Multivariate CLT follows from strong Rayleigh property

Subhroshekhar Ghosh* Thomas M. Liggett Robin Pemantle†

29 October, 2016

Abstract

Let $(X_1, \ldots, X_d)$ be a random nonnegative integer vector. Many conditions are known to imply a central limit theorem for a sequence of such random vectors, for example, independence and convergence of the normalized covariances, or various combinatorial conditions allowing the application of Stein's method, couplings, etc. Here, we prove a central limit theorem directly from hypotheses on the probability generating function $f(z_1, \ldots, z_d)$. In particular, we show that the f being **real stable** (meaning no zeros with all coordinates in the open upper half plane) is enough to imply a CLT under a nondegeneracy condition on the variance. Known classes of distributions with real stable generating polynomials include spanning tree measures, conditioned Bernoullis and counts for determinantal point processes. Soshnikov [Sos02] showed that occupation counts of disjoint sets by a determinantal point process satisfy a multivariate CLT. Our results extend Soshnikov's to the class of real stable laws. The class of real stable laws is much larger than the class of determinantal laws, being defined by inequalities rather than identities. Along the way we investigate the related problem of *stable multiplication*.

1 Introduction

In the analysis of random combinatorial objects, one frequently encounters random variables taking values in a bounded set of nonnegative integers. The bound will depend on a size parameter, which will be taken to infinity, and limit laws will be sought for the behavior of the random variables. Perhaps the most common such scenario is when the variables obey a central limit theorem. Formally, if $\mathbb{P}_n$ is the probability law with size parameter n and $\mathbf{X}^{(n)} := (X_1^{(n)}, \ldots, X_d^{(n)})$ is a random integer vector with law $\mathbb{P}_n$, one would like to know conditions under which there are centering and scale constants for which $(\mathbf{X}^{(n)} - \mu^{(n)})/\sigma^{(n)}$ converges in distribution to a given multivariate normal.

One way to answer this is with the tools of classical probability theory, showing that the variables satisfy the hypotheses of a classical central limit theorem such as the Lindeberg-Feller, or a martingale CLT, or Stein's method, etc. Often the information available in a combinatorial problem is of a different nature. In particular, rather than the sort of analytic estimates needed in the hypotheses of classical probabilistic results, combinatorial applications often come with a generating function that is either exactly known or has certain verifiable properties. The question then arises of whether information from the generating function can be converted to a limit law.

In the univariate case, one very well known result is CLT behavior for real-rooted generating functions. Suppose $f(z) := \sum_{k=0}^n \mathbb{P}(X = k) z^k$ is the generating function for a random variable X. If f has only real roots then f factors as

$$f(z) = \prod_{j=1}^n (1 - p_j + p_j z)$$

where p_j are numbers in the unit interval. It follows that X is distributed as a sum of independent Bernoulli variables, which implies the self-normalized limit theorem

$$(1) \qquad \frac{X - \mathbb{E}X}{(\mathrm{Var}\, X)^{1/2}} \to N(0,1) \quad \text{as} \quad \mathrm{Var}\,(X) \to \infty\,.$$

In the multivariate case, a strand of research initiated in [Ben73, BR83, GR92] shows that a multivariate central limit theorem follows when the the multivariate generating function is a quasi-power (asymptotically $C_n f g^n$). More recently central limit behavior has been shown to follow when the generating function is rational and obeys a smoothness hypothesis [PW04, PW08], or in certain cases when the generating function is algebraic [Gre15].

A new development is the emergence of the class of *real stable* generating functions [COSW04, BB09a, BB09b, BBL09]. A real polynomial in d-variables is said to be stable if it has no zeros all of whose coordinates are strictly in the complex upper half-plane. In other words, f is stable if there are no solutions to $f(z_1, \ldots, z_d) = 0$ such that $\Im\{z_i\} > 0$ for all $i \in \{1, \ldots, d\}$. Often it is

*Research supported in part by ARO grant # W911NF-14-1-0094

†Research supported in part by NSF grant # DMS-1612674

possible to verify that a generating function is in this class without having an explicit description. Examples collected in [BBL09] include Tutte polynomials, spanning tree polynomials, and matrix polynomials such as $(\sum_{i=1}^d z_i A_i)$ for collections of positive definite matrices. In the case of spanning trees, consequences for the distribution of the number of edges in each of a finite collection of sets have been applied to TSP approximation [GSS11].

For univariate polynomials, stability reduces to real-rootedness, whence (1) is automatic for univariate stable polynomials. One important class of stable generating functions in several variables are the occupation counts of determinantal processes. Let $(X_1, \ldots, X_d)$ be the occupation counts of disjoint sets $B_1, \ldots, B_d$ for a determinantal point process; see [Sos00] for definitions. Soshnikov [Sos02, p. 174] proved a normal limit theorem for linear combinations $\sum_{j=1}^d \alpha_j X_j$, which is equivalent to a multivariate CLT. This generalized an earlier result for several specific determinantal kernels arising in random spectra [Sos00].

Determinantal measures are known to have stable generating functions. This might suggest that perhaps the CLT follows directly from stability of the generating function. Such a result would be important for the following reason. Determinantal measures are in some sense a very small set of measures. For example, determinantal measures supported on a set of cardinality d are parametrized by $d \times d$ Hermitian matrices, and therefore occupy a d^2-dimensional set in the $(2^d - 1)$-dimensional space of probability laws on $\{0,1\}^d$. The set of strong Rayleigh measures, by contrast, has full dimension, being constrained by inequalities rather than identities.

Our main results, Theorem 2.1 in the bivariate case and Theorem 2.1′ in the multivariate case, show this to be the case. The next section gives a statement of this and outlines the proof. The subsequent two sections discuss extensions and some theoretical questions about the class of real stable distributions which are raised by the arguments of the paper and partially answered. The last section discusses some applications to sampling, in which the generating functions are stable but not determinantal.

2 Main result

Our first result in this direction is a bivariate CLT valid when the variance grows faster than the 2/3 power of the maximum value. Because real stable variables are known to be negatively correlated, the covariances are denoted by negative quantities.

Theorem 2.1. *Let $\{(X_n, Y_n) : n \geq 1\}$ be a sequence of random integer pairs each of whose bivariate generating polynomials $f_n(x, y)$ is real stable and has degree at most M_n in each variable. Let*

$$A_n = \begin{bmatrix} \alpha_n & -\beta_n \\ -\beta_n & \gamma_n \end{bmatrix}$$

denote the covariance matrix of (X_n, Y_n). Suppose there is a sequence $s_n \to \infty$ and a fixed matrix $A = \begin{bmatrix} \alpha & -\beta \\ -\beta & \gamma \end{bmatrix}$ such that $s_n^{-2} A_n \to A$ and $s_n^{-1} M_n^{1/3} \to 0$. Then

$$\frac{(X_n, Y_n) - (\mathbb{E}X_n, \mathbb{E}Y_n)}{s_n} \to N(0, A) \tag{2}$$

in distribution as $n \to \infty$.

An outline of the proof is as follows. Let a and b be positive integers. From the definition of stability it may be shown that the generating polynomial for $aX_n + bY_n$ has no zeros near 1 (this is Lemma 6.1 below). A result of [LPRS16] then implies a Gaussian approximation for $aX_n + bY_n$ (Lemma 6.4 below). Tightness and continuity could be used to extend this to positive real (a, b), however the usual Cramér-Wold argument requires this for all real (a, b) regardless of sign. Instead, the argument is finished instead by invoking an improved Cramér-Wold result (Lemma 6.2 and Corollary 6.3). The complete proof is given in the Appendix.

A natural question is whether the nondegeneracy hypothesis on the variance is necessary. This condition is present in Lemma 6.4 from [LPRS16]. Already there, we do not know whether the condition is necessary.

3 Extensions

Higher dimensions The following extension to more than two variables requires only small generalizations of two of the lemmas.

Theorem (2.1′). *Let $\{\mathbf{X}^{(n)}\}$ be a sequence of random vectors of the same length, with real stable generating polynomials, degree at most M_n in each variable, and covariance matrices A_n. Suppose $s_n \to \infty$ with $s_n^{-2} A_n \to A$ and $s_n^{-1} M_n^{1/3} \to 0$. Then $(\mathbf{X} - \mathbb{E}\mathbf{X})/s_n \to N(0, A)$ in distribution as $n \to \infty$.* □

Singularity of A When A is singular, say $\mathbf{v}A = 0$ where $\mathbf{v} = (a, b)$, the conclusion of Theorem 2.1, namely a bivariate Gaussian limit, implies only that $(aX + bY)/s_n \to 0$, not that $aX + bY$ has a normal limit. This can be improved to the following result; intuitively, the only way to get degeneracy in the covariance matrix of a jointly real stable law is to condition a subset sum to be (nearly) constant.

Theorem 3.1. *In the notation of Theorem 2.1′, suppose A is singular and let $\mathcal{N}$ denote the nullspace of A. Let $\mathbf{1}_G$ denote the vector whose j component is 1 if $j \in G$ and 0 otherwise. The space $\mathcal{N}$ is spanned by a collection $\{1_G : G \in \mathcal{M}\}$ where $\mathcal{M}$ is a collection of disjoint sets. The quantities $Z_G^{(n)} := \mathbf{1}_G \cdot \mathbf{X}^{(n)}$ all have normal limits, provided the variances $\sigma_G^{(n)} := \mathrm{Var}\,(Z_G^{(n)})^{1/2}$ go to infinity; assuming this, $(\sigma_G^{(n)})^{-1}(Z_G^{(n)} - \mathbb{E}Z_G^{(n)}) \to N(0,1)$.*

Remark. This gives a CLT for a collection of linear functionals spanning the null space of A. More generally, one might want a CLT for every element of the null space. If the null space has dimension r then one may construct $\{Z_1, \ldots, Z_r\}$ as above. The vectors $\{\mathbf{Z}^{(n)}\}$ are real stable with covariance matrices A'_n for which $s_n^{-2}A'_n \to 0$. If it is possible to find s'_n for which $(s'_n)^{-2}A'_n \to A'$ then one obtains a finer multivariate CLT. The covariance matrices A'_n may or may not have a rescaled limit.

Quantitative version Suppose f_n is a sequence of bivariate real stable generating functions and that M_n/s_n^3 goes to zero, where M_n is the maximum degree of f_n in either variable and s_n^2 is the maximum variance of either variable. Let Q_n denote the probability law represented by f_n and let A_n denote the covariance matrix for this law. Suppose that Q_n, centered and divided by s_n, stays at least ϵ away from the bivariate Gaussian with mean zero and covariance $s_n^{-2}A_n$. Taking a subsequence $\{n_k\}$, there is a matrix A such that $s_n^{-2}A_n \to A$, contradicting Theorem 2.1. We conclude that there is a quantitative version of this result: namely a function g going to zero at zero such that

$$||Q - N(\mathbf{v}, \Sigma)|| < g(M^{1/3}/||\Sigma||^{1/2}) \tag{3}$$

whenever Q is a bivariate real stable law with mean $\mathbf{v}$, covariance Σ and maximum M. We do not known the best possible function g in (3).

Lemmas 6.1 and 6.4 are quantitative and sharp. Therefore, establishing (3) without giving up too much in the choice of function g would rest on a quantitative version of Corollary 6.3. Inverting the characteristic function is inherently quantitative, however the use of uniform continuity so as to use only values on a finite mesh is messy. Furthermore, while Lemma 6.1 is sharp, its use is certainly not: for example, if $f(z)$ generates a distribution within ϵ of normal, then so does $f(z^k)$, even though the nearest zero to 1 becomes nearer by a factor of k.

Non-uniformity of the estimates as the denominator of the rational slope increases is an annoying artifact of the proof and points to the need to replace Lemma 6.1 with something uniform over sets of directions. One possibility is to replace the exact combination $aX + bY$ with $a, b \in \mathbb{Z}^+$ by a probabilistic approximation. One somewhat crude approximation is to let $Z := \mathrm{Bin}\,(X, a) + \mathrm{Bin}\,(Y, b)$ be the sum of binomial distributions, conditionally independent given (X, Y). This has generating polynomial $g(z) = f(1 - a + az, 1 - b + bz)$ if $f(x, y)$ is the generating polynomial for (X, Y). When f is stable, so is g, thereby achieving uniformity in direction. Conditioned on (X, Y), the difference $Z - aX - bY$ is normal with variance $a(1-a)X + b(1-b)Y$, which has order M. The size parameter M cannot be less than a constant times s^2, where s^2 is the norm of the covariance matrix, but in the regime where $M = O(s^2)$, the added noise does not swamp the signal and near normality of Z implies near normality of the true $aX + bY$. This works equally well in any dimension.

To extend beyond the regime where M and s^2 are comparable, we would need to find a random variable Z with real stable law that approximates $aX + bY$ to within a smaller error than $M^{1/2}$. This motivates a one-dimensional version of this problem, which we now discuss.

4 Approximate multiplication

We use the term "stable multiplication by a" to denote an algorithm for constructing Z given X, where X and Z are positive integer random variables with stable generating polynomials and $|Z - aX| = O(1)$.

Proposition 4.1 (stable division by 2)**.** *Conditional on X, if X is even let $Z = X/2$, while if X is odd, flip a fair coin to decide whether $Z = \lfloor X/2 \rfloor$ or $Z = \lceil X/2 \rceil$. Then Z stably multiplies X by $1/2$.*

Proposition 4.2 (stable division by k)**.** *For any $k \geq 2$, $\lfloor X/k \rfloor$ stably multiplies X by $1/k$.*

The engine for proving both of these is the following result concerning interlacing roots. This result, proved in the Appendix, is of independent interest because of the power of interlacing results when trying to establish stability. Let NR be the collection of polynomials all of whose roots are simple and strictly negative. If f is a polynomial of degree n and $k \geq 1$, write

$$f(x) = \sum_{i=0}^{k-1} x^i g_i(x^k), \tag{4}$$

where g_i is a polynomial of degree $\lfloor \frac{n-i}{k} \rfloor$.

Theorem 4.3. *If $f \in \mathrm{NR}$ has degree n, the corresponding polynomials g_i are in NR as well. Furthermore, their roots are interlaced in the sense that if the collection of all $n - k + 1$ roots s_j of the g_i's are placed in* *increasing order,*

$$s_{n-k} < \cdots < s_4 < s_3 < s_2 < s_1 < s_0 < 0,$$

then the roots of g_i are $s_i, s_{i+k}, s_{i+2k}, \ldots$. $\square$

With this result in hand, the particular stability results are easily completed.

Proof of Proposition 4.1: The generating polynomial for Z is $\sum_k a_k z^k$ where $a_k = (1/2)\mathbb{P}(X = 2k+1) + \mathbb{P}(X = 2k) + (1/2)\mathbb{P}(X = 2k-1)$. Let $g(z) = (1/2)(1+z)^2 f(z)$ where f is the generating polynomial for X. Then $f \in \mathrm{NR}$ implies $g \in \mathrm{NR}$. Applying Theorem 4.3 with g in place of f, we have $g = g_0 + z g_1$ where $g_0, g_1 \in \mathrm{NR}$. The z^k coefficient of g_1 is the z^{2k+1} coefficient of g, which we see is equal to a_k. Thus Z has generating polynomial g_1, which is stable. $\square$

Proof of Proposition 4.2: The generating polynomial for $Z := \lfloor X/k \rfloor$ is

$$h(y) = \sum_{i=0}^{k-1} g_i(y) \tag{5}$$

where $g_0, \ldots, g_k$ are defined from the generating polynomial f for X by (4).

From the proof of Theorem 4.3, we see that $(-1)^m h(s_{mk}) > 0$ for each $0 \le m \le \frac{n-k}{k}$ (since the smallest root is s_{n-k}). Therefore, h has a root in each of the intervals of the form $(s_{(m+1)k}, s_{mk})$ for each $0 \le m \le \frac{n-2k}{k}$. This shows that h at least $\lfloor \frac{n}{k} \rfloor - 1$ negative roots. The degree of h is the largest of the degrees of the g_i's, which is the degree of g_0, i.e. $\lfloor \frac{n}{k} \rfloor$. To capture the final negative root, we observe that

$$(-1)^{\lfloor \frac{n}{k} \rfloor} h(s_{(\lfloor \frac{n}{k} \rfloor - 1)k}) < 0$$

and

$$(-1)^{\lfloor \frac{n}{k} \rfloor} h(s) > 0 \text{ for large negative } s.$$

$\square$

We do not know the extent to which multiplication by a can be accomplished when $a \in (0,1)$ is not a unit fraction. The same construction does not work. For example, if X has pgf

$$\frac{1}{20}(x+1)^2(x+4),$$

then the pgf of $Y = \lfloor \frac{2}{3} X \rfloor$ is $\frac{1}{20}(y^2 + 6y + 13)$, which has roots $-3 \pm 2\imath$. Thus an approach analogous to the one for unit fractions, does not work for non-unit fractions, the simplest unknown case being $a = 2/3$.

5 Discussion

Suppose one wishes to sample a random subset $S \subseteq [n]$ with prescribed marginals $p_1, \ldots, p_n$, that is, p_j is the desired value of $\mathbb{P}(j \in S)$. Among desirable properties of the law of S are as follows. (1) If $\sum_{j=1}^n p_j = k$, then one might ask for S always to contain precisely k elements; more generally, one might constrain the sample to k elements and replace each p_j by $k p_j / \sum_{i=1}^n p_i$. (2) For each pair $i \neq j \in [n]$, $\mathbb{P}(i, j \in S) \le p_i p_j$. This second property is desirable if the sample is used to estimate averages of more complicated functionals, where the pairwise negative correlation of elements of the sample will lead to a concentration inequality for the functional being estimated. Sampling schemes with prescribed marginals are known as πps sampling. Dozens of such schemes are known [BH83]; a number are shown in [BJ12] to have stable generating functions and are therefore pairwise negatively correlated. By Theorem 2.1$'$, we may conclude a central limit theorem. Two of the most common and useful examples or schemes are as follows.

Example 5.1 (Conditioned Bernoulli sampling)**.** *Given nonnegative $p_1, \ldots, p_n$ summing to k, it is shown how to find $\gamma_1, \ldots, \gamma_n$ such that the conditional law $\mathbb{P}$ of independent Bernoullis $\{X_1, \ldots, X_n\}$ with means $\gamma_1, \ldots, \gamma_n$, given $\sum_{j=1}^n X_j = k$, will have marginals $p_1, \ldots, p_n$. For instance one such choice for $\gamma_1, \ldots, \gamma_n$ may be characterized as a maximum entropy measure with the given marginals [Che00]. Conditioned Bernoullis are known to have a stable generating function (see, e.g., [BBL09] or [Pem12, Example 6.2]). It follows that if $B_1, \ldots, B_d$ are disjoint subsets of $[n]$, then the joint law of $(Y_1, \ldots, Y_d)$ has stable generating function where Y_j is the number of elements of B_j present in the sample. A multivariate CLT for these counts follows from Theorem 2.1$'$.*

Example 5.2 (pivot sampling)**.** *One of the quickest schemes for sampling a subset with given marginals is* **pivot sampling**. *This is a linear time algorithm in which elements are chosen or rejected sequentially, requiring the update of only one marginal probability at each step. The corresponding multivariate generating function is shown to be stable in [BJ12]. Again, a central limit theorem follows for the joint counts of disjoint subsets.*

We close with some open questions concerning sharpening the variance requirements in Theorem 2.1 as well as stable multiplication.

Question 1. *Can the hypothesis $s_n^{-1} M_n^{1/3} \to 0$ in Theorem 2.1 be weakened, preferably to $s_n \to \infty$?*

Question 2. *What is the best possible function g in* (3)*?*

Question 3. *Is there an $O(1)$ stable multiplication by $2/3$?*

A solution to the following more general stable multiplication question would improve the hypotheses for the CLT by lowering the variance requirement below $M^{2/3}$.

Question 4. *Let $\mathbf{X}$ have real stable probability generating polynomial with maximum value M and let $\mathbf{a}$ be a positive rational vector. Is there a stable $o(M^{1/3})$ approximation to $\mathbf{a} \cdot \mathbf{x}$?*

References

[BB09a] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur programs, I: Linear operators preserving stability. *Invent. Math.*, 177:541–569, 2009.

[BB09b] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur programs, II: Theory of stable polynomials and applications. *Comm. Pure Appl. Math.*, 62:1595–1631, 2009.

[BBL09] J. Borcea, P. Brändén, and T. Liggett. Negative dependence and the geometry of polynomials. *J. AMS*, 22:521–567, 2009.

[Ben73] E. Bender. Central and local limit theorems applied to asymptotic enumeration. *J. Combinatorial Theory Ser. A*, 15:91–111, 1973.

[BH83] K.R.W. Brewer and M. Hanif. *Sampling with unequal probabilities.* Springer, New York-Berlin, 1983.

[BJ12] P. Brändén and J. Jonasson. Negative dependence in sampling. *Scand. J. Stat.*, 39:830–838, 2012.

[BMR97] C. Bélisle, J.-C. Massé, and T. Ransford. When is a probability measure determined by infinitely many projections? *Ann. Probab.*, 25:767–786, 1997.

[BR83] E. Bender and L. B. Richmond. Central and local limit theorems applied to asymptotic enumeration, II: multivariate generating functions. *J. Combinatorial Theory Ser. A*, 34:255–265, 1983.

[Che00] S. X. Chen. General properties and estimation of conditional Bernoulli models. *J. Multivariate Anal.*, 74:69–87, 2000.

[COSW04] Y.-B. Choe, J. Oxley, A. Sokal, and D. Wagner. Homogeneous multivariate polynomials with the half-plane property. *Adv. Appl. Math.*, 32:88–187, 2004.

[GR92] Z. Gao and L. B. Richmond. Central and local limit theorems applied to asymptotic enumeration. IV. Multivariate generating functions. *J. Comput. Appl. Math.*, 41(1-2):177–186, 1992. Asymptotic methods in analysis and combinatorics.

[Gre15] T. Greenwood. *Coefficients of Bivariate Analytic Functions with Algebraic Singularities.* PhD thesis, University of Pennsylvania, 2015.

[GSS11] S. Gharan, A. Saberi, and M. Singh. A randomized rounding approach to the Traveling Salesman Problem. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 550–559. IEEE Computer Science: Los Alamitos, CA, 2011.

[HLP59] G.H. Hardy, J. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, Cambridge, 1959.

[LPRS16] J. Lebowitz, B. Pittel, D. Ruelle, and D. Speer. Central limit theorem, Lee-Yang zeros, and graph counting polynomials. *J. Comb. Theory, Ser. A*, 141:147–183, 2016.

[Pem12] R. Pemantle. Hyperbolicity and stable polynomials in combinatorics and probability. In *Current Developments in Mathematics*, pages 57–124. International Press, Somerville, MA, 2012.

[PP14] R. Pemantle and Y. Peres. Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Comb. Prob. Comput.*, 23:140–160, 2014.

[PW04] R. Pemantle and M.C. Wilson. Asymptotics of multivariate sequences, II. Multiple points of the singular variety. *Combin. Probab. Comput.*, 13:735–761, 2004.

[PW08] R. Pemantle and M.C. Wilson. Twenty combinatorial examples of asymptotics derived from multivariate generating functions. *SIAM Review*, 50:199–272, 2008.

[Ros89] S. Rosset. Normalized symmetric functions, Newton's inequalities and a new set of stronger inequalities. *Amer. Math. Monthly*, 96:815–819, 1989.

[Sos00] A. Soshnikov. Determinantal random point fields. *Uspekhi Mat. Nauk*, 55:107–160, 2000.

[Sos02] A. Soshnikov. Gaussian limit for determinantal random point fields. *Ann. Probab.*, 30:171–187, 2002.

6 Appendix

Proof of Theorem 2.1

Lemma 6.1. *Whenever (X,Y) is stable and $b \geq a$ are positive integers, the probability generating function for $aX+bY$ has no zeros in the open disk of radius δ about 1, where $\delta := \sin(\pi/b)$.*

PROOF: If $f(x,y)$ is the pgf for (X,Y) then the pgf for $aX+bY$ is $f(z^a,z^b)$. Stability of f implies that $f(z^a,z^b)$ has no zeros whose argument z lies in the open interval $(0,\pi/b)$. Invariance under conjugation and the fact that a probability generating function can never have positive real zeros implies that $f(z^a,z^b)$ is in fact zero-free on the sector $\{z : |\mathrm{Arg}\,(z)| < \pi/b\}$. The nearest point to 1 in this sector is at distance δ. $\square$

Lemma 6.2 ([BMR97, Corollary 4.3])**.** *Let $\mathcal{L}$ be an infinite family of $(d-1)$-dimensional subspaces of $\mathbb{R}^d$. Let π_L denote projection of measures onto L, in other words $\pi_L\mu := \mu \circ \pi_L^{-1}$. Let μ be a probability measure on $\mathbb{R}^d$ with finite moment generating function in a neighborhood of the origin and let ν be any probability measure on $\mathbb{R}^d$. Suppose that the projections $\pi_L\mu$ and $\pi_L\nu$ coincide for every $L \in \mathcal{L}$. Then $\mu = \nu$.* $\square$

Corollary 6.3. *Let μ be a centered Gaussian law on $\mathbb{R}^d$ and let $\mathcal{L}$ be an infinite family of $(d-1)$-dimensional subspaces of $\mathbb{R}^d$. Suppose $\{\mu_n\}$ is a sequence of probability measures on $\mathbb{R}^d$ such that for each $L \in \mathcal{L}$, the projections $\pi_L\mu_n$ converge in the weak topology as $n \to \infty$ to $\pi_L\mu$. Then $\mu_n \to \mu$.*

PROOF: Convergence of $\pi_L\mu_n$ for more than one hyperplane L implies tightness of the family $\{\mu_n\}$. Therefore, any subsequence of $\{\mu_n\}$ has a convergent subsubsequence; denote its limit by ν. It suffices to show that $\nu = \mu$. Each π_L is continuous, therefore $\pi_L\nu = \lim_{n\to\infty} \pi_L\mu_n = \pi_L\mu$. Noting that μ has moment generating function defined everywhere, the conclusion now follows from Lemma 6.2. $\square$

Lemma 6.4 ([LPRS16, Theorem 2.1])**.** *Let f be the generating polynomial for a probability law Q on the nonnegative integers. Let N denote the degree of f. Let m and σ^2 respectively denote the mean and variance of Q and let F denote the self-normalized cumulative distribution function defined by*

$$F(x) := \sum_{k \leq m + x\sigma} Q(k)\,.$$

Let $\mathcal{N}(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,dt$ denote the standard normal CDF. Given $\delta > 0$, there exists a constant C_δ depending only on δ such that if f has no roots in the ball $\{z : |z-1| < \delta\}$ then

$$\sup_{x\in\mathbb{R}} |F(x) - G(x)| \leq C_\delta \frac{N^{1/3}}{\sigma}\,.$$

PROOF: The result as stated in [LPRS16, Theorem 2.1], in the special case $z_0 = 1$ has the upper bound $B_1N/\sigma^3 + B_2N^{1/3}/\sigma$ with B_1 and B_2 depending on δ. Because $|F - \mathcal{N}|$ is never more than 1, we may assume that $N^{1/3}/\sigma \leq B_2^{-1}$, whence $B_1N/\sigma^3 \leq (B_1/B_2^2)N^{1/3}/\sigma$. Setting $C = C_\delta = B_2 + B_1/B_2^2$ recovers the result in our form. The result as stated holds for $N > N_0(\delta)$, but with CN_0 in place of C it holds for all N. $\square$

PROOF OF THEOREM 2.1: We will apply Corollary 6.3 with $\mu = N(0,A)$ and $\mathcal{L}$ equal to the set of lines through the origin with positive rational slope. Given $L \in \mathcal{L}$, let (a,b) be a positive integer pair in L. Then $\pi_L(X,Y) = (aX+bY)/\sqrt{a^2+b^2}$ and $\pi_L\mu = N(0,V)$ where

$$V := V(a,b) := \frac{\alpha\,a^2 - 2\beta\,ab + \gamma\,b^2}{a^2+b^2}\,.$$

According to Corollary 6.3, the theorem will follow if we can show that

$$\frac{a}{\sqrt{a^2+b^2}}\frac{X_n - \mathbb{E}X_n}{s_n} + \frac{b}{\sqrt{a^2+b^2}}\frac{Y_n - \mathbb{E}Y_n}{s_n} \to N(0,V(a,b)) \tag{6}$$

weakly for fixed positive integers a and b as $n \to \infty$. We proceed to show this.

First, if $V(a,b) = 0$, we observe that the left-hand side of (6) has mean zero and variance

$$\frac{\alpha_n\,a^2 - 2\beta_n\,ab + \gamma_n\,b^2}{(a^2+b^2)s_n^2} = o(1)$$

by the assumption that $A_n/s_n^2 \to A$. Weak convergence to δ_0, which is the right-hand-side of (6), follows from Chebyshev's inequality.

Assume now that $V \neq 0$. By Lemma 6.1, for all n, the generating polynomial g_n for aX_n+bY_n has no zeros within distance $\delta := \sin(\pi/b)$ of 1. Apply Lemma 6.4 to the generating polynomial g_n with $N = (a+b)M_n$. In the notation of Lemma 6.4,

$$\begin{aligned} m &= a\mathbb{E}X_n + b\mathbb{E}Y_n\,; \\ \sigma^2 &= a^2\alpha_n - 2ab\beta_n + b^2\gamma_n\,. \end{aligned}$$

The assumption $s_n^{-2}A_n \to A$ implies that

$$\sigma^2/s_n^2 \to (a^2+b^2)V\,. \tag{7}$$

The conclusion of the lemma is that $[a(X_n - \mathbb{E}X_n) + b(Y_n - \mathbb{E}Y_n)]/\sigma$ differs from a standard normal by at most

$$C_\delta \frac{N^{1/3}}{\sigma} = (1+o(1)) C_\delta \frac{(a+b)^{1/3}}{\sqrt{a^2+b^2}\sqrt{V}} \frac{M_n^{1/3}}{s_n}\,.$$

By hypothesis $M_n^{1/3}/s_n \to 0$. Thus $[a(X_n - \mathbb{E}X_n) + b(Y_n - \mathbb{E}Y_n)]/\sigma \to N(0,1)$; multiplying through by $\sqrt{V}$ and plugging in (7) gives (6). $\square$

Proof of Theorem 2.1′ More or less the same argument works to prove a multivariate CLT for real stable distributions in d variables. The first of the two necessary generalizations is immediate.

Lemma (6.1′). *If* $(\mathbf{X})$ *is stable and* $\mathbf{a}$ *is a positive integer vector, then the probability generating function for* $\mathbf{a}\cdot\mathbf{x}$ *has no zeros in the open disk of radius* δ *about 1, where* $\delta := \sin(\pi/\max_j a_j)$. $\square$

Corollary (6.3′). *Let* μ *be a centered Gaussian law on* $\mathbb{R}^d$. *If* $\{\mu_n\}$ *is a sequence of probability measures such that for all positive rational lines* L, *the projections* $\pi_L \mu_n$ *converge to* $\pi_L \mu$, *then* $\mu_n \to \mu$.

PROOF: We prove by induction on m that $\pi_L\mu_n \to \pi_L\mu$ for all m-dimensional subspaces containing a positive rational point. It is true by hypothesis when $m = 1$. Assume for induction that it is true for dimensions smaller than m. Fix any m-dimensional subspace L containing a positive rational point and apply Corollary 6.3 with $\pi_L\mu$ in place of μ and the infinite family of subspaces $L' \subseteq L$ having a basis of $m-1$ positive rational vectors in place of $\mathcal{L}$. By the induction hypothesis, each $\pi_{L'}\mu_n \to \pi_{L'}\mu$, so by Corollary 6.3, $\pi_L\mu_n \to \pi_L\mu$, completing the induction. Once $m = d$, the corollary is proved. $\square$

These two results imply the extension of Theorem 2.1 to d variables.

Proof of Theorem 3.1

Lemma 6.5. *If* $(X_1, \ldots, X_r)$ *is a random integer vector whose* r*-variate generating function is real stable, then its covariance matrix has nonnegative row and column sums.*

PROOF: The row sums of the covariance matrix are the values $\mathbb{E}(X_i - \mu_i)\sum_j (X_j - \mu_j)$. The argument may be reduced to the case $r = 2$ by considering the pair (X_i, Y_i) where $Y_i := \sum_{i \neq j} X_j$. Without loss of generality, we therefore assume $r = 2$ and denote the pair (X_1, X_2) by (X, Y). We first claim that for all k,

(8) $\mathbb{E}(Y|X=k) \leq \mathbb{E}(Y|X=k-1) \leq \mathbb{E}(Y|X=k)+1\,.$

This follows from the strong Rayleigh property for the polarization $(X_1, \ldots, X_m, Y_1, \ldots, Y_m)$ of (X, Y). Indeed, suppose the polarization is coupled to (X, Y) so that for each k, ℓ, the conditional law of $(X_1, \ldots, X_m, Y_1, \ldots, Y_m)$ given $X = k, Y = \ell$ is the product $\nu_{m,k} \times \nu_{m,\ell}$ where $\nu_{m,j}$ is uniform on sequences of zeros and ones of length m summing to j. Then

$$\mathbb{E}(Y|X=k) = E\left(\sum_{j=1}^m Y_j \,\middle|\, X_j = \mathbf{1}_{j \leq k} : 1 \leq j \leq m \right)$$

and the claim follows from the stochastic covering property for strong Rayleigh measures [PP14, Proposition 2.2]. In fact it is only the right-hand inequality of (8) that we need. Adding X gives $\mathbb{E}(X+Y|X = k-1) \leq \mathbb{E}(X+Y|X=k)$. Thus $\mathbb{E}(X+Y|X)$ is a monotone increasing function of X. This immediately implies nonnegative correlation of the bounded variables X and $X+Y$, which is the conclusion of the lemma. $\square$

The next lemma is stated generally though it is used for one specific purpose, namely for the covariance matrix of a collection of random integers with real stable generating function.

Lemma 6.6. *Let* M *be any symmetric matrix with nonnegative diagonal entries, nonpositive off-diagonal entries and nonnegative row sums. Then the index set* $[m]$ *may be partitioned into disjoint sets* T *and* $\{S_\alpha\}$ *such that* $M_{i,j} = 0$ *when* i *and* j *are in different sets of the partition. This can be done in such a way that the restriction* $M|_T$ *is nonsingular, while the restrictions* $M|_{S_i}$ *have one-dimensional null spaces containing the vectors with all entries equal.*

PROOF: Recall that $\mathcal{N}$ denotes the null space of M. Choose any nonzero vector $\mathbf{v} \in \mathcal{N}$ with minimal support set S, meaning that no vector whose support is a proper subset of S is in the null space of M. Suppose $\mathbf{v}$ has coordinates of mixed sign. Let E be the set of indices of positive coordinates and F the set indices of negative coordinates. Let M' be the 2×2 matrix indexed by the set $\{E, F\}$ whose (G, G')-element is $\sum_{i \in G, j \in G'} M_{ij}$. This matrix also has nonnegative diagonal entries (follows from nonnegativity of row sums and nonpositivity of off-diagonal elements), nonpositive off-diagonal entries (obvious) and nonnegative row sums. It has a vector of mixed signs in its null space, namely $(\sum_{j \in E} v_j, \sum_{j \in F} v_j)$, hence must be the 2×2 zero matrix. This means that the $\mathbf{v}_E$ and $\mathbf{v}_F$ are each separately in the null space (where $\mathbf{v}_G$ denotes the vector whose j^{th} coordinate is $v_j \mathbf{1}_G(j)$). This contradicts the minimality of the support of $\mathbf{v}$. We conclude that all elements of the null space with minimal support have coordinates all of one sign.

Still assuming $\mathbf{v}$ to have minimal support set $S \subseteq \mathcal{N}$, consider the sub-collection $\{X_j : j \in S\}$, which inherits the properties in the hypotheses. Its covariance matrix M' is the submatrix of M indexed by S. Assume for contradiction that the coordinates of $\mathbf{v}$ are not equal. Let $\mathbf{w}$ be the all ones vector of the same length as $\mathbf{v}$. Scale $\mathbf{v}$ so that its minimum coordinate is equal to 1. If $v_i = 1$ then

$$0 = (M'\mathbf{v})_i \geq (M'\mathbf{w})_i \geq 0\,,$$

the last inequality following from nonnegativity of the row sums. It follows that $M_{ij} = 0$ for all i, j such that $v_i = 1 < v_j$. Thus $S' := \{i : v_i = 1\}$ is a proper subset of S whose indicator vector is in the null space of M. By contradiction, $\mathbf{v} = \mathbf{w}$ as desired.

Finally, if $\mathbf{w}_S$ and $\mathbf{w}_T$ are vectors of ones and zeros with support sets S and T respectively and these are not disjoint, then $\mathbf{w}_S - \mathbf{w}_T \in \mathcal{N}$ and is of mixed sign, a contradiction. This finishes the proof. $\square$

Proof of Theorem 3.1: The conclusions of Lemma 6.5 pass to the limit: the limiting covariance matrix A has nonnegative row sums as well as being symmetric with nonnegative diagonal entries and nonpositive off-diagonal entries. The conclusions of Lemma 6.6 then follow as well. Fix $\mathbf{v}$ such that $\mathbf{v}A = 0$. It follows from Lemma 6.6 that $\mathbf{w}A = 0$ as well. The random variables $Z_n = \mathbf{w} \cdot \mathbf{X}^{(n)}$ are univariate real stable, hence subject to the real stable CLT (1). In particular, $\sigma_n^{-1}(Z^{(n)} - \mathbb{E}Z^{(n)}) \to N(0,1)$ weakly whenever $\sigma_n := \mathrm{Var}\,(Z_n)^{1/2} \to \infty$. $\square$

Proof of Theorem 4.3 The proof is by induction on the degree n of f. Let $r_1, \ldots, r_n$ be the negatives of the roots of f, and let $e_j = e_j(r_1, \ldots, r_n)$ be the elementary symmetric functions:

$$e_0 = 1, \quad e_1 = \sum_i r_i, \quad e_2 = \sum_{i<j} r_i r_j, \ldots.$$

Assuming without loss of generality that f is monic,

$$f(x) = \prod_{i=1}^{n}(x + r_i) = \sum_{j=0}^{n} x^j e_{n-j}. \tag{9}$$

Then

$$g_i(y) = \sum_{j=0}^{\lfloor \frac{n-i}{k} \rfloor} y^j e_{n-kj-i}.$$

For the base step of the induction, take $n < 2k$, so that the g_i's are linear or constant. In fact, $g_i(y) = e_{n-i}$ if $i > n-k$ and $g_i(y) = e_{n-i} + y e_{n-k-i}$ if $i \leq n-k$. In the latter case, the root is $-e_{n-i}/e_{n-k-i}$, so the interlacement property is a consequence of the log concavity of the sequence e_m:

$$\frac{e_{m+1}}{e_m} \downarrow .$$

This statement is a consequence of Newton's inequalities; see [HLP59] and [Ros89].

Now assume the result for a given n, let f be as in (9), consider the polynomial of degree $n+1$

$$F(x) = (x + r)f(x), \quad r > 0,$$

and its decomposition

$$F(x) = \sum_{i=0}^{k-1} x^i G_i(x^k).$$

If $e'_j = e'_j(r_1, \ldots, r_n, r)$ are the elementary symmetric functions corresponding to the longer sequence, $e'_j = e_j + r e_{j-1}$, so

$$\begin{aligned} G_i(y) &= \sum_{j=0}^{\lfloor \frac{n+1-i}{k} \rfloor} y^j e'_{n+1-kj-i} \\ &= \sum_{j=0}^{\lfloor \frac{n+1-i}{k} \rfloor} y^j [e_{n+1-kj-i} + r e_{n-kj-i}] \\ &= r g_i(y) + \begin{cases} y g_{k-1}(y) & \text{if } i = 0; \\ g_{i-1}(y) & \text{if } i \geq 1. \end{cases} \end{aligned} \tag{10}$$

Now we use this to determine the sign of $G_i(s_j)$. The signs of g_i alternate between intervals separated by the roots of g_i, since all roots are simple. Also, $g_i(0) > 0$ for each i.

We describe the argument in the following array, in case $k = 3$:

$$\begin{pmatrix} & \cdots & s_6 & s_5 & s_4 & s_3 & s_2 & s_1 & s_0 & 0 \\ \\ g_0 & \cdots & 0 & + & + & 0 & - & - & 0 & + \\ g_1 & \cdots & + & + & 0 & - & - & 0 & + & + \\ g_2 & \cdots & + & 0 & - & - & 0 & + & + & + \\ \\ G_0 & \cdots & - & + & + & + & - & - & - & + \\ G_1 & \cdots & + & + & + & - & - & - & + & + \\ G_2 & \cdots & + & + & - & - & - & + & + & + \end{pmatrix}.$$

Note that each row is periodic of period 6, and each row within the two groups is obtained from the previous row via a shift. Here are some examples of the computation for the bottom rows:

$$\begin{aligned} G_0(s_2) &= r g_0(s_2) + s_2 g_2(s_2) = r g_0(s_2) < 0\,, \\ G_2(s_3) &= r g_2(s_3) + g_1(s_3) < 0. \end{aligned}$$

More generally note that the induction hypothesis implies that

(11)

$$g_i(s_j)\begin{cases} <0 & \text{if } \frac{j-i}{k}\in\cup_{m=0}^{\infty}(2m,2m+1);\\ =0 & \text{if } \frac{j-i}{k}\in\{0,1,2,\dots\};\\ >0 & \text{if } \frac{j-i}{k}\in(-\infty,0)\cup\cup_{m=0}^{\infty}(2m+1,2m+2).\end{cases}$$

We would like to show that

(12)

$$G_i(s_j)\begin{cases} <0 & \text{if } \frac{j-i}{k}\in\cup_{m=0}^{\infty}[2m,2m+1);\\ >0 & \text{if } \frac{j-i}{k}\in(-\infty,0)\cup\cup_{m=0}^{\infty}[2m+1,2m+2).\end{cases}$$

There are several cases to consider. First take $i=0, 2mk\le j<(2m+1)k$ for some $m\ge 0$. Then by (10),

$$G_0(s_j)=rg_0(s_j)+s_jg_{k-1}(s_j).$$

By (11), $g_0(s_j)=0$ if $j=2mk$ and is <0 otherwise, while $g_{k-1}(s_j)=0$ if $j=(2m+1)k-1$ and >0 otherwise. Since $r>0$ and $s_j<0$, $G_0(s_j)<0$ as required. The next case is $i=0,(2m+1)k\le j<(2m+2)k$ for some $m\ge 0$. Now $g_0(s_j)=0$ if $j=(2m+1)k$ and >0 otherwise, while $g_{k-1}(s_j)=0$ if $j=(2m+2)k-1$ and <0 otherwise, so $G_0(s_j)>0$.

Next take $i\ge 1$ and $2mk\le j-i<(2m+1)k$ for some $m\ge 0$. Now

$$G_i(s_j)=rg_i(s_j)+g_{i-1}(s_j),$$

$g_i(s_j)=0$ if $j-i=2mk$ and is <0 otherwise, and $g_{i-1}(s_j)=0$ if $j-i=2(m+1)k-1$ and is <0 otherwise, so $G_i(s_j)<0$. If, on the other hand, $i\ge 1$ and $j<i$ or $(2m+1)k\le j-i<(2m+2)k$ for some $m\ge 0$, $g_i(s_j)=0$ if $j-i=(2m+1)k$ and is >0 otherwise, and $g_{i-1}(s_j)=0$ if $j-i=(2m+2)k-1$ and is >0 otherwise, so $G_i(s_j)>0$.

From (12) we see that G_i has a root in each interval of the form

$$(s_{mk+i},s_{mk+i-1}) \tag{13}$$

for $0\le m\le\frac{n-k-i}{k}$. (By convention, we set $s_{-1}=0$.) This shows that G_i has at least $\lfloor\frac{n-i}{k}\rfloor$ negative roots. The degree of G_i is $\lfloor\frac{n+1-i}{k}\rfloor$. We see that all roots of G_i are negative, except possibly in case $\lfloor\frac{n-i+1}{k}\rfloor=\lfloor\frac{n-i}{k}\rfloor+1$. In this case, the extra root is recovered by noting that, with $m=\lfloor\frac{n-i}{k}\rfloor$,

$$\begin{aligned}(-1)^mG_i(s_{(m-1)k+i}) &> 0 \quad\text{and}\\ (-1)^mG_i(s) &< 0 \quad\text{for large negative } s.\end{aligned}$$

Therefore, G_i has the correct number of negative roots. The interlacement property follows from the form of the intervals in (13):

$$t_{n-k+1}<s_{n-k}<t_{n-k}\cdots<s_1<t_1<s_0<t_0<0$$

where the roots of G_i are $t_i,t_{i+k},t_{i+2k},\dots$. This completes the induction step. $\square$