}(\epsilon-\ln(1+\epsilon))}}{\sqrt{\pi n}(\epsilon+\frac{2}{n})}. \qquad (67)$$

Similarly, Eq. (31) follows from Eq. (66) with $x_0 = n(1-\epsilon)$

$$\Pr\left[\frac{W_n}{n} < 1-\epsilon\right] \leq C_n\int_0^{n(1-\epsilon)} e^{g(x)}dx \leq \frac{e^{\frac{n}{2}(\epsilon+\ln(1-\epsilon))}}{\sqrt{\pi n}(\epsilon-\frac{2}{n})}\cdot\left(1-e^{1-\frac{n\epsilon}{2}}\right). \qquad (68)$$

The last step is to prove the lower bound (32). To this end, note that for $n > 2$ and $x > 0$

$$\frac{d^3h}{dx^3} = 2\left(\tfrac{n}{2}-1\right)\tfrac{1}{x^3} > 0.$$

Therefore for any $x_0 > 0$ and $x > x_0$

$$h(x) \geq \tilde{g}(x) = h_0 + h_0'(x-x_0) + h_0''\frac{(x-x_0)^2}{2}.$$

Next, using the inequality $e^{bx} \geq 1+bx$ and the identity $\int e^{ax}x^2dx = e^{ax}(a^2x^2-2ax+2)/a^3$ gives that

$$\begin{aligned}\Pr[W_n > x_0] \geq C_n\int_{x_0}^{\infty} e^{\tilde{g}(x)}dx &= C_n e^{h_0}\int_{x_0}^{\infty} e^{h_0'(x-x_0)+h_0''\frac{(x-x_0)^2}{2}}dx \\ &\geq C_n e^{h_0}\int_{x_0}^{\infty} e^{h_0'(x-x_0)}\left(1+h_0''\frac{(x-x_0)^2}{2}\right)dx \\ &= C_n e^{h_0}\left(\frac{1}{|h_0'|}-\frac{h_0''}{{h_0'}^3}\right) = \frac{C_n e^{h_0}}{|h_0'|}\left(1-\frac{|h_0''|}{{h_0'}^2}\right) \qquad (69)\end{aligned}$$

To conclude the proof we need an upper bound on $\Gamma(z)$, which appears in the denominator of C_n. The following auxiliary lemma, proven below, provides such a bound:

LEMMA A.1. *Let $\Gamma(z) = \int_o^{\infty} t^{z-1}e^{-t}dt$ be the Gamma function. Then for $z \in \mathbb{R}$,*

$$\Gamma(z) \leq \sqrt{2\pi}z^{z-1/2}e^{-z}\left(1+\frac{1}{2z}\right). \qquad (70)$$

Choosing $x_0 = n(1+\epsilon)$ and using the bound (70) on $\Gamma(x_0)$ we get

$$\begin{aligned}\frac{C_n e^{h(n(1+\epsilon))}}{|h'(n(1+\epsilon))|} &= \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\frac{n}{2}\Gamma\left(\frac{n}{2}\right)} \cdot \frac{e^{-\frac{n}{2}(1+\epsilon-\ln(1+\epsilon))}}{\left(\epsilon+\frac{2}{n}\right)} \geq \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}-1}}{\sqrt{2\pi}\left(\frac{n}{2}\right)^{\frac{n}{2}-\frac{1}{2}} e^{-\frac{n}{2}}} \cdot \frac{e^{-\frac{n}{2}(1+\epsilon-\ln(1+\epsilon))}}{\left(\epsilon+\frac{2}{n}\right)} \frac{1}{1+\frac{1}{n}} \\ &\geq \frac{e^{-\frac{n}{2}(\epsilon-\ln(1+\epsilon))}}{\sqrt{\pi n}\left(\epsilon+\frac{2}{n}\right)}\left(1-\frac{1}{n}\right). \qquad (71)\end{aligned}$$

Assuming that $\epsilon \geq \sqrt{\frac{2}{n}}z$ for some $z > 1$ we get that

$$\frac{|h_0''|}{{h_0'}^2} = \frac{\left(\frac{n}{2}-1\right)\frac{1}{n^2(1+\epsilon)^2}}{\frac{(2+n\epsilon)^2}{4n^2(1+\epsilon)^2}} = \frac{2n-4}{(2+n\epsilon)^2} \leq \frac{1}{z^2}.$$

Inserting this inequality into (69) proves Eq. (32). □

A.1.1. *Proof of Lemma A.1.* We start from the following upper bound on the Gamma function given by [2].

$$\Gamma(z) \leq \sqrt{2\pi}\left(\frac{z-\frac{1}{2}}{e}\right)^{z-\frac{1}{2}} = \sqrt{2\pi}z^{z-\frac{1}{2}}e^{-z}\left[\left(\frac{z-\frac{1}{2}}{z}\right)^{z-\frac{1}{2}}\sqrt{e}\right]$$

To prove the proposition we bound the right term in the last equation. Using the fact that $\ln(1-\epsilon) \leq -\epsilon$ and that $1/\sqrt{1-\frac{1}{2z}} \leq 1+\frac{1}{2z}$ for all $z > \frac{1}{2}$ gives that

$$\left(\frac{z-\frac{1}{2}}{z}\right)^{z-\frac{1}{2}} = \left(1-\frac{1}{2z}\right)^{z-\frac{1}{2}} = e^{z\ln\left(1-\frac{1}{2z}\right)} \cdot \frac{1}{\sqrt{1-\frac{1}{2z}}} \leq e^{-\frac{1}{2}} \cdot \left(1+\frac{1}{2z}\right)$$

Combining the above two equations proves (70).

A.2. Derivation of Eq. (39). Taking logarithms on both sides of Eq. (38) and making a change of variable $\epsilon = \sqrt{\frac{2}{n}}z$ yields

$$\ln p - \tfrac{n}{2}\left(\sqrt{\tfrac{2}{n}}z - \ln\left(1+\sqrt{\tfrac{2}{n}}z\right)\right) - \tfrac{1}{2}\ln 2\pi - \ln z - \ln\left(1+\sqrt{\tfrac{2}{n}}\tfrac{1}{z}\right) + \ln\left(1-O\left(\frac{1}{z_0^2},\frac{1}{n}\right)\right) = \ln|\ln(1-\alpha)|.$$

Replacing $\sqrt{\frac{2}{n}}z - \ln(1+\sqrt{\frac{2}{n}}z)$ with its Taylor expansion gives Eq. (39), up to a factor of -2,

$$-\frac{z^2}{2} + \sqrt{\frac{2}{n}}\frac{z^3}{3} - \ln z + O\left(\frac{z^4}{n}, \frac{1}{\sqrt{n}z}, \frac{1}{z_0^2}, \frac{1}{n}\right) = -\ln p + \frac{1}{2}\ln 2\pi + \ln|\ln(1-\alpha)|.$$

A.3. Proof of Lemma 3.2. The leading order term in the r.h.s of (39) is $2\ln p$, so the solution has the form of $z^2 = 2\ln p \cdot (1+\delta)$ where $\delta = o(1)$. Plugging this solution into the Eq. (39) and dividing by $\ln p$ gives

$$2\delta - \frac{8}{3}\sqrt{\frac{\ln p}{n}}(1+\delta)^{3/2} + \frac{\ln\ln p}{\ln p} + \frac{\ln(1+\delta)}{\ln p} = \frac{-\ln 4\pi - 2\ln|\ln(1-\alpha)|}{\ln p}.$$

The assumption of the lemma that $(\ln p)^{3/2} \ll \sqrt{n}$, implies that the leading order term now is $-\frac{\ln\ln p}{2\ln p}$. Thus $\delta = -\frac{\ln\ln p}{2\ln p} + \delta_2$ where $\delta_2 = o\left(\frac{\ln\ln p}{2\ln p}\right)$. The equation for δ_2 is

$$2\delta_2 - \frac{8}{3}\sqrt{\frac{\ln p}{n}}(1+\delta)^{3/2} + \frac{\ln(1+\delta)}{\ln p} = \frac{-\ln 4\pi - 2\ln|\ln(1-\alpha)|}{\ln p}.$$

This gives $\delta_2 = \frac{-\ln 4\pi - 2\ln|\ln(1-\alpha)|}{2\ln p}$ and the expression for z^2 is:

$$z^2 = 2\ln p - \ln\ln p - \ln 4\pi - 2\ln|\ln(1-\alpha)|$$

which proves the lemma. □

A.4. Edgeworth Expansion for the Sample Variance. Our proof of lemma 3.3 is based on the work of [12] on the Edgeworth expansion for the sample variance. Let Z denote some random variable with zero mean, unit variance, and finite 8-th moments. Further assume that $\limsup_{\|t\|\to\infty} \left|\mathbb{E}\left[e^{i(t_1 Z + t_2 Z^2)}\right]\right| < 1$. Let S denote the unbiased sample variance computed from $n+1$ i.i.d. samples of Z. Define the random variable

$$y = \sqrt{n}(S-1). \tag{72}$$

The asymptotic variance σ^2 and skewness κ of y are defined as

$$\mathbb{E}[(y - \mathbb{E}[y])^2] = \mathbb{E}[y^2] = \sigma^2 + O(n^{-1}) \tag{73}$$

$$\mathbb{E}[(y - \mathbb{E}[y])^3] = \mathbb{E}[y^3] = n^{-1/2}\kappa + O(n^{-3/2}). \tag{74}$$

From Eq.(3.3) of [12] we have that asymptotically in n:

$$\Pr\left[\frac{y}{\sigma} \le z\right] = \Phi(z) - \frac{\phi(z)}{\sqrt{n}}\frac{\kappa}{6\sigma^3}(z^2-1) + O\left(n^{-1}\right) \tag{75}$$

where $\Phi(x)$ and $\phi(x)$ are the distribution and density functions of a $N(0,1)$ r.v.. The Edgeworth expansion for a specific distribution follows from the following claim, proven below:

Claim A.2. *For any distribution with zero mean and unit variance it holds that*

$$\sigma^2 = \mu_4 - 1 \tag{76}$$

$$\kappa = \mu_6 - 3(\mu_4 + 2\mu_3^2) + 2 \tag{77}$$

Lemma 3.3 follows from Eq.(75) together with claim A.2. □

Remark: A similar derivation for the Edgeworth expansion of the biased sample variance can be found in [10].

A.4.1. *Proof of Claim A.2.* To compute the explicit value of σ we first calculate the expected value of S^2. It can be easily verified that

(78)
$$S^2 = \left(\frac{1}{n}\left(\sum_{i=1}^{n+1} x_i^2 - (n+1)\bar{x}^2\right)\right)^2 = \frac{1}{n^2}\left(\left(\sum_{i=1}^{n+1} x_i^2\right)^2 - 2(n+1)\bar{x}^2\sum_{i=1}^{n+1} x_i^2 + (n+1)^2\bar{x}^4\right)$$

The first term in Eq.(78) is $\left(\sum_i x_i^2\right)^2 = \sum_i x_i^4 + \sum_{i\neq j} x_i^2 x_j^2$ and its expected value is

$$\mathbb{E}\left[\left(\sum_{i=1}^{n+1} x_i^2\right)^2\right] = (n+1)\mathbb{E}\left[x^4\right] + (n+1)n\mathbb{E}\left[x^2\right]\mathbb{E}\left[x^2\right] = (n+1)(\mu_4 + n). \tag{79}$$

To compute the mean value of the second term in Eq.(78) we first calculate

$$\bar{x}^2 = \frac{1}{(n+1)^2}\left(\sum x_i\right)^2 = \frac{1}{(n+1)^2}\left(\sum_i x_i^2 + \sum_{i\neq j} x_i x_j\right) \tag{80}$$

Hence, the second term is equal to

$$\bar{x}^2\sum x_i^2 = \frac{1}{(n+1)^2}\left(\sum_i x_i^4 + \sum_{i\neq j} x_i^2 x_j^2 + 2\sum_{i\neq j} x_i^3 x_j + \sum_{i\neq j\neq k} x_i^2 x_j x_k\right) \tag{81}$$

and its expected value is

$$\mathbb{E}\left[\bar{x}^2\sum x_i^2\right] = \frac{(n+1)\mu_4 + (n+1)n}{(n+1)^2} = \frac{\mu_4 + n}{n+1}. \tag{82}$$

The following formula will be helpful for evaluation of the last term in (78)

$$\left(\sum_{i\neq j} x_i x_j\right)^2 = 2\sum_{i\neq j} x_i^2 x_j^2 + 4\sum_{i\neq j\neq k} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l} x_i x_j x_k x_l. \tag{83}$$

For the last term in (78), we have

(84)
$$\begin{aligned}(n+1)^4\bar{x}^4 &= \left(\sum_i x_i^2 + \sum_{i\neq j} x_i x_j\right)^2 = \sum_i x_i^4 + \sum_{i\neq j} x_i^2 x_j^2 + 2\left(2\sum_{i\neq j} x_i^3 x_j + \sum_{i\neq j\neq k} x_i^2 x_j x_k\right) + \left(\sum_{i\neq j} x_i x_j\right)^2 \\ &= \sum_i x_i^4 + 3\sum_{i\neq j} x_i^2 x_j^2 + 4\sum_{i\neq j} x_i^3 x_j + 6\sum_{i\neq j\neq k} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l} x_i x_j x_k x_l\end{aligned}$$

Hence its mean is

$$\mathbb{E}[\bar{x}^4] = \frac{1}{(n+1)^4}\left((n+1)\mu_4 + 3(n+1)n\right) = \frac{\mu_4 + 3n}{(n+1)^3}. \tag{85}$$

Combining (79) with (82) and (85) yields

$$\begin{aligned}\mathbb{E}\left[S^2\right] &= \frac{1}{n^2}\left((n+1)(\mu_4+n) - 2(\mu_4+n) + \frac{\mu_4+3n}{n+1}\right) \\ &= \frac{n-1+\frac{1}{n+1}}{n^2}\cdot\mu_4 + \frac{n-1+\frac{3}{n+1}}{n}\end{aligned} \tag{86}$$

We now compute the explicit expression for σ

(87)
$$\mathbb{E}[y^2] = (n+1)\left(E[S^2 - 2S + 1]\right) = (n+1)\left(\frac{n-1+\frac{1}{n+1}}{n^2}\cdot\mu_4 + \frac{n-1+\frac{3}{n+1}}{n} - 1\right) = \mu_4 - 1 + \frac{2}{n}$$

therefore, Eq.(48) follows. Next, to compute κ we also need the explicit expression for $\mathbb{E}[S^3]$.

(88)
$$S^3 = \frac{1}{n^3}\left(\left(\sum_i x_i^2\right)^3 - 3\left(\sum_i x_i^2\right)^2 (n+1)\bar{x}^2 + 3\left(\sum_i x_i^2\right)\left((n+1)\bar{x}^2\right)^2 - (n+1)^3\bar{x}^6\right)$$

We analyze each term in the r.h.s. separately. The mean of the first term is

(89)
$$\mathbb{E}\left[\left(\sum_i x_i^2\right)^3\right] = \mathbb{E}\left[\sum_i x_i^6 + 3\sum_{i\neq j} x_i^4 x_j^2 + \sum_{i\neq j\neq k} x_i^2 x_j^2 x_k^2.\right] = (n+1)\mu_6 + 3(n+1)n\mu_4 + (n+1)n(n-1).$$

The second term in Eq. (88) is

$$\begin{aligned}\left(\sum_i x_i^2\right)^2 \bar{x}^2 &= \frac{1}{(n+1)^2}\left(\sum_i x_i^4 + \sum_{i\neq j} x_i^2 x_j^2\right)\left(\sum_i x_i^2 + \sum_{i\neq j} x_i x_j\right) \\ &= \frac{1}{(n+1)^2}\left(\sum_i x_i^6 + \sum_{i\neq j} x_i^4 x_j^2 + \sum_i x_i^4 \sum_{j\neq k} x_j x_k + 2\sum_{i\neq j} x_i^4 x_j^2 + \sum_{i\neq j\neq k} x_i^2 x_j^2 x_k^2\right. \\ &+ \left. 2\sum_{i\neq j} x_i^3 x_j^3 + \sum_{i\neq j\neq k}(x_i^3 x_j^2 x_k + x_i^2 x_j^3 x_k) + \sum_{i\neq j\neq k\neq l} x_i^2 x_j^2 x_k x_l\right).\end{aligned}$$

Its expected value is

(90)
$$\mathbb{E}\left[\left(\sum_i x_i^2\right)^2 \bar{x}^2\right] = \frac{1}{(n+1)^2}\left((n+1)\mu_6 + 3(n+1)n\mu_4 + 2(n+1)n\mu_3^2 + (n+1)n(n-1)\right)$$

For the third term in Eq.(88) we use Eq.(84), and obtain that

$$\begin{aligned}\left(\sum_i x_i^2\right)\left(\bar{x}^2\right)^2 &= \frac{\sum_i x_i^2}{(n+1)^4}\left(\sum_i x_i^4 + 3\sum_{i\neq j} x_i^2 x_j^2 + 4\sum_{i\neq j} x_i^3 x_j + 6\sum_{i\neq j\neq k} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l} x_i x_j x_k x_l\right) \\ &= \frac{1}{(n+1)^4}\left(\sum_i x_i^6 + \sum_{i\neq j} x_i^2 x_j^4 + 6\sum_{i\neq j} x_i^4 x_j^2 + 3\sum_{i\neq j\neq k} x_i^2 x_j^2 x_k^2\right. \\ &+ \left. 4\sum_{i\neq j}(x_i^5 x_j + x_i^3 x_j^3) + 4\sum_{i\neq j\neq k} x_i^2 x_j^3 x_k + 6\sum_{i,j\neq k\neq l} x_i^2 x_j^2 x_k x_l + \sum_{i,j\neq k\neq l\neq m} x_i^2 x_j x_k x_l x_m\right)\end{aligned}$$

Its expected value is

(91)
$$\mathbb{E}\left[\left(\sum_i x_i^2\right)\left(\bar{x}^2\right)^2\right] = \frac{1}{(n+1)^4}\left((n+1)\mu_6 + 7(n+1)n\mu_4 + 4(n+1)n\mu_3^2 + 3(n+1)n(n-1)\right)$$

Finally, the last term in (88) involves $\bar{x}^6$.

(92)
$$\begin{aligned}\bar{x}^6 &= \tfrac{1}{(n+1)^6}\left(\textstyle\sum_i x_i^2 + \sum_{i\neq j} x_i x_j\right)^3 \\ &= \tfrac{1}{(n+1)^6}\left(\left(\textstyle\sum_i x_i^2\right)^3 + 3\left(\textstyle\sum_i x_i^2\right)^2\left(\sum_{i\neq j} x_i x_j\right) + 3\left(\textstyle\sum_i x_i^2\right)\left(\sum_{i\neq j} x_i x_j\right)^2 + \left(\sum_{i\neq j} x_i x_j\right)^3\right)\end{aligned}$$

The mean of the first term in Eq.(92) is given by Eq. (89). The second term is

(93)
$$\begin{aligned}\left(\textstyle\sum_i x_i^2\right)^2\left(\sum_{i\neq j} x_i x_j\right) &= \Big(\sum_i x_i^4 + \sum_{i\neq j} x_i^2 x_j^2\Big)\Big(\sum_{i\neq j} x_i x_j\Big) \\ &= 2\sum_{i\neq j} x_i^5 x_j + \sum_{i\neq j\neq k} x_i^4 x_j x_k + 2\sum_{i\neq j} x_i^3 x_j^3 + 4\sum_{i\neq j\neq k} x_i^3 x_j^2 x_k + \sum_{i\neq j\neq k\neq l} x_i^2 x_j^2 x_k x_l\end{aligned}$$

Its expected value is

$$\mathbb{E}\left[(\sum_i x_i^2)^2(\sum_{i\neq j} x_i x_j)\right] = 2(n+1)n\mu_3^2. \tag{94}$$

The third term in Eq.(92) is

$$\begin{aligned}\left(\sum_i x_i^2\right)\left(\sum_{i\neq j} x_i x_j\right)^2 &= \left(\sum_i x_i^2\right)\left(2\sum_{i\neq j} x_i^2 x_j^2 + 2\sum_{i\neq j\neq k} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l} x_i x_j x_k x_l\right) \\ &= 4\sum_{i\neq j} x_i^4 x_j^2 + 2\sum_{i\neq j\neq k} x_i^2 x_j^2 x_k^2 + \left(\sum_i x_i^2\right)\left(2\sum_{i\neq j\neq k} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l} x_i x_j x_k x_l\right)\end{aligned}$$

and its mean is equal to

$$\mathbb{E}\left[\left(\sum_i x_i^2\right)\left(\sum_{i\neq j} x_i x_j\right)^2\right] = 4(n+1)n\mu_4 + 2(n+1)n(n-1). \tag{95}$$

The last term in Eq.(92) is

(96)
$$\Big(\sum_{i\neq j} x_i x_j\Big)^3 = \sum_{i\neq j} x_i^3 x_j^3 + 3\sum_{i\neq j\neq k}(x_i^3 x_j^2 x_k + x_i^2 x_j^3 x_k) + 8\sum_{i\neq j\neq k} x_i^2 x_j^2 x_k^2 + 3\sum_{i\neq j\neq k\neq l} x_i^2 x_j^2 x_k x_l + \ldots$$

The dots in the last equation represent additional terms that are not of our interest since their expected value is zero. Hence,

$$\mathbb{E}\left[\Big(\sum_{i\neq j} x_i x_j\Big)^3\right] = (n+1)n\mu_3^2 + 8(n+1)n(n-1) \tag{97}$$

and the expected value of the last term in (92) is

$$\mathbb{E}\left[\bar{x}^6\right] = \frac{1}{(n+1)^6}\Big((n+1)\mu_6 + 15(n+1)n\mu_4 + 7(n+1)n\mu_3^2 + 15(n+1)n(n-1)\Big) \tag{98}$$

Finally we have that

$$\begin{aligned}(99)\qquad \mathbb{E}\left[S^3\right] &= \frac{n-2+\frac{3}{n+1}-\frac{1}{(n+1)^2}}{n^3}\cdot\mu_6+\frac{3(n+1)-9+\frac{15}{n+1}-\frac{15}{(n+1)^2}}{n^2}\cdot\mu_4\\ &+ \frac{-6+\frac{12}{n+1}-\frac{7}{(n+1)^2}}{n^2}\cdot\mu_3^2+\frac{(n-1)(n-2+\frac{6}{n+1}-\frac{15}{(n+1)^2})}{n^2}.\end{aligned}$$

Now we are ready to compute the explicit expression for κ:

$$\begin{aligned}\mathbb{E}[y^3] &= (n+1)\sqrt{n+1}\cdot\mathbb{E}\left[S^3-3S^2+3S-1\right]=(n+1)\sqrt{n+1}\cdot\mathbb{E}\left[S^3-3S^2+2\right]\\ &= \frac{(n+1)^2\sqrt{n+1}}{n^3}\mu_6-\frac{3(n+1)\sqrt{n+1}}{n^2}\mu_4-\frac{6(n+1)\sqrt{n+1}}{n^2}\cdot\mu_3^2+\frac{2(n+1)\sqrt{n+1}}{n^2}+O(n^{-\frac{3}{2}})\end{aligned}$$

Therefore Eq.(49) follows. □

A.5. Proof of claim 4.1. Our proof of Eq. (53), is similar to [26], which considered only Gaussian observations. We introduce the following $s=p(p-1)/2$ random variables,

$$(100)\qquad w_{ij}(t)=\begin{cases}1 & |R_{ij}|>t\\ 0 & \text{otherwise}\end{cases}$$

and define

$$(101)\qquad T=\sum_{i<j}w_{ij}.$$

Then by definition we have that

$$\Pr[L_n<t]=\Pr[T=0]=p_0.$$

The main idea is thus deriving bounds on the r.h.s. of the last equation using the first and second moments of T, which is equivalent to an inclusion-exclusion principle. Our starting point is the definition of the mean of T,

$$\mathbb{E}[T]=\sum_{j=0}^{\infty}j\Pr[T=j]=p_1+2p_2+\ldots$$

from which it immediately follows that

$$\mathbb{E}[T]\geq p_1+p_2+\ldots=1-p_0.$$

On the other hand, $\mathbb{E}[T]=s\mathbb{E}[w_{ij}]=sA$. Hence

$$(102)\qquad p_0\geq 1-sA.$$

Next, we wish to find an upper bound on $p_0=\Pr[L_n<t]$. To this end, we consider the random variable T^2. By definition, we have that

(103)

$$\mathbb{E}[T^2]=\mathbb{E}\left[\sum_{i<j}w_{i,j}\sum_{k<l}w_{k,l}\right]=\mathbb{E}\left[\sum_{i<j}w_{ij}^2\right]+\mathbb{E}\left[\sum_{\text{disjoint }i,j,k,l}w_{ij}w_{kl}\right]+\mathbb{E}\left[\sum_{i\neq j,k\neq(i,j)}w_{ij}w_{ik}\right]$$

The total number of disjoint indexes with $i \neq j \neq k \neq l$ is $s(p-2)(p-3)/2$. Note that under the null hypothesis, when the indexes i, j, k, l are all disjoint, the random variables w_{ij} and w_{kl} are independent. Hence, simple calculations show that

$$\begin{aligned} \mathbb{E}[T^2] &= sA + s\frac{(p-2)(p-3)}{2}A^2 + s\left(s - \frac{(p-2)(p-3)}{2} - 1\right)A_2 \\ &= sA + s(s-1)A^2 + 2s(p-2)(A_2 - A^2) \end{aligned} \tag{104}$$

where $A_2 = \mathbb{E}[w_{ij}w_{ik}] = \Pr\left[|R_{i,j}| > t \cap |R_{i,k}| > t\right]$. Next, note that there exist positive coefficients c_j such that

$$\frac{1}{2}\left[3\mathbb{E}[T] - \mathbb{E}[T^2]\right] = p_1 + p_2 - \sum_{j=3}^{\infty} c_j p_j \leq p_1 + p_2 + p_3 + \ldots = 1 - p_0$$

Inserting the expressions for $\mathbb{E}[T]$ and $\mathbb{E}[T^2]$ gives

$$p_0 \leq 1 - sA + \frac{1}{2}s^2A^2 - \frac{1}{2}sA^2 + s(p-2)(A_2 - A^2). \tag{105}$$

□

A.6. Edgeworth Expansion for the correlation coefficient. To prove lemma 4.2 we follow closely the notation and results of Ogasawara [29]. It is worth mentioning that in fact [29] considered the more general case where X_1 and X_2 may have different distributions, as well as a non-zero population correlation ρ_{12}. In our analysis, we consider the case where X_1 and X_2 are equally distributed and *independent*, which in particular implies that they are uncorrelated, $\rho_{12} = 0$. Furthermore, w.l.g. we may assume that X_1 has zero mean and unit variance, as these do not affect the sample correlation coefficient. This leads to a considerable simplification in the expression for the Edgeworth expansion of the correlation coefficient.

The starting point is Eq. 2.5 in [29], which gives the asymptotic expansion for the distribution of a single correlation coefficient, in terms of coefficients α_1,$\alpha_2, \alpha_3, \alpha_4$ and $\Delta\alpha_2$, which are all rather complicated expressions of the moments of the underlying distribution of X.

$$\begin{aligned} \Pr\left[\frac{R_{12}}{\sqrt{\alpha_2}} < \frac{t}{\sqrt{n}}\right] &= \Phi(t) - \frac{1}{\sqrt{n}}\left\{\alpha_1 + \frac{\alpha_3}{6}(t^2-1)\right\}\phi(t) \\ &\quad - \frac{t\phi(t)}{n}\left\{(\Delta\alpha_2 + \alpha_1^2)\frac{1}{2\alpha_2} + (\frac{\alpha_4}{24} + \frac{\alpha_1\alpha_3}{6})\frac{t^2-3}{\alpha_2^2} + \frac{\alpha_3^2}{72\alpha_2^2}(t^4 - 10t^2 + 15)\right\} \\ &\quad + o(1/n). \end{aligned} \tag{106}$$

First, note that since we are interested in $\Pr[|R_{12}| > t/\sqrt{n}\,]$, the leading correction term, of the form $f(t)/\sqrt{n}$, has no contribution, as $f(t)$ is an *even* function of t. Next, as discussed near Eq. (5.6) in [29], and following the results of Pitman [31], when X_1 and X_2 are independent, the bias and variance of R_{12} are asymptotically robust. Namely, the values of $\alpha_1, \alpha_2, \Delta\alpha_2$ are equal to those in the Gaussian case,

$$\alpha_2 = \alpha_{2,G} = 1, \quad \alpha_1 = \alpha_{1,G} = 0, \quad \Delta\alpha_2 = 0.$$

Therefore, Eq. (61) readily follows. To conclude the proof of the lemma, it remains to determine the values of the two coefficients α_3 and α_4.

Expressions for α_3, α_4 appear in [29], and depend on various quantities which we analyze below. As we consider a single correlation coefficient, we study the expressions in [29] with a dimension $p = 2$, so in the formulas below, all indices $a, b, c, \ldots$ take values in $\{1, 2\}$.

To study the values of α_3, α_4, we first introduce the following notation. Let $\Sigma = (\sigma_{ab})$ be the 2×2 population covariance matrix of the two random variables X_1, X_2, and denote by $\theta = \rho_{12} = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$ their population correlation coefficient. Further, denote by $\hat{\theta} = r_{12}$ their sample correlation coefficient. As in [29], we introduce the following additional notation: $\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{12}, \sigma_{22})$, with higher order moments defined as follows,

$$\sigma_{ab\ldots f} = \mathbb{E}[X_a X_b \cdots X_f].$$

We also denote by $\boldsymbol{\Omega}$ a 3×3 matrix with entries $\Omega_{ab,cd} = \sigma_{abcd} - \sigma_{ab}\sigma_{cd}$. The expressions for α_3 and α_4 also depend on cumulants of various orders. The first few are given by

$$\kappa_{ab} = \sigma_{ab}, \quad \kappa_{abc} = \sigma_{abc}, \quad \kappa_{abcd} = \sigma_{abcd} - \sigma_{ab}\sigma_{cd} - \sigma_{ac}\sigma_{bd} - \sigma_{ad}\sigma_{bc}.$$

where $a, b, c, d \in \{1, 2\}$. However, since $p = 2$, many of these cumulants vanish. For example, $\kappa_{1212} = \sigma_{1212} - \sigma_{11}\sigma_{22} = 0$. The only potentially non-zero 4th order cumulants are

$$\kappa_{1111} = \kappa_{2222} = \mathbb{E}[X^4] - 3. \tag{107}$$

Similarly, other than κ_{11111} and κ_{22222} which do not appear in our expressions, $\kappa_{abcde} = 0$ for all other values of $abcde$. The relevant 6th order cumulant is

$$\kappa_{112222} = \sigma_{112222} - \sigma_{11}\sigma_{2222} - 6\sigma_{22}\sigma_{1122} + 2 \cdot 3\sigma_{11}\sigma_{22}^2 = 0. \tag{108}$$

Similarly, the only relevant 8-th order cumulant also vanishes,

$$\begin{aligned} \kappa_{12121212} &= \sigma_{12121212} - 12\sigma_{11}\sigma_{112222} - \sigma_{1111}\sigma_{2222} - 18\sigma_{1122}\sigma_{1122} \\ &\quad +2(3\sigma_{1111}\sigma_{22}^2 + 3\sigma_{2222}\sigma_{11}^2 + 36\sigma_{1122}\sigma_{11}\sigma_{22}) - 6 \cdot 9\sigma_{11}^2\sigma_{22}^2 \\ &= \mathbb{E}[X^4]^2 - 12\mathbb{E}[X^4] - \mathbb{E}[X^4]^2 - 18 + 12\mathbb{E}[X^4] + 72 - 54 = 0. \end{aligned} \tag{109}$$

The formula below for α_4 also depends on M-functions defined as follows: For pairs of indices, we have $M(ab, cd) = \kappa_{abcd} + \kappa_{ac}\kappa_{bd} + \kappa_{ad}\kappa_{bc}$. Therefore,

$$\begin{aligned} M(11, 12) &= 0, \quad & M(11, 22) &= 0, \\ M(12, 12) &= 1, \quad & M(11, 11) &= \mathbb{E}[X^4] - 1. \end{aligned} \tag{110}$$

Eq. 3.13 in [29] contains the expression for $M(ab, cd, ef)$. Below are those relevant to us,

$$\begin{aligned} M(12, 12, 12) &= \sigma_{121212} = \mathbb{E}[X^3]^2 \\ M(11, 12, 12) &= \sigma_{111122} - \sigma_{1212}\sigma_{11} = \mathbb{E}[X^4] - 1. \end{aligned} \tag{111}$$

With the above auxiliary results at hand, we now consider the expression for α_3, Eq. 3.2 in [29]. This formula depends on $\boldsymbol{\Omega}$ and on first and second order derivatives $\partial\theta/\partial\sigma_{ab}$ and $\partial^2\theta/\partial\sigma_{ab}\sigma_{cd}$. However, when the two random variables X_1 and X_2 are independent and equally distributed,

many terms vanish and the resulting expression simplifies considerably. In particular, at a correlation coefficient $\rho_{12} = 0$, or equivalently $\sigma_{12} = 0$, we have that (see Eq. 4.1 in [29])

$$\left.\frac{\partial\theta}{\partial\boldsymbol{\sigma}}\right|_{\sigma_{12}=0} = (0,1,0)^T \tag{112}$$

$$\left.\frac{\partial^2\theta}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\right|_{\sigma_{12}=0} = \begin{pmatrix} 0 & -1/2 & 0 \\ -1/2 & 0 & -1/2 \\ 0 & -1/2 & 0 \end{pmatrix}. \tag{113}$$

Thus, inserting Eq. (112) and Eq. (113) into Eq. 3.2 in [29] simplifies to

$$\alpha_3 = \left(\frac{\partial\theta}{d\sigma_{12}}\right)^3 \mathbb{E}[X^3]^2 + 3\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial^2\theta}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\theta}{\partial\boldsymbol{\sigma}} \tag{114}$$

Note that

$$\boldsymbol{\Omega}\frac{\partial\theta}{\partial\boldsymbol{\sigma}} = (\Omega_{11,12}, \Omega_{11,22}, \Omega_{22,12})^T$$

However, when X_1 and X_2 are independent, all of these entries vanish, since

$$\begin{aligned} \Omega_{11,22} &= \mathbb{E}[X_1^2X_2^2] - \sigma_1^2\sigma_2^2 = 0 \\ \Omega_{11,12} &= \mathbb{E}[X_1^3X_2] - \mathbb{E}[X_1^2]\mathbb{E}[X_1X_2] = 0 \\ \Omega_{22,12} &= \mathbb{E}[X_2^3X_1] - \mathbb{E}[X_2^2]\mathbb{E}[X_1X_2] = 0. \end{aligned}$$

Therefore, the second term in Eq. (114) vanishes, and we obtain Eq. (62).

Next, we consider the formula for α_4, given by Eq. (3.11) in [29]. First, recall that for $a \neq b$, $\partial\theta/\partial\sigma_{ab} = 1$, whereas $\partial\theta/\partial\sigma_{aa} = \partial\theta/\partial\sigma_{bb} = 0$. Hence, in the outer summation for the first four lines in Eq. (3.1) only the single term with $a > b, c > d, e > f, g > h$ remains. That is,

$$\begin{aligned} \alpha_4 = {} & \kappa_{12121212} + \sum^{24}\kappa_{ac}\kappa_{bdefgh} + \sum^{32}\kappa_{ace}\kappa_{bdfgh} \\ & + \sum^{8}\kappa_{aceg}\kappa_{bdfh} + \sum^{24}\kappa_{abeg}\kappa_{cdfh} + \sum^{96}\kappa_{ac}\kappa_{be}\kappa_{dfgh} + \sum^{48}\kappa_{ac}\kappa_{eg}\kappa_{bdfh} \\ & + \sum^{96}\kappa_{ac}\kappa_{beg}\kappa_{dfh} + \sum^{48}\kappa_{bc}\kappa_{de}\kappa_{fg}\kappa_{ha} - \sum^{6}\kappa_{abcd}M(ef,gh) \\ & + \sum_{a\geq b}\sum_{c\geq d}\sum_{e\geq f}\sum_{g\geq h}\sum_{j\geq k} 2\frac{\partial\theta^2}{\partial\sigma_{ab}\partial\sigma_{cd}}\frac{\partial\theta}{\partial\sigma_{ef}}\frac{\partial\theta}{\partial\sigma_{gh}}\frac{\partial\theta}{\partial\sigma_{jk}}\sum^{10}M(ab,cd)M(ef,gh,jk) \\ & + \sum_{a\geq b}\sum_{c\geq d}\sum_{e\geq f}\sum_{g\geq h}\sum_{j>k,l>m}\left(\tfrac{3}{2}\frac{\partial^2\theta}{\partial\sigma_{ab}\sigma_{cd}}\frac{\partial\theta^2}{\partial\sigma_{ef}\partial\sigma_{gh}} + \tfrac{2}{3}\frac{\partial^3\theta}{\partial\sigma_{ab}\partial\sigma_{cd}\partial\sigma_{ef}}\frac{\partial\theta}{\partial\sigma_{gh}}\right)\times \\ & \frac{\partial\theta}{\partial\sigma_{jk}}\frac{\partial\theta}{\partial\sigma_{lm}}\sum^{15}M(ab,cd)M(ef,gh)M(jk,lm) \\ & - \left(4\alpha_1\alpha_3 + 6\alpha_2\Delta\alpha_2 + 6\alpha_2\alpha_1^2\right). \end{aligned} \tag{115}$$

We now separately analyze each of the terms in the equation above. First, according to Eq. (109), $\kappa_{12121212} = 0$. Next, consider the first sum, $\sum^{24}\kappa_{ac}\kappa_{bdefgh}$. It contains 24 terms, which account for the 8 choices for the index a multiplied by 6 choices for c and divided by 2 as

the order does not matter. Given a choice of an index a, for κ_{ac} not to vanish, we must have $c = a$. Suppose $a = 1$, then the multiplying factor is κ_{112222}, which according to Eq. (108) vanishes. Hence, all terms in the first sum are zero.

We proceed to the second sum, $\sum^{32} \kappa_{ace}\kappa_{bdfgh}$. For a term to be non-zero, we must have $ace = 111$ or $ace = 222$, but then the multiplying term is either κ_{12222} or κ_{21111} both of which vanish. Hence, the second sum also yields no contribution. In the third sum with 8 terms, we finally encounter a non-zero term, $\kappa_{1111}\kappa_{2222}$, hence this sum equals $(\mathbb{E}[X^4] - 3)^2$. Both the 4th and 5th sums vanish, as all their terms contain 4th order cumulants of the form κ_{12**}, which are all zero. In the 6th sum with 48 terms, all but six terms vanish, three terms of the form κ_{1111} and another three of the form κ_{2222}, so overall this sum contributes $6(\mathbb{E}[X^4] - 3)$.

The 7th sum vanishes, since the only potentially non-zero contribution is from terms of the form $\kappa_{111}\kappa_{222}$, but then the remaining factor is $\kappa_{12} = 0$. Next, consider the 8th sum, $\sum^{48} \kappa_{bc}\kappa_{de}\kappa_{fg}\kappa_{ha}$. For each of the 8 possible choices for b, only 3 choices for c give a non-zero κ_{11} or κ_{22}. Then, there are two valid choices for $e = d$. This has to be divided by 2^3, to account for the order in the first three pairs, so overall we have 6 terms each contributing a value of 1. The 9th sum vanishes as all its 6 terms contain the factor κ_{1212}.

Next we consider sum on the 4th row in Eq. (115). Here for the first order derivatives such as $\partial\theta/\partial\sigma_{ef}$ not to vanish, we must have $e \neq f, g \neq h, j \neq k$. For second derivatives to be non-zero, either $a = b, c \neq d$, or $a \neq b, c = d$, which gives an overall of 4 choices. For each such choice, out of the 10 arrangements for $M(ab, cd)M(ef, gh, jk)$, a non-zero contribution is obtained only when the equal pair is in the triplet $M(\cdot, \cdot, \cdot)$, which occurs in 6 out of the 10 terms, and gives $\mathbb{E}[X^4] - 1$. So the overall contribution of this sum is $2(-1/2) \cdot 6(\mathbb{E}[X^4] - 1)$, multiplied by the 4 choices for (ab, cd), that is $-24(\mathbb{E}[X^4] - 1)$.

Next, we open the parenthesis on the 5th row of Eq. (115), and analyze each term separately. The first one is

$$\frac{3}{2}\frac{\partial^2\theta}{\partial\sigma_{ab}\partial\sigma_{cd}}\frac{\partial^2\theta}{\partial\sigma_{ef}\partial\sigma_{gh}}\frac{\partial\theta}{\partial\sigma_{jk}}\frac{\partial\theta}{\partial\sigma_{lm}}\sum^{15} M(ab, cd)M(ef, gh)M(jk, lm)$$

Here we must have $j \neq k, l \neq m$. As in the previous case either $a = b, c \neq d$ or $a \neq b, c = d$, for an overall of 4 choices. Suppose $(ab, cd) = (11, 12)$. To obtain a nonzero contribution from the $M(\cdot, \cdot)$ terms, we must have either $(ef = 11)$ or $(gh = 11)$, so 2 choices. Furthermore, the equal pair must be in the same M-group as (ab). There are $3 = 15/5$ such configurations. So, the overall contribution from this sum is $3/2(-1/2)^2 \cdot 4 \cdot 2 \cdot 3(\mathbb{E}[X^4] - 1) = 9(\mathbb{E}[X^4] - 1)$.

Finally we consider the second term in this sum,

$$\frac{2}{3}\frac{\partial^3\theta}{\partial\sigma_{ab}\partial\sigma_{cd}\partial\sigma_{ef}}\frac{\partial\theta}{\partial\sigma_{gh}}\frac{\partial\theta}{\partial\sigma_{jk}}\frac{\partial\theta}{\partial\sigma_{lm}}\sum^{15} M(ab, cd)M(ef, gh)M(jk, lm)$$

At a correlation coefficient $\rho_{12} = 0$, the cases where the third derivative does not vanish are

$$\frac{\partial^3\theta}{\partial\sigma_{ab}\partial\sigma_{aa}^2} = 3/4, \quad \frac{\partial^3\theta}{\partial\sigma_{12}\partial\sigma_{11}\partial\sigma_{22}} = 1/4.$$

However, in the second case we will have only terms of the form $M(11, 12)$ or $M(11, 22)$ both of which vanish. We thus focus now on the first case. There are 3 choices for which pair, out of (ab, cd, ef) will have different indices, and 2 choices for the equal indices (11 or 22). Suppose

$ab = 11, cd = 11$. Then, there are $3 = 15/5$ cases where both belong to the same M-group. Overall, $2/3 \cdot 3 \cdot 2 \cdot 3(\mathbb{E}[X^4] - 1) \cdot 3/4$ or $9(\mathbb{E}[X^4] - 1)$.

To conclude, we use all of the above in Eq. (115) for α_4:

$$\begin{aligned} \alpha_4 &= (\mathbb{E}[X^4] - 3)^2 + 6(\mathbb{E}[X^4] - 3) + 6 - 24(E[X^4] - 1) + 9(\mathbb{E}[X^4] - 1) + 9(\mathbb{E}[X^4] - 1) \\ (116) &= -6 + (\mathbb{E}[X^4] - 3)^2. \end{aligned}$$

□

REFERENCES

[1] Amini, A. A. and Wainwright, M. J. (2009). High-dimensional analysis of semidefinite relaxations for sparse principal components. *Ann. Statist.* **37** 2877-2921.

[2] Batir, N. (2008). Inequalities for the Gamma function. *Arch. Math. (Basel)* **91** 554–563.

[3] Bender, C. M. and Orszag, S. A. (1999). *Advanced mathematical methods for scientists and engineers. I.* Springer-Verlag, New York.

[4] Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604.

[5] Birnbaum, A., Johnstone, I. M., Nadler, B. and Paul, D. (2012). Minimax bounds for sparse PCA for noisy high-dimensional data. *submitted.*

[6] Cai, T. and Jiang, T. (2012). Limiting Laws of Coherence of Random Matrices with Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices. *Ann. Statist.* **39** 1496-1525.

[7] Cai, T. and Jiang, T. (2012). Phase transition in limiting distributions of coherence of high dimensional random matrices. *Journal of Multivariate Analysis* **107** 24-39.

[8] Cai, T. T., Zhang, C. H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144.

[9] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society B.* 849-911.

[10] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion. Springer Series in Statistics.* Springer, New York.

[11] Hero, A. and Rajaratnam, B. (2011). Large Scale Correlation Screening. *Journal of the American Statistical Association* **106** 1540-1552.

[12] Ichikawa, M. and Konishi, S. (2002). Asymptotic expansions and bootstrap approximations in factor analysis. *J. Multivar. Anal.* **81** 47–66.

[13] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880.

[14] Johnson, J. T. and Potter, L. C. (2009). Performance Study of Algorithms for Detecting Pulsed Sinusoidal Interference in Microwave Radiometry. *IEEE Transactions On Geoscience and Remote Sensing* **47** 628 - 636.

[15] Johnstone, I. M. and Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Phil. Trans. R. Soc. A* **367** 4237–4253.

[16] Johnstone, I. M. (2001). Chi-square oracle inequalities. In *State of the art in probability and statistics (Leiden, 1999). IMS Lecture Notes Monogr. Ser.* **36** 399–418. Inst. Math. Statist., Beachwood, OH.

[17] Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. . MR2751448

[18] Karoui, N. E. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. . MR2485011

[19] Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions.* Imperial College Press, London. Theory and applications. . MR1892574

[20] Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254-4278.

[21] Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. . MR1805785

[22] Li, D., Liu, W.-D. and Rosalsky, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probability Theory and Related Fields* **148** 5-35.

[23] Liu, W. D., Lin, Z. and Shao, Q. M. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. . MR2474539

[24] Logan, B. R. and Rowe, D. B. (2004). An evaluation of thresholding techniques in fMRI analysis. *NeuroImage* **22** 95–108.

[25] Ma, Z. (2011). Sparse principal component analysis and iterative thresholding Technical Report, University of Pennsylvania.

[26] Moran, P. A. P. (1980). Testing the largest of a set of correlation coefficients. *Austral. J. Statist.* **22** 289–297. MR611045

[27] Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics. MR652932

[28] Noh, J. and Solo, V. (2008). Activation Detection in Functional MRI Based On Non-Separable Space-Time Noise Models. *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* 580-583.

[29] Ogasawara, H. (2006). Asymptotic expansion of the sample correlation coefficient under nonnormality. *Comput. Statist. Data Anal.* **50** 891-910. . MR2210737

[30] Ogasawara, H. (2012). Cornish - Fisher expansions using sample cumulants and monotonic transformations. *J. Mult. Anal.*

[31] Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test,. *J. Roy. Stat. Soc.* **4** 225-232.

[32] Sricharan, K., Hero, A. O. and Rajaratnam, B. (2011). A local dependence measure and its application to screening for high correlations in large data sets. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)* 1-8.

[33] Sun, S. and Nobel, A. B. (2011). On the maximal size of Large-Average and ANOVA-t Submatrices in a Gaussian Random Matrix. submitted.

[34] van de Geer, S. and Lederer, J. (2012). The Bernstein-Orlicz norm and deviation inequalities. *submitted.*

[35] Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345-5363.

School of Computer Science and Engineering
Hebrew University of Jerusalem
The Edmond J. Safra Campus
Jerusalem, 91904, Israel
E-mail: aharon.birnbaum@mail.huji.ac.il

Department of Computer Science
and Applied Mathematics
Weizmann Institute of Science
P.O.Box 26, Rehovot, 76100, Israel
E-mail: boaz.nadler@weizmann.ac.il