ABSTRACT

Independence Models for Integer Points of Polytopes

by
Austin Warren Shapiro

Chair: Alexander I. Barvinok

The integer points of a high-dimensional polytope P are generally difficult to count or sample uniformly. We consider a class of low-complexity random models for these points which arise from an entropy maximization problem. From these models, by way of "anti-concentration" results for sums of independent random variables, we derive general, efficiently computable upper bounds on the number of integer points of P.

We make a detailed study of contingency tables with bounded entries, which are the integer points of a transportation polytope truncated by a cuboid. We provide efficiently computable estimates for the logarithm of the number of $m \times n$ tables with specified row and column sums $r_1, \ldots, r_m, c_1, \ldots, c_n$ and bounds on the entries. These estimates are asymptotic as $m, n \to \infty$ simultaneously, given that no r_i (resp., c_j) is allowed to exceed a fixed multiple of the average row sum (resp., column sum).

As an application, we consider a random, uniformly selected table with entries $\leq \kappa$ having a given sum. Responding to questions raised by Diaconis and Efron in the context of statistical significance testing, we show that the occurrence of row

sums $r_1, \ldots, r_m$ is positively correlated with the occurrence of column sums $c_1, \ldots, c_n$ when $\kappa \geq 2$ and $r_1, \ldots, r_m, c_1, \ldots, c_n$ are sufficiently extreme. We give evidence that the opposite is true for near-average values of $r_1, \ldots, r_m, c_1, \ldots, c_n$.

INDEPENDENCE MODELS FOR INTEGER POINTS OF POLYTOPES

by

Austin Warren Shapiro

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)
in The University of Michigan
2011

Doctoral Committee:

Professor Alexander Barvinok, Chair

Professor Mark Rudelson

Professor Roman Vershynin

Assistant Professor Seth Pettie

ACKNOWLEDGMENTS

The influence of Sasha Barvinok on this dissertation comprises two parts, of which the lesser is attested by the twenty-five mentions of his name herein (not counting the title page or bibliography). The greater but less visible part is the encouragement he gave me (occasionally rising, as needed, to mild compulsion) to press on through difficulties and complete the work. Sasha also enlarged my ambitions by convincing me that, if I would be a good combinatoricist, it wouldn't hurt to be a good analyst as well.

I am grateful for the work of my committee, and especially to Roman Vershynin for acquainting me with Littlewood-Offord theory and the work of Gábor Halász.

Keith Ball's course on convex analysis, which he taught during a visit to University of Michigan in Fall 2007, held great sway over my subsequent interests—greater than I realized at the time. The same is true of my 2002 REU at UC Davis, during which I worked peripherally on the LattE software project under the guidance of Jesús De Loera. Besides introducing me to integer points of polytopes and to (the work of) my future advisor, Jesús also set me a fine example of a mathematician who is busy *doing* (I had thought *thinking* their main activity). When I first met him to discuss the REU, he asked me, "Can you program in Maple?" When I answered that I had never used Maple in my life, he said, "Well, you'll have a couple of days before the REU begins—that's long enough to learn."

My interest in Sperner theory was stoked by a single, highly enjoyable conversation

with John Goldwasser.

My fascination with mathematics has been evident for about as long as I have been sentient; this, at least, is the story according to my parents, Ren and Art, who thus deny themselves the credit. However, they deserve all the greater credit for their constant nourishment of my rather demanding appetites, intellectual and other. They also raised me (or tried their best) to be a *mensch*, which is something more than a mathematician.

Essential contributors to my education are too many for me to name. Today I am thinking of (my uncle) Neale Austin, George Bergman, Greg Kuperberg, Joanne Moldenhauer, Motohico Mulase, Deanne Quinn, Karen Rhea, Tom Sallee, and Zvezda Stankova, and my peers Andrew Dudzik, Paul Shearer, and Jeremy Tauzer. Tomorrow, I am sure to regret some omissions from that list.

I might follow custom by concluding, "I couldn't have done this work without the love, support, patience, etc. of my spouse, Mandy," but is this true? Had I been a loveless hermit, I might have done just as much mathematics—even a little more, for I would have known fewer tender cares and delightful pastimes outside it. Accounting only for productivity, Mandy's true and steady companionship has demanded more of me than her (many) helps can compensate. I am compensated by happiness.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

CHAPTER I

Introduction: Integer Points of Polytopes

A **polytope**, here used interchangeably with **bounded convex polytope**, may be variously defined as

(i) the convex hull of finitely many points in $\mathbb{R}^n$,

(ii) a bounded region formed by the intersection of half-spaces in $\mathbb{R}^n$, or

(iii) a bounded region formed as the locus of solutions $\mathbf{x} \in \mathbb{R}^n$ to a system of linear inequalities $A\mathbf{x} \leq \mathbf{b}$, where A is a real $m \times n$ matrix, $\mathbf{b}$ is a real m-vector, and the inequality is understood componentwise.

Definitions (ii) and (iii) are easily seen to be equivalent. Their equivalence to (i) is only slightly more difficult (a proof is given in [37]), but it can be quite hard to recover a description of a specific polytope in form (i) from a description in form (ii) or (iii), or *vice versa*. This problem lies beyond the scope of our efforts, and we will assume that the polytopes we work with are given in a form similar to (iii):

Definition I.1. A **polytope in standard form** is a bounded region of the form

$$\{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{x} \geq \mathbf{0},\ A\mathbf{x} = \mathbf{b}\},$$

where A is a real $m \times n$ matrix, $\mathbf{b}$ is a real m-vector, and (in)equality of vectors is understood componentwise.

Clearly, a polytope in standard from *is* a polytope as defined in (iii), but the converse is true only in a special sense, which we now explain. A polytope is called **rational** if it can be written in form (iii) with all entries of A and $\mathbf{b}$ integers. (This turns out to be equivalent to having a description of type (i) in which all points have rational coordinates. A polytope whose vertices are integer points is called a **lattice polytope**.) We borrow a definition from [23]:

Definition I.2. Let $P \subset \mathbb{R}^p$, $Q \subset \mathbb{R}^q$ be polytopes, where $p \leq q$. We say that Q **represents** P if there is an injection $\sigma : \{1, \ldots, p\} \to \{1, \ldots, q\}$ such that the coordinate-erasing projection $\pi : \mathbb{R}^q \to \mathbb{R}^p$ taking $(x_1, \ldots, x_q)$ to $(x_{\sigma(1)}, \ldots, x_{\sigma(p)})$ induces a bijection of Q onto P. If, moreover, π induces a bijection between the integer points of Q and the integer points of P, then we say that Q **represents** P **with respect to integer points**.

For every rational polytope P, there is a polytope Q in standard form which represents P with respect to integer points. We may obtain Q by translating P by an integer vector so that it lies in the principal orthant, and by introducing "slack variables" which turn inequalities into equations. For instance, the inequality $a_1x_1 + \cdots + a_nx_n \leq b$ may be rewritten as $a_1x_1 + \cdots + a_nx_n + y = b$, where $y \geq 0$. When our purpose is to count the integer points of P, its representation Q will do just as well.

1.1 Why count integer points of polytopes?

Many objects of combinatorial interest can be expressed as the integer points of some polytope. We give a tour of a few well-known examples, with applications of counting interspersed throughout.

1.1.1 Feasible flows

A **network** is a triple (G, b, k), where $G = (V, E)$ is a finite directed graph, $b : V \to \mathbb{R}$ is a function on the vertices (called the *excess* or *demand*), and $k : E \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is a function on the edges (called the *capacity*).[1] A **feasible flow** on this network is a function $x : E \to \mathbb{R}_{\geq 0}$ such that

(i) For every $v \in V$, we have $\sum_{\substack{e \in E: \\ v = \text{head}(e)}} x(e) - \sum_{\substack{e \in E: \\ v = \text{tail}(e)}} x(e) = b(v)$.

(ii) For every $e \in E$, we have $x(e) \leq k(e)$.

Note that for condition (i) to be satisfiable, the total excess on all vertices must equal zero.

Conditions (i) and (ii) are linear. If G is acyclic, then these conditions define a bounded region (hence a polytope) called the **flow polytope** of the network; it may be concisely described as

$$\{x \in \mathbb{R}^E : \quad Ax = b,\ 0 \leq x \leq k\},$$

where A is the signed vertex-edge incidence matrix of G. The integer points of this polytope are (sensibly enough) called **integer feasible flows**. Exact counting of integer feasible flows is a $\#P$-complete problem in terms of the length of the input A, $b(, k)$.[2] An algorithm is given in [2], where applications of counting flows are also discussed. Several of the objects to follow in this list are instances of feasible flows.

1.1.2 Contingency tables

A contingency table is defined as a nonnegative integer matrix with specified row and column sums, called the *margins*. Given vectors

$$R = (r_1, r_2, \ldots, r_m) \in \mathbb{Z}^m_{\geq 0} \quad \text{and} \quad C = (c_1, c_2, \ldots, c_n) \in \mathbb{Z}^n_{\geq 0}$$

[1] The capacity is conventionally denoted by the letter c, but we wish to reserve this letter for other purposes later.

[2] The class $\#P$ consists of counting problems for which the corresponding decision problems are NP. A $\#P$ problem is $\#P$-complete if every $\#P$ problem can be reduced to it.

such that

$$r_1 + r_2 + \cdots + r_m = c_1 + c_2 + \cdots + c_n = N,$$

we denote by $\Pi(R, C)$ the set of all $X = \big(x_{ij}\big) \in \mathbb{R}_{\geq 0}^{m \times n}$ such that

$$\sum_{j=1}^{n} x_{ij} = r_i \ (1 \leq i \leq m) \quad \text{and} \quad \sum_{i=1}^{m} x_{ij} = c_j \ (1 \leq j \leq n).$$

Then $\Pi(R, C)$ is a polytope, called the **transportation polytope** associated to R and C, and its integer points are the contingency tables with margins R and C. (We may call N the 0-*margin*.)

The name of the transportation polytope comes from its interpretation as the flow polytope of a complete bipartite graph $K_{m,n}$ with all edges directed from the vertices of the first component ("sources") to the vertices of the second component ("sinks"). Source i is assigned negative excess $-r_i$, sink j is assigned positive excess c_j, and x_{ij} is understood as the flow from source i to sink j, so that a feasible flow across the network represents a schedule for transporting goods from sites of supply to sites of demand.[3] Because the underlying graph of the network is bipartite, we may eliminate all signs from A and $\mathbf{b}$ in the standard form of the transportation polytope. The matrix A then has the characteristic form

$$\begin{pmatrix} 1 & 1 & \cdots & 1 & & & & & & & & & & \\ & & & & 1 & 1 & \cdots & 1 & & & & & & \\ & & & & & & & & \ddots & & & & & \\ & & & & & & & & & & & 1 & 1 & \cdots & 1 \\ 1 & & & & 1 & & & & & & & 1 & & \\ & 1 & & & & 1 & & & \cdots & & & 1 & & \\ & & \ddots & & & & \ddots & & & & & & \ddots & \\ & & & 1 & & & & 1 & & & & & & 1 \end{pmatrix}. \tag{1.1}$$

[3] As an aside, given a cost function $w : E \to \mathbb{R}_{\geq 0}$ on the edges of G, and defining the cost of a flow x as $\sum_{e \in E} w(e)x(e)$, we may ask what is the cheapest feasible flow satisfying the excess constraints; this is the *transportation problem*. In this context, the integer feasible flows are the natural candidates in case the goods in question can only be transported in discrete units. But even if the goods are arbitrarily divisible, it turns out [34] that the optimal flow is integer-valued whenever the same is true of the excess and capacity functions.

As with networks in general, we may consider a capacity-constrained version of the problem. Given $K \in (\mathbb{R}_{\geq 0} \cup \{\infty\})^{m \times n}$, let

$$\Pi_K(R, C) := \{X \in \Pi(R, C) : \quad X \leq K \text{ entrywise}\}.$$

We call the integer points of $\Pi_K(R, C)$ **K-bounded contingency tables**. By setting some entries of K equal to zero, we obtain tables representing feasible flows on an arbitrary subgraph of $K_{m,n}$, hence on an arbitrary bipartite (source-sink) graph. In fact, given *any* acyclic (not necessarily bipartite) network on n vertices, there is a bijective encoding of integer feasible flows on that network as contingency tables (see [6]); thus these two objects are essentially equivalent.

Contingency tables arise in the empirical sciences, where they represent the joint distribution of categorical variables (e.g., hair color and eye color) in a sample. The problems of counting and sampling contingency tables are intimately related to statistical significance testing. We will say more about this connection in Section 3.1.

Enumeration of bounded contingency tables is the main "case study" in the present dissertation. For previous work on this subject, see [21], where the complexity of the problem is addressed.

1.1.3 Multi-way tables and flows on hypergraphs

As we have seen, contingency tables can represent the joint distribution of two categorical variables. We can extend this idea to more than two variables. Let $X = \left(x_{j_1 j_2 \cdots j_r}\right)$ be an order-r tensor of dimensions $n_1 \times n_2 \times \cdots \times n_r$. By a *partial index specification* (or p.i.s.), we mean an element of the set

$$\{\cdot, 1, 2, \ldots, n_1\} \times \{\cdot, 1, 2, \ldots, n_2\} \times \cdots \times \{\cdot, 1, 2, \ldots, n_r\},$$

where the symbol '$\cdot$' is understood as an unspecified index. The number of specified indices is called the *order* of the p.i.s. We say that a p.i.s. *masks* all entries $x_{j_1 j_2 \cdots j_r}$

of X whose indices agree with those specified by the p.i.s. The sum of all entries of X masked by a given order-k p.i.s. is called a *k-margin* of X, and a **k-margin r-way contingency table** is defined as a nonnegative integer order-r tensor whose k-margins are equal to some specified values. (Thus an ordinary contingency table is a 1-margin 2-way table.) Dropping the integrality condition, the set of r-way tables with given margins is a polytope, called the **multi-index transportation polytope** [54].

Multi-way tables are poorly behaved; for example, the set of integers obtainable in a given position of a 3-way table with given 2-margins is not necessarily an interval of $\mathbb{Z}$ [24], and the existence of a $3 \times m \times n$ table with given 2-margins is an NP-complete problem. De Loera and Onn [23] put this fact into context by showing that *every* rational polytope is represented with respect to integer points by a multi-index transportation polytope whose points are $3 \times m \times n$ tables with specified 2-margins. Therefore, the problem of counting integer points of polytopes reduces to counting such tables.

A **hypergraph** is a pair (V, E), where V is a set whose elements are called *vertices* and E is a set of subsets of V having arbitrary size, which are known as *edges*. There are multiple notions of directed hypergraphs in the literature. Cambini, Gallo, and Scutellà [19] consider flows on hypergraphs in which each edge has a single "head" but (possibly) several "tails." These flows are again the points of a polytope, but they do not correspond to multi-way tables and we will not consider them further.

1.1.4 Knapsack packings

Even the integer points of a right-angled simplex are of interest, as the following problem shows. Suppose we are going camping with a knapsack which will bear weight $b \in \mathbb{R}_{\geq 0}$. Subject to this limitation, we wish to pack the most useful set of

supplies from a store of n distinct items with weights $a_1, a_2, \ldots, a_n > 0$. If these items are available in unlimited quantity, then the feasible packings are the integer points of the simplex

$$\{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n_{\geq 0} : \quad \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}.$$

We may introduce additional constraints $0 \leq x_i \leq k_i$ to represent finite availability of the items; in this case, the underlying polytope is not a simplex, but a cuboid (i.e., a right-angled parallelepiped) truncated by a hyperplane.[4] Integer points of these polytopes have other interpretations as well, for instance in homology theory [55] and number theory [70]. (Notably, the integer points of the simplex

$$\{\mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d_{\geq 0} : \quad x_1 + 2x_2 + \cdots + dx_d = n\}$$

correspond to **partitions** of the integer n into parts not greater than d.)

The problem of counting knapsack packings is $\#P$-complete in terms of the dimension n or the full input length [68, 35]. Polynomial-time randomized approximation schemes exist [53, 28], whereas the fastest known algorithms which give an exact answer require time exponential in n [56]. Some recent bounds are given in [70].

1.1.5 Perfect matchings of graphs

Given a graph $G = (V, E)$, a **perfect matching** of G is a subset $M \subseteq E$ of the edges such that each vertex $v \in V$ belongs to exactly one edge in M. The indicator functions of perfect matchings of G are the integer points of the polytope

$$\{x \in \mathbb{R}^E_{\geq 0} : \quad Ax = \mathbf{1}_V\}, \tag{1.2}$$

where A is the (unsigned) vertex-edge incidence matrix of G, and $\mathbf{1}_V$ denotes the vector of length $|V|$ with all entries equal to 1. (This polytope should not be confused

[4] To complete the specification of the programming problem we have alluded to, we should assign each item a *value* as well; the objective is to maximize total value over the set of feasible packings. However, we will restrict our attention here to the packings themselves.

with the smaller **perfect matching polytope**, defined as the convex hull of the indicator functions of perfect matchings. A presentation of that polytope is given in a well-known paper of Edmonds [29].)

Given the similarity of polytope (1.2) to the other polytopes we have described, it comes as no surprise that counting perfect matchings is, again, $\#P$-complete [67]. (However, a polynomial-time randomized approximation scheme is given in [44]. Also, the special case of G planar and bipartite is more tractable [46, 65].) This counting problem is of major importance in statistical physics (we cannot hope to encompass the literature here, but see e.g. [58], [51], [50]). Counting also has an application to computing matrix permanents. The permanent of an $n \times n$ matrix $X = \big(x_{ij}\big)$ is defined as

$$\operatorname{per} X := \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)},$$

where S_n is the symmetric group. If X is a 0-1 matrix, then there is a bipartite graph on $n + n$ vertices whose biadjacency matrix is X; the permanent of X is then equal to the number of perfect matchings of that graph. As we will see in Section 3.2, matrix permanents play a role in the enumeration of contingency tables.

Like network flows, perfect matchings may be generalized to hypergraphs, see e.g. [1].

1.1.6 Magic squares, Latin squares, etc.

Among contingency tables, some special margins have attracted interest. Most fundamental are the 2-way tables with margins $R = C = (1, 1, \ldots, 1)$—otherwise known as permutation matrices. The corresponding polytope, $\Pi(\mathbf{1}, \mathbf{1})$, is known as the **Birkhoff polytope**; that the permutation matrices are its vertices is the statement of the *Birkhoff–von Neumann theorem*.

Although the problem of enumerating permutation matrices may be considered safely dead, it has some simple generalizations which, though old, are very much alive. Of the various classes of objects known as **magic squares**, the most basic are $n \times n$ tables with constant margins, $R = C = (t, t, \ldots, t)$.[5] These are discussed in [14], where a quasi-polynomial-time randomized approximation algorithm for the number of magic squares is given. An asymptotic formula appeared in [20].

More general than magic squares are contingency tables with "smooth" margins, a class defined in [13] which includes tables with sufficiently near-constant margins. An algorithm approximately counting such tables is given in [13], and an asymptotic enumeration appears in [11].

A **Latin square of order** n is an $n \times n$ matrix with entries in $\{1, 2, \ldots, n\}$, arranged so that each row and each column contains each of $1, 2, \ldots, n$ exactly once. Latin squares are a basic object in the theory of experimental design; the essential treatise on the subject is [25]. A Latin square of order n contains the same information as an $n \times n \times n$ 3-way table with all 2-margins equal to unity; thus Latin squares are a natural analogue of permutation matrices. However, the obvious analogue of the Birkhoff–von Neumann theorem for these $n \times n \times n$ tables does not hold, since there are non-integer tables with all 2-margins equal to 1 which do not lie in the convex hull of the integer tables with 2-margins equal to 1.

Euler appears to have been the first to investigate the number of Latin squares of order n. The best known upper and lower bounds on this number appear in [69], where they are shown to differ by an $e^{O(n^2)}$ factor. The analysis is improved in [66], where it is shown that the bounds of [69] actually differ by a factor of $e^{O(n \log^2 n)}$. (This same paper proposes a number of conjectures which would improve the error to

[5] These are sometimes called *semi-magic squares* by authors who reserve the term *magic squares* for those whose diagonal sums are equal to their row and column sums.

simply exponential or better, but we are not aware of any strong evidence supporting these claims.)

1.2 The challenge of counting: a brief (and partial) history

One of the oldest results concerning integer points of polytopes is

Theorem I.3 (Pick [57]). *If P is a convex polygon with vertices in $\mathbb{Z}^2$, then*

$$\text{Area}(P) = I + \frac{1}{2}B - 1,$$

where I is the number of interior integer points of P and B is the number of integer points on the boundary of P.

It follows from Pick's theorem that the number of integer points in tP (the dilatation of P by a factor of t) is a polynomial in t. Much of the modern theory of integer points of polytopes stems from the following generalization:

Theorem I.4 (Ehrhart [30]). *Given a lattice polytope $P \subset \mathbb{R}^n$, let*

$$\ell_P(t) := |tP \cap \mathbb{Z}^n|, \quad t \in \mathbb{Z}_{\geq 0}.$$

Then $\ell_P(t)$ is a polynomial in t (now called the **Ehrhart polynomial***).*

The Ehrhart polynomial encodes a wealth of combinatorial information about P. Its degree is the intrinsic dimension of P; its leading coefficient is the volume of P, up to a trivial normalization. As shown by Macdonald [52], for $t \in \mathbb{Z}_{>0}$, the value $|\ell_P(-t)|$ gives the number of integer points in the relative interior of tP. For a good introduction to "Ehrhart theory," the reader is referred to [71].

Using complex analysis, Beck and Pixton [15] computed the Ehrhart polynomial of the Birkhoff polytope, which counts magic squares (see Section 1.1.6). Generalizing

their approach, Baldoni-Silva *et al.* computed the Ehrhart polynomials of transportation and flow polytopes in [2]. Their algorithms are tractable (i.e., polynomial-time) in fixed dimension, but when the dimension n is allowed to vary, they run aground on the fundamental hardness (specifically $\#P$-completeness) of the counting problems which they solve. The same is true of an algorithm of Barvinok [5], which uses a decomposition of P into cones to compute a short rational function representation for a generating function encoding the integer points of P.

Because of this obstacle, there is a need for approximations and bounds on $|P \cap \mathbb{Z}^n|$ which can be computed quickly when n is large. One approach is **Monte Carlo simulation**, which in its most basic form consists of "throwing darts" at the integer points of a low-complexity region Q (such as a box) containing P and observing how often the darts hit integer points of P. Thanks to the law of large numbers, the frequency of "hits" almost surely converges to the ratio $|P \cap \mathbb{Z}^n|/|Q \cap \mathbb{Z}^n|$.

The problem with this method is that, when n is large, this ratio may be so minuscule that the time until the first "hit" is impractically large, to say nothing of the convergence rate! For example, the smallest coordinate-axis-aligned box containing the standard unit simplex

$$\{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n_{\geq 0} : \quad x_1 + x_2 + \cdots + x_n = 1\}$$

is $[0, 1]^n$, which has 2^n integer points; the simplex, by comparison, has $n + 1$ integer points. Clearly, a more refined approach is needed.

We have already mentioned the paper of Dyer [28], which combines dynamic programming with "dart-throwing" to approximately count knapsack packings and contingency tables with a fixed number of rows. The idea may be glossed as follows: For a polytope

$$P = \{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{x} \geq \mathbf{0},\ A\mathbf{x} = \mathbf{b}\}$$

with $A, \mathbf{b}$ integral, we substitute

$$P' = \{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{x} \geq \mathbf{0},\ A'\mathbf{x} = \mathbf{b}'\}$$

where $A', \mathbf{b}'$ are integral, of fixed magnitude (relative to n), and as close to a proportional scaling of $A, \mathbf{b}$ as the preceding conditions will allow. Thus P' may be thought of as a "low-resolution" simulacrum of P whose integer points may be counted via dynamic programming[6] in time depending only on n. Dyer shows that (in the cases he discusses) P and P' have the same number of integer points up to a small factor (e.g., this factor is bounded by $n+1$ in the case of knapsack packings). The tabulated data may then be used to throw darts uniformly at the integer points of P' (which contains P), improving the estimate of the relative error. In the case of knapsack packings and contingency tables with a fixed number of rows, this algorithm is a **fully-polynomial randomized approximation scheme** (or FPRAS), meaning that for any fixed $p \in (0, 1)$, it estimates $|P \cap \mathbb{Z}^n|$ to within a factor of $1 \pm \varepsilon$ with probability p in time polynomial in both n and ε^{-1}. Dyer's method is apparently too weak to produce an FPRAS for contingency tables of arbitrary dimension.

Another randomized approach to integer point enumeration is **Markov chain Monte Carlo** (MCMC) simulation, which aims to sample the integer points of P (almost) uniformly by means of a random walk. Such walks have been constructed, e.g., for Latin squares [41] and for perfect matchings of a bipartite graph [44]; the latter construction proved sufficient for an FPRAS which computes the permanent of a 0-1 matrix. Jerrum, Valiant, and Vazirani showed [45] that approximate counting of the integer points of a polytope is of equivalent complexity to "almost uniform sampling" from that set. The main difficulty of MCMC typically lies not in the

[6] That is, by iteratively solving subproblems—in this case, tabulating a function which counts solutions to truncations of the system $A'\mathbf{x} = \mathbf{b}''$, for $\mathbf{b}'' \leq \mathbf{b}'$.

construction of a random walk, but in establishing a good mixing rate [22]. For a more detailed introduction to MCMC simulation, the reader is directed to [26].

Recently, following up a series of papers [7], [8], [9] suggesting a role for entropy in the enumeration of contingency tables, Barvinok and Hartigan [12] proposed a general approach to integer point counting (and sampling) based on the *maximum entropy principle*. This approach forms the background for the present work, and we discuss it further in Section 2.2.

1.2.1 Objectives and organization of this thesis

One of the principal advantages of Barvinok and Hartigan's maximum-entropy method is its generality. Random walks on integer points (and similar stratagems) are often highly dependent on the special properties of the class of polytopes under observation; although very effective in individual cases, these methods give little idea of how to tackle arbitrary P. Designing and analyzing a random walk on $P \cap \mathbb{Z}^n$ seems to grow in difficulty as the complexity of P increases. In contrast, the Barvinok–Hartigan approach actually produces better estimates for the number of r-way contingency tables as r increases, thanks to central limit-like behavior in the geometry of high-dimensional convex bodies [12].

Our objective in Chapter II is to derive efficiently computable upper bounds on $|P \cap \mathbb{Z}^n|$ using maximum-entropy methods, under very weak assumptions regarding P. We show that if P is presented in standard form with matrix A being $m \times n$ (i.e., P is defined by n linear inequalities and m linear equations), then for m fixed and under mild conditions ensuring that A is "essentially full-rank" and P does not shrink toward the origin, we can bound $|P \cap \mathbb{Z}^n|$ by a computable Gaussian heuristic.

In Chapter III, we refine these methods for application to K-bounded contingency tables (see Section 1.1.2). We show that the logarithm of the number of such tables is

approximated by a concave function of the row and column sums. We give efficiently computable estimators for this function, which we show are asymptotically exact as the dimension of the tables goes to ∞. As an application, we show that for fixed $\kappa \geq 2$ and for sufficiently small row and column margins R and C, the number of contingency tables with these margins and with entries $\leq \kappa$ is greater by an exponential factor than predicted by a heuristic of independence; in other words, the margins are strongly positively correlated. We present numerical evidence that the opposite correlation occurs when R and C are *not* "sufficiently small." Such correlations contribute to the doubts raised by Diaconis and Efron [27] regarding standard χ^2 significance testing for contingency tables; this is discussed further in Section 3.1.

CHAPTER II

Maximum-Entropy Methods

2.1 Independence models

What is the easiest class of polytopes from which to (uniformly) sample integer points? We think the reader will not object if we claim this honor for the axis-aligned *cuboids*, that is to say, the right-angled parallelepipeds formed as the Cartesian product of intervals on the line[1]. Of course, the convenient feature of the integer points of a cuboid is that their coordinates vary independently: if $X = (X_1, X_2, \ldots, X_n)$ is such a point drawn at random, then for $1 \leq j_1 < j_2 < \cdots < j_r \leq n$ and $a_1, a_2, \ldots, a_r \in \mathbb{Z}$, we have

$$\mathbf{Pr}\left[\bigwedge_{i=1}^{r} X_{j_i} = a_i\right] = \prod_{i=1}^{r} \mathbf{Pr}\left[X_{j_i} = a_i\right]. \tag{2.1}$$

Any convex polytope for which this property holds is necessarily a cuboid. Yet for X drawn uniformly from the integer points of an arbitrary polytope $P \subset \mathbb{Z}^n$, we may reasonably ask whether (2.1) holds *approximately*. It is intuitively appealing to guess that this does occur when $\dim P$ is large, $r \ll \dim P$, and the projection of P on coordinates $X_1, \ldots, X_r$ is of full dimension r. For example, given a sufficiently large random contingency table with known margins, one might surmise that there is very little dependence between a small number of entries. Some vague support for

[1] Or in common parlance, *boxes*.

this idea comes from high-dimensional convex geometry. One theme of that subject, emphasized in [3], is that "all convex bodies behave a bit like Euclidean balls," for instance, in that they have either ball-like sections or ball-like projections in low dimension.[2] High-dimensional Euclidean balls do approximately satisfy a version of (2.1): for fixed r, the projection of the uniform measure on the n-dimensional ball to a dimension-r subspace is asymptotic (when appropriately scaled) to the Gaussian measure on $\mathbb{R}^r$, which is the r-fold product of measures on $\mathbb{R}$ (see [4]).

Thus inspired, we propose

Definition II.1. An **independence model** is a random vector $X=(X_1, X_2, \ldots, X_n)$, supported on $\mathbb{Z}^n$, which satisfies (2.1) for all $1 \leq j_1 < j_2 < \cdots < j_r \leq n$ and $a_1, a_2, \ldots, a_r \in \mathbb{Z}$.

The term *model* may strike the reader as premature. We offer the preceding definition with a view toward "fitting" the best independence model to the uniform distribution on the integer points of a polytope P. However, we do not want to build a particular philosophy of "best fit" into the definition at this point.

Nevertheless, in examples with a lot of symmetry, the best independence model may be self-evident. Consider the simplicially truncated cuboid

$$TC(n,r) := \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in [0,1]^n : \quad x_1 + x_2 + \cdots + x_n = r\},$$

whose integer points are all 0-1 vectors with r entries equal to 1 and $n-r$ 0's. If Y is a random point drawn uniformly from that $TC(n,r) \cap \mathbb{Z}^n$, then $Y_1, Y_2, \ldots, Y_n$ are each Bernoulli with support $\{0,1\}$ and expectation r/n. They are not independent, but it is natural to consider an independence model X for Y such that $X_1, X_2, \ldots, X_n$ are also Bernoulli with expectation r/n, but *are* independent. By means of such a

[2] This claim can be made precise for sections or projections of dimension at most $\log \dim P$. However, at the cost of some generality, we will find support for approximate versions of (2.1) when r is not nearly so small as that.

model, we can explicate an estimate for $|TC(n,r) \cap \mathbb{Z}^n|$ which is usually derived from Stirling's formula:

Proposition II.2. *Let n, r be integers ($n > 0,\ 0 \leq r \leq n$), and let s vary in $\mathbb{Z}_{>0}$. Then*

$$\ln \binom{sn}{sr} = sn \cdot h\left(\frac{r}{n}\right) - \Theta(\ln s), \quad [3] \tag{2.2}$$

where $h : [0,1] \to \mathbb{R}$ is the **binary entropy function**[4]

$$h(x) := x \ln\left(\frac{1}{x}\right) + (1-x) \ln\left(\frac{1}{1-x}\right).$$

In order to interpret (and prove) this proposition, we must first acquaint the reader with some concepts from information theory.

2.1.1 Entropy and counting

Entropy is a statistic associated to a random variable and commonly identified with its *information content* (an interpretation which we will not formalize, but which will give some intuitive feel for results to be stated later). Apart from variation in the choice of logarithm base, the definition of entropy is essentially unchanged since its introduction by Claude Shannon in the famous papers [61], [62].

Definition II.3. Let X be a random variable and x a value in the support of X. The **Shannon self-information** of the pair (X, x) is

$$I(X,x) := \ln \frac{1}{\mathbf{Pr}[X = x]}.$$

[3] We adhere to conventional Landau notation. The statement $g(n) = O(f(n))$ means that there exists a constant c such that $|g(n)/f(n)| < c$ for all sufficiently large n. The statement $g(n) = \Theta(f(n))$ means that $g(n) = O(f(n))$ and $f(n) = O(g(n))$. We will also write $g(n) = o(f(n))$ to indicate that $g(n)/f(n) \to 0$, and $g(n) = \Omega(f(n))$ to indicate that $f(n) = O(g(n))$.

[4] A graph is provided in Figure 2.1.

The **entropy** of X is

$$\begin{aligned}\mathbf{H}[X] &:= \mathbf{E}_x[I(X,x)] \\ &= \sum_{x \in \operatorname{supp} X} \mathbf{Pr}[X=x] \ln \frac{1}{\mathbf{Pr}[X=x]}.\end{aligned}$$

If Y is another random variable and y a value in its support, then we define the **conditional entropies**

$$\mathbf{H}[X|Y=y] := \sum_{x \in \operatorname{supp} X} \mathbf{Pr}[X=x|Y=y] \ln \frac{1}{\mathbf{Pr}[X=x|Y=y]}$$

and

$$\begin{aligned}\mathbf{H}[X|Y] &:= \mathbf{E}_y\big[\mathbf{H}[X|Y=y]\big] \\ &= \sum_{y \in \operatorname{supp} Y} \mathbf{Pr}[Y=y]\mathbf{H}[X|Y=y].\end{aligned}$$

We also define the **joint entropy** $\mathbf{H}[X,Y]$ as the entropy of the vector (X,Y).

(We will only be concerned with random variables having discrete support; there are other definitions of entropy for continuous distributions. Note that when X has countably infinite support, the value of $\mathbf{H}[X]$ may be finite or infinite.)

The following properties of entropy are fundamental:

- $\mathbf{H}[X]$ is a concave function of the probability mass function associated to X. In particular, among all distributions on n-point support, the maximum entropy is achieved by the uniform distribution (and is equal to $\ln n$).
- For random variables X and Y, we have $\mathbf{H}[X|Y] = \mathbf{H}[X,Y] - \mathbf{H}[Y] \leq \mathbf{H}[X]$, with equality if and only if X and Y are independent.

For proofs and discussion, see Khinchin's excellent introduction to information theory [47].

Thanks to the first property, if we know the entropy of the *uniform* distribution on a finite set, then we have as good as counted that set. Now let us return to Proposition II.2 and see how this equivalence helps us estimate $\binom{n}{r}$.

The left-hand side of (2.2) is the entropy of a random integer point of $TC(sn, sr)$, drawn uniformly. The dominant term on the right-hand side is the entropy of the corresponding independence model which we discussed earlier.[5] The proposition asserts that the difference between these quantities is small. Although $\binom{sn}{sr}$ grows exponentially with s, the proposition can be used to estimate $\binom{sn}{sr}$ to within polynomial error. The proof, although simple, will serve as a useful prototype when we evaluate other independence models.

Proof of Proposition II.2. Let $X_1, X_2, \ldots$ be independent 0-1 Bernoulli random variables, each with expectation r/n. Let $X = (X_1, \ldots, X_{sn})$.

Observe that if $\mathbf{x}, \mathbf{x}' \in \{0,1\}^{sn}$, then

$$\frac{\mathbf{Pr}[X = \mathbf{x}']}{\mathbf{Pr}[X = \mathbf{x}]} = \left(\frac{r}{n-r}\right)^{|\mathbf{x}'|-|\mathbf{x}|}$$

(where $|\mathbf{x}| := \sum_{i=1}^{sn} x_i$). In particular, all values of X with equal sum of coordinates are equiprobable. Let $\mathbf{x}_*$ denote an arbitrary value of X satisfying $|\mathbf{x}_*| = sr$. Thus

$$\begin{aligned} sn \cdot h\left(\frac{r}{n}\right) = \mathbf{H}[X] &= \mathbf{E}_{\mathbf{x}}[I(X, \mathbf{x})] \\ &= I(\mathbf{x}_*) - \left(\ln \frac{r}{n-r}\right) \mathbf{E}\big[|X| - sr\big] \\ &= I(\mathbf{x}_*) \\ &= -\ln\left[\binom{sn}{sr}^{-1} \cdot \mathbf{Pr}\big[|X| = sr\big]\right] \\ &= \ln\binom{sn}{sr} - \ln \mathbf{Pr}\big[|X| = sr\big]. \end{aligned}$$

[5] As its name suggests, the binary entropy function $h(x)$ is the entropy of a Bernoulli random variable which takes value 1 with probability x and value 0 with probability $1 - x$.

By the local limit theorem of de Moivre and Laplace,

$$\mathbf{Pr}\big[|X| = sr\big] \sim \big[2\pi sn\mathbf{Var}[X_1]\big]^{-1/2} = \left(2\pi s \cdot \frac{r(n-r)}{n}\right)^{-1/2} = \Theta(s^{-1/2}),$$

proving the proposition. □

The proof we have just presented asserts somewhat more than the proposition: it also tells us the asymptotic relative error of the "independence estimate" for $|TC(sn, sr) \cap \mathbb{Z}^{sn}|$. If we express $TC(sn, sr)$ in standard form $A\mathbf{x} = \mathbf{b}$ (an exercise), then this relative error measures the volume of the range of typical variation of AX (where X is the independence model)—a foretaste of things to come.

Remark II.4. How good is the obvious (symmetric) independence model for permutation matrices? This is not an idle question: although we know that there are exactly $n!$ permutation matrices of order n, we do *not* have a good estimate of the number L_n of Latin squares of order n, for which the independence model (per the cubic representation described in Section 1.1.6) is quite similar.

The model we have in mind has n^2 Bernoulli coordinates with support $\{0, 1\}$ and expectation $1/n$. Its entropy thus works out to

$$n^2 h\left(\frac{1}{n}\right) = n[n \ln n - (n-1)\ln(n-1)],$$

whereas the actual entropy of the uniform distribution on permutation matrices is $\ln(n!)$. We may compare the two:

n	$\ln(n!)$	$n^2 h(1/n)$	Difference
2	0.693	2.773	2.079
3	1.792	5.729	3.937
4	3.178	8.997	5.819
5	4.787	12.510	7.723
6	6.579	16.220	9.641

Evidently the predicted entropy and the actual entropy diverge linearly. A calculation with Stirling's formula reveals the error to be equal to $2n - \frac{1}{2} - \ln\sqrt{2\pi n} + o(1)$. Can we account for this? The independence model is a random contingency table with margins of *expected* value 1. In the limit as $n \to \infty$, the margins behave as Poisson random variables of mean 1, and thus each achieves its expected value exactly with probability $\sim 1/e$. There are $2n - 1$ linearly independent margins (not $2n$, because the sum of the row margins and the sum of the column margins are necessarily equal). Thus we might expect the actual number of permutation matrices to differ from the independence estimate roughly by a factor of $e^{-(2n-1)}$—and this is in fact what happens, up to a lower-order term in the exponent. However, it is only thanks to Stirling's formula that we know this for a fact. We cannot justify our estimate of $e^{-(2n-1)}$, because the row margins are not independent from the column margins in the *probabilistic* sense. A theory to justify such estimates is much to be desired, as it holds the promise of estimating L_n to within a simply exponential factor or better.

2.2 The maximum-entropy independence model

In the seminal papers [42], [43], E. T. Jaynes proposed a rule for guessing the probability distribution of a random variable about which one has only partial information. Jaynes' work was motivated by the problem of assigning *prior distributions*

for use in Bayes' rule, which computes updated *posterior* probabilities on the basis of additional observations. Bayesian methods in statistics are controversial because of their explicit reliance on apparently arbitrary "priors,"[6] and many writers have considered how to choose the most neutral (or "non-informative") priors. In the most basic case, where one wishes to assign a distribution on n mutually exclusive events in the absence of any evidence distinguishing them, it is traditional, at least since Laplace, to assign each event a uniform probability of $1/n$. (This is the "Principle of Indifference.") Recall that the uniform distribution on a finite set is the distribution which maximizes entropy. Interpreting entropy as a measure of non-informativeness, Jaynes proposed the following generalization: given the constraints of known data, the best prior is that which attains maximum entropy *subject to those constraints*.[7] Naturally, this rule has come to be known as the Principle of Maximum Entropy. There is a large literature discussing its justification, as well as extensions such as the cross-entropy principle; we suggest the article [38] or the book [60] to the reader interested in these issues.

Suppose $P \subset \mathbb{R}^n$ is a polytope in standard form

$$P := \{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{x} \geq \mathbf{0},\ A\mathbf{x} = \mathbf{b}\}$$

(with A an $m \times n$ matrix). Following Barvinok and Hartigan [12], who were apparently the first to do so, we study the random vector X with maximum entropy subject to two conditions:

- X is supported on $\mathbb{Z}^n_{\geq 0}$, and
- $\mathbf{E}[AX] = \mathbf{b}$ (or, equivalently, $\mathbf{E}[X] \in P$).

[6] A controversy which we feel no need of trying to resolve here.

[7] This principle may be taken in the spirit of Einstein's often-paraphrased remark that "the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience."

The inspiration for this choice is from Jaynes, but we claim no justification for it beyond what we are able to prove about the model.

Definition II.5. The random vector X with the above properties is called a[8] **maximum-entropy independence model** (MEIM) associated to P.

We also wish to define a MEIM for 0-1 polytopes, and more generally for polytopes truncated by a cuboid (which we will consider extensively in Chapter III). Although such polytopes can be written in standard form, doing so comes at the cost of increasing the dimension (via slack variables), which will degrade the quality of the model. Hence the following definitions:

Definition II.6. A **polytope in standard truncated form** is a bounded region of the form

$$\{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{0} \leq \mathbf{x} \leq \mathbf{k},\ A\mathbf{x} = \mathbf{b}\},$$

where $\mathbf{k} \in (\mathbb{Z}_{\geq 0} \cup \{\infty\})^n$, $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and (in)equality of vectors is understood componentwise. Given P a polytope in standard truncated form, let X be the random vector with maximum entropy subject to the conditions $\operatorname{supp} X \subseteq \{\mathbf{x} \in \mathbb{Z}^n : \ \mathbf{0} \leq \mathbf{x} \leq \mathbf{k}\}$ and $\mathbf{E}[AX] = \mathbf{b}$. Then we call X a MEIM associated to P.

Convention II.7. For the remainder of this chapter, we will assume all polytopes are given either in standard form or standard truncated form; if we wish to distinguish between these two cases, we will do so explicitly. We also fix the usage of m, n, $A = \big(a_{ij}\big)$, $\mathbf{b} = (b_1, \ldots, b_m)$, $\mathbf{k} = (k_1, \ldots, k_n)$ (when mentioned in relation to a polytope) according to their usage in Definition II.6, and assume that A always has rank m. We denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

[8] Actually *the* maximum-entropy independence model, as we shall justify shortly.

Per the following basic proposition, every P has a unique MEIM, which is in fact an independence model (as its name suggests):

Proposition II.8. *Let $P \subset \mathbb{R}^n$ be a polytope. Then there exists a unique MEIM $X = (X_1, \ldots, X_n)$ associated to P. Moreover:*

(i) *X is an independence model.*

(ii) *X has constant mass on all integer points of P.*

The existence and uniqueness of X are well-known, while the other properties given above are proved in [12]. Nevertheless, we give our own self-contained proof of the proposition.

Proof. Suppose $Y = (Y_1, \ldots, Y_n)$ is a random vector supported on $\mathbb{Z}^n_{\geq 0}$, such that $\mathbf{E}[Y] \in P$. Let

$$|Y| := \|Y\|_\infty = \max\{Y_1, \ldots, Y_n\}.$$

Since P is bounded, there exists some integer N such that $\mathbf{E}\big[|Y|\big] < N$. By Markov's inequality,

$$\mathbf{Pr}\big[|Y| \geq 2^k N\big] \leq 2^{-k}$$

for each $k = 1, 2, \ldots$. Thus

$$\begin{aligned}
\mathbf{H}[Y] &\leq \ln\big((2N)^n\big) + \frac{1}{2}\ln\big((4N)^n\big) + \frac{1}{4}\ln\big((8N)^n\big) + \cdots \\
&\leq n\ln(2N) + \frac{n}{2}\ln(4N) + \frac{n}{4}\ln(8N) + \cdots \\
&= 2n\ln N + 4n\ln 2.
\end{aligned}$$

In particular, $\mathbf{H}[Y]$ is finite. Entropy is therefore a well-defined function on the space of probability mass functions associated to random variables Y as above. This space is compact, so the entropy attains its maximum, proving the existence of a MEIM

for P. Moreover, the entropy is a strictly concave function of the probability mass function, so the MEIM is unique; we call it X during the remainder of this proof.

Now let $Y = (Y_1, \ldots, Y_n)$ be the independence model such that Y_i is distributed identically to X_i, $1 \leq i \leq n$. Then $\mathbf{E}[AY] = \mathbf{b}$, and

$$\mathbf{H}[Y] = \mathbf{H}[Y_1] + \cdots + \mathbf{H}[Y_n] = \mathbf{H}[X_1] + \cdots + \mathbf{H}[X_n] \geq \mathbf{H}[X],$$

with equality if and only if $X = Y$. Since X was chosen to maximize entropy, it follows that $X = Y$, hence (i).

To see (ii), let Y be a random vector distributed identically to X on points not lying in P, but having constant mass $\mathbf{Pr}[X \in P]/|P \cap \mathbb{Z}^n|$ at each integer point of P. It is clear that $\mathbf{E}[AY] = \mathbf{b}$ and that $\mathbf{H}[Y] \geq \mathbf{H}[X]$. Again, since X was chosen to maximize entropy (subject to the constraint $\mathbf{E}[AX] = \mathbf{b}$), we have $X = Y$. $\square$

2.2.1 The maximum-entropy distribution with a given mean

As we shall see shortly, the coordinates of X are drawn from the following class of distributions.

Definition II.9. Let $\kappa \in \mathbb{Z}_{>0}$. A random variable X is **truncated geometric** with support $\{0, 1, 2, \ldots, \kappa\}$ if there are parameters $p \in (0, 1]$ and $q \in [0, \infty)$, such that

$$\mathbf{Pr}[X = t] = pq^t \qquad \text{for } t = 0, 1, \ldots, \kappa.$$

For symmetry, we also say that X is truncated geometric with parameters $p = 0$ and $q = \infty$ if $\mathbf{Pr}[X = \kappa] = 1$; however, in what follows, explicit treatment of this case will sometimes be left to the reader.

A random variable X on support $\mathbb{Z}_{\geq 0}$ is **geometric** if there are parameters $p \in (0, 1]$ and $q \in [0, 1)$ (in this case necessarily satisfying $p + q = 1$), such that

$$\mathbf{Pr}[X = t] = pq^t \qquad \text{for } t = 0, 1, 2, \ldots.$$

To avoid unnecessary duplication of results, we regard this as a special case of the truncated geometric distribution for which $\kappa = \infty$. (When writing $\{0, 1, 2, \ldots, \kappa\}$, we allow that $\kappa = \infty$, in which case $\{0, 1, 2, \ldots, \kappa\}$ is to be interpreted as $\mathbb{Z}_{\geq 0}$.)

Proposition II.10. *Given $\kappa \in \mathbb{Z}_{\geq 0}$ and $x \in [0, \kappa]$, or given $\kappa = \infty$ and $x \in [0, \infty)$, there is a unique truncated geometric distribution with support $\{0, 1, 2, \ldots, \kappa\}$ and expected value equal to x.*

Proof. Let X denote the truncated geometric distribution on $\{0, 1, 2, \ldots, \kappa\}$ with parameters p, q. These parameters satisfy

$$1 = p(1 + q + q^2 + \cdots + q^\kappa)$$

if $\kappa < \infty$, or

$$1 = p(1 + q + q^2 + \cdots)$$

if $\kappa = \infty$; thus p is determined by q, so the truncated geometric distributions on $\{0, 1, 2, \ldots, \kappa\}$ form a family of one parameter (q). It is clear that $\mathbf{E}[X]$ is a strictly increasing (hence one-to-one) function of q, with range $[0, \kappa]$ (or $[0, \infty)$ if $\kappa = \infty$). Thus for the given x, there is a unique choice of q so that $\mathbf{E}[X] = x$. $\square$

Definition II.11. Let κ and x be as in the previous proposition. We denote the truncated geometric distribution on $\{0, 1, 2, \ldots, \kappa\}$ with expected value x by $TG(x; \kappa)$, its parameters p, q by $p(x; \kappa)$ and $q(x; \kappa)$, and its entropy by $H_\kappa^{\max}(x)$.

The parameters $p = p(x; \kappa)$ and $q = q(x; \kappa)$ are given implicitly by the equations

$$1 = p(1 + q + q^2 + \cdots + q^\kappa), \tag{2.3}$$

$$x = p(q + 2q^2 + \cdots + \kappa q^\kappa), \tag{2.4}$$

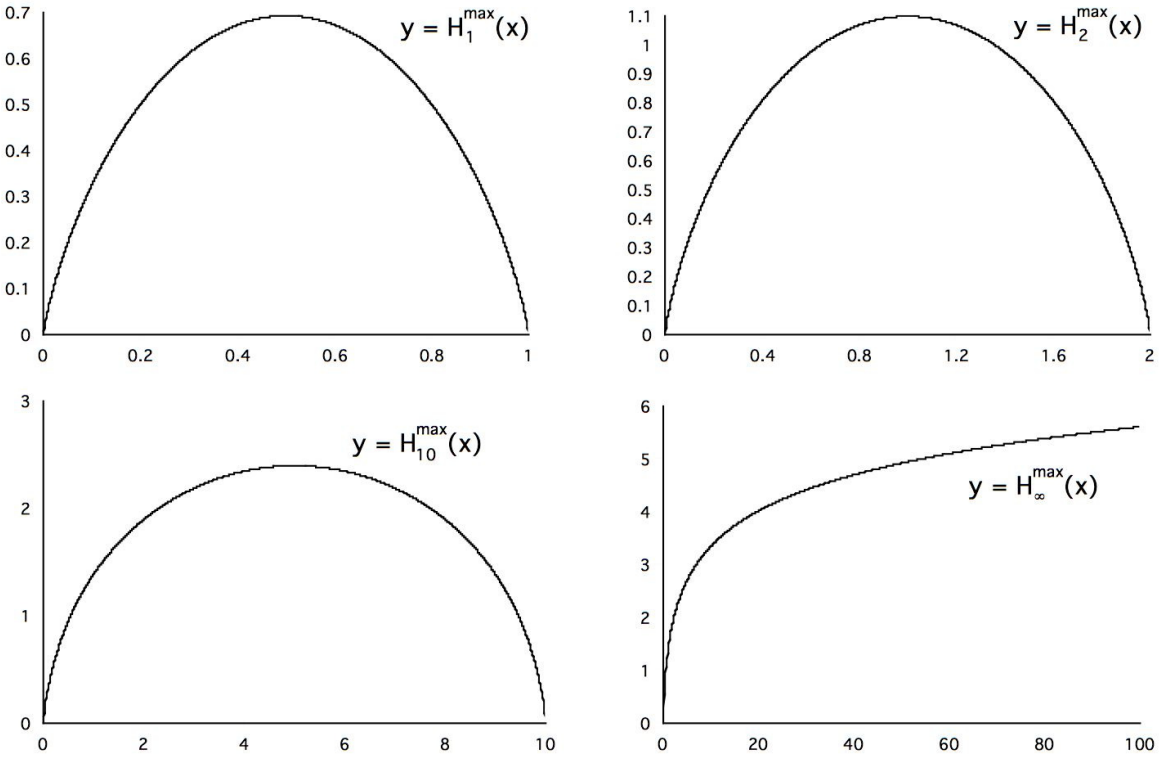

Figure 2.1: Graphs of $H_\kappa^{\max}(x)$, $\kappa = 1, 2, 10, \infty$

which, to the author's knowledge, cannot be neatly solved in general. There are, however, simple expressions when $\kappa = 1$ or $\kappa = \infty$:

(2.5)

$$H_1^{\max}(x) = -x \ln x - (1-x)\ln(1-x) \qquad p(x;1) = 1 - x \qquad q(x;1) = \frac{x}{1-x}$$

(2.6)

$$H_\infty^{\max}(x) = (x+1)\ln(x+1) - x \ln x \qquad p(x;\infty) = \frac{1}{x+1} \qquad q(x;\infty) = \frac{x}{x+1}$$

(We've seen $H_1^{\max}$ before, under the name "binary entropy"; cf. Proposition II.2.)

Proposition II.12. *Among all probability distributions supported in* $\{0, 1, 2, \ldots, \kappa\}$ *and having expected value* x*, the greatest entropy is attained by* $TG(x;\kappa)$*.*

Proof. By Proposition II.8, there exists a maximum-entropy distribution X on $\{0, 1, 2, \ldots, \kappa\}$ with expected value x. For $t \in \{0, 1, 2, \ldots, \kappa\}$, let $p_t := \mathbf{Pr}[X = t]$.

We have

$$\mathbf{H}[X] = \sum_{t=0}^{\kappa} p_t \ln\left(\frac{1}{p_t}\right).$$

Let us regard the expression on the right-hand side as a function of $p_0, p_1, \ldots, p_\kappa$. Its partial derivatives are finite where all $p_t > 0$, but its partial derivative with respect to p_t is $+\infty$ where $p_t = 0$. It follows that, for the maximum-entropy distribution, all $p_t > 0$. Introducing Lagrange multipliers for the relations (2.3), (2.4), we determine that $(\ln p_0, \ln p_1, \ldots, \ln p_\kappa)$ is a linear combination of the vectors $(1, 1, \ldots, 1)$ and $(0, 1, 2, \ldots, \kappa)$. Thus $p_0, p_1, \ldots, p_\kappa$ are in geometric progression. □

Corollary II.13. *Let $P \subset \mathbb{R}^n$ be a polytope in standard (truncated) form, and let $X = (X_1, \ldots, X_n)$ be its associated MEIM. Then each coordinate X_j has truncated geometric distribution.*

Proof. Immediate from Proposition II.12. □

Corollary II.13 does not fully characterize the maximum-entropy independence model for P. There is a unique independence model $X = (X_1, \ldots, X_n)$ with truncated geometric coordinates for each value of $\mathbf{E}[X]$. We know $\mathbf{E}[X] \in P$, so we can take the polytope P itself as a parameter space for the distribution of X; our objective is to maximize $\mathbf{H}[X]$. To see why this is feasible, we now study the entropy of $TG(x; \kappa)$ as a function of x.

2.2.2 The function $H_\kappa^{\max}$

Proposition II.14 (Properties of $H_\kappa^{\max}$)**.** *Let $p = p(x; \kappa)$, $q = q(x; \kappa)$. Then:*

(i) *$H_\kappa^{\max}$ is strictly concave on its domain.*

(ii) $H_\kappa^{\max}(x) = -[\ln p + x \ln q]$.

(iii) *For* $0 < x < \kappa$, $\frac{d}{dx} H_\kappa^{\max}(x) = -\ln q$.

Proof. First we prove claim (i). Let $x, y \in [0, \kappa]$ and $\alpha, \beta > 0$ such that $\alpha + \beta = 1$. We wish to prove that

$$H_\kappa^{\max}(\alpha x + \beta y) > \alpha H_\kappa^{\max}(x) + \beta H_\kappa^{\max}(y).$$

Let X and Y be independent random variables with distributions $TG(x;\kappa)$ and $TG(y;\kappa)$, respectively. Define a random variable Z whose distribution is a *mixture* of X and Y with weights α and β; that is,

$$\mathbf{Pr}[Z = t] = \alpha p(x;\kappa) q(x;\kappa)^t + \beta p(y;\kappa) q(y;\kappa)^t \qquad \text{for } t = 0, 1, \ldots, \kappa.$$

Then

$$\mathbf{E}[Z] = \alpha x + \beta y$$

and

$$\mathbf{H}[Z] > \alpha \mathbf{H}[X] + \beta \mathbf{H}[Y]$$

(since entropy is well-known to be strictly concave with respect to mixture). But

$$H_\kappa^{\max}(\alpha x + \beta y) \geq \mathbf{H}[Z],$$

since $H_\kappa^{\max}(\alpha x + \beta y)$ is the maximum entropy achieved by any random variable supported on $\{0, 1, 2, \ldots, \kappa\}$ with expectation $\alpha x + \beta y$. This concludes the proof of (i).

Claim (ii) is the result of a simple calculation:

$$\begin{aligned}
H_\kappa^{\max}(x) &= -[p \ln p + pq \ln(pq) + pq^2 \ln(pq^2) + \cdots + pq^\kappa \ln(pq^\kappa)] \\
&= -[p \ln p + pq(\ln p + \ln q) + pq^2(\ln p + 2 \ln q) + \cdots + pq^\kappa(\ln p + \kappa \ln q)] \\
&= -[(p + pq + pq^2 + \cdots + pq^\kappa)(\ln p) + (pq + 2pq^2 + \cdots + \kappa pq^\kappa)(\ln q)] \\
&= -[\ln p + x \ln q],
\end{aligned}$$

where we have used equations (2.3), (2.4) in the last step.

Differentiating this formula with respect to x, and again applying equations (2.3) and (2.4), we obtain

$$\begin{aligned}
(H_\kappa^{\max})'(x) &= -\frac{p'}{p} - x \cdot \frac{q'}{q} - \ln q \\
&= p \cdot \left(\frac{1}{p}\right)' - p(q + 2q^2 + \cdots + \kappa q^\kappa) \cdot \frac{q'}{q} - \ln q \\
&= p \cdot \left(\frac{1}{p}\right)' - pq'(1 + 2q + \cdots + \kappa q^{\kappa-1}) - \ln q \\
&= p \cdot \left(\frac{1}{p}\right)' - p \cdot \left(\frac{1}{p}\right)' - \ln q \\
&= -\ln q.
\end{aligned}$$

This proves claim (iii). □

The entropy of an independence model $(X_1, \ldots, X_n)$ with truncated geometric coordinates is equal to

$$\sum_{j=1}^{n} H_\kappa^{\max}(z_j),$$

where $z_j := \mathbf{E}[X_j]$. The entropy is thus a strictly concave function of the parameters $z_1, \ldots, z_n$, which are located in domain P; such a function can be maximized in polynomial time by interior point methods, as mentioned in [12].

In Proposition II.8 (ii), we showed that the MEIM of a polytope in standard (truncated) form has constant mass on the integer points of that polytope. Now we determine this mass. First, however, we append the following notation to the aforementioned Conventions II.7:

Convention II.15. Let $P \in \mathbb{R}^n$ be a polytope in standard truncated form and

$X = (X_1, \ldots, X_n)$ its associated MEIM. Then we write

$$z_j := \mathbf{E}[X_j],$$
$$p_j := p(z_j; k_j), \quad \text{and}$$
$$q_j := q(z_j; k_j).$$

Then we have

Proposition II.16. *Observe Conventions II.7 and II.15. Then for every* $\mathbf{x} \in P \cap \mathbb{Z}^n$*, we have*

$$\mathbf{Pr}[X = \mathbf{x}] = e^{-\mathbf{H}[X]}.$$

Proof. Let $\mathbf{x} \in P \cap \mathbb{Z}^n$. Let $\mathbf{z} := \mathbf{E}[X] = (z_1, \ldots, z_n)$, and set $\mathbf{u} := \mathbf{x} - \mathbf{z} \in \ker A$.

The distribution of X depends on $\mathbf{z}$. Regarding $\mathbf{H}[X]$ as a function of $\mathbf{z}$, we have

$$\frac{\partial}{\partial z_j}\mathbf{H}[X] = -\ln q_j \tag{2.7}$$

by Proposition II.14 (iii). Since X is the independence model of maximum entropy subject to $\mathbf{E}[AX] = \mathbf{b}$, it follows that $\mathbf{H}[X]$ has zero directional derivative in any direction belonging to $\ker A$. Thus by (2.7), we have $\sum_j u_j \ln q_j = 0$ and hence $\prod_j q_j^{u_j} = 1$.

It follows that

$$\begin{aligned}\mathbf{Pr}[X = \mathbf{x}] &= \prod_{j=1}^{n} p_j q_j^{x_j} \\ &= \left(\prod_{j=1}^{n} p_j q_j^{z_j}\right)\left(\prod_{j=1}^{n} q_j^{u_j}\right) \\ &= e^{-\mathbf{H}[X]},\end{aligned}$$

where the last equality follows from Proposition II.14 (ii). □

2.3 Upper bounds on $|P \cap \mathbb{Z}^n|$

Proposition II.16 implies that

$$|P \cap \mathbb{Z}^n| = e^{\mathbf{H}[X]}\mathbf{Pr}[X \in P], \tag{2.8}$$

where X is the MEIM associated to P. The factor $e^{\mathbf{H}[X]}$ is efficiently computable, so, for the remainder of the chapter, our objective is to estimate $\mathbf{Pr}[X \in P]$. In [12], Barvinok and Hartigan consider a Gaussian heuristic for this factor, which can be proven to give good results for certain special classes of polytopes: for example, they use it to produce an asymptotic formula for the number of r-way contingency tables, $r \geq 5$, with given 1-margins. However, the general effectiveness of the Gaussian heuristic is unclear. By contrast, we present some definite upper bounds on $\mathbf{Pr}[X \in P]$ which pertain to a very general range of polytopes, including all of the standard (nontruncated) polytopes surveyed in Section 1.1.[9] We make use of the following concept:

Definition II.17. The **point concentration** of a discrete random variable Y is

$$\operatorname{conc}(Y) := \max_{y \in \operatorname{supp} Y} \mathbf{Pr}[Y = y].$$

An upper bound on $\operatorname{conc}(AX)$ is, necessarily, also an upper bound on $\mathbf{Pr}[AX = \mathbf{b}] = \mathbf{Pr}[X \in P]$. Therefore, we have

$$|P \cap \mathbb{Z}^n| \leq e^{\mathbf{H}[X]} \operatorname{conc}(AX). \tag{2.9}$$

It is convenient to use $\operatorname{conc}(AX)$ (i.e., concentration at the mode) as a proxy for $\mathbf{Pr}[X \in P]$ (concentration at the mean). *A priori,* there seems to be no reason to expect a large difference between the two.

[9] In fairness, none of these bounds come remotely as close to the correct count as the Gaussian heuristic does in the cases in which the latter is known to be effective; so there is an apparent trade-off, for the time being, between generality and accuracy.

2.3.1 Anti-concentration and the Littlewood-Offord problem

The concentration of sums of random variables is such a basic and richly studied subject that it would be folly to attempt a history of it here. Instead, we will confine our remarks to the particular project of obtaining *upper* bounds on concentration (sometimes called "anti-concentration" results), and especially the precedents for the upper bounds to be presented here.

First in this line is the **Littlewood-Offord problem**, which asked for the maximum point concentration of

$$\varepsilon_1 a_1 + \varepsilon_2 a_2 + \cdots + \varepsilon_n a_n$$

when $a_1, a_2, \ldots, a_n$ are nonzero integers and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are symmetric Bernoulli random variables. (In fact, Littlewood and Offord asked, equivalently, how many subsums of $a_1 + a_2 + \cdots + a_n$ may coincide.) Unsurprisingly, the maximum concentration is achieved when $a_1 = \cdots = a_n$, in which case the concentration is of order $O(n^{-1/2})$ (of course, we may write down the exact formula as well). The proof of this fact, using poset theory, is due to Erdős [32].

Halász [39] extended this result to random sums

$$\varepsilon_1 \mathbf{a}_1 + \varepsilon_2 \mathbf{a}_2 + \cdots + \varepsilon_n \mathbf{a}_n$$

of m-vectors (again with symmetric Bernoulli coefficients), obtaining a bound of order $O(n^{-m/2})$—consistent with the behavior of a Gaussian distribution—under conditions ensuring that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are reasonably "spread out" in $\mathbb{R}^m$ (i.e., not excessively close to a proper subspace). As stated in [39], Halász's results actually pertain to the small ball concentration of $\varepsilon_1 \mathbf{a}_1 + \varepsilon_2 \mathbf{a}_2 + \cdots + \varepsilon_n \mathbf{a}_n$, but can be specialized to point concentration by a scaling argument. These results, which Halász

proved using a Fourier-theoretic lemma of Esséen, were subsequently reproduced by Oskolkov [40, notes by Howard], who gave a simpler proof using **rearrangement inequalities**. Here is the precise result of Halász:

Theorem II.18 (Halász [39])**.** *Let* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. *Let* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ *be independent symmetric Bernoulli random variables, and let*

$$S := \varepsilon_1 \mathbf{a}_1 + \varepsilon_2 \mathbf{a}_2 + \cdots + \varepsilon_n \mathbf{a}_n.$$

Define

$$\mathrm{conc}_1(S) := \max_{\mathbf{y} \in \mathbb{R}^m} \mathbf{Pr}[|S - \mathbf{y}| < 1].$$

Suppose that there exists a constant $\delta > 0$ *such that for any* $|\mathbf{e}| = 1$ *one can select at least* δn *vectors* $\mathbf{a}_k$ *with* $|\langle \mathbf{a}_k, \mathbf{e} \rangle| \geq 1$. *Then*

$$\mathrm{conc}_1(S) \leq c(\delta, m) n^{-m/2},$$

where $c(\delta, m)$ *depends only on* δ *and* m.

Our stated problem of bounding $\mathbf{Pr}[AX = \mathbf{b}]$ (for $A, X, \mathbf{b}$ in accord with Conventions II.7 and II.15) is essentially the problem Halász solved, except that the coefficients ε_j are replaced by geometric (or truncated geometric) random variables. This is not a trivial distinction: symmetric Bernoulli random variables are all alike, having concentration $1/2$, whereas the concentration of our $X_1, \ldots, X_n$ depends on $\mathbf{E}[X]$. We should expect a result similar to that of Halász, but with constant depending on $z_1, \ldots, z_n$ as well as m and δ (or an analogous parameter). This expectation is realized in Theorem II.37 (which we call the **H-bound** in recognition of Halász).[10] Its proof, which is the major undertaking of this chapter, owes much to the method of Oskolkov [40].

[10] In fact, Halász also gave a result (Theorem 4 in [39]) which applies to random sums with coefficients of arbitrary distribution, but in the case of $X_1, \ldots, X_n$ truncated geometric, the constant in Halász's result is generally very poor compared to the constant we will obtain. See Remark II.49.

Before coming to the H-bound, we propose two simpler (but somewhat more specialized) upper bounds on $\mathrm{conc}(AX)$. One of these, the **I-bound** (Theorem II.19), is designed to show the influence of the parameters $z_1, \ldots, z_n$ as plainly as possible. This bound is easy to compute, easy to understand, and almost trivial to prove, all at the cost of neglecting the large-n central limit phenomena captured by the H-bound. The I-bound is obtained by discarding all columns of A except a linearly independent set (hence the letter "I"), and is thus maximally effective when $n-m$ is small. Our other result, the **E-bound**, is an adaptation of Erdős's Littlewood-Offord result (and his poset-theoretic methods) to the case of geometric random variables, or, more generally, to random variables with individually bounded concentration. Essentially effective only in the case $m = 1$ (for reasons to be discussed), the E-bound may be trivially extended to the case $m > 1$ when A has only m *distinct* columns up to scaling, which form a basis for $\mathbb{R}^m$. We state the E-bound in this form (Theorem II.25). Although limited, it has application to counting knapsack packings (see Section 1.1.4).

2.4 The I-bound

Theorem II.19 (I-bound)**.** *Assume Conventions II.7 and II.15, with P in standard form.*[11] *Then*

$$\begin{aligned} |P \cap \mathbb{Z}^n| &\leq e^{\mathbf{H}[X]} \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} (1-q_{j_1})(1-q_{j_2})\cdots(1-q_{j_m}) \\ &= e^{\mathbf{H}[X]} \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} \prod_{i=1}^{m} \frac{1}{z_{j_i}+1}. \end{aligned}$$

Remark II.20. The selection of indices $j_1, \ldots, j_m$ which minimize $\prod_{i=1}^{m}(1-q_{j_i})$ is an instance of choosing a minimum-cost base of a matroid. This problem is solved by

[11] For the remainder of this chapter, we generally only treat polytopes in standard (nontruncated) form, although we expect similar results can be derived for polytopes in standard truncated form. We will revisit truncated polytopes in Chapter III.

the greedy algorithm: for $i = 1, \ldots, m$ in turn, we choose j_i such that q_{j_i} is maximal under the constraint that $a_{j_i} \notin \operatorname{span}\{a_{j_1}, \ldots, a_{j_{i-1}}\}$. Thus the I-bound is easy to compute.[12]

We prove Theorem II.19 by means of the following simple fact:

Lemma II.21. *If X, Y are independent discrete random variables, then*

$$\operatorname{conc}(X + Y) \leq \operatorname{conc}(X).$$

Proof. Observe that $\operatorname{conc}(X+Y)$ is a weighted average of values of the probability mass function of X, of which the largest is $\operatorname{conc}(X)$. □

Proof of Theorem II.19. By Lemma II.21 and the previously mentioned properties of geometric random variables,

$$\begin{aligned}
\operatorname{conc}(X_1\mathbf{a}_1 + \cdots + X_n\mathbf{a}_n) &\leq \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} \operatorname{conc}(X_{j_1}\mathbf{a}_{j_1} + \cdots + X_{j_m}\mathbf{a}_{j_m}) \\
&\leq \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} \mathbf{Pr}[X_{j_1} = \cdots = X_{j_m} = 0] \\
&= \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} (1 - q_{j_1})(1 - q_{j_2}) \cdots (1 - q_{j_m}) \\
&= \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} \prod_{i=1}^{m} \frac{1}{z_{j_i} + 1}.
\end{aligned} \tag{2.10}$$

By (2.9), it follows that

$$\begin{aligned}
|P \cap \mathbb{Z}^n| &\leq e^{\mathbf{H}[X]} \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} (1 - q_{j_1})(1 - q_{j_2}) \cdots (1 - q_{j_m}) \\
&= e^{\mathbf{H}[X]} \min_{\substack{\mathbf{a}_{j_1},\ldots,\mathbf{a}_{j_m} \\ \text{lin.indep.}}} \prod_{i=1}^{m} \frac{1}{z_{j_i} + 1}. \qquad \blacksquare
\end{aligned}$$

[12] This was pointed out by Alexander Barvinok (private communication).

Remark II.22. Perhaps inequality (2.10) can be improved by a factor on the order of $n^{-m/2}$ under conditions guaranteeing that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are sufficiently well-distributed in $\mathbb{R}^m$. This seems to the author the most promising path toward unification of the ideas behind the I- and H-bounds.

2.4.1 The symmetrized I-bound

We also prove a "symmetrized" version of the I-bound:

Theorem II.23. *Let $I_1, I_2 \ldots, I_p$ be m-element subsets of $\{1, 2, \ldots, n\}$,*

$$I_k = \{j_{k1}, j_{k2}, \ldots, j_{km}\},$$

such that $\mathbf{a}_{j_{k1}}, \ldots, \mathbf{a}_{j_{km}}$ form a basis for $\mathbb{R}^m$ $(1 \leq k \leq p)$, and such that $I_1 \cup I_2 \cup \cdots \cup I_p = \{1, 2, \ldots, n\}$. Then

$$|P \cap \mathbb{Z}^n| \leq e^{\mathbf{H}[X]} \left(\frac{1}{\mathbf{E}[\bar{X}] + 1} \right)^m,$$

where $\bar{X}$ is a geometrically distributed random variable with entropy equal to $\frac{1}{pm}\mathbf{H}[X]$.

(Cf. (2.6) for a formula giving $\mathbf{H}[X]$ in terms of $\mathbf{E}[X]$. The inverse is apparently not elementary, but is easy to compute in practice.)

Underlying Theorem II.23 is the following observation:

Lemma II.24. *Among all vectors $Y := (Y_1, Y_2, \ldots, Y_m)$ of independent, geometrically distributed random variables with fixed joint entropy Ω, the highest concentration* $\mathrm{conc}(Y)$ *is achieved when $Y_1, Y_2, \ldots, Y_m$ are identically distributed.*

Proof. Since Y_i is geometrically distributed $(1 \leq i \leq m)$, there exist parameters $r_i \in [0, 1)$ such that

$$\mathbf{Pr}[Y_i = k] = (1 - r_i) r_i^k \quad \text{for} \quad k \in \mathbb{Z}_{\geq 0}.$$

The concentration of Y is $\prod_{i=1}^{m}(1-r_i)$, so we must show that this expression is maximized (for fixed Ω) when $r_1 = \ldots = r_m$.

We introduce the changes of variable $s_i := \frac{1}{1-r_i}, \quad t_i := \ln s_i$. (Thus $1 - r_i = \frac{1}{s_i}$, and $s_i = e^{t_i}$, where $t_i \in [0, \infty)$.) Also, let

$$\omega(t) := (1 - e^t)\ln(1 - e^{-t}) + t.$$

Now

$$\begin{aligned}
\Omega &= \sum_{i=1}^{m} \frac{r_i}{1-r_i} \ln \frac{1}{r_i} + \ln \frac{1}{1-r_i} \\
&= \sum_{i=1}^{m} (s_i - 1) \ln \frac{s_i}{s_i - 1} + \ln s_i \\
&= \sum_{i=1}^{m} (e^{t_i} - 1) \ln \frac{e^{t_i}}{e^{t_i} - 1} + t_i \\
&= \sum_{i=1}^{m} (1 - e^{t_i}) \ln(1 - e^{-t_i}) + t_i \\
&= \sum_{i=1}^{m} \omega(t_i),
\end{aligned}$$

and

$$\prod_{i=1}^{m}(1-r_i) = \exp\Big(-\sum_{i=1}^{m} t_i\Big).$$

The following three statements are equivalent:

(i) For Ω fixed, $\prod_i (1 - r_i)$ is maximized when $r_1 = \cdots = r_m$.

(ii) For Ω fixed, $\sum_i t_i$ is minimized when $t_1 = \ldots = t_m$.

(iii) If $\sum_i t_i$ is fixed and Ω free to vary, then Ω is maximized when $t_1 = \ldots = t_m$.

The equivalence of statements (i) and (ii) is clear. To see that (ii) and (iii) are equivalent, it is enough to observe that Ω is increasing with respect to each of $t_1, \ldots, t_m$. Thus to prove (i), which is the assertion of the lemma, it will suffice for us to prove (iii).

Writing $s := e^t$, we obtain

$$\begin{aligned}\frac{d\omega}{dt} &= (1-e^t)\left(\frac{e^{-t}}{1-e^{-t}}\right) - e^t \ln(1-e^{-t}) + 1 \\ &= -e^t \ln(1-e^{-t})\end{aligned}$$

and

$$\begin{aligned}\frac{d^2\omega}{dt^2} &= -e^t \cdot \frac{e^{-t}}{1-e^{-t}} - e^t \ln(1-e^{-t}) \\ &= -\frac{1}{1-\frac{1}{s}} - s\ln\left(1-\frac{1}{s}\right) \\ &= -\frac{s}{s-1} + s\ln\frac{s}{s-1} \\ &= -s\left(\frac{1}{s-1}\right) + s\ln\left(1+\frac{1}{s-1}\right) \\ &\leq 0,\end{aligned}$$

since $\ln(1+x) \leq x$ for $x \geq 0$. This shows that $\omega(t)$ is concave for $t \geq 0$, which implies (iii) and so completes the proof of the lemma. □

Proof of Theorem II.23. For $I \subset \{1, 2, \ldots, n\}$, let $\mathbf{H}[X_I]$ denote the joint entropy of $\{X_j : j \in I\}$. Since $X_1, \ldots, X_n$ are independent, we have $\mathbf{H}[X_I] = \sum_{j\in I} \mathbf{H}[X_j]$.

Since the sets $I_1, I_2, \ldots, I_p$ cover $\{1, 2, \ldots, n\}$, we have

$$\mathbf{H}[X] \leq \sum_{k=1}^{p} \mathbf{H}[X_{I_k}],$$

and thus by the pigeonhole principle

$$\mathbf{H}[X_{I_k}] \geq \frac{1}{p}\mathbf{H}[X]$$

for some $k \in \{1, \ldots, p\}$. By Lemma II.24, the concentration of the vector $(X_{j_{k1}}, \ldots, X_{j_{km}})$ is maximized when $X_{j_{k1}}, \ldots, X_{j_{km}}$ are identically distributed. In this case, each has entropy equal to $\frac{1}{m}\mathbf{H}[X_{I_k}]$, which is greater than or equal to

$\mathbf{H}[\bar{X}] = \frac{1}{pm}\mathbf{H}[X]$; we pause to recall that the entropy and the expectation of a geometric random variable are monotonically increasing functions of one another. Thus (as in the proof of Theorem II.19),

$$\begin{aligned}\operatorname{conc}(AX) &\leq \operatorname{conc}(X_{j_{k1}}\mathbf{a}_{j_{k1}} + \cdots + X_{j_{km}}\mathbf{a}_{j_{km}}) \\ &\leq \left(\frac{1}{\mathbf{E}[\bar{X}]+1}\right)^m.\end{aligned}$$

The theorem follows by (2.9). ■

2.5 Sperner theory and the E-bound

We now turn to the following Erdős-inspired bound:

Theorem II.25 (E-bound)**.** *Assume Conventions II.7 and II.15, with P in standard form. Let N be an integer such that $2 \leq \mathbf{E}[X_j] < N$ for $j = 1, 2, \ldots, n$. Additionally, suppose that $n = pm$ for some integer p and that, for each $i = 1, 2, \ldots, m$, we have $\frac{\mathbf{a}_i}{\|\mathbf{a}_i\|} = \frac{\mathbf{a}_{m+i}}{\|\mathbf{a}_{m+i}\|} = \frac{\mathbf{a}_{2m+i}}{\|\mathbf{a}_{2m+i}\|} = \cdots = \frac{\mathbf{a}_{(p-1)m+i}}{\|\mathbf{a}_{(p-1)m+i}\|}$, where $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ is a basis for $\mathbb{R}^m$. (That is to say, the columns of A cycle through a basis of $\mathbb{R}^m$ periodically, up to scaling.)*

Then for fixed m and N, we have

$$|P \cap \mathbb{Z}^n| \leq (1 + o(1))e^{\mathbf{H}[X]} \prod_{i=1}^{m} \left(\frac{\pi}{6}\sum_{t=1}^{p}\left(\lfloor \mathbf{E}[X_{(t-1)m+i}] + 1\rfloor^2 - 1\right)\right)^{-1/2}$$

as $p \to \infty$.

The E-bound is actually just the application to polytopes of a more general concentration result, Theorem II.27. To state this result, we must introduce some notions from the branch of poset theory known as **Sperner theory**.[13]

Definitions II.26. Let S be a finite poset (partially ordered set) and $x, y \in S$. We say that x covers y if $x > y$ and if $x \geq z \geq y \Rightarrow z \in \{x, y\}$.

[13] To our knowledge, the most complete handbook on this still-evolving subject is Engel [31].

A **rank function** on S is a function $\mathrm{rk} : S \to \mathbb{Z}_{\geq 0}$, such that for all $x, y \in S$, if x covers y, then $\mathrm{rk}(x) = \mathrm{rk}(y) + 1$. A **ranked poset** is a pair (S, rk) where S is a poset and rk is a rank function on S. (By abuse of notation, we also call S a ranked poset when there is no ambiguity about the rank function.) We say that $\mathrm{rk}(x)$ is the rank of element x. A **layer** of a ranked poset is a level set of the rank function.

We denote by $[N]$ the **chain** (i.e., totally ordered set) of cardinality N together with the unique rank function which assigns its least element rank 0. If (S, rk) and (S', rk') are ranked posets, then $S \times S'$ is a ranked poset with rank function $\mathrm{rk} + \mathrm{rk}'$.

An **antichain** in a poset is a collection of pairwise incomparable elements. The **width** of a poset S, denoted by $w(S)$, is the cardinality of its largest antichain(s). The i^{th} **Whitney number** W_i of a ranked poset is the cardinality of its layer of rank i. If the width of a ranked poset is equal to its largest Whitney number, then we say that the poset has the **Sperner property**.

For instance, the "Boolean cube"[14] $[2] \times [2] \times [2]$ has Whitney numbers $W_0 = 1$, $W_1 = 3$, $W_2 = 3$, $W_3 = 1$ and width 3, so it has the Sperner property. Note that the width of any poset is greater than or equal to its largest Whitney number, because all layers are necessarily antichains.

Now we are ready to state

Theorem II.27. *Let $X_1, X_2, \ldots, X_p$ be independent, integer-valued random variables such that*

$$\mathrm{conc}(X_j) \leq \frac{1}{N_j} \qquad \text{for } 1 \leq j \leq p,$$

where $N_1, N_2, \ldots, N_p$ are positive integers. Then

$$\mathrm{conc}(X_1 + \cdots + X_p) \leq \frac{w\big([N_1] \times \cdots \times [N_p]\big)}{N_1 N_2 \cdots N_p}.$$

[14] A stock example.

Moreover, given any fixed N such that $2 \leq N_1, N_2, \ldots, N_p < N$, we have

$$\frac{w\big([N_1] \times \cdots \times [N_p]\big)}{N_1 N_2 \cdots N_p} \sim \Big(\frac{\pi}{6} \sum_{j=1}^{p} (N_j^2 - 1)\Big)^{-1/2}$$

as $p \to \infty$.

This theorem will be easiest to prove under the assumption that each X_j is uniformly supported on N_j points (with mass $1/N_j$ at each). To justify passing to this case, we will use the following definition and the two lemmas after it:

Definition II.28. A discrete random variable is a **mixture** of random variables $Y_1, Y_2, \ldots$ if its probability mass function lies in the convex hull of the probability mass functions of $Y_1, Y_2, \ldots$.

Lemma II.29. *Let Y be a random variable, supported on $\mathbb{Z}_{\geq 0}$, such that $\mathrm{conc}(Y) \leq \frac{1}{N}$. Then Y can be written as a mixture of random variables $Y_1, Y_2, \ldots$, such that each Y_k is uniformly supported on N points, i.e., has an N-point support with probability mass $\frac{1}{N}$ at each point in its support.*

Proof. Let $\mathcal{M}$ be the space of probability measures on $\mathbb{Z}_{\geq 0}$. Let

$$\mathcal{M}(N) := \left\{ \mu \in \mathcal{M} : \ \max_k \mu(\{k\}) \leq \frac{1}{N} \right\}$$

and

$$\mathcal{M}_{\mathrm{u}}(N) := \{\mu \in \mathcal{M} : \ \mu \text{ is uniformly supported on } N \text{ points}\}.$$

By the Krein-Milman theorem [59], $\mathcal{M}(N)$ is the convex hull of its extreme points. We claim that the extreme points are precisely the points of $\mathcal{M}_{\mathrm{u}}(N)$. It is immediately evident that each point of $\mathcal{M}_{\mathrm{u}}(N)$ is an extreme point of $\mathcal{M}(n)$. To check the converse inclusion, we suppose $\mu \in \mathcal{M}(N) \backslash \mathcal{M}_{\mathrm{u}}(N)$. Thus there is some $k \in \mathbb{Z}_{\geq 0}$ such that $0 < \mu(\{k\}) < \frac{1}{N}$, but in fact, there must be at least two distinct such

k, since the total mass of μ is 1 (an integer multiple of $\frac{1}{N}$). Therefore, μ is not an extreme point of $\mathcal{M}(N)$.

This proves our claim. Hence the probability measure associated to Y can be written as a countable convex combination of points of $\mathcal{M}_{\mathrm{u}}(N)$, each of which defines the distribution of a random variable Y_k (proving the lemma). □

Lemma II.30 (Properties of mixtures)**.** *If Y is a mixture of random variables $Y_1, Y_2, \ldots$, then:*

(i) *There is some $k \geq 1$ for which* $\mathrm{conc}(Y) \leq \mathrm{conc}(Y_k)$.

(ii) *If Z is a random variable and f is a function such that $Z = f(Y)$, then Z is a mixture of random variables $Z_1, Z_2, \ldots$, where $Z_k = f(Y_k)$.*

Proof. By the definition of *mixture*, there exist nonnegative $\alpha_1, \alpha_2, \ldots$ such that $\alpha_1 + \alpha_2 + \cdots = 1$ and such that

$$\mathbf{Pr}[Y = y] = \sum_{k=1}^{\infty} \alpha_k \mathbf{Pr}[Y_k = y].$$

Thus by the pigeonhole principle, for arbitrary y, there exists $k = k(y)$ such that

$$\mathbf{Pr}[Y = y] \leq \mathbf{Pr}[Y_k = y].$$

Choosing y such that $\mathrm{conc}(Y) = \mathbf{Pr}[Y = y]$, we conclude that $\mathrm{conc}(Y) \leq \mathrm{conc}(Y_k)$ for this k. This proves claim (i) in the lemma. Claim (ii) is self-evident. □

The last ingredient we need to prove Theorem II.27 is a borrowed local limit theorem for log-concave sequences.

Definition II.31. A sequence $(\ldots, b_{-1}, b_0, b_1, b_2, \ldots)$ of nonnegative real numbers is **properly log-concave** if it is log-concave (i.e., $b_{t-1}b_{t+1} \leq b_t^2$ for all t) and has no internal zeroes (i.e., if $b_t > 0$ and $b_{t+k} > 0$, then $b_{t+1}, b_{t+2}, \ldots, b_{t+k-1} > 0$).

Theorem II.32 (Bender [16])**.** [15] *Suppose that* $\left(\zeta_n : n \in \mathbb{Z}_{>0}\right)$ *is a sequence of integer-valued random variables and* $\left(\sigma_n\right)$ *and* $\left(\mu_n\right)$ *are sequences of real numbers, such that*

$$\lim_{n\to\infty} \mathbf{Pr}\left[\zeta_n < \sigma_n x + \mu_n\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

for all $x \in \mathbb{R}$*. Also suppose that* $\sigma_n \to \infty$ *as* $n \to \infty$*. Further, suppose that, for every* n*, the sequence* $b_n(t) := \mathbf{Pr}(\zeta_n = t)$ *is properly log-concave with respect to* t*. Then*

$$\lim_{n\to\infty} \sigma_n \mathbf{Pr}\left[\zeta_n = \lfloor \sigma_n x + \mu_n \rfloor\right] = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

uniformly for all $x \in \mathbb{R}$*.*

To satisfy the hypotheses of Bender's local limit theorem, one must generally first apply a central limit theorem. We will use the following standard one (see, e.g., [17] or [63] for a proof):

Theorem II.33 (Lyapunov)**.** *Suppose that* $\left(X_n : n \in \mathbb{Z}_{>0}\right)$ *is a sequence of independent random variables, such that* $\mu_n := \mathbf{E}[X_n]$ *and* $\sigma_n^2 := \mathbf{Var}[X_n]$ *are finite. Let* $\zeta_n = X_1 + \cdots + X_n$*, and define*

$$m_n := \mathbf{E}[\zeta_n] = \mu_1 + \cdots + \mu_n,$$

$$s_n^2 := \mathbf{Var}[\zeta_n] = \sigma_1^2 + \cdots + \sigma_n^2.$$

If

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{E}\left[|X_k - \mu_k|^{2+\delta}\right] = 0$$

for some $\delta > 0$*, then*

$$\lim_{n\to\infty} \mathbf{Pr}\left[\zeta_n < s_n x + m_n\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

for all $x \in \mathbb{R}$*.*

[15] Our statement of this result is based on the treatment in [31], where a proof also appears.

Proof of Theorem II.27. For $j = 1, 2, \ldots, p$, we are given to assume that $\mathrm{conc}(X_j) \leq \frac{1}{N_j}$. By Lemma II.29, each X_j is a mixture of some random variables which are each uniformly supported on some N_j points. Thus the random vector $X = (X_1, \ldots, X_p)$ is a mixture of random vectors each of the form $X^{(k)} := (X_1^{(k)}, \ldots, X_p^{(k)})$, where the coordinates are independent and each $X_j^{(k)}$ is uniformly supported on N_j points. The sum $X_1 + \cdots + X_p$ is a function of X, so by using both parts of Lemma II.30, we see that

$$\mathrm{conc}(X_1 + \cdots + X_p) \leq \mathrm{conc}(X_1^{(k)} + \cdots + X_p^{(k)})$$

for some k. Since we are seeking an upper bound on $\mathrm{conc}(X_1 + \cdots + X_p)$, we assume with no loss of generality that $X = X^{(k)}$, or, more to the point, that each coordinate X_j is uniformly supported on N_j points (with mass $\frac{1}{N_j}$ on each).

Denote the support of X_j by $\{a_{j1}, a_{j2}, \ldots, a_{jN_j}\}$, where $a_{j1} < a_{j2} < \cdots < a_{jN_j}$. Then

$$a_{1i_1} + a_{2i_2} + \cdots + a_{pi_p} = a_{1i'_1} + a_{2i'_2} + \cdots + a_{pi'_p}$$

implies that the p-tuples $(i_1, i_2, \ldots, i_p)$ and $(i'_1, i'_2, \ldots, i'_p)$ are identical or incomparable in $[N_1] \times \cdots \times [N_p]$. It follows that

$$\mathrm{conc}(X_1 + \cdots + X_p) \leq \frac{w\big([N_1] \times \cdots \times [N_p]\big)}{N_1 N_2 \cdots N_p}.$$

This proves the first claim of Theorem II.27.

For the remainder of the proof, assume that $2 \leq N_1, N_2, \ldots, N_p < N$ for some integer N. We are going to apply Bender's local limit theorem (Theorem II.32). Let ζ_p denote the rank of a uniformly distributed random element of $[N_1] \times [N_2] \times \cdots \times [N_p]$. Set $\mu_p := \frac{N_1 + \cdots + N_p}{2}$ and $\sigma_p^2 = \sum_{j=1}^{p} \frac{N_j^2 - 1}{12}$. It is easily verified that μ_p and σ_p^2 are respectively the mean and the variance of ζ_p. By Lyapunov's central limit theorem

(Theorem II.33), the hypothesis

$$\lim_{p\to\infty} \mathbf{Pr}\left[\zeta_p < \sigma_p x + \mu_p\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

in Bender's local limit theorem is satisfied. The hypothesis $\sigma_p \to \infty$ is plainly also satisfied.

To see that the sequence $b_p(t) := \mathbf{Pr}(\zeta_p = t)$ is properly log-concave, we note that this sequence is proportional to the Whitney numbers of the chain product $[N_1] \times [N_2] \times \cdots \times [N_p]$, which is the convolution of the sequences of Whitney numbers for the factor chains. Each factor chain has Whitney numbers $1, 1, \ldots, 1, 0, 0, \ldots$ (a properly log-concave sequence). Furthermore, the convolution of properly log-concave sequences is again properly log-concave, see e.g. [48]. Thus, $\big(b_p(t)\big)$ is properly log-concave.

All antecedents of Bender's theorem have been verified, so the conclusion holds:

$$\lim_{p\to\infty} \sigma_p \mathbf{Pr}\big(\zeta_p = \lfloor \sigma_p x + \mu_p \rfloor\big) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

uniformly for all $x \in \mathbb{R}$. Setting $x = 0$, we obtain

$$\begin{aligned} \mathbf{Pr}\big(\zeta_p = \lfloor \mu_p \rfloor\big) &\sim \frac{1}{\sqrt{2\pi\sigma_p}} \\ &= \Big(\frac{\pi}{6} \sum_{j=1}^{p} (N_j^2 - 1)\Big)^{-1/2}. \end{aligned}$$

Finally, we observe that chain products have the Sperner property [31].[16] In particular, the width in the above formula is equal to the Whitney number $W_{\lfloor \mu_p \rfloor}$, so that

$$\frac{w\big([N_1] \times \cdots \times [N_p]\big)}{N_1 N_2 \cdots N_p} = \mathbf{Pr}\big(\zeta_p = \lfloor \mu_p \rfloor\big).$$

This completes the proof of the proposition. ■

[16] There is a pretty proof of this fact using *symmetric chain decompositions*.

We obtain the E-bound as an instance of Theorem II.27:

Proof of Theorem II.25. As noted in the proof of Theorem II.19, we have

$$\operatorname{conc}(X_j \mathbf{a}_j) = \frac{1}{\mathbf{E}(X_j)+1} \leq \frac{1}{\lfloor \mathbf{E}(X_j)+1 \rfloor}$$

for $1 \leq j \leq n$. Since $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ are linearly independent, we have

$$\begin{aligned}\operatorname{conc}(AX) &= \prod_{i=1}^{m} \operatorname{conc}(X_i \mathbf{a}_i + X_{m+i}\mathbf{a}_{m+i} + X_{2m+i}\mathbf{a}_{2m+i} + \cdots + X_{(p-1)m+i}\mathbf{a}_{(p-1)m+i}) \\ &= (1+o(1)) \prod_{i=1}^{m} \left(\frac{\pi}{6} \sum_{t=1}^{p} \left(\lfloor \mathbf{E}[X_{(t-1)m+i}] + 1 \rfloor^2 - 1 \right) \right)^{-1/2},\end{aligned}$$

where the last claim follows by Theorem II.27. Finally, by (2.9), we infer Theorem II.25. ■

Remark II.34. As previously noted, the E-bound is essentially a dimension-1 result. One obstacle to a full generalization is the lack of a well-developed Sperner theory for posets with multi-dimensional rank functions.

It is plausible to guess that something of the following sort might be true:

Hypothesis II.35. *Let $n = pm$. Let $X_1, X_2, \ldots, X_n$ be geometric random variables with $\mathbf{E}[X_{(k-1)m+1}] \geq \mathbf{E}[X_{(k-1)m+2}] \geq \cdots \geq \mathbf{E}[X_{km}]$ for every $k = 1, \ldots, p$. Then among all sequences of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ such that $\mathbf{a}_{(k-1)m+1}, \ldots, \mathbf{a}_{km}$ are linearly independent for every $k = 1, \ldots, p$, the maximum value of $\operatorname{conc}(AX)$ is achieved when $\mathbf{a}_i = \mathbf{a}_{m+i} = \mathbf{a}_{2m+i} = \cdots = \mathbf{a}_{(p-1)m+i}$ for every $i = 1, \ldots, m$.*

Were this the case, the rather restrictive hypotheses of the E-bound would represent the worst case and so become universal. We would also thus obtain the $n^{-m/2}$ factor wished for in Remark II.22. Unfortunately, we have discovered a counterexample[17] to Hypothesis II.35 in dimension 2, but the validity of a weakened form of

[17] Such counterexamples do not appear to be rare, but for the record, here is ours: Let $m = 2$, $p = 4$. Given $X_1, \ldots, X_8$ geometric with $\mathbf{E}[X] = (8, 7, 6, 5, 4, 3, 2, 1)$, we have $\operatorname{conc}(AX) = 1.940 \times 10^{-3}$ when $A = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$, but $\operatorname{conc}(AX) = 2.046 \times 10^{-3}$ when $A = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$.

Hypothesis II.35 (perhaps with an approximation factor) remains plausible. The idea of "aligning" or "co-rectifying" the columns of A will reappear in the next section.

2.6 The H-bound

The statement of the H-bound with explicit constants is complicated enough that we are driven to invert usual protocol and state the corollary first:

Corollary II.36. *Fix an integer $m \geq 1$, and fix $\varepsilon > 0$. Then there exists a positive constant $\delta = \delta(m, \varepsilon)$, such that the following is true:*

Assume Conventions II.7 and II.15, with P in standard form. If A has integer entries, and a subset of its columns can be partitioned into p bases for $\mathbb{R}^m$, and if $\min_j q_j \geq \varepsilon$, then

$$|P \cap \mathbb{Z}^n| \leq e^{\mathbf{H}[X]}(\delta p^{-m/2}).$$

In informal terms, $\mathbf{Pr}[AX = \mathbf{b}]$ (or actually the point concentration of AX) is bounded by a Gaussian estimate as $p \to \infty$, so long as $\min_j q_j$ is uniformly bounded away from zero.

Note that, apart from the integrality of A, the hypotheses of Corollary II.36 are not restrictive; we do not insist that $p \approx n/m$, though the conclusion is strongest in that case.[18] The role of the parameter p here is "honest," analogous to the role of δ in the theorem of Halász (Theorem II.18).

Here is the full result:

Theorem II.37 (H-bound)**.** *Assume Conventions II.7 and II.15, with P in standard form. Assume that A has integer entries, and that $q_j > 0$ for $1 \leq j \leq n$.* [19]

[18] In the full result to follow, we *do* assume $n = pm$, but this can be achieved by ignoring extra columns of A, in effect projecting P to dimension pm. Per (2.10), the concentration of AX may go up but not down under this operation, so the resulting bounds are valid for the original P.

[19] Instead of assuming $q_j > 0$ for all j, we may assume that $\langle \mathbf{a}_j, \mathbf{b} \rangle > 0$ for $1 \leq j \leq n$. To see why these assumptions are equivalent, refer to the proof of Proposition II.12. Note that these assumptions are not restrictive: if $\langle \mathbf{a}_j, \mathbf{b} \rangle = 0$ for any j, then P represents with respect to integer points (see Definition I.2) a lower-dimensional polytope for which this is not the case.

Suppose that $n = pm$ for some integer p, and that $\mathbf{a}_{(k-1)m+1}, \mathbf{a}_{(k-1)m+2}, \ldots, \mathbf{a}_{km}$ are linearly independent for $1 \le k \le p$. Let $\gamma > 0$. [20] *Define constants*

$$\alpha_j := \frac{2q_j}{(1-q_j)^2} \qquad (1 \le j \le n),$$

$$\alpha_i^\vee := \min\{\alpha_{(k-1)m+i} : 1 \le k \le p\} \qquad (1 \le i \le m),$$

$$q_i^\vee := \min\{q_{(k-1)m+i} : 1 \le k \le p\} \qquad (1 \le i \le m),$$

$$c_i := \max\left\{\frac{1}{\gamma^2}\ln\left[1+\alpha_i^\vee\left(1-\cos\frac{\gamma}{\sqrt{\alpha_i^\vee}}\right)\right],\quad \frac{1}{\alpha_i^\vee \pi^2}\ln\left[1+2\alpha_i^\vee\right]\right\} \ (1 \le i \le m),$$

$$C := \prod_{i=1}^{m}(2\pi c_i \alpha_i^\vee)^{-1/2},$$

$$C' := \max_{1\le i\le m} e^{-\gamma^2 c_i/2}.$$

Then

$$|P \cap \mathbb{Z}^n| \le e^{\mathbf{H}[X]}\big(Cp^{-m/2} + (C')^p\big).$$

All notation introduced in the statement of Theorem II.37 is used throughout this section, and all its hypotheses (importantly, the integrality of A) are assumed to hold.

2.6.1 Lemmas supporting the proof of the H-bound

In the lemmas stated in this section, the proof of Theorem II.37 can be seen in outline; it will be made explicit in the following section. These lemmas are proved in Section 2.6.3.

Definition II.38. For $1 \le k \le p$, define the function $\Pi_k : \mathbb{R}^m \to \mathbb{R}$ by

$$\Pi_k(\mathbf{t}) := \begin{cases} \prod_{j=(k-1)m+1}^{km} \frac{1}{\sqrt{1+\alpha_j(1-\cos\langle \mathbf{t}, \mathbf{a}_j\rangle)}} & \text{for } \mathbf{t} \in (-\pi,\pi]^m \\ 0 & \text{for } \mathbf{t} \notin (-\pi,\pi]^m \end{cases}.$$

[20] The parameter γ is "at the discretion of the user." See Remark II.49.

Lemma II.39.

$$\mathbf{Pr}[AX = \mathbf{b}] \leq \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p \; d\mathbf{t}.$$

Definition II.40. Given a measurable function $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$, we define its **upper level sets**

$$\Gamma_{\geq \tau}(\Phi) := \{\mathbf{t} \in \mathbb{R}^m : \Phi(\mathbf{t}) \geq \tau\}$$

for all $\tau > 0$.

Suppose that Φ **vanishes at infinity**, meaning that $\Gamma_{\geq \tau}(\Phi)$ has finite volume for every $\tau > 0$. Then we define its **symmetrically decreasing arrangement** as the function $\Phi^* : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ given by

$$\Phi^*(\mathbf{t}) := \max\Big\{\tau : \ \mathrm{vol}\left(\Gamma_{\geq \tau}(\Phi)\right) \geq \|\mathbf{t}\|^m v_m\Big\},$$

where v_m denotes the volume of the unit ball in $\mathbb{R}^m$.

The theory of symmetrically decreasing rearrangements is treated in [18] (also [49]), and we do not develop it fully here. The important properties of Φ^* are that

- Φ^* is symmetrically decreasing, i.e., $\|\mathbf{t}\| \geq \|\mathbf{s}\| \Rightarrow \Phi^*(\mathbf{t}) \leq \Phi^*(\mathbf{s})$; and
- Φ^* is equimeasurable with Φ, i.e., $\mathrm{vol}(\Gamma_{\geq \tau}(\Phi^*)) = \mathrm{vol}(\Gamma_{\geq \tau}(\Phi))$ for all $\tau > 0$.

(In fact, Φ^* is the unique function with these properties, up to difference on a set of measure zero.)

Lemma II.41 (Hardy-Littlewood)**.** *If $\Phi_1, \ldots, \Phi_n : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ are measurable functions vanishing at infinity, then*

$$\int_{\mathbb{R}^m} \Phi_1(\mathbf{t}) \cdots \Phi_n(\mathbf{t}) \; d\mathbf{t} \leq \int_{\mathbb{R}^m} \Phi_1^*(\mathbf{t}) \cdots \Phi_n^*(\mathbf{t}) \; d\mathbf{t},$$

provided that the integral on the right-hand side converges.

Definition II.42. For $1 \leq k \leq p$, define the function $\Pi_k^{\text{rect}} : \mathbb{R}^m \to \mathbb{R}$ by

$$\Pi_k^{\text{rect}}(\mathbf{t}) := \begin{cases} \prod_{i=1}^m \frac{1}{\sqrt{1+\alpha_{(k-1)m+i}(1-\cos \mathbf{t}_i)}} & \text{for } \mathbf{t} \in (-\pi, \pi]^m, \\ 0 & \text{for } \mathbf{t} \notin (-\pi, \pi]^m. \end{cases}$$

The formula for Π_k^{rect} differs from that for Π_k in that the linear form $\langle \mathbf{t}, \mathbf{a}_{(k-1)m+i} \rangle$ in the denominator of Π_k is replaced by $\mathbf{t}_i$. Effectively, each basis $\mathbf{a}_{(k-1)m+1}, \mathbf{a}_{(k-1)m+2}, \ldots, \mathbf{a}_{km}$ of $\mathbb{R}^m$ is replaced by a standard basis[21]. This will make Π_k^{rect} easier to work with than Π_k.

Lemma II.43. *Let $1 \leq k \leq p$. Then*

$$\text{vol}\left(\Gamma_{\geq \tau}(\Pi_k^{\text{rect}})\right) = \text{vol}\left(\Gamma_{\geq \tau}(\Pi_k)\right)$$

for all $\tau > 0$, and $(\Pi_k^{\text{rect}})^* \equiv \Pi_k^*$.

Lemma II.44 (Isotonicity of rearrangement)**.** *Suppose $\Phi, \Psi : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ are measurable functions vanishing at infinity. Let τ denote a constant. Then:*

(i) *If $\Phi(\mathbf{t}) \geq \Psi(\mathbf{t})$ for all $\mathbf{t}$, then $\Phi^*(\mathbf{t}) \geq \Psi^*(\mathbf{t})$ for all $\mathbf{t}$.*

(ii) *If $\Phi(\mathbf{t}) \geq \max\{\Psi(\mathbf{t}), \tau\}$ for all $\mathbf{t} \in \operatorname{supp} \Phi$, then $\Phi^*(\mathbf{t}) \geq \max\{\Psi^*(\mathbf{t}), \tau\}$ for all $\mathbf{t} \in \operatorname{supp} \Phi^*$.*

Lemma II.45. [22] *For $0 \leq t \leq \min\left\{\frac{\gamma}{\sqrt{\alpha_i^\vee}},\ \pi\right\}$, we have*

$$1 + \alpha_i^\vee(1 - \cos t) \geq e^{c_i \alpha_i^\vee t^2}.$$

Lemma II.46. *For each $k = 1, 2, \ldots, p$, and for all $\mathbf{t} \in \mathbb{R}^m$, we have*

$$\Pi_k^{\text{rect}}(\mathbf{t}) \leq \max\left\{\prod_{i=1}^m e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2},\quad C'\right\}.$$

[21] As prefigured in Remark II.34.

[22] Recall the definitions of $\alpha_i^\vee$ and c_i from the statement of Theorem II.37.

2.6.2 Proof of the H-bound

Using Lemmas II.39, II.41, and II.43, we have

$$\begin{aligned}\mathbf{Pr}[AX = \mathbf{b}] &\leq \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} \Pi_1 \Pi_2 \cdots \Pi_p \; d\mathbf{t} \\ &\leq \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \Pi_1^* \Pi_2^* \cdots \Pi_p^* \; d\mathbf{t} \\ &= \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} (\Pi_1^{\text{rect}})^* (\Pi_2^{\text{rect}})^* \cdots (\Pi_p^{\text{rect}})^* \; d\mathbf{t}.\end{aligned}$$

We may instead take either of the last two integrals over B, the closed ball of volume $(2\pi)^m$ centered at the origin in $\mathbb{R}^m$, since the integrands are zero outside this ball.

By Lemmas II.44 and II.46, we have

$$\begin{aligned}&\frac{1}{(2\pi)^m} \int_B (\Pi_1^{\text{rect}})^* (\Pi_2^{\text{rect}})^* \cdots (\Pi_p^{\text{rect}})^* \; d\mathbf{t} \\ &\qquad \leq \frac{1}{(2\pi)^m} \int_B \prod_{k=1}^{p} \left(\max \left\{ \left(\prod_{i=1}^{m} e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2} \right)^*, \quad C' \right\} \right) \; d\mathbf{t} \\ &\qquad = \frac{1}{(2\pi)^m} \int_B \left(\max \left\{ \left(\prod_{i=1}^{m} e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2} \right)^*, \quad C' \right\} \right)^p \; d\mathbf{t} \\ &\qquad = \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} \left(\max \left\{ \prod_{i=1}^{m} e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2}, \quad C' \right\} \right)^p \; d\mathbf{t}.\end{aligned}$$

This last integral is bounded above by

$$\begin{aligned}&\frac{1}{(2\pi)^m} \left[\int_{\mathbb{R}^m} \left(\prod_{i=1}^{m} e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2} \right)^p \; d\mathbf{t} \quad + \quad \int_{(-\pi,\pi]^m} (C')^p \; d\mathbf{t} \right] \\ &= \frac{1}{(2\pi)^m} \left[\int_{\mathbb{R}^m} \exp\left(-p \sum_{i=1}^{m} -c_i \alpha_i^\vee \mathbf{t}_i^2/2 \right) \; d\mathbf{t} \quad + \quad (2\pi)^m (C')^p \right] \\ &= \frac{1}{(2\pi)^m} \cdot (2\pi)^{m/2} p^{-m/2} \prod_{i=1}^{m} (c_i \alpha_i^\vee)^{-1/2} \quad + \quad (C')^p \\ &= C p^{-m/2} + (C')^p.\end{aligned}$$

(Note that in integrating the Gaussian term, we used the assumption that $q_j > 0$ for all $1 \leq j \leq n$, which implies that $c_i \alpha_i^\vee > 0$ for $1 \leq i \leq m$.)

Theorem II.37 (the H-bound) now follows by (2.8). ■

Proof of Corollary II.36. Fix arbitrary $\gamma > 0$. Let m, ε be fixed. Continuing the notation of Theorem II.37, we have $C' < 1$, so $(C')^p = o(Cp^{-m/2})$ as $p \to \infty$. Both C and C' were defined in such a way that they depend only on ε. The corollary follows straightforwardly. ■

Remark II.47. Our strategy for bounding $\mathbf{Pr}[AX = \mathbf{b}]$, carried out above, may be summarized/motivated as follows. First, we obtain an integral formula for the probability mass function of AX, derived from its Fourier transform (Lemma II.39). The integrand splits into n factors, which we then group into maximal subproducts such that the factors in each subproduct behave like independent random variables on the domain of integration. The worst case is now that these subproducts themselves are "completely non-independent," that is, that they decay identically; this is the significance of Lemmas II.41 and II.43, and of the definitions of $q_i^\vee$ and $\alpha_i^\vee$. We bound the decay of the integrand near the origin by a Gaussian (Lemma II.46), explaining the appearance of the $Cp^{-m/2}$ term in the conclusion of Theorem II.37. Away from the origin, we simply bound each subproduct by the constant C', producing the $(C')^p$ term. The parameter γ controls the boundary between the two approximation regimes.

This two-regime bound (with arbitrary parameter γ) is strong enough to imply Corollary II.36, but for non-asymptotic computations, the crudity of the approximation away from the origin is quite noticeable. The $(C')^p$ term can be replaced by a more sensitive approximation, at the cost of simplicity: for example, one could carve the domain of integration into 2^n regions (treating each variable separately), or (as a compromise) into $n + 1$ regions according to the number of variables which

are removed from the origin. These considerations are important if one wishes to compute good "H-bounds" for individual specimens, but they are mostly irrelevant if one only wants to confirm the asymptotic Gaussian behavior of the H-bound in families of polytopes with dimension approaching ∞.

2.6.3 Proofs of the supporting lemmas

Proof of Lemma II.39. In [12], Lemma 8.1, the following integral representation is proved:

$$\mathbf{Pr}[AX = \mathbf{b}] = \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} e^{-i\langle \mathbf{t},\mathbf{b}\rangle} \prod_{j=1}^{n} \frac{1-q_j}{1-q_j e^{i\langle \mathbf{t},\mathbf{a}_j\rangle}}\, d\mathbf{t},$$

where $\mathbf{b}$ is an arbitrary $\mathbb{Z}_{\geq 0}$-vector. It follows that

$$\begin{aligned}
\mathbf{Pr}[AX = \mathbf{b}] &\leq \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} \left| e^{-i\langle \mathbf{t},\mathbf{b}\rangle} \prod_{j=1}^{n} \frac{1-q_j}{1-q_j e^{i\langle \mathbf{t},\mathbf{a}_j\rangle}} \right| d\mathbf{t} \\
&= \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} \prod_{j=1}^{n} \frac{1-q_j}{\sqrt{1+q_j^2-2q_j\cos\langle \mathbf{t},\mathbf{a}_j\rangle}}\, d\mathbf{t} \\
&= \frac{1}{(2\pi)^m} \int_{(-\pi,\pi]^m} \Pi_1\Pi_2\cdots\Pi_p\, d\mathbf{t},
\end{aligned}$$

where the last two steps are straightforward simplification. □

Proof of Lemma II.41. See Theorem 3.8 in [49]. □

Proof of Lemma II.43. Let A_* be the $m \times m$ matrix whose rows are $\mathbf{a}_{(k-1)m+1}^T, \mathbf{a}_{(k-1)m+2}^T, \ldots, \mathbf{a}_{km}^T$, and define $\mathcal{A}_* : \mathbb{R}^m \to \mathbb{R}^m$ as the linear map $\mathbf{t} \mapsto A_*\mathbf{t}$. Thus,

$$\mathcal{A}_*(\mathbf{t})_i = \langle \mathbf{t}, \mathbf{a}_{(k-1)m+i}\rangle \quad (1 \leq i \leq m).$$

This map $\mathcal{A}_*$ scales the volume of measurable sets uniformly by a factor of $d := |\det(A_*)|$, and takes the lattice $\Lambda := (2\pi\mathbb{Z})^m$ to the lattice

$$\Lambda' := 2\pi\mathbb{Z}[\mathbf{col}_1(A_*), \mathbf{col}_2(A_*), \ldots, \mathbf{col}_m(A_*)].$$

Let $K := (-\pi, \pi]^m$ and let $K' := \mathcal{A}_*(K)$. Since K is a fundamental region of Λ, it follows that K' is a fundamental region of Λ'. Moreover, we assumed A to have integer entries, so Λ' is a sublattice of index d in Λ, and the induced map of tori $\phi : \mathbb{R}^m/\Lambda' \to \mathbb{R}^m/\Lambda$ is an even covering of order d.

Identifying K with $\mathbb{R}^m/\Lambda$ and K' with $\mathbb{R}^m/\Lambda'$, we may regard ϕ is a map from K' to K, and $\phi \circ \mathcal{A}_*$ as a self-map of K. If $U \subseteq K$ is a measurable set, then $(\phi \circ \mathcal{A}_*)^{-1}(U)$ is the union of d disjoint preimages each of volume $\frac{\mathrm{vol}(U)}{d}$. Thus, $\mathrm{vol}((\phi \circ \mathcal{A}_*)^{-1}(U)) = \mathrm{vol}(U)$.

Observe that $\cos \mathbf{t}_i = \cos(\phi(\mathbf{t})_i)$ for all $\mathbf{t}$. Therefore

$$\begin{aligned}\Gamma_{\geq\tau}(\Pi_k) &= {\mathcal{A}_*}^{-1}(\Gamma_{\geq\tau}(\Pi_k^{\mathrm{rect}})) \\ &= (\phi \circ \mathcal{A}_*)^{-1}(\Gamma_{\geq\tau}(\Pi_k^{\mathrm{rect}}))\end{aligned}$$

from which it follows that

$$\mathrm{vol}\left(\Gamma_{\geq\tau}(\Pi_k^{\mathrm{rect}})\right) = \mathrm{vol}\left(\Gamma_{\geq\tau}(\Pi_k)\right).$$

This conclusion holds for all $\tau > 0$, so it follows from the definition of the symmetrically decreasing rearrangement that $(\Pi_k^{\mathrm{rect}})^* \equiv \Pi_k^*$. $\square$

Proof of Lemma II.44. We prove (i) by contradiction. Suppose that $\Phi(\mathbf{t}) \geq \Psi(\mathbf{t})$ for all $\mathbf{t}$, but suppose $\Phi^*(\mathbf{t}_0) < \Psi^*(\mathbf{t}_0)$ for some $\mathbf{t}_0$. Let $\tau_0 := \Psi^*(\mathbf{t}_0)$. Then

$$\mathrm{vol}\left(\Gamma_{\geq\tau_0}(\Phi)\right) < \|\mathbf{t}_0\|^m v_m \leq \mathrm{vol}\left(\Gamma_{\geq\tau_0}(\Psi)\right),$$

where v_m is the volume of the unit ball in $\mathbb{R}^m$. It follows that $\Gamma_{\geq\tau_0}(\Psi) \backslash \Gamma_{\geq\tau_0}(\Phi)$ has positive measure, contradicting our assumption that $\Phi(\mathbf{t}) \geq \Psi(\mathbf{t})$ for all $\mathbf{t}$.

To see that Statement (ii) holds, define $\Psi_\tau(\mathbf{t})$ as the function equal to $\max\{\Psi(\mathbf{t}), \tau\}$ on $\operatorname{supp}\Psi$, and to zero elsewhere; also define $(\Psi^*)_\tau(\mathbf{t})$ as the function equal to

$\max\{\Psi^*(\mathbf{t}), \tau\}$ on $\operatorname{supp}\Psi^*$ and to zero elsewhere. Then it is easily verified that $(\Psi^*)_\tau = (\Psi_\tau)^*$, so (ii) follows from (i). $\square$

Proof of Lemma II.45. One may check that

$$c_i := \begin{cases} \frac{1}{\gamma^2} \ln\left[1 + \alpha_i^\vee \left(1 - \cos\frac{\gamma}{\sqrt{\alpha_i^\vee}}\right)\right] & \text{if } \alpha_i^\vee \geq \frac{\gamma^2}{\pi^2} \\ \frac{1}{\alpha_i^\vee \pi^2} \ln\left[1 + 2\alpha_i^\vee\right] & \text{if } \alpha_i^\vee \leq \frac{\gamma^2}{\pi^2} \end{cases}. \tag{2.11}$$

Define $t_0 := \min\left\{\frac{\gamma}{\sqrt{\alpha_i^\vee}},\ \pi\right\}$, and define $f(t) := 1 + \alpha_i^\vee(1 - \cos t) - e^{c_i \alpha_i^\vee t^2}$ for $-t_0 \leq t \leq t_0$.

Note that $f(0) = 0$. Also, we claim that $f(t_0) = 0$. This must be verified in two cases, according to whether $\alpha_i^\vee \geq \frac{\gamma^2}{\pi^2}$ or $\alpha_i^\vee \leq \frac{\gamma^2}{\pi^2}$.

If $\alpha_i^\vee \geq \frac{\gamma^2}{\pi^2}$, then $t_0 = \frac{\gamma}{\sqrt{\alpha_i^\vee}}$, so

$$\begin{aligned} f(t_0) &= 1 + \alpha_i^\vee \left(1 - \cos\frac{\gamma}{\sqrt{\alpha_i^\vee}}\right) - \exp\left(\frac{\alpha_i^\vee}{\gamma^2} \cdot \ln\left[1 + \alpha_i^\vee \left(1 - \cos\frac{\gamma}{\sqrt{\alpha_i^\vee}}\right)\right] \cdot \frac{\gamma^2}{\alpha_i^\vee}\right) \\ &= 0. \end{aligned}$$

If $\alpha_i^\vee \leq \frac{\gamma^2}{\pi^2}$, then $t_0 = \pi$, and

$$f(t_0) \quad = \quad 1 + 2\alpha_i^\vee - \exp\left(\frac{1}{\alpha_i^\vee \pi^2} \cdot \ln\left[1 + 2\alpha_i^\vee\right] \cdot \alpha_i^\vee \pi^2\right) \quad = \quad 0.$$

This proves the claim that $f(t_0) = 0$. It follows that the average value of $f'(t)$ on $[0, t_0]$ is zero.

Finally, we observe that $f'(0) = 0$, and that $f(t)$ has nonpositive third derivative on $[0, t_0]$ (indeed, on $[0, \pi]$). The verification of these claims is routine and is omitted. We infer that either $f'(t) \equiv 0$ on $[0, t_0]$, or $f''(t)$ has exactly one sign change on $[0, t_0]$, from positive to negative. In the latter case, $f'(t)$ must also have exactly one sign change on $[0, t_0]$ (also from positive to negative), since its average value on the interval

is zero. It follows in either case that $f(t) \geq 0$ on $[0, t_0]$, and thus on $[-t_0, t_0]$ (since $f(t)$ is an even function). This proves the lemma. $\square$

Proof of Lemma II.46. Let

$$K := \left\{ \mathbf{t} \in \mathbb{R}^m : |\mathbf{t}_i| \leq \min\left\{ \frac{\gamma}{\sqrt{\alpha_i^\vee}},\ \pi \right\} \text{ for } i = 1, 2, \ldots, m \right\}.$$

If $\mathbf{t} \in K$, then by Lemma II.45,

$$\begin{aligned} \Pi_k^{\text{rect}}(\mathbf{t}) &= \prod_{i=1}^{m} \frac{1}{\sqrt{1 + \alpha_{(k-1)m+i}(1 - \cos \mathbf{t}_i)}} \\ &\leq \prod_{i=1}^{m} \frac{1}{\sqrt{1 + \alpha_i^\vee (1 - \cos \mathbf{t}_i)}} \\ &\leq \prod_{i=1}^{m} e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2}. \end{aligned}$$

Now suppose $\mathbf{t} \notin K$. Thus, there exists some i such that $\mathbf{t}_i > \min\left\{ \frac{\gamma}{\sqrt{\alpha_i^\vee}},\ \pi \right\}$. If $\mathbf{t}_i > \pi$, then we trivially have $\Pi_k^{\text{rect}}(\mathbf{t}) = 0 \leq C'$.

Otherwise, we have $\mathbf{t}_i > \frac{\gamma}{\sqrt{\alpha_i^\vee}}$, and therefore

$$\begin{aligned} \Pi_k^{\text{rect}}(\mathbf{t}) &\leq \frac{1}{\sqrt{1 + \alpha_i^\vee (1 - \cos \mathbf{t}_i)}} \\ &\leq \frac{1}{\sqrt{1 + \alpha_i^\vee \left(1 - \cos\left(\gamma / \sqrt{\alpha_i^\vee}\right)\right)}} \\ &= e^{-\gamma^2 c_i/2} \\ &\leq C'. \end{aligned}$$

Thus whether $\mathbf{t} \in K$ or $\mathbf{t} \notin K$, we have

$$\Pi_k^{\text{rect}}(\mathbf{t}) \leq \max\left\{ \prod_{i=1}^{m} e^{-c_i \alpha_i^\vee \mathbf{t}_i^2/2},\quad C' \right\},$$

proving the lemma. $\square$

2.6.4 Analysis of the constants

The constants C, C' in the statement of the H-bound are awkward. Although we have given "explicit" formulas for both, these formulas are too complicated to understand at a glance, and their behavior relative to γ is unclear. In this section, we give upper bounds on both constants, then discuss optimization of the H-bound.

Theorem II.48. *Defining all notation as in the statement of Theorem II.37, we have*

$$C \leq \left[\frac{\gamma}{2\sqrt{\pi \ln\left(1 + \frac{2\gamma^2}{\pi^2}\right)}} \right]^m \prod_{i=1}^{m} \frac{1 - q_i^\vee}{\sqrt{q_i^\vee}}$$

and

$$C' \leq \frac{1}{\sqrt{1 + \frac{2\gamma^2}{\pi^2}}}.$$

Proof. We may understand equation (2.11) as expressing c_i as a function of $\alpha_i^\vee$. We claim that this function is minimized at $\alpha_i^\vee = \frac{\gamma^2}{\pi^2}$. To demonstrate this claim, it suffices to check that:

1. The function $f(x) := \frac{\ln(1+2x)}{x}$ is decreasing for $0 < x \leq \frac{\gamma^2}{\pi^2}$.

2. The function $g(x) := x(1 - \cos\frac{\gamma}{\sqrt{x}})$ is increasing for $\frac{\gamma^2}{\pi^2} \leq x < \infty$.

Proof of (1): Differentiating, we obtain $f'(x) = \frac{1}{x^2}\left[\frac{2x}{1+2x} - \ln(1+2x)\right]$. In general, $\ln(1+u) > \frac{u}{1+u}$ for $u > 0$, so we have $f'(x) < 0$ for all $x > 0$. In particular, $f(x)$ is decreasing for $0 < x \leq \frac{\gamma^2}{\pi^2}$.

Proof of (2): Differentiating, we obtain $g'(x) = 1 - \cos\frac{\gamma}{\sqrt{x}} - \frac{\gamma}{2\sqrt{x}}\sin\frac{\gamma}{\sqrt{x}}$. It will be convenient to define $y := y(x) = \frac{\gamma}{\sqrt{x}}$. This change of variable bijectively transforms the interval $\frac{\gamma^2}{\pi^2} \leq x < \infty$ into the interval $0 < y \leq \pi$. We may hence write $g'(x) = h(y)$, where

$$h(y) := 1 - \cos y - \frac{y}{2}\sin y.$$

Differentiating twice *with respect to* y, we obtain

$$\frac{dh}{dy} = \frac{1}{2}\sin y - \frac{y}{2}\cos y \quad \text{and} \quad \frac{d^2h}{dy^2} = \frac{y}{2}\sin y.$$

In particular, note that $h(0) = 0$, $h'(0) = 0$, and $h''(y) > 0$ for $0 < y < \pi$. It follows that $h(y) > 0$ for $0 < y \leq \pi$. Equivalently, $g'(x) > 0$ (and $g(x)$ is increasing) for $\frac{\gamma^2}{\pi^2} \leq x < \infty$.

We have thus proved that c_i is minimized when $\alpha_i^\vee = \frac{\gamma^2}{\pi^2}$, in which case $c_i = \frac{1}{\gamma^2}\ln\left(1 + \frac{2\gamma^2}{\pi^2}\right)$. That is to say,

$$c_i \geq \frac{1}{\gamma^2}\ln\left(1 + \frac{2\gamma^2}{\pi^2}\right)$$

for all values of $\alpha_i^\vee$. It follows that

$$\begin{aligned} C = \prod_{i=1}^m (2\pi c_i \alpha_i^\vee)^{-1/2} &\leq \prod_{i=1}^m \left(\frac{2\pi}{\gamma^2}\ln\left(1 + \frac{2\gamma^2}{\pi^2}\right) \cdot \frac{2q_i^\vee}{(1-q_i^\vee)^2}\right)^{-1/2} \\ &= \left[\frac{\gamma}{2\sqrt{\pi \ln\left(1 + \frac{2\gamma^2}{\pi^2}\right)}}\right]^m \prod_{i=1}^m \frac{1-q_i^\vee}{\sqrt{q_i^\vee}} \end{aligned}$$

and

$$\begin{aligned} C' = \max_{1\leq i\leq m} e^{-\gamma^2 c_i/2} &\leq \exp\left(-\frac{\ln\left(1 + \frac{2\gamma^2}{\pi^2}\right)}{2}\right) \\ &= \frac{1}{\sqrt{1 + \frac{2\gamma^2}{\pi^2}}}, \end{aligned}$$

proving Theorem II.48. ■

Remark II.49. For fixed γ and for values of $q_i^\vee$ bounded away from zero, the constant C is essentially a constant multiple of the I-bound for $\mathrm{conc}(AX)$. For example, fixing $\gamma = 1$, we have

$$C \leq (.657)^m \prod_{i=1}^m \frac{1-q_i^\vee}{\sqrt{q_i^\vee}},$$

suggesting that the H-bound outperforms the I-bound when $q_i^\vee$ is not very small and p is large enough for the Gaussian term of the H-bound to dominate the exponential term. We note in passing that Theorem 4 in [39] gives an asymptotic result similar to the H-bound, but with the constant C replaced by a much worse constant, which (up to a factor depending only on m) is at least as large as $\mathrm{conc}(X_1)+\cdots+\mathrm{conc}(X_n)$.

The H-bound can be improved further by letting γ vary and optimizing the result. As $\gamma \to \infty$, all other inputs being fixed, we have $C = O\Big(\big(\frac{\gamma}{\ln \gamma}\big)^m\Big)$ and $C' = O\big(\frac{1}{\gamma}\big)$. There is thus a trade-off between optimizing the $Cp^{-m/2}$ term in Theorem II.37 and optimizing the $(C')^p$ term. Exact optimization of the H-bound is perhaps best performed by a computer, but we can use some simple heuristics to estimate the optimal choice of γ. Let

$$\Gamma := Q\left[\frac{\gamma}{2\sqrt{\pi \ln\left(1+\frac{2\gamma^2}{\pi^2}\right)}}\right]^m \qquad \text{and} \qquad \Delta := \frac{1}{\sqrt{1+\frac{2\gamma^2}{\pi^2}}}$$

denote the bounds on C and C' from Theorem II.48, where

$$Q := \prod_{i=1}^{m} \frac{1-q_i^\vee}{\sqrt{q_i^\vee}}.$$

The global minimum of $\Gamma p^{-m/2} + \Delta^p$ occurs at the unique $\gamma > 0$ satisfying

$$Qm\gamma^{m-2}\left(1+\frac{2\gamma^2}{\pi^2}\right)^{p/2}\left[(\pi^2+2\gamma^2)\ln\left(1+\frac{2\gamma^2}{\pi^2}\right)-2\gamma^2\right] = 4p^{1+m/2}\ln\left(1+\frac{2\gamma^2}{\pi^2}\right)\sqrt{\pi\ln\left(1+\frac{2\gamma^2}{\pi^2}\right)} \tag{2.12}$$

By inspection, we see that this γ must approach 0 as $p \to \infty$, given that m and Q are fixed. Thus we may plausibly substitute $\frac{2\gamma^2}{\pi^2}$ for $\ln\left(1+\frac{2\gamma^2}{\pi^2}\right)$ in equation (2.12). After simplifying the resulting equation, we obtain

$$\gamma^{m-1}(\pi^2+2\gamma^2)^{p/2} = \frac{\pi^p}{Qm}\cdot\frac{2\sqrt{2}}{\sqrt{\pi}}p^{1+m/2}.$$

The solution γ to this equation is $\Omega(p^{-1/2})$, but $o(p^\delta)$ for $\delta > -1/2$.

2.6.5 Numerical examples

We are compelled to give some examples of computed I-, H-, and E-bounds, knowing that they do not impress when juxtaposed with actual enumerations of integer points; for the value of these bounds is not that they are especially sharp, but precisely that they are applicable in settings (such as very high dimension) in which exact computation is not feasible.[23] All of the dimensions and estimates in the examples which follow, except for the last, we regard as "small." (The reader may find this label jarring when applied to numbers on the order of 10^{44}, but for perspective, in Chapter III we will consider families of polytopes whose integer points grow at the rate $e^{\Omega(n^2)}$.)

A standard benchmark among transportation polytopes is that corresponding to the margins $R = (220, 215, 93, 64)$, $C = (108, 286, 71, 127)$; cf. Table 3.1 (p. 65). The actual number of contingency tables with these margins is 1.23×10^{15}. Let X be the corresponding MEIM. Optimization yields

$$\mathbf{E}[X] = \begin{pmatrix} 36.4 & 36.0 & 20.6 & 14.9 \\ 117.2 & 113.3 & 34.3 & 21.2 \\ 22.2 & 22.0 & 15.1 & 11.7 \\ 44.2 & 43.6 & 23.0 & 16.2 \end{pmatrix}$$

and $\mathbf{H}[X] = 2.96 \times 10^{30}$. The I-bound then yields

$$\begin{aligned} |P \cap \mathbb{Z}^n| &\leq \frac{2.96 \times 10^{30}}{(1+36.4)(1+117.2)(1+113.4)(1+34.3)(1+21.2)(1+22.2)(1+44.2)} \\ &= 7.14 \times 10^{18}, \end{aligned}$$

off by between three and four orders of magnitude. This level of relative error seems to be typical for the I-bound applied to 4×4 tables, regardless of the magnitude

[23] The H-bound in particular was designed with an eye toward asymptotic behavior as the dimension goes to ∞.

of the margins. The transportation polytope studied here is defined by 7 equations in 16 variables; accordingly, the best we can do in the H-bound is $p = 2$ (after a suitable reordering of the variables so that the columns of matrix A begin with two bases of $\mathbb{R}^7$). This yields $|P \cap \mathbb{Z}^n| \leq 8.01 \times 10^{26}$.

We also computed the H- and I-bounds for the number of 5×5 tables with margins $R = C = (60, 20, 20, 20, 20)$. The actual number of tables is 2.46×10^{15}. These tables are defined by 9 equations in 25 variables, so we still have $p = 2$ in the H-bound, and performance is only slightly improved (relative to the previous example) due to the greater uniformity in the margins: here the H-bound is 1.26×10^{25}, while the I-bound is 1.04×10^{20}.

Our third example is the 3-way 1-margin transportation polytope whose integer points are $3 \times 3 \times 3$ cubic arrays with all layer sums (1-margins) equal to 20. These arrays are defined by 7 equations in 27 variables, so $p = 3$ in the H-bound, which already yields noticeable improvement: the H-bound is 3.66×10^{20}, almost catching up to the I-bound, here 7.00×10^{19}. (The actual number of arrays is 6.43×10^{14}.) In the three examples considered so far, we note that the H-bound is optimized when the parameter γ goes to ∞.

Now we consider some simplices. Our first simplex comes from [22]: Let $A = (2, 11, 18, 4, 17, 19, 6, 9, 2, 10, 16, 4, 18, 1, 15, 6, 17, 2, 8, 10, 7, 19, 7, 10, 14)$. Then the simplex

$$\{\mathbf{x} = (x_1, x_2, \ldots, x_{25}) \in \mathbb{R}^{25}_{\geq 0} : \ \langle A, \mathbf{x} \rangle \leq 5000\}$$

has 8.57×10^{42} integer points; the H-, I-, and E-bounds are respectively 2.00×10^{44}, 1.07×10^{44}, and 1.04×10^{44}. (The H-bound is optimized at $\gamma = 1.40$.)

Finally, consider the simplex

$$\Sigma^n(r) := \{(x_1, \ldots, x_n) \in \mathbb{R}^n_{\geq 0} : \ x_1 + \cdots + x_n = r\},$$

which has $\binom{n+r-1}{r}$ integer points. For $r = 10$ and $n = 1000$, the optimal H-bound (achieved at $\gamma = 0.172$) is $|\Sigma^n(r) \cap \mathbb{Z}^n| \leq 3.14 \times 10^{23}$, whereas actually $|\Sigma^n(r) \cap \mathbb{Z}^n| = 2.88 \times 10^{23}$. By comparison, when $r = 100$ and $n = 10000$, the optimal H-bound (at $\gamma = 0.0645$) is 1.774×10^{242} integer points, while the actual number of points is 1.755×10^{242}; the relative error is about 1.1%.[24] It can be shown that $|\Sigma^n(r) \cap \mathbb{Z}^n|$ is asymptotically computed by the H-bound at $\gamma = \frac{\pi r^\delta}{\sqrt{n}}$ given that $0 < \delta < \frac{1}{2}$, $n \to \infty$, and $r = \Theta(n^\varepsilon)$ for some $\varepsilon \in (0, 1)$.

[24]The E-bound is not applicable to these simplices, since the coordinates of the typical integer point are too close to 0. The I-bound is generally not recommended for simplices; for $\Sigma^n(r)$, the I-bound is $(n+r)^{n+r-1} n^{-(n-1)} r^{-r}$, which exceeds the actual value of $|\Sigma^n(r) \cap \mathbb{Z}^n|$ by a factor of approximately $\sqrt{2\pi(n-1)r/(n+r-1)}$.

CHAPTER III

Bounded Contingency Tables

Contingency tables and K-bounded contingency tables were introduced in Section 1.1.2. As in that section, let

$$\Pi_K(R,C) := \left\{ X \in \mathbb{R}_{\geq 0}^{m\times n} : \begin{array}{l} \displaystyle\sum_{j=1}^{n} x_{ij} = r_i \quad (i = 1,\ldots,m), \\ \displaystyle\sum_{i=1}^{m} x_{ij} = c_j \quad (j = 1,\ldots,n), \\ \text{and} \quad x_{ij} \leq k_{ij} \quad \text{for all } i,j \end{array} \right\}$$

where

$$R = (r_1,\ldots,r_m) \in \mathbb{Z}_{\geq 0}^m, \qquad C = (c_1,\ldots,c_n) \in \mathbb{Z}_{\geq 0}^n,$$

$$N = r_1 + \cdots + r_m = c_1 + \cdots + c_n, \qquad \text{and} \quad K = \left(k_{ij}\right) \in (\mathbb{Z}_{\geq 0} \cup \{\infty\})^{m\times n}.$$

Let $T_K(R,C) := |\Pi_K(R,C) \cap \mathbb{Z}^{m\times n}|$ denote the number of K-bounded contingency tables with margins R, C. We abuse the notation slightly, writing $\Pi_\kappa(R,C)$ and $T_\kappa(R,C)$ with $\kappa \in \mathbb{Z}_{>0}$ in case K is the matrix with all entries equal to κ. We write $T(R,C)$ for the number of unbounded tables with the given margins (i.e., the case of $k_{ij} = \infty$ for all i,j). As in Definition II.9, we will avoid writing many results twice simply by letting the notation $\{0,1,2,\ldots,\kappa\}$ refer to $\mathbb{Z}_{\geq 0}$ when $\kappa = \infty$.

3.1 Significance testing and the independence heuristic

The following table has become a standard example in the literature on contingency tables since its first appearance in a paper of Snee [64], whose students collected the data:

	Black	**Brown**	**Red**	**Blond**	**Total**
Brown	68	119	26	7	**220**
Blue	20	84	17	94	**215**
Hazel	15	54	14	10	**93**
Green	5	29	14	16	**64**
Total	**108**	**286**	**71**	**127**	**592**

Table 3.1: Cross-tabulation of eye and hair color in a population

A geneticist wishing to decide whether there is a correlation between eye and hair color would traditionally compute the Pearson X^2 statistic for this table (with 9 degrees of freedom) and check the p-value of the corresponding χ^2 value under a hypothesis of independence. In this case, $X^2 \approx 138.29$ and $p < .01$—a conventional benchmark for strong rejection of the independence hypothesis.

It would seem, therefore, that eye color and hair color are strongly related. However, Diaconis and Efron [27] noticed that approximately 10% of all distinct 4×4 tables with $N = 592$ have X^2 smaller than that achieved by the above table. Thus (at a significance level of, say, $p = .05$) we cannot reject the hypothesis that this table was generated at random from a uniform distribution on the set of tables with $N = 592$. Diaconis and Efron discuss this and a spectrum of other hypotheses which, taking the Jaynesian view (cf. Section 2.2), may be plausibly regarded as unbiased (or "non-informative").

As suggested by the preceding example, the independence hypothesis and the uniformity hypothesis may be largely (and surprisingly) incompatible. Following a

heuristic of Good [36], let us consider the set of $m \times n$ nonnegative integer matrices with sum of entries equal to N; there are $\binom{N+mn-1}{mn-1}$ such tables. Equip this set with the uniform probability measure. Then the probability that a random sample from this set has row margin R is

$$\binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{r_i+n-1}{n-1},$$

while the probability that a random sample has column margin C is

$$\binom{N+mn-1}{mn-1}^{-1} \prod_{j=1}^{n} \binom{c_j+m-1}{m-1}.$$

If these two events were independent, then the number of tables satisfying both constraints would be

$$I(R,C) := \binom{N+mn-1}{mn-1}^{-1} \prod_{i=1}^{m} \binom{r_i+n-1}{n-1} \prod_{j=1}^{n} \binom{c_j+m-1}{m-1}.$$

However, as observed by Barvinok [8], the actual number $T(R,C)$ of tables is larger than this for most choices of R and C, even by an $\Omega(\gamma^{mn})$ factor ($\gamma > 1$) when the margins grow with m and n in a natural way (see Section 3.4 for the precise statement). This result may be interpreted as showing that most row and column margins are strongly positively correlated.

Moreover, as shown in [10], the tables with given margins R, C are in a certain sense concentrated around a (not necessarily integral) table which, in our vocabulary from Chapter II, is the expected value of the MEIM for $\Pi(R,C)$.[1] The one table with margins R, C which satisfies the independence hypothesis is the **rank 1 table**

$$X^{\text{ind}} = X^{\text{ind}}_{R,C} := \left(\frac{r_i c_j}{N}\right)_{i,j}, \tag{3.1}$$

but according to the concentration result from [10], the rank 1 table may be wildly *atypical:* for example, as $n \to \infty$, the top-left entry of the typical $n \times n$ table with

[1] The "independence" (of coordinates) in the maximum-entropy independence model is not to be confused with the "independence" in the independence hypothesis for the margins!

margins $R = C = (3n, n, n, \ldots, n)$ is known to grow linearly with n, while the corresponding entry of the rank 1 table with those margins is $O(1)$ [10].

Good's heuristic can be adapted to 0-1 contingency tables (i.e., tables with a bound of $k_{ij} = 1$ on each entry). In this case, there are $\binom{mn}{N}$ tables with the given 0-margin N. If the appearance of row margin R and column margin C were independent events, then the number of tables would be

$$I_1(R, C) := \binom{mn}{N}^{-1} \prod_{i=1}^{m} \binom{n}{r_i} \prod_{j=1}^{n} \binom{m}{c_j}.$$

Barvinok [9] showed that the actual value of $T(R, C)$ is typically *smaller* than this prediction, again by a factor exponential in mn: most row and column margins are strongly negatively correlated.

This raises a question. Unbounded contingency tables and 0-1 tables are extreme cases of uniformly bounded tables (i.e., those tables counted by $T_\kappa(R, C)$). What is the cause of the opposite correlation effects when $\kappa = \infty$ and when $\kappa = 1$, and how does the transition occur? In this chapter, we use maximum-entropy independence models to interpret, re-prove, and extend Barvinok's results; we show, in particular, that there exist families of (R, C) which are asymptotically strongly positively correlated in the presence of any entry bound $\kappa \geq 2$, though not for $\kappa = 1$. The precise statement of this result is Theorem III.21.

We also present evidence that asymptotic negative correlation can be extended to some families of margins (R, C) in the presence of any entry bound $\kappa \leq \infty$. This claim, paradoxically, is (seemingly) harder to prove because, in the presence of entry bounds, there is no analogue of the independence hypothesis to which we might compare the uniformity hypothesis. Even under the assumption $T_K(R, C) > 0$, the rank 1 table with margins R and C (3.1) does not necessarily lie in $\Pi_K(R, C)$.[2]

[2]For example, although there is a 0-1 table $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ with margins $R = C = (2, 1)$, there is no rank 1 table with

3.1.1 The independence heuristic for K-bounded tables

Inspired by Good's estimate $I(R,C)$, we consider the following estimate $I_K(R,C)$ for the number of contingency tables in $\Pi_K(R,C)$.

Notation III.1. Let $f(x_1,x_2,\ldots)$ be a polynomial or formal power series, and $\mathbf{x}^\alpha$ a monomial. Then we denote the coefficient of $\mathbf{x}^\alpha$ in $f(x_1,x_2,\ldots)$ by

$$[\mathbf{x}^\alpha]\ f(x_1,x_2,\ldots).$$

The number of K-bounded tables with given 0-margin N is

$$\mathcal{T}_K(N) := [x^N]\ \prod_{i=1}^{m}\prod_{j=1}^{n}(1+x+x^2+\cdots+x^{k_{ij}}).$$

The proportion of these having row margins R is

$$\mathcal{T}_K(N)^{-1}\ \prod_{i=1}^{m}\left\{[x^{r_i}]\ \prod_{j=1}^{n}(1+x+x^2+\cdots+x^{k_{ij}})\right\}.$$

The proportion having column margins C is

$$\mathcal{T}_K(N)^{-1}\ \prod_{j=1}^{n}\left\{[x^{c_j}]\ \prod_{i=1}^{m}(1+x+x^2+\cdots+x^{k_{ij}})\right\}.$$

If these two events were independent, then the number of K-bounded tables with margins R and C would be

$$I_K(R,C) := \frac{\prod_i\left\{[x^{r_i}]\ \prod_j(1+x+\cdots+x^{k_{ij}})\right\}\ \prod_j\left\{[x^{c_j}]\ \prod_i(1+x+\cdots+x^{k_{ij}})\right\}}{[x^N]\ \prod_i\prod_j(1+x+\cdots+x^{k_{ij}})}.$$

In case the matrix K is constant ($k_{ij}=\kappa$ for all i,j), we can write the above estimate in a nicer form by means of the following notation:

Definition III.2 ("$(\kappa+1)$-nomial coefficients" [33])**.** Let κ be a positive integer. For integers $n\geq 0$ and $0\leq r\leq n\kappa$, we denote by $\binom{n}{r}_\kappa$ the coefficient of x^r in the polynomial expansion of $(1+x+x^2+\cdots+x^\kappa)^n$.

those margins and with entries ≤ 1. The rank 1 table with those margins is $\left(\begin{smallmatrix}4/3 & 2/3\\ 2/3 & 1/3\end{smallmatrix}\right)$. The absence of a viable independence hypothesis for K-bounded contingency tables makes the uniformity hypothesis all the more attractive.

For integers $n \geq 0$, $r \geq 0$, we define $\binom{n}{r}_\infty$ to be the coefficient of x^r in the power series expansion of $(1 + x + x^2 + \cdots)^n$.

Given this definition, we have

$$I_\kappa := I_{\kappa \cdot \mathbf{1}} = \binom{mn}{N}_\kappa^{-1} \prod_{i=1}^{m} \binom{n}{r_i}_\kappa \prod_{j=1}^{n} \binom{m}{c_j}_\kappa. \tag{3.2}$$

Note that $\binom{n}{r}_1 = \binom{n}{r}$ and $\binom{n}{r}_\infty = \binom{r+n-1}{r} = \binom{r+n-1}{n-1}$. For $\kappa \neq 1, \infty$, there is (to the author's knowledge) no comparably neat exact formula for $\binom{n}{r}_\kappa$. The problem seems to be related to the difficulty of expressing $H_\kappa^{\max}(x)$, $p(x;\kappa)$, and $q(x;\kappa)$ in terms of x, which in general requires solving a degree-κ equation. The connection to entropy appears in a logarithmically asymptotic formula for $\binom{n}{r}_\kappa$, analogous in both statement and proof to Proposition II.2:

Proposition III.3. *Let $\kappa \in \mathbb{Z}_{>0} \cup \{\infty\}$. Let n, r be integers ($n > 0,\ 0 \leq r \leq n\kappa$). Then*

$$\ln \binom{sn}{sr} = snH_\kappa^{\max}\left(\frac{r}{n}\right) - \Theta(\ln s).$$

Proof. Let $X_1, X_2, \ldots$ be independent random variables, each with distribution $TG\left(\frac{r}{n};\kappa\right)$. Let $X = (X_1, \ldots, X_{sn})$.

Observe that if $\mathbf{x}, \mathbf{x}' \in \{0, 1, 2, \ldots, \kappa\}^{sn}$, then

$$\frac{\mathbf{Pr}[X = \mathbf{x}']}{\mathbf{Pr}[X = \mathbf{x}]} = q\left(\frac{r}{n};\kappa\right)^{|\mathbf{x}'|-|\mathbf{x}|}$$

(where $|\mathbf{x}| := \sum_{i=1}^{sn} x_i$). In particular, all values of X with equal sum of coordinates

are equiprobable. Let $\mathbf{x}_*$ denote an arbitrary value of X satisfying $|\mathbf{x}_*| = sr$. Thus

$$\begin{aligned} snH_\kappa^{\max}\left(\frac{r}{n}\right) = \mathbf{H}[X] &= \mathbf{E}_{\mathbf{x}}[I(X,\mathbf{x})] \\ &= I(\mathbf{x}_*) - \left(\ln\left(\frac{r}{n};\kappa\right)\right)\mathbf{E}\big[|X| - sr\big] \\ &= I(\mathbf{x}_*) \\ &= -\ln\left[\binom{sn}{sr}_\kappa^{-1} \cdot \mathbf{Pr}\big[|X| = sr\big]\right] \\ &= \ln\binom{sn}{sr}_\kappa - \ln \mathbf{Pr}\big[|X| = sr\big]. \end{aligned} \tag{3.3}$$

Note that the probability mass function for each X_i is log-concave on $\mathbb{Z}$. We apply Theorem II.32 (Bender's local limit theorem) using

$$\zeta_p = X_1 + \cdots + X_p, \qquad \sigma_p^2 = p \cdot \mathbf{Var}(X_1), \qquad \mu_p = p \cdot \frac{r}{n}, \qquad \text{and} \qquad x = 0,$$

with the normality hypothesis secured via Theorem II.33, to infer

$$\lim_{p\to\infty} \sigma_p \mathbf{Pr}\left[\zeta_p = \left\lfloor p \cdot \frac{r}{n} \right\rfloor\right] = \frac{1}{\sqrt{2\pi}},$$

and thence

$$\mathbf{Pr}\big[|X| = sr\big] \quad \sim \quad (2\pi sn\mathbf{Var}(X_1))^{-1/2} \quad = \quad \Theta(s^{-1/2}).$$

Substituting into (3.3), we conclude that

$$snH_\kappa^{\max}\left(\frac{r}{n}\right) = \ln\binom{sn}{sr}_\kappa + \Theta(\ln s),$$

proving the proposition. □

3.2 Counting contingency tables via permanents

The following result is due to Barvinok:

Theorem III.4 ([7], paraphrased)**.** *Take m, n, R, C, N as heretofore, and let $W = \big(w_{ij}\big) \in \mathbb{R}^{m\times n}$. Let $\Gamma = \big(\gamma_{ij}\big)$ be a random $m\times n$ matrix whose entries are independent exponential random variables of mean 1.*[3] *Let $A = A(\Gamma)$ be the $N \times N$ matrix formed*

[3]The distribution function of an exponential random variable of mean t is $F(x) = 1 - e^{-x/t}$ $(x \geq 0)$.

by replacing each entry (i, j) of Γ by an $r_i \times c_j$ block with all entries equal to $w_{ij}\gamma_{ij}$. Let each contingency table X with margins R and C be counted with the weight

$$w(X) = \prod_{i=1}^{m} \prod_{j=1}^{n} w_{ij}^{x_{ij}}.$$

Then the total weight of all such tables is

$$T(R, C; W) = \frac{\mathbf{E}[\operatorname{per} A]}{r_1! \cdots r_m! c_1! \cdots c_n!}.$$

Exact computation of the factor $\mathbf{E}[\operatorname{per} A]$ is intractable, but estimation is possible. The following strategy is again due to Barvinok (ibid.): For A with all entries positive (which occurs with probability 1), there exist "scaling factors" $\xi_1, \ldots, \xi_N, \eta_1, \ldots, \eta_N$ such that the matrix $A^{\text{scaled}} := \left(\xi_i^{-1} \eta_j^{-1} a_{ij}\right)$ is doubly stochastic, that is, has all row and column margins equal to 1. Letting $\sigma(A) := \prod_{i=1}^{N} \xi_i \prod_{j=1}^{N} \eta_j$, we have

$$\operatorname{per} A = \sigma(A) \operatorname{per} A^{\text{scaled}}$$

(by row- and column-linearity of the permanent). It turns out that $\sigma(A)$ is log-concave and efficiently computable (and integrable), while $\operatorname{per} A^{\text{scaled}}$ can be bounded to within a relative error of $N^{O(m+n)}$ by means of van der Waerden-Falikman-Egorychev's and Minc-Brégman's permanent inequalities (see [8]). The number of contingency tables with margins at least linear in m and n is exponential in mn, so the above strategy succeeds in estimating this number asymptotically in the logarithm.

3.2.1 Counting K-bounded tables

Barvinok pointed out[4] that, for $K \in \mathbb{Z}_{\geq 0}^{m \times n}$, the number $T_K(R, C)$ of bounded tables can also be expressed as the expectation of a random permanent. Let us

[4] Private communication, October 2008.

define

$$\tilde{c}_j = \left(\sum_{i=1}^{m} k_{ij} \right) - c_j \tag{3.4}$$

for $1 \leq j \leq n$. Then

$$T_K(R,C) = [x_1^{r_1} \cdots x_m^{r_m} y_1^{\tilde{c}_1} \cdots y_n^{\tilde{c}_n}] \; \mathbf{E} \left[\prod_{i=1}^{m} \prod_{j=1}^{n} \frac{(\xi_{ij} x_i + \eta_{ij} y_j)^{k_{ij}}}{k_{ij}!} \right],$$

where ξ_{ij}, η_{ij} $(1 \leq i \leq m,\ 1 \leq j \leq n)$ are independent exponential random variables of mean 1. The coefficient of a monomial in a product of $|K|$ linear forms can be expressed as the permanent of a $|K| \times |K|$ matrix whose entries are the coefficients of the forms.

However, we do not take this approach, instead preferring to represent a K-bounded contingency table by an enlarged $((m+n) \times (mn))$ table with enforced zeroes, in the following fashion:

Define vectors $\mathcal{R} \in \mathbb{Z}_{\geq 0}^{m+n}, \mathcal{C} \in \mathbb{Z}_{\geq 0}^{mn}$ by

$$\mathcal{R} = (r_1, \ldots, r_m,\ \tilde{c}_1, \ldots, \tilde{c}_n),$$

$$\mathcal{C} = (k_{11}, \ldots, k_{1n},\ k_{21}, \ldots, k_{2n},\ \ldots,\ k_{m1}, \ldots, k_{mn}).$$

Observe that $\mathcal{R}$ and $\mathcal{C}$ have equal sum of entries.

Let $W = (w_{\cdot,\cdot})$ be the $(m+n) \times (mn)$ matrix with

$$w_{i,(i-1)n+j} = 1 \qquad \text{for all } i = 1, \ldots, m \text{ and } j = 1, \ldots, n,$$

$$w_{m+j,(i-1)n+j} = 1 \qquad \text{for all } i = 1, \ldots, m \text{ and } j = 1, \ldots, n,$$

and zeroes in all other positions (which we've seen before: cf. (1.1)).

Given a contingency table $X = \left(x_{ij}\right) \in \Pi_K(R,C)$, we may construct a table $\mathcal{X} = \left(x'_{\cdot,\cdot}\right) \in \Pi(\mathcal{R},\mathcal{C})$ by assigning

$$x'_{i,(i-1)n+j} = x_{ij} \qquad \text{for all } i = 1, \ldots, m \text{ and } j = 1, \ldots, n,$$

$$x'_{m+j,(i-1)n+j} = k_{ij} - x_{ij} \qquad \text{for all } i = 1, \ldots, m \text{ and } j = 1, \ldots, n,$$

and zeroes in all other positions. This conversion is easily reversed, and thus gives a bijection between tables $X \in \Pi_K(R, C)$ and tables $\mathcal{X} \in \Pi(\mathcal{R}, \mathcal{C})$ which have enforced zeroes in all zero positions of W. That is,

$$T_K(R, C) \quad = \quad T(\mathcal{R}, \mathcal{C}; W). \tag{3.5}$$

Therefore, we can count K-bounded tables using Theorem III.4. *A priori,* we might expect the quality of the estimate to be degraded by the enlargement of the dimensions. However, we will show that the estimates produced by this approach are still asymptotic in the logarithm when the margins grow linearly with m and n, and are still accurate enough to detect a correlation phenomenon as announced earlier.

3.2.2 Approximate log-concavity of $T_K(R, C)$

In fact, we do not use Theorem III.4 directly, but one of its consequences:

Notation III.5. For a vector or matrix V, let $|V|$ denote the sum of the entries of V. For an integer $n \geq 0$, let $\omega(n) := \frac{n^n}{n!}$ (agreeing that $0^0 = 1$). For a vector or matrix V with nonnegative integer entries, let $\Omega(V)$ denote the sum of $\omega(v)$ over all entries v of V.

Theorem III.6 (Barvinok [6])**.** *Define $T(R, C; W)$ as in Theorem III.4. Let $R^1, \ldots, R^p \in \mathbb{Z}^m_{\geq 0}$ and $C^1, \ldots, C^p \in \mathbb{Z}^n_{\geq 0}$, such that*

$$|R^1| = \cdots = |R^p| = |C^1| = \cdots = |C^p| = N.$$

Let

$$R := \alpha_1 R^1 + \alpha_2 R^2 + \cdots + \alpha_p R^p \quad \textit{and} \quad C := \alpha_1 C^1 + \alpha_2 C^2 + \cdots + \alpha_p C^p,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_p \geq 0$ satisfy $\alpha_1 + \alpha_2 + \cdots + \alpha_p = 1$. Then

$$\frac{\omega(N) T(R, C; W)}{\Omega(R)\Omega(C)} \quad \geq \quad \prod_{t=1}^{p} \left[\frac{T(R^t, C^t; W)}{\min\{\Omega(R^t), \Omega(C^t)\}} \right]^{\alpha_t}.$$

By means of the "enlargement" discussed in the prior section, we derive as a corollary of this theorem the following version for $T_K(R, C)$:

Theorem III.7. [5] *Take R_t, C_t, α_t $(1 \le t \le p)$ and R, C as in the hypotheses of Theorem III.6. Define*

$$\tilde{C} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$$

as in (3.4), and similarly define $\tilde{C}^1 = (\tilde{c}_1^1, \ldots, \tilde{c}_n^1), \ldots, \tilde{C}^p = (\tilde{c}_1^p, \ldots, \tilde{c}_n^p) \in \mathbb{Z}_{\ge 0}^n$ by

$$\tilde{c}_j^t = \left(\sum_{i=1}^{m} k_{ij} \right) - c_j^t \qquad (1 \le t \le p,\ 1 \le j \le n).$$

Then

$$\frac{\omega(|K|) T_K(R, C)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \quad \ge \quad \prod_{t=1}^{p} \left[\frac{T_K(R^t, C^t)}{\min\{\Omega(R^t)\Omega(\tilde{C}^t),\ \Omega(K)\}} \right]^{\alpha_t}.$$

This theorem is somewhat opaque in itself, due to the confounding factors $\Omega(R^t)$, $\Omega(\tilde{C}^t)$, etc. However, some analysis in Section 3.3 will reveal that these factors typically grow more slowly than the numbers $T_K(R, C)$ themselves.

Proof of Theorem III.7. Define vectors $\mathcal{R} \in \mathbb{Z}_{\ge 0}^{m+n}$ and $\mathcal{C} \in \mathbb{Z}_{\ge 0}^{mn}$ as in Section 3.2.1, and define $\mathcal{R}^1, \mathcal{R}^2, \ldots, \mathcal{R}^p \in \mathbb{Z}_{\ge 0}^{m+n}$ analogously:

$$\mathcal{R}^t = (r_1^t, \ldots, r_m^t,\ \tilde{c}_1^t, \ldots, \tilde{c}_n^t).$$

Observe that

$$|\mathcal{R}^1| = |\mathcal{R}^2| = \cdots = |\mathcal{R}^p| = |\mathcal{C}| = |K|$$

and that

$$\mathcal{R} = \alpha_1 \mathcal{R}^1 + \alpha_2 \mathcal{R}^2 + \cdots + \alpha_p \mathcal{R}^p.$$

Take W as in Section 3.2.1, so that (as discussed there)

$$T_K(R, C) \quad = \quad T(\mathcal{R}, \mathcal{C}; W) \tag{3.6}$$

[5] The theorem can be stated in slightly greater generality with only trivial modifications to the proof. Specifically, $T_K(R, C)$ can be replaced by a weighted function $T_K(R, C; W)$, analogous to the function $T(R, C; W)$ in the statement of Theorem III.6; also, given

$$K = \alpha_1 K^1 + \alpha_2 K^2 + \cdots + \alpha_p K^p,$$

Theorem III.7 remains true when each instance of K on the right-hand side is replaced by K^t.

and

$$T_K(R^t, C^t) \quad = \quad T(\mathcal{R}^t, \mathcal{C}; W) \tag{3.7}$$

for $1 \leq t \leq p$.

Substituting $\mathcal{R}^1, \ldots, \mathcal{R}^p, \mathcal{R}$ for $R^1, \ldots, R^p, R$ in the statement of Theorem III.6, as well as $\mathcal{C}$ for each of $C^1, \ldots, C^p, C$ and $|K|$ for N, we obtain the conclusion

$$\frac{\omega(|K|)T(\mathcal{R}, \mathcal{C}; W)}{\Omega(\mathcal{R})\Omega(\mathcal{C})} \quad \geq \quad \prod_{t=1}^{p} \left[\frac{T(\mathcal{R}^t, \mathcal{C}; W)}{\min\{\Omega(\mathcal{R}^t), \Omega(\mathcal{C})\}} \right]^{\alpha_t}.$$

Using equations (3.6) and (3.7), we rewrite the above result as

$$\frac{\omega(|K|)T_K(R, C)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \quad \geq \quad \prod_{t=1}^{p} \left[\frac{T_K(R^t, C^t)}{\min\{\Omega(R^t)\Omega(\tilde{C}^t),\ \Omega(K)\}} \right]^{\alpha_t},$$

proving Theorem III.7. ■

Remark III.8. It is no exaggeration to state that all of our results about $T_K(R, C)$, including Theorems III.16, III.19, and III.21, flow from the above theorem. Log-concavity turns out to be a powerful property. Although Barvinok derived the approximate log-concavity of $T(R, C)$ from the permanental bounds, our arguments work in the opposite direction, showing that these results are essentially equivalent. Thus if Theorem III.7 could be proved by purely combinatorial means, as seems not implausible, then the other results about $T_K(R, C)$ would also be placed on a combinatorial foundation.

It may be possible to strengthen Theorem III.7 considerably. We are not aware of any counterexamples to the hypothesis that $T_K(R, C)$ is actually (rather than approximately) log-concave as a function of R, C.

3.2.3 An honestly concave proxy for $\ln T_K(R, C)$

We define a function which "smooths over" $\ln T_K(R, C)$:

Definition III.9. For $R \in \mathbb{R}^m_{\geq 0}$, $C \in \mathbb{R}^n_{\geq 0}$, and $K \in \mathbb{Z}^{m\times n}_{>0}$, let

$$f(R,C) = f_K(R,C) := \max_{\substack{\alpha_1,\ldots,\alpha_p \geq 0 \\ \alpha_1+\cdots+\alpha_p=1 \\ \alpha_1 R^1+\cdots+\alpha_p R^p=R \\ \alpha_1 C^1+\cdots+\alpha_p C^p=C}} \sum_{t=1}^{p} \alpha_t \ln T_K(R^t, C^t).$$

(To be clear, the maximum is taken over choices of $p \geq 1$, $\alpha_1, \ldots, \alpha_p$, $R^1, \ldots, R^p$, and $C^1, \ldots, C^p$ which satisfy the indicated constraints, and for which the summation on the right is defined. If the maximum is taken over an empty set, then we regard it as $-\infty$.)

Note that the maximum in Definition III.9 is well-defined (allowing $-\infty$ as "well-defined"), because there are finitely many pairs (R, C) for which $T_K(R, C) > 0$. It is redundant to allow any repetition among $R^1, \ldots, R^p$ or $C^1, \ldots, C^p$, so the summation on the right takes on finitely many values.

Lemma III.10. (i) $f(R,C) \geq \ln T_K(R,C)$.

(ii) *f is concave.*

(iii) *The domain of f (i.e., where $f > -\infty$) is a subset of $\Pi_K(R,C)$.*

Proof.[6] Claim (i) is trivial, since we can set $p = 1$, $\alpha_1 = 1$, $R^1 = R$, $C^1 = C$ in Definition III.9.

For claim (ii), it suffices to show that if $\alpha + \beta = 1$, then

$$\alpha f(R^1, C^1) + \beta f(R^2, C^2) \leq f(\alpha R^1 + \beta R^2, \alpha C^1 + \beta C^2). \tag{3.8}$$

By Definition III.9, there exist $\gamma_1, \ldots, \gamma_p \geq 0$; $R^{11}, \ldots, R^{1p}$; and $C^{11}, \ldots, C^{1p}$ such

[6] The unavoidably cumbersome notation used in this proof may distract the reader from the fact that the proof is utterly conventional.

that

$$\sum_{t=1}^{p} \gamma_t = 1, \qquad \sum_{t=1}^{p} \gamma_t R^{1t} = R, \qquad \sum_{t=1}^{p} \gamma_t C^{1t} = C,$$
$$\text{and} \quad f(R^1, C^1) = \sum_{t=1}^{p} \gamma_t \ln T_K(R^{1t}, C^{1t}).$$

Likewise, there exist $\delta_1, \ldots, \delta_q \geq 0$; $R^{21}, \ldots, R^{2q}$; and $C^{21}, \ldots, C^{2q}$ such that

$$\sum_{t=1}^{q} \delta_t = 1, \qquad \sum_{t=1}^{q} \delta_t R^{2t} = R, \qquad \sum_{t=1}^{q} \delta_t C^{2t} = C,$$
$$\text{and} \quad f(R^2, C^2) = \sum_{t=1}^{q} \delta_t \ln T_K(R^{2t}, C^{2t}).$$

Note that

$$\sum_{t=1}^{p} \alpha\gamma_t + \sum_{t=1}^{q} \beta\delta_t = 1, \qquad \sum_{t=1}^{p} \alpha\gamma_t R^{1t} + \sum_{t=1}^{q} \beta\delta_t R^{2t} = \alpha R^1 + \beta R^2,$$
$$\text{and} \qquad \sum_{t=1}^{p} \alpha\gamma_t C^{1t} + \sum_{t=1}^{q} \beta\delta_t C^{2t} = \alpha C^1 + \beta C^2;$$

applying Definition III.9 to $f(\alpha R^1 + \beta R^2, \alpha C^1 + \beta C^2)$, we obtain equation (3.8) and thus claim (ii).

It is clear that f is defined only on the convex hull of all (R, C) for which $T_K(R, C) > 0$; this region is a subset of $\Pi_K(R, C)$, proving claim (iii). □

Lemma III.11 (Quality of approximation)**.** *Suppose $R \in \mathbb{Z}^m_{>0}$, $C \in \mathbb{Z}^n_{\geq 0}$, and $K \in \mathbb{Z}^{m\times n}_{>0}$. Define $\tilde{C} = (\tilde{c}_1, \ldots, \tilde{c}_n)$ as in (3.4), and suppose that $\tilde{C} \in \mathbb{Z}^n_{>0}$. Then*

$$f_K(R, C) - \ln T_K(R, C) \quad \leq \quad -\ln\sqrt{2\pi|K|} + \sum_{i=1}^{m} \ln\sqrt{2\pi r_i} + \sum_{j=1}^{n} \ln\sqrt{2\pi\tilde{c}_j} + (m+n)\ln\left(\frac{e}{\sqrt{2\pi}}\right).$$

Proof. By Stirling's formula,

$$n - \ln\sqrt{2\pi n} - \ln\left(\frac{e}{\sqrt{2\pi}}\right) \quad \leq \quad \ln\omega(n) \quad \leq \quad n - \ln\sqrt{2\pi n} \tag{3.9}$$

for $n \geq 1$.

Choose $\alpha_1, \ldots, \alpha_p, R^1, \ldots, R^p, C^1, \ldots, C^p$ which achieve the maximum in Definition III.9. Now apply Theorem III.7 and (3.9):

$$\begin{aligned}
f_K(R,C) - \ln T_K(R,C) \;&\leq\; \ln\left[\frac{\omega(|K|)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \cdot \prod_{t=1}^{p}\Big(\min\{\Omega(R^t)\Omega(\tilde{C}^t),\ \Omega(K)\}\Big)^{\alpha_t}\right] \\
&\leq \quad \ln\left[\frac{\omega(|K|)}{\Omega(R)\Omega(\tilde{C})\Omega(K)} \cdot \prod_{t=1}^{p}\Omega(K)^{\alpha_t}\right] \\
&= \quad \ln\frac{\omega(|K|)}{\Omega(R)\Omega(\tilde{C})} \\
&= \quad \ln\omega(|K|) - \sum_{i=1}^{m}\ln\omega(r_i) - \sum_{j=1}^{n}\ln\omega(\tilde{c}_j) \\
&\leq \quad |K| - \ln\sqrt{2\pi|K|} - \sum_{i}\left(r_i - \ln\sqrt{2\pi r_i} - \ln\left(\frac{e}{\sqrt{2\pi}}\right)\right) \\
&\qquad\qquad - \sum_{j}\left(\tilde{c}_j - \ln\sqrt{2\pi\tilde{c}_j} - \ln\left(\frac{e}{\sqrt{2\pi}}\right)\right) \\
&\leq \quad -\ln\sqrt{2\pi|K|} + \sum_{i=1}^{m}\ln\sqrt{2\pi r_i} + \sum_{j=1}^{n}\ln\sqrt{2\pi\tilde{c}_j} + (m+n)\ln\left(\frac{e}{\sqrt{2\pi}}\right). \quad \square
\end{aligned}$$

3.3 Asymptotic formulas for $\ln T_K(R,C)$

In this section, we present two approximate formulas for $T_K(R,C)$ (Theorems III.16, III.19), analogous to results for unbounded tables appearing in [8]. Both formulas are logarithmically asymptotic to the actual count in an asymptotic regime which we now define:

Definition III.12 (Cloning)**.** Let

$$R = (r_1, \ldots, r_m) \in \mathbb{Z}_{\geq 0}^m \quad \textit{and} \quad C = (c_1, \ldots, c_m) \in \mathbb{Z}_{\geq 0}^n.$$

Then we define

$$R^{(s)} = (sr_1, \ldots, sr_m,\ sr_1, \ldots, sr_m,\ \ldots,\ sr_1, \ldots, sr_m)$$

and

$$C^{(s)} = (sc_1, \ldots, sc_n,\ sc_1, \ldots, sc_n,\ \ldots,\ sc_1, \ldots, sc_n),$$

where the number of repetitions is s (thus $R^{(s)} \in \mathbb{Z}_{\geq 0}^{sm}$ and $C^{(s)} \in \mathbb{Z}_{\geq 0}^{sn}$). We refer to these vectors as the **s-fold clonings** of R and C.

If $K \in \mathbb{Z}_{\geq 0}^{m \times n}$, then we define $K^{(s)}$ as the $sm \times sn$ matrix of form

$$\begin{pmatrix} K & K & \cdots & K \\ K & K & \cdots & K \\ \vdots & \vdots & \ddots & \vdots \\ K & K & \cdots & K \end{pmatrix}$$

(with s blocks in either direction). We call this the **s-fold cloning** of K.

Note that the clonings are defined so that, if X is a contingency table with margins R and C, then the $sm \times sn$ matrix

$$\begin{pmatrix} X & X & \cdots & X \\ X & X & \cdots & X \\ \vdots & \vdots & \ddots & \vdots \\ X & X & \cdots & X \end{pmatrix}$$

has margins $R^{(s)}$ and $C^{(s)}$.

3.3.1 Exact and approximate generating functions for tables

Definition III.13. Given $K = \left(k_{ij}\right) \in \mathbb{Z}_{\geq 0}^{m \times n}$, define the polynomial

$$G(\mathbf{x}, \mathbf{y}) = G_K(\mathbf{x}, \mathbf{y}) := \prod_{i=1}^{m} \prod_{j=1}^{n} \left[1 + x_i y_j + (x_i y_j)^2 + \cdots + (x_i y_j)^{k_{ij}}\right]$$

(where $\mathbf{x} = (x_1, \ldots, x_m),\ \mathbf{y} = (y_1, \ldots, y_n)$).

Trivially, G is a generating function for K-bounded contingency tables; that is,

$$G(\mathbf{x}, \mathbf{y}) = \sum_{R} \sum_{C} T_K(R, C) \mathbf{x}^R \mathbf{y}^C, \tag{3.10}$$

where the sum is taken over all margins (R, C) of lengths m and n (of which finitely many contribute a nonzero coefficient).[7] In principle, we can "compute" $T_K(R, C)$ by expanding $G(\mathbf{x}, \mathbf{y})$ and extracting the coefficient of $\mathbf{x}^R\mathbf{y}^C$. This is of course not practical, but we might estimate this coefficient by

$$\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R\mathbf{y}^C};$$

indeed, this is an *upper* bound on $T_K(R, C)$, as may be readily seen by dividing both sides of (3.10) by $\mathbf{x}^R\mathbf{y}^C$. To bound $T_K(R, C)$ from the other side, we replace $G(\mathbf{x}, \mathbf{y})$ by an approximate version with smoother coefficients:

Definition III.14. Let

$$\tilde{G}(\mathbf{x}, \mathbf{y}) := \sum_R \sum_C e^{f(R,C)}\mathbf{x}^R\mathbf{y}^C,$$

where the sum is taken over all integer margins (R, C) such that $f(R, C) > -\infty$.

(See Definition III.9 for the meaning of $f(R, C)$.)

We will find the following lemma useful, as it will allow us to pick out any nonzero term of $\tilde{G}(\mathbf{x}, \mathbf{y})$ as the largest:

Lemma III.15. [8] *For any (R_*, C_*) in the relative interior of the domain of f, there exist $\mathbf{x}_*, \mathbf{y}_* > 0$ such that the function*

$$\Phi(R, C) := e^{f(R,C)}\mathbf{x}_*^R\mathbf{y}_*^C$$

attains its maximum at $R = R_*$, $C = C_*$.

Proof. Recall that f is concave; therefore, its graph has a supporting hyperplane over (R_*, C_*). Let such a hyperplane have outward-pointing normal vector

[7] We use $\mathbf{x}^R$ and $\mathbf{y}^C$ as shorthand for $x_1^{r_1}x_2^{r_2}\cdots x_m^{r_m}$ and $y_1^{c_1}y_2^{c_2}\cdots y_n^{c_n}$, respectively.

[8] We call this the *tilting lemma*, as it merely reflects the fact that a convex body can be tilted so as to designate any arbitrary point as the summit.

$(u_1, \ldots, u_m, v_1, \ldots, v_n, 1)$. Set

$$\mathbf{x}_* = (x_1, \ldots, x_m) = (e^{-u_1}, \ldots, e^{-u_m}) \quad \text{and} \quad \mathbf{y}_* = (y_1, \ldots, y_n) = (e^{-v_1}, \ldots, e^{-v_n}).$$

Then

$$\phi(R, C) := f(R, C) + \sum_{i=1}^{m} r_i \ln x_i + \sum_{j=1}^{n} c_j \ln y_j$$

is concave with respect to R and C, and attains a critical point (hence its unique global maximum) at (R_*, C_*). Therefore, so does $\Phi(R, C) = e^{\phi(R,C)}$. $\square$

3.3.2 A generating-function-based formula for $\ln T_K(R, C)$

We now give the first of our two estimates:

Theorem III.16. *Let $R \in \mathbb{Z}_{>0}^m$, $C \in \mathbb{Z}_{>0}^n$, and $K \in \mathbb{Z}_{>0}^{m \times n}$.*

Assume that $T_K(R, C) > 0$, that is, there is at least one contingency table with margins R and C, bounded entrywise by K. Then

$$\lim_{s \to \infty} \frac{1}{s^2} \ln T_{K^{(s)}}(R^{(s)}, C^{(s)}) \quad = \quad \ln \left(\inf_{\substack{x_1, \ldots, x_m > 0 \\ y_1, \ldots, y_n > 0}} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right),$$

where $G(\mathbf{x}, \mathbf{y})$ is as in Definition III.13 and $R^{(s)}, C^{(s)}, K^{(s)}$ are as in Definition III.12.

Proof. Using Lemma III.15, choose $\mathbf{x}_*, \mathbf{y}_*$ so that $e^{f(R,C)} \mathbf{x}^R \mathbf{y}^C$ is the largest term in the expansion of $\tilde{G}(\mathbf{x}, \mathbf{y})$, evaluated at $\mathbf{x} = \mathbf{x}_*$ and $\mathbf{y} = \mathbf{y}_*$. Thus

$$\frac{\tilde{G}(\mathbf{x}_*, \mathbf{y}_*)}{\mathbf{x}_*^R \mathbf{y}_*^C} \leq [\text{\# of terms of } \tilde{G} \text{ with nonzero coeffs.}] \cdot e^{f(R,C)}.$$

The number of terms of $\tilde{G}$ is at most

$$\mathcal{N} := \prod_{i=1}^{m} \left(1 + \sum_{j=1}^{n} k_{ij} \right) \cdot \prod_{j=1}^{n} \left(1 + \sum_{i=1}^{m} k_{ij} \right),$$

since $T_K(R, C) > 0$ implies that R and C do not exceed the margins of K.

Let the symbol $\heartsuit$ denote the quantity

$$- \ln \sqrt{2\pi |K|} + \sum_{i=1}^{m} \ln \sqrt{2\pi r_i} + \sum_{j=1}^{n} \ln \sqrt{2\pi \tilde{c}_j} + (m + n) \ln \left(\frac{e}{\sqrt{2\pi}} \right),$$

last seen in Lemma III.11.

We deduce the following chain of inequalities:

$$\begin{aligned}
\ln\left(\inf_{x_i,y_j>0}\frac{G(\mathbf{x},\mathbf{y})}{\mathbf{x}^R\mathbf{y}^C}\right) &\geq \ln T_K(R,C)\\
&\geq f(R,C)-\heartsuit\\
&\geq \ln\left(\frac{\tilde{G}(\mathbf{x}_*,\mathbf{y}_*)}{\mathbf{x}_*^R\mathbf{y}_*^C\cdot\mathcal{N}}\right)-\heartsuit\\
&\geq \ln\left(\inf_{x_i,y_j>0}\frac{\tilde{G}(\mathbf{x},\mathbf{y})}{\mathbf{x}^R\mathbf{y}^C\cdot\mathcal{N}}\right)-\heartsuit\\
&\geq \ln\left(\inf_{x_i,y_j>0}\frac{G(\mathbf{x},\mathbf{y})}{\mathbf{x}^R\mathbf{y}^C}\right)-\ln\mathcal{N}-\heartsuit. \qquad (3.11)
\end{aligned}$$

Now we consider the cloning of the margins. Let $G^{(s)}$ denote the generating function for $K^{(s)}$-bounded contingency tables. Letting

$$\begin{aligned}
\mathbf{x}^{(s)} &:= (\mathbf{x}^1,\mathbf{x}^2,\ldots,\mathbf{x}^s)\\
&= (x_1^1,\ldots,x_m^1,\ x_1^2,\ldots,x_m^2,\ \ldots,\ x_1^s,\ldots,x_m^s),
\end{aligned}$$

and defining $\mathbf{y}^{(s)}$ similarly, we note that

$$\frac{G^{(s)}(\mathbf{x}^{(s)},\mathbf{y}^{(s)})}{[\mathbf{x}^{(s)}]^{R^{(s)}}[\mathbf{y}^{(s)}]^{C^{(s)}}} = \prod_{k=1}^{s}\prod_{\ell=1}^{s}\frac{G(\mathbf{x}^k,\mathbf{y}^\ell)}{(\mathbf{x}^k)^R(\mathbf{y}^\ell)^C}.$$

From this it follows that

$$\frac{1}{s^2}\ln\left(\inf_{x_i^k,y_j^\ell>0}\frac{G^{(s)}(\mathbf{x}^{(s)},\mathbf{y}^{(s)})}{[\mathbf{x}^{(s)}]^{R^{(s)}}[\mathbf{y}^{(s)}]^{C^{(s)}}}\right) = \ln\left(\inf_{x_i,y_j>0}\frac{G(\mathbf{x},\mathbf{y})}{\mathbf{x}^R\mathbf{y}^C}\right)$$

for all $s\geq 1$.

Inspection of the formulas for $\ln\mathcal{N}$ and $\heartsuit$ shows that both of these terms from (3.11) have growth of order $O(s\ln s)$ as $s\to\infty$. Therefore, by (3.11),

$$\frac{1}{s^2}\ln T_{K^{(s)}}(R^{(s)},C^{(s)}) = \ln\left(\inf_{x_i,y_j>0}\frac{G(\mathbf{x},\mathbf{y})}{\mathbf{x}^R\mathbf{y}^C}\right)+O\left(\frac{\ln s}{s}\right),$$

from which Theorem III.16 follows. ■

Remark III.17. The "cloning" limit, which will also appear in the statement of our second estimate (Theorem III.19), is an artifice designed to enforce the linear growth of the margins as $m, n \to \infty$, so that we can state our estimates for $\ln T_K(R, C)$ as asymptotic formulas. One may wonder if there is a less rigid limit in which the estimate

$$\ln T_K(R, C) \sim \ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right)$$

(or

$$\ln T_K(R, C) \sim \max_{Z \in \Pi_K(R,C)} \sum_i \sum_j H_{k_{ij}}^{\max}(z_{ij}),$$

anticipating Theorem III.19) holds.

We offer the following answer. Suppose $d|n$. Given a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ satisfying

$$\max\{v_1, v_2, \ldots, v_n\} \leq \frac{d}{n}(v_1 + v_2 + \cdots + v_n), \tag{3.12}$$

Lemma II.29 (which we previously used for a totally different purpose) implies that we can obtain $\mathbf{v}$ as a convex combination of (n/d)-fold clonings of d-vectors.[9] It thus follows from Theorems III.7 and III.16 that

$$\ln T_K(R, C) = \ln \left(\inf_{x_i, y_j > 0} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right) + O(\max\{m, n\} \ln \max\{m, n\}) \tag{3.13}$$

uniformly for k_{ij} varying between fixed positive bounds and R, C satisfying

$$\max\{r_1, \ldots, r_m\} \leq \frac{d}{m}(r_1 + \cdots + r_m), \quad \max\{c_1, \ldots, c_n\} \leq \frac{d}{n}(c_1 + \cdots + c_n)$$

with d fixed. This condition also ensures (barring the degenerate case of $R = C = \mathbf{0}$) that $T_K(R, C) = e^{\Omega(mn)}$, so that the "main term" on the right-hand side of (3.13) is in fact dominant as $m, n \to \infty$ simultaneously.

[9] We can obtain $\mathbf{v}$ as a combination of at most n such vectors, since polytopes are triangulable. Moreover, efficient algorithms for this decomposition exist.

3.3.3 A maximum-entropy formula for $\ln T_K(R, C)$

Per Proposition II.13, the MEIM[10] associated to $\Pi_K(R, C)$ is a matrix $X = \big(x_{ij}\big)$ whose entries are independent $TG(z_{ij}; k_{ij})$ random variables, where $Z = \big(z_{ij}\big)$ is whichever point of $\Pi_K(R, C)$ maximizes the entropy

$$\mathbf{H}[X] = \sum_{i=1}^{m} \sum_{j=1}^{n} H_{k_{ij}}^{\max}(z_{ij}). \tag{3.14}$$

We know that the MEIM assigns equal mass to all *bona fide* integer points of $\Pi_K(R, C)$, while also awarding some mass to impostors outside this polytope. Thus formula (3.14) must overestimate the entropy of the uniform distribution on $\Pi_K(R, C) \cap \mathbb{Z}^{m \times n}$, and so provides an upper bound on $\ln T_K(R, C)$. However, this upper bound turns out to be asymptotically accurate in the cloning limit, as the following result implies:

Lemma III.18. *Extending the notation of Theorem III.16, we have*

$$\ln \left(\inf_{\substack{x_1,\ldots,x_m>0 \\ y_1,\ldots,y_n>0}} \frac{G(\mathbf{x}, \mathbf{y})}{\mathbf{x}^R \mathbf{y}^C} \right) = \max_{Z \in \Pi_K(R,C)} \sum_{i=1}^{m} \sum_{j=1}^{n} H_{k_{ij}}^{\max}(z_{ij}). \tag{3.15}$$

Proof. By Proposition II.14, $H_{k_{ij}}^{\max}(x)$ is strictly concave for all i, j. Also, $(H_{k_{ij}}^{\max})'(x) = -\ln q(x; k_{ij})$ approaches ∞ as $x \to 0$ and $-\infty$ as $x \to k_{ij}$. It follows that the maximum on the right-hand side of (3.15) is well-defined and is attained in the relative interior of $\Pi_K(R, C)$. For the remainder of this proof, let Z denote the (unique) point at which the maximum is attained, and let $p_{ij} := p(z_{ij}; k_{ij})$, $q_{ij} := q(z_{ij}; k_{ij})$.

Since Z is in the interior of $\Pi_K(R, C)$, the local defining equations for $\Pi_K(R, C)$ at Z are just

$$\sum_{j=1}^{n} a_{ij} = r_i \ (1 \le i \le m) \quad \text{and} \quad \sum_{i=1}^{m} a_{ij} = c_j \ (1 \le j \le n).$$

[10] See Definition II.6.

Introducing Lagrange multipliers for these constraints, we infer that $\ln q_{ij} = \lambda_i + \mu_j$ for some constants $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_n$. Define $\xi_i := e^{\lambda_i}$, $\eta_j = e^{\mu_j}$; thus $q_{ij} = \xi_i \eta_j$. Dividing equation (2.4) by equation (2.3), we obtain

$$z_{ij} = \frac{\xi_i\eta_j + 2(\xi_i\eta_j)^2 + \cdots + k_{ij}(\xi_i\eta_j)^{k_{ij}}}{1 + \xi_i\eta_j + (\xi_i\eta_j)^2 + \cdots + (\xi_i\eta_j)^{k_{ij}}}.$$

For real-valued $\mathbf{t} = (t_1, \ldots, t_m)$ and $\mathbf{s} = (s_1, \ldots, s_n)$, let

$$\begin{aligned}\psi(\mathbf{t},\mathbf{s}) &:= \ln \frac{G(\mathbf{x},\mathbf{y})}{\mathbf{x}^R\mathbf{y}^C}\bigg|_{\substack{x_i=e^{t_i}\\ y_j=e^{s_j}}} \\ &= -\sum_{i=1}^m r_i t_i - \sum_{j=1}^n c_j s_j + \sum_{i=1}^m\sum_{j=1}^n \ln\left(1 + e^{t_i+s_j} + e^{2(t_i+s_j)} + \cdots + e^{k_{ij}(t_i+s_j)}\right).\end{aligned}$$

This function is strictly convex, and has a critical point (hence a global minimum) at $(\mathbf{t},\mathbf{s})$ if and only if the gradient is zero, that is, if

$$\begin{aligned} r_i &= \sum_{j=1}^n \frac{e^{t_i+s_j} + 2e^{2(t_i+s_j)} + \cdots + k_{ij}e^{k_{ij}(t_i+s_j)}}{1 + e^{t_i+s_j} + e^{2(t_i+s_j)} + \cdots + e^{k_{ij}(t_i+s_j)}}, \qquad 1 \le i \le m \\ \text{and } c_j &= \sum_{i=1}^m \frac{e^{t_i+s_j} + 2e^{2(t_i+s_j)} + \cdots + k_{ij}e^{k_{ij}(t_i+s_j)}}{1 + e^{t_i+s_j} + e^{2(t_i+s_j)} + \cdots + e^{k_{ij}(t_i+s_j)}}, \qquad 1 \le j \le n. \end{aligned}$$

These conditions are satisfied at $\mathbf{t} = (\lambda_1, \ldots, \lambda_m)$ and $\mathbf{s} = (\mu_1, \ldots, \mu_n)$. The minimum value of ψ is thus

$$\begin{aligned}\psi(\mathbf{t},\mathbf{s}) &= -\sum_{i=1}^m r_i\lambda_i - \sum_{j=1}^n c_j\mu_j + \sum_{i=1}^m\sum_{j=1}^n \ln\left(1 + \xi_i\eta_j + (\xi_i\eta_j)^2 + \cdots + (\xi_i\eta_j)^{k_{ij}}\right) \\ &= \sum_{i=1}^m\sum_{j=1}^n \left[-z_{ij}(\lambda_i+\mu_j) + \ln(1 + q_{ij} + q_{ij}^2 + \cdots + q_{ij}^{k_{ij}})\right] \\ &= \sum_{i=1}^m\sum_{j=1}^n \left[-z_{ij}\ln q_{ij} + \ln\left(\frac{1}{p_{ij}}\right)\right] \\ &= \sum_{i=1}^m\sum_{j=1}^n H_{k_{ij}}^{\max}(z_{ij}).\end{aligned}$$

This proves the lemma. □

Combining Lemma III.18 with Theorem III.16, we have the following second asymptotic estimate for $\ln T_K(R,C)$:

Theorem III.19. *Let $R \in \mathbb{Z}^m_{>0}$, $C \in \mathbb{Z}^n_{>0}$, and $K \in \mathbb{Z}^{m\times n}_{\geq 0}$.*

Assume that $T_K(R,C) > 0$, that is, there is at least one contingency table with margins R and C, bounded entrywise by K. Then

$$\lim_{s\to\infty} \frac{1}{s^2} \ln T_{K^{(s)}}(R^{(s)}, C^{(s)}) \quad = \quad \max_{Z\in\Pi_K(R,C)} \sum_{i=1}^{m}\sum_{j=1}^{n} H^{\max}_{k_{ij}}(z_{ij}).$$

Proof. Immediate corollary of the results just mentioned. ■

Notice that this estimate is efficiently computable, as it is the maximum of a strictly concave function over a convex polytope. See Remark III.17 for a more general setting in which this estimate holds asymptotically.

3.4 Correlation phenomena

In the language we have developed so far, Barvinok's correlation results for row and column margins may be stated as follows:

Theorem III.20 ([8], [9])**.** *Let $R \in \mathbb{Z}^m_{>0}$ and $C \in \mathbb{Z}^n_{>0}$.*

If $T(R,C) > 0$, then

$$\lim_{s\to\infty} \frac{1}{s^2} \ln T(R^{(s)}, C^{(s)}) \quad \geq \quad \lim_{s\to\infty} \frac{1}{s^2} \ln I(R^{(s)}, C^{(s)})$$

(where $I(R,C)$ is Good's independence heuristic; cf. Section 3.1).

If $T_1(R,C) > 0$, then

$$\lim_{s\to\infty} \frac{1}{s^2} \ln T_1(R^{(s)}, C^{(s)}) \quad \leq \quad \lim_{s\to\infty} \frac{1}{s^2} \ln I_1(R^{(s)}, C^{(s)}).$$

Both inequalities are strict if neither R nor C is a constant vector (i.e., if it is not the case that $r_1 = \cdots = r_m$ or $c_1 = \cdots = c_n$).

We will use the entropy-based estimate for $\ln T_K(R,C)$ (Theorem III.19) to prove the following extension:

Theorem III.21. *Let $R \in \mathbb{Z}_{>0}^m$, $C \in \mathbb{Z}_{>0}^n$, and $\kappa \in \{2,3,4,\ldots\}$. Then there exists $\delta = \delta(\kappa) \in (0,1)$, such that if (R,C) satisfy*

$$\left(\max_{1\le i\le m} r_i\right)\left(\max_{1\le j\le n} c_j\right) \quad < \quad \delta\kappa N$$

then

$$\lim_{s\to\infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) \quad \ge \quad \lim_{s\to\infty} \frac{1}{s^2} \ln I_\kappa(R^{(s)}, C^{(s)}),$$

with strict inequality if neither R nor C is a constant vector.

3.4.1 Estimate for the independence heuristic

The following result is the counterpart of Theorem III.19 for the independence heuristic $I_\kappa(R,C)$.

Lemma III.22. *Let $R \in \mathbb{Z}_{>0}^m$, $C \in \mathbb{Z}_{>0}^n$, $N = |R| = |C|$, and $\kappa \in \mathbb{Z}_{>0} \cup \{\infty\}$. Then*

$$\lim_{s\to\infty} \frac{1}{s^2} \ln\, I_\kappa(R^{(s)}, C^{(s)}) \quad = \\ - mnH_\kappa^{\max}\left(\frac{N}{mn}\right) \;+\; n\sum_{i=1}^m H_\kappa^{\max}\left(\frac{r_i}{n}\right) \;+\; m\sum_{j=1}^n H_\kappa^{\max}\left(\frac{c_j}{m}\right).$$

Proof. By (3.2), we have

$$I_\kappa(R^{(s)}, C^{(s)}) = \binom{s^2mn}{s^2N}_\kappa^{-1} \left(\prod_{i=1}^m \binom{sn}{sr_i}_\kappa\right)^s \left(\prod_{j=1}^n \binom{sm}{sc_j}_\kappa\right)^s.$$

Applying Lemma III.3, we obtain

$$\begin{aligned}\ln I_\kappa(R^{(s)},C^{(s)}) \;&=\; -\left[s^2mnH_\kappa^{\max}\left(\frac{N}{mn}\right) + o(s^2)\right] + s\sum_{i=1}^m \left[snH_\kappa^{\max}\left(\frac{r_i}{n}\right) + o(s)\right] \\ &\qquad + s\sum_{j=1}^n \left[smH_\kappa^{\max}\left(\frac{c_j}{m}\right) + o(s)\right] \\ &=\; s^2\left[-mnH_\kappa^{\max}\left(\frac{N}{mn}\right) + n\sum_{i=1}^m H_\kappa^{\max}\left(\frac{r_i}{n}\right) + m\sum_{j=1}^n H_\kappa^{\max}\left(\frac{c_j}{m}\right) + o(1)\right],\end{aligned}$$

proving the lemma. □

3.4.2 A measure of surprise

The following function plays a key role in the proof of Theorem III.21:

Definition III.23. Fix $\kappa \in \mathbb{Z}_{>0} \cup \{\infty\}$. Given nonnegative $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1$, let

$$J(r) \;=\; J_{\alpha,\kappa}(r) \;:=\; nH_\kappa^{\max}\left(\frac{r}{n}\right) - \sum_{j=1}^{n} H_\kappa^{\max}(r\alpha_j)$$

for all $r \geq 0$ such that $r\alpha_1, r\alpha_2, \ldots, r\alpha_n \leq \kappa$.

To interpret this function, we consider four independence models for a random contingency table. Let $X = \left(x_{ij}\right)$, $X^R = \left(x_{ij}^R\right)$, $X^C = \left(x_{ij}^C\right)$, and $X^{R,C} = \left(x_{ij}^{R,C}\right)$ be the $m \times n$ random matrices with independent $TG(\cdot, \kappa)$ entries satisfying the following expectations:

$$\mathbf{E}[x_{ij}] = \frac{N}{mn}, \qquad \mathbf{E}[x_{ij}^R] = \frac{r_i}{n}, \qquad \mathbf{E}[x_{ij}^C] = \frac{c_j}{m}, \qquad \mathbf{E}[x_{ij}^{R,C}] = \frac{r_i c_j}{N}.$$

The first three of these are MEIMs for contingency tables about which we know only the 0-margin, the row margins, and the column margins respectively. The fourth model is generally *not* the maximum-entropy model for a table with margins R and C (discussed in Section 3.3.3). It is, rather, a naïve guess at the MEIM (in the same sense that the rank 1 table[11] is a naïve guess at the "typical" table), which we study despite its flaws because we can actually write it down.[12] Note that in order for $X^{R,C}$ to be well-defined, the rank 1 table $X_{R,C}^{\text{ind}}$ must have all entries $\leq \kappa$, which is not guaranteed to be the case. We will essentially will this problem away by means of the stipulation $\delta < 1$ in the statement of Theorem III.21.

[11] See (3.1).

[12] The MEIM is efficiently computable for individual choices of R and C, but this is not in itself sufficient for the analysis we intend to do.

Now, letting $\alpha_j := c_j/N$ $(1 \le j \le n)$, we have

$$\mathbf{H}[X^R] - \mathbf{H}[X^{R,C}] \;=\; \sum_{i=1}^{m} J(r_i), \tag{3.16}$$

$$\mathbf{H}[X] - \mathbf{H}[X^C] \;=\; mJ\left(\frac{N}{m}\right). \tag{3.17}$$

Assuming that we model an unknown contingency table by the four independence models described above, quantities (3.16) and (3.17) represent the loss of entropy (or "surprise") when we learn the row margins of the table, respectively with or without prior knowledge of the column margins. If less surprise occurs under the former circumstance, that is, if

$$J(r_1) + J(r_2) + \cdots + J(r_m) \le mJ\left(\frac{N}{m}\right), \tag{3.18}$$

then that implies (informally) that R and C are positively correlated. This is the strategy for proving Theorem III.21, in a nutshell.

3.4.3 Proof of Theorem III.21

As in the previous section, let

$$\alpha_j \;:=\; \frac{c_j}{N} \qquad (1 \le j \le n).$$

Consider the function

$$\phi(x) \;:=\; x^2 (H_\kappa^{\max})''(x) \;=\; -x^2 \cdot \frac{q'(x;\kappa)}{q(x;\kappa)}$$

(all derivatives being with respect to x). The second equality here follows from Lemma II.14(iii).

The above formula defines $\phi(x)$ only for $0 \le x \le \kappa$, but we claim that $\phi(x)$ can be extended analytically to a neighborhood of $x = 0$.

Proof of claim: Equations (2.3) and (2.4) yield

$$x = \frac{q + 2q^2 + \cdots + \kappa q^\kappa}{1 + q + q^2 + \cdots + q^\kappa},$$

where $q = q(x;\kappa)$. Although this formula has only been assigned meaning for $q \geq 0$, it shows that x (as a function of q) can be extended analytically to a neighborhood of $q = 0$; the Maclaurin series is $x = q + q^2 + O(q^3)$. Since $\frac{dx}{dq} \neq 0$ at $q = 0$, it follows that the inverse function $q(x;\kappa)$ is also defined and analytic in a neighborhood of $x = 0$, with Maclaurin series $q = x - x^2 + O(x^3)$. Applying l'Hôpital's rule, we see that the singularity of ϕ at $x = 0$ is removable, so $\phi(x)$ is locally analytic there, proving the claim. $\square$

We compute the Maclaurin series of $\phi(x)$:

$$\phi(x) \;=\; -x \cdot \frac{1 - 2x + O(x^2)}{1 - x + O(x^2)} \;=\; -x + x^2 + O(x^3).$$

Since the coefficient of x^2 is positive, $\phi(x)$ is strictly convex in a neighborhood of $x = 0$. Choose $\delta \in (0,1)$ such that $\phi(x)$ is strictly convex in the interval $|x| \leq \delta\kappa$.

Because $\delta < 1$, $J(r)$ is defined and differentiable at $r = r_1, \ldots, r_m$. Differentiating, we have

$$J'(r) \;=\; (H_\kappa^{\max})'\left(\frac{r}{n}\right) - \sum_{j=1}^{n} \alpha_j (H_\kappa^{\max})'(r\alpha_j)$$

and

$$\begin{aligned} J''(r) \;&=\; \frac{1}{n}(H_\kappa^{\max})''\left(\frac{r}{n}\right) - \sum_{j=1}^{n} \alpha_j^2 (H_\kappa^{\max})''(r\alpha_j) \\ &=\; \frac{n}{r^2}\phi\left(\frac{r}{n}\right) - \sum_{j=1}^{n} \frac{1}{r^2}\phi(r\alpha_j). \end{aligned}$$

By the (local) convexity of $\phi(x)$, we have $J''(r) \leq 0$ for $0 < r \leq \frac{\delta\kappa}{\max\{\alpha_1,\ldots,\alpha_n\}}$; the inequality is strict if $\alpha_1, \ldots, \alpha_n$ are not all equal. Therefore, $J(r)$ is concave on (the closure of) that interval, and strictly concave if $\alpha_1, \ldots, \alpha_n$ are not all equal. By our assumption that $r_i c_j \leq \delta\kappa N$, it follows that $r_1, \ldots, r_m$ are in that interval.

Thus, inequality (3.18) holds, and holds strictly if $\alpha_1, \ldots, \alpha_n$ are not all equal and $r_1, \ldots, r_m$ are also not equal. When the function J is evaluated throughout this inequality, we obtain

$$n \sum_{i=1}^{m} H_\kappa^{\max}\left(\frac{r_i}{m}\right) - \sum_{i=1}^{m}\sum_{j=1}^{n} H_\kappa^{\max}\left(\frac{r_i c_j}{N}\right) \;\leq\; mn H_\kappa^{\max}\left(\frac{N}{mn}\right) - m \sum_{j=1}^{n} H_\kappa^{\max}\left(\frac{c_j}{m}\right).$$

Combining this with Theorem III.19 and Lemma III.22, we have

$$\begin{aligned}
\lim_{s\to\infty} \frac{1}{s^2} \ln T_\kappa(R^{(s)}, C^{(s)}) \;&\geq\; \max_{Z \in \Pi_\kappa(R,C)} \sum_{i=1}^{m}\sum_{j=1}^{n} H_\kappa^{\max}(z_{ij}) \\
&\geq\; \sum_{i=1}^{m}\sum_{j=1}^{n} H_\kappa^{\max}\left(\frac{r_i c_j}{N}\right) \\
&\geq\; \lim_{s\to\infty} \frac{1}{s^2} \ln I_\kappa(R^{(s)}, C^{(s)}).
\end{aligned}$$

If $\alpha_1, \ldots, \alpha_n$ are not all equal and $r_1, \ldots, r_m$ are not all equal, then the last inequality in this chain is strict. This completes the proof of Theorem III.21. ■

3.4.4 Negative correlation of margins: evidence and prospects

Recall that for $\kappa = 1$, all pairs of margins (R, C) have either zero or negative asymptotic correlation under cloning (specifically, negative correlation unless either R or C is a constant vector). For $\kappa = \infty$, the sign of correlation is reversed. We expect that these are the only "pure" cases: that is, when $1 < \kappa < \infty$, there are some positively correlated pairs of margins as well as some negatively correlated pairs. Theorem III.21 asserts half of this conjecture: for $\kappa \geq 2$, any sufficiently sparse margins are asymptotically positively correlated. (By symmetry, "co-sparse margins"—those which force most entries to be close to κ—are also positively correlated.)

Numerical evidence and heuristic arguments suggest that, for all $\kappa < \infty$, margins which are neither sparse nor co-sparse—or, more specifically, margins which are close

$(\kappa = 2)$		γ							
		0.12	0.18	0.24	0.30	0.36	0.42	0.48	$\cdots$
	0.06	+	+	+	+	−	−	−	$\cdots$
	0.12		+	+	+	−	−	−	$\cdots$
ε	0.18			+	+	−	−	−	$\cdots$
	0.24				+	−	−	−	$\cdots$
	0.30					−	−	−	$\cdots$

$(\kappa = 4)$		γ										
		0.24	0.36	0.48	0.60	$\cdots$	1.32	1.44	1.56	1.68	1.8	$\cdots$
	0.12	+	+	+	+	$\cdots$	+	−	−	−	−	$\cdots$
	0.24		+	+	+	$\cdots$	+	−	−	−	−	$\cdots$
	0.36			+	+	$\cdots$	+	−	−	−	−	$\cdots$
	0.48				+	$\cdots$	+	+	−	−	−	$\cdots$
	$\vdots$					$\ddots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
ε	1.20						+	+	+	−	−	$\cdots$
	1.32							+	+	−	−	$\cdots$
	1.44								+	−	−	$\cdots$
	1.56									+	−	$\cdots$
	1.68										−	$\cdots$

Table 3.2: Sign of $\lim_{s\to\infty} \frac{1}{s^2}\left[\ln T_\kappa(R^{(s)}, C^{(s)}) - \ln I_\kappa(R^{(s)}, C^{(s)})\right]$ for margins of the form $R = C = (\gamma - \varepsilon, \gamma, \gamma + \varepsilon)$, where $\kappa = 2$ or 4 and γ, ε take various values. Sign corresponds to asymptotic correlation of the (cloned) margins. All omitted entries between +'s are +'s. Were these tables to be continued to the right or downward, all omitted entries would be −'s except for mirror images of the +'s shown.

to $R = (\frac{n\kappa}{2}, \ldots, \frac{n\kappa}{2})$ and $C = (\frac{m\kappa}{2}, \ldots, \frac{m\kappa}{2})$—are negatively correlated. For example, we have used Theorem III.19 to compute

$$\lim_{s\to\infty} \frac{1}{s^2}\left[\ln T_\kappa(R^{(s)}, C^{(s)}) - \ln I_\kappa(R^{(s)}, C^{(s)})\right] \tag{3.19}$$

for margins of the form $R = C = (\gamma-\varepsilon, \gamma, \gamma+\varepsilon)$ and $\kappa = 2, 4, 6, 8, 10$. The results for $\kappa = 2, 4$ are shown in Table 3.2. (Note that γ and ε are not required to be integers; as long as they are rational, the cloned margins will be integral for some values of s.) For every value of the increment ε we tested, we found that the values of γ for which (3.19) is negative form an interval centered at $\gamma = \frac{3\kappa}{2}$. Our computations with small ε allow us to estimate that the largest possible values of δ in Theorem III.21 are $\delta_{\text{cr}} \approx 0.05, 0.11, 0.14, 0.15, 0.16$ when $\kappa = 2, 4, 6, 8, 10$ respectively. Nonconstant margins (R, C) satisfying $\delta_{\text{cr}}\kappa n < r_i < (1 - \delta_{\text{cr}})\kappa n$ and $\delta_{\text{cr}}\kappa m < c_j < (1 - \delta_{\text{cr}})\kappa m$

appear to always exhibit negative correlation.

An intuitive gloss on this phenomenon is that the distribution $TG(x;\kappa)$ "looks like" a geometric distribution when $x \approx 0$ (or $x \approx \kappa$), but looks more like a Bernoulli distribution when x is at neither extreme. In the former case, the "lid" κ (or the floor 0) is remote from typical values, so the behavior observed when $\kappa = \infty$ dominates. In the latter case, the $\kappa = 1$ behavior seems to dominate.

The fundamental difference between these cases is hinted at by the function $\phi(x)$ which appears in the proof of Theorem III.21. When $\kappa = \infty$, this function is convex throughout its domain; when $\kappa = 1$, it is concave; and when $1 < \kappa < \infty$, this function is convex near the origin, but has an inflection point.[13] See Figure 3.1.

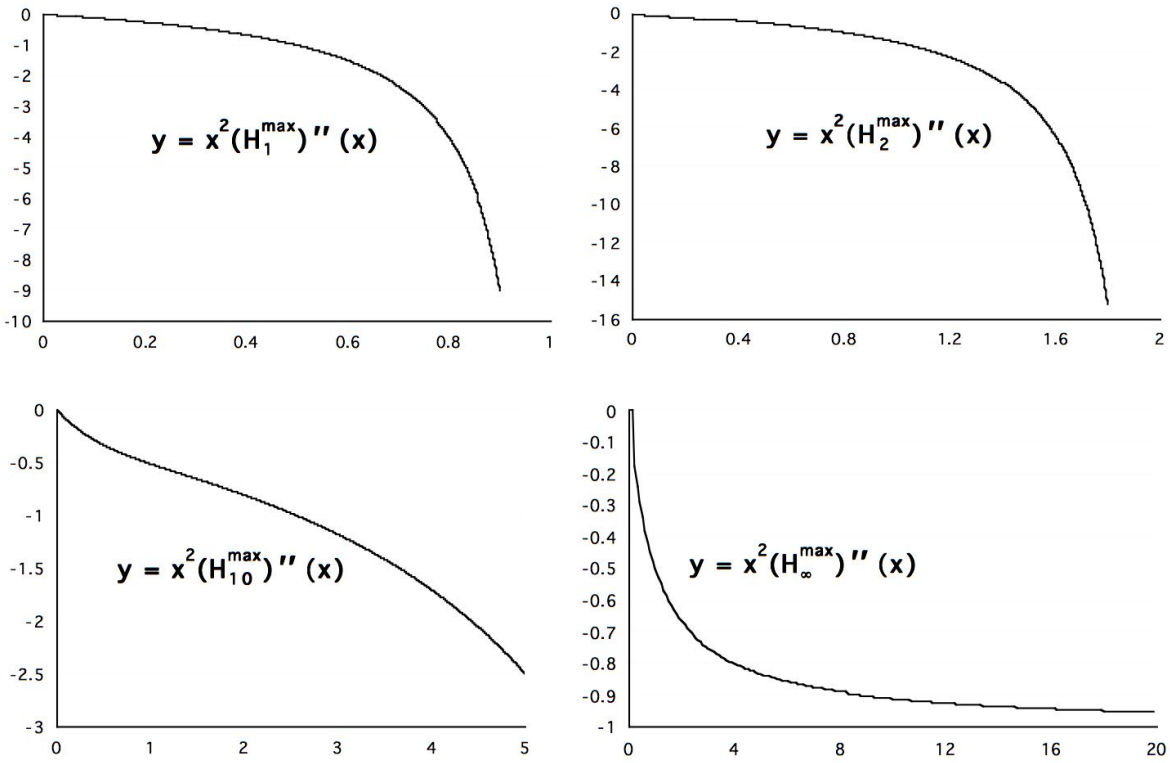

Figure 3.1: Graphs of $\phi(x)$, $\kappa = 1, 2, 10, \infty$

Given $\kappa < \infty$, we can show that $\phi(x)$ is concave for $x \approx \frac{\kappa}{2}$; so what are the obsta-

[13]The subtlety of this matter can be appreciated by recalling that $\phi(x)$ was defined in terms of the second derivative of $H_\kappa^{\max}(x)$; therefore the concavity of $\phi(x)$ is influenced by the *fourth* derivative of $H_\kappa^{\max}(x)$.

cles to a reversed Theorem III.21? There are two. In the proof of Theorem III.21, we relied on the fact that

$$\max_{Z\in\Pi_\kappa(R,C)} \sum_{i=1}^{m}\sum_{j=1}^{n} H_\kappa^{\max}(z_{ij}) \;\geq\; \sum_{i=1}^{m}\sum_{j=1}^{n} H_\kappa^{\max}\left(\frac{r_i c_j}{N}\right), \tag{3.20}$$

a triviality whose opposite (i.e., the inequality with reversed sign) is of course false. It is this triviality which fructified our use of the rank 1 matrix $X^{\mathrm{ind}} = \left(\frac{r_i c_j}{N}\right)$ as a proxy for the unknown Z which achieves the maximum. Even so, in order to give existence to this proxy, we had to assume that X^{ind} has entries $\leq \kappa$; it happened that the hypothesis of sparse margins in Theorem III.21 served a double purpose by underwriting this assumption. Neither of these helps is available toward proving a negative correlation result. To do that, we believe it will be necessary to understand something about *where* the maximum on the left-hand side of (3.20) is achieved.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] R. Aharoni, A. Georgakopoulos, and P. Sprüssel. Perfect matchings in r-partite r-graphs. *European Journal of Combinatorics*, 30:39–42, 2009.

[2] W. Baldoni-Silva, J. A. De Loera, and M. Vergne. Counting integer flows in networks. *Foundations of Computational Mathematics*, 4:277–314, 2004.

[3] K. Ball. An elementary introduction to modern convex geometry. In *MSRI Publications #31: Flavors of Geometry*, pages 1–58. Cambridge University Press, 1997. `http://library.msri.org/books/Book31/files/ball.pdf`.

[4] A. Barvinok. Notes on measure concentration. `http://www.math.lsa.umich.edu/~barvinok/total710.pdf`.

[5] A. Barvinok. A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed. *Mathematics of Operations Research*, 19:769–779, 1994.

[6] A. Barvinok. Brunn–Minkowski inequalities for contingency tables and integer flows. *Advances in Mathematics*, 211:105–122, 2007.

[7] A. Barvinok. Enumerating contingency tables via random permanents. *Combinatorics, Probability, and Computing*, 17:1–19, 2008.

[8] A. Barvinok. Asymptotic estimates for the number of contingency tables, integer flows, and volumes of transportation polytopes. *International Mathematics Research Notices*, 2:348–385, 2009.

[9] A. Barvinok. On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries. *Advances in Mathematics*, 224:316–339, 2010.

[10] A. Barvinok. What does a random contingency table look like? *Combinatorics, Probability, and Computing*, 19:517–539, 2010.

[11] A. Barvinok and J. Hartigan. An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums. *Transactions of the AMS*, 2010. To appear; preprint available at `http://arxiv.org/abs/0910.2477`.

[12] A. Barvinok and J. Hartigan. Maximum entropy Gaussian approximation for the number of integer points and volumes of polytopes. *Advances in Applied Mathematics*, 45:252–289, 2010.

[13] A. Barvinok, Z. Luria, A. Samorodnitsky, and A. Yong. An approximation algorithm for counting contingency tables. *Random Structures and Algorithms*, 37:25–66, 2010.

[14] A. Barvinok, A. Samorodnitsky, and A. Yong. Counting magic squares in quasi-polynomial time. Preprint (2007), available at `http://www.math.uiuc.edu/~ayong/squares.final.ps`.

[15] M. Beck and D. Pixton. The Ehrhart polynomial of the Birkhoff polytope. *Discrete Computational Geometry*, 30:623–637, 2003.

[16] E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A*, 15:91–111, 1973.

[17] P. Billingsley. *Probability and Measure, 2nd ed.* Wiley, New York, 1986.

[18] A. Burchard. A short course on rearrangement inequalities. `http://www.math.utoronto.ca/almut/rearrange.pdf`.

[19] R. Cambini, G. Gallo, and M. G. Scutellà. Flows on hypergraphs. *Mathematical Programming*, 78:195–217, 1997.

[20] E. R. Canfield and B. D. McKay. Asymptotic enumeration of integer matrices with constant row and column sums. *Combinatorica*, 2007. To appear; preprint available at `http://arxiv.org/abs/math/0703600`.

[21] M. Cryan, M. Dyer, and D. Randall. Approximately counting integral flows and cell-bounded contingency tables. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, pages 413–422. ACM, 2005.

[22] J. De Loera. Counting and estimating lattice points: tools from algebra, analysis, convexity, and probability. *Optima: Newsletter of the Mathematical Programming Society*, Dec. 2009. With appendix.

[23] J. De Loera and S. Onn. All linear and integer programs are slim 3-way transportation programs. *SIAM Journal on Optimization*, 17:806–821, 2006.

[24] J. De Loera and S. Onn. Markov bases of three-way tables are arbitrarily complicated. *Journal of Symbolic Computation*, 41:173–181, 2006.

[25] J. H. Dénes and A. D. Keedwell. *Latin Squares: New Developments in the Theory and Applications.* Academic Press, Amsterdam, 1991.

[26] P. Diaconis. The Markov chain Monte Carlo revolution. *Bulletin of the American Mathematical Society*, 46:179–205, 2009.

[27] P. Diaconis and B. Efron. Testing for independence in a two-way table: new interpretations of the chi-square statistic. *Annals of Statistics*, 13:845–874, 1985.

[28] M. Dyer. Approximate counting by dynamic programming. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 693–699. ACM, 2003.

[29] J. Edmonds. Maximum matching and a polyhedron with 0, 1 vertices. *Journal of the National Bureau of Standards, Section B*, 69:125–130, 1965.

[30] E. Ehrhart. Sur un problème de géométrie diophantienne linéaire. II. Systèmes diophantiens linéaires. *Journal für die reine und angewandte Mathematik*, 227:25–49, 1967.

[31] K. Engel. *Sperner Theory.* Cambridge University Press, Cambridge, 1997.

[32] P. Erdős. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51:898–902, 1945.

[33] L. Euler. De evolutione potestatis polynomialis cuiuscunque $(1+x+x^2+x^3+x^4+\text{etc.})^n$. *Nova Acta Academiae Scientarum Imperialis Petropolitinae*, 12:47–57, 1801. Translation available at `http://arxiv.org/abs/math.HO/0505425`.

[34] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[35] M. R. Garey and S. J. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, San Francisco, 1979.

[36] I.J. Good. On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Annals of Statistics*, 4:1159–1189, 1976.

[37] B. Grünbaum. *Convex Polytopes, 2nd ed.* Springer, New York, 2003.

[38] S. Guiasu and A. Shenitzer. The principle of maximum entropy. *The Mathematical Intelligencer*, 7:42–48, 1985.

[39] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8:197–211, 1977.

[40] R. Howard. Estimates on the concentration function of sets in $\mathbb{R}^d$: Notes on lectures of Oskolkov. `http://www.math.sc.edu/~howard/Notes/concentration.pdf`.

[41] M. T. Jacobson and P. Matthews. Generating uniformly distributed random Latin squares. *Journal of Combinatorial Designs*, 4:405–437, 1996.

[42] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

[43] E. T. Jaynes. Information theory and statistical mechanics II. *Physical Review*, 108:171–190, 1957.

[44] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, 51:671–697, July 2004.

[45] M. Jerrum, L. Valiant, and V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[46] P. W. Kasteleyn. The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice. *Physica*, 27:1209–1225, 1961.

[47] A. I. Khinchin. *Mathematical Foundations of Information Theory.* Dover, New York, 1957.

[48] W. Kook. On the product of log-concave polynomials. *INTEGERS: Electronic Journal of Combinatorial Number Theory*, 6, 2006.

[49] E. Lieb and M. Loss. *Analysis, 2nd ed.* AMS, Providence, 2001.

[50] M. Loebl and L. Zdeborová. The 3d dimer and Ising problems revisited. *European Journal of Combinatorics*, 29:966–978, 2008.

[51] L. Lovász and M. D. Plummer. *Matching Theory.* AMS Chelsea, Providence, 2009.

[52] I. G. Macdonald. Polynomials associated with finite cell-complexes. *Journal of the London Mathematical Society*, 4:181–192, 1971.

[53] B. Morris and A. Sinclair. Random walks on truncated cubes and sampling 0-1 knapsack solutions. In *40th Annual Symposium on Foundations of Computer Science, proceedings*, pages 230–240. IEEE, 1999.

[54] T. S. Motzkin. The multi-index transportation problem. *Bulletin of the American Mathematical Society*, 58:494, 1952.

[55] L. Nicolaescu. Lattice points inside rational simplices and the Casson invariant of Brieskorn spheres. *Geometriae Dedicata*, 88:37–53, 2001.

[56] T. E. O'Neil and S. Kerlin. Sub-exponential algorithms for 0/1 knapsack and bin packing. Preprint available at `http://people.aero.und.edu/~oneil/pubs/cocoon11-10pt.pdf`.

[57] G. Pick. Geometrisches zur Zahlentheorie. *Sitzungsberichte des deutschen naturwissenschaftlich-medicinischen Vereines für Böhmen "Lotos"*, 19:311–319, 1899.

[58] J. Propp. Enumeration of matchings: problems and progress. *New Perspectives in Geometric Combinatorics*, 38:255–291, 1999.

[59] H. L. Royden. *Real Analysis*. Prentice-Hall, Englewood Cliffs, 1988.

[60] R. Y. Rubenstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. Springer, New York, 2004.

[61] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[62] C. Shannon. A mathematical theory of communication II. *Bell System Technical Journal*, 27:623–656, 1948.

[63] A. V. Skorokhod. *Basic Principles and Applications of Probability Theory*. Springer-Verlag, Heidelberg, 2005.

[64] R. Snee. Graphical display of two-way contingency tables. *The American Statistician*, 28:9–12, 1974.

[65] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics—an exact result. *Philosophical Magazine*, 6:1061–1063, 1961.

[66] A. N. Timashov. On permanents of random doubly stochastic matrices and on asymptotic estimates for the number of Latin rectangles and Latin squares. *Discrete Mathematics and Applications*, 12:431–452, 2002.

[67] L. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[68] L. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8:410–421, 1979.

[69] J. H. Van Lint and R. M. Wilson. *A Course in Combinatorics, 2nd ed.* Cambridge University Press, 2001.

[70] X. Wang and S. Yau. On the GLY conjecture of upper estimate of positive integral points in real right-angled simplices. *Journal of Number Theory*, 122:184–210, 2007.

[71] G. Ziegler. *Lectures on Polytopes*. Springer-Verlag, New York, 1995.