ISRAEL JOURNAL OF MATHEMATICS **125** (2001), 369–380

PROOF OF A HYPERCONTRACTIVE ESTIMATE VIA ENTROPY

BY

EHUD FRIEDGUT

Institute of Mathematics, The Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: ehudf@math.huji.ac.il

AND

VOJTECH RÖDL*

Department of Mathematics, Emory University
Atlanta, GA 30033, USA
e-mail: rodl@mathcs.emory.edu

ABSTRACT

Consider the probability space $W = \{-1,1\}^n$ with the uniform (=product) measure. Let $f\colon W \to \mathbf{R}$ be a function. Let $f = \sum f_I X_I$ be its unique expression as a multilinear polynomial where $X_I = \prod_{i\in I} x_i$. For $1 \le m \le n$ let $f_{\hat{m}} = \sum_{|I|=m} f_I X_I$. Let $T_\varepsilon(f) = \sum f_I \varepsilon^{|I|} X_I$ where $0 < \varepsilon < 1$ is a constant. A hypercontractive inequality, proven by Bonami and independently by Beckner, states that

$$|T_\varepsilon(f)|_2 \le |f|_{1+\varepsilon^2}.$$

This inequality has been used in several papers dealing with combinatorial and probabilistic problems. It is equivalent to the following inequality via duality: For any $q \ge 2$

$$|f_{\hat{m}}|_q \le (\sqrt{q-1})^m |f_{\hat{m}}|_2.$$

In this paper we prove a special case with a slightly weaker constant, which is sufficient for most applications. We show

$$|f_{\hat{m}}|_4 \le c^m |f_{\hat{m}}|_2$$

where $c = \sqrt[4]{28}$. Our proof uses probabilistic arguments, and a generalization of Shearer's Entropy Lemma, which is of interest in its own right.

* Supported partially by NSF Award Abstract #0071261.
Received July 11, 2000

1. Introduction

Consider the probability space $W = \{-1,1\}^n$ with the uniform (=product) measure. For two functions $f, g\colon W \to \mathbf{R}$ define their inner product by $\langle f, g\rangle = 2^{-n}\sum_{w\in W} f(w)g(w)$. The set of multilinear monomials

$$\{X_I = \prod_{i\in I} x_i\}_{I\subseteq\{1,\ldots,n\}}$$

is an orthonormal basis for the space of real functions on W. Hence any $f\colon W \to \mathbf{R}$ has a unique representation of the form $f = \sum f_I X_I$. For a fixed constant $0 < \varepsilon < 1$ define the operator T_ε by $T_\varepsilon(f) = \sum f_I \varepsilon^{|I|} X_I$. As usual, let $|f|_p = (\int |f|^p)^{1/p}$. Bonami [2] and independently Beckner [1] proved the following hypercontractive inequality:

THEOREM 1.1: *For any $0 < \varepsilon < 1$ and any $f : W \to \mathbf{R}$*

$$|T_\varepsilon f|_2 \le |f|_{1+\varepsilon^2}. \tag{1}$$

It turns out, perhaps surprisingly, that this result is extremely useful in studying combinatorial problems related to collective coin flipping, and influences of variables on Boolean functions on product spaces. A striking example of this appears in the paper [11] by Kahn Kalai and Linial, probably the first application of this inequality in this setting. This approach was later extended in [5]. In several related papers such as [13], [4], and [3], inequality (1) again plays a key role. The power of this approach is often revealed to anyone who attempts to solve these problems using alternative methods: in all the above cases (except perhaps a semi-exception in the case [3]) no proofs are known that do not use hypercontractivity. From a combinatorialist's point of view this is a state of affairs that is less than satisfactory, since the inequality is used as a "black box" in the key stages of the proofs. Studying the original proof of the inequality is not usually enlightening in the combinatorial setting. The proof uses induction on n, the dimension of the probability space, and submultiplicativity of norms of product operators. The base case of a two point space boils down to some elementary yet cumbersome calculus. It seems quite hard to translate this information back into combinatorial intuition concerning the problems at hand.

In this paper we prove a special case of (1), with a slightly worse constant. However, this weakened version is sufficient to derive all the bounds proved in the papers mentioned above (of course, with some loss in the constants). Our proof reduces the inequality to a combinatorial problem, which is then solved via an information-theoretic approach, analyzing the entropy of appropriate random variables. To this end we present a generalization of Shearer's Entropy Lemma.

Let $f\colon W \to \mathbf{R}$ be a function, $f = \sum f_I X_I$ its unique representation as a multilinear polynomial and $1 \le m \le n$ an integer. Let $f_{\hat{m}} = \sum_{|I|=m} f_I X_I$. It is observed, e.g. in [10], [13], that Theorem 1.1 is equivalent via duality to the following

THEOREM 1.2: *For any $q > 2$ and any $fW \to \mathbf{R}$*

$$|f_{\hat{m}}|_q \le (\sqrt{q-1})^m |f_{\hat{m}}|_2.$$

Our main result is

THEOREM 1.3: *For any $fW \to \mathbf{R}$*

$$|f_{\hat{m}}|_4 \le c^m |f_{\hat{m}}|_2 \tag{2}$$

where $c = \sqrt[4]{28}$.

The paper is organized as follows: In section 2 we present some elementary mathematical facts that will be needed throughout the paper. In section 3 we present the generalization of Shearer's Entropy Lemma. In section 4 we give the proof of Theorem 1.3, and in the final section we point out the thematic connection between the present paper and [9].

2. Background; Orthogonal functions and entropy

2.1 ORTHONORMALITY. Let $W = \{-1,1\}^n$ with the uniform product measure. Let $\mathbf{E}(X)$ denote the expectation of a random variable X.

For any monomial $X = \prod x_i^{d_i}$ we have

$$\mathbf{E}(X) = \begin{cases} 1 & \text{if all } d_i \text{ are even,} \\ 0 & \text{otherwise.} \end{cases}$$

This is because the x_i's are independent (this is a product space!) and, for any x_i, $\mathbf{E}(x_i) = 0, \mathbf{E}(x_i)^2 = 1$. For a set $I \subseteq \{1,\ldots,n\}$ let

$$X_I = \prod_{i \in I} x_i.$$

It follows from (3) that

$$\mathbf{E}(X_I X_J) = \delta_{IJ},$$

i.e., the family $\{X_I\}_{I \subseteq \{1,\ldots,n\}}$ is an orthonormal set with respect to the standard inner product. Since its cardinality is 2^n it is a basis for the space of real functions on W, hence any function $f\colon W \to \mathbf{R}$ has a unique expansion of the form $f =$

$\sum f_I X_I$. For a function $f\colon W \to \mathbf{R}$ let $|f|_p = (\mathbf{E}(|f|^p))^{1/p}$. Using (3) and linearity of expectation we get

$$|f|_2^2 = E(f^2) = \sum f_I^2 \tag{4}$$

(this is Parseval's identity), and

$$|f|_4^4 = E(f^4) = \sum_{I \Delta J \Delta K \Delta L = \emptyset} f_I f_J f_K f_L. \tag{5}$$

Recall that we will be interested in the 4-norm and the 2-norm of $f_{\hat{m}}$. In effect we confine ourselves to the case where f_I is non-zero only if $|I| = m$.

2.2 Entropy. The standard facts we quote below regarding entropy may be found in any book concerning information theory, e.g. [7]. In what follows X, Y etc. are discrete random variables taking values in any finite set. In our setting $\log = \log_2$.

The **entropy** of a random variable X is

$$H(X) = \sum_x p(x) \log \frac{1}{p(x)},$$

where we write $p(x)$ for $\Pr(X = x)$ and extend this convention in natural ways below. Note that the entropy of X does not depend on the values that X attains, but only on the probabilities of attaining them. The intuitive meaning of $H(X)$ is the number of bits of information conveyed by X on average. It is always true that

$$H(X) \le \log |\mathrm{Support}(X)|, \tag{6}$$

where $\mathrm{Support}(X)$ is the set of values that X attains. Equality is attained if and only if X is uniformly distributed on its support.

The **conditional entropy** of X given Y is

$$H(X|Y) = \mathbf{E}H(X|Y = y) = \sum_y p(y) \sum_x p(x|y) \log \frac{1}{p(x|y)}.$$

Intuitively, $H(X|Y)$ measures the expected amount of information X conveyed to one who knows the value of Y, where the expectation is taken over the values of Y. Clearly,

(7) If X and Y are independent then $H(X|Y) = H(X)$.

For a random vector $X = (X_1, \ldots, X_n)$ (note this is also a random variable), we have

$$H(X) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1, \ldots, X_{n-1}). \tag{8}$$

We will also use the following inequality:

$$H(X|Y) \le H(X), \tag{9}$$

and more generally, if $I \supseteq J$,

$$H(X|Y_i : i \in I) \le H(X|Y_i : i \in J). \tag{10}$$

This inequality, which is central to our proof, has the intuitive meaning that the more information one knows, the less information is conveyed by X.

3. The Entropy Lemma

Let us begin by recalling the definition of a hypergraph.

A hypergraph $H = (V, E)$ consists of a set of vertices, V, and a family E of subsets of V — the edges. A graph, for example, is a hypergraph where all edges are of size 2.

The following lemma is due to J. Shearer [6].

LEMMA 3.1: *Let t be a positive integer. Let $H = (V, E)$ be a hypergraph, and let $F_1, \ldots, F_r \subseteq V$ such that every vertex in V belongs to at least t of the sets F_i. Let $H_i, i = 1, \ldots, r$ be the projection hypergraphs: $H_i = (V, E_i)$ where $E_i = \{e \cap F_i : e \in E\}$. Then*

$$|E|^t \le \prod |E_i|.$$

The original proof of this lemma is quite easy, and uses induction on t. However, there exists a slightly different proof which is even more transparent. This proof is folklore, probably first discovered by Jaikumar Radhakrishnan [12]. The proof we give of the generalized lemma is a generalization of the folklore proof.

LEMMA 3.2: *Let $H, E, V, t, F_1, \ldots, F_r, E_1, \ldots, E_r$ be as in the previous lemma. Denote $e \cap F_i$ by e_i. Let every edge $e_i \in E_i$ be endowed with a nonnegative weight $w_i(e_i)$. Then*

$$\left(\sum_{e \in E} \prod_{i=1}^{r} w_i(e_i) \right)^t \le \prod_i \sum_{e_i \in E_i} w_i(e_i)^t.$$

Of course, setting all weights equal to 1 gives Shearer's Lemma.

Proof: Clearly we may assume all weights are positive integers. For simplicity of notation assume $V = \{1, \ldots, n\}$. Define a new multihypergraph $H' = (V, E')$ by creating $\prod_i w_i(e_i)$ copies, $\{e^{(c_1,\ldots,c_r)}\}$ of each edge e, where $1 \le c_i \le w(e_i)$ for $1 \le i \le r$. Similarly, for every $1 \le i \le r$ define H_i' by creating $w_i(e_i)$ copies of every edge e_i.

Let e be an edge of H' chosen uniformly at random from all edges. Let

$$Y = Y(e) = (X, C) = (x_1, \ldots, x_n, c_1, \ldots, c_r),$$

where $X = (x_1, \ldots, x_n)$ is the characteristic vector of e (i.e., $x_k = 1$ if $k \in e$ and 0 otherwise) and $C = (c_1, \ldots, c_r)$ gives the index of the copy of the given edge. Note that $(c_1|X), \ldots, (c_r|X)$ are mutually independent. Y is uniformly distributed on $\left(\sum_{e\in E} \prod_i w_i(e_i)\right)$ values, hence

$$H(Y) = \log\left(\sum_{e\in E} \prod_i w_i(e_i)\right). \tag{11}$$

Let $X^i = (x_1^i, \ldots, x_n^i)$ be the characteristic vector of $e_i = e \cap F_i$. Note that this vector is derived from X by setting the coordinates not in F_i to 0, hence the variables $\{x_k^i : k \in F_i\}$ have the same joint distribution as the corresponding $\{x_k : k \in F_i\}$. For $1 \le i \le r$ let $c_i^1, \ldots, c_i^t$ be t independent random variables such that the joint distribution of (X, c_i^s) is the same as that of (X, c_i). Note that $1 \le c_i^s \le w_i(e_i)$ for $1 \le s \le t$. Let

$$Y^i = (X^i, C^i) = (X^i, c_i^1, \ldots, c_i^t).$$

Y^i corresponds to picking an edge e uniformly from H', observing its projection e_i, and then choosing with replacement t independent copies of e_i from the $w_i(e_i)$ possible copies. Note that Y^i can take on at most $\sum_{e_i\in E_i} w_i(e_i)^t$ different values, hence, for $1 \le i \le r$, by (6)

$$H(Y^i) \le \log\left(\sum_{e_i\in E_i} w_i(e_i)^t\right).$$

So to prove our result we must show that

$$tH(Y) \le \sum H(Y^i).$$

By (8)

$$H(Y) = \sum_{m=1}^{n} H(x_m|x_l : l < m) + \sum_{i=1}^{r} H(c_i|x_1, \ldots, x_n).$$

Similarly

$$H(Y^i) = \sum_{m \in F_i} H(x_m | x_l : l < m, l \in F_i) + tH(c_i | x_l : l \in F_i).$$

Here we have used the fact that $(c_i^s | X)$ has the distribution of $(c_i | X)$ and depends only on $\{x_l : l \in F_i\}$. So,

$$\sum H(Y^i) - tH(Y) =$$

$$\sum_{m=1}^{n} \Bigg(\sum_{i: m \in F_i} (H(x_m | x_l : l < m, l \in F_i)) - tH(x_m | x_l : l \leq m) \Bigg)$$

$$+ t \Bigg(\sum_{i=1}^{r} H(c_i | x_l, l \in F_i) - H(c_i | x_1, \ldots, x_n) \Bigg).$$

Finally, using the fact that every m belongs to at least t of the F_i's, and by (10) all summands on the right-hand side above are positive, therefore $tH(Y) \leq \sum H(Y^i)$ as required. ∎

It turns out that this lemma, when applied to various specifically chosen hypergraphs, yields some interesting inequalities, such as Cauchy-Schwarz, Hölder, the Arithmetic-Geometric mean inequality, and others. More about this will appear in a paper in preparation [8].

4. The Proof of Theorem 1.3

Let us begin with some notation. Let $\mathcal{M}$ be a family of subsets of $\{1, \ldots, n\}$, all of size m. Let $Q = (I, J, K, L)$ be an ordered quadruplet of sets. We will say the quadruplet is **even** if $I \Delta J \Delta K \Delta L = \emptyset$. Note that for an even quadruplet the Venn diagram of all four sets is determined by the Venn diagram of three of them, say, I, J, K. In what follows the role played by I, J, K and L is symmetric even if we use only I, J and K for indexing purposes, letting L be determined by the parity condition. Let P be a partition of our ground set $\{1, \ldots, n\}$ into seven parts denoted by $P_I, P_J, P_K, P_{IJ}, P_{IK}, P_{JK}, P_{IJK}$. We will say an even quadruplet $Q = (I, J, K, L)$ is **consistent** with P if the corresponding elements of the sets belong to the "correct" parts of P, e.g., $I \cap J \cap K \subseteq P_{IJK}$, $(I \cap J) \setminus K \subseteq P_{IJ}$, etc. Let

$$\mathcal{Q}_P = \{Q \colon Q \text{ is even and consistent with } P\}.$$

Let $\mathcal{Q} = \bigcup \mathcal{Q}_P$.

Since the symbols I, J, K, L are now used as indices we shift to M, N, R, T for concrete sets. We will say a set M is I-consistent with P if there exist sets N, R, T such that $(M, N, R, T) \in \mathcal{Q}_P$. Let

$$\mathcal{M}_P^I = \{M \in \mathcal{M}\colon M \text{ is } I\text{-consistent with } P\}.$$

Define $\mathcal{M}_P^J, \mathcal{M}_P^K, \mathcal{M}_P^L$ similarly. Let

$$\mathcal{M}_P = \mathcal{M}_P^I \cup \mathcal{M}_P^J \cup \mathcal{M}_P^K \cup \mathcal{M}_P^L.$$

Recall now that we wish to show

$$|f_{\hat{m}}|_4 \leq c^m |f_{\hat{m}}|_2$$

with $c = \sqrt[4]{28}$. Using (4) and (5) this is equivalent to the following:

LEMMA 4.1: *Let $\mathcal{M}$ be a family of subsets of $\{1, \ldots, n\}$, each of size exactly m. For every set $M \in \mathcal{M}$ let f_M be a weight associated with it. Then*

$$\sum_{(M,N,R,T)\in\mathcal{Q}} f_M f_N f_R f_T \leq (28)^m \Big(\sum_{M\in\mathcal{M}} f_M^2 \Big)^2. \tag{13}$$

This formulation indeed suggests using Lemma 3.2, but in order to use the Lemma we must first fix a partition P and restrict ourselves to the consistent sets and quadruplets.

LEMMA 4.2: *Let*

$$P = (P_I, P_J, P_K, P_{IJ}, P_{IK}, P_{JK}, P_{IJK})$$

be a partition of $\{1, \ldots, n\}$. Then

$$\sum_{(M,N,R,T)\in\mathcal{Q}_P} f_M f_N f_R f_T \leq \frac{1}{16} \Big(\sum_{M\in\mathcal{M}_P} f_M^2 \Big)^2.$$

Proof: We wish to invoke Lemma 3.2. The hypergraph $H = (V, E)$ is simple to define:

$$V(H) = \{1, \ldots, n\},$$
$$E(H) = \{e_{MNRT} = M \cup N \cup R \cup T : (M, N, R, T) \in \mathcal{Q}_P\}.$$

Note that the consistency with P ensures that if $(M, N, R, T) \neq (M', N', R', T')$ then $e_{MNRT} \neq e_{M'N'R'T'}$. Define the four subsets F_I, F_J, F_K, F_L in the obvious way:

$$\begin{aligned} F_I &= P_I \cup P_{IJ} \cup P_{IK} \cup P_{IJK}, \\ F_J &= P_J \cup P_{IJ} \cup P_{JK} \cup P_{IJK}, \\ F_K &= P_K \cup P_{IK} \cup P_{JK} \cup P_{IJK}, \\ F_L &= P_I \cup P_J \cup P_K \cup P_{IJK}. \end{aligned}$$

Note that every $m \in \{1, \dots, n\}$ is covered by at least two of these subsets. Recall that $e_I = e \cap F_I$, etc. Note that $(e_{MNRT})_I = M$, $(e_{MNRT})_J = N$, etc. Naturally we define $w_I((e_{MNRT})_I) = w_I(M) = f_M$, etc. We now can invoke Lemma 3.2 with $t = 2$. This gives

$$\Bigg(\sum_{(M,N,R,T)\in\mathcal{Q}_P} f_M f_N f_R f_T\Bigg)^2 \le$$
$$\Bigg(\sum_{M\in\mathcal{M}_P^I} f_M^2\Bigg)\Bigg(\sum_{M\in\mathcal{M}_P^J} f_M^2\Bigg)\Bigg(\sum_{M\in\mathcal{M}_P^K} f_M^2\Bigg)\Bigg(\sum_{M\in\mathcal{M}_P^L} f_M^2\Bigg).$$

By Jenssen's inequality the left-hand side of this inequality is less than $\frac{1}{256}\left(\sum_{M\in\mathcal{M}_P} f_M^2\right)^4$, which proves Lemma 4.2. ∎

Lemma 4.2 dealt with a fixed partition P. We now deduce Lemma 4.1 by an averaging argument.

LEMMA 4.3: *Let $P_{rand} = (P_I, P_J, P_K, P_{IJ}, P_{IK}, P_{JK}, P_{IJK})$ be a random partition of $\{1, \dots, n\}$, chosen uniformly from all partitions. Then*

$$\mathbf{E}\Bigg(\Bigg(\sum_{M\in\mathcal{M}_{P_{rand}}} f_M^2\Bigg)^2\Bigg) \le 4(4/7)^m\Bigg(\sum_{M\in\mathcal{M}} f_M^2\Bigg)^2 \tag{14}$$

and

$$7^{-2m}\sum_{(M,N,R,T)\in\mathcal{Q}} f_M f_N f_R f_T \le \mathbf{E}\Bigg(\sum_{(M,N,R,T)\in\mathcal{Q}_{P_{rand}}} f_M f_N f_R f_T\Bigg). \tag{15}$$

Lemma 4.2, (14) and (15) now give

$$\sum_{(M,N,R,T)\in\mathcal{Q}} f_M f_N f_R f_T \le \frac{1}{4}(28)^m\Bigg(\sum_{M\in\mathcal{M}} f_M^2\Bigg)^2,$$

completing the proof of Theorem 1.3 with a factor of 1/4 to spare. ∎

Proof of Lemma 4.3: The model we have in mind for picking the random partition P is that of making an independent choice for every $k \in \{1, \ldots, n\}$ of which of the seven subsets of the partition it belongs to. Given a set $M \in \mathcal{M}$, the probability that it is consistent with the random partition is at most $4(4/7)^m$. The factor of 4 comes from choosing whether $M \in \mathcal{M}_P^I, \mathcal{M}_P^J, M_P^K$ or $\mathcal{M}_P^L$. The 4^m comes from the different ways to partition M into the four parts. The factor of $(1/7)^m$ comes from the probability that every element in M falls into the "correct" part of the partition. We now write

$$\left(\sum_{M\in\mathcal{M}} f_M^2\right)^2 = \sum_{M,N\in\mathcal{M}} f_M^2 f_N^2.$$

Bounding the probability that both M and N are consistent with P by the probability that one of them is consistent (a bound which is exact if $M = N$), and using linearity of expectation, we get (14).

The bound in (15) is much the same. For a given (M, N, R, T) the probability of being consistent with P is exactly $7^{-|M\cup N\cup R\cup T|}$. Since every point in this union is covered at least twice, $|M \cup N \cup R \cup T| \leq 2m$. ∎

Remark: The choice of the 4-norm in this proof was somewhat arbitrary; a similar proof works for any *even* integer larger than 2.

5. On the number of copies of one hypergraph in another

In this section we want to point out the connection between the present paper and paper [9], which carries the same name as this section. In that paper the following theorem is proven:

THEOREM 5.1: *Let H and G be hypergraphs, and let $|E(G)| \leq l$. Then the number of copies of H in G is at most l^{ρ^*}, where ρ^* is the fractional covering number of H.*

The paper presents two different proofs of this fact, one via Shearer's Entropy Lemma and the other (which gives a weaker result) using the Beckner–Bonami inequality.

One way of stating Theorem 1.2 very roughly is that for any $q > 2$ the q-norm of $f_{\hat{m}}$ is comparable to the 2-norm. In the present paper, concentrating on the case $q = 4$ we reduce this statement to a setting very similar to that of Theorem 5.1. Essentially we have a hypergraph, albeit a weighted one, whose edges are the sets of size m. We are trying to bound the number of occurrences of a certain pattern, the even quadruplets, which are hypergraphs with four edges, which must intersect in a certain manner. It is not hard to see that $\rho^* = 2$ for these patterns (except for the trivial pattern consisting of four copies of the same edge). This value of ρ^* explains the power 2 appearing in the right-hand side of (13), which luckily coincides with the expression for the 2-norm of f.

The bound in (13) also contains the exponential factor c^m. This factor is the price paid when going from the general hypergraph to a 7-partite subhypergraph corresponding to a given partition. Readers familiar with [9] may notice that there too one must restrict to a mutipartite hypergraph before applying the entropy technique.

We hope the above remarks have helped to shed some light on the contents of this paper.

References

[1] W. Beckner, *Inequalities in Fourier analysis,* Annals of Mathematics **102** (1975), 139–182.

[2] A. Bonami, *Etude des coefficients Fourier des fonctiones de $L^p(G)$,* Annales de l'Institut Fourier (Grenoble) **20** (1970), 335–402.

[3] J. Bourgain, *An appendix to Sharp thresholds of graph properties, and the k-sat problem by E. Friedgut,* Journal of the American Mathematical Society **12** (1999), 1017–1054.

[4] J. Bourgain and G. Kalai, *Influences of variables and threshold intervals under group symmetries,* Geometric and Functional Analysis **7** (1997), 438–461.

[5] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson and N. Linial, *The influence of variables in product spaces,* Israel Journal of Mathematics **77** (1992), 55–64.

[6] F. R. K. Chung, P. Frankl, R. L. Graham and J. B. Shearer, *Some intersection theorems for ordered sets and graphs,* Journal of Combinatorial Theory **43** (1986), 23–37.

[7] M. M. Cover and J. A. Thomas, *Elements of Information Theory,* Wiley, New York, 1991.

[8] E. Friedgut, *Hypergraphs, entropy and inequalities,* in preparation.

[9] E. Friedgut and J. Kahn, *On the number of copies of one hypergraph in another,* Israel Journal of Mathematics **105** (1998), 251–256.

[10] R. F. Gundy, *Some Topics in Probability and Analysis,* American Mathematical Society, Providence, RI, 1989.

[11] J. Kahn, G. Kalai and N. Linial, *The influence of variables on Boolean functions,* in *Proceedings of the 29-th Annual Symposium on Foundations of Computer Science,* Computer Society Press, White Planes, 1988, pp. 68–80.

[12] J. Radhakrishnan, private communication.

[13] M. Talagrand, *On Russo's approximate 0-1 law*, The Annals of Probability **22** (1994), 1376–1387.