

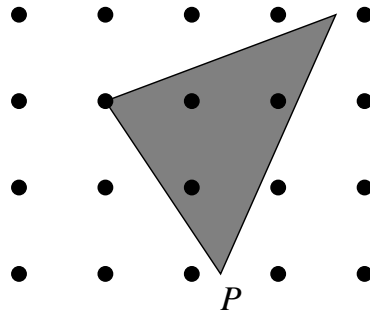
COUNTING INTEGER POINTS IN POLYHEDRA

ALEXANDER BARVINOK

Papers are available at

<http://www.math.lsa.umich.edu/~barvinok/papers.html>

Let $P \subset \mathbb{R}^d$ be a polytope. We want to compute (exactly or approximately) the number $|P \cap \mathbb{Z}^d|$ of integer points in P .



PART I: EXACT COUNTING IN FIXED DIMENSION

RATIONAL POLYHEDRA

$$P = \left\{ (\xi_1, \dots, \xi_d) : \sum_{j=1}^d a_{ij} \xi_j \leq b_i, \quad i = 1, \dots, n \right\},$$

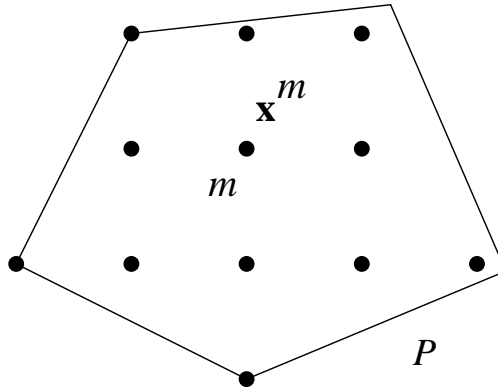
where $a_{ij}, b_i \in \mathbb{Z}$.

The input size of P :

$$\mathcal{L}(P) = n(d+1) + \sum_{ij} \lceil \log_2(|a_{ij}| + 1) \rceil + \sum_i \lceil \log_2(|b_i| + 1) \rceil,$$

the number of bits needed to define P .

GENERATING FUNCTIONS



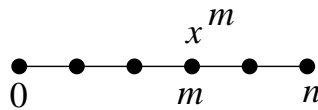
We consider the sum

$$\sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m, \quad \text{where}$$

$$\mathbf{x}^m = x_1^{\mu_1} \cdots x_d^{\mu_d} \quad \text{for } m = (\mu_1, \dots, \mu_d)$$

The motivating example

$$\sum_{m=0}^n x^m = \frac{1 - x^{n+1}}{1 - x}$$



The sum over the integer points on an interval.

Theorem. *Let us fix d . There exists a polynomial time algorithm, which, given a rational polyhedron $P \subset \mathbb{R}^d$ without lines, computes the generating function*

$$f(P, \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$$

in the form

$$f(P, \mathbf{x}) = \sum_{i \in I} \epsilon_i \frac{\mathbf{x}^{v_i}}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $\epsilon_i \in \{-1, 1\}$, $v_i \in \mathbb{Z}^d$, $u_{ij} \in \mathbb{Z}^d \setminus \{0\}$.

The complexity of the algorithm is $\mathcal{L}^{O(d)}$, where $\mathcal{L} = \mathcal{L}(P)$ is the input size of P .

In particular, $|I| = \mathcal{L}^{O(d)}$.

Proved in 1993 (with $\mathcal{L}^{O(d^2)}$ bound) and then again in 1999.

APPLICATIONS

Having computed

$$f(P, \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m,$$

we can

1) efficiently count $|P \cap \mathbb{Z}^d|$, the number of integer points in a given rational polytope P . Carefully substitute $x_1 = \dots = x_d = 1$ into $f(P, \mathbf{x})$;

2) solve *integer programming problems* of optimizing a given linear function on the set $P \cap \mathbb{Z}^d$ of integer points in P .

There are at least two implementations:

`LattE` (lattice point enumerator) by J. De Loera et al., now subsumed by `LattE macchiato` by M. Köppe, and `barvinok` by S. Verdoolaege

<http://www.math.ucdavis.edu/~latte>

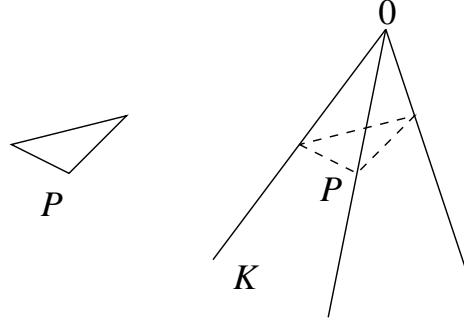
<http://www.kotnet.org/~skimo/barvinok>

<http://freshmeat.net/projects/barvinok/>

SOME IDEAS OF THE PROOF

Reducing polyhedra to cones

We can represent the polyhedron as a hyperplane section of a higher-dimensional cone.



We have $P = K \cap H$, where $P \subset \mathbb{R}^d$, $K \subset \mathbb{R}^{d+1}$, and $H \subset \mathbb{R}^{d+1}$ is an affine hyperplane identified with \mathbb{R}^d .

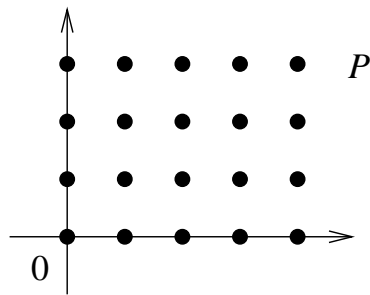
Consequently, we have

$$f(P, \mathbf{x}) = \frac{\partial}{\partial x_{d+1}} f(K, \widehat{\mathbf{x}}) \Big|_{x_{d+1}=0}, \quad \text{where } \widehat{\mathbf{x}} = (\mathbf{x}, x_{d+1}).$$

Dealing with cones: unimodular cones

A cone $K \subset \mathbb{R}^d$ generated by a basis u_1, \dots, u_d of \mathbb{Z}^d is called *unimodular*. For such a cone, we have

$$f(K, \mathbf{x}) = \sum_{m \in K \cap \mathbb{Z}^d} \mathbf{x}^m = \frac{1}{(1 - \mathbf{x}^{u_1}) \cdots (1 - \mathbf{x}^{u_d})}.$$



Integer points in the non-negative orthant.

$$\sum_{m \in \mathbb{Z}_+^d} \mathbf{x}^m = \frac{1}{(1 - x_1) \cdots (1 - x_d)}.$$

A unimodular cone differs from the non-negative orthant by a unimodular transformation.

The main construction (1993):

For any fixed d there is a polynomial time algorithm which decomposes a given rational cone $K \subset \mathbb{R}^d$ into a combination of unimodular cones K_i :

$$[K] = \sum_{i \in I} \epsilon_i [K_i],$$

where $\epsilon_i \in \{-1, 1\}$, K_i are unimodular cones, and

$$[A] : \mathbb{R}^d \longrightarrow \mathbb{R}, \quad [A](x) = \begin{cases} 1 & \text{if } x \in \mathbb{R}^d \\ 0 & \text{otherwise} \end{cases}$$

is the indicator of a set A .

Example: continued fractions in dimension 2

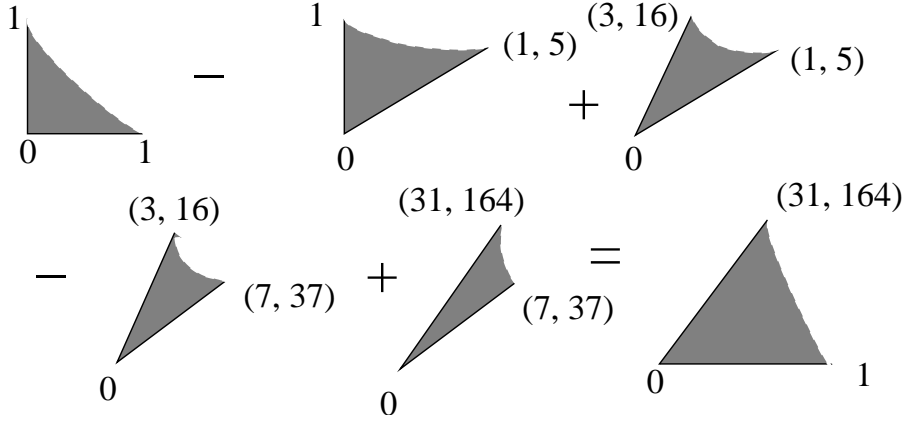
Suppose that $K \subset \mathbb{R}^2$ is generated by $(1, 0)$ and $(31, 164)$.
First, we compute

$$\frac{164}{31} = 5 + \frac{9}{31} = 5 + \frac{1}{3 + \frac{4}{9}} = 5 + \frac{1}{3 + \frac{1}{2 + \frac{1}{4}}}.$$

Hence $164/31 = [5; 3, 2, 4]$. Now we compute the convergents:

$$[5; 3, 2] = 5 + \frac{1}{3 + \frac{1}{2}} = \frac{37}{7}, [5; 3] = 5 + \frac{1}{3} = \frac{16}{3}, \quad [5] = \frac{5}{1}.$$

Then we do cutting and pasting of cones:



Let K_0 be the cone generated by $(1, 0)$ and $(0, 1)$.

Starting with K_0 , we

- cut the cone generated by $(0, 1)$ and $(1, 5)$;
- paste the cone generated by $(1, 5)$ and $(3, 16)$;
- cut the cone generated by $(3, 16)$ and $(7, 37)$;
- paste the cone generated by $(7, 37)$ and $(31, 164)$

to finally get K generated by $(1, 0)$ and $(31, 164)$.

The four cones we cut and paste are unimodular.

So we get

$$\begin{aligned}
 f(K, \mathbf{x}) &= \frac{1}{(1-x_1)(1-x_2)} - \frac{1}{(1-x_2)(1-x_1x_2^5)} \\
 &\quad + \frac{1}{(1-x_1x_2^5)(1-x_1^3x_2^{16})} \\
 &\quad - \frac{1}{(1-x_1^3x_2^{16})(1-x_1^7x_2^{37})} \\
 &\quad + \frac{1}{(1-x_1^7x_2^{37})(1-x_1^{31}x_2^{164})}.
 \end{aligned}$$

Higher dimensions

For K generated by linearly independent $u_1, \dots, u_d \in \mathbb{Z}^d$, let $\text{ind } K$ be the volume of the parallelepiped spanned by u_1, \dots, u_d (so $\text{ind } K = 1$ if and only if K is unimodular).

Using Minkowski's Convex Body Theorem, compute $w \in \mathbb{Z}^d \setminus \{0\}$ such that

$$w = \sum_{i=1}^d \alpha_i u_i \quad \text{where} \quad |\alpha_i| \leq (\text{ind } K)^{-1/d}.$$

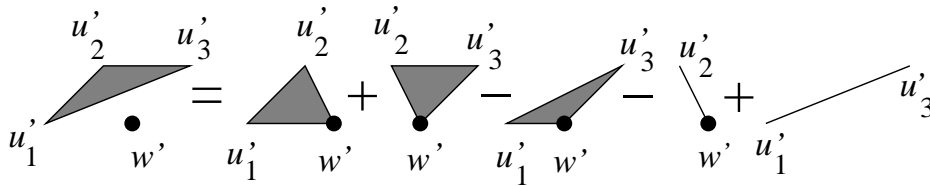
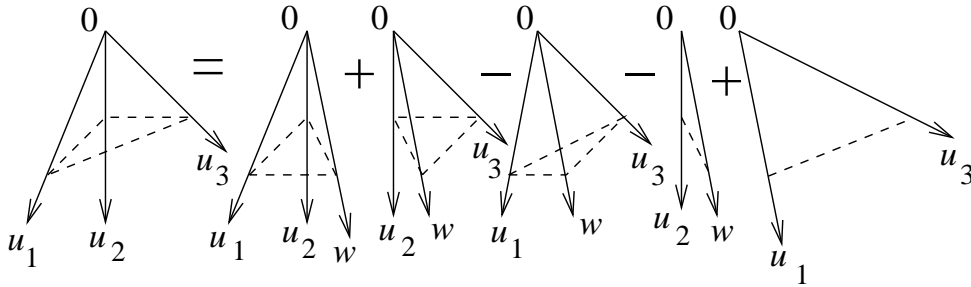
Let K_i be the cone generated by $u_1, \dots, u_{i-1}, w, u_{i+1}, \dots, u_d$. Then

$$[K] = \sum_{i=1}^d \epsilon_i [K_i] \quad + \text{ indicators of lower-dimensional cones,}$$

where $\epsilon_i \in \{-1, 1\}$ and

$$\text{ind } K_i \leq (\text{ind } K)^{\frac{d-1}{d}}.$$

Now iterate.



PART II : APPROXIMATE COUNTING IN HIGHER DIMENSIONS

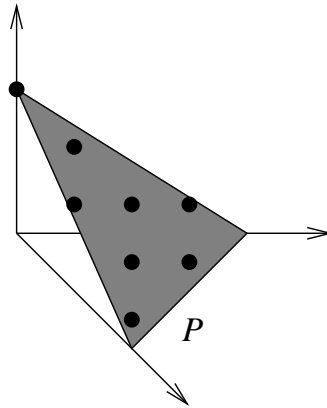
We assume that P is defined by a system of linear equations

$$Ax = b$$

and inequalities

$$x \geq 0,$$

so the picture looks more like this



Here $A = (a_{ij})$ is an integer $k \times n$ matrix of rank $k < n$ and b is an integer n -vector.

$$\begin{array}{c}
 \text{integer entries} \\
 k \left[\begin{array}{cccc}
 * & * & * & * \\
 * & * & * & * \\
 * & * & * & *
 \end{array} \right] A \\
 n
 \end{array}$$

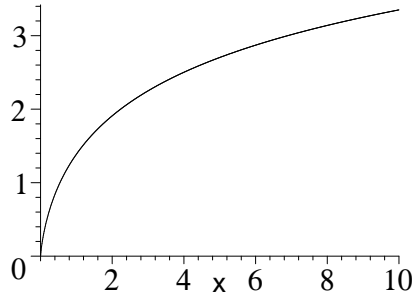
Let $\Lambda = Ax : x \in \mathbb{Z}^n$ be the lattice in \mathbb{Z}^k . Unless $b \in \Lambda$, we have $P \cap \mathbb{Z}^n = \emptyset$.

THE GAUSSIAN FORMULA

Joint work with J.A. Hartigan (Yale).

Let us consider the function

$$g(\xi) = (\xi + 1) \ln(\xi + 1) - \xi \ln \xi \quad \text{for } \xi \geq 0.$$



Let us solve the optimization problem:

$$\begin{aligned} \text{Find} \quad & \max g(x) = \sum_{j=1}^n g(\xi_j) \\ \text{Subject to:} \quad & x = (\xi_1, \dots, \xi_n) \in P. \end{aligned}$$

Since g is strictly concave, the maximum point

$$z = (\zeta_1, \dots, \zeta_n)$$

is unique and can be found efficiently by interior point methods, for example. Computing z is easy both in theory and in practice.

Let us compute a $k \times k$ matrix $Q = (q_{ij})$ by

$$q_{ij} = \sum_{m=1}^n a_{im} a_{jm} (\zeta_m^2 + \zeta_m).$$

We approximate the number of integer points in P by

$$|P \cap \mathbb{Z}^n| \approx \frac{e^{g(z)} \det \Lambda}{(2\pi)^{k/2} (\det Q)^{1/2}}.$$

A COUPLE OF EXAMPLES

Let us compute the number of 4×4 non-negative integer matrices with row sums 220, 215, 93 and 64 and column sums 108, 286, 71 and 127.

	108	286	71	127
220	*	*	*	*
215	*	*	*	*
93	*	*	*	*
64	*	*	*	*

The number of such matrices is

$$1225914276768514 \approx 1.23 \times 10^{15}.$$

We have $n = 16$ variables and $k = 7$ equations (one equation can be thrown out).

The Gaussian formula overestimates by about 6%.

J. De Loera computed more examples. Here is one of them: the exact number of $3 \times 3 \times 3$ arrays of non-negative integers with the sums

$$[31, 22, 87],$$

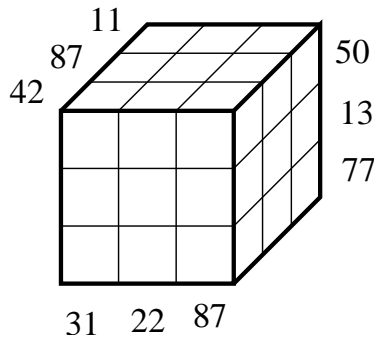
$$[50, 13, 77],$$

$$[42, 87, 11]$$

along the affine coordinate hyperplanes is

$$8846838772161591 \approx 8.84 \times 10^{15}.$$

The Gaussian formula gives the relative error of 0.185%.



Here we have $n = 27$ variables and $k = 7$ equations (two equations can be thrown out).

THE (FIRST LEVEL OF) INTUITION

A random variable x is *geometric* if

$$\Pr \{x = m\} = pq^m \quad \text{for } m = 0, 1, \dots$$

where $p + q = 1$ and $p, q > 0$. We have

$$\mathbf{E} x = \frac{q}{p} \quad \text{and} \quad \mathbf{var} x = \frac{q}{p^2}.$$

Conversely,

$$\text{if } \mathbf{E} x = z \quad \text{then} \quad p = \frac{1}{1+z}, \quad q = \frac{z}{1+z} \quad \text{and} \quad \mathbf{var} x = z^2 + z.$$

Theorem. Let $P \subset \mathbb{R}^n$ be a polytope that is the intersection of an affine subspace in \mathbb{R}^n and the non-negative orthant \mathbb{R}_+^n . Suppose that P has a non-empty interior (contains a point with strictly positive coordinates).

Then the strictly concave function

$$g(x) = \sum_{j=1}^n \left((\xi_j + 1) \ln (\xi_j + 1) - \xi_j \ln \xi_j \right)$$

attains its maximum on P at a unique point $z = (\zeta_1, \dots, \zeta_n)$ with positive coordinates.

Suppose that x_1, \dots, x_n are independent geometric random variables with expectations ζ_1, \dots, ζ_n and let $X = (x_1, \dots, x_n)$. Then the probability mass function of X is constant on $P \cap \mathbb{Z}^n$ and equal to $e^{-g(z)}$ at every $x \in P \cap \mathbb{Z}^n$. In particular,

$$|P \cap \mathbb{Z}^n| = e^{g(z)} \Pr \{X \in P\}.$$

Now, suppose that

$$P = \left\{ x : Ax = b, \quad x \geq 0 \right\}.$$

Let $Y = AX$, that is, $Y = (y_1, \dots, y_k)$, where

$$y_i = \sum_{j=1}^n a_{ij} x_j.$$

Theorem implies that

$$|P \cap \mathbb{Z}^n| = e^{g(z)} \mathbf{Pr} \{Y = b\}.$$

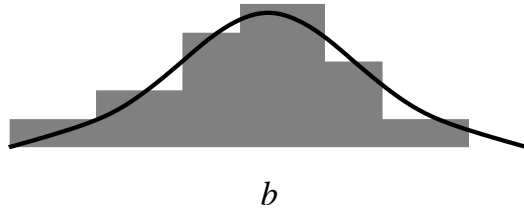
We note that

$$\mathbf{E}Y = b$$

and that

$$\mathbf{cov}(y_i, y_j) = \sum_{m=1}^n a_{im} a_{jm} \mathbf{var} x_k = \sum_{m=1}^n a_{im} a_{jm} (\zeta_m^2 + \zeta_m).$$

Now, we observe that Y is the sum of n independent random vectors $x_j A_j$, where A_j is the j -th column of A , so we make a leap of faith and assume that the distribution of Y in the vicinity of its expectation is close to the distribution of a Gaussian random vector Y^* with the same expectation b and the covariance matrix Q .



Hence it is not unreasonable to assume that

$$\begin{aligned} \mathbf{Pr} \{Y = b\} &\approx \mathbf{Pr} \{Y^* \in b + \text{fundamental domain of } \Lambda\} \\ &\approx \frac{\det \Lambda}{(2\pi)^{k/2} (\det Q)^{1/2}}. \end{aligned}$$

In 1957, E.T. Jaynes formulated a general principle. Let Ω be a large but finite probability space with an unknown measure μ , let $f_1, \dots, f_k : \Omega \rightarrow \mathbb{R}$ be random variables with known expectations

$$\mathbf{E} f_i = \alpha_i \quad \text{for } i = 1, \dots, k$$

and let $g : \Omega \rightarrow \mathbb{R}$ be yet another random variable. Then to compute or estimate $\mathbf{E} g$ one should assume that μ is the probability measure on Ω of the largest entropy such that that $\mathbf{E} f_i = \alpha_i$ for $i = 1, \dots, k$.

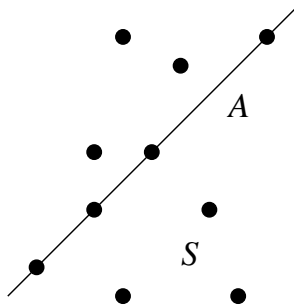
Works great for the Maxwell-Boltzmann distribution.

In 1963, I.J. Good argued that the “null hypothesis” concerning an unknown probability distribution from a given class should be the one stating that the distribution is the maximum entropy distribution in the class.

In our case, Ω is the set \mathbb{Z}_+^n of non-negative integer vectors, f_i are the linear equations defining polytope P , and μ is the counting probability measure on $P \cap \mathbb{Z}_+^n$. We approximate μ by the maximum entropy distribution on \mathbb{Z}_+^n subject to the constraints $\mathbf{E} f_i = \alpha_i$, where f_i are the linear equations defining P .

Fact: Among all distributions on \mathbb{Z}_+ with a given expectation, the geometric distribution has the maximum entropy. The entropy of a geometric distribution with expectation x is

$$g(x) = (x + 1) \ln(x + 1) - x \ln x.$$



Let us choose a maximum entropy distribution among all probability distributions on a given set $S \subset \mathbb{R}^n$ with the expectation in a given affine subspace A . It is not hard to argue that the conditional distribution on $S \cap A$ must be uniform.

MULTI-WAY TRANSPORTATION POLYTOPES
AND MULTI-WAY CONTINGENCY TABLES

Transportation polytopes. Let us choose positive r_1, \dots, r_m and c_1, \dots, c_n such that

$$r_1 + \dots + r_m = c_1 + \dots + c_n = N.$$

The polytope P of non-negative $m \times n$ matrices (x_{ij}) with row sums r_1, \dots, r_m and column sums c_1, \dots, c_n is called a (two-index) *transportation polytope*. We have

$$\dim P = (m - 1)(n - 1).$$

Suppose r_i and c_j are integer. Integer points in P are called (two-way) *contingency tables*.

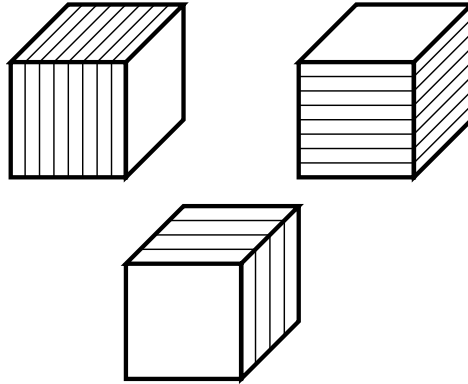
ν -way transportation polytopes. Let us fix an integer $\nu \geq 2$. The polytope P of ν -dimensional

$$k_1 \times \dots \times k_\nu$$

arrays $(x_{j_1 \dots j_\nu})$ with prescribed sectional sums

$$\sum_{\substack{1 \leq j_1 \leq k_1 \\ \dots \\ 1 \leq j_{i-1} \leq k_{i-1} \\ 1 \leq j_{i+1} \leq k_{i+1} \\ \dots \\ 1 \leq j_\nu \leq k_\nu}} x_{j_1 \dots j_{i-1}, j, j_{i+1} \dots j_\nu}$$

are called ν -way transportation polytopes.



As long as the natural balance conditions are met, P is a polytope with

$$\dim P = k_1 \cdots k_\nu - (k_1 + \dots + k_\nu) + \nu - 1.$$

Integer points in P are called ν -way contingency tables.

SOME RESULTS

Fix ν and let k_1, \dots, k_ν grow roughly proportionately.

We proved:

- The number of integer points in P is asymptotically Gaussian provided $\nu \geq 5$. We suspect it is Gaussian already for $\nu \geq 3$.

- In particular, the number of non-negative integer $k \times \dots \times k$ *magic cubes*, that is, contingency tables with all sectional sums equal to $r = \alpha k^{\nu-1}$ is

$$\left(1 + o(1)\right) \left((\alpha + 1)^{\alpha+1} \alpha^{-\alpha}\right)^{k^\nu} \left(2\pi\alpha^2 + 2\pi\alpha\right)^{-(k\nu - \nu + 1)/2} k^{(\nu - \nu^2)(k-1)/2}$$

as $k \rightarrow +\infty$,

provided $\nu \geq 5$, r is integer and α is separated away from 0;

- For $\nu = 2$, the asymptotic of the number of integer contingency tables with equal row sums and equal column sums was recently computed by E.R. Canfield and B. McKay. It differs from the Gaussian estimate by a constant factor (generally, greater than 1). This corresponds to the *Edgeworth correction* to the Gaussian distribution.

THE CURIOUS CASE OF $\nu = 2$

The Gaussian formula should be multiplied by the correction factor, computed as follows.

Let $Z = (z_{ij})$ be the matrix maximizing

$$g(x) = \sum_{ij} \left((x_{ij} + 1) \ln(x_{ij} + 1) - x_{ij} \ln x_{ij} \right)$$

on the polytope of non-negative $m \times n$ matrices with row sums (r_1, \dots, r_m) and column sums (c_1, \dots, c_n) .

Let us consider the quadratic form $q : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$:

$$q(s, t) = \frac{1}{2} \sum_{ij} (z_{ij}^2 + z_{ij}) (s_i + t_j)^2 \quad \text{for } (s, t) = (s_1, \dots, s_m; t_1, \dots, t_n).$$

Let

$$u = \left(\underbrace{1, \dots, 1}_{m \text{ times}}; \underbrace{-1, \dots, -1}_{n \text{ times}} \right)$$

and let $L \subset \mathbb{R}^{m+n}$ be any hyperplane not containing u . Then the restriction of q onto L is strictly positive definite and we consider the Gaussian probability measure on L with the density proportional to e^{-q} .

Let us define random variables $f, h : L \rightarrow \mathbb{R}$ by

$$f(s, t) = \frac{1}{6} \sum_{ij} z_{ij} (z_{ij} + 1) (2z_{ij} + 1) (s_i + t_j)^3$$

and

$$h(s, t) = \frac{1}{24} \sum_{ij} z_{ij} (z_{ij} + 1) (z_{ij}^2 + 6z_{ij} + 1) (s_i + t_j)^4$$

for $(s, t) = (s_1, \dots, s_m; t_1, \dots, t_n)$.

Let

$$\mu = \mathbf{E} f^2 \quad \text{and} \quad \nu = \mathbf{E} h.$$

Then the correction factor is

$$\exp \left\{ -\frac{\mu}{2} + \nu \right\}.$$

RAMIFICATIONS: COUNTING 0-1 POINTS IN POLYTOPES

A similar approach works for the set $P \cap \{0, 1\}^n$ of points with 0-1 coordinates in $P \subset \mathbb{R}^n$, where P is defined by the system

$$P = \left\{ x : Ax = b, 0 \leq x \leq 1 \right\}.$$

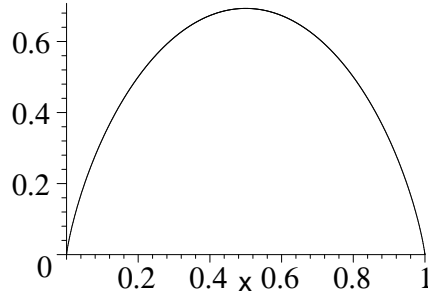
Here $A = (a_{ij})$ is a $k \times n$ integer matrix of rank $k < n$ and $b \in \mathbb{Z}^k$. Let $\Lambda = A(\mathbb{Z}^n) \subset \mathbb{Z}^k$.

The geometric distribution is replaced by the Bernoulli distribution and function

$$g(\xi) = (\xi + 1) \ln(\xi + 1) - \xi \ln \xi$$

is replaced by the standard entropy function

$$h(x) = \xi \ln \frac{1}{\xi} + (1 - \xi) \ln \frac{1}{1 - \xi} \quad \text{for } 0 \leq \xi \leq 1.$$



We solve a convex optimization problem:

$$\begin{aligned} \text{Find} \quad & \max h(x) = \sum_{j=1}^n h(\xi_j) \\ \text{Subject to:} \quad & x = (\xi_1, \dots, \xi_n) \in P. \end{aligned}$$

Since h is strictly concave, the maximum point

$$z = (\zeta_1, \dots, \zeta_n)$$

is unique and can be found efficiently by interior point methods, for example.

We compute the $k \times d$ matrix $Q = (q_{ij})$ by

$$q_{ij} = \sum_{m=1}^n a_{im} a_{jm} (\zeta_m - \zeta_m^2).$$

The Gaussian formula:

$$|P \cap \{0, 1\}^n| \approx \frac{e^{h(z)} \det \Lambda}{(2\pi)^{k/2} (\det Q)^{1/2}}.$$