ISRAEL JOURNAL OF MATHEMATICS **158** (2007), 159–191
DOI: 10.1007/s11856-007-0008-8

RANDOM WEIGHTING, ASYMPTOTIC COUNTING, AND INVERSE ISOPERIMETRY

BY

Alexander Barvinok*

Department of Mathematics, University of Michigan
Ann Arbor, MI 48109-1043, USA
e-mail: barvinok@umich.edu

AND

Alex Samorodnitsky**

Department of Computer Science, The Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: salex@cs.huji.ac.il

ABSTRACT

For a family X of k-subsets of the set $\{1,\dots,n\}$, let $|X|$ be the cardinality of X and let $\Gamma(X,\mu)$ be the expected maximum weight of a subset from X when the weights of $1,\dots,n$ are chosen independently at random from a symmetric probability distribution μ on $\mathbb{R}$. We consider the inverse isoperimetric problem of finding μ for which $\Gamma(X,\mu)$ gives the best estimate of $\ln|X|$. We prove that the optimal choice of μ is the logistic distribution, in which case $\Gamma(X,\mu)$ provides an asymptotically tight estimate of $\ln|X|$ as $k^{-1}\ln|X|$ grows. Since in many important cases $\Gamma(X,\mu)$ can be easily computed, we obtain computationally efficient approximation algorithms for a variety of counting problems. Given μ, we describe families X of a given cardinality with the minimum value of $\Gamma(X,\mu)$, thus extending and sharpening various isoperimetric inequalities in the Boolean cube.

* The research of the first author was partially supported by NSF Grants DMS 9734138 and DMS 0400617.
** The research of the second author was partially supported by ISF Grant 039-7165 and by GIF grant I-2052.
Received May 24, 2005

1. Introduction

Let X be a family of k-subsets of the set $\{1,\ldots,n\}$. Geometrically, we think of X as a set of points $x=(\xi_1,\ldots,\xi_n)$ in the *Hamming sphere* of radius k

$$\xi_1+\cdots+\xi_n=k \quad \text{where } \xi_i \in \{0,1\} \text{ for } i=1,\ldots,n.$$

We also consider general families X of subsets of $\{1,\ldots,n\}$, which we view as sets $X \subset \{0,1\}^n$ of points in the Boolean cube.

Let us fix a Borel probability measure μ in $\mathbb{R}$. We require μ to be symmetric, that is, $\mu(A)=\mu(-A)$ for any Borel set $A \subset \mathbb{R}$, and to have finite variance.

In this paper, we relate two quantities associated with X. The first quantity is the cardinality $|X|$ of X. The second quantity $\Gamma(X,\mu)$ is defined as follows. Let us fix a measure μ as above and let $\gamma_1,\ldots,\gamma_n$ be independent random variables having the distribution μ. Then

$$\Gamma(X,\mu)=\mathbf{E}\max_{x\in X}\sum_{i\in x}\gamma_i.$$

In words: we sample weights of $1,\ldots,n$ independently at random from the distribution μ, define the weight of a subset $x \in X$ as the sum of the weights of its elements and let $\Gamma(X,\mu)$ be the expected maximum weight of a subset from X.

Often, when the choice of μ is clear from the context or not important, we write simply $\Gamma(X)$.

It is easy to see that $\Gamma(X)$ is well defined, that $\Gamma(X)=0$ if X consists of a single point (recall that μ is symmetric) and that $\Gamma(X)\geq\Gamma(Y)$ provided $Y\subset X$. In some respects, $\Gamma(X)$ behaves rather like $\ln|X|$. For example, if $X\subset\{0,1\}^n$ and $Y\subset\{0,1\}^m$, we can define the direct product $X\times Y\subset\{0,1\}^{m+n}$. In this case, $|X\times Y|=|X|\cdot|Y|$ and $\Gamma(X\times Y)=\Gamma(X)+\Gamma(Y)$. Thus, in a sense, $\Gamma(X)$ measures how large X is.

One of our goals is to solve the following *inverse isoperimetric problem* (the choice of the name should become clear by the end of this section):

(1.1) Problem. Find a measure μ for which $\Gamma(X,\mu)$ gives the best estimate of $\ln|X|$ over all families X.

Our motivation comes from problems of efficient combinatorial counting. For many interesting families X, given a set $\gamma_1,\ldots,\gamma_n$ of weights, we can easily find the maximum weight of a subset $x\in X$ using well-known optimization algorithms. The value of $\Gamma(X,\mu)$ can be efficiently computed through averaging of several sample maxima for randomly chosen weights $\gamma_1,\ldots,\gamma_n$. At the same time, counting elements in X can be a hard and interesting problem. Thus, for such families, $\Gamma(X, \mu)$ provides a quick estimate for $\ln |X|$. We give some examples in Section 2. As is discussed in Section 2.7, it follows from our results that the problems of optimization (computing $\Gamma(X, \mu)$) and counting (computing $\ln |X|$) are asymptotically equivalent.

(1.2) THE LOGISTIC MEASURE. It turns out that in some well-defined sense to be made precise later, the optimal choice of μ in Problem 1.1 is the *logistic measure* $\mu = \mu_0$ with density

$$\frac{1}{e^{\gamma} + e^{-\gamma} + 2} \quad \text{for } \gamma \in \mathbb{R}.$$

In this case, for any non-empty family X of k-subsets of $\{1, \ldots, n\}$, the value of $\Gamma(X, \mu_0)$ provides an asymptotically tight estimate for $\ln |X|$ provided $\ln |X|$ grows faster than a linear function of k. Namely, we prove that for any $\alpha > 1$ there exists $\beta = \beta(\alpha) > 0$ such that

$$\beta\Gamma(X, \mu_0) \leq \ln |X| \leq \Gamma(X, \mu_0) \quad \text{provided } |X| \geq \alpha^k$$

and

$$\beta(\alpha) \longrightarrow 1 \quad \text{as } \alpha \longrightarrow +\infty.$$

Moreover, we prove that for $t = k^{-1}\Gamma(X, \mu_0)$ we have

$$t - \ln t - 1 \leq k^{-1} \ln |X| \leq t$$

for all sufficiently large t. Note that the bounds do not depend on n at all.

Geometrically, if we fix the cardinality $|X|$ of a set X in the Hamming sphere of radius k in the n-dimensional Boolean cube, we expect $\Gamma(X)$ to be large if X is "random" and small if X is tightly packed. It turns out that as $|X|$ grows with respect to k though not necessarily with respect to n, the difference between dense and sparse sets in the Hamming sphere disappears as long as the functional $\Gamma(X, \mu_0)$ is concerned. There are some other probability measures μ which share this property with the logistic measure μ_0. In Sections 4 and 5 we prove some general asymptotically tight inequalities relating $\Gamma(X, \mu)$ and $\ln |X|$, from which it follows, for example, that if μ is the measure with density $|\gamma|e^{-|\gamma|}/2$ then $\Gamma(X, \mu)$ and $\ln |X|$ are asymptotically equivalent, whereas if μ is the Gaussian or Bernoulli measure then there is no asymptotic equivalence.

We prove that the logistic distribution μ_0 is, in a well-defined sense, optimal: in the class of all distributions μ for which $\Gamma(X, \mu)$ provides an *upper* bound for

$\ln|X|$, given a *lower* bound for $\Gamma(X,\mu)$, we get the best lower bound for $\ln|X|$ when $\mu=\mu_0$, cf. Theorem 3.3.

In addition, we prove that the logistic distribution has an interesting extremal property: the inequality $\ln|X| \leq \Gamma(X,\mu_0)$ which holds for all non-empty subsets $X \subset \{0,1\}^n$ turns into equality if X is a face (subcube) of the Boolean cube $\{0,1\}^n$.

We state our results in Section 3.

The problems we are dealing with have obvious connections to some central questions in probability and combinatorics, such as discrete isoperimetric inequalities (cf. [ABS98], [Le91], and [T95]) and estimates of the supremum of a stochastic process, see [T94]. In particular, in [T94], M. Talagrand considers the functional $\Gamma(X,\mu_1)$, where X is a family of subsets of the set $\{1,\ldots,n\}$ and μ_1 is the symmetric exponential distribution with density $e^{-|\gamma|}/2$. He proves that $\ln|X| \leq c\Gamma(X,\mu_1)$ for some absolute constant c, see also [La97]. As another application of our method, in Section 7 we prove that the optimal value of this constant is $c = 2\ln 2$ (the equality is obtained when X is a face of the Boolean cube $\{0,1\}^n$). We also prove that $\ln|X| \leq \Gamma(X,\mu_1) + k\ln 2$ provided X lies in the Hamming ball of radius k (the inequality is asymptotically sharp).

(1.3) Isoperimetric inequalities. Suppose that μ is the Bernoulli measure:

$$\mu\{1\} = \mu\{-1\} = \frac{1}{2}.$$

This case was studied in our paper [BS01]. It turns out that $\Gamma(X)$ has a simple geometric interpretation: the value of $0.5n - \Gamma(X)$ is the average Hamming distance from a point x in the Boolean cube $\{0,1\}^n$ to the subset $X \subset \{0,1\}^n$. The classical isoperimetric inequality in the Boolean cube, Harper's Theorem (see [Le91]), implies that among all sets X of a given cardinality, the smallest value of $\Gamma(X)$ is attained when X is the sphere in the Hamming metric. More precisely, let us fix $0<\alpha<\ln 2$. Then there exists $\beta=\beta(\alpha)$, $0<\beta<1/2$, such that if Y_n is the Hamming sphere of radius $\beta n + o(n)$ in $\{0,1\}^n$ then we have $\ln|Y_n| = \alpha n + o(n)$ and for any set $X_n \subset \{0,1\}^n$ with $\ln|X_n| = \alpha n + o(n)$, we have $\Gamma(Y_n) \leq \Gamma(X_n) + o(n)$. We determine β from the equation

$$\beta \ln\frac{1}{\beta} + (1-\beta)\ln\frac{1}{1-\beta} = \alpha$$

and note that $\Gamma(Y_n) = \beta n + o(n)$.

In Section 8, we construct sets Y_n with asymptotically the smallest value of $\Gamma(Y_n)$ for an *arbitrary* symmetric probability measure μ with finite variance.

It is no longer true that Y_n is a Hamming sphere in $\{0,1\}^n$. For example, if $\mu\{1\} = \mu\{-1\} = \mu\{0\} = 1/3$ then Y_n has to be the direct product of two Hamming spheres. It turns out that for any symmetric μ with finite variance Y_n can be chosen to be the direct product of at most two Hamming spheres. More precisely, let us fix a symmetric probability measure μ and a number $0 < \alpha < \ln 2$. Then we construct numbers $\lambda_i = \lambda_i(\alpha,\mu) \geq 0$ and $0 \leq \beta_i = \beta_i(\alpha,\mu) \leq 1/2$ for $i = 1, 2$, such that $\lambda_1 + \lambda_2 = 1$ and the following holds: if Y_n is the direct product of the Hamming sphere of radius $\beta_1 n + o(n)$ in the Boolean cube of dimensions $\lambda_1 n + o(n)$ and the Hamming sphere of radius $\beta_2 n + o(n)$ in the Boolean cube of dimension $\lambda_2 n + o(n)$ (so that Y_n is a subset of the Boolean cube of dimension n) then $\ln|Y_n| = \alpha n + o(n)$ and $\Gamma(Y_n) \leq \Gamma(X_n) + o(n)$ for any set $X_n \subset \{0,1\}^n$ such that $\ln|X_n| = \alpha n + o(n)$.

(1.4) THE INVERSE ISOPERIMETRIC PROBLEM. In view of the discussion above, we can consider Problem 1.1 as the problem of finding a measure with prescribed isoperimetric properties. The relation between $\Gamma(X,\mu)$ and $\ln|X|$ is, in fact, quite sensitive to the choice of the measure μ. Let us consider the class of all non-empty families X of k-subsets of $\{1,\ldots,n\}$ for all k and n. For the inequality $\ln|X| \leq c\Gamma(X,\mu)$ to hold in this class with some constant $c = c(\mu)$, the measure μ has to have a tail that is at least exponential, see Remark 4.7.3. For the lower bound of $k^{-1}\ln|X|$ to be positive given a positive lower bound of $k^{-1}\Gamma(X,\mu)$, the measure μ has to have a tail that is at most exponential, see Remark 5.3.2. Thus for the solution of the inverse isoperimetric Problem 1.1, we are interested in measures with exponential tails.

2. Applications to combinatorial counting

This research is a continuation of [B97] and [BS01], where the idea to use optimization algorithms for counting problems was developed.

First, we discuss how to compute $\Gamma(X)$ for many interesting families of subsets.

Let us assume that the family X of subsets of $\{1,\ldots,n\}$ is given by its *Optimization Oracle*.

(2.1) OPTIMIZATION ORACLE.
Input: Real vector $c = (\gamma_1,\ldots,\gamma_n)$.
Output: Real number

$$w(X,c) = \max_{x \in X} \sum_{i \in x} \gamma_i.$$

Thus, we input real weights of the elements $1, \dots, n$ and output the maximum weight $w(X, c)$ of a subset $x \in X$ in this weighting. As is discussed in [B97] and [BS01], for many interesting families X Optimization Oracle 2.1 can be easily constructed. We provide two examples below.

(2.2) BASES IN MATROIDS. Let A be a $k \times n$ matrix of rank k over a field $\mathbb{F}$. We assume that $k < n$. Let $X = X(A)$ be the set of all k-subsets x of $\{1, \dots, n\}$ such that the columns of A indexed by the elements of x are linearly independent. Thus X is the set of all non-zero $k \times k$ minors of A, or, in other words, the set of *bases* of the matroid represented by A. It is an interesting and apparently hard problem to compute or to approximate the cardinality of X, cf. [JS97].

On the other hand, it is very easy to construct the Optimization Oracle for X. Indeed, given real weights $\gamma_1, \dots, \gamma_n$, we construct a linearly independent set $a_{i_1}, \dots, a_{i_k}$ of columns of the largest total weight one-by-one. First, we choose a_{i_1} to be a non-zero column of A with the largest possible weight γ_{i_1}. Then we choose a_{i_2} to be a column of the maximum possible weight such that a_{i_1} and a_{i_2} are linearly independent. We proceed as above, and finally select a_{i_k} to be a column of the maximum possible weight such that $a_{i_1}, \dots, a_{i_k}$ are linearly independent; cf., for example, Chapter 12 of [PS98] for "greedy algorithms". Particular cases of this problem include counting forests and spanning subgraphs in a given graph.

Let A and B be $k \times n$ matrices of rank $k < n$ and let X be the set of all k-subsets x of $\{1, \dots, n\}$ such that the columns of A indexed by the elements of x are linearly independent and the columns of B indexed by the elements of x are linearly independent. Then there exists a much more complicated than above but still polynomial time algorithm, which, given weights $\gamma_1, \dots, \gamma_n$, computes the largest weight of a subset x from X, see Chapter 12 of [PS98].

(2.3) PERFECT MATCHINGS IN GRAPHS. Let G be a graph with $2k$ vertices and n edges. A collection of k pairwise disjoint edges in G is called a *perfect matching* (known to physicists as a *dimer cover*). It is a hard and interesting problem to count perfect matchings in a given graph, see [JS97]. Recently, using the Markov chain approach, M. Jerrum, A. Sinclair and E. Vigoda constructed a polynomial time approximation algorithm to count perfect matchings in a given bipartite graph [JSV04], but for general graphs no such algorithms are known.

There is a classical $O(n^3)$ algorithm for finding a perfect matching of the maximum weight in any given edge-weighted graph, see Section 11.3 of [PS98], so Oracle 2.1 is readily available.

For any set X given by its Optimization Oracle 2.1, the value of $\Gamma(X)$ can be well approximated by the sample mean of a moderate size.

(2.4) Algorithm for computing $\Gamma(X, \mu)$.

Input: A family X of subsets of $\{1, \ldots, n\}$ given by its Optimization Oracle 2.1;

Output: A number w approximating $\Gamma(X, \mu)$;

Algorithm: Choose a positive integer m (see Section 2.5 for details). Sample independently m random vectors c_i from the product measure $\mu^{\otimes n}$ on $\mathbb{R}^n$. For each vector c_i, using Optimization Oracle 2.1, compute the maximum weight $w(X, c_i)$ of a subset from X. Output

$$w = \frac{1}{m} \sum_{i=1}^{m} w(X, c_i).$$

(2.5) Choosing the number of samples m. Let us consider the output

$$w = w(X; c_1, \ldots, c_m)$$

of Algorithm 2.4 as a random variable on the space

$$\mathbb{R}^{nm} = \underbrace{\mathbb{R}^n \oplus \cdots \oplus \mathbb{R}^n}_{m \text{ times}}$$

endowed with the product measure $\mu^{\otimes mn}$. Clearly, the expectation of w is $\Gamma(X, \mu)$.

Let $D = \mathbf{E}(\gamma^2)$ be the variance of μ. Using the estimates

$$\left(\sum_{i \in x} \gamma_i \right)^2 \leq \left(\sum_{i=1}^{n} |\gamma_i| \right)^2 \leq n \sum_{i=1}^{n} \gamma_i^2 \quad \text{for } x \subset \{1, \ldots, n\},$$

we conclude that the variance of w does not exceed $n^2 D/m$. Therefore, by Chebyshev's inequality, for the output w to satisfy $|w - \Gamma(X, \mu)| \leq \epsilon$ with probability at least 2/3, we can choose $m = \lceil 3\epsilon^{-2} n^2 D \rceil$.

As usual, to achieve a higher probability $1 - \delta$ of success, we can run the algorithm $O(\ln \delta^{-1})$ times and then find the median of the computed estimates.

For many measures μ the bound for m can be essentially improved. In particular, we are interested in the case of the logistic measure μ_0 with density $(2 + e^{\gamma} + e^{-\gamma})^{-1}$. In this case, to estimate $\Gamma(X, \mu_0)$ within error ϵ it suffices to choose $m = O(k\epsilon^{-2})$, where k is the size of every set from X. In particular, the number m of samples is independent of the size n of the ground set. To obtain the estimate we use a concentration property of the symmetric exponential measure μ_1 with density $e^{-|\gamma|}/2$, see Section 4.5 of [Led01].

Let us define

$$\psi(\gamma) = \begin{cases} \gamma - \ln(2 - e^{\gamma}) & \text{if } \gamma \leq 0 \\ \gamma + \ln(2 - e^{-\gamma}) & \text{if } \gamma > 0 \end{cases}$$

and

$$\Psi(c) = (\psi(\gamma_1), \ldots, \psi(\gamma_n)) \quad \text{for } c = (\gamma_1, \ldots, \gamma_n).$$

Then $\psi(\gamma)$ has the logistic distribution μ_0 if γ has the exponential distribution μ_1. Thus we can write

$$\Gamma(X, \mu_0) = \mathbf{E} w(X; \Psi(c_1), \ldots, \Psi(c_m)),$$

where vectors $(c_1, \ldots, c_m)$ are sampled from the exponential distribution $\mu_1^{\otimes mn}$ in $\mathbb{R}^{nm}$. If X is a family of k-subsets then the Lipschitz coefficient of

$$f(c_1, \ldots, c_m) = w(X; \Psi(c_1), \ldots, \Psi(c_m))$$

with respect to the ℓ^2 metric of $\mathbb{R}^{nm}$ does not exceed $2\sqrt{k/m}$ while the Lipschitz coefficient with respect to the ℓ^1 metric does not exceed $2/m$. Applying Proposition 4.18 of [Led01], we conclude that for the output w of Algorithm 2.4 to satisfy $|w - \Gamma(X, \mu_0)| \leq \epsilon$ with probability at least 2/3, we can choose $m = O(k\epsilon^{-2})$.

We observe that it is easy to sample a random weight γ from the logistic distribution provided sampling from the uniform distribution on the interval $[0, 1]$ is available (which is the case for many computer packages). Indeed, if ξ is uniformly distributed on the interval $[0, 1]$, then $\gamma = \ln \xi - \ln(1 - \xi)$ has the logistic distribution.

(2.6) Counting with multiplicities. Suppose that every element i of the ground set $\{1, \ldots, n\}$ has a positive integer *multiplicity* q_i. Let X be a family of k-subsets of $\{1, \ldots, n\}$ and let

$$p_X(q_1, \ldots, q_n) = \sum_{x \in X} \prod_{i \in x} q_i.$$

It may be of interest to compute or approximate p_X.

For instance, let $A = (a_{ij})$ be a $2k \times 2k$ symmetric matrix of non-negative integers a_{ij}. Let us construct an (undirected) graph G on $2k$ vertices $\{1, \ldots, 2k\}$ where the vertices i and j are connected by an edge if and only if $a_{ij} > 0$. We identify the edges of G with the set $\{1, \ldots, n\}$. Let X be the set of all perfect matchings in G identified with a family of k-subsets of $\{1, \dots, n\}$, see Example 2.3. If we assign multiplicities a_{ij} to the edges of G, then the value of $p_X(a_{ij})$ is called the *hafnian* $\operatorname{haf} A$ of A, a polynomial of considerable interest which generalizes permanent.

Computing $p_X(q_1, \dots, q_n)$ is reduced to counting in the following straightforward way. Let $N = q_1 + \dots + q_n$ and let us view the set $\{1, \dots, N\}$ as the multiset consisting of q_1 copies of 1, q_2 copies of 2, …, q_n copies of n. Let us construct a family Y of k-subsets of $\{1, \dots, N\}$ as follows: for each k-subset $x \in X$ we construct $\prod_{i \in x} q_i$ k-subsets $y \in Y$ by replacing every $i \in x$ by any of its q_i copies. It is clear that $|Y| = p_X(q_1, \dots, q_n)$.

To construct Optimization Oracle 2.1 for Y, we apply the oracle for X with the input $c = (\gamma_1, \dots, \gamma_n)$, where γ_i is the maximum of q_i weights assigned to the q_i copies of i. Moreover, Algorithm 2.4 is easily modified for computing $\Gamma(Y, \mu)$ instead of $\Gamma(X, \mu)$. We still work with the underlying family X, but instead of sampling weights from the distribution μ, we sample the i-th weight γ_i from the distribution μ_{q_i} of the maximum of q_i independent random variables with the distribution μ (note that μ_{q_i} is not symmetric for $q_i > 1$). Thus, if μ is the logistic distribution, to sample γ_i, we sample ξ from the uniform distribution on $[0, 1]$ and let $\gamma_i = -\ln(\xi^{-1/q_i} - 1)$. Luckily, for the logistic distribution the required number m of calls to Oracle 2.1 does not depend on the size of the ground set, hence we use the same number m of calls whether we consider counting with or without multiplicities.

In [BS01] we discuss how our approach fits within the general framework of the Monte Carlo method. The estimates we get are not nearly as precise as those obtained by the Markov chain based Monte Carlo Method (see, for example, [JS97]), but supply non-trivial information and are easily computed for a wide variety of problems for which Optimization Oracle 2.1 is available. Even for the much-studied problem of counting perfect matchings in general (non-bipartite) graphs our approach produces new theoretical results. For some of the problems, such as counting bases in the intersection of two general matroids (see Example 2.2), our estimates seem to be the only ones that can be efficiently computed at the moment. If X is a family of k-subsets of $\{1, \dots, n\}$ and $|X| = e^{k\lambda}$ for some $\lambda = \lambda(X)$ then, in polynomial time, we estimate $\lambda(X)$ within a constant multiplicative factor as long as $\lambda(X)$ is separated from 0 and all sufficiently large $\lambda(X)$ are estimated with an additive error of $1 + \ln \lambda(X)$, see Section 3. Similar estimates hold for counting with multiplicities of Section 2.6. On the other hand, the Markov chain approach, if successful, allows one to estimate the cardinality $|X|$ within any prescribed relative error. We note that for truly large problems the correct scale is logarithmic because $|X|$ can be prohibitively large to deal with. The Markov chain approach relies on the local structure of X (it requires "high connectivity" of X needed for "rapid mixing"), whereas our method uses some global structure of X (it requires the ability to efficiently optimize on X).

(2.7) Asymptotic equivalence of counting and optimization. One can view the optimization functional $\max_{x \in X} \sum_{i \in x} \gamma_i$ as the "tropical version" (cf. [M04]) of the polynomial $p_X(q_1, \ldots, q_n)$ of Section 2.6: we get the former if we replace "+" with "max" and product with sum in the latter. Thus our results establish a weak asymptotic equivalence of the counting and optimization problems: if we can optimize, we can estimate $\ln p_X$ with a relative error which approaches 0 as $k^{-1} \ln p_X$ grows. Vice versa, if we can approximate $\ln p_X$, we can optimize (at least approximately): choosing $q_i(t) = 2^{t\gamma_i}$, we get

$$\lim_{t \longrightarrow +\infty} t^{-1} \log_2 p_X(q_1(t), \ldots, q_n(t)) = \max_{x \in X} \sum_{i \in x} \gamma_i.$$

A. Yong [Y03] implemented our algorithms for some counting problems, such as estimating the number of forests in a given graph, computing the permanent and the hafnian of a given non-negative integer matrix and performed a number of numerical experiments. The algorithm produces the upper and lower bounds for the logarithm of the cardinality of the family in question, see Section 3. The upper bound is attained when the family is "tightly packed" as a subset of the Boolean cube whereas the lower bound is attained on sparse families. It appears that there is some metric structure inherent to various families of combinatorially defined sets. For example, when we applied our methods to estimate the logarithm of the number of spanning trees in a given connected graph, the exact value (which can be easily computed by the matrix-tree formula) turns out to be very close to the upper bound obtained by our algorithm. Informally, spanning trees appear to be "tightly packed". On the other hand, when we estimated the logarithm of the number of perfect matchings in a graph, the true value (when we were able to find it by other methods) seems to lie close to the middle point between the upper and lower bounds. We also observed that the number m of samples of random weights we have to choose to get a good approximation of $\Gamma(X)$ is much smaller than the theoretical bound of Section 2.5 (in many cases just one sample sufficed).

3. The logistic measure: Results

Let us choose μ_0 with density

$$\frac{1}{e^\gamma + e^{-\gamma} + 2} \quad \text{for } \gamma \in \mathbb{R}.$$

The cumulative distribution function F of μ_0 is given by

$$F(\gamma) = \frac{1}{1 + e^{-\gamma}} \quad \text{for } \gamma \in \mathbb{R}.$$

The variance of μ_0 is $\pi^2/3$ [M85]. Our first main result is as follows.

(3.1) Theorem:

(1) *For every non-empty set* $X \subset \{0,1\}^n$, *we have*

$$\ln |X| \leq \Gamma(X).$$

(2) *Let*

$$h(t) = \sup_{0 \leq \delta < 1} \left(\delta t + \ln \frac{\sin \pi\delta}{\pi\delta} \right) \quad \textit{for } t \geq 0.$$

Then $h(t)$ *is a convex increasing function and for any non-empty family* X *of* k*-subsets of* $\{1, \ldots, n\}$, *we have*

$$h(t) \leq k^{-1} \ln |X| \quad \textit{where } t = k^{-1}\Gamma(X).$$

From the expansion

$$\ln \frac{\sin \pi x}{\pi x} = -\frac{\pi^2}{6} x^2 + O(x^4) \quad \text{for } x \approx 0,$$

we deduce that

$$h(t) = \frac{3}{2\pi^2} t^2 + O(t^4) \quad \text{for } t \approx 0$$

(we substitute $\delta = (3t/\pi^2)$). From the expansion

$$\ln \frac{\sin \pi(1-x)}{\pi(1-x)} = \ln x + x + O(x^2) \quad \text{for } x \approx 0,$$

we deduce that

$$h(t) \geq t - \ln t - 1 \quad \text{as } t \longrightarrow +\infty$$

(we substitute $\delta = 1 - t^{-1}$).

A Maple plot of $h(t)$ is shown on Figure 1.

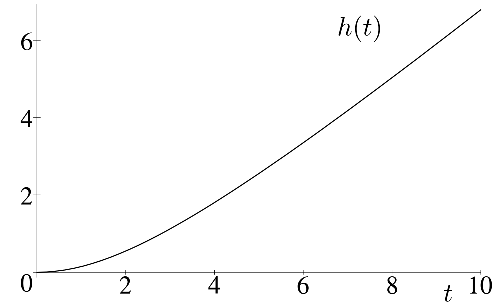

Figure 1.

We obtain the following corollary.

(3.2) Corollary: *For any $\alpha > 1$ there exists $\beta = \beta(\alpha) > 0$ such that for any non-empty family X of k-subsets of $\{1, \ldots, n\}$ with $|X| \geq \alpha^k$ we have*

$$\beta\Gamma(X) \leq \ln|X| \leq \Gamma(X).$$

Moreover,

$$\beta(\alpha) \longrightarrow 1 \quad \textit{as } \alpha \longrightarrow +\infty.$$

Proof: From Part (1) of Theorem 3.1, we have $k^{-1}\Gamma(X) \geq \ln\alpha$. Since $h(t)$ is convex, we have $h(t) \geq \beta t$ for some $\beta = \beta(\alpha) > 0$ and all $t \geq \ln\alpha$. The asymptotics of $\beta(\alpha)$ as $\alpha \longrightarrow +\infty$ follows from the asymptotics of $h(t)$ as $t \longrightarrow +\infty$. ∎

Thus, using the logistic distribution allows us to estimate $\ln|X|$ within a constant factor and the approximation factor approaches 1 as $k^{-1}\ln|X|$ grows.

We note that the bound $\ln|X| \leq \Gamma(X)$ is sharp. For example, if X is an m-dimensional face of the Boolean cube then $\ln|X| = m\ln 2$ and one can show that $\Gamma(X) = m\ln 2$ as well. Indeed, because $\Gamma(X)$ is invariant under coordinate permutations, we may assume that X consists of the points $(\xi_1, \ldots, \xi_m, 0, \ldots, 0)$, where $\xi_i \in \{0,1\}$ for $i = 1, \ldots, m$. The set X can be written as the Minkowski sum $X = X_1 + \cdots + X_m$, where X_i consists of the origin and the i-th basis vector e_i. Hence $\Gamma(X) = m\Gamma(X_1)$ (cf. Section 4.1) and $\Gamma(X_1)$ is computed directly as

$$\Gamma(X_1) = \int_0^{+\infty} \frac{x}{e^x + e^{-x} + 2} dx = \ln 2$$

(we substitute $e^x = y$ and then integrate by parts).

It turns out that the logistic measure is optimal in a well-defined sense.

(3.3) THEOREM: *Let $\mathcal{M}$ be the set of all measures μ such that*

$$\ln|X| \leq \Gamma(X, \mu)$$

for any non-empty family X of k-subsets of $\{1, \dots, n\}$, any $n \geq 1$, and any $1 \leq k \leq n$.

For a measure $\mu \in \mathcal{M}$ and a number $t > 0$, let $c(t, \mu)$ be the infimum of $k^{-1} \ln|X|$ taken over all $n \geq 1$, all $1 \leq k \leq n$, and all non-empty families X of k-subsets $\{1, \dots, n\}$ such that $k^{-1}\Gamma(X, \mu) \geq t$. Then for all $t > 0$

$$c(t, \mu) \leq c(t, \mu_0),$$

where μ_0 is the logistic distribution.

(3.4) DISCUSSION. Unless μ is concentrated in 0, for $X = \{0, 1\}$ we have $\Gamma(X, \mu) = c \ln 2$ for some $c > 0$ and hence $\Gamma(X, \mu) = c \ln|X|$ if X is a face of the Boolean cube $\{0, 1\}^n$, cf. Section 4.1. As we are looking for the best measure μ in Problem 1.1, it is only natural to assume that $\Gamma(X, \mu) \geq c_1 \ln|X|$ for all $X \subset \{0, 1\}^n$, which, after scaling, becomes $\Gamma(X, \mu) \geq \ln|X|$. This explains the definition of $\mathcal{M}$.

Let us choose $\mu \in \mathcal{M}$. Then any upper bound for $\Gamma(X, \mu)$ is automatically an upper bound for $\ln|X|$. The function $c(t, \mu)$ measures the quality of the lower bound estimate for $\ln|X|$ given a lower bound for $\Gamma(X, \mu)$.

Incidentally, it follows from our proof that the logistic measure is the measure of the smallest variance in $\mathcal{M}$.

We prove Theorems 3.1 and 3.3 in Section 6.

4. General estimates: The upper bound

It is convenient to think about families X geometrically, as subsets of the Boolean cube $\{0, 1\}^n \subset \mathbb{R}^n$. Let us fix a symmetric probability measure μ on $\mathbb{R}$ with finite variance and let $\mu^{\otimes n}$ be the product measure on $\mathbb{R}^n$. For a finite set $X \subset \mathbb{R}^n$ we write

$$\Gamma(X) = \mathbf{E} \max_{x \in X} \langle c, x \rangle,$$

where $c = (\gamma_1, \dots, \gamma_n)$ is a random vector sampled from the distribution $\mu^{\otimes n}$ on $\mathbb{R}^n$ and $\langle \cdot, \cdot \rangle$ is the standard scalar product in $\mathbb{R}^n$.

(4.1) Preliminaries. It is easy to check that

$$\Gamma(X) \geq \Gamma(Y) \quad \text{provided } Y \subset X$$

and that

$$\Gamma(X) = 0 \quad \text{if } |X| = 1,$$

that is, if X is a point (μ is symmetric). It follows that $\Gamma(X) \geq 0$ for any finite non-empty subset $X \subset \mathbb{R}^n$. Moreover,

$$\Gamma(X+Y) = \Gamma(X) + \Gamma(Y) \quad \text{where} \quad X+Y = \{x+y : x \in X,\ y \in Y\}$$

is the Minkowski sum of X and Y. In particular, $\Gamma(X+y) = \Gamma(X)$ for any set X and any point y. We note that

$$\Gamma(\lambda X) = |\lambda| \Gamma(X) \quad \text{where } \lambda X = \{\lambda x :\ x \in X\}$$

is a dilation of X and that $\Gamma(X)$ is invariant under the action of the hyperoctahedral group, which permutes and changes signs of the coordinates.

Let $S(k,n)$ be the Hamming sphere of radius k centered at the origin, that is, the set of points $x = (\xi_1, \ldots, \xi_n) \in \{0,1\}^n$ such that $\xi_1 + \cdots + \xi_n = k$. Combinatorially, $S(k,n)$ is the family of all k-subsets of $\{1, \ldots, n\}$.

The main result of this section is Theorem 4.2 below. For a parameter $\tau > 0$, we define the "coupling" $G(X,\tau) = \ln|X| - \tau\Gamma(X)$ of the two main quantities we consider. Inspired by Talagrand's method [T95], we prove an upper bound for $G(X,\tau)$ by induction on the dimension. A computation shows then that this bound is asymptotically sharp on Hamming spheres. In the subsequent sections, we will tune up the parameter τ to obtain the best upper bound of $\ln|X|$ in terms of $\Gamma(X)$.

(4.2) Theorem: *Let F be the cumulative distribution function of μ. For a non-empty set $X \subset \{0,1\}^n$ and a number $\tau > 0$, let*

$$G(X,\tau) = \ln|X| - \tau\Gamma(X).$$

Let

$$g_\tau(a) = \ln(1 + e^{-\tau a}) - \tau \int_a^{+\infty} (1 - F(t))dt \quad \text{for } a \in \mathbb{R}.$$

(1) *For any non-empty set $X \subset \{0,1\}^n$, we have*

$$G(X,\tau) \leq n \sup_{a \geq 0} g_\tau(a).$$

(2) *Suppose that*

$$\sup_{a \geq 0} g_\tau(a) = g_\tau(a_0) > 0 \quad \textit{for some, necessarily finite, } a_0 \geq 0.$$

Then there exists a sequence $X_n = S(k_n, n) \subset \{0,1\}^n$ of Hamming spheres such that

$$\lim_{n \longrightarrow +\infty} \frac{G(X_n, \tau)}{n} = g_\tau(a_0).$$

Assuming that F is continuous and strictly increasing, we can choose $k_n = \alpha n + o(n)$ for $\alpha = 1 - F(a_0)$.

Before we embark on the proof of Theorem 4.2, we summarize some useful properties of $g_\tau(a)$.

(4.3) PROPERTIES OF g_τ. We observe that

$$g_\tau(0) = \ln 2 - \tau \int_0^{+\infty} (1 - F(t))dt = \ln 2 - \tau \int_0^{+\infty} tdF(t).$$

Furthermore,

$$\lim_{a \longrightarrow +\infty} g_\tau(a) = 0,$$

since μ has expectation. If $F(t)$ is continuous then g_τ is differentiable and

$$g'_\tau(a) = \tau(\frac{e^{\tau a}}{1 + e^{\tau a}} - F(a)).$$

In particular, a is a critical point of $g_\tau(a)$ if and only if a is a solution of the equation

$$\frac{e^{\tau a}}{1 + e^{\tau a}} = F(a)$$

or, in other words, if

$$a\tau = \ln F(a) - \ln(1 - F(a)).$$

In particular, $a = 0$ is always a critical point of g_τ.

We prove Part (1) of Theorem 4.2 by induction on n.

(4.4) LEMMA: *Suppose that the cumulative distribution function F is continuous. For a non-empty set $X \subset \{0,1\}^n$, $n > 1$, let*

$$X_1 = \{x \in \{0,1\}^{n-1} : (x,1) \in X\} \quad \textit{and} \quad X_0 = \{x \in \{0,1\}^{n-1} : (x,0) \in X\}.$$

Then, for any $a \in \mathbb{R}$, we have

$$\Gamma(X) \geq (1 - F(a))\Gamma(X_1) + F(a)\Gamma(X_0) + \int_a^{+\infty} tdF(t).$$

Proof: Let $c = (\overline{c}, \gamma)$, where $\overline{c} \in \mathbb{R}^{n-1}$, $\gamma \in \mathbb{R}$, and let

$$w(X, c) = \max_{x \in X} \langle x, c \rangle \quad \text{for } c \in \mathbb{R}^n.$$

Clearly,

$$w(X, c) \geq w(X_1, \overline{c}) + \gamma \quad \text{and} \quad w(X, c) \geq w(X_0, \overline{c}).$$

Therefore,

$$\begin{aligned}
\Gamma(X) =& \int_{\mathbb{R}^n} w(X, c) d\mu^{\otimes n}(c) \\
=& \int_{\mathbb{R}^n : \gamma > a} w(X, c) d\mu^{\otimes n}(c) + \int_{\mathbb{R}^n : \gamma \leq a} w(X, c) d\mu^{\otimes n}(c) \\
\geq& \int_{\mathbb{R}^n : \gamma > a} (w(X_1, \overline{c}) + \gamma) d\mu^{\otimes n}(c) + \int_{\mathbb{R}^n : \gamma \leq a} w(X_0, \overline{c}) d\mu^{\otimes n}(c) \\
=& (1 - F(a)) \int_{\mathbb{R}^{n-1}} w(X_1, \overline{c}) d\mu^{\otimes n-1}(\overline{c}) + \int_a^{+\infty} \gamma dF(\gamma) \\
&+ F(a) \int_{\mathbb{R}^{n-1}} w(X_0, \overline{c}) d\mu^{\otimes n-1}(\overline{c}) \\
=& (1 - F(a))\Gamma(X_1) + F(a)\Gamma(X_0) + \int_a^{+\infty} \gamma dF(\gamma),
\end{aligned}$$

and the proof follows. ∎

(4.5) Lemma: *Suppose that the cumulative distribution function F is continuous. For a non-empty set $X \subset \{0,1\}^n$ and a number $\tau > 0$ let $G(X, \tau)$ and $g_\tau(a)$ be defined as in Theorem 4.2. Then for any non-empty set $X \subset \{0,1\}^n$, $n > 1$, there exists a non-empty set $Y \subset \{0,1\}^{n-1}$ such that*

$$G(X, \tau) \leq G(Y, \tau) + \sup_{a \geq 0} g_\tau(a).$$

Proof: Let us construct X_1 and X_0 as in Lemma 4.4. We have

$$|X_1| = \lambda |X| \quad \text{and} \quad |X_0| = (1 - \lambda)|X| \quad \text{for some } 0 \leq \lambda \leq 1.$$

Without loss of generality, we assume that $0 \leq \lambda \leq 1/2$. Otherwise, we replace X by X', where

$$X' = \{(\xi_1, \ldots, 1 - \xi_n) : \quad (\xi_1, \ldots, \xi_n) \in X\}.$$

Clearly, $|X| = |X'|$ and by Section 4.1, $\Gamma(X) = \Gamma(X')$.

If $\lambda = 0$ we choose $Y = X_0$. Identifying $\mathbb{R}^{n-1}$ with the hyperplane $\xi_n = 0$ in $\mathbb{R}^n$, we observe that $X = Y$ and so $G(X,\tau) = G(Y,\tau)$. Since by Section 4.3 we have

$$\sup_{a \geq 0} g_\tau(a) \geq 0,$$

the result follows.

Thus we assume that $0 < \lambda \leq 1/2$. Let $Y \in \{X_0, X_1\}$ be the set with the larger value of $G(\cdot,\tau)$, where the ties are broken arbitrarily. We have

$$|X| = \frac{1}{\lambda}|X_1| \quad \text{and} \quad |X| = \frac{1}{1-\lambda}|X_0|.$$

For any $a \geq 0$

$$\begin{aligned} G(X,\tau) =& \ln|X| - \tau\Gamma(X) = (1-F(a))\ln|X| + F(a)\ln|X| - \tau\Gamma(X) \\ =&(1-F(a))\ln|X_1| + F(a)\ln|X_0| \\ &+ ((1-F(a))\ln\frac{1}{\lambda} + F(a)\ln\frac{1}{1-\lambda} - \tau\Gamma(X). \end{aligned}$$

By Lemma 4.4 we conclude that

$$\begin{aligned} G(X,\tau) \leq&(1-F(a))\ln|X_1| + F(a)\ln|X_0| + ((1-F(a))\ln\frac{1}{\lambda} + F(a)\ln\frac{1}{1-\lambda} \\ &- (1-F(a))\tau\Gamma(X_1) - F(a)\tau\Gamma(X_0) - \tau\int_a^{+\infty} tdF(t) \\ =&(1-F(a))G(X_1,\tau) + F(a)G(X_0,\tau) \\ &+ (1-F(a))\ln\frac{1}{\lambda} + F(a)\ln\frac{1}{1-\lambda} - \tau\int_a^{+\infty} tdF(t) \\ \leq& G(Y,\tau) + (1-F(a))\ln\frac{1}{\lambda} + F(a)\ln\frac{1}{1-\lambda} - \tau\int_a^{+\infty} tdF(t). \end{aligned}$$

Optimizing in a, we choose

$$a = \frac{1}{\tau}\ln(\frac{1-\lambda}{\lambda}), \quad \text{so } a \geq 0.$$

Then

$$\frac{1}{\lambda} = 1 + e^{\tau a} \quad \text{and} \quad \frac{1}{1-\lambda} = \frac{1+e^{\tau a}}{e^{\tau a}}.$$

Hence

$$\begin{aligned} G(X,\tau) \leq & G(Y,\tau) + \ln(1+e^{\tau a}) - \tau a F(a) - \tau \int_a^{+\infty} t dF(t) \\ = & G(Y,\tau) + \ln(1+e^{-\tau a}) + \tau a(1-F(a)) + \tau \int_a^{+\infty} t d(1-F(t)) \\ = & G(Y,\tau) + \ln(1+e^{-\tau a}) - \tau \int_a^{+\infty} (1-F(t))\, dt \\ = & G(Y,\tau) + g_\tau(a), \end{aligned}$$

as claimed. ∎

Now we are ready to prove Part (1) of Theorem 4.2.

Proof of Part (1) of Theorem 4.2: Without loss of generality, we may assume that the cumulative distribution function F is continuous. The proof follows by induction on n. For $n = 1$, there are two possibilities. If $|X| = 1$ then $G(X,\tau) = 0$ (see Section 4.1) and the result holds since

$$\sup_{a \geq 0} g_\tau(a) \geq 0,$$

see Section 4.3. If $|X| = 2$ then $X = \{0,1\}$ and

$$G(X,\tau) = \ln 2 - \tau \int_0^{+\infty} t dF(t) = g_\tau(0),$$

so the inequality holds as well.

The induction step follows by Lemma 4.5. ∎

Let $S(k,n)$ be the Hamming sphere of radius k, that is, the set of all k-subsets of $\{1,\ldots,n\}$. Given weights $\gamma_1,\ldots,\gamma_n$, the maximum weight of a subset $x \in S(k,n)$ is the sum of the first k largest weights among $\gamma_1,\ldots,\gamma_n$.

The proof of Part (2) of Theorem 4.2 is based on the following lemma.

(4.6) Lemma: *Suppose that the cumulative distribution function F of μ is strictly increasing and continuous. Let us choose $0 < \alpha < 1$ and let X_n be the Hamming sphere of radius $\alpha n + o(n)$ in $\{0,1\}^n$.*

Then

$$\lim_{n \longrightarrow +\infty} \frac{\Gamma(X_n)}{n} = \int_{F^{-1}(1-\alpha)}^{+\infty} t dF(t).$$

Proof: Let $\gamma_1,\ldots,\gamma_n$ be independent random variables with the distribution μ and let $u_{1:n} \leq u_{2:n} \leq \cdots \leq u_{n:n}$ be the corresponding order statistics, that is, the permutation of $\gamma_1, \ldots, \gamma_n$ in the increasing order. Then

$$\max_{x \in X_n} \sum_{i \in x} \gamma_i = \sum_{m=n-\alpha n+o(n)}^{n} u_{m:n}.$$

Consequently, $\Gamma(X_n)$ is the expectation of the last sum.

The corresponding asymptotics for the order statistics is well known; see, for example, [S73]. ∎

Now we are ready to complete the proof of Theorem 4.2.

Proof of Part (2) of Theorem 4.2: Without loss of generality, we assume that the cumulative distribution function F of μ is continuous and strictly increasing. Let us choose α and k_n as described, so $X_n \subset \{0,1\}^n$ is the Hamming sphere of radius $\alpha n + o(n)$ in $\{0,1\}^n$.

As is known (see, for example, Theorem 1.4.5 of [Li99]),

$$\begin{aligned}\lim_{n \longrightarrow +\infty} \frac{\ln |X_n|}{n} =& \alpha \ln \frac{1}{\alpha} + (1-\alpha) \ln \frac{1}{1-\alpha} \\ =& (1-F(a_0)) \ln \frac{1}{1-F(a_0)} + F(a_0) \ln \frac{1}{F(a_0)}.\end{aligned}$$

Moreover, by Lemma 4.6,

$$\lim_{n \longrightarrow +\infty} \frac{\Gamma(X_n)}{n} = \int_{a_0}^{+\infty} t dF(t).$$

Hence

$$\lim_{n \longrightarrow +\infty} \frac{G(X_n, \tau)}{n} = (1-F(a_0)) \ln \frac{1}{1-F(a_0)} + F(a_0) \ln \frac{1}{F(a_0)} - \tau \int_{a_0}^{+\infty} t dF(t).$$

On the other hand,

$$\begin{aligned}g_\tau(a) =& \ln(1+e^{-\tau a}) - \tau \int_a^{+\infty} (1-F(t)) dt \\ =& \ln(1+e^{-\tau a}) + \tau a(1-F(a)) - \tau \int_a^{+\infty} t dF(t).\end{aligned}$$

Since a_0 is a critical point of g_τ, we have

$$\tau a_0 = \ln F(a_0) - \ln(1-F(a_0)),$$

cf. Section 4.3. Therefore,

$$\begin{aligned}g_\tau(a_0) =& -\ln F(a_0) + (\ln F(a_0) - \ln(1-F(a_0)))(1-F(a_0)) - \tau \int_{a_0}^{+\infty} t dF(t) \\ =& (1-F(a_0)) \ln \frac{1}{1-F(a_0)} + F(a_0) \ln \frac{1}{F(a_0)} - \tau \int_{a_0}^{+\infty} t dF(t)\end{aligned}$$

and the proof follows. ∎

Some remarks are in order.

(4.7) Remarks:

(4.7.1) Optimizing in a in Lemma 4.4, we substitute $a = \Gamma(X_0) - \Gamma(X_1)$ and obtain the inequality

$$\begin{aligned}\Gamma(X) \geq &(1 - F(a))\Gamma(X_1) + F(a)\Gamma(X_0) + \int_a^{+\infty} t dF(t) \\ =&\Gamma(X_0) + \int_{\Gamma(X_0)-\Gamma(X_1)}^{+\infty} (1 - F(t))dt.\end{aligned}$$

This inequality is harder to work with than with that of Theorem 4.2 but it sometimes leads to more delicate estimates, see Section 7.

(4.7.2) M. Talagrand proved in [T94] that for every non-empty set X of subsets of $\{1, \ldots, n\}$ there is a "shifted" set X' of subsets of $\{1, \ldots, n\}$ such that $|X'| = |X|$, $\Gamma(X') \leq \Gamma(X)$, X' is *hereditary* (that is, if $x \in X'$ and $y \subset x$ then $y \in X'$) and *left-hereditary* (that is, if $x \in X'$, $i \in x$, $j \notin x$ and $j < i$ then the subset $x \cup \{j\} \setminus \{i\}$ also lies in X').

(4.7.3) Suppose that μ is such that the inequality $\ln |X| \leq c\Gamma(X, \mu)$ holds for some constant $c = c(\mu)$ and every non-empty family X of k-subsets of $\{1, \ldots, n\}$ with arbitrary k and n. Choosing $\tau = c$ in Theorem 4.2, we conclude that we must have $g_c(a) \leq 0$ for all $a \geq 0$, from which it follows that for any $c' > c$ one can find an arbitrarily large a such that $1 - F(a) > e^{-c'a}$. In other words, μ should have a tail that is at least exponential.

5. General Estimates: the Lower Bound

Let us fix a symmetric probability measure μ with the cumulative distribution function F. In this section, we prove the following main result.

(5.1) Theorem: *Assume that the moment generating function*

$$L(\delta, \mu) = L(\delta) = \mathbf{E}e^{\delta x} = \int_{-\infty}^{+\infty} e^{\delta x} d\mu(x)$$

is finite in some neighborhood of $\delta = 0$. Let

$$h(t, \mu) = h(t) = \sup_{\delta \geq 0}(\delta t - \ln L(\delta)) \quad \textit{for } t \geq 0.$$

(1) *For any non-empty family X of k-subsets of $\{1, \ldots, n\}$, we have*

$$k^{-1} \ln |X| \geq h(t) \quad \textit{for } t = k^{-1}\Gamma(X).$$

(2) *For any* $t > 0$ *such that* $F(t) < 1$ *and for any* $0 < \epsilon < 0.1$ *there exist* $k = k(t, \epsilon, \mu)$, $n = n(k)$, *and a family of* k*-subsets of the set* $\{1, \ldots, n\}$ *such that*

$$k^{-1}\Gamma(X) \geq (1-\epsilon)t \quad \textit{and} \quad k^{-1}\ln|X| \leq h(t) + \epsilon.$$

Before proving Theorem 5.1, we summarize some properties of $L(\delta)$ and $h(t)$.

(5.2) Preliminaries. Let $f(\delta) = \ln L(\delta)$. Thus we assume that $f(\delta)$ is finite on some interval in $\mathbb{R}$, possibly on the whole line. It is known that $f(\delta)$ is convex and continuous on the interval where it is finite; see, for example, Section 5.11 of [GS01]. Since μ is symmetric, we have $f(0) = 0$ and from Jensen's inequality we conclude that $f(\delta) \geq 0$ for all δ.

The function $h(t)$ is convex conjugate to $f(\delta)$. Therefore, $h(t)$ is finite on some interval where it is convex, continuous and approaches $+\infty$ as t approaches a boundary point not in the interval. Besides,

$$h(t) = \frac{t^2}{2D} + O(t^4) \quad \text{for } t \approx 0,$$

where D is the variance of μ. In particular, $h(0) = 0$ and $h(t)$ is increasing for $t \geq 0$, see Section 5.11 of [GS01].

Now we are ready to prove Theorem 5.1.

Proof of Theorem 5.1: Let us prove Part (1). Without loss of generality, we assume that $\Gamma(X) > 0$. Let us choose a positive integer m, let $N = nm$, $K = km$ and let

$$Y = \underbrace{X \times \ldots \times X}_{m \text{ times}} \subset \{0,1\}^N.$$

Let us pick a point $y = (x_1, \ldots, x_m)$ from Y, where $x_i \in X$ for $i = 1, \ldots, m$. Thus some K coordinates of y are 1's and the rest are 0's. Let us endow $\mathbb{R}^N$ with the product measure $\mu^{\otimes N}$ and let $\gamma_1, \ldots, \gamma_K$ be independent random variables with the distribution μ. Then, for any $t > 0$

$$\mathbf{P}\{c \in \mathbb{R}^N : \langle c, y\rangle > mt\} = \mathbf{P}\left\{\sum_{i=1}^K \gamma_i > mt\right\} = \mathbf{P}\left\{\sum_{i=1}^K \gamma_i > K\frac{t}{k}\right\}.$$

By the Large Deviations Inequality (see, for example, Section 5.11 of [GS01])

$$\mathbf{P}\left\{\sum_{i=1}^K \gamma_i \geq K\frac{t}{k}\right\} \leq \exp\{-Kh(t/k)\}.$$

Therefore,

$$\mathbf{P}\{c \in \mathbb{R}^N : \quad \max_{y \in Y} \langle c, y \rangle > mt\} \leq |Y| \exp\{-Kh(t/k)\} = (|X| \exp\{-kh(t/k)\})^m.$$

Since a vector $c \in \mathbb{R}^N$ is an m-tuple $c = (c_1, \ldots, c_m)$ with $c_i \in \mathbb{R}^n$ and

$$\max_{y \in Y} \langle c, y \rangle = \sum_{i=1}^m \max_{x \in X} \langle c_i, x \rangle,$$

the last inequality can be written as

$$\mathbf{P}\{c_1, \ldots, c_m : \frac{1}{m} \sum_{i=1}^m \max_{x \in X} \langle c_i, x \rangle > t\} \leq (|X| \exp\{-kh(t/k)\})^m.$$

However, by the Law of Large Numbers

$$\frac{1}{m} \sum_{i=1}^m \max_{x \in X} \langle c_i, x \rangle \longrightarrow \Gamma(X) \quad \text{in probability}$$

as $m \longrightarrow +\infty$. Therefore, for any $0 < t < \Gamma(X)$,

$$\mathbf{P}\{c_1, \ldots, c_m : \quad \frac{1}{m} \sum_{i=1}^m \max_{x \in X} \langle c_i, x \rangle > t\} \longrightarrow 1 \quad \text{as } m \longrightarrow +\infty.$$

Therefore, we must have

$$|X| \exp\{-kh(t/k)\} \geq 1 \quad \text{for every } t < \Gamma(X).$$

Hence

$$k^{-1} \ln |X| \geq h(t) \quad \text{for every } t < k^{-1}\Gamma(X),$$

and the proof follows by the continuity of h, cf. Section 5.2.

Let us prove Part (2). Let $\gamma_1, \ldots, \gamma_k$ be independent random variables having the distribution μ. By the Large Deviations Theorem (see Section 5.11 of [GS01]), if $k = k(\epsilon, t, \mu)$ is sufficiently large then

$$\mathbf{P}\left\{ \sum_{i=1}^k \gamma_i > kt \right\} \geq \exp\{-k(h(t) + \epsilon/2)\}.$$

We make k large enough to ensure, additionally, that $(\ln 3 + \ln \ln(1/\epsilon))/k \leq \epsilon/2$.

Let $|X|$ be the largest integer not exceeding

$$3 \ln \frac{1}{\epsilon} \exp\{k(h(t) + \epsilon/2)\},$$

so $k^{-1} \ln |X| \leq h(t) + \epsilon$, and let X consist of $|X|$ pairwise disjoint k-subsets of $\{1, \dots, n\}$ for a sufficiently large $n = n(k)$.

Suppose that $c = (\gamma_1, \dots, \gamma_n)$ is a random vector of independent weights with the distribution μ. Since $x \in X$ are disjoint, the weights $\sum_{i \in x} \gamma_i$ of subsets from X are independent random variables. Let $w(X, c)$ be the largest weight of a subset $x \in X$. We have

$$\mathbf{P}\{c : w(X, c) \leq kt\} \leq (1 - \exp\{-k(h(t) + \epsilon/2)\})^{|X|} \leq \epsilon/2.$$

Similarly (since μ is symmetric):

$$\mathbf{P}\{c : w(X, -c) \leq kt\} \leq \epsilon/2,$$

and, therefore,

$$\mathbf{P}\{c : w(X, c) + w(X, -c) \leq 2kt\} \leq \epsilon.$$

Since $w(X, c) + w(X, -c)$ is always non-negative, its expectation is at least $(1 - \epsilon)2kt$. On the other hand, this expectation is $2\Gamma(X)$. Hence we have constructed a family X of k-subsets such that

$$k^{-1}\Gamma(X) \geq (1 - \epsilon)t \quad \text{and} \quad k^{-1} \ln |X| \leq h(t) + \epsilon. \qquad \blacksquare$$

(5.3) Remarks:

(5.3.1) Using the convexity of $h(t)$, one can extend the bound of Part (1) of Theorem 5.1 to families X of *at most* k-element subsets of $\{1, \dots, n\}$.

(5.3.2) Suppose that the moment generating function $L(\delta, \mu)$ is infinite for all δ except for $\delta = 0$. Let us choose $t > 0$ and $0 < \epsilon < 0.1$. We claim that there exists a family X of k-subsets of $\{1, \dots, n\}$ such that

$$k^{-1}\Gamma(X) \geq (1 - \epsilon)t \quad \text{and} \quad k^{-1} \ln |X| \leq \epsilon$$

(in other words, we can formally take $h(t) \equiv 0$ in Part (2) of Theorem 5.1). Let γ be a random variable with the distribution μ. For $c > 0$, let γ_c be the *truncation* of γ:

$$\gamma_c = \begin{cases} \gamma, & \text{if } |\gamma| \leq c, \\ 0, & \text{if } |\gamma| > c. \end{cases}$$

Let μ_c be the distribution of γ_c. It is not hard to see that $\Gamma(X, \mu) \geq \Gamma(X, \mu_c)$ (consider $\Gamma(X)$ as the expectation of $0.5w(X, c) + 0.5w(X, -c)$, where $w(X, c)$ is the maximum weight of a subset $x \in X$ for the vector $c = (\gamma_1, \dots, \gamma_n)$ of weights). Choosing a sufficiently large c brings $h(t, \mu_c)$ arbitrarily close to 0. Then we construct a set X as in Part (2) of Theorem 5.1.

We conclude that to have a positive lower bound of $k^{-1} \ln |X|$ given a positive lower bound of $k^{-1}\Gamma(X, \mu)$ in the class of all non-empty families X of k-subsets of $\{1, \ldots, n\}$ for all k and n, we must have $L(\delta, \mu)$ finite for all δ in some open interval containing 0. In other words, μ must have a tail that is at most exponential: $1 - F(t) \leq e^{-\delta_1 t}$ for some $\delta_1 > 0$ and all sufficiently large t.

(5.3.3) Our proof of Part (2) of Theorem 5.1 seems to require n to be exponentially large in k. This is not so, since every suitable pair n, k can be rescaled to a suitable pair $N = nm$, $K = km$ for a positive integer m. Let X be a family of k-subsets of $\{1, \ldots, n\}$ constructed in the proof of Part (2) and let

$$Y = \underbrace{X \times \ldots \times X}_{m \text{ times}} \subset \{0, 1\}^N.$$

Then Y is a family of K-subsets of $\{1, \ldots, N\}$ and

$$K^{-1}\Gamma(Y) \geq (1 - \epsilon)t \quad \text{and} \quad K^{-1} \ln |Y| \leq h(t) + \epsilon.$$

6. The logistic measure: Proofs

In this section, we prove Theorems 3.1 and 3.3.

Proof of Theorem 3.1: To prove Part (1), let us choose $\tau = 1$ in Part (1) of Theorem 4.2. We have

$$g_1(a) = \ln(1 + e^{-a}) - \int_a^{+\infty} \frac{e^{-t}}{1 + e^{-t}} dt = 0 \quad \text{for all } a.$$

Hence

$$\ln |X| \leq \Gamma(X)$$

as claimed.

To prove Part (2), we use Part (1) of Theorem 5.1. The moment generating function of the logistic distribution is given by

$$L(\delta) = \int_{-\infty}^{+\infty} \frac{e^{\delta x}}{e^x + e^{-x} + 2} dx = \frac{\pi\delta}{\sin \pi\delta} \quad \text{for } -1 < \delta < 1,$$

see [M85]. Hence the formula for $h(t)$ follows.

It follows from Section 5.2 that h is convex and increasing. ∎

Now we are ready to prove optimality of the logistic distribution.

Proof of Theorem 3.3: Let us choose $\mu \in \mathcal{M}$ and let F_μ be the cumulative distribution function of μ. We claim that $F_\mu(t) < 1$ for all $t \in \mathbb{R}$. To see that, we let $\tau = 1$ in Theorem 4.2. If $F_\mu(t) = 1$ then $g_1(t) > 0$ and, by Part (2) of Theorem 4.2, there is a set $X \subset \{0,1\}^n$ with $\ln|X| > \Gamma(X)$, which contradicts the definition of $\mathcal{M}$.

Let us assume first that the moment generating function $L(\delta, \mu)$ is finite in some neighborhood of $\delta = 0$. Then, by Theorem 5.1, we have $c(t,\mu) = h(t,\mu)$ and hence we must prove that $h(t,\mu) \leq h(t,\mu_0)$, where μ_0 is the logistic distribution.

Let

$$T(a,\mu) = \int_a^{+\infty} (1 - F_\mu(t))dt.$$

We can write

$$\int_0^{+\infty} e^{\delta x} dF_\mu(x) = -\int_0^{+\infty} e^{\delta x} d(1 - F_\mu(x)) = \frac{1}{2} + \int_0^{+\infty} \delta e^{\delta x}(1 - F_\mu(x))dx$$

$$= \frac{1}{2} + \int_0^{+\infty} \delta e^{\delta x} d(-T(x,\mu)) = \frac{1}{2} + \delta T(0,\mu) + \int_0^{+\infty} \delta^2 e^{\delta x} T(x,\mu)dx.$$

Similarly,

$$\int_{-\infty}^0 e^{\delta x} dF_\mu(x) = \int_0^{+\infty} e^{-\delta x} dF_\mu(x) = \frac{1}{2} - \delta T(0,\mu) + \int_0^{+\infty} \delta^2 e^{-\delta x} T(x,\mu)dx.$$

Therefore,

$$L(\delta,\mu) = 1 + \delta^2 \int_0^{+\infty} (e^{-\delta x} + e^{\delta x})T(x,\mu)dx.$$

Since $\ln|X| \leq \Gamma(X)$, by Part (2) of Theorem 4.2 we conclude that

$$T(a,\mu) \geq \ln(1 + e^{-a}) = T(a,\mu_0) \quad \text{for all } a \geq 0.$$

Therefore, $L(\delta,\mu) \geq L(\delta,\mu_0)$ and $h(t,\mu) \leq h(t,\mu_0)$ for all $t \geq 0$, as claimed.

Suppose now that the moment generating function $L(\delta,\mu)$ is infinite for $\delta \neq 0$. Then, as follows from Remark 5.3.2, $c(t,\mu) = 0$ for all $t > 0$, which completes the proof. ∎

7. The exponential measure

Let us choose μ_1 to be the measure with density

$$\frac{1}{2}e^{-|\gamma|} \quad \text{for } \gamma \in \mathbb{R}.$$

As we have already mentioned, one of the results of [T94] is the estimate

$$\ln|X| \leq c\Gamma(X, \mu_1) = c\Gamma(X)$$

for some absolute constant c. In this section, we find the optimal value of c and establish some general isoperimetric inequalities which, we believe, are interesting in their own right.

(7.1) THEOREM: *Let μ_1 be the measure with density $e^{-|\gamma|}/2$ for $\gamma \in \mathbb{R}$.*

(1) *Let $X \subset \{0,1\}^n$ be a non-empty subset of the Boolean cube. Then*

$$\ln|X| \leq (2\ln 2)\Gamma(X).$$

(2) *Let $X \subset \{0,1\}^n$ be a non-empty subset of the Boolean cube such that $\xi_1 + \cdots + \xi_n \leq k$ for every $(\xi_1, \ldots, \xi_n) \in X$. That is, X lies in the Hamming ball of radius k and we may interpret X as a family of at most k-element subsets of $\{1, \ldots, n\}$. Then*

$$\ln|X| \leq \Gamma(X) + k\ln 2.$$

Before we prove Theorem 7.1, we note that $c = 2\ln 2$ is the best possible value in Part (1). If X is a m-dimensional face of the Boolean cube then $\ln|X| = m\ln 2$ and we show that $\Gamma(X) = m/2$, so the equality holds. As in Section 3, it suffices to check the formula for $X = \{0,1\}$, in which case

$$\Gamma(X) = \frac{1}{2}\int_0^{+\infty} xe^{-x}\,dx = \frac{1}{2}.$$

The inequality of Part (2) is asymptotically sharp: if X is the Hamming sphere of radius $k = o(n)$ in $\{0,1\}^n$, then

$$\Gamma(X) = \ln|X| - k\ln 2 + o(k) \quad \text{as } k \longrightarrow +\infty,$$

cf. Lemma 4.6.

As for the lower bound, using Part (1) of Theorem 5.1 one can show that for any non-empty family X of k-subsets of $\{1, \ldots, n\}$, we have

$$k^{-1}\ln|X| \geq h(k^{-1}\Gamma(X)),$$

where

$$\begin{aligned} h(t) =&\sqrt{1+t^2}+\ln(\sqrt{1+t^2}-1)-2\ln t+\ln 2-1 \\ =&t-\ln t-O(1) \quad \text{for large } t. \end{aligned}$$

Thus the exponential distribution also allows us to estimate $\ln|X|$ up to a constant factor. However, the estimates are not as good as for the logistic distribution.

Proof of Theorem 7.1: To prove Part (1), we use Part (1) of Theorem 4.2.

The function $g_\tau(a)$ is given by

$$g_\tau(a)=\ln(1+e^{-\tau a})-\frac{\tau}{2}e^{-a} \quad \text{for } a\in\mathbb{R}.$$

Let us consider the critical points of g_τ.

We have

$$g'_\tau(a)=\frac{\tau}{2}\Big(\frac{e^{(\tau-1)a}+e^{-a}-2}{1+e^{\tau a}}\Big).$$

Since the numerator of the fraction is a linear combination of two exponential functions and a constant, it can have at most two real zeros. We observe that $a=0$ is a zero and that $g'_\tau(a)<0$ for small $a>0$ provided $\tau<2$.

Hence for $\tau<2$ the function g_τ has at most one critical point $a>0$ which has to be a point of local minimum.

Therefore

$$\sup_{a\geq 0} g_\tau(a)=\max\{g_\tau(0),\ 0\} \quad \text{for all } \tau<2.$$

Let us choose $\tau=2\ln 2$. Then $g_\tau(0)=0$ and we conclude that

$$\sup_{a\geq 0} g_\tau(a)=0.$$

By Part (1) of Theorem 4.2, we conclude that

$$\ln|X|\leq\tau\Gamma(X)=(2\ln 2)\Gamma(X).$$

We prove Part (2) by induction on n. If $n=1$, there are two cases. If X consists of a single point then $\Gamma(X)=0$, $\ln|X|=0$ and the inequality is satisfied. If $X=\{0,1\}$ then $k=1$ and $\Gamma(X)=1/2$, hence the inequality holds as well.

Suppose that $n>1$. Clearly, we can assume that $k>0$. Without loss of generality, we may assume that X is hereditary, see Remark 4.7.2. Let us construct sets $X_0, X_1\subset\{0,1\}^{n-1}$ as in Lemma 4.4. We note that X_0 lies in the Hamming ball of radius k and X_1 lies in the Hamming ball of radius $k-1$. Since X is hereditary, $X_1 \subset X_0$. Therefore,

$$|X_0| \geq |X_1| \quad \text{and} \quad \Gamma(X_0) \geq \Gamma(X_1).$$

The inequality of Remark 4.7.1 gives us

$$\Gamma(X) \geq \Gamma(X_0) + \frac{1}{2}\exp\{\Gamma(X_1) - \Gamma(X_0)\}.$$

Let us consider a function

$$f(a,b) = a + \frac{1}{2}e^{b-a}.$$

It is easy to see that for every a the function is increasing in b and that for every b it is increasing on the interval $a \geq b - \ln 2$.

Applying the induction hypothesis to X_0 and X_1, we conclude

$$\begin{aligned} f(\Gamma(X_0), \Gamma(X_1)) \geq & f(\Gamma(X_0), \ln|X_1| - (k-1)\ln 2) \\ \geq & f(\ln|X_0| - k\ln 2, \ln|X_1| - (k-1)\ln 2). \end{aligned}$$

Therefore,

$$\begin{aligned} \Gamma(X) \geq & \ln|X_0| - k\ln 2 + \frac{|X_1|}{|X_0|} = (\ln|X| - k\ln 2) + \Big(\frac{|X_1|}{|X_0|} - \ln\frac{|X_1| + |X_0|}{|X_0|}\Big) \\ = & (\ln|X| - k\ln 2) + (t - \ln(1+t)) \quad \text{for } t = \frac{|X_1|}{|X_0|} \\ \geq & \ln|X| - k\ln 2. \end{aligned}$$

The proof now follows. ∎

8. An asymptotic solution to the isoperimetric problem

In this section, we discuss what sets $X_n \subset \{0,1\}^n$ with the smallest ratio $\Gamma(X_n,\mu)/\ln|X_n|$ may look like. We claim that for any symmetric probability measure μ with finite variance and for a sufficiently large n we can choose X_n to be the product of at most two Hamming spheres.

(8.1) Theorem: *Let us fix a symmetric probability measure μ and a number*

$$0 < \alpha < \ln 2.$$

Then there exist numbers β_i, λ_i, $i = 1,2$, depending on α and μ only, such that

$$\begin{aligned} & 0 \leq \beta_i \leq \lambda_i \quad \textit{for } i = 1,2, \\ & \lambda_1 + \lambda_2 = 1 \end{aligned}$$

and the following holds.

Let S_n^i be the Hamming sphere of radius $\beta_i n + o(n)$ in the Boolean cube of dimension $d_i = \lambda_i n + o(n)$, $i = 1, 2$, such that $d_1 + d_2 = n$, and let $Y_n = S_n^1 \times S_n^2$ be the direct product of the spheres considered as a subset of the Boolean cube of dimension n.

Then

$$\ln |Y_n| = \alpha n + o(n)$$

and for any sequence of sets $X_n \subset \{0,1\}^n$ such that

$$\ln |X_n| = \alpha n + o(n),$$

we have

$$\Gamma(Y_n, \mu) \leq \Gamma(X_n, \mu) + o(n).$$

Proof: Let F be the cumulative distribution function of μ. Without loss of generality, we assume that F is continuous and strictly increasing. Given μ and α, let us consider the function

$$H(\tau, x) = \frac{\alpha}{\tau} - \frac{\ln(1 + e^{-\tau x})}{\tau} + \int_x^{+\infty} (1 - F(t))dt$$

of two variables $\tau > 0$ and $x \geq 0$.

By Part (1) of Theorem 4.2, for any $\tau > 0$,

(8.1.1) $$n^{-1}\Gamma(X_n) \geq \inf_{x \geq 0} H(\tau, x) + o(1) \quad \text{provided } \ln |X_n| = \alpha n + o(n).$$

We claim that there exists $0 < \tau_0 < +\infty$ such that

$$\inf_{x \geq 0} H(\tau_0, x) \geq \inf_{x \geq 0} H(\tau, x) \quad \text{for all } \tau \geq 0.$$

Indeed, since $\alpha < \ln 2$,

$$\inf_{x \geq 0} H(\tau, x) \longrightarrow -\infty \quad \text{as } \tau \longrightarrow 0+.$$

Also,

$$\inf_{x \geq 0} H(\tau, x) \longrightarrow 0 \quad \text{as } \tau \longrightarrow +\infty.$$

On the other hand, choosing $x_1 > 0$ such that

$$\int_{x_1}^{+\infty} (1 - F(t))dt = \delta > 0$$

and τ_1 such that

$$\alpha - \ln(1+e^{-\tau_1 x_1}) > 0 \quad \text{and} \quad \tau_1^{-1}|\alpha - \ln 2| < \delta$$

we observe that

$$\inf_{x \geq 0} H(\tau_1, x) > 0,$$

which implies that there exists $0 < \tau_0 < +\infty$ maximizing $\inf_{x\geq 0} H(\tau, x)$.

Our next goal is to show that one can find $0 \leq x_1, x_2 \leq +\infty$ such that

$$H(\tau_0, x_1) = H(\tau_0, x_2) = \inf_{x\geq 0} H(\tau_0, x) \tag{8.1.2}$$

and such that

$$\frac{\tau_0 x_1}{e^{\tau_0 x_1}+1} + \ln(1+e^{-\tau_0 x_1}) \geq \alpha \quad \text{and} \quad \frac{\tau_0 x_2}{e^{\tau_0 x_2}+1} + \ln(1+e^{-\tau_0 x_2}) \leq \alpha \tag{8.1.3}$$

(it is possible that $x_1 = x_2$ or that $x_2 = +\infty$).

For ϵ in a small neighborhood of 0, we define $x_\epsilon \geq 0$ as a point such that

$$H(\tau_0+\epsilon,\ x_\epsilon) = \inf_{x\geq 0} H(\tau_0+\epsilon, x)$$

(possibly $x_\epsilon = +\infty$). We obtain x_1 as a limit point of x_ϵ as $\epsilon \longrightarrow 0-$ and x_2 as a limit point of x_ϵ as $\epsilon \longrightarrow 0+$. Clearly, (8.1.2) holds and it remains to show that (8.1.3) holds as well.

Indeed,

$$\begin{aligned} H(\tau_0, x_i) \geq & H(\tau_0+\epsilon, x_\epsilon) = H(\tau_0, x_\epsilon) + \epsilon \frac{\partial}{\partial \tau} H(\overline{\tau}, x_\epsilon) \\ \geq & H(\tau_0, x_i) + \epsilon \frac{\partial}{\partial \tau} H(\overline{\tau}, x_\epsilon) \end{aligned}$$

for some $\overline{\tau}$ between τ_0 and $\tau_0+\epsilon$ and $i = 1, 2$.

Besides,

$$\frac{\partial}{\partial \tau} H(\tau, x) = \frac{1}{\tau^2}\left(\frac{\tau x}{1+e^{\tau x}} + \ln(1+e^{-\tau x}) - \alpha\right),$$

from which we deduce (8.1.3).

Additionally, from (8.1.2) we deduce that if $0 < x_i < +\infty$, we must have

$$\frac{\partial}{\partial x} H(\tau_0, x_i) = 0,$$

that is,

$$\frac{1}{e^{\tau_0 x_i}+1} = 1 - F(x_i), \quad \text{for } i = 1, 2, \tag{8.1.4}$$

which also holds for $x_i = 0$ and $x_i = +\infty$.

Now we are ready to define λ_i and β_i. Namely, we write

$$\alpha = \sum_{i=1,2} \lambda_i \left(\frac{\tau_0 x_i}{e^{\tau_0 x_i} + 1} + \ln(1 + e^{-\tau_0 x_i})\right) \quad \text{where } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1,$$

cf. (8.1.3).

Next, we define β_1 and β_2 by

$$\beta_i = \frac{\lambda_i}{e^{\tau_0 x_i} + 1} \quad \text{for } i = 1, 2.$$

Let S_n^i be the Hamming sphere of dimension $\lambda_i n + o(n)$ and radius $\beta_i n + o(n)$. Using Theorem 1.4.5 of [Li99], we obtain

$$\begin{aligned} \frac{1}{\lambda_i n} \ln |S_n^i| =& \frac{1}{e^{\tau_0 x_i} + 1} \ln(e^{\tau_0 x_i} + 1) + \frac{e^{\tau_0 x_i}}{e^{\tau_0 x_i} + 1} \ln(1 + e^{-\tau_0 x_i}) + o(1) \\ =& \frac{\tau_0 x_i}{e^{\tau_0 x_i} + 1} + \ln(1 + e^{-\tau_0 x_i}) + o(1). \end{aligned}$$

Thus for $Y_n = S_n^1 \times S_n^2$, we have

$$\ln |Y_n| = \alpha n + o(n),$$

as claimed.

By Part (2) of Theorem 4.2 and (8.1.4)

$$\frac{\Gamma(S_n^i)}{\lambda_i n} = H(\tau_0, x_i) + o(1).$$

Using (8.1.2) we conclude that for $Y_n = S_n^1 \times S_n^2$, we have

$$\begin{aligned} \frac{\Gamma(Y_n)}{n} =& \frac{\Gamma(S_n^1) + \Gamma(S_n^2)}{n} = \lambda_1 H(\tau_0, x_1) + \lambda_2 H(\tau_0, x_2) + o(1) \\ =& \inf_{x \geq 0} H(\tau_0, x) + o(1). \end{aligned}$$

Hence, by (8.1.1),

$$\Gamma(Y_n) \leq \Gamma(X_n) + o(n),$$

which completes the proof. ∎

Acknowledgement: The authors are very grateful to Nathan Linial for many helpful discussions and encouragement, in particular, for his suggestion to look for an optimal solution to the "inverse isoperimetric problem".

References

[ABS98] N. Alon, R. Boppana and J. Spencer, *An asymptotic isoperimetric inequality,* Geometric and Functional Analysis **8** (1998), 411–436.

[B97] A. Barvinok, *Approximate counting via random optimization,* Random Structures & Algorithms **11** (1997), 187–198.

[BS01] A. Barvinok and A. Samorodnitsky, *The distance approach to approximate combinatorial counting,* Geometric and Functional Analysis **11** (2001), 871–899.

[GS01] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes,* Third edition, The Clarendon Press, Oxford University Press, New York, 2001.

[JS97] M. Jerrum and A. Sinclair, *The Markov chain Monte Carlo method: an approach to approximate counting and integration,* in *Approximation Algorithms for NP-hard Problems* (D. S. Hochbaum, ed.), PWS, Boston, 1997, pp. 483–520.

[JSV04] M. Jerrum, A. Sinclair and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries,* Journal of the Association for Computing Machinery **51** (2004), 671–697.

[La97] R. Latała, *Sudakov minoration principle and supremum of some processes,* Geometric and Functional Analysis **7** (1997), 936–953.

[Le91] I. Leader, *Discrete isoperimetric inequalities,* in *Probabilistic Combinatorics and its Applications (San Francisco, CA 1991)*, Proceedings of Symposia in Applied Mathematics, Vol. 44, American Mathematical Society, Providence, RI, 1991, pp. 57–80.

[Led01] M. Ledoux, *The Concentration of Measure Phenomenon,* Mathematical Surveys and Monographs, Vol. 89, American Mathematical Society, Providence, RI, 2001.

[Li99] J. H. van Lint, *Introduction to Coding Theory,* Third edition, Graduate Texts in Mathematics, Vol. 86, Springer-Verlag, Berlin, 1999.

[M85] H. J. Malik, *Logistic distribution,* in *Encyclopedia of Statistical Sciences* (S. Kotz, N. L. Johnson and C. B. Read, eds.), Vol. 5, Wiley-Interscience, New York, 1985, pp. 123–129.

[M04] G. Mikhalkin, *Amoebas of algebraic varieties and tropical geometry,* in *Different Faces of Geometry*, Int. Math. Ser. (N.Y.), Kluwer/Plenum, New York, 2004, pp. 257–300.

[PS98] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity,* Dover, New York, 1998.

[S73] S. M. Stigler, *The asymptotic distribution of the trimmed mean,* Annals of Statistics **1** (1973), 472–477.

[T94] M. Talagrand, *The supremum of some canonical processes,* American Journal of Mathematics **116** (1994), 283–325.

[T95] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces,* Publications Mathématiques de l'Institut des Hautes Études Scientifiques **81** (1995), 73–205.

[Y03] A. Yong, *Experimental C++codes for estimating permanents, hafnians and the number of forests in a graph,* available at http://www.math.lsa.umich.edu/~barvinok/papers.html, 2003.