Lower Bound on the Correlation Between Monotone Families in the Average Case

Nathan Keller
Einstein Institute of Mathematics, Hebrew University
Jerusalem 91904, Israel
nkeller@math.huji.ac.il

February 21, 2008

Abstract

A well-known inequality due to Harris and Kleitman [4, 9] states that any two monotone subsets of $\{0,1\}^n$ are non-negatively correlated with respect to the uniform measure on $\{0,1\}^n$. In [14], Talagrand established a lower bound on the correlation in terms of how much the two sets depend simultaneously on the same coordinates. In this paper we show that when the correlation is averaged over all the pairs $A, B \in T$ for any family T of monotone subsets of $\{0,1\}^n$, the lower bound asserted in [14] can be improved, and more precise estimates on the average correlation can be given. Furthermore, we generalize our results to the correlation between monotone functions on $[0,1]^n$ with respect to the Lebesgue measure.

1 Introduction

Correlation inequalities between monotone functions play an important role in numerous areas, including probability, combinatorics, mathematical physics, etc. In this paper we consider monotone functions defined on the discrete cube $\{0,1\}^n$ endowed with the uniform measure μ, and especially Boolean functions that can be treated as characteristic functions of subsets of the discrete cube.

Definition 1 *A function $f : \{0,1\}^n \rightarrow \mathbb{R}$ is monotone if for all $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$,*

$$(\forall i : x_i \leq y_i) \Rightarrow (f(x) \leq f(y)).$$

A subset $A \subset \{0,1\}^n$ is called monotone if its characteristic function is monotone.

One of the first correlation inequalities established for such functions is the following inequality, due to Harris [4] and Kleitman [9]:

Theorem 2 (Harris, Kleitman) *Let A, B be monotone subsets of $\{0,1\}^n$ endowed with the uniform measure μ. Then*

$$\mu(A \cap B) \geq \mu(A)\mu(B), \tag{1}$$

i.e., the correlation of A and B is nonnegative.

Clearly, the inequality in Theorem 2 is tight, since the correlation between independent monotone subsets of the discrete cube is zero. However, if A and B are dependent, the inequality is not tight, and hence it seems possible that one can obtain a lower bound on the correlation in terms of the dependence between A and B. Such bound was indeed established by Talagrand [14], where the measure of dependence is how much the two sets depend simultaneously on the same coordinates.

Definition 3 *Let $A \subset \{0,1\}^n$ be monotone. For all $1 \leq i \leq n$, define*

$$A_i = \{(x_1, \ldots, x_n) \in A : (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n) \notin A\}.$$

The influence of the i-th coordinate on A is $\mu(A_i)$.

Theorem 4 (Talagrand) *Let A, B be monotone subsets of $\{0,1\}^n$ endowed with the uniform measure μ. Then*

$$\mu(A \cap B) - \mu(A)\mu(B) \geq K\varphi(\sum_{i \leq n} \mu(A_i)\mu(B_i)), \tag{2}$$

where $\varphi(x) = x/log(e/x)$ and K is a universal constant.

Whereas the term $\sum_{i \leq n} \mu(A_i)\mu(B_i)$ seems a natural measure of the dependence between A and B, the *log* factor seems unnatural. However, it was shown in [14] by calculating the correlation between the sets $A = \{x : \sum_{i \leq n} x_i \geq t\}$ and $B = \{x : \sum_{i \leq n} x_i > n - t\}$, that the *log* term cannot be removed in general.

We show that the *log* term can be removed in the "average case".

Theorem 5 *Let T be a family of monotone subsets of the discrete cube. Then*

$$\sum_{A,B \in T} (\mu(A \cap B) - \mu(A)\mu(B)) \geq \sum_{A,B \in T} \sum_{i \leq n} \mu(A_i)\mu(B_i). \tag{3}$$

Unlike the proof of Talagrand's result [14], the proof of Theorem 5 is very simple and uses only the basic properties of the Fourier-Walsh expansion of functions on the discrete cube.

We generalize Theorem 5 to non-Boolean functions defined on the continuous cube $[0,1]^n$ endowed with the Lebesgue measure. Unlike the discrete case, in the continuous case there is no single natural definition of the influences, and at least three different definitions were proposed in previous papers [2, 5, 12]. We show that for the definition presented in [5, 12], Theorem 5 can be generalized to the continuous case, where the generalization of the influences is the first-level Fourier coefficients with respect to the shifted Legendre polynomials ([8], p. 121).

Theorem 6 *Let T be a family of monotone functions on the continuous cube $[0,1]^n$ endowed with the Lebesgue measure λ. Then*

$$\sum_{f,g \in T} (\int fg d\lambda - \int f d\lambda \int g d\lambda) \geq \sum_{f,g \in T} \sum_{i \leq n} \hat{f}(\{i\})\hat{g}(\{i\}), \tag{4}$$

where $\hat{f}(\{i\}) = \int f r_i d\lambda$ and $r_i(x_1, \ldots, x_n) = \sqrt{3}(2x_i - 1)$ are the first degree shifted Legendre polynomials on $[0,1]$.

The paper is organized as follows: In Section 2 we recall some preliminaries related to the Fourier-Walsh expansion of functions on the discrete cube, and present the proof of Theorem 5. In Section 3 we present the generalization of our results to non-Boolean functions on the continuous cube. In Section 4 we show that Theorem 4 is tight even for the correlation of symmetric monotone subsets of the discrete cube, and discuss the tightness of Theorem 5. Finally, in the appendix we present a direct (though, more complicated) proof of Theorem 5 using an inductive approach.

2 Fourier-Walsh Expansion of Functions on the Discrete Cube and Proof of Theorem 5

Consider the discrete cube $\{0,1\}^n$ endowed with the uniform measure μ. Denote the set of all real-valued functions on the discrete cube by Y. The inner product of functions $f, g \in Y$ is defined as usual as

$$\langle f, g\rangle = \int fg d\mu = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x).$$

This inner product induces a norm on Y:

$$||f||_2 = \sqrt{\langle f, f\rangle} = \sqrt{\int f^2 d\mu}.$$

Consider the Rademacher functions $\{r_i\}_{i=1}^n$, defined as:

$$r_i(x_1, \ldots, x_n) = 2x_i - 1.$$

These functions constitute an orthonormal system in Y. Moreover, this system can be completed to an orthonormal basis in Y by defining

$$r_S = \prod_{i \in S} r_i$$

for all $S \subset \{1, \ldots, n\}$. Every function $f \in Y$ can be represented by its Fourier expansion with respect to the system $\{r_S\}_{S \subset \{1,\ldots,n\}}$:

$$f = \sum_{S \subset \{1,\ldots,n\}} \langle f, r_S\rangle r_S.$$

The coefficients in this expansion are denoted

$$\hat{f}(S) = \langle f, r_S\rangle.$$

By the Parseval identity, for all $f \in Y$ we have

$$\sum_{S \subset \{1,\ldots,n\}} \hat{f}(S)^2 = ||f||_2^2.$$

More generally, for all $f, g \in Y$,

$$\langle f, g\rangle = \sum_{S\subset\{1,\ldots,n\}} \hat{f}(S)\hat{g}(S).$$

Finally, we note that for all $f \in Y$,

$$\hat{f}(\emptyset) = \int (fr_{\emptyset})d\mu = \int (f\cdot 1)d\mu = \int f d\mu.$$

2.1 Proof of Theorem 5

Consider the function $F(x) = \sum_{A\in T} 1_A(x)$. Note that for all $A, B \in T$ we have

$$\mu(A\cap B) - \mu(A)\mu(B) = Exp(1_A 1_B) - Exp(1_A)Exp(1_B) = Cov(1_A, 1_B).$$

Hence,

$$\sum_{A,B\in T} (\mu(A\cap B) - \mu(A)\mu(B)) = \sum_{A,B\in T} Cov(1_A, 1_B) = Var(\sum_{A\in T} 1_A) = Var(F). \tag{5}$$

By the Parseval identity,

$$Var(F) = Exp(F^2) - (Exp(F)^2) = \sum_{S} \hat{F}(S)^2 - \hat{F}(\emptyset)^2 = \sum_{S\neq\emptyset} \hat{F}(S)^2,$$

where $\hat{f}(S)$ is the coefficient of r_S in the Fourier-Walsh expansion of F. Thus, in the left hand side of Inequality 3 we have

$$\sum_{A,B\in T} (\mu(A\cap B) - \mu(A)\mu(B)) = \sum_{S\neq\emptyset} \hat{F}(S)^2.$$

We turn now to the right hand side.

$$\sum_{A,B\in T}\sum_{i\leq n} \mu(A_i)\mu(B_i) = \sum_{i\leq n}\sum_{A,B\in T} \mu(A_i)\mu(B_i) = \sum_{i\leq n}(\sum_{A\in T}\mu(A_i))^2.$$

Note that for a monotone subset A of the discrete cube and for all $1 \leq i \leq n$,

$$\hat{1}_A(\{i\}) = \mu(A_i).$$

Thus, by the linearity of the Fourier transform,

$$(\sum_{A\in T}\mu(A_i))^2 = (\sum_{A\in T}\hat{1}_A(\{i\}))^2 = \hat{F}(\{i\})^2.$$

Hence,

$$\sum_{A,B\in T}\sum_{i\leq n} \mu(A_i)\mu(B_i) = \sum_{i\leq n}\hat{F}(\{i\})^2 \leq \sum_{S\neq\emptyset}\hat{F}(S)^2 = \sum_{A,B\in T}(\mu(A\cap B) - \mu(A)\mu(B)). \tag{6}$$

This completes the proof of Theorem 5.

3 Generalization to non-Boolean Functions on the Continuous Cube

In this section we consider non-Boolean functions defined on a probability space X with measure μ. In this case, the correlation between two functions f and g is represented by the covariance:

$$Cov(f,g) = \int fg d\mu - \int f d\mu \int g d\mu.$$

We start with a lemma formulated in a more general setting:

Lemma 7 *Let H be a Hilbert space and let U be an orthonormal system in H. Then for every family $T \subset H$,*

$$\sum_{f,g \in T} (\langle f, g \rangle - \sum_{u \in U} \hat{f}(u)\hat{g}(u)) \geq 0,$$

where the Fourier coefficients are with respect to the system U.

Proof First, we note that by Zorn's lemma, the system U can be completed into a complete orthonormal system in H ([11], Corollary 13.6.1). Furthermore, any complete orthonormal system in a Hilbert space is an orthonormal basis ([11], Theorem 13.6.5.), and hence U can be completed into an orthonormal basis of H. Denote this basis by V. Every element $f \in H$ can be represented in the form

$$f = \sum_{v \in V} \hat{f}(v) v,$$

where $\hat{f}(v) = \langle f, v \rangle$. By the Parseval identity, for all $f, g \in H$,

$$\langle f, g \rangle = \sum_{v \in V} \hat{f}(v)\hat{g}(v).$$

Clearly, $U \subset V$. Hence, for all $f, g \in H$,

$$\langle f, g \rangle - \sum_{u \in U} \hat{f}(u)\hat{g}(u) = \sum_{v \in (V \setminus U)} \hat{f}(v)\hat{g}(v).$$

Therefore, for every family $T \subset H$, we have

$$\sum_{f,g \in T} (\langle f, g \rangle - \sum_{u \in U} \hat{f}(u)\hat{g}(u)) = \sum_{f,g \in T} \sum_{v \in (V \setminus U)} \hat{f}(v)\hat{g}(v) =$$

$$= \sum_{v \in (V \setminus U)} \sum_{f,g \in T} \hat{f}(v)\hat{g}(v) = \sum_{v \in (V \setminus U)} (\sum_{f \in T} \hat{f}(v))^2 \geq 0,$$

as asserted. ■

In order to establish a generalization of Theorem 5 to functions defined on $[0,1]$, we first need to find the appropriate generalization of the influences to the continuous case.

The most common definition is the following, introduced in [2]:

Definition 8 *Let $f : [0,1]^n \to \{0,1\}$ be a measurable function. For all $x \in [0,1]^n$, the fiber of x in the k-th direction is $l_k(x) = \{y \in [0,1]^n : y_j = x_j, \forall j \neq k\}$. Denote by $S_k(f)$ the set of all $x \in [0,1]^n$ for which f is non-constant on the set $l_k(x)$. The influence of the k-th coordinate on f is $I_f(k) = \lambda(S_k(f)))$.*

Neither Talagrand's Theorem 4 nor our Theorem 5 cannot be generalized to the continuous case under this definition of influence. This can be seen in the following example:
Example Let $f : [0,1]^n \to \{0,1\}$ be defined by

$$(f(x) = 1) \Longleftrightarrow (\forall i : x_i \geq 1/n).$$

Clearly, $Exp(f^2) = Exp(f) = (1 - 1/n)^n \approx 1/e$. The function f is non-constant on a fiber $l_k(x)$ if and only if $x_j \geq 1/n$ for all $j \neq k$. Hence, the influence of the k-th coordinate on f is $I_f(k) = (1 - 1/n)^{n-1} \approx 1/e$. Consider the correlation between f and itself. We have

$$Cov(f,f) = Exp(f^2) - (Exp(f)^2) \approx 1/e - 1/e^2.$$

On the other hand, the natural generalization of the term $\sum_{i \leq n} \mu(A_i)\mu(B_i)$ appearing in the right hand side of Inequality 2 is

$$\sum_{k \leq n} I_f(k) I_f(k) \approx n(1/e^2).$$

Therefore, the natural generalizations of Theorems 4 and 5 are far from being correct in these settings.

Another natural definition of the influences in the continuous case was introduced recently in [5, 12]:

Definition 9 *Let $f : [0,1]^n \to \{0,1\}$ be a measurable function. Denote by $f_k^x : [0,1] \to \{0,1\}$ the restriction of f to the fiber of x in the k-th direction. That is, $f_k^x(t) = f(x_1, \ldots, x_{k-1}, t, x_{k+1}, \ldots, x_n)$. The influence of the k-th coordinate on f is*

$$\tilde{I}_f(k) = Exp_x(Var(f_k^x)).$$

In some sense, this definition is more natural than the former one, since it is more sensitive to the behavior of f on each fiber, and not only checks whether f is constant on it.

It appears that under the second definition, there is a natural Fourier-theoretic realization of the influences. Consider the first degree shifted Legendre polynomials ([8], p. 121):

$$r_i'(x_1, \ldots, x_n) = 2x_i - 1,$$

for $x \in [0,1]^n$. Since

$$\int_0^1 (2x - 1) dx = 0,$$

the functions $\{r_i'\}_{i=1}^n$ are orthogonal. By normalizing the functions, we get the orthonormal system $\{r_i\}_{i=1}^n$, where

$$r_i(x_1, \ldots, x_n) = \sqrt{3}(2x_i - 1).$$

The Fourier coefficients with respect to this system are a natural generalization of the influences, up to multiplication by a constant. Indeed, if f is Boolean and monotone, then on each fiber there exists t_0 such that $f_k^x(t) = 0$ for all $t < t_0$, and $f_k^x(t) = 1$ for all $t > t_0$. In this case, the variance of f_k^x is $t_0(1-t_0)$. On the other hand, we have

$$\int_{t=0}^{1} (fr'_k)(x_1, \ldots, x_{k-1}, t, x_{k+1}, \ldots, x_n)dt = \int_{t=0}^{1} f_k^x(t)(2t-1)dt = \int_{t=t_0}^{1} (2t-1)dt = t_0(1-t_0).$$

Hence, by the Fubini theorem

$$\hat{f}(\{k\}) = \sqrt{3}\int_{x\in[0,1]^n} fr'_k d\lambda = \sqrt{3}\int_{x\in[0,1]^{n-1}}\int_{t=0}^{1} (f_k^x r'_k)(t)dtd\lambda = \sqrt{3}Exp_x(Var(f_k^x)),$$

where $(x \in [0,1]^{n-1})$ means that the k-th coordinate of x is neglected. Therefore, up to the normalization constant, the Fourier coefficients with respect to the system $\{r_i\}_{i\leq n}$ are equal to the influences for monotone Boolean functions. Since the Fourier coefficients are defined in the same way for non-Boolean functions, they can be considered a natural generalization of the influences to general functions on the continuous cube.

After finding the appropriate orthonormal basis, Theorem 6 follows immediately from Lemma 7. Indeed, we apply the lemma to the space of all real-valued functions on the continuous cube with the inner product

$$\langle f, g\rangle = \int fgd\lambda$$

and the orthonormal system $U = \{\emptyset, r_1, \ldots, r_n\}$ and get

$$\sum_{f,g\in T}\left(\int fgd\lambda - \int fd\lambda\int gd\lambda - \sum_{i\leq n}\hat{f}(\{i\})\hat{g}(\{i\})\right) = \sum_{f,g\in T}\left(\langle f,g\rangle - \hat{f}(\emptyset)\hat{g}(\emptyset) - \sum_{i\leq n}\hat{f}(\{i\})\hat{g}(\{i\})\right) \geq 0,$$

as asserted. This completes the proof of Theorem 6.

Remark A more complicated definition of the influences in the continuous case was presented in [2]. The definition is based on discretizing the function and measuring the (discrete) influences of the new coordinates. We don't know whether a generalization of Theorem 5 holds under this definition.

We conclude this section with a remark about subsets of the discrete cube endowed with the product measure μ_p defined by

$$\mu_p(x) = p^{\sum_{i\leq n} x_i}(1-p)^{n-\sum_{i\leq n} x_i}.$$

In this case, Definition 3 is still the natural definition of influence. In order to get a generalization of Theorem 5 to this setting, we can replace the subsets in a standard way by functions defined on $[0,1]^n$ and use Theorem 6. The resulting formula is

$$\sum_{A,B\in T}(\mu_p(A\cap B) - \mu_p(A)\mu_p(B)) \geq 3(1-p)^2\sum_{A,B\in T}\sum_{i\leq n}\mu_p(A_i)\mu_p(B_i). \tag{7}$$

A stronger result can be achieved by using an orthonormal basis of functions defined on the discrete cube with the measure μ_p. This basis was probably first presented in [13]. Let

$$s_i(x_1,\ldots,x_n) = \begin{cases} \sqrt{\frac{1-p}{p}}, & x_i = 1 \\ -\sqrt{\frac{p}{1-p}}, & x_i = 0. \end{cases}$$

These functions can be completed into an orthonormal basis by defining

$$s_T = \prod_{i \in T} s_i,$$

for all $T \subset \{1,\ldots,n\}$, and $s_\emptyset \equiv 1$. Applying Lemma 7 to the space of real-valued functions on the discrete cube with the inner product

$$\langle f, g \rangle = \int fg d\mu_p$$

and the orthonormal system $U = \{s_T\}_{T \subset \{1,\ldots,n\}}$ we get

Proposition 10 *Let T be a family of monotone subsets of the discrete cube endowed with the product measure μ_p. Then*

$$\sum_{A,B \in T} (\mu(A \cap B) - \mu(A)\mu(B)) \geq \frac{1-p}{p} \sum_{A,B \in T} \sum_{i \leq n} \mu(A_i)\mu(B_i). \tag{8}$$

Note that for the uniform measure (i.e., $p = 1/2$), Proposition 10 is identical to Theorem 5, while Inequality 7 yields a weaker result.

4 Tightness of Results

We conclude this paper with several remarks regarding the tightness of Talagrand's results [14] and our results.

4.1 Symmetric Subsets of the Discrete Cube

Definition 11 *A set $A \subset \{0,1\}^n$ is symmetric if it is invariant under a transitive permutation group Γ on $\{1,\ldots,n\}$.*

Clearly, if A is symmetric then the influences of all the coordinates on A are equal. Hence, by the KKL theorem [7], if A is also balanced (i.e., $\mu(A) = 1/2$) then all the influences are $\Omega(\log n/n)$. Therefore, by Theorem 4, we get:

Proposition 12 *Let A and B be balanced, monotone, and symmetric subsets of the discrete cube (endowed with the uniform measure). Then*

$$\mu(A \cap B) - \mu(A)\mu(B) \geq K \log n/n,$$

where K is a universal constant.

The assertion of Proposition 12 is tight, as can be seen in the following example, based on the "tribes" function presented in [1].

Example Let $r \approx \log n - \log\log n$. Subdivide the set $\{1,\ldots,n\}$ into n/r disjoint sets $\{S_j\}$ of size r. The set A is defined as follows:

$$(x \in A) \Longleftrightarrow (\exists j : x_i = 1, \forall i \in S_j).$$

The set B is the dual of A, that is,

$$(x \in B) \Longleftrightarrow (\forall j, \exists i \in S_j : x_i = 1).$$

We note that while for $r = \log n - \log\log n$, we have $\mu(A) \approx 1 - 1/e$, r can be easily modified such that $\mu(A) = 1/2$. Hence, we assume that $\mu(A) = 1/2$, and thus $\mu(B) = 1/2$ (since B is the dual of A). Let us compute $\mu(A \cap B) - \mu(A)\mu(B)$. We have

$$\mu(B \setminus A) = (1 - 2 \cdot 2^{-r})^{n/r},$$

and hence

$$\mu(A \cap B) - \mu(A)\mu(B) = \mu(B) - \mu(B \setminus A) - \mu(B)(1 - \mu(B)) = \mu(B)^2 - (1 - 2 \cdot 2^{-r})^{n/r}.$$

Since $\mu(B)^2 = (1 - 2 \cdot 2^{-r} + 2^{-2r})^{n/r} = 1/4$, we get

$$\mu(A \cap B) - \mu(A)\mu(B) = \frac{1}{4}(1 - (1 - \frac{2^{-2r}}{1 - 2 \cdot 2^{-r} + 2^{2r}})^{n/r}) \approx \frac{1}{4}(1 - (1 - 2^{-2r})^{n/r})$$
$$\approx \frac{1}{4}(1 - (\frac{\log^2 n}{n^2})^{n/\log n}) \approx \frac{1}{4}\log n/n.$$

Hence, the assertion of Proposition 12 is tight in this case.

By Theorem 5 applied to the family of all balanced "tribes" functions, we get:

Proposition 13 *The expected correlation between two balanced tribes functions is* $\Theta(\log^2(n)/n)$.

The lower bound on the correlation between two balanced tribes functions asserted by Proposition 12 is $\Theta(\log n/n)$. It seems interesting to find out whether this lower bound is tight.

4.2 Tightness of Theorem 5

1. It is clear from Inequality 6 that the assertion of Theorem 5 is tight if and only if $\hat{F}(S) = \sum_{A \in T} \hat{1}_A(S) = 0$ for all $|S| > 1$, which is equivalent to the condition that F is linear. The same reasoning holds for Theorem 6. Hence, both theorems are tight if and only if the function F is linear. A simple example of this instance is when T consists only of linear functions. However, this is not the only possible example. For any linear function $L : \{0,1\}^n \to \{0,1,\ldots,M\}$, the inequality is tight for the family $T = \{A^k(L)\}_{k=1}^M$, where $A^k(L) = \{x \in \{0,1\}^n : L(x) \geq k\}$, since in this case $F(T) = L$ is linear. In this example, T consists of weighted majority functions. It seems interesting to further characterize the families for which Theorem 5 is tight.

2. If T is the family of all the monotone functions, the variance of $F(T)$ can be computed asymptotically using the asymptotic characterization of monotone Boolean functions obtained in [10] as part of the solution of Dedekind's problem [3]. As a result, we get:

   Proposition 14 *When $n \to \infty$, the expected correlation between two monotone Boolean functions on $\{0,1\}^n$ is $1/4 - o(1)$.*

   This result is interesting in view of the fact that by the Cauchy-Schwarz inequality, the maximal possible correlation between two monotone Boolean functions on the discrete cube is $1/4$.

References

[1] M. Ben-Or and N. Linial, Collective Coin Flipping, in *Randomness and Computation* (S. Micali, ed.), Academic Press, New York, 1990, pp. 91–115.

[2] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, and N. Linial, The Influence of Variables in Product Spaces, *Israel J. Math.*, **77** (1992), pp. 55–64.

[3] R. Dedekind, Über Zerlegungen von Zahlen Durch Ihre Grössten Gemeinsammen Teiler, *Gesammelte Werke*, Bd. 1, pp. 103–148, 1897.

[4] T.E. Harris, A Lower Bound for the Critical Probability in a Certain Percolation Process, *Proc. Cambridge Phil. Soc.* **56** (1960), pp. 13–20.

[5] H. Hatami, Influences and Decision Trees, submitted. Available online at arXiv:math/0612405v1.

[6] G.H. Hardy, J.E. Littlewood and G. Polya, Inequalities, Cambridge University Press, 1934.

[7] J. Kahn, G. Kalai, and N. Linial, The Influence of Variables on Boolean Functions, Proc. 29-th Ann. Symp. on Foundations of Comp. Sci., pp. 68–80, Computer Society Press, 1988.

[8] L.V. Kantorovich and G.P. Akilov, Functional Analysis, 2-nd Ed., Pergamon Press, 1982.

[9] D.J. Kleitman, Families of Non-Disjoint Subsets, *J. Combin. Theory* **1** (1966), pp. 153–155.

[10] A.D. Korshunov, Monotone Boolean Functions, *Russian Math. Surveys* **58** (2003), no. 5, pp. 929–1001.

[11] R. Larsen, Functional Analysis, Marcel Dekker inc., 1973.

[12] E. Mossel, R. O'Donnel, and K. Oleszkiewicz, Noise Stability of Functions with Low Influences: Invariance and Optimality, *Annals of Math.*, to appear.

[13] M. Talagrand, On Russo's Approximate 0-1 Law, *Ann. Prob.* **22** (1994), pp. 1576–1587.

[14] M. Talagrand, How Much are Increasing Sets Positively Correlated?, *Combinatorica* **16** (1996), no. 2, pp. 243–258.

A Inductive Proof of Theorem 5

We start with an inductive approach to the proof of the Harris-Kleitman theorem.

Definition 15 *Let $A \subseteq \{0,1\}^n$ be a monotone family. For every $1 \leq k \leq n-1$ and for every $\alpha \in \{0,1\}^{n-k}$, denote*

$$A_k^\alpha = \{x = (x_1, \ldots, x_n) \in A_k | (x_{k+1}, \ldots, x_n) = \alpha\},$$

where A_k is defined as in Section 1.

By the definition, the set A_k (consisting of the points for which the k-th coordinate has influence on A) is divided to 2^{n-k} sets, according to the last $n-k$ coordinates. Note that since A_k is a disjoint union of $\{A_k^\alpha\}_{\alpha \in \{0,1\}^{n-k}}$, we have

$$\mu(A_k) = \sum_{\alpha \in \{0,1\}^{n-k}} \mu(A_k^\alpha).$$

Lemma 16 *Let A, B be monotone subsets of the discrete cube (endowed with the uniform measure). Then*

$$\mu(A \cap B) - \mu(A)\mu(B) = \sum_{k=1}^{n} 2^{n-k} \sum_{\alpha \in \{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha), \tag{9}$$

where the term corresponding to $k = n$ in the right hand side is $\mu(A_n)\mu(B_n)$.

Proof The proof is by induction on n. For $n = 1$ the claim is reduced to

$$\mu(A \cap B) - \mu(A)\mu(B) = \mu(A_1)\mu(B_1),$$

and can be easily verified by checking all the possible pairs (A, B). Assume now that the claim holds for $n - 1$. Denote

$$A^0 = \{x \in \{0,1\}^{n-1} | (x, 0) \in A\},$$

and

$$A^1 = \{x \in \{0,1\}^{n-1} | (x, 1) \in A\}.$$

Note that since A is monotone, we have $A^0 \subset A^1$. Denote by μ' the measure induced by μ on $\{0,1\}^{n-1}$. It is clear that $\mu(A) = (\mu'(A^0) + \mu'(A^1))/2$, and similarly for B and for $A \cap B$. Hence,

$$\begin{aligned}
\mu(A \cap B) - \mu(A)\mu(B) &= \frac{1}{2}(\mu'(A^0 \cap B^0) + \mu'(A^1 \cap B^1)) - \frac{1}{4}(\mu'(A^0) + \mu'(A^1))(\mu'(B^0) + \mu'(B^1)) \\
&= \frac{1}{2}(\mu'(A^0 \cap B^0) - \mu'(A^0)\mu'(B^0)) + \frac{1}{2}(\mu'(A^1 \cap B^1) - \mu'(A^1)\mu'(B^1)) \\
&\quad + \frac{1}{4}(\mu'(A^0)\mu'(B^0) + \mu'(A^1)\mu'(B^1) - \mu'(A^0)\mu'(B^1) - \mu'(A^1)\mu'(B^0)) \\
&= \frac{1}{2}(\mu'(A^0 \cap B^0) - \mu'(A^0)\mu'(B^0)) + \frac{1}{2}(\mu'(A^1 \cap B^1) - \mu'(A^1)\mu'(B^1)) \\
&\quad + \frac{1}{4}(\mu'(A^1) - \mu'(A^0))(\mu'(B^1) - \mu'(B_0)).
\end{aligned}$$

We note that

$$\mu'(A^1) - \mu'(A^0) = \mu'(A^1 \setminus A^0) = 2\mu(A_n),$$

and similarly

$$\mu'(B^1) - \mu'(B^0) = 2\mu(B_n).$$

Thus,

$$\mu(A\cap B)-\mu(A)\mu(B) = \frac{1}{2}(\mu'(A^0\cap B^0)-\mu'(A^0)\mu'(B^0))+\frac{1}{2}(\mu'(A^1\cap B^1)-\mu'(A^1)\mu'(B^1))+\mu(A_n)\mu(B_n). \tag{10}$$

By the induction assumption we have

$$\mu'(A^0 \cap B^0) - \mu'(A^0)\mu'(B^0) = \sum_{k=1}^{n-1} 2^{n-1-k} \sum_{\alpha\in\{0,1\}^{n-1-k}} \mu'((A^0)_k^\alpha)\mu'((B^0)_k^\alpha),$$

and similarly for A^1 and B^1. Note that $(A^0)_k^\alpha = A_k^{(\alpha,0)}$, where $(\alpha, 0)$ is the concatenation of the binary string α with 0 in the end, and similarly $(A^1)_k^\alpha = A_k^{(\alpha,1)}$. Hence,

$$\sum_{\alpha\in\{0,1\}^{n-1-k}} \mu'((A^0)_k^\alpha)\mu'((B^0)_k^\alpha) = 4 \sum_{\alpha\in\{0,1\}^{n-1-k}} \mu(A_k^{(\alpha,0)})\mu(B_k^{(\alpha,0)}),$$

and similarly

$$\sum_{\alpha\in\{0,1\}^{n-1-k}} \mu'((A^1)_k^\alpha)\mu'((B^1)_k^\alpha) = 4 \sum_{\alpha\in\{0,1\}^{n-1-k}} \mu(A_k^{(\alpha,1)})\mu(B_k^{(\alpha,1)}).$$

Since all the binary strings of length $n-k$ are either of the form $\{(\alpha,0) : \alpha \in \{0,1\}^{n-1-k}\}$ or $\{(\alpha,1) : \alpha \in \{0,1\}^{n-1-k}\}$, we get

$$\sum_{\alpha\in\{0,1\}^{n-1-k}} \mu'((A^0)_k^\alpha)\mu'((B^0)_k^\alpha) + \sum_{\alpha\in\{0,1\}^{n-1-k}} \mu'((A^1)_k^\alpha)\mu'((B^1)_k^\alpha) = 4 \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha).$$

Substituting into Equation 10 we obtain

$$\begin{aligned}
\mu(A \cap B) - \mu(A)\mu(B) &= \frac{1}{2}(\sum_{k=1}^{n-1} 2^{n-1-k} 4 \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha)) + \mu(A_n)\mu(B_n) \\
&= \sum_{k=1}^{n-1} 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha) + \mu(A_n)\mu(B_n) \\
&= \sum_{k=1}^{n} 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha),
\end{aligned}$$

as asserted. ■

Remark Note that since for all k and all $\alpha \in \{0,1\}^{n-k}$, we have $\mu(A_k^\alpha)\mu(B_k^\alpha) \geq 0$, Lemma 16 implies the Harris-Kleitman theorem.

In the proof of Theorem 5 we use the following form of the Cauchy-Schwarz inequality ([6], p. 16):

Proposition 17 *Let* $\{a_k\}_{k=1}^n$ *and* $\{b_k\}_{k=1}^n$ *be two sequences of real numbers. Then*

$$(\sum_{k=1}^n a_k b_k)^2 \leq (\sum_{k=1}^n a_k^2)(\sum_{k=1}^n b_k^2).$$

Now we are ready to present the proof of Theorem 5.
Let T be a family of monotone subsets of the discrete cube. By Lemma 16, it is sufficient to show that for every $1 \leq k \leq n$ we have

$$\sum_{A,B\in T} 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha) \geq \sum_{A,B\in T} \mu(A_k)\mu(B_k). \tag{11}$$

We start with the right hand side.

$$\sum_{A,B\in T} \mu(A_k)\mu(B_k) = \sum_{A\in T}(\mu(A_k)\sum_{B\in T}\mu(B_k)) = (\sum_{B\in T}\mu(B_k))(\sum_{A\in T}\mu(A_k)) = (\sum_{A\in T}\mu(A_k))^2.$$

Similarly, for the left hand side we have

$$\begin{aligned}\sum_{A,B\in T} 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha) &= 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} (\sum_{A,B\in T} \mu(A_k^\alpha)\mu(B_k^\alpha)) \\ &= 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} (\sum_{A\in T} \mu(A_k^\alpha))^2.\end{aligned}$$

Define a sequence $\{z_\alpha\}_{\alpha\in\{0,1\}^{n-k}}$ by

$$z_\alpha = \sum_{A\in T} \mu(A_k^\alpha).$$

Note that we have

$$\sum_{A\in T}\mu(A_k) = \sum_{A\in T}(\sum_{\alpha\in\{0,1\}^{n-k}}\mu(A_k^\alpha)) = \sum_{\alpha\in\{0,1\}^{n-k}}(\sum_{A\in T}\mu(A_k^\alpha)) = \sum_{\alpha\in\{0,1\}^{n-k}} z_\alpha.$$

Hence, Inequality 11 is equivalent to

$$(\sum_{\alpha\in\{0,1\}^{n-k}} z_\alpha)^2 \leq 2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} z_\alpha^2, \tag{12}$$

and this inequality is a direct application of the Cauchy-Schwarz inequality to the sequence $\{z_\alpha\}_{\alpha\in\{0,1\}^{n-k}}$ and the constant sequence. This completes the proof of Theorem 5.

A.1 Application to the Tightness of Talagrand's Results

Lemma 16 can be used to shed some light on the cases in which Talagrand's Theorem 4 is tight. It follows from Equation 9 that the tightness of Talagrand's theorem depends on the relation between the quantities

$$2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)\mu(B_k^\alpha)$$

and $\mu(A_k)\mu(B_k)$, for all $1 \leq k \leq n$. For a fixed k, consider the sequences $\{x_\alpha\}_{\alpha\in\{0,1\}^{n-k}}$ and $\{y_\alpha\}_{\alpha\in\{0,1\}^{n-k}}$ defined by

$$x_\alpha = \mu(A_k^\alpha), \qquad y_\alpha = \mu(B_k^\alpha)$$

for all $\alpha \in \{0,1\}^{n-k}$. Since

$$\mu(A_k) = \sum_{\alpha\in\{0,1\}^{n-k}} \mu(A_k^\alpha)$$

and similarly for B_k, we are interested in the relation between the quantities

$$2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} x_\alpha y_\alpha$$

and

$$(\sum_{\alpha\in\{0,1\}^{n-k}} x_\alpha)(\sum_{\alpha\in\{0,1\}^{n-k}} y_\alpha).$$

This relation is connected to the Rearrangement inequality ([6], p. 261) and in general depends on whether the elements of $\{x_\alpha\}_{\alpha\in\{0,1\}^{n-k}}$ and $\{y_\alpha\}_{\alpha\in\{0,1\}^{n-k}}$ are arranged in the same order. More precisely, if the sequences $\{x_\alpha\}$ and $\{y_\alpha\}$ are fixed except for the order, the expression

$$2^{n-k} \sum_{\alpha\in\{0,1\}^{n-k}} x_\alpha y_\alpha$$

assumes its maximal possible value when the sequences are arranged in the same order, and assumes its minimal value when the sequences are arranged in opposite orders. The expression

$$(\sum_{\alpha\in\{0,1\}^{n-k}} x_\alpha)(\sum_{\alpha\in\{0,1\}^{n-k}} y_\alpha).$$

is the average over all possible orders of the former expression. In the example presented by Talagrand in [14],

$$A = \{(x_1, \ldots, x_n) : \sum_{i\leq n} x_i \geq t\}$$

and

$$B = \{(x_1, \ldots, x_n) : \sum_{i\leq n} x_i > n - t\}.$$

If $t = o(n)$, then for most of the values of k (more precisely, for all k such that $t \leq (n-k)/2$), the corresponding sequences are arranged in opposite order. Hence,

$$\mu(A \cap B) - \mu(A)\mu(B)$$

is relatively small and thus the inequality asserted by Theorem 4 is relatively tight.