

eScholarship

Combinatorial Theory

Title

Maximum entropy and integer partitions

Permalink

<https://escholarship.org/uc/item/0nw2d8hb>

Journal

Combinatorial Theory, 3(1)

ISSN

2766-1334

Authors

McKinley, Gweneth
Michelen, Marcus
Perkins, Will

Publication Date

2023

DOI

10.5070/C63160420

Supplemental Material

<https://escholarship.org/uc/item/0nw2d8hb#supplemental>

Copyright Information

Copyright 2023 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

MAXIMUM ENTROPY AND INTEGER PARTITIONS

Gweneth McKinley¹, Marcus Michelen^{*2}, and Will Perkins^{†3}

¹*Department of Mathematics, University of California San Diego, California, U.S.A.
gmckinley@ucsd.edu*

²*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Illinois, U.S.A.
michelen.math@gmail.com*

³*School of Computer Science, Georgia Institute of Technology, Georgia, U.S.A.
math@willperkins.org*

Submitted: Jul 21, 2021; Accepted: Nov 1, 2022; Published: Mar 15, 2023

© The authors. Released under the CC BY license (International 4.0).

Abstract. We derive asymptotic formulas for the number of integer partitions with given sums of j th powers of the parts for j belonging to a finite, non-empty set $J \subset \mathbb{N}$. The method we use is based on the ‘principle of maximum entropy’ of Jaynes. This principle leads to an intuitive variational formula for the asymptotics of the logarithm of the number of constrained partitions as the solution to a convex optimization problem over real-valued functions. Finding the polynomial corrections and leading constant involves two steps: quantifying the error in approximating a discrete optimization problem by a continuous one and proving a multivariate local central limit theorem.

Keywords. Integer partitions, maximum entropy, asymptotic enumeration, local central limit theorem, limit shape

Mathematics Subject Classifications. 05A17, 05A16, 60F05

1. Introduction

An integer partition is a finite multiset λ of positive integers. It is a partition of n if $\sum_{x \in \lambda} x = n$. The partition number $p(n)$ counts the number of different partitions of n . A classical result of Hardy and Ramanujan [HR18], obtained using Euler’s generating function and the Hardy–Littlewood circle method, gives the asymptotics of $p(n)$:

$$p(n) = \frac{1 + o(1)}{4\sqrt{3}n} e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}} \quad (1.1)$$

*Supported in part by NSF grant DMS-2137623.

†Supported in part by NSF grant DMS-1847451.

as $n \rightarrow \infty$. Since then, partitions and partition numbers have been extensively studied, and analytic, probabilistic, and combinatorial methods for analyzing partition numbers have been developed and refined.

In this paper we prove asymptotic formulas for the number of partitions of a very general type: those that prescribe the sums of the j th powers of the parts of the partition for j belonging to some finite set of non-negative integers J . In doing so we also provide an intuitive explanation for three factors that lead to (1.1) and other asymptotic formulas.

In particular, given such a set J , and a vector $\alpha \in \mathbb{R}_+^J$, and $n \in \mathbb{N}$, let $\tilde{p}_n(\alpha)$ denote the number of integer partitions λ so that

$$\sum_{x \in \lambda} x^j = \lfloor \alpha_j n^{(j+1)/2} \rfloor \quad \text{for all } j \in J. \quad (1.2)$$

The scaling by $n^{(j+1)/2}$ is chosen to allow the existence of non-trivial limit shapes, as we will see below in Section 1.4.

A special case of this problem is the classical Hardy–Ramanujan case of $J = \{1\}$ (prescribing the sum of the parts). In the same paper they stated asymptotics for partitions of n into k th powers for each fixed k ; this is the case $J = \{k\}$ in the general formulation. These asymptotic formulas were proven rigorously by Wright in 1934 [Wri34]. The general problem also includes that of partitioning n into a given number of parts [Sze53, Can97, Rom05] (obtained by taking $J = \{0, 1\}$).

While including these long studied problems as special cases, the generalization we consider here can also exhibit dramatically different behavior. When multiple sums of positive powers are constrained several new wrinkles to the problem arise, and to the best of our knowledge such cases have not been considered before. For one, a wide variety of exotic limit shapes can be obtained, including those with multiple inflection points (Section 1.4). More broadly, the generalization emphasizes and reveals connections with convex optimization and information theory, in both the asymptotic formulas obtained and in the methods applied. In the next section we outline the method we apply here, based on Jaynes’ ‘principle of maximum entropy’, and discuss these connections.

1.1. Maximum entropy approach

Several different (and more elementary) proofs of the Hardy–Ramanujan formula have been given since the original paper [Erd42, New62], but one can ask for an intuitive explanation of the formula: why is the exponent $\pi \sqrt{\frac{2}{3}} \sqrt{n}$ and why is the polynomial correction n^{-1} ? We give an explanation here by following Jaynes’ principle of maximum entropy [Jay57]. Following this approach cleanly separates partition numbers into three factors, each of which has an intuitive explanation and whose asymptotics can be computed.

Jaynes’ principle of maximum entropy is a kind of axiom about probabilistic inference: given some measurements of observed data, the best estimate for the generating distribution, in the sense of making the fewest additional assumptions, is the distribution of maximum entropy consistent with these measurements. More concretely, given the values of one or more statistics,

the best estimate for the unknown distribution generating the data is the distribution of maximum entropy whose expectations match the observed statistics. Beyond statistical inference, maximum entropy can be used to derive many important probability distributions (Gaussian, geometric, exponential, etc.) as well as the probabilistic models that arise in statistical mechanics. Due to its connection to convex optimization, maximum entropy has also proved very useful in algorithmic settings.

Applied to our main topic, moment constrained integer partitions, the principle suggests one should consider probability distributions on the countably infinite set of all integer partitions, and in particular, one should consider the unique probability distribution on all partitions that satisfies (1.2) in expectation and maximizes entropy over all such distributions. This maximum entropy distribution μ will turn out to have some remarkable properties that will help us approximate $\tilde{p}_n(\alpha)$.

The first useful property of maximum entropy distributions is that there is an exact formula for $\tilde{p}_n(\alpha)$ in terms of μ . Let \mathcal{P} denote the set of partitions that satisfy the constraints (1.2) exactly. Then, as we prove below in Section 3,

$$\tilde{p}_n(\alpha) = e^{H(\mu)} \mu(\mathcal{P}), \tag{1.3}$$

where $H(\mu) = -\sum_x \mu(x) \log \mu(x)$ is the Shannon entropy of μ and $\mu(\mathcal{P})$ is the probability that a partition drawn according to μ is a partition of n . This formula is a consequence of a much more general fact about maximum entropy distributions (given in Lemmas 3.1 and 3.2 below) and is inspired by the work of Barvinok and Hartigan [BH10, BH12, Bar17] on counting integer points in polytopes.

The second useful property of constrained maximum entropy distributions is that they can be determined via convex programming. This property has been used to great effect in several recent results in theoretical computer science [SV14, AGM⁺17, AGV18] and is also used in [BH10]. In the case of integer partitions, the description of μ is relatively explicit.

We now sketch a derivation of this distribution. A probability distribution on partitions is a joint distribution of non-negative integer-valued random variables $(Y_k)_{k \geq 1}$ indexed by the natural numbers. The expectation constraints are that for each $j \in J$, the sum of the means of these distributions times their indices to the j th power equals $\lfloor \alpha_j n^{(j+1)/2} \rfloor$; that is,

$$\sum_{k \geq 1} k^j \eta_k = \lfloor \alpha_j n^{(j+1)/2} \rfloor \quad \text{for all } j \in J \tag{1.4}$$

where $\eta_k = \mathbb{E}Y_k$ is the expected number of parts of size k . The constraints are given entirely in terms of the means η_k ; nothing else about the joint distribution of the Y_k 's enter in. Since entropy is maximized by a product measure, and a geometric random variable has the greatest entropy of any non-negative integer-valued random variable with a given mean, the maximizing distribution must be a collection of independent geometric random variables. Their means can be determined by solving a discrete convex program. To compute the asymptotics of $H(\mu)$, we scale this discrete optimization problem and take a limit, obtaining a continuous convex program ((1.7) below) on real valued functions whose optimum determines the exponential growth rate of $\tilde{p}_n(\alpha)$. We then use Euler–Maclaurin summation to measure the approximation error.

The final step is approximating the probability that μ yields a partition satisfying all constraints by proving a multivariate local central limit theorem. The generality of the types of partitions we enumerate necessitates some new technical ideas here. As with many local central limit theorems, we write a probability as an integral. The elimination of so-called *minor arcs*—i.e. the portion of the integral that contributes an essentially negligible amount—requires a quantitative equidistribution result of Green and Tao [GT12]. Establishing this multivariate local central limit is perhaps the most technically challenging part of the proof; it appears in Section 6.

1.2. Main results

We now define precisely the class of partitions we enumerate. Let J be a finite set of non-negative integers containing at least one positive integer, and let $\mathbf{N} = (\mathbf{N}_j)_{j \in J}$ be a vector of positive integers indexed by J . A partition λ has *profile* \mathbf{N} if

$$\sum_{x \in \lambda} x^j = \mathbf{N}_j \quad \text{for all } j \in J.$$

We call J the *profile set*.

Let $\mathcal{P}(\mathbf{N})$ denote the set of partitions with profile \mathbf{N} and let $p(\mathbf{N}) = |\mathcal{P}(\mathbf{N})|$. For instance, with $J = \{1\}$ and $\mathbf{N}_1 = n$, we have $p(\mathbf{N}) = p(n)$, the usual partition number. To study the asymptotics of $p(\mathbf{N})$ we normalize the profile. For $\alpha \in \mathbb{R}_+^J$ and $n \in \mathbb{N}$, let

$$\mathbf{N}(\alpha, n) = (\lfloor \alpha_j n^{(j+1)/2} \rfloor)_{j \in J}.$$

Then let $\tilde{p}_n(\alpha) = p(\mathbf{N}(\alpha, n))$. We will study the asymptotics of $\tilde{p}_n(\alpha)$ as $n \rightarrow \infty$. As mentioned above, this class of partition problems includes the classic case of partitions of n ; partitions of n into sums of k th powers; and partitions of n with a given number of parts. While these cases are all covered by our main result, the cases in which J includes more than one positive integer have not been considered before, and it is here that many of the interesting features in the problem and our approach emerge.

One new feature is that certain profiles are impossible, either for number-theoretic reasons or because the values in α are incompatible. For example, the constraints may violate the Cauchy–Schwartz inequality. The compatibility of constraints depends on the vector α and can be expressed in terms of the Stieltjes moment problem [Sti94]; this is discussed further in Section 1.5.

Other profiles are impossible for number-theoretic reasons that depend on n . For instance, since $k^2 \equiv k \pmod{2}$ for all integers k , we have that $p(\mathbf{N}) = 0$ if $\mathbf{N}_1 \not\equiv \mathbf{N}_2 \pmod{2}$. A concise way of describing this particular constraint is that the polynomial $\frac{1}{2}x^2 + \frac{1}{2}x$ is *integer-valued* meaning that $\frac{1}{2}m^2 + \frac{1}{2}m \in \mathbb{Z}$ for all $m \in \mathbb{Z}$. Thus a necessary condition for $\mathcal{P}(\mathbf{N}) \neq \emptyset$ is that $\frac{1}{2}\mathbf{N}_2 + \frac{1}{2}\mathbf{N}_1 \in \mathbb{Z}$. It will turn out that *all* number-theoretic obstructions can be defined in this way. Let

$$\mathcal{Q}_J = \left\{ \sum_{j \in J} t_j x^j : t_j \in (-1/2, 1/2] \text{ and } \sum_{j \in J} t_j m^j \in \mathbb{Z} \text{ for all } m \in \mathbb{Z} \right\} \quad (1.5)$$

be the set of integer-valued polynomials using only powers in J and having coefficients in $(-1/2, 1/2]$. We subsequently define the set

$$\text{NT} := \text{NT}(J) = \left\{ (m_j)_{j \in J} \in \mathbb{Z}^J : a_j m_j \in \mathbb{Z} \text{ for all } \sum_{j \in J} a_j x^j \in \mathcal{Q}_J \right\}.$$

It follows from the definition that if $\mathbf{N} \notin \text{NT}$ then $p(\mathbf{N}) = 0$. We say α is n -feasible if $\mathbf{N}(\alpha, n) \in \text{NT}$.

To connect $p(\mathbf{N})$ to the principle of maximum entropy, we define μ to be the maximum entropy distribution on the set of all partitions so that

$$\mathbb{E}_{\lambda \sim \mu} \sum_{x \in \lambda} x^j = \mathbf{N}_j$$

for all $j \in J$. We will see below that under one assumption, the maximum entropy distribution exists, is unique, and can be represented as a collection of independent geometric random variables. We then prove the identity (1.3) relating $p(\mathbf{N})$ to μ .

Equipped with (1.3), we pose a continuous convex program that determines the exponential growth rate of $\tilde{p}_n(\alpha)$ (in \sqrt{n}). Recall from above that the maximum entropy distribution will consist of independent geometric random variables. The Shannon entropy of a geometric random variable with mean η is

$$G(\eta) := (\eta + 1) \log(\eta + 1) - \eta \log \eta. \tag{1.6}$$

We now define

$$\begin{aligned} M(\alpha) &= \max_{f \in \mathcal{F}} \int_0^\infty G(f(x)) dx \\ \text{subject to } & \int_0^\infty x^j f(x) dx = \alpha_j \text{ for } j \in J, \end{aligned} \tag{1.7}$$

where \mathcal{F} is the set of all integrable functions $f : [0, \infty) \rightarrow [0, \infty)$. Note that the objective function to be maximized is strictly concave, and the constraints linear, so we have a convex program.

The assumption we make on α ensures feasibility of this continuous convex program and the existence of an optimal solution.

Assumption 1. *There exists $\beta \in \{\mathbb{R} \setminus 0\}^J$ so that*

$$\int_0^\infty \frac{x^j}{\exp(\sum_{\ell \in J} \beta_\ell x^\ell) - 1} dx = \alpha_j \text{ for } j \in J. \tag{1.8}$$

In fact, for these integrals to converge, β must belong to a certain convex subset of \mathbb{R}^J : those vectors for which the polynomial $\sum_{j \in J} \beta_j x^j$ is positive on $(0, \infty)$. We discuss Assumption 1 and its connections to other problems in optimization and information theory in Section 1.5.

Assumption 1 implies the optimization problem (1.7) is feasible; in fact via convex duality the vector β is a certificate that the optimizer is

$$f^*(x) := \frac{1}{\exp\left(\sum_{j \in J} \beta_j x^j\right) - 1}. \quad (1.9)$$

The optimum $M(\alpha)$ determines the growth rate of the entropy of the maximum entropy distribution μ from (1.3), and thus the growth rate of $\tilde{p}_n(\alpha)$. We can now state our main result.

Theorem 1.1. *For all profile sets J and all $\alpha \in \mathbb{R}_+^J$ satisfying Assumption 1,*

$$\tilde{p}_n(\alpha) = (c(\alpha) + o(1)) \cdot \frac{e^{\sqrt{n}M(\alpha)}}{n^{b(J)}}$$

when α is n -feasible (and 0 otherwise).

The constant $b(J)$ is given by

$$b(J) = \frac{j_* + |J|}{4} + \frac{1}{2} \sum_{j \in J} j,$$

where $j_* := \min J$. The constant $c(\alpha)$ depends on α implicitly through the vector β guaranteed by Assumption 1 and is given by

$$c(\alpha) = \frac{|Q_J|}{(2\pi)^{\frac{j_* + |J|}{2}} (\det \Sigma)^{1/2}} \cdot \beta_{j_*}^{\frac{1_{j_* \geq 1}}{2}} \cdot \exp\left(\frac{1_{j_* = 0}}{2} \left(\frac{\beta_0}{e^{\beta_0} - 1} - G\left(\frac{1}{e^{\beta_0} - 1}\right)\right)\right)$$

where

$$\Sigma = \left(\int_0^\infty x^{i+j} \frac{\exp\left(\sum_{\ell \in J} \beta_\ell x^\ell\right)}{\left(\exp\left(\sum_{\ell \in J} \beta_\ell x^\ell\right) - 1\right)^2} dx \right)_{i,j \in J} \quad (1.10)$$

and Q_J is defined in (1.5).

In Section 2 we make things more concrete by sketching the maximum-entropy-based proof of Theorem 1.1 in the special case of $J = \{1\}$ before proving the general result in the remaining sections.

1.3. Related work

While maximizing entropy is implicit in other methods of enumerating partitions (including saddlepoint and large deviations methods), our proof method is a direct application of maximum entropy to counting (inspired by Barvinok and Hartigan [BH10] and presented in Section 3). The maximum entropy method works by analyzing the maximum entropy distribution μ . Distributions on partitions with independent coordinates have often arisen in the study of the structure of typical integer partitions. Indeed Fristedt identified the distribution μ above from the form of the generating function of $p(n)$ [Fri93], though he did not connect it with maximum entropy. Vershik [Ver96, Ver97] and Vershik and Yakubovich [VY01], in the context of finding limiting

shapes of partitions, considered related distributions and noted that they can be interpreted as grand canonical distributions from statistical physics. See also [DVZ00] in which large deviations for limit shapes are approached via the same type of distribution.

Melczer, Panova, and Pemantle [MPP20] noted that while such distributions have been commonly used to determine limit shapes, they have only rarely been used to prove asymptotic enumeration results (their results and Takács [Tak86] being the exceptions). The approach of [MPP20] shares some important steps in common with our approach here: both are probabilistic approaches to enumeration, and both solve a limiting variational problem. Melczer, Panova, and Pemantle enumerate partitions by proving a local large deviation principle, which, when solved, produces a probability distribution on partitions given by independent geometric random variables with specified means. In fact computing a large deviation rate function in this setting is equivalent to minimizing the Kullback–Leibler divergence between probability measures (as in Sanov’s Theorem [San58]), which is essentially an entropy maximization problem (see the discussion in, e.g. [Csi75]). Where the maximum entropy and large deviation approaches differ is that a large deviation approach requires a prior distribution on partitions and thus is restricted to settings such as that of bounded Young diagrams, while the principle of maximum entropy works in general, without a prior (and in fact this is an important motivation for the principle itself: to be able to generate a prior when one does not exist). The identity (1.3) provides a direct and very general link between enumeration and probability distributions on partitions.

Finally, the last step of our approach is to prove a multivariate local central limit theorem; the use of a local central limit theorems is a common step in many approaches to asymptotic enumeration of partition numbers [Tak86, Fri93, Pit97, CCH01, Rom05, MPP20].

1.4. Limit Shapes

The scaling $\mathbf{N}(\boldsymbol{\alpha}, n) = (\lfloor \alpha_j n^{(j+1)/2} \rfloor)_{j \in J}$ is chosen so that a typical partition λ in $\mathcal{P}(\mathbf{N})$ has a limit shape. In particular, if we rescale the Young diagram of $\lambda \in \mathcal{P}(\mathbf{N})$ by \sqrt{n} in each direction, then the area of the rescaled diagram will be of roughly constant order; indeed, in the case that $1 \in J$, the rescaled area will be exactly α_1 . Informally, we say that there is a *limit shape* if the rescaled Young diagram of a uniformly random $\lambda \in \mathcal{P}(\mathbf{N})$ converges in distribution (in an appropriate sense) to a constant shape. In the classical case of $J = \{1\}$, a limit shape was shown to exist by Szalay and Turán [ST77a, ST77b] (see also [Ver96]). This shape is shown in Figure 1.2. Similarly, a limit shape for partitions whose Young diagram fit in a rectangle of constant aspect ratio was found by Petrov [Pet10].

We will show that there is a limit shape for all the cases covered by Theorem 1.1. In order to state precisely what is meant by “limit shape” some preliminaries are required. Following Vershik [Ver96], define the space $\mathcal{D} = \{\phi(t)\}$ where $\phi : \mathbb{R}_+ \rightarrow [0, \infty)$ with $\int \phi(t) dt < \infty$ and ϕ non-increasing. Endow \mathcal{D} with the topology of uniform convergence on compact sets. We will think of \mathcal{D} as the space of scaled Young diagrams and their limits, where we consider Young diagrams in French notation. For a partition λ and $n \in \mathbb{N}$, define the function

$$\phi_{\lambda,n}(t) = n^{-1/2} |\{a \in \lambda : a \geq t\sqrt{n}\}| .$$

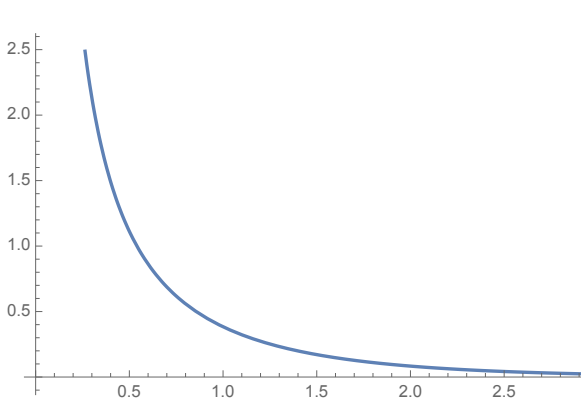


Figure 1.1: figure
The optimizer $f^* = (e^{\frac{\pi}{\sqrt{6}}x} - 1)^{-1}$ for partitions
of n .

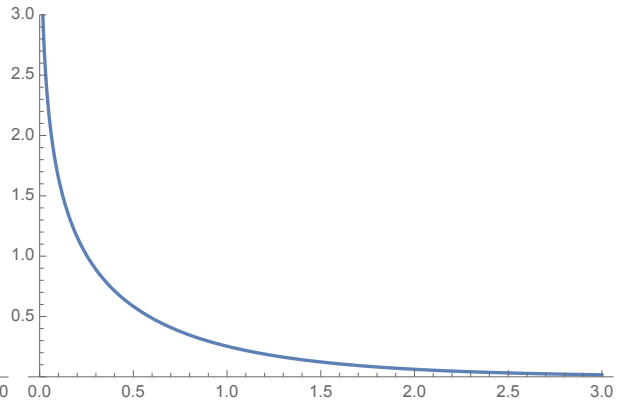


Figure 1.2: figure
The limit shape for partitions of n .

The function $\phi_{\lambda,n}(t)$ is simply the boundary of the Young diagram of λ in French notation, rescaled by \sqrt{n} in each direction. Our goal is to identify the limit shape when λ is chosen from $\mathcal{P}(\mathbf{N})$ uniformly at random; intuitively, the law of large numbers states that if λ is chosen from the maximum entropy measure μ_n instead, then we have

$$n^{-1/2} |\{a \in \lambda : a \geq x\sqrt{n}\}| \approx \int_x^\infty \frac{1}{\exp(\sum_{j \in J} \beta_j s^j) - 1} ds. \quad (1.11)$$

With this in mind, the function $\phi_\infty(t)$ defined via

$$\phi_\infty(t) = \int_t^\infty \frac{1}{\exp(\sum_{j \in J} \beta_j s^j) - 1} ds = \int_t^\infty f^*(s) ds,$$

where f^* is as in (1.9), is a strong candidate for the limit shape. This will turn out to be the case.

Theorem 1.2. *In the context of Theorem 1.1, let λ be an element of $\mathcal{P}(\mathbf{N}(\alpha, n))$ chosen uniformly at random. Then $\phi_{\lambda,n}$ converges in distribution to ϕ_∞ as $n \rightarrow \infty$.*

Note that the value $M(\alpha)$ may be viewed as a functional of the limit shape ϕ_∞ itself, since f^* can be obtained by differentiation. The relationship between the growth rate of $\exp(M(\alpha)\sqrt{n})$ and the limit shape ϕ_∞ is not new and has a long history in statistical mechanics and its adjacent fields. For instance, in a survey on the limit shapes, Shlosman shows that the asymptotic $\log p(n) \sim \pi\sqrt{2/3}\sqrt{n}$ follows from the shape theorem for partitions [Shl01]. The survey [Oko16] by Okounkov discusses many other examples of relationships—both heuristic and rigorous—between limit shapes, asymptotic enumeration and large deviation principles (see also, e.g., [CKP01]).

We now give two examples of limit shapes obtainable in Theorem 1.2. These examples were obtained by choosing β first then calculating the corresponding α .

Example 1.3. Let $J = \{0, 1, 2, 3, 4\}$ and let $\alpha = \{12.8748, 6.698, 4.66192, 3.72617, 3.15877\}$. Then $\beta \approx \{.95, -10.1, 36.5, -49.5, 22.4\}$. The limit shape is given in Figure 1.4.

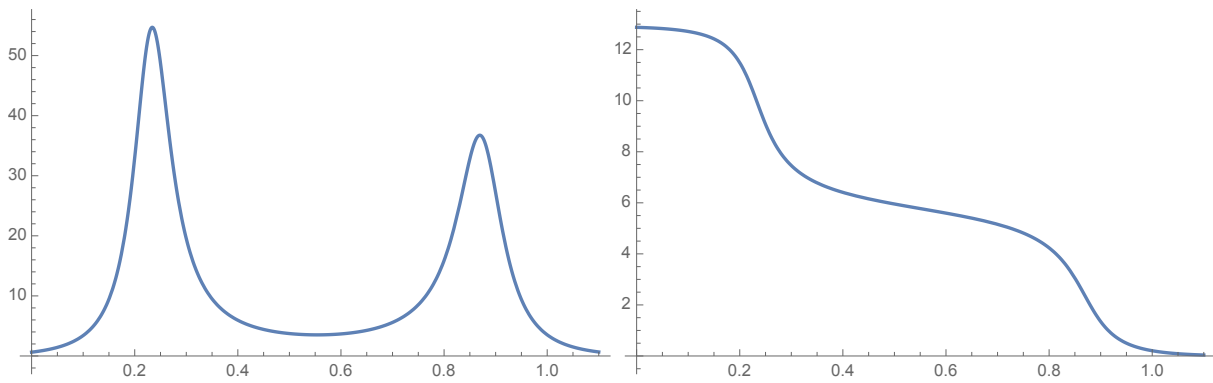


Figure 1.3: The optimizer f^* for Example 1.3. Figure 1.4: The limit shape for Example 1.3.

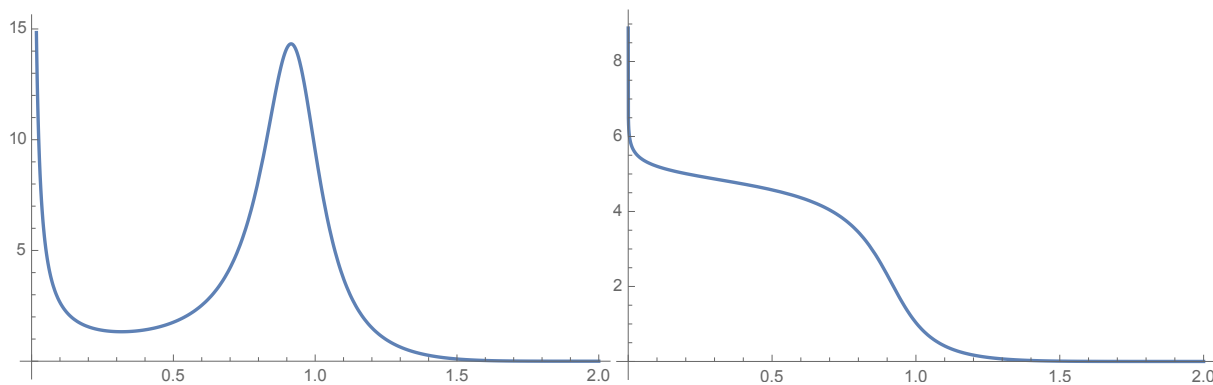


Figure 1.5: The optimizer f^* for Example 1.4. Figure 1.6: The limit shape for Example 1.4.

Example 1.4. Let $J = \{1, 2, 3\}$ and let $\alpha = \{4.31168, 3.86652, 3.65774\}$. Then $\beta \approx \{4.0, -8.5, 4.6\}$. The optimizer f^* is shown in Figure 1.5 and the limit shape is shown in Figure 1.6. Note that both have a vertical asymptote at 0 indicating that the typical number of parts of such a partition is $\omega(\sqrt{n})$.

These examples indicate some of the rich behavior possible in the setting of moment-constrained integer partitions. By specifying k moments we can obtain a limit shape with up to $k - 1$ inflection points. In fact, it is not hard to show that the set of limit shapes obtainable in the framework of Theorem 1.1 is dense in the set of all integrable, non-negative and non-increasing functions on $[0, \infty)$.

1.5. Remarks on Assumption 1

The Stieltjes moment problem [Sti94] is the problem of finding a density function of a continuous random variable supported on $[0, \infty)$ given its moments. In other words, given a sequence of positive numbers $\alpha_1, \alpha_2, \dots$, determine if there is a density f on $[0, \infty)$ with $\int x^k f(x) dx = \alpha_k$, $k \geq 1$, and if so, whether it is uniquely determined. The truncated or reduced Stieltjes moment problem is the same but with only the first d moments specified. The closely related

Hausdorff and Hamburger moment problems are the analogous problems with support $[0, 1]$ and $(-\infty, \infty)$ respectively. Stieltjes determined necessary and sufficient conditions on the sequence $\alpha_1, \alpha_2, \dots$ for the existence of a solution to the moment problem. Let A_n and \bar{A}_n be the Hankel matrices generated from $1 = \alpha_0, \alpha_1, \dots, \alpha_n$ and $\alpha_1, \alpha_2, \dots, \alpha_n$ respectively, where we recall that the Hankel matrix generated by a sequence $a_1, a_2, \dots, a_{2n-1}$ is $(a_{i+j-1})_{i,j=1}^n$. Then the Stieltjes moment problem has a solution if and only if $\det A_n > 0$ and $\det \bar{A}_n > 0$ for all n . Note that by interlacing, this is equivalent to the matrices A_n and \bar{A}_n being positive definite for all n . The conditions for the truncated Stieltjes problem are the same but only constraining the Hankel matrices formed using the specified moments. These conditions are equivalent to certain matrices being positive semidefinite [Cur91] and thus the problem of determining feasibility can be posed as a semidefinite program.

When the (truncated) moment problem is feasible there may be infinitely many solutions, and so one can ask for a principled approach to select one distribution satisfying the given constraints. Jaynes' principle of maximum entropy suggests choosing the distribution with maximum entropy subject to the constraints. Such a choice is very natural: many widely used distributions, both discrete and continuous, are maximum entropy distributions subject to constraints on the support and a small number of moments, e.g. Gaussian, exponential, geometric, and uniform (see e.g. [Cov99, Chapter 12]).

One can pose the maximum entropy Stieltjes moment problem as a continuous convex program.

$$\begin{aligned} \max_{f \in \mathcal{F}} & - \int_0^\infty f(x) \log f(x) dx & (1.12) \\ \text{subject to} & \int_0^\infty x^j f(x) dx = \alpha_j \text{ for } j = 0, \dots, d, \end{aligned}$$

where we set $\alpha_0 = 1$ to ensure that f is a probability density function. The objective function is the entropy of the distribution with density f . Note the similarity of the maximum entropy moment program (1.12) to the program (1.7), which we will call the *maximum geometric entropy moment problem*. The only difference is in the objective functions which are different strictly concave functions (also in (1.7) we may only specify a subset of the first m moments, but we could do the same in (1.12)). The two problems share essentially all of their qualitative features. To describe these features, let us assume for now that $J = \{0, \dots, d\}$ and that $\alpha_0 = 1$ (the latter is simply a normalization). Then the feasible sets of (1.7) and (1.12) are identical and non-empty if and only if the Stieltjes condition holds. Csiszár [Csi75] shows that if a maximum entropy solution to (1.12) exists then it must be of the form $f(x) = \exp(-\sum_{j=0}^d \beta_j x^j)$ for some $\beta \in \mathbb{R}^{d+1}$. Moreover, from any solution β to the system of equations

$$\int_0^\infty x^j \exp\left(-\sum_{\ell=0}^d \beta_\ell x^\ell\right) dx = \alpha_j \text{ for } j = 0, \dots, d \quad (1.13)$$

we can generate an optimal solution to (1.12) via $f(x) = \exp\left(-\sum_{j=0}^d \beta_j x^j\right)$. We will show below in Lemma 3.6 that from any solution β to the system (1.8) we can also generate an optimal solution to (1.7).

On the other hand, it can happen that α satisfies the Stieltjes condition for feasibility but no solution to (1.13) exists [Jun00, Tag03]. A simple example is $\alpha = (\alpha_0, \alpha_1, \alpha_2) = (1, 1, 3)$. This set of moments is feasible for the truncated Stieltjes problem since the matrices $\{\{1, 1\}, \{1, 3\}\}$ and $\{\{1\}\}$ have positive determinants. In this case there is a least upper bound to the maximization problem but it is not achieved by any density function.

Similarly, in the case of the maximum geometric entropy moment problem there are vectors α feasible for the Stieltjes moment problem that nevertheless do not satisfy Assumption 1 and thus do not have an optimal solution to (1.7). It is straightforward to generate such an α . First pick $\beta = (\beta_0, \beta_1, \dots, \beta_d)$ so that $\sum_{j=0}^d x^j \beta_j$ is positive on $(0, \infty)$. Let $\alpha'_j, j = 0, \dots, d+1$ be given by

$$\alpha'_j = \int_0^\infty \frac{x^j}{\exp\left(\sum_{\ell=0}^d x^\ell \beta_\ell\right) - 1}.$$

Then let $\alpha_j = \alpha'_j$ for $j = 1, \dots, d$ and $\alpha_{d+1} = \alpha'_{d+1} + \varepsilon$ where $\varepsilon > 0$ is chosen small enough so that the corresponding Hankel matrices have positive determinant (this mirrors the construction in [Jun00]). This results in a feasible optimization problem without an optimal solution. Such vectors α are not covered by our results and we ask if it is still possible to determine the asymptotics of $\tilde{p}_n(\alpha)$.

Question 1.5. Fix J and suppose α is feasible for the Stieltjes moment problem but no optimal solution to (1.7) exists. What are the asymptotics of $\tilde{p}_n(\alpha)$ as $n \rightarrow \infty$?

One can also ask a computational question: given α satisfying Assumption 1, can we efficiently compute the corresponding β and thus the growth constant $M(\alpha)$? Again essentially all of the work devoted to solving the analogous maximum entropy moment problem can be applied here, since both objective functions are strictly concave. We refer the reader to [Las10, Chapter 12].

1.6. Organization and notation

In Section 3 we discuss the connection between maximum entropy distributions and counting and prove a version of the exact formula (1.3). In Section 4 we compute the asymptotics of $e^{H(\mu)}$ by comparing a discrete optimization problem to the continuous optimization problem (1.7). In Section 5 we prove Theorem 1.2 which shows the existence of a limit shape. In Section 6 we prove a multivariate local central limit theorem to estimate the factor $\mu(\mathcal{P}(\mathbb{N}))$. Finally in Section 7 we indicate some possible extensions and future directions related to the results and methods of this paper.

All logarithms in this paper are base e . The Shannon entropy of a discrete random variable X with probability mass function f_X is $H(X) = -\sum_x f_X(x) \log f_X(x)$. A geometric random variable X with parameter $p \in (0, 1)$ has probability mass function $f_X(k) = p(1-p)^k$ for $k \geq 0$. Its mean is $\eta = \frac{1-p}{p}$ and its entropy is $(\eta + 1) \log(\eta + 1) - \eta \log \eta$. We let \mathcal{P} denote the set of all partitions, which we identify with the set $L_0(\mathbb{N}) := \bigoplus_{j=1}^\infty \mathbb{N}_0$, i.e. the set of sequences in $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ that converge to 0 (in particular, \mathcal{P} is a countable set). We let $\mathbb{R}_+ = (0, \infty)$. We will use the convention that bold symbols $(\mathbb{N}, \alpha, \beta, \dots)$ denote vectors indexed by a profile set J or by the integers $\{1, \dots, d\}$.

2. An extended example: the Hardy–Ramanujan formula

To give a feel for the overall approach, we show how to derive (1.1) using the method outlined in Section 1.1. Let Ω denote the set of all integer partitions λ , i.e. all finite multisets of natural numbers; and for each n , let $\mathcal{P}_n = \{\lambda \in \Omega : \sum_{x \in \lambda} x = n\}$ denote the set of partitions of n .

Now consider all probability measures μ on Ω that satisfy

$$\mathbb{E}_{\lambda \sim \mu} \sum_{x \in \lambda} x = n, \quad (2.1)$$

i.e. distributions on partitions with expected size n . Define μ_n to be the probability measure on Ω satisfying (2.1) that maximizes entropy among all such measures. We will establish in Section 3 (Lemma 3.2) an exact identity relating μ_n to the size of \mathcal{P}_n :

$$p(n) = |\mathcal{P}_n| = e^{H(\mu_n)} \mu_n(\mathcal{P}_n). \quad (2.2)$$

Thus to deduce the asymptotics of $p(n)$, it will be sufficient to analyze the entropy $H(\mu_n)$ as well as the probability $\mu_n(\mathcal{P}_n)$. To do this we need to understand the maximum entropy measure. In fact, we can describe μ_n explicitly following the two observations made above. Let λ be a random partition chosen according to μ_n and let Y_k denote the number of parts of λ of size k . Then the constraint (2.1) is a linear constraint on the means $(\eta_k)_{k \geq 1}$ of the random variables $(Y_k)_{k \geq 1}$: $\sum_k k \eta_k = n$. To maximize entropy, the Y_k 's must be independent geometric random variables. Then to compute $H(\mu_n)$ we solve the following convex optimization problem with variables $(\eta_k)_{k \geq 1}$.

$$\begin{aligned} H(\mu_n) &= \max \sum_{k \geq 1} G(\eta_k) dx \\ &\text{subject to } \sum_{k \geq 1} k \eta_k = n. \end{aligned} \quad (2.3)$$

To understand the asymptotics of $H(\mu_n)$, we scale by \sqrt{n} and approximate a Riemann sum by an integral to obtain the following continuous convex optimization problem over real-valued functions:

$$\begin{aligned} M &= \max_f \int_0^\infty G(f(x)) dx \\ &\text{subject to } \int_0^\infty x f(x) dx = 1, \end{aligned} \quad (2.4)$$

over all integrable functions $f : [0, \infty) \rightarrow [0, \infty)$. The optimizer $f^*(x) = \frac{1}{e^{\beta x} - 1}$ with $\beta = \frac{\pi}{\sqrt{6}}$ can be found using Lagrange multipliers. This yields $M = \pi \sqrt{\frac{2}{3}}$, the constant in the exponent of the Hardy–Ramanujan formula. The solution to the discrete problem is also given by Lagrange multipliers: for $k \geq 1$, $\eta_k = \frac{1}{e^{\hat{\beta} k} - 1}$, where the Lagrange multiplier $\hat{\beta}$ satisfies

$$\hat{\beta} = \frac{\beta}{\sqrt{n}} - \frac{1}{4n} + O(n^{-3/2}). \quad (2.5)$$

We now can approximate $\mu_n(\mathcal{P}_n)$. With the above description of μ_n in place, this is a question about the probability that a linear combination of independent (but not identically distributed) geometric random variables equals its mean. Using the above expression for η_k and (2.5) we see that

$$\text{Var} \left(\sum_k kY_k \right) \sim n^3 \int_0^1 \frac{x^2 e^{\beta x}}{(e^{\beta x} - 1)^2} dx = n^3 \cdot \frac{2\sqrt{6}}{\pi}. \tag{2.6}$$

After proving a sufficiently strong local central limit theorem (Theorem 6.2), this implies

$$\mu_n(\mathcal{P}_n) \sim \left(2\pi \cdot \text{Var} \left(\sum_k kY_k \right) \right)^{-1/2} \sim (96n^3)^{-1/4}. \tag{2.7}$$

This provides asymptotics for one of the two quantities in (2.2). We now need to estimate the entropy $H(\mu_n)$ of the measure μ_n . By independence, we have $H(\mu_n) = \sum H(Y_k)$; using the expression for $G(\cdot)$ in (1.6) we see that

$$H(\mu_n) = \sum_{k \geq 1} G \left((e^{\hat{\beta}k} - 1)^{-1} \right). \tag{2.8}$$

This is a Riemann sum, and since $\hat{\beta} \sim \beta/\sqrt{n}$, the leading order of this sum is a corresponding integral:

$$\sum_{k \geq 1} G \left(\frac{1}{e^{\hat{\beta}k} - 1} \right) \sim \sqrt{n} \int_0^\infty G \left(\frac{1}{e^{\beta x} - 1} \right) dx \sim \pi \sqrt{\frac{2}{3}} \sqrt{n}. \tag{2.9}$$

Already, this shows a weaker version of (1.1), namely

$$\log p(n) = \pi \sqrt{\frac{2}{3}} \sqrt{n} (1 + o(1)). \tag{2.10}$$

Note that in order to prove (2.10) from (2.2) it was necessary to prove that $\mu(\mathcal{P}_n)$ does not decrease too quickly, as in (2.7).

To upgrade (2.10) up to the asymptotic formula (1.1), we need to upgrade our asymptotic expression (2.9) beyond just the leading term; in particular, we will need asymptotics for $H(\mu_n)$ up to an additive error of $o(1)$. Using more careful Euler–Maclaurin summation (Lemma 4.1) one can show

$$H(\mu_n) = \pi \sqrt{\frac{2}{3}} \sqrt{n} - \frac{1}{4} \log n - \frac{1}{2} \log(2\sqrt{6}) + o(1). \tag{2.11}$$

Plugging in the asymptotic expressions (2.7) and (2.11) into (2.2) proves (1.1). Or in other words, the formula (1.1) arises from three factors: a factor exponential in \sqrt{n} with constant M in the exponent given by a continuous optimization problem; a polynomial factor (with specified constant) arising from a local central limit theorem; and a factor arising from the approximation of a Riemann sum by an integral.

3. Maximum entropy distributions

In this section we derive the maximum entropy distribution on partitions given moment constraints and give an exact formula for $p(\mathbf{N})$ in terms of this distribution.

We first give an elementary and completely general statement connecting counting to maximum entropy.

Lemma 3.1. *Let Ω be a finite set and let $f : \Omega \rightarrow \mathbb{R}^d$. For $\mathbf{B} \in \mathbb{R}^d$ define $\Lambda_{\mathbf{B}} = \{\omega \in \Omega : f(\omega) = \mathbf{B}\}$ and suppose $\Lambda_{\mathbf{B}}$ is finite and non-empty. Let μ be the maximum entropy distribution on Ω so that $\mathbb{E}_{\mu} f = \mathbf{B}$. Then*

$$|\Lambda_{\mathbf{B}}| = e^{H(\mu)} \cdot \mu(\Lambda_{\mathbf{B}}). \quad (3.1)$$

This is inspired by, e.g. [BH10, Theorem 3.1], and is a simple consequence of the form of maximum entropy distributions subject to mean constraints. The usefulness of convex duality in constrained entropy maximization has appeared in several previous works, including [CC75, BTC77, BTTC88, BV04, SV14].

Proof. Let $K \subseteq \mathbb{R}^d$ be the convex hull of the set $\{f(\omega) : \omega \in \Omega\}$. We have $\mathbf{B} \in K$ since $\Lambda_{\mathbf{B}}$ is non-empty. We may assume that \mathbf{B} lies in the relative interior of K ; otherwise we can restrict ourselves to the proper face F of K in which \mathbf{B} lies and consider distributions on $\Omega' = \{\omega \in \Omega : f(\omega) \in F\}$, since any distribution on Ω satisfying $\mathbb{E} f = \mathbf{B}$ must be supported on Ω' . If $F = \{\mathbf{B}\}$ then the lemma follows from the fact that the maximum entropy distribution on a finite set is the uniform distribution.

We can determine the maximum entropy distribution by solving the convex optimization problem

$$\begin{aligned} & \max - \sum_{\omega \in \Omega} \mu(\omega) \log \mu(\omega) \\ & \text{subject to } \sum_{\omega \in \Omega} \mu(\omega) f(\omega) = \mathbf{B} \\ & \sum_{\omega \in \Omega} \mu(\omega) = 1 \\ & \mu(\omega) \geq 0 \forall \omega \in \Omega. \end{aligned}$$

The convex dual to this program is

$$\begin{aligned} & \min \mathbf{b} \cdot \mathbf{B} + \log \sum_{\omega \in \Omega} e^{-\mathbf{b} \cdot f(\omega)} \\ & \text{where } \mathbf{b} \in \mathbb{R}^d. \end{aligned}$$

Our assumption that \mathbf{B} is in the relative interior of K means that the primal problem is strictly feasible; i.e. there exists a strictly positive feasible solution $\mu > 0$. Slater's condition (see e.g. [BV04]) then guarantees strong duality: the optima of the primal and dual are equal. This gives an optimal primal solution

$$\mu(\omega) = \frac{1}{Z} e^{-\mathbf{b} \cdot f(\omega)}$$

where $\mathbf{b} \in \mathbb{R}^d$ and Z are chosen such that $\sum_{\omega} \mu(\omega) = 1$ and $\sum \mu(\omega) f(\omega) = \mathbf{B}$. The normalizing constant Z is called the partition function in statistical mechanics. The existence of such a \mathbf{b} follows from strong duality.

We then compute

$$\begin{aligned} H(\mu) &= - \sum_{\omega \in \Omega} \mu(\omega) \log \mu(\omega) \\ &= \frac{1}{Z} \sum_{\omega \in \Omega} e^{-\mathbf{b} \cdot f(\omega)} (\log Z + \mathbf{b} \cdot f(\omega)) \\ &= \log Z + \mathbf{b} \cdot \mathbf{B}. \end{aligned}$$

On the other hand, $\mu(\Lambda_{\mathbf{B}}) = \frac{1}{Z} |\Lambda_{\mathbf{B}}| e^{-\mathbf{b} \cdot \mathbf{B}}$, and putting these together yields (3.1). □

If the set Ω is countably infinite, a maximum entropy distribution subject to a given mean constraint may not exist. See, e.g. [Csi75, Cen00], for some sufficient conditions. The following lemma will suffice for our application.

Lemma 3.2. *Let Ω be a countably infinite set and let $f : \Omega \rightarrow \mathbb{R}^d$. For $\mathbf{B} \in \mathbb{R}^d$ define $\Lambda_{\mathbf{B}} = \{\omega \in \Omega : f(\omega) = \mathbf{B}\}$ and suppose $\Lambda_{\mathbf{B}}$ is finite and non-empty. Suppose further that there exists some $\mathbf{b} \in \mathbb{R}^d$ so that $Z = \sum_{\omega \in \Omega} e^{-\mathbf{b} \cdot f(\omega)} < \infty$, and with $\mu(\omega) = \frac{1}{Z} e^{-\mathbf{b} \cdot f(\omega)}$, we have $\mathbb{E}_{\mu} f = \mathbf{B}$. Then μ is the maximum entropy distribution on Ω so that $\mathbb{E}_{\mu} f = \mathbf{B}$, and*

$$|\Lambda_{\mathbf{B}}| = e^{H(\mu)} \cdot \mu(\Lambda_{\mathbf{B}}). \tag{3.2}$$

Proof. The fact that μ is the maximum entropy distribution follows from the strict convexity of the entropy function. The calculation of $H(\mu)$ and the verification of (3.2) then follow exactly as in the proof of Lemma 3.1. □

Before applying Lemma 3.2 to our setting, we need a lemma relating Assumption 1 to the existence of a solution to a system of equations.

Lemma 3.3. *Suppose there exists $\beta = (\beta_j)_{j \in J} \in \mathbb{R}^J$ so that*

$$\int_0^{\infty} \frac{x^j}{\exp(\sum_{i \in J} \beta_i x^i) - 1} dx = \alpha_j$$

for all $j \in J$. Then for n sufficiently large, there exists $\hat{\beta} = (\hat{\beta}_j)_{j \in J}$ so that

$$\sum_{k \geq 1} \frac{k^j}{\exp(\sum_{i \in J} \hat{\beta}_i k^i) - 1} = \lfloor \alpha_j n^{(j+1)/2} \rfloor. \tag{3.3}$$

Further, as n tends to infinity $\hat{\beta}_j n^{j/2} \rightarrow \beta_j$ for each $j \in J$.

To prove Lemma 3.3 we need the following basic calculus fact.

Lemma 3.4. *Let $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be continuously differentiable with $g(0) = 0$ and assume $g'(0) = M$ is invertible. Suppose there is a $\delta > 0$ so that for $\|x\| \leq \delta$ we have $\|M^{-1}\| \cdot \|g'(x) - M\| \leq 1/2$. Then for all y with $\|M^{-1}\| \cdot \|y\| \leq \delta/2$ there is some $\|x\| \leq \delta$ so that $g(x) = y$.*

Proof. Set $x_0 = 0$ and $x_k = M^{-1}(Mx_{k-1} + y - g(x_{k-1}))$ for $k \geq 1$. Then we first note that for any x with $\|x\| \leq \delta$ we have

$$\begin{aligned} \|M^{-1}(Mx - g(x))\| &\leq \|M^{-1}\| \cdot \left\| Mx - \int_0^1 g'(tx) \cdot x dt \right\| \\ &\leq \|M^{-1}\| \|x\| \cdot \max_{t \in [0,1]} \|g'(tx)\| \\ &\leq \delta/2. \end{aligned}$$

By induction, we claim that $\|x_k\| \leq \delta$. Indeed

$$\|x_k\| \leq \|M^{-1}(Mx_{k-1} - g(x_{k-1}))\| + \|M^{-1}\| \|y\| \leq \delta.$$

We now want to show that the sequence $\{x_k\}$ is Cauchy. By the mean-value theorem, we have

$$\begin{aligned} \|x_k - x_{k-1}\| &\leq \|M^{-1}\| \|M(x_{k-1} - x_{k-2}) - (g(x_{k-1}) - g(x_{k-2}))\| \\ &= \|M^{-1}\| \left\| M(x_{k-1} - x_{k-2}) - \int_0^1 g'(x_{k-2} + t(x_{k-1} - x_{k-2})) \cdot (x_{k-1} - x_{k-2}) dt \right\| \\ &\leq \|M^{-1}\| \max_{t \in [0,1]} \|g'(x_{k-2} + t(x_{k-1} - x_{k-2})) - M\| \cdot \|x_{k-1} - x_{k-2}\|. \end{aligned}$$

We claim that $\|x_k - x_{k-1}\| \leq \delta 2^{-k+1}$ and prove so by induction. Since $\|x_k\| \leq \delta$ and the ball of radius δ is convex, we have $\|x_{k-2} + t(x_{k-1} - x_{k-2})\| \leq \delta$ and so by the above we may bound

$$\|x_k - x_{k-1}\| \leq \frac{1}{2} \|x_{k-1} - x_{k-2}\|$$

completing the proof that $\|x_k - x_{k-1}\| \leq \delta 2^{-k+1}$.

This shows that $\{x_k\}$ is Cauchy and thus converges to some x_∞ . Taking limits of both sides of the definition of x_k then shows that $g(x_\infty) = y$. \square

Now we prove Lemma 3.3.

Proof of Lemma 3.3. Rewrite our desired equality as

$$f_j(\widehat{\beta}) := n^{-1/2} \sum_{k \geq 1} \frac{(kn^{-1/2})^j}{\exp\left(\sum_{i \in J} (\widehat{\beta}_i n^{i/2}) (kn^{-1/2})^i\right) - 1} = n^{-(j+1)/2} [\alpha_j n^{(j+1)/2}].$$

Note that evaluating at $\widehat{\beta}_0 := (\beta_j n^{-j/2})_{j \in J}$ gives

$$f_j(\widehat{\beta}_0) \rightarrow \alpha_j$$

as $n \rightarrow \infty$. Further, if we define $\mathbf{f} = (f_j)_{j \in J}$ to be a function from $\mathbb{R}^J \rightarrow \mathbb{R}^J$ then observe that as $n \rightarrow \infty$ we have

$$(f'(\widehat{\beta}_0))_{i,j} \rightarrow - \int_0^\infty \frac{x^{j+i} \exp(\sum_{i \in J} \beta_i x^i)}{(\exp(\sum_{i \in J} \beta_i x^i) - 1)^2} dx = -\Sigma_{i,j}.$$

Since Σ is a Gram matrix of linearly independent entries, it is positive definite and thus invertible. This means that we can find a δ so that for n sufficiently large we have $\|f'(\widehat{\beta}_0)^{-1}\| \cdot \|f'(\widehat{\beta}) - f'(\widehat{\beta}_0)\| \leq 1/2$ for $\|\widehat{\beta} - \widehat{\beta}_0\| \leq \delta$. For n sufficiently large we have

$$\|f'(\widehat{\beta}_0)^{-1}\| \cdot \|(n^{-(j+1)/2}[\alpha_j n^{(j+1)/2}])_{j \in J} - (\alpha_j)_{j \in J}\| \leq \delta/2,$$

and so we may apply Lemma 3.4 to find the desired solution. Noting that we may take $\delta \rightarrow 0$ slowly shows convergence. □

As a corollary of Lemmas 3.2 and 3.3 we derive a formula for $p(\mathbf{N})$.

Corollary 3.5. *Let J be a profile set and suppose $\alpha \in \mathbb{R}^J$ satisfies Assumption 1. Let $\mathbf{N} = \mathbf{N}(\alpha, n)$. Then for large enough n ,*

$$p(\mathbf{N}) = e^{H(\mu_n)} \mu_n(\mathcal{P}(\mathbf{N})), \tag{3.4}$$

where μ_n is the maximum entropy distribution on \mathcal{P} so that

$$\mathbb{E}_{\lambda \sim \mu_n} \sum_{x \in \lambda} x^j = \mathbf{N}_j \tag{3.5}$$

for all $j \in J$. In particular, μ_n is the product measure on $\mathbb{N}_0^{\mathbb{N}}$ where the projection μ_n^k to coordinate k is given by a geometric random variable with parameter $p_k := 1 - \exp(-\sum_{j \in J} \widehat{\beta}_j k^j)$ where $\widehat{\beta}$ is the solution to (3.3) guaranteed by Lemma 3.3.

Proof. Recall that the set of all integer partitions, \mathcal{P} , is a countable set. Let $f : \mathcal{P} \rightarrow \mathbb{R}^J$ be defined by $f_j(\lambda) = \sum_{x \in \lambda} x^j$. Since α satisfies Assumption 1, there exists β solving (1.8), and so by Lemma 3.3, there exists $\widehat{\beta}$ solving the system (3.3). Let μ_n be the distribution on $\mathbb{N}_0^{\mathbb{N}}$ described in the last sentence of the statement of the lemma.

For a partition λ , let a_k be the multiplicity of k in λ . We write $f_j(\lambda) = \sum_{k \geq 1} a_k k^j$ and compute

$$\begin{aligned} \log \mu_n(\lambda) &= \sum_{k \geq 1} \log p_k + a_k \log(1 - p_k) \\ &= \sum_{k \geq 1} \log \left(1 - \exp\left(-\sum_{j \in J} \widehat{\beta}_j k^j\right) \right) - \sum_{k \geq 1} a_k \sum_{j \in J} \widehat{\beta}_j k^j \\ &= -\log Z - \sum_{j \in J} \widehat{\beta}_j \sum_{k \geq 1} a_k k^j \end{aligned} \tag{3.6}$$

where we have defined $\log Z := -\sum_{k \geq 1} \log \left(1 - \exp(-\sum_{j \in J} \widehat{\beta}_j k^j) \right)$. Exponentiating (3.6) gives

$$\mu_n(\lambda) = \frac{1}{Z} e^{-\widehat{\beta} \cdot f(\lambda)},$$

as required for Lemma 3.2. Moreover, since f_j is additive over coordinates for each $j \in J$,

$$\begin{aligned} \mathbb{E}_{\lambda \sim \mu_n} [f_j(\lambda)] &= \sum_{k \geq 1} \frac{\exp(-\sum_{\ell \in J} \widehat{\beta}_\ell k^\ell)}{1 - \exp(-\sum_{\ell \in J} \widehat{\beta}_\ell k^\ell)} k^j \\ &= \sum_{k \geq 1} \frac{k^j}{\exp(\sum_{\ell \in J} \widehat{\beta}_\ell k^\ell) - 1} = N_j \end{aligned}$$

by our assumption on $\widehat{\beta}$. Therefore by Lemma 3.2, μ_n is the maximum entropy distribution on \mathcal{P} such that $\mathbb{E}_{\lambda \sim \mu_n} \sum_{x \in \lambda} x^j = N_j$ for $j \in J$ and $p(\mathbf{N}) = e^{H(\mu_n)} \mu_n(\mathcal{P}(\mathbf{N}))$. \square

Finally we determine the optimum and optimizer of the continuous convex program (1.7).

Lemma 3.6. *Suppose α satisfies Assumption 1 and let β be a solution to (1.8). Let*

$$f^*(x) = \frac{1}{\exp(\sum_{j \in J} \beta_j x^j) - 1}.$$

Then f^* is an optimal solution to (1.7), and thus

$$M(\alpha) = \int_0^\infty G(f^*(x)) dx.$$

Proof. This is a consequence of duality in infinite dimensional convex programming (see e.g. [Roc74]). The program (1.7) is an infinite-dimensional convex program since the constraints are linear in $f(x)$ and the objective function $\int_0^\infty G(f(x)) dx$ is strictly concave (this follows from the fact that $G(\eta)$ is a strictly concave function of η). The Lagrangian associated to this program is

$$L(f, y, \beta) = \int_0^\infty G(f(x)) dx - \sum_{j \in J} \beta_j \left(\int x^j f(x) dx - \alpha_j \right).$$

A sufficient condition for optimality of f^* is that f^* is feasible and the function derivative of $L(f, y, \beta)$ with respect to f vanishes at f^* . This is the condition

$$0 = \log \left(1 + \frac{1}{f(x)} \right) - \sum_{j \in J} \beta_j x^j$$

for all $x \in [0, \infty)$. This is satisfied by $f^*(x)$ with β as given by Assumption 1 since $\log \left(1 + \frac{1}{f(x)} \right) = \sum_{j \in J} \beta_j x^j$, and so f^* is an optimal solution. \square

4. Asymptotics of $e^{H(\mu_n)}$

In this section we compute the asymptotics of the first term in the formula (3.4). In what follows we fix the profile set J and $\alpha \in \mathbb{R}_+^J$ satisfying Assumption 1. We let $N = N(\alpha, n)$ and let β be the solution to (1.8) guaranteed by Assumption 1.

Lemma 4.1. *With μ_n defined as in Corollary 3.5, we have*

$$H(\mu_n) = \sqrt{n}M(\alpha) + b_1(J) \log n + c_1(\alpha) + o(1) \tag{4.1}$$

as $n \rightarrow \infty$, where $b_1(J) = -\frac{j_*}{4}$ and

$$c_1(\alpha) = -\frac{j_*}{2} \log(2\pi) + \frac{\mathbf{1}_{j_*=0}}{2} \left(\frac{\beta_0}{e^{\beta_0} - 1} - G\left(\frac{1}{e^{\beta_0} - 1}\right) \right) + \frac{\mathbf{1}_{j_* \geq 1}}{2} \log(\beta_{j_*}).$$

Let $\widehat{\beta}$ be as in Lemma 3.3. For $x > 0$ let $f^*(x) = (\exp(\sum_{j \in J} \beta_j x^j) - 1)^{-1}$ and for $k \geq 1$ let $\widehat{f}(k) = (\exp(\sum_{j \in J} \widehat{\beta}_j k^j) - 1)^{-1}$. Then from Lemma 3.6 and Corollary 3.5, we have

$$M(\alpha) = \int_0^\infty G(f^*(x)) dx,$$

$$H(\mu_n) = \sum_{k \geq 1} G(\widehat{f}(k)).$$

We start by relating the parameters $\widehat{\beta}_j$ to their analogues β_j .

Lemma 4.2. *Define ε via $\widehat{\beta}_j n^{j/2} = \beta_j + \varepsilon_j / \sqrt{n}$. Then*

$$\varepsilon = \Sigma^{-1} \left(-\frac{\mathbf{v}_{j=j_*}}{2} \left(\frac{\mathbf{1}_{j_*=0}}{e^{\beta_0} - 1} + \frac{\mathbf{1}_{j_* \geq 1}}{\beta_{j_*}} \right) \right)^T + O(n^{-1/2})$$

where $\mathbf{v}_{j=j_*}$ is the vector with 1 in the $j = j_*$ coordinate and 0 in all other coordinates.

Proof. First write

$$\begin{aligned} \alpha_j &= n^{-1/2} \sum_{k \geq 1} \frac{(kn^{-1/2})^j}{\exp\left(\sum_{i \in J} (\widehat{\beta}_i n^{i/2})(k/\sqrt{n})^i\right) - 1} \\ &= -\frac{\mathbf{1}_{j=j_*}}{2\sqrt{n}} \left(\left(\frac{\mathbf{1}_{j_*=0}}{e^{\beta_0} - 1} \right) + \frac{\mathbf{1}_{j_* \geq 1}}{\widehat{\beta}_{j_*} n^{j_*/2}} \right) + \int_0^\infty \frac{x^j}{\exp\left(\sum_{i \in J} (\widehat{\beta}_i n^{i/2})x^i\right) - 1} dx + O(1/n). \end{aligned}$$

Taylor expanding the integrand at β_i gives

$$\begin{aligned} & \int_0^\infty \frac{x^j}{\exp\left(\sum_{i \in J} (\widehat{\beta}_i n^{i/2}) x^i\right) - 1} dx \\ &= \alpha_j - \sum_{i \in J} \frac{\varepsilon_i}{\sqrt{n}} \int_0^\infty x^{i+j} \frac{\exp\left(\sum_{k \in J} (\widehat{\beta}_k n^{k/2}) x^k\right)}{\left(\exp\left(\sum_{k \in J} (\widehat{\beta}_k n^{k/2}) x^k\right) - 1\right)^2} dx + O(1/n) \\ &= \alpha_j - \sum_{i \in J} (1 + o(1)) \frac{\varepsilon_i}{\sqrt{n}} \Sigma_{i,j} + O(1/n) \end{aligned}$$

and so

$$\begin{aligned} \sum_{i \in J} (1 + o(1)) \varepsilon_i \Sigma_{i,j} &= -\frac{\mathbf{1}_{j=j_*}}{2} \left(\left(\frac{\mathbf{1}_{j_*=0}}{e^{\widehat{\beta}_0} - 1} \right) + \frac{\mathbf{1}_{j_* \geq 1}}{\widehat{\beta}_{j_*} n^{j_*/2}} \right) + O(n^{-1/2}) \\ &= -\frac{\mathbf{1}_{j=j_*}}{2} \left(\left(\frac{\mathbf{1}_{j_*=0}}{e^{\beta_0} - 1} \right) + \frac{\mathbf{1}_{j_* \geq 1}}{\beta_{j_*}} \right) + O(n^{-1/2}). \end{aligned}$$

Solving for ε completes the Lemma. \square

We need an Euler–Maclaurin summation calculation that we postpone to the Appendix A.

Lemma 4.3. For $\gamma \in \mathbb{R}^J$ with $\sum_{j \in J} \gamma_j t^j \geq 0$ for all $t \geq 0$ and $\gamma_{j_*} > 0$, we have as $t \rightarrow 0^+$,

$$\begin{aligned} \sum_{k \geq 1} G\left(\frac{1}{\exp(\sum_{j \in J} \gamma_j (tk)^j) - 1}\right) &= t^{-1} \int_0^\infty G\left(\frac{1}{\exp(\sum_{j \in J} \gamma_j x^j) - 1}\right) dx - \frac{j_*}{2} \log(2\pi/t) \\ &\quad - \frac{\mathbf{1}_{j_*=0}}{2} G\left(\frac{1}{e^{\gamma_0} - 1}\right) + \frac{\mathbf{1}_{j_* \geq 1}}{2} (\log \gamma_{j_*} - 1) + o(1) \end{aligned}$$

where the error is uniform for γ in a compact set $K \subset \mathbb{R}^J$ satisfying the hypotheses.

Using this we compute the asymptotics of $H(\mu_n)$.

Proof of Lemma 4.1. Write

$$\begin{aligned} H(\mu_n) &= \sum_{k \geq 1} G\left(\frac{1}{\exp\left(\sum_{j \in J} \widehat{\beta}_j k^j\right) - 1}\right) \\ &= \sum_{k \geq 1} G\left(\frac{1}{\exp\left(\sum_{j \in J} (\widehat{\beta}_j n^{j/2}) (kn^{-1/2})^j\right) - 1}\right) \end{aligned}$$

and apply Lemma 4.3 to $\gamma = (\widehat{\beta}_j n^{j/2})_{j \in J}$ with $t = n^{-1/2}$ to obtain

$$\begin{aligned}
 H(\mu_n) &= \sqrt{n} \int_0^\infty G \left(\frac{1}{\exp \left(\sum_{j \in J} (\widehat{\beta}_j n^{j/2}) x^j \right) - 1} \right) dx - \frac{j^*}{2} \log(2\pi n^{1/2}) \\
 &\quad - \frac{\mathbf{1}_{j^*=0}}{2} G \left(\frac{1}{e^{\widehat{\beta}_0} - 1} \right) + \frac{\mathbf{1}_{j^* \geq 1}}{2} \left(\log(\widehat{\beta}_{j^*} n^{j^*/2}) - 1 \right) + o(1).
 \end{aligned} \tag{4.2}$$

Now recall $\widehat{\beta}_j n^{j/2} = \beta_j + \varepsilon_j n^{-1/2}$, then Lemma 4.2 states

$$\varepsilon = \Sigma^{-1} \left(-\frac{\mathbf{v}_{j=j^*}}{2} \left(\frac{\mathbf{1}_{j^*=0}}{e^{\beta_0} - 1} + \frac{\mathbf{1}_{j^* \geq 1}}{\beta_{j^*}} \right) \right)^T + O(n^{-1/2})$$

where $\mathbf{v}_{j=j^*} \in \mathbb{R}^J$ is the vector with a 1 in the $j = j^*$ coordinate and zeros elsewhere, and Σ given in (1.10). Then we have

$$\begin{aligned}
 &\int_0^\infty G \left(\frac{1}{\exp \left(\sum_{j \in J} (\widehat{\beta}_j n^{j/2}) x^j \right) - 1} \right) dx \\
 &= \int_0^\infty G \left(\frac{1}{\exp \left(\sum_{j \in J} (\beta_j + \varepsilon_j n^{-1/2}) x^j \right) - 1} \right) dx
 \end{aligned} \tag{4.3}$$

$$\begin{aligned}
 &= M(\alpha) - n^{-1/2} (\varepsilon \Sigma \beta) + o(n^{-1/2}) \\
 &= M(\alpha) + \frac{1}{2n^{1/2}} \left(\mathbf{1}_{j^*=0} \left(\frac{\beta_0}{e^{\beta_0} - 1} \right) + \mathbf{1}_{j^* \geq 1} \right) + o(n^{-1/2}).
 \end{aligned} \tag{4.4}$$

Putting (4.2) and (4.3) together gives the lemma. □

5. Limit Shapes

In order to show convergence in distribution of $\{\phi_{\lambda,n}\}$ to $\{\phi_\infty\}$ we need to show that for each $0 < t_1 < t_2 < \infty$ and $\varepsilon > 0$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\lambda \sim \mathcal{P}(\mathbf{N}(\alpha, n))} \left(\max_{t \in [t_1, t_2]} |\phi_{\lambda,n} - \phi_\infty| \geq \varepsilon \right) \rightarrow 0 \tag{5.1}$$

where the probability takes λ uniformly from $\mathcal{P}(\mathbf{N}(\alpha, n))$.

The main idea is that if we sample λ according to the maximum entropy measure μ_n and condition on the profile of λ to be $\mathbf{N}(\alpha, n)$, then this is the same as choosing λ uniformly from $\mathcal{P}(\mathbf{N}(\alpha, n))$. Lemma 6.1 will provide a polynomial lower bound for $\mathbb{P}_{\lambda \sim \mu_n}(\lambda \in \mathcal{P}(\mathbf{N}(\alpha, n)))$. As such, it is sufficient to show that if λ is chosen according μ_n , then the convergence in (5.1) is exponentially small in \sqrt{n} .

Adopting the notation of Theorem 6.2, let $\{X_k\}_{k \geq 1}$ be independent geometric random variables where X_k has mean $(\exp(\sum_{j \in J} \widehat{\beta}_j k^j) - 1)^{-1}$ and recall that this is the number of parts of size k of a partition λ chosen according to the maximum entropy measure μ_n .

The core of the proof is to show that the heuristic (1.11) holds on the scale of \sqrt{n} holds:

Lemma 5.1. *Let $K \subset (0, \infty)$ be a compact set. Then for each $t \in K$ and $\varepsilon > 0$ there exists a constant $C_{K,\varepsilon}$ so that*

$$\mathbb{P} \left(\left| \sum_{k \geq t\sqrt{n}} X_k - \int_t^\infty \frac{1}{\exp\left(\sum_{j \in J} \beta_j x^j\right) - 1} dx \right| \geq \varepsilon\sqrt{n} \right) \leq e^{-C_{K,\varepsilon}\sqrt{n}}.$$

Proof. Note that for n sufficiently large (uniformly in K) we have

$$\begin{aligned} & \left| \sum_{k \geq t\sqrt{n}} \mathbb{E}X_k - \int_t^\infty \frac{1}{\exp\left(\sum_{j \in J} \beta_j x^j\right) - 1} dx \right| \\ &= \left| \sum_{k \geq t\sqrt{n}} \frac{1}{\exp\left(\sum_{j \in J} \hat{\beta}_j k^j\right) - 1} - \int_t^\infty \frac{1}{\exp\left(\sum_{j \in J} \beta_j x^j\right) - 1} dx \right| \\ &\leq \varepsilon\sqrt{n}/2 \end{aligned}$$

and so it is sufficient to show

$$\mathbb{P} \left(\left| \sum_{k \geq t\sqrt{n}} (X_k - \mathbb{E}X_k) \right| \geq \varepsilon\sqrt{n}/2 \right) \leq \exp(-C_{K,\varepsilon}\sqrt{n}). \quad (5.2)$$

This will follow from a standard Chernoff bound argument. By Lemma 6.18, for each $k \geq t\sqrt{n}$ there are constants $C, c > 0$ (depending on K) so that for $|\theta| \leq c$ we have

$$\mathbb{E} \exp(\theta(X_k - \mathbb{E}X_k)) \leq \exp(C\theta^2 \text{Var}(X_k)).$$

This implies for $|\theta| \leq c$ we have

$$\mathbb{E} \exp \left(\theta \sum_{k \geq t\sqrt{n}} (X_k - \mathbb{E}X_k) \right) \leq \exp \left(C\theta^2 \sum_{k \geq t\sqrt{n}} \text{Var}(X_k) \right) \leq \exp(C'\theta^2\sqrt{n}).$$

Applying this bound along with Markov's inequality for $\theta \in (0, c]$ to be chosen small enough with respect to ε

$$\begin{aligned} \mathbb{P} \left(\sum_{k \geq t\sqrt{n}} (X_k - \mathbb{E}X_k) \geq \varepsilon\sqrt{n}/2 \right) &= \mathbb{P} \left(\exp \left(\theta \sum_{k \geq t\sqrt{n}} (X_k - \mathbb{E}X_k) \right) \geq \exp(\varepsilon\theta\sqrt{n}/2) \right) \\ &\leq \exp(C'\theta^2\sqrt{n} - \varepsilon\theta\sqrt{n}/2) \\ &\leq \exp(-C_\varepsilon\sqrt{n}). \end{aligned}$$

The corresponding lower bound follows by an identical argument. \square

We now show that (5.1) holds for λ chosen according to μ_n :

Corollary 5.2. For $0 < t_1 < t_2 < \infty$ and $\varepsilon > 0$, there is a constant $c = c(t_1, t_2, \varepsilon) > 0$ so that

$$\mathbb{P}_{\lambda \sim \mu_n} \left(\max_{t \in [t_1, t_2]} |\phi_{\lambda, n}(t) - \phi_\infty(t)| \geq \varepsilon \right) \leq e^{-c\sqrt{n}}.$$

Proof. For each $t \in [t_1, t_2]$ so that $t\sqrt{n} \in \mathbb{Z}$, apply Lemma 5.1 to see

$$|\phi_{\lambda, n}(t) - \phi_\infty(t)| \geq \varepsilon \leq e^{-c'\sqrt{n}}.$$

Since there are $O(\sqrt{n})$ many such t , we may union bound

$$\mathbb{P}_{\lambda \sim \mu_n} \left(\max_{t \in [t_1, t_2]} |\phi_{\lambda, n}(t) - \phi_\infty(t)| \geq \varepsilon \right) \leq C\sqrt{n}e^{-c'\sqrt{n}} \leq e^{-c''\sqrt{n}}.$$

□

Proof of Theorem 1.2. Recall that in order to show Theorem 1.2, it is sufficient to show (5.1); let A denote the event in (5.1) and note

$$\mathbb{P}_{\lambda \sim \mathcal{P}(\mathbf{N})}(A) = \mathbb{P}_{\lambda \sim \mu_n}(A \mid \lambda \in \mathcal{P}(\mathbf{N})) \leq \mu_n(\mathcal{P}(\mathbf{N}))^{-1} \mathbb{P}_{\lambda \sim \mu_n}(A) = O(n^C e^{-c\sqrt{n}}) = o(1).$$

□

6. The Local CLT

In the previous two sections, we gave an exact formula (3.4) for $p(\mathbf{N})$ and computed the asymptotics of its first factor, $e^{H(\mu_n)}$. In this section, we compute the asymptotics of the second factor, $\mu_n(\mathcal{P}(\mathbf{N}))$, the probability a random partition generated by the entropy-maximizing distribution μ_n has profile \mathbf{N} .

Lemma 6.1. For all α satisfying Assumption 1 and n -feasible,

$$\mu_n(\mathcal{P}(\mathbf{N})) = (1 + o(1)) \cdot \frac{|\mathcal{Q}_J|}{(2\pi)^{|J|/2} \det(\Sigma)^{1/2}} \cdot n^{-\sum_{j \in J} (j/2 + 1/4)} \tag{6.1}$$

as $n \rightarrow \infty$, where $\mathbf{N} = (\lfloor \alpha_j n^{(j+1)/2} \rfloor)_{j \in J}$. Further, the error is uniform for α varying in a compact set of \mathbb{R}^J satisfying Assumption 1.

To prove Lemma 6.1 we prove a multivariate local central limit theorem. While the proof of this local central limit theorem can be easily adapted to a broader setting—such as for random variables other than geometrics—we state it for the random variables of interest.

Theorem 6.2. Fix a compact set $K \subset \mathbb{R}^J$. Suppose $\{Y_k\}_{k \geq 1}$ are independent geometric random variables with parameters p_k where $p_k = 1 - \exp\left(-\sum_{j \in J} \hat{\beta}_j k^j\right)$ and $\hat{\beta} = (\hat{\beta}_j)_{j \in J}$ with $(\hat{\beta}_j n^{j/2})_{j \in J} \in K$ for all n . Define

$$\mathbf{X} = \left(\sum_{k \geq 1} Y_k k^j \right)_{j \in J}$$

and set $S = \text{Cov}(\mathbf{X})$. Then

$$\sup_{\mathbf{a}} \left| \mathbb{P}(\mathbf{X} = \mathbf{a}) - \frac{\mathbf{1}_{\mathbf{a} \in \text{NT}} |\mathcal{Q}_J|}{\sqrt{(2\pi)^{|J|} \det(S)}} \exp\left(-\frac{1}{2}(\mathbf{a} - \mathbb{E}\mathbf{X})^T (S)^{-1} (\mathbf{a} - \mathbb{E}\mathbf{X})\right) \right| \\ = o(\det(S)^{-1/2}) = o(n^{-\sum_{j \in J} (j/2+1/4)})$$

where the error term depends only on K and J .

Before discussing the proof of this theorem, we will attempt to give some intuition about the statement itself. First, we may think of \mathbf{X} as the profile of a random partition sampled from the maximum entropy distribution μ_n (where Y_k is the number of times k appears as a part in the random partition). Our goal in this section is to estimate the probability that \mathbf{X} is equal to its mean, $\mathbb{E}\mathbf{X} = \mathbf{N}$ (Lemma 6.1). In fact, the theorem above estimates the probability mass function $\mathbb{P}(\mathbf{X} = \mathbf{a})$ for *each* possible profile \mathbf{a} , and not only for $\mathbf{a} = \mathbf{N}$ (although the multiplicative error given by this estimate is large when \mathbf{a} is far from \mathbf{N}). We recover Lemma 6.1 by taking $\mathbf{a} = \mathbf{N}$ and noting $\det(S) \sim n^{\sum_{j \in J} (j/2+1/4)} \det(\Sigma)$ as $n \rightarrow \infty$, as detailed below.

The conclusion of the theorem approximates the probability mass function $\mathbb{P}(\mathbf{X} = \mathbf{a})$ in terms of the density of a multivariate Gaussian. Indeed, the expression above is very nearly the density of a $|J|$ -dimensional Gaussian with mean $\mathbb{E}\mathbf{X}$ and covariance matrix S ; the only difference is the factor $\mathbf{1}_{\mathbf{a} \in \text{NT}} |\mathcal{Q}_J|$.

We now explain the factor $\mathbf{1}_{\mathbf{a} \in \text{NT}} |\mathcal{Q}_J|$. The random variable \mathbf{X} is constrained by number-theoretic identities. For instance, for any integer k we have $k^2 \equiv k \pmod{2}$ and so if $1, 2 \in J$ then we must have that $X_1 \equiv X_2 \pmod{2}$. The set NT describes these constraints. While the vector \mathbf{X} is close to a Gaussian when centered and scaled, its support is a subset of this smaller set NT; this means that the atoms of \mathbf{X} must be assigned larger probability by some quantity roughly reflecting the density of the set NT in \mathbb{Z}^J . This density is precisely $1/|\mathcal{Q}_J|$, thus giving the factor $|\mathcal{Q}_J|$.

We will show that the rescaled random variable defined via

$$\widehat{\mathbf{X}} = \left(n^{-j/2-1/4} \sum_{k \geq 1} (Y_k - \mathbb{E}Y_k) k^j \right)_{j \in J}$$

converges in distribution to a centered multivariate Gaussian provided we have $(\beta_j n^{j/2})_{j \in J} \rightarrow \boldsymbol{\alpha}$. This is a consequence of (6.9).

Before proving Theorem 6.2, we show that Lemma 6.1 follows from Theorem 6.2.

Proof of Lemma 6.1. Apply Theorem 6.2 with the $\widehat{\boldsymbol{\beta}}$ guaranteed by Lemma 3.3. By hypothesis, we have $\mathbf{N} = \mathbb{E}\mathbf{X}$. Taking $\mathbf{a} = \mathbf{N}$ then yields

$$\mu_n(\mathcal{P}(\mathbf{N})) = \mathbb{P}(\mathbf{X} = \mathbf{N}) = (1 + o(1)) \frac{|\mathcal{Q}_J|}{(2\pi)^{|J|/2} \det S^{1/2}}.$$

since α is n -feasible. Recalling the definition of Σ from (1.10) and that $\widehat{\beta}_j n^{j/2} \rightarrow \beta_j$ for each j , the computation

$$\begin{aligned} S_{i,j} &= \sum_{k \geq 0} k^{i+j} \frac{\exp\left(\sum_{\ell \in J} \widehat{\beta}_\ell k^\ell\right)}{\left(\exp\left(\sum_{\ell \in J} \widehat{\beta}_\ell k^\ell\right) - 1\right)^2} \\ &= n^{(i+j+1)/2} \left(\frac{1}{\sqrt{n}} \sum_{k \geq 0} (k/\sqrt{n})^{i+j} \frac{\exp\left(\sum_{\ell \in J} \widehat{\beta}_\ell n^{\ell/2} (k/\sqrt{n})^\ell\right)}{\left(\exp\left(\sum_{\ell \in J} \widehat{\beta}_\ell n^{\ell/2} (k/\sqrt{n})^\ell\right) - 1\right)^2} \right) \\ &= n^{(i+j+1)/2} \Sigma_{i,j} (1 + o(1)) \end{aligned}$$

shows $\det S = n^{\sum_{j \in J} (j+1/2)} \det \Sigma (1 + o(1))$, which completes the proof. □

6.1. Outline and Preliminaries

Theorem 6.2 is proved via Fourier analysis. Specifically, for $\mathbf{t} \in \mathbb{R}^J$ and $x \in \mathbb{R}$, define $Q_{\mathbf{t}}(x) := \sum_{j \in J} t_j x^j$ and set $\varphi_k(t) := \mathbb{E} e^{itY_k}$.

The characteristic function $\varphi(\mathbf{t}) := \mathbb{E} e^{i\langle \mathbf{t}, \mathbf{X} \rangle}$ may then be written as

$$\varphi(2\pi\mathbf{t}) = \prod_{k \geq 1} \varphi_k(2\pi Q_{\mathbf{t}}(k)). \tag{6.2}$$

Fourier inversion (e.g. [Fel71, Theorem XV.3.4]) gives

$$\mathbb{P}(\mathbf{X} = \mathbf{a}) = \int_{(-1/2, 1/2]^J} \varphi(2\pi\mathbf{t}) e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} dt. \tag{6.3}$$

Theorem 6.2 will be proven by analyzing the integral in (6.3). Our method is a variant on the Hardy–Littlewood circle method. The set $(-1/2, 1/2]^J$ is broken up into *major* and *minor* arcs: the major arcs are sets of small volume that contribute the bulk of the mass of the integral in (6.3) while the minor arcs are the rest.

When proving a local central limit theorem, it is often the case that the only major arc is a neighborhood of $\mathbf{t} = 0$. In our problem, however, this does not occur; the number-theoretic obstructions force $\mathbb{P}(\mathbf{X} = \mathbf{a}) = 0$ for $\mathbf{a} \notin \text{NT}$. On the Fourier side, this means that there are multiple major arcs, and in the case of $\mathbf{a} \notin \text{NT}$, the integral over them cancels out.

Before beginning the analysis, we give an outline of the proof of Theorem 6.2: the general flow is that the size of the set we are integrating over decreases as the proof goes on. First, for each $\delta > 0$, we define the following small neighborhood around 0:

$$U = U(\delta) = \{\mathbf{t} \in (-1/2, 1/2]^J : |t_j| \leq \delta n^{-j/2} \text{ for all } j \in J\}.$$

We then define

$$R(\delta) = (-1/2, 1/2]^J \setminus \left(\bigcup_{q \in \mathcal{Q}_J} (q + U) \right),$$

where for each polynomial $q(z) \in \sum q_j z^j \in \mathcal{Q}_J$, we interpret $q = (q_j)_{j \in J}$ as a vector in $(-1/2, 1/2]^J$. We show that for each $R(\delta)$, the integral on $R(\delta)$ is exponentially small. In the language of the circle method, this states that the major arcs are exactly the neighborhoods $q + U$ for $q \in \mathcal{Q}_J$. This is carried out in Section 6.2, and a more detailed summary is given there, including an intuitive description of how the polynomials in \mathcal{Q}_J arise.

Next, in Section 6.3 we show that the integral over each neighborhood $q + U$ is equal provided $\mathbf{a} \in \text{NT}$, and so we may combine the $|\mathcal{Q}_J|$ -many integrals over $\bigcup_{q \in \mathcal{Q}_J} (q + U)$ into $|\mathcal{Q}_J|$ times the integral over U .

Section 6.4 compares the integral of our characteristic function over U to the integral over \mathbb{R}^d of the characteristic function of the corresponding Gaussian by showing upper and lower bounds on the matrix S when viewed as a quadratic form (Lemma 6.16).

Section 6.5 reduces our integral even further to the set $V = \{\mathbf{t} : |t_j| \leq (\log n)n^{-j/2-1/4}\}$ by comparing the characteristic function of each geometric variable of bounded mean to the characteristic function of the corresponding Gaussian (Lemma 6.18); while not all of our geometric variables have bounded mean, the bulk of the contribution to the covariance matrix S comes from the parameters Y_k of bounded mean (Lemma 6.17), which is sufficient for this purpose.

Finally, 6.6 evaluates the integral over V ; this is equivalent to an ordinary multivariate central limit theorem with more careful tracking of error terms.

6.2. Bounding the minor arcs: reducing to a neighborhood of \mathcal{Q}_J

Recall that our goal is to estimate the following integral

$$\mathbb{P}(\mathbf{X} = \mathbf{a}) = \int_{(-1/2, 1/2]^J} \left(\prod_{k \geq 1} \varphi_k(2\pi Q_{\mathbf{t}}(k)) \right) e^{-i(2\pi \mathbf{t}, \mathbf{a})} d\mathbf{t}$$

(obtained by combining (6.2) and (6.3)). To help identify the major arcs, we will see in Lemma 6.5 we have that if t is not close to an integer, then $|\varphi_k(2\pi t)|$ is uniformly bounded away from 1. This shows that if there are many integer values k for which the polynomial $Q_{\mathbf{t}}(k)$ is not near an integer, then the integrand above is small.

This means that we have to understand when the polynomial $Q_{\mathbf{t}}(x) = \sum_{j \in J} x^j t_j$ is close to an integer-valued polynomial, i.e. a polynomial q so that $q(\mathbb{Z}) \subset \mathbb{Z}$. Perhaps surprisingly, there are many such polynomials, even if we omit those with integer coefficients; as an example, the binomial coefficients $\binom{z}{k}$ are integer-valued polynomials but do not have integer coefficients. Motivated by this, for a given J , recall that we define

$$\mathcal{Q}_J = \left\{ \sum_{j \in J} t_j x^j : \sum_{j \in J} t_j m^j \in \mathbb{Z} \text{ for all } m \in \mathbb{Z}, t_j \in (-1/2, 1/2] \right\}$$

to be the set of integer-valued polynomials of interest.

Pólya [Pól15] showed that all integer-valued polynomials are integer linear combinations of binomial coefficients; this may be proved by induction on the degree and examining finite differences. This shows, for instance, that $|\mathcal{Q}_J|$ is finite. To better understand \mathcal{Q}_J , we look at an extreme case:

Lemma 6.3. For each $d \geq 1$, $|\mathcal{Q}_{[d]}| = \prod_{j=1}^d (j!)$.

Proof. Since every element of \mathbb{R}/\mathbb{Z} has a unique representative in $(-1/2, 1/2]$, it is sufficient to count the number of polynomials of degree at most d in \mathbb{R}/\mathbb{Z} that are integer-valued. Each such equivalent class is equal to

$$\sum_{j=1}^d m_j \binom{z}{j}$$

for some choice of integers m_j . For each selection of (m_1, \dots, m_d) with $1 \leq m_j \leq j!$ we obtain a distinct representative. Conversely, each integer polynomial may be written in the above form for *some* integers (m_j) ; further, for two integers n_1 and n_2 , we have $n_1 \binom{z}{j} \equiv n_2 \binom{z}{j}$ as polynomials with coefficients in \mathbb{R}/\mathbb{Z} if and only if $n_1 \equiv n_2 \pmod{j!}$. This means that we may uniquely choose the representatives (m_j) to satisfy $1 \leq m_j \leq j!$ for each j . \square

Now, recall our definitions of U and R : for $\delta > 0$, define

$$U = U(\delta) = \{\mathbf{t} \in (-1/2, 1/2]^J : |t_j| \leq \delta n^{-j/2} \text{ for all } j \in J\}$$

and

$$R(\delta) = (-1/2, 1/2]^J \setminus \left(\bigcup_{q \in \mathcal{Q}_J} (q + U) \right).$$

Thus $R(\delta)$ is the set of points that are far from the coefficients of any integer-valued polynomial. Our goal for this section is to show that the contribution of $R(\delta)$ is negligible for *any* choice of $\delta > 0$.

Lemma 6.4. In the context of Theorem 6.2, for each $\delta > 0$ there exists a constant $c = c(\delta, K)$ so that

$$\int_{R(\delta)} |\varphi(2\pi\mathbf{t})| d\mathbf{t} = O(e^{-c\sqrt{n}}).$$

Lemma 6.4 will be accomplished by a pointwise bound on $|\varphi(2\pi\mathbf{t})|$ on the set $R(\delta)$. In light of the infinite product in (6.2), in order to show that $|\varphi(2\pi\mathbf{t})|$ is exponentially small in \sqrt{n} , it is enough to show that on the order of \sqrt{n} many $|\varphi_k(2\pi Q_t(k))|$ are uniformly less than 1. The following elementary fact is a step in this direction. For a real number $t \in \mathbb{R}$ define $\|t\|_{\mathbb{R}/\mathbb{Z}} = \min_{z \in \mathbb{Z}} |t - z|$ to be the distance to the nearest integer. We want to show that for if $\|t\|_{\mathbb{R}/\mathbb{Z}}$ is large, then $|\mathbb{E}e^{itY}|$ is bounded away from 1.

Lemma 6.5. Fix $\varepsilon > 0$, and let $p \in [\varepsilon, 1 - \varepsilon]$. Suppose Y is a geometric variable with parameter p . Then for each $\delta_1 > 0$ there exists a $\delta_2 > 0$ so that if $\|t\|_{\mathbb{R}/\mathbb{Z}} \geq \delta_1$ then $|\mathbb{E}e^{i2\pi tY}| \leq 1 - \delta_2$.

Proof. Since Y is integer-valued, we may assume without loss of generality that $t \in (-1/2, 1/2]$. In each case we may uniformly bound the modulus of the characteristic function using compactness and continuity. \square

From Lemma 6.5 we extract the following simple consequence, which is the engine behind the proof of Lemma 6.4.

Lemma 6.6. *In the context of Theorem 6.2, for each $\varepsilon_1, \varepsilon_2 > 0$ there exists a $c > 0$ so that the following holds: if $\mathbf{t} \in \mathbb{R}^J$ satisfies*

$$\frac{|\{m \in [\sqrt{n}] : \|Q_{\mathbf{t}}(m)\|_{\mathbb{R}/\mathbb{Z}} \geq \varepsilon_1\}|}{\sqrt{n}} \geq \varepsilon_2$$

then

$$|\varphi(2\pi\mathbf{t})| \leq e^{-c\sqrt{n}}$$

where c may be chosen uniformly depending only on K, ε_1 and ε_2 .

Proof. There exists an $\varepsilon_3 > 0$ so that for $k \in [\varepsilon_2\sqrt{n}/2, \sqrt{n}]$, the geometric parameters p_k lie in the interval $[\varepsilon_3, 1 - \varepsilon_3]$. Thus, for at least $\varepsilon_2\sqrt{n}/2$ values of $k \in [\varepsilon_2\sqrt{n}/2, \sqrt{n}]$ we have

$$|\varphi_k(2\pi Q_{\mathbf{t}}(k))| \leq 1 - \varepsilon_4$$

for some $\varepsilon_4 > 0$ depending only on ε_1 and ε_3 . We then may bound

$$|\varphi(2\pi\mathbf{t})| \leq \prod_{k=\varepsilon_2\sqrt{n}/2}^{\sqrt{n}} |\varphi_k(2\pi Q_{\mathbf{t}}(k))| \leq (1 - \varepsilon_4)^{\varepsilon_2\sqrt{n}/2} \leq e^{-\varepsilon_2\varepsilon_4\sqrt{n}/2}.$$

□

We now need a structural result which will say that either $\|Q_{\mathbf{t}}(x)\|_{\mathbb{R}/\mathbb{Z}}$ is either bounded below quite often, or $Q_{\mathbf{t}}$ is close to an element of \mathcal{Q}_J . A quantitative equidistribution theorem of Green and Tao [GT12, Proposition 4.3] will make explicit that these are the only two cases.

Theorem 6.7 (Green–Tao). *Let $d \geq 0$ and suppose that g is a polynomial with real coefficients of degree d . Suppose that $\delta \in (0, 1/2)$. Then either $(g(x) \bmod \mathbb{Z})_{x \in [N]}$ is δ -equidistributed or else there is an integer k satisfying $1 \leq k \ll \delta^{-O_d(1)}$ so that $\|kg \bmod \mathbb{Z}\|_{C^\infty[N]} \ll \delta^{-O_d(1)}$.*

Some definitions are in order.

Definition 6.8. Let g be a polynomial of degree d . Then there exist unique s_j so that

$$g(m) = s_0 + s_1 \binom{m}{1} + \cdots + s_d \binom{m}{d}$$

for each m . Define

$$\|g\|_{C^\infty(N)} := \sup_{1 \leq j \leq d} N^j \|s_j\|_{\mathbb{R}/\mathbb{Z}}$$

where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ is the nearest distance from x to an integer.

Further, a sequence $\{g(m)\}_{m \in [N]}$ is δ -equidistributed if for all Lipschitz functions $F : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ and arithmetic progressions $P \subset [N]$ with $|P| \geq \delta N$ we have

$$\left| \frac{1}{|P|} \sum_{m \in P} F(g(m)) - \int_{\mathbb{R}/\mathbb{Z}} F(x) dx \right| \leq \delta \|F\|_{\text{Lip}}.$$

Theorem 6.7 is proven via an effective version of Weyl’s equidistribution theorem together with iterating the van der Corput difference trick. Anticipating an application of Theorem 6.7, we first translate the two possibilities of its dichotomy into our setting, beginning with the structured case:

Lemma 6.9. *Let $g(z) = t_0 + t_1z + \dots + t_dz^d$ be of degree d and have coefficients in $(-k/2, k/2]$. There exists a constant C so that if $\|g\|_{C^\infty[N]} \leq R$ then there is an integer-valued polynomial $q(z) = \sum q_jz^j$ so that $\|t_j - q_j\|_{\mathbb{R}/(k\mathbb{Z})} \leq CRN^{-j}$ for all $j \in [d]$.*

Proof. By assumption we have $N^j\|s_j\|_{\mathbb{R}/\mathbb{Z}} \leq R$ for all $1 \leq j \leq d$ where s_j is defined by $g(n) = \sum s_i \binom{n}{i}$. Thus there are integers m_j so that $s_i = m_j + \varepsilon_j$ where $|\varepsilon_j| \leq N^{-j}R$. Define the linear transformation T to be the map that takes as input (t_0, \dots, t_d) and outputs (s_0, \dots, s_d) defined by

$$\sum_{j=0}^d t_jz^j = \sum_{j=0}^d s_j \binom{z}{j}.$$

Since T is linear and invertible, there is a constant C so that $C^{-1} \leq \frac{\|Tx\|}{\|x\|} \leq C$ for all $x \neq 0$. For a given vector of integers $(m_j)_{j=0}^d$, define q_j to be $T^{-1}((m_j)_j)$ so that

$$\sum_{j=0}^d q_jz^j = \sum_{j=0}^d m_j \binom{z}{j},$$

and note that the above is an integer-valued polynomial. By linearity together with the fact that q_j depends only on s_j, s_{j+1}, \dots, s_d , we have that

$$T^{-1}\{x : |x_j - m_j| \leq \varepsilon_j\} \subset \{y : |y_j - q_j| \leq C \sum_{i=j}^d \varepsilon_i\}.$$

By taking the rational numbers q_j modulo $k\mathbb{Z}$ and replacing $|\cdot|$ with $\|\cdot\|_{\mathbb{R}/(k\mathbb{Z})}$, we have Thus, we have that

$$\|t_j - q_j\|_{\mathbb{R}/(k\mathbb{Z})} \leq C \sum_{i=j}^d \varepsilon_i \leq 2CRN^{-j}.$$

□

We now show that the Green–Tao definition of equidistributed is good enough for the case at hand.

Lemma 6.10. *If a sequence $\{g(m)\}_{m \in [N]}$ is $\frac{1}{32}$ -equidistributed then*

$$\frac{|\{m \in [N] : \|g(m)\|_{\mathbb{R}/\mathbb{Z}} \geq 1/4\}|}{N} \geq \frac{1}{8}.$$

Proof. Let F denote the Lipschitz function on \mathbb{R}/\mathbb{Z} that is piecewise linear with $F(0) = F(1/4) = F(3/4) = F(1) = 0$ and $F(1/2) = 1$. Then $\int_{\mathbb{R}/\mathbb{Z}} F(x) dx = \frac{1}{4}$, $\|F\|_{\text{Lip}} = 4$ and $F(x) \leq \mathbf{1}_{\|x\|_{\mathbb{R}/\mathbb{Z}} \geq 1/4}$. By the definition of $1/32$ -equidistributed we have

$$\left| \frac{1}{N} \sum_{m \in [N]} F(g(m)) - \frac{1}{4} \right| \leq \frac{1}{8}$$

implying

$$\frac{1}{N} \sum_{m \in [N]} \mathbf{1}_{\|g(m)\|_{\mathbb{R}/\mathbb{Z}} \geq 1/4} \geq \frac{1}{N} \sum_{m \in [N]} F(g(m)) \geq \frac{1}{8}.$$

□

We are now prepared to make use of Theorem 6.7; rather than proving Lemma 6.4 straight away, it will show that Lemma 6.4 holds for *some* $\delta > 0$ rather than all $\delta > 0$:

Lemma 6.11. *There exist constants $c, C > 0$ so that*

$$\int_{R(C)} |\varphi(2\pi \mathbf{t})| dt = O(e^{-c\sqrt{n}}).$$

Proof. Let $\mathbf{t} \in R(C)$ for $C \geq 1$ to be determined later. Apply Theorem 6.7 and Lemma 6.9 with $\delta = 1/32$ to obtain constants C_2, C_3 so that either $(Q_{\mathbf{t}}(x) \bmod \mathbb{Z})_{x \in [\sqrt{n}]}$ is $1/32$ -equidistributed or there is a k with $1 \leq k \leq C_2$ so that

$$\|kQ_{\mathbf{t}}\|_{C^\infty[\sqrt{n}]} \leq C_3.$$

We have three cases that we address separately:

Case 1: Approximately equidistributed. If $(Q_{\mathbf{t}}(x) \bmod \mathbb{Z})_{x \in [\sqrt{n}]}$ is $1/32$ -equidistributed, then

$$\frac{|\{x \in [\sqrt{n}] : \|Q_{\mathbf{t}}(x)\|_{\mathbb{R}/\mathbb{Z}} \geq 1/4\}|}{\sqrt{n}} \geq \frac{1}{8}$$

by Lemma 6.10. Lemma 6.6 shows $|\varphi(2\pi \mathbf{t})| \leq e^{-c\sqrt{n}}$ for some c depending only on K .

Case 2: Not equidistributed, but $k = 1$. In this case, Lemma 6.9 implies that there is an integer-valued polynomial q so that $\|t_j - q_j\|_{\mathbb{R}/\mathbb{Z}} \leq C_4 n^{-j/2}$ for all $j \in J$. If we require $C \geq C_4$, then this would imply $\mathbf{t} \notin R(C)$, thus completing this case.

Case 3: Not equidistributed, $k > 1$. Suppose that $\|kQ_{\mathbf{t}}\|_{C^\infty[\sqrt{n}]} \leq C_3$ for some $k > 1$ and *not* $k = 1$. Then there is an integer-valued polynomial q so that $Q_{\mathbf{t}}$ is close to q/k ; further, since we know that $\|Q_{\mathbf{t}}\|_{C^\infty[\sqrt{n}]} > C_3$, we have that q is *not* integer-valued.

Since q is integer-valued but q/k is not, there exists some n and integer a so that $q(n) = a$ where $a \not\equiv 0 \pmod{k}$. Further, note that the polynomial $(q(z) - a)/k$ cannot be integer-valued

since it has non-zero constant term. Thus, there must be some m and other value b with $b \not\equiv a \pmod k$ so that $q(m) = b$. Write $q(z)/k = \frac{1}{B} \sum_{j \in J} a_j z^j$ where B and each a_j are integers. Then note that for each integer r with $r \equiv n \pmod B$ we have $q(r)/k \equiv a/k \pmod{\mathbb{Z}}$; similarly, if $r \equiv m \pmod B$ then $q(r)/k \equiv b/k \pmod{\mathbb{Z}}$. Moreover, $\|b - a\|_{\mathbb{R}/(k\mathbb{Z})} \geq \frac{1}{k}$. Therefore for each t_0 , we have that $\max\{\|t_0 + b/k\|_{\mathbb{R}/\mathbb{Z}}, \|t_0 + a/k\|_{\mathbb{R}/\mathbb{Z}}\} \geq \frac{1}{2k} \geq \frac{1}{C_5}$. Therefore, on some set of positive density, Q_t is uniformly bounded away from 0 on the torus \mathbb{R}/\mathbb{Z} . Applying Lemma 6.6 completes the proof. \square

We are now ready to prove the lemma.

Proof of Lemma 6.4. In light of Lemma 6.11, it is sufficient to find a constant $c = c(\delta, K)$ so that

$$\int_{R(\delta) \setminus R(C)} |\varphi(2\pi \mathbf{t})| \, d\mathbf{t} = O(e^{-c\sqrt{n}}).$$

Write

$$R(\delta) \setminus R(C) = \bigcup_{q \in \mathcal{Q}_J} (q + U(C) \setminus U(\delta)).$$

Thus, it is sufficient to show the bound for each set in the above union. Fix some $q \in \mathcal{Q}_J$ and $\mathbf{t} \in q + U(C) \setminus U(\delta)$. In the case where $0 \notin J$, simply set $t_0 = 0$ so that the following two cases make sense.

Case 1: $\|t_0\|_{\mathbb{R}/\mathbb{Z}} \geq \delta$. Then for $m \leq \delta\sqrt{n}/(2dC)$, we have

$$\begin{aligned} \left\| \sum_{j \in J} t_j m^j \right\|_{\mathbb{R}/\mathbb{Z}} &= \left\| t_0 + \sum_{j \in J \setminus \{0\}} (t_j - q_j) m^j \right\|_{\mathbb{R}/\mathbb{Z}} \\ &\geq \delta - \frac{\delta}{2} \\ &= \delta/2. \end{aligned}$$

Applying Lemma 6.6 completes this case.

Case 2: $\|t_0\|_{\mathbb{R}/\mathbb{Z}} < \delta$. The idea will be to look in the limiting setting and use a compactness argument. Define the set of polynomials

$$T := \left\{ \sum_{j \in J} c_j z^j : |c_i| \leq C \text{ for all } i \text{ and } |c_i| \geq \delta \text{ for some } i, |c_0| < \delta \right\}.$$

Note that T is compact; further, for $|z| \leq \delta/(2C)$ and all $p \in T$ we have

$$|p(z)| < \delta + \delta = 2\delta.$$

Since $R(\delta)$ increases as δ decreases, we may assume without loss of generality that $\delta \leq 1/8$; in particular, this implies that $|p(z)| \leq 1/4$ for all $|z| \leq \delta/(2C)$.

For each $p \in T$, p is not identically zero on $[0, \delta/(2C)]$ and so $|p|$ must attain a non-zero maximum $M(p)$. Since $p \mapsto M(p)$ is a continuous function of the coefficients of p and T is compact, we must have that there is a value M_0 so that $M(p) \geq M_0$ for all $p \in S$. Additionally, let $L(p)$ be the length of the maximum interval in $[0, \delta/(2C)]$ on which $|p| \geq M_0/2$. Note that L is non-zero on T and continuous, and so we must have $|L(p)| \geq L_0$ for all $p \in T$.

For any \mathbf{t} in the desired set, find $s_j \in (-1/2, 1/2]$ so that $(t_j - q_j) \equiv s_j \pmod{\mathbb{Z}}$. Then the polynomial

$$p(z) := \sum_{j \in J} (s_j \cdot n^{j/2}) z^j$$

lies in the set T . Thus, there is an interval $I \subset [0, \delta/(2C)]$ of length at least L_0 on which $|p(z)| \geq M_0/2$. For any x for which we have $x/N \in I$ compute

$$\begin{aligned} \left\| \sum_{j \in J} t_j x^j \right\|_{\mathbb{R}/\mathbb{Z}} &= \left\| \sum_{j \in J} s_j x^j \right\|_{\mathbb{R}/\mathbb{Z}} \\ &= \left\| \sum_{j \in J} (s_j n^{j/2}) (x/\sqrt{n})^j \right\|_{\mathbb{R}/\mathbb{Z}} \\ &= \left| \sum_{j \in J} (s_j n^{j/2}) (x/\sqrt{n})^j \right| \\ &\geq M_0/2 \end{aligned}$$

where in the last equality we used the fact that $|p(z)| \leq 1/4$ for $|z| \leq \delta/(2C)$. Since I is an interval, we have that the number of x for which $x/\sqrt{n} \in I$ is at least $\sqrt{n} \cdot |I| - 1$. Since $|I|$ is bounded below by L_0 , the proof is complete after applying Lemma 6.6. \square

6.3. Combining the integrals

The primary goal of this section is to show the following lemma.

Lemma 6.12. *For each $\delta > 0$ there is a constant $c = c(\varepsilon, \delta, d) > 0$ so that*

$$\sup_{\mathbf{a}} \left| \mathbb{P}(\mathbf{X} = \mathbf{a}) - \mathbf{1}_{\mathbf{a} \in \mathcal{NT}} |\mathcal{Q}_J| \int_U \varphi(2\pi \mathbf{t}) e^{-i\langle 2\pi \mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} \right| = O(e^{-c\sqrt{n}}).$$

A first step is a simple lemma that changes coordinates to combine the integrals from the previous section. For a polynomial $q \in \mathcal{Q}_J$, we write $\mathbf{q} \in \mathbb{R}^J$ for the vector of coefficients of q .

Lemma 6.13. *For n sufficiently large we have*

$$\int_{\cup_{q \in \mathcal{Q}_J} (q+U)} \varphi(2\pi \mathbf{t}) e^{-i\langle 2\pi \mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} = \left(\sum_{q \in \mathcal{Q}_J} e^{-i\langle 2\pi \mathbf{q}, \mathbf{a} \rangle} \right) \int_U \varphi(2\pi \mathbf{t}) e^{-i\langle 2\pi \mathbf{t}, \mathbf{a} \rangle} d\mathbf{t}.$$

Proof. For n sufficiently large, we have that the union is in fact disjoint. This means that we may first write

$$\int_{\cup_{q \in \mathcal{Q}_J} (q+U)} \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} = \sum_{q \in \mathcal{Q}_J} \int_{q+U} \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t}.$$

For each $q \in \mathcal{Q}_J$, we have

$$\begin{aligned} \int_{q+U} \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} &= \int_U \varphi(2\pi(\mathbf{t} + \mathbf{q}))e^{-i\langle 2\pi(\mathbf{t} + \mathbf{q}), \mathbf{a} \rangle} d\mathbf{t} \\ &= e^{-i\langle 2\pi\mathbf{q}, \mathbf{a} \rangle} \int_U \prod_{x \geq 1} \varphi_x(2\pi i\langle \mathbf{t} + \mathbf{q}, \mathbf{x} \rangle) e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} \\ &= e^{-i\langle 2\pi\mathbf{q}, \mathbf{a} \rangle} \int_U \prod_{x \geq 1} \varphi_x(2\pi i\langle \mathbf{t}, \mathbf{x} \rangle) e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} \\ &= e^{-i\langle 2\pi\mathbf{q}, \mathbf{a} \rangle} \int_U \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t}. \end{aligned}$$

□

Lemma 6.14. For each vector of integers $\mathbf{m} = (m_j)_{j \in J}$ if $\mathbf{m} \in \text{NT}$ then we have

$$\sum_{q \in \mathcal{Q}_J} e^{-2\pi i\langle \mathbf{q}, \mathbf{m} \rangle} = |\mathcal{Q}_J|. \tag{6.4}$$

Proof. Let $q \in \mathcal{Q}_J$. Then since $\mathbf{m} \in \text{NT}$ we have

$$\sum_{j \in J} q_j m_j \equiv 0 \pmod{\mathbb{Z}}$$

and so

$$\sum_{q \in \mathcal{Q}_J} e^{-2\pi i\langle \mathbf{q}, \mathbf{m} \rangle} = \sum_{q \in \mathcal{Q}_J} 1 = |\mathcal{Q}_J|.$$

□

Proof of Lemma 6.12. Note that if $\mathbf{a} \notin \text{NT}$ then $\mathbb{P}(\mathbf{X} = \mathbf{a}) = 0$. We may thus write

$$\begin{aligned} \mathbb{P}(\mathbf{X} = \mathbf{a}) &= \mathbf{1}_{\mathbf{a} \in \text{NT}} \int_{(-1/2, 1/2]^J} \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} \\ &= \mathbf{1}_{\mathbf{a} \in \text{NT}} \int_{\cup_{q \in \mathcal{Q}_J} (q+U)} \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} + O(e^{-c\sqrt{n}}) \\ &= \mathbf{1}_{\mathbf{a} \in \text{NT}} \left(\sum_{q \in \mathcal{Q}_J} e^{-i\langle 2\pi\mathbf{q}, \mathbf{a} \rangle} \right) \int_U \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} + O(e^{-c\sqrt{n}}) \\ &= \mathbf{1}_{\mathbf{a} \in \text{NT}} |\mathcal{Q}_J| \int_U \varphi(2\pi\mathbf{t})e^{-i\langle 2\pi\mathbf{t}, \mathbf{a} \rangle} d\mathbf{t} + O(e^{-c\sqrt{n}}) \end{aligned}$$

by applying (6.3) and then Lemmas 6.4, 6.13 and 6.14 in succession.

□

6.4. Approximating the integral over U

With Lemma 6.12 established, we turn to the integral over U . First we show that an expression in the statement of Theorem 6.2 can be written as a Gaussian integral.

Lemma 6.15. *For any vectors $\mathbf{t}, \mathbf{a} \in \mathbb{R}^d$ and positive definite $d \times d$ matrix S we have*

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp\left(-i\langle \mathbf{t}, \mathbf{a} - \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) d\mathbf{t} = \frac{\exp\left(-\frac{1}{2}(\mathbf{a} - \mathbb{E}\mathbf{X})^T S^{-1}(\mathbf{a} - \mathbb{E}\mathbf{X})\right)}{\sqrt{(2\pi)^d \det(S)}}. \quad (6.5)$$

Proof. Since S is positive definite, there is an invertible matrix M so that $S = M^T M$. We may thus write

$$\begin{aligned} -i\mathbf{t}^T(\mathbf{a} - \mathbb{E}\mathbf{X}) - \frac{1}{2} \mathbf{t}^T S \mathbf{t} &= -\frac{1}{2}(M\mathbf{t} + i(M^{-1})^T(\mathbf{a} - \mathbb{E}\mathbf{X}))^T (M\mathbf{t} + i(M^{-1})^T(\mathbf{a} - \mathbb{E}\mathbf{X})) \\ &\quad - \frac{1}{2}(\mathbf{a} - \mathbb{E}\mathbf{X})^T (S)^{-1}(\mathbf{a} - \mathbb{E}\mathbf{X}). \end{aligned}$$

Setting $\mathbf{s} = M\mathbf{t} + i(M^{-1})^T(\mathbf{a} - \mathbb{E}\mathbf{X})$ and using Cauchy's integral theorem shows

$$\begin{aligned} \int_{\mathbb{R}^d} \exp\left(-i\langle \mathbf{t}, \mathbf{a} - \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) d\mathbf{t} \\ = \frac{\exp\left(-\frac{1}{2}(\mathbf{a} - \mathbb{E}\mathbf{X})^T (S)^{-1}(\mathbf{a} - \mathbb{E}\mathbf{X})\right)}{\det(M)} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} \|\mathbf{s}\|^2\right) d\mathbf{s}. \end{aligned}$$

Evaluating the Gaussian integral as $(2\pi)^{d/2}$ and recalling $\det(M) = \det(S)^{1/2}$ completes the proof of (6.5). \square

Recall that for a given $\delta > 0$, we have defined $U = U(\delta) = \{\mathbf{t} : |t_j| \leq \delta n^{-j/2}\}$. With Lemmas 6.12 and 6.15 in tow, it is sufficient to show that for some $\delta > 0$, we have

$$\int_{\mathbb{R}^J} \left| \varphi(\mathbf{t}) \cdot \mathbf{1}_{\{\mathbf{t} \in U\}} - \exp\left(i\langle \mathbf{t}, \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) \right| d\mathbf{t} = o(n^{-\sum_{j \in J} (j/2+1/4)}). \quad (6.6)$$

This bound will follow from showing the following estimates:

$$\int_{U^c} \exp\left(-\frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) d\mathbf{t} = O(e^{-c\sqrt{n}}) \quad (6.7)$$

$$n^{-\sum_{j \in J} (j/2+1/4)} \int_{U \setminus V} \left| \varphi(\mathbf{t}) - \exp\left(i\langle \mathbf{t}, \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) \right| d\mathbf{t} = O(e^{-c'(\log n)^2}) \quad (6.8)$$

$$n^{-\sum_{j \in J} (j/2+1/4)} \int_V \left| \varphi(\mathbf{t}) - \exp\left(i\langle \mathbf{t}, \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) \right| d\mathbf{t} = O\left(\frac{(\log n)^3}{n^{1/8}}\right) \quad (6.9)$$

where we have defined $V := \{\mathbf{t} : |t_j| \leq (\log n)n^{-j/2-1/4}\}$.

To show (6.7) along with the bound on $\det(S)$ stated in Theorem 6.2, we demonstrate upper and lower bounds on S when viewed as a quadratic form. In what follows, we set $\sigma_k^2 = \text{Var}(Y_k) = (1 - p_k)/p_k^2$ and $S_k = \text{Cov}((Y_k k^j)_{j \in J})$.

Lemma 6.16. *In the context of Theorem 6.2, define the matrix \tilde{S} via $\tilde{S}_{i,j} = S_{i,j}/n^{(i+j+1)/2}$ for all $i, j \in J$. Then there is a constant $C = C(K) > 0$ so that*

$$C^{-1}\|\mathbf{a}\|^2 \leq \mathbf{a}^T \tilde{S} \mathbf{a} \leq C\|\mathbf{a}\|^2 \tag{6.10}$$

for all $\mathbf{a} \in \mathbb{R}^J$. In particular, $\det(S) = \Theta(n^{\sum_{j \in J} (j+1/2)})$.

Proof. By rescaling, assume without loss of generality that $\|\mathbf{a}\| = 1$. Compute

$$\mathbf{a}^T \tilde{S} \mathbf{a} = \frac{1}{\sqrt{n}} \sum_{k=1}^{\infty} \sigma_k^2 \left(\sum_{j \in J} a_j (k/\sqrt{n})^j \right)^2 \tag{6.11}$$

where $\sigma_k^2 = \text{Var}(Y_k)$. If we define

$$f_{\tilde{\beta}}(t) = \frac{\exp\left(-\sum_{j \in J} \tilde{\beta}_j t^j\right)}{\left(1 - \exp\left(-\sum_{j \in J} \tilde{\beta}_j t^j\right)\right)^2}$$

then $\sigma_k^2 = f_{\tilde{\beta}}(k/\sqrt{n})$, where we have written $\tilde{\beta} = (\tilde{\beta}_j n^{j/2})_{j \in J}$. We then have

$$\mathbf{a}^T \tilde{S} \mathbf{a} = \frac{1}{\sqrt{n}} \sum_{k \geq 1} f_{\tilde{\beta}}(k/\sqrt{n}) \left(\sum_{j \in J} a_j (k/\sqrt{n})^j \right)^2 = \int_0^{\infty} f_{\tilde{\beta}}(x) \left(\sum_{j \in J} a_j x^j \right)^2 dx + E \tag{6.12}$$

where

$$E \leq \frac{1}{\sqrt{n}} (g_{\tilde{\beta}}(1/\sqrt{n}) + g'_{\tilde{\beta}}(1/\sqrt{n})) + \int_0^{1/\sqrt{n}} g_{\tilde{\beta}}(x) dx + \frac{1}{n} \int_0^{\infty} |g''_{\tilde{\beta}}(x)| dx$$

where $g_{\tilde{\beta}}(x) = f_{\tilde{\beta}}(x) (\sum_{j \in J} a_j x^j)^2$. Bounding each term in a straightforward manner gives $|E| = O(1/\sqrt{n})$ where the error is uniform since $\tilde{\beta}$ varies in a compact set.

Since \mathbf{a} and $\tilde{\beta}$ vary over compact sets, the integral on the right-hand-side of (6.12) is bounded above and below away from 0; thus, for n sufficiently large, (6.12) demonstrates uniform bounds above and below on $\mathbf{a}^T \tilde{S} \mathbf{a}$. For each remaining small n , the sum

$$\frac{1}{\sqrt{n}} \sum_{k \geq 1} f_{\tilde{\beta}}(k/\sqrt{n}) \left(\sum_{j \in J} a_j (k/\sqrt{n})^j \right)^2$$

may be uniformly bounded above and below by compactness and continuity, thereby completing the proof. \square

The bound (6.7) now follows easily from Lemma (6.16): for a given \mathbf{t} define \mathbf{s} via $s_j = t_j n^{j/2+1/4}$; then

$$\begin{aligned} \int_{U^c} \exp\left(-\frac{1}{2} \mathbf{t}^T S \mathbf{t}\right) dt &\leq \int_{\mathbb{R}^J \setminus [-\delta n^{1/4}, \delta n^{1/4}]^J} \exp(-c\|\mathbf{s}\|^2) ds \\ &= O(e^{-c'\sqrt{n}}) \end{aligned}$$

for some constant $c' > 0$. The next two sections show (6.8) and (6.9).

6.5. Establishing (6.8)

To show (6.8), we will show that for a geometric variable Y with mean bounded above we can compare the characteristic function of Y to that of a corresponding Gaussian with an error depending $\text{Var}(Y)$ (this is Lemma 6.18 below). In the problem at hand, our geometric variables actually have unbounded means; our first step is to show that while the means can be unbounded, the bulk of the contribution to the variance S comes from variables of bounded mean. In this direction, we alter the proof of Lemma 6.16 to show that the contribution of the variance from the first $\varepsilon\sqrt{n}$ terms can be made to be less than half provided ε is small enough:

Lemma 6.17. *There exists an $\varepsilon > 0$ so that for all $\mathbf{t} \in U$ we have*

$$\mathbf{t}^T S \mathbf{t} \geq \frac{1}{2} \mathbf{t}^T S^{(\varepsilon)} \mathbf{t}$$

where

$$S^{(\varepsilon)} := \sum_{k \geq \sqrt{n}\varepsilon} S_k.$$

Proof. Define \mathbf{s} via $s_j = t_j n^{j/2+1/4}$. Then

$$\mathbf{t}^T S \mathbf{t} = \mathbf{s}^T \tilde{S} \mathbf{s}.$$

We then want to show

$$\mathbf{s}^T \tilde{S} \mathbf{s} \geq \frac{1}{2} \mathbf{s}^T \tilde{S}^{(\varepsilon)} \mathbf{s}$$

where $\tilde{S}^{(\varepsilon)}$ is defined in the analogous way. Since both sides are homogeneous in \mathbf{s} , assume without loss of generality that \mathbf{s} is a unit vector.

The proof of Lemma 6.16 shows

$$\mathbf{s}^T \tilde{S}^{(\varepsilon)} \mathbf{s} = \int_{\varepsilon}^{\infty} \left(\sum_{j \in J} s_j x^j \right)^2 \sigma_x^2 dx + O(1/\sqrt{n})$$

where the error may be taken uniformly over ε . By compactness we have that for each $\varepsilon > 0$ that

$$\max_{\|\mathbf{s}\|=1} \int_0^{\varepsilon} \left(\sum_{j \in J} s_j x^j \right)^2 \sigma_x^2 dx$$

is bounded above and tends to zero as $\varepsilon \rightarrow 0^+$. Similarly, we also have

$$\min_{\|\mathbf{s}\|=1} \int_0^{\infty} \left(\sum_{j \in J} s_j x^j \right)^2 \sigma_x^2 dx =: c > 0.$$

Choosing ε small enough then gives the desired bound. □

Seeking to show our characteristic function of \mathbf{X} is close the characteristic function of the corresponding Gaussian, we need a tail bound on the cumulant generating function $\log \mathbb{E}e^{itY}$ for geometric random variables. This will allow us to approximate the characteristic function for each Y with that of a Gaussian with mean $\mathbb{E}Y$ and variance $\text{Var}(Y)$.

Lemma 6.18. Fix $\varepsilon > 0$. Let Y be a geometric random variable for some $p \in [\varepsilon, 1)$. Then there are constants $C, c > 0$ so that for all $|t| \leq c$, we have

$$\left| \log \mathbb{E}e^{itY} - \left(it\mathbb{E}Y - \frac{t^2}{2}\text{Var}(Y) \right) \right| \leq C\text{Var}(Y)|t|^3.$$

Proof. For convenience, write $\mu = \mathbb{E}Y$ and $\sigma^2 = \text{Var}(Y)$ and set $\hat{Y} = Y - \mu$. Then

$$\left| \mathbb{E}e^{it(Y-\mu)} - (1 - \sigma^2 t^2/2) \right| \leq |t|^3 \mathbb{E}|Y - \mu|^3.$$

We bound

$$\mathbb{E}|Y - \mu|^3 \leq 8(\mathbb{E}|Y^3| + \mu^3) = 8 \left(\frac{(1-p)(p^2 - 6p + 6)}{p^3} + \frac{(1-p)^3}{p^3} \right) \leq C_\varepsilon \frac{1-p}{p^2} = C_\varepsilon \sigma^2.$$

Note $\mathbb{E}e^{itY} = \frac{p}{1+(1-p)e^{it}}$ and so $|\arg \mathbb{E}e^{itY}| = |\arg(1 + (1-p)e^{it})|$ which is uniformly bounded since $1-p$ is uniformly bounded away from 1. Write

$$\begin{aligned} \left| \log \mathbb{E}e^{it(Y-\mu)} + \frac{t^2}{2}\sigma^2 \right| &= \left| \log \left(\mathbb{E}e^{it(Y-\mu)} e^{\sigma^2 t^2/2} \right) \right| \\ &= \left| \log \left(1 - \left(\frac{e^{-t^2\sigma^2/2} - \mathbb{E}e^{it(Y-\mu)}}{e^{-t^2\sigma^2/2}} \right) \right) \right| \\ &\leq C \left| \frac{\mathbb{E}e^{it(Y-\mu)} - e^{-t^2\sigma^2/2}}{e^{-t^2\sigma^2/2}} \right| \\ &\leq C' \left| \mathbb{E}e^{it(Y-\mu)} - (1 - t^2\sigma^2/2) \right| + O(|t|^4\sigma^4) \\ &\leq C''|t|^3\sigma^2, \end{aligned}$$

where each inequality uses the fact that $|t|$ and σ are uniformly bounded above. □

In a similar vein to Lemma 6.16, we need the following simple bound whose proof is omitted since it is essentially the same as that of Lemma 6.16.

Lemma 6.19.

$$\frac{1}{\sqrt{n}} \sum_{k \geq 1} \sigma_k^2 \left(\sum_{j \in J} (k/\sqrt{n})^j \right)^3 = O(1).$$

Now define $T_k := \{t : k \leq \max_j n^{j/2+1/4}|t_j| \leq k+1\}$ and note that $U \setminus V = \bigcup_{k=1}^{\delta n^{1/4}} T_k$.

Lemma 6.20. *There exist constants c, C so that for $\mathbf{t} \in T_k$ we have*

$$\mathbf{t}^T S \mathbf{t} \geq ck^2$$

and

$$\sum_{x \geq \varepsilon \sqrt{n}} \left| \log \varphi_r(\mathbf{t}) - \left(i \langle \mathbf{t}, \mu_r \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t} \right) \right| \leq C \delta k^2.$$

Proof. The first statement follows from (6.10). For the second, note that for $\mathbf{t} \in T_k$ we have

$$\begin{aligned} \sum_{x \geq \varepsilon \sqrt{n}} \left| \log \varphi_x(\mathbf{t}) - \left(i \langle \mathbf{t}, \mathbb{E} Y_x \rangle - \frac{1}{2} \mathbf{t}^T S_x \mathbf{t} \right) \right| &\leq C \sum_{x \geq \varepsilon \sqrt{n}} \sigma_x^2 |Q_{\mathbf{t}}(x)|^3 \\ &\leq C \sum_{x \geq 1} n^{-3/4} \sigma_x^2 \left(\sum_{j \in J} (x/\sqrt{n})^j (k+1) \right)^3 \\ &\leq C' k^3 n^{-3/4} \sum_{x \geq 1} \sigma_x^2 \left(\sum_{j \in J} (x/\sqrt{n})^j \right)^3 \\ &\leq C' \delta k^2 n^{-1/2} \sum_{x \geq 1} \sigma_x^2 \left(\sum_{j \in J} (x/\sqrt{n})^j \right)^3 \\ &\leq C'' \delta k^2, \end{aligned}$$

where the first bound is by Lemma 6.18 and the last bound is via Lemma 6.19. \square

As an immediate result, we see that if $\mathbf{t} \in T_k \cap U$ for large k then $|\varphi(\mathbf{t})|$ is quite small.

Corollary 6.21. *For δ sufficiently small, there is a constant $c > 0$ so that for $\mathbf{t} \in T_k \cap U$ we have*

$$|\varphi(\mathbf{t})| \leq e^{-ck^2}.$$

Proof. Bound

$$\begin{aligned} |\varphi(\mathbf{t})| &\leq \prod_{x \geq \varepsilon n^{1/4}} |\varphi_x(\mathbf{t})| \\ &\leq \exp(-(c/4 - C\delta)k^2) \\ &\leq \exp(-c'k^2), \end{aligned}$$

where we have chosen δ sufficiently small and used Lemma 6.17. \square

Since the measure of the set $\{\mathbf{b} : k \leq |b_j| \leq k+1\}$ is equal to some polynomial $p(k)$, the measure of T_k is $|T_k| = p(k)n^{-\sum_{j \in J} (j/2+1/4)}$.

With these preliminaries in place, we are ready to tackle (6.8):

Proof of (6.8). To begin with, bound

$$\int_{U \setminus V} \left| \varphi(\mathbf{t}) - \exp(i\langle \mathbf{t}, \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}) \right| dt \leq \sum_{k=\log n}^{\lceil \delta n^{1/4} \rceil} |T_k| \max_{\mathbf{t} \in T_k} \left| \varphi(\mathbf{t}) - \exp(i\langle \mathbf{t}, \mathbb{E}\mathbf{X} \rangle - \frac{1}{2} \mathbf{t}^T S \mathbf{t}) \right|,$$

then compute

$$\begin{aligned} \sum_{\log n \leq k \leq \delta n^{1/4}} |T_k| \max_{\mathbf{t} \in T_k} \left| \varphi(\mathbf{t}) - \exp(-\frac{1}{2} \mathbf{t}^T S \mathbf{t}) \right| &\leq n^{-\sum_{j \in J} (j/2+1/4)} \sum_{k \geq \log n} p(k) \exp(-ck^2) \\ &= O\left(e^{-c'(\log(n))^2}\right). \end{aligned}$$

□

6.6. Establishing (6.9)

Our proof of (6.9) can be viewed as an adaptation of the classical proof of the Lindeberg–Feller central limit theorem with an explicit error bound; see, for instance, [Dur10, Chapter 3.4] for a similar proof and discussion of the classic theorem.

We first bound contribution of the tail to the second moment of a geometric variable.

Lemma 6.22. *Let X be a geometric random variable with parameter p . Then for each $C > 0$ we have*

$$\mathbb{E} \left(|X - \mathbb{E}X|^2 \mathbf{1}_{\{|X - \mathbb{E}X| \geq C\}} \right) \leq \frac{3\text{Var}(X)}{Cp}.$$

Proof. By the Cauchy–Schwartz inequality we have

$$\mathbb{E} \left(|X - \mathbb{E}X|^2 \mathbf{1}_{\{|X - \mathbb{E}X| \geq C\}} \right) \leq \sqrt{\mathbb{E} \left(|X - \mathbb{E}X|^4 \right) \mathbb{P} \left(|X - \mathbb{E}X| \geq C \right)}.$$

Compute

$$\begin{aligned} \mathbb{E} |X - \mathbb{E}X|^4 &= \sum_{k \geq 0} \left(k - \frac{1-p}{p} \right)^4 p \cdot (1-p)^k = \frac{(1-p)(9-9p+p^2)}{p^4} \\ &\leq \frac{9(1-p)}{p^4} = \frac{9\text{Var}(X)}{p^2}. \end{aligned}$$

By Chebyshev’s inequality, bound

$$\mathbb{P} \left(|X - \mathbb{E}X| \geq C \right) \leq \frac{\text{Var}(X)}{C^2}.$$

Putting the three equations together completes the proof.

□

We are now ready to show (6.9).

Proof of (6.9). Fix $\mathbf{t} \in \mathbb{R}^J$ with $\|\mathbf{t}\| \leq 1$ and let $\varepsilon > 0$ to be chosen later. Consider the random variables

$$Z_k := Z_k(\mathbf{t}) := (Y_k - \mathbb{E}Y_k) \sum_{j \in J} k^j t_j n^{-j/2-1/4}.$$

Then for $|\theta| \leq \log n$ note that

$$\varphi_{\widehat{\mathbf{X}}}(\theta \mathbf{t}) = \prod_{k \geq 1} \mathbb{E} e^{i\theta Z_k}$$

and so

$$\begin{aligned} |\varphi_{\widehat{\mathbf{X}}}(\theta \mathbf{t}) - \exp(-\theta^2 \mathbf{t}^T S \mathbf{t} / 2)| &\leq \left| \exp(-\theta^2 \mathbf{t}^T S \mathbf{t} / 2) - \prod_{k \geq 1} (1 - \theta^2 \mathbb{E} Z_k^2 / 2) \right| \\ &\quad + \sum_{k \geq 1} |\mathbb{E}[e^{i\theta Z_k}] - (1 - \theta^2 \mathbb{E} Z_k^2 / 2)|. \end{aligned}$$

Note that

$$\begin{aligned} \prod_{k \geq 1} (1 - \theta^2 \mathbb{E} Z_k^2 / 2) &= \exp\left(\sum_{k \geq 1} \log(1 - \theta^2 \mathbb{E} Z_k^2 / 2)\right) \\ &= \exp\left(-\theta^2 \mathbf{t}^T S \mathbf{t} / 2 + \sum_{k \geq 1} O(\theta^4 (\mathbb{E} Z_k^2)^2)\right) \\ &= \exp\left(-\theta^2 \mathbf{t}^T S \mathbf{t} / 2 + O(1/\sqrt{n})\right) \end{aligned}$$

where the third line is obtained by comparing a sum to an integral as in the proof of Lemma 6.16. It is thus sufficient to show

$$\max_{\|\mathbf{t}\| \leq 1, |\theta| \leq \log n} \sum_{k \geq 1} |\mathbb{E}[e^{i\theta Z_k}] - (1 - \theta^2 \mathbb{E} Z_k^2 / 2)| = O\left(\frac{(\log n)^3}{n^{1/8}}\right). \quad (6.13)$$

Working towards this, recall that for any mean-zero variable X we have

$$|\mathbb{E}[e^{itX}] - (1 - t^2 \mathbb{E} X^2 / 2)| \leq \mathbb{E}[\min\{|tX|^3, 2|tX|^2\}]. \quad (6.14)$$

See, for instance, [Dur10, Equation (3.3.3)]. Applying this bound to Z_k gives

$$\begin{aligned} |\mathbb{E}[e^{i\theta Z_k}] - (1 - \theta^2 \mathbb{E} Z_k^2 / 2)| &\leq \mathbb{E}[\min\{|\theta Z_k|^3, 2|\theta Z_k|^2\}] \\ &\leq \mathbb{E}[|\theta Z_k|^3 \mathbf{1}\{|Z_k| \leq \varepsilon\}] + \mathbb{E}[2|\theta Z_k|^2 \mathbf{1}\{|Z_k| > \varepsilon\}] \\ &\leq |\theta|^3 \varepsilon \mathbb{E}[|Z_k|^2 \mathbf{1}\{|Z_k| \leq \varepsilon\}] + 2\theta^2 \mathbb{E}[|Z_k|^2 \mathbf{1}\{|Z_k| > \varepsilon\}] \\ &\leq |\theta|^3 \varepsilon \text{Var}(Z_k) + 2\theta^2 \mathbb{E}[|Z_k|^2 \mathbf{1}\{|Z_k| > \varepsilon\}]. \end{aligned}$$

Using (6.14) and Lemma 6.22, bound

$$\begin{aligned} \sum_{k \geq 1} |\mathbb{E}[e^{i\theta Z_k}] - (1 - \theta^2 \mathbb{E}Z_k^2/2)| &\leq \varepsilon |\theta|^3 \mathbf{t}^T \tilde{S} \mathbf{t} + 2\theta^2 \sum_{k \geq 1} \mathbb{E}[|Z_k|^2 \mathbf{1}\{|Z_k| > \varepsilon\}] \\ &\leq \varepsilon |\theta|^3 \mathbf{t}^T \tilde{S} \mathbf{t} + 6 \frac{\theta^2}{n^{1/4} \varepsilon} \cdot \left(\frac{1}{\sqrt{n}} \sum_{k \geq 1} \left(\sum_{j \in J} k^j t_j n^{-j/2} \right)^3 \frac{\sigma_k^2}{p_k} \right) \\ &= O \left(\varepsilon |\theta|^3 + \frac{\theta^2}{n^{1/4} \varepsilon} \right), \end{aligned}$$

where in the last line we compared a sum to an integral to obtain this error term. Choosing $\varepsilon = n^{-1/8}$ and bounding $|\theta| \leq \log n$ completes the proof. \square

Proof of Theorem 6.2. Chaining together Lemmas 6.12 and 6.15 with (6.6) completes the proof. \square

Theorem 1.1 follows immediately from Corollary 3.5, Lemma 4.1, and Lemma 6.1.

7. Extensions and future directions

There is a wealth of extensions and generalizations of the problem of enumerating integer partitions (see e.g. [And98]). Many of these extensions can be framed in the maximum entropy framework, leading to continuous convex programs like (1.7) that express the exponential growth rate of the partition number. Once we have such a program it is natural to modify it by adding additional constraints or restricting the domain of the candidate functions. These new optimization problems can be translated back to give new classes of integer partitions problems. Posing the problems in the maximum entropy framework gives a natural, unifying explanation of many methods and formulas in the literature, but also illuminates some new connections between integer partitions and infinite-dimensional convex programming. Below we indicate some possible extensions of these methods to previously studied classes of integer partitions.

7.1. Distinct partitions: changing the objective function

One way to restrict a class of partitions is to insist that each part appear with one of a set of prescribed multiplicities. The simplest such case is that of distinct partitions: partitions in which each part occurs with multiplicity at most 1; equivalently a distinct partition is a subset of \mathbb{N} (as opposed to a multiset). Let $q(n)$ denote the number of distinct partitions of n . The asymptotics of $q(n)$ have been studied in, e.g. [Mei54, Rom05].

Given a profile \mathbf{N} indexed by the profile set J , let $q(\mathbf{N})$ be the number of distinct partitions λ so that $\sum_{x \in \lambda} x^j = \mathbf{N}_j$ for all $j \in J$, and for $\alpha \in \mathbb{R}_+^J$ let $\tilde{q}_n(\alpha) = q(\mathbf{N}(\alpha, n))$. Using the methods of this paper, and making an assumption analogous to Assumption 1, we believe one could show the exponential growth rate of $\tilde{q}_n(\alpha)$ is again given by the optimum of a continuous

convex program:

$$M_{\text{dist}}(\boldsymbol{\alpha}) = \max_{f \in \mathcal{F}_1} \int_0^\infty H(f(x)) dx$$

subject to $\int_0^\infty x^j f(x) dx = \alpha_j$ for $j \in J$,

where \mathcal{F}_1 is the set of all integrable functions $f : [0, \infty) \rightarrow [0, 1]$ and $H(p) = -p \log p - (1 - p) \log(1 - p)$ is the entropy of a Bernoulli random variable with parameter p . This is exactly the same optimization problem as in (1.7) but with a different objective function; both objective functions are strictly concave, however, and so share essentially all the same qualitative properties.

7.2. Bounded Young diagrams: restricting the support

Another class of restricted partitions is the class of partitions with bounded Young diagrams. That is, partitions with restrictions on the size of the largest part and on the number of parts. Asymptotic enumeration of partitions of n with largest part and number of parts both $\Theta(\sqrt{n})$ has been carried out in [JW19, MPP20].

One can follow the methods of this paper to enumerate partitions with a given profile and largest part at most $A\sqrt{n}$. The continuous convex program giving the growth rate is the following:

$$M_A(\boldsymbol{\alpha}) = \max_f \int_0^A G(f(x)) dx$$

subject to $\int_0^A x^j f(x) dx = \alpha_j$ for $j \in J$,

In regards to the discussion in Section 1.5 about the existence of a solution, something different happens in this setting. There are again feasibility conditions on the moments $\boldsymbol{\alpha}$; this time related to the Hausdorff moment problem [Hau21] (that of finding a probability distribution on a bounded interval with given moments). Because the interval $[0, A]$ is compact, if a collection of moments are feasible for the truncated Hausdorff moment problem, then a unique maximum entropy distribution with the given moments always exists [MP84]. The same is true if we maximize geometric entropy, and so the only assumption on $\boldsymbol{\alpha}$ needed is feasibility: the situation discussed in Section 1.5 cannot occur.

Question 7.1. Can Theorem 1.1 be shown for partitions in a rectangle under the weakened assumption that $\boldsymbol{\alpha}$ is feasible for the Stieltjes moment problem?

7.3. Plane Partitions and Higher Dimensions

The Young diagram of a partition gives us a two-dimensional representation of the partition n ; the natural generalization to dimension three — functions $f : \mathbb{N}^3 \rightarrow \{0, 1\}$ that are weakly decreasing in each coordinate—give the structures known as *plane partitions*. The asymptotics for

the number of plane partitions of weight n was given by Wright [Wri31], although enumeration of restricted plane partitions appears to not have been explored.

Question 7.2. What are the asymptotics for the number of plane partitions that fit in an appropriately scaled rectangular box?

Acknowledgements

We thank the anonymous reviewer for their very helpful comments on the presentation and organization of the paper.

References

- [AGM⁺17] Arash Asadpour, Michel X Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n / \log \log n)$ -approximation algorithm for the asymmetric traveling salesman problem. *Operations Research*, 65(4):1043–1061, 2017. doi:10.1287/opre.2017.1603.
- [AGV18] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 35–46. IEEE, 2018.
- [And98] George E Andrews. *The Theory of Partitions*. Cambridge university press, 1998.
- [Bar17] Alexander Barvinok. Counting integer points in higher-dimensional polytopes. In *Convexity and concentration*, pages 585–612. Springer, 2017.
- [BH10] Alexander Barvinok and JA Hartigan. Maximum entropy Gaussian approximations for the number of integer points and volumes of polytopes. *Advances in Applied Mathematics*, 45(2):252–289, 2010. doi:10.1016/j.aam.2010.01.004.
- [BH12] Alexander Barvinok and John A Hartigan. The number of graphs and a random graph with a given degree sequence. *Random Structures & Algorithms*, 42(3):301–348, 2012. doi:10.1002/rsa.20409.
- [BTC77] Aharon Ben-Tal and Abraham Charnes. A dual optimization framework for some problems of information theory and statistics. Technical report, University of Texas at Austin Center for Cybernetic Studies, 1977.
- [BTTC88] Aharon Ben-Tal, Marc Teboulle, and Abraham Charnes. The role of duality in optimization problems involving entropy functionals with applications to information theory. *Journal of optimization theory and applications*, 58(2):209–223, 1988.
- [BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [Can97] E Rodney Canfield. From recursions to asymptotics: on Szekeres’ formula for the number of partitions. *Electron. J. Combin*, 4(2):19, 1997.

- [CC75] Abraham Charnes and William W Cooper. Constrained Kullback–Leibler estimation; generalized Cobb–Douglas balance, and unconstrained convex programming. *Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti*, 59:568–576, 1975.
- [CCH01] Rod Canfield, Sylvie Corteel, and Pawel Hitczenko. Random partitions with non-negative r th differences. *Advances in Applied Mathematics*, 27(2-3):298–317, 2001. doi:10.1006/aama.2001.0736.
- [Cen00] Nikolai Nikolaevich Cencov. *Statistical decision rules and optimal inference*. Number 53 in Translations of Mathematical Monographs. American Mathematical Soc., 2000.
- [CKP01] Henry Cohn, Richard Kenyon, and James Propp. A variational principle for domino tilings. *Journal of the American Mathematical Society*, 14(2):297–346, 2001. doi:10.1090/S0894-0347-00-00355-6.
- [Cov99] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.
- [Csi75] Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.
- [Cur91] Raúl E Curto. Recursiveness, positivity and truncated moment problems. *Houston Journal of Mathematics*, 17:603–635, 1991.
- [Dur10] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010. doi:10.1017/CB09780511779398.
- [DVZ00] A. Dembo, A. Vershik, and O. Zeitouni. Large deviations for integer partitions. *Markov Process. Related Fields*, 6(2):147–179, 2000.
- [Erd42] Pál Erdős. On an elementary proof of some asymptotic formulas in the theory of partitions. *Annals of Mathematics*, pages 437–450, 1942.
- [Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.
- [Fri93] Bert Fristedt. The structure of random partitions of large integers. *Transactions of the American Mathematical Society*, 337(2):703–735, 1993.
- [GT12] Ben Green and Terence Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Annals of Mathematics*, pages 465–540, 2012. doi:10.4007/annals.2012.175.2.2.
- [Hau21] Felix Hausdorff. Summationsmethoden und momentfolgen. I. *Mathematische Zeitschrift*, 9(1-2):74–109, 1921.
- [HR18] Godfrey H Hardy and Srinivasa Ramanujan. Asymptotic formulæ in combinatory analysis. *Proceedings of the London Mathematical Society*, 2(1):75–115, 1918.
- [Jay57] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [Jun00] Michael Junk. Maximum entropy for reduced moment problems. *Mathematical Models and Methods in Applied Sciences*, 10(07):1001–1025, 2000.

- [JW19] Tiefeng Jiang and Ke Wang. A generalized Hardy-Ramanujan formula for the number of restricted integer partitions. *Journal of Number Theory*, 201:322–353, 2019. doi:10.1016/j.jnt.2019.02.006.
- [Las10] Jean-Bernard Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2010.
- [Mei54] Günter Meinardus. Über partitionen mit differenzenbedingungen. *Mathematische Zeitschrift*, 61(1):289–302, 1954.
- [MP84] Lawrence R Mead and Nikos Papanicolaou. Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8):2404–2417, 1984.
- [MPP20] Stephen Melczer, Greta Panova, and Robin Pemantle. Counting partitions inside a rectangle. *SIAM Journal on Discrete Mathematics*, 34(4):2388–2410, 2020. doi:10.1137/20m1315828.
- [New62] DJ Newman. A simplified proof of the partition formula. *The Michigan Mathematical Journal*, 9(3):283–287, 1962.
- [Oko16] Andrei Okounkov. Limit shapes, real and imagined. *Bull. Amer. Math. Soc. (N.S.)*, 53(2):187–216, 2016. doi:10.1090/bull/1512.
- [Pet10] F. Petrov. Limit shapes of young diagrams. two elementary approaches. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 370(Kraevye Zadachi Matematicheskoi Fiziki i Smezhnye Voprosy Teorii Funktsii. 40):111–131, 221, 2010. doi:10.1007/s10958-010-9845-9.
- [Pit97] Boris Pittel. On a likely shape of the random Ferrers diagram. *Advances in Applied Mathematics*, 18(4):432–488, 1997.
- [Pó115] Georg Pólya. Über ganzwertige ganze funktionen. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 40(1):1–16, 1915.
- [Roc74] R Tyrrell Rockafellar. *Conjugate duality and optimization*. SIAM, 1974.
- [Rom05] Dan Romik. Partitions of n into $t\sqrt{n}$ parts. *European Journal of Combinatorics*, 26(1):1–17, 2005. doi:10.1016/j.ejc.2004.02.005.
- [San58] Ivan N Sanov. *On the probability of large deviations of random variables*. United States Air Force, Office of Scientific Research, 1958.
- [Shl01] S. B. Shlosman. The Wulff construction in statistical mechanics and combinatorics. *Uspekhi Mat. Nauk*, 56(4(340)):97–128, 2001. doi:10.1070/RM2001v056n04ABEH000417.
- [ST77a] M. Szalay and P. Turán. On some problems of the statistical theory of partitions with application to characters of the symmetric group. I. *Acta Math. Acad. Sci. Hungar.*, 29(3-4):361–379, 1977. doi:10.1007/BF01895857.
- [ST77b] M. Szalay and P. Turán. On some problems of the statistical theory of partitions with application to characters of the symmetric group. II. *Acta Math. Acad. Sci. Hungar.*, 29(3-4):381–392, 1977. doi:10.1007/BF01895858.

- [Sti94] T-J Stieltjes. Recherches sur les fractions continues. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, Série 6, Tome 4(1):J1–J122, 1894. doi:10.5802/afst.789.
- [SV14] Mohit Singh and Nisheeth K Vishnoi. Entropy, optimization and counting. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 50–59, 2014.
- [Sze53] George Szekeres. Some asymptotic formulae in the theory of partitions (II). *The Quarterly Journal of Mathematics*, 4(1):96–111, 1953.
- [Tag03] Aldo Tagliani. Maximum entropy solutions and moment problem in unbounded domains. *Applied mathematics letters*, 16(4):519–524, 2003. doi:10.1016/S0893-9659(03)00030-2.
- [Tak86] Lajos Takács. Some asymptotic formulas for lattice paths. *Journal of Statistical Planning and Inference*, 14(1):123–142, 1986.
- [Ver96] Anatolii Moiseevich Vershik. Statistical mechanics of combinatorial partitions, and their limit shapes. *Funktsional’nyi Analiz i ego Prilozheniya*, 30(2):19–39, 1996.
- [Ver97] Anatolii Moiseevich Vershik. Limit distribution of the energy of a quantum ideal gas from the viewpoint of the theory of partitions of natural numbers. *Russian Mathematical Surveys*, 52(2):379, 1997.
- [VY01] Anatolii Moiseevich Vershik and Yurii Vladimirovich Yakubovich. The limit shape and fluctuations of random partitions of naturals with fixed number of summands. *Moscow Mathematical Journal*, 1(3):457–468, 2001. doi:10.17323/1609-4514-2001-1-3-457-468.
- [Wri31] E Wright. Asymptotic partition formulae i. plane partitions. *The Quarterly Journal of Mathematics*, (1):177–189, 1931.
- [Wri34] E Maitland Wright. Asymptotic partition formulae. III. Partitions into k th powers. *Acta Mathematica*, 63:143–191, 1934.

A. Proof of Lemma 4.3

We make use of the Euler–Maclaurin summation formula which allows us to compare a sum to a corresponding integral with an explicit error term.

Lemma A.1. (*Euler–Maclaurin Formula*). *Let $f : [a, b] \rightarrow \mathbb{C}$ be a smooth function. Then*

$$\sum_{k=a}^b f(k) = \int_a^b f(x) dx + \frac{f(a) + f(b)}{2} + \frac{f'(b) - f'(a)}{12} - \int_a^b f''(x) \frac{P_2(x)}{2} dx$$

where P_2 is the second periodized Bernoulli polynomial

$$P_2(x) = (x - [x])^2 - (x - [x]) + \frac{1}{6}.$$

The following easy consequence of Stirling’s formula is closely related to the Wallis product.

Fact A.2.

$$\int_1^\infty \frac{P_2(x)}{2x^2} dx = \frac{1}{2} \log(2\pi) - \frac{11}{12}.$$

Proof. Stirling’s formula states

$$\sum_{k=1}^N \log k = N(\log N - 1) + \frac{1}{2} \log N + \frac{1}{2} \log(2\pi) + o(1).$$

Applying Lemma A.1 to the left-hand side and taking $N \rightarrow \infty$ completes the proof. □

To prove Lemma 4.3, first note

$$G\left(\frac{1}{e^a - 1}\right) = \frac{a}{e^a - 1} - \log(1 - e^{-a}). \tag{A.1}$$

We will set $a = \sum_{j \in J} \gamma_j (tk)^j$ and will find asymptotics for each of these two terms. To simplify notation in this section we will set

$$p(x) = p_\gamma(x) = \sum_{j \in J} \gamma_j x^j. \tag{A.2}$$

Lemma A.3. *As $t \rightarrow 0^+$ we have*

$$\sum_{k \geq 1} \frac{p(tk)}{e^{p(tk)} - 1} = t^{-1} \int_0^\infty \frac{p(x)}{e^{p(x)} - 1} dx - \frac{1}{2} \left(\mathbf{1}_{j_* \geq 1} + \frac{\gamma_0}{e^{\gamma_0} - 1} \mathbf{1}_{j_* = 0} \right) + o(1).$$

Proof. Set

$$f_j(t, x) := \frac{d^j}{dx^j} \frac{p(tx)}{e^{p(tx)} - 1}$$

for $j \in \{1, 2\}$ and

$$E := -\frac{f_1(t, 1)}{12} - \frac{1}{2} \int_1^\infty f_2(t, x) P_2(x) dx.$$

By the Euler–Maclaurin formula (Lemma A.1) we have

$$\begin{aligned} \sum_{k \geq 1} \frac{p(tk)}{e^{p(tk)} - 1} &= \int_1^\infty \frac{p(tx)}{e^{p(tx)} - 1} dx + \frac{1}{2} \frac{p(t)}{e^{p(t)} - 1} + E \\ &= t^{-1} \int_0^\infty \frac{p(x)}{e^{p(x)} - 1} dx - \frac{1}{2} \left(\mathbf{1}_{j_* \geq 1} + \frac{\gamma_0}{e^{\gamma_0} - 1} \mathbf{1}_{j_* = 0} \right) + o(1) + E. \end{aligned}$$

To show $E = o(1)$ note first that $f_1(t, 1) = O(t)$. Define $g(s) := \frac{s}{e^s - 1}$ and compute directly that g' and g'' decay exponentially as $s \rightarrow \infty$ and are both uniformly bounded. We then see that

$$f_2(t, x) = t^2 (g''(p(tx))(p'(tx))^2 + g'(p(tx))p''(tx)).$$

Changing variables by setting $s = tx$ we see

$$\left| \int_1^\infty f_2(t, x) P_2(x) dx \right| \leq t \int_0^\infty |g''(p(s))(p'(s))^2 + g'(p(s))p''(s)| ds \\ = O(t),$$

since g'' and g' are uniformly bounded and decay exponentially. \square

Before dealing with the other term in (A.1), we need two lemmas.

Lemma A.4. *We have*

$$\lim_{t \rightarrow 0} \int_1^\infty \frac{t^2 p''(tx)}{e^{p(tx)} - 1} P_2(x) dx = j_*(j_* - 1) \int_1^\infty \frac{P_2(x)}{x^2} dx.$$

Proof. We first claim that there is a constant M so that for all $x > 0$ we have

$$\frac{t^2 p''(tx)}{e^{p(tx)} - 1} \leq \frac{M}{x^2}.$$

Multiplying both sides by x^2 and setting $b = tx$ we need to show that the function

$$\frac{b^2 p''(b)}{e^{p(b)} - 1}$$

is uniformly bounded for all $b > 0$. For b near zero, the denominator is $\Omega(b^{j_*})$ and the numerator is $O(b^2 \cdot b^{j_* - 2}) = O(b^{j_*})$. This shows that the function is bounded for b in a neighborhood of 0. Conversely, the function tends to zero as b tends to infinity, thus showing the desired inequality.

To prove the lemma, we apply the dominated convergence theorem along with the fact that for each fixed x we have the desired limit. \square

By a similar argument, we see the following.

Lemma A.5. *We have*

$$\lim_{t \rightarrow \infty} \int_1^\infty \frac{e^{p(tx)} t^2 (p'(tx))^2}{(e^{p(tx)} - 1)^2} P_2(x) dx = j_*^2 \int_1^\infty \frac{P_2(x)}{x^2} dx.$$

Lemma A.6. *As $t \rightarrow 0^+$ we have*

$$\sum_{k \geq 1} \log(1 - \exp(-p(tk))) = t^{-1} \int_0^\infty \log(1 - \exp(-p(y))) dy - \frac{\mathbf{1}_{j_* = 0}}{2} \log(1 - e^{-\gamma_0}) \\ + \mathbf{1}_{j_* \geq 1} \left(\frac{j_*}{2} \log(2\pi) - \frac{1}{2} \log(\gamma_{j_*}) + \frac{j_*}{2} \log(1/t) \right) + o(1).$$

Proof. Note that

$$\begin{aligned} \frac{d}{dx} \log(1 - e^{-p(tx)}) &= \frac{tp'(tx)}{e^{p(tx)} - 1} =: f_1(t, x) \\ \frac{d^2}{dx^2} \log(1 - e^{-p(tx)}) &= \frac{t^2 p''(tx)}{e^{p(tx)} - 1} - \frac{e^{p(tx)} t^2 (p'(tx))^2}{(e^{p(tx)} - 1)^2} \\ &=: f_2(t, x). \end{aligned}$$

By the Euler–Maclaurin formula, we then have

$$\begin{aligned} \sum_{k \geq 1} \log(1 - \exp(-p(tk))) &= \int_1^\infty \log(1 - \exp(-p(tx))) dx + \frac{1}{2} \log(1 - e^{-p(t)}) \\ &\quad - \frac{1}{12} f_1(t, 1) - \frac{1}{2} \int_1^\infty f_2(t, x) P_2(x) dx. \end{aligned}$$

Each piece will be dealt with separately. Write

$$\int_1^\infty \log(1 - \exp(-p(tx))) dx = t^{-1} \int_0^\infty \log(1 - \exp(-p(y))) dy - \int_0^1 \log(1 - \exp(-p(tx))) dx.$$

Note if $j_* = 0$ then

$$\int_0^1 \log(1 - \exp(-p(tx))) dx = \log(1 - e^{-\gamma_0}) + o(1)$$

and if $j_* \geq 1$ then

$$\begin{aligned} \int_0^1 \log(1 - \exp(-p(tx))) dx &= \int_0^1 \log(p(tx)) dx + o(1) \\ &= \int_0^1 \log(\gamma_{j_*} t^{j_*} x^{j_*}) dx + o(1) \\ &= \log(\gamma_{j_*}) + j_* \log t - j_* + o(1). \end{aligned}$$

Now compute

$$\frac{1}{2} \log(1 - e^{-p(t)}) = \frac{\mathbf{1}_{j_*=0}}{2} \log(1 - e^{-\gamma_0}) + \frac{\mathbf{1}_{j_* \geq 1}}{2} \log(\gamma_{j_*} t^{j_*}) + o(1)$$

and

$$-\frac{1}{12} f_1(t, 1) \sim -\frac{1}{12} \cdot \frac{j_* t \gamma_{j_*} t^{j_*-1}}{\gamma_{j_*} t^{j_*}} = -\frac{j_*}{12}.$$

Applying the previous two lemmas to deal with the P_2 term finishes the proof. □

Proof of Lemma 4.3. Combining (A.1) with Lemmas A.3 and A.6 completes the proof. □