Improved FKG Inequality for Product Measures on the Discrete Cube

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

September 4, 2008

Abstract

The basic instance of the FKG correlation inequality [3] is the Harris-Kleitman lemma [4, 6], stating that any two monotone subsets of $\{0,1\}^n$ are non-negatively correlated with respect to the uniform measure on $\{0,1\}^n$. In [9], Talagrand established a lower bound on the correlation in the Harris-Kleitman lemma in terms of how much the two sets depend simultaneously on the same coordinates. In this paper we generalize Talagrand's result to the correlation between monotone functions on the discrete cube endowed with a general product measure, thus establishing an improvement of the FKG inequality for product measures on the discrete cube.

1 Introduction

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.

Correlation inequalities between monotone functions on the discrete cube play an important role in numerous areas, including probability, combinatorics, mathematical physics, etc. One of the most well-known correlation inequalities is the FKG inequality [3]:[1]

Theorem 2 (Fortuin, Kasteleyn, and Ginibre) *Let $f, g : \{0,1\}^n \rightarrow \mathbb{R}$ be monotone functions, and let μ be a log-supermodular measure on $\{0,1\}^n$, i.e.,*

$$\mu(x)\mu(y) \leq \mu(x \cup y)\mu(x \cap y) \tag{1}$$

for all $x, y \in \{0,1\}^n$. Then

$$\int fg d\mu \geq \int f d\mu \int g d\mu. \tag{2}$$

In the partial case of monotone subsets of the discrete cube, the FKG inequality was preceded by the Harris-Kleitman lemma [4, 6]:

[1] The original formulation of the FKG inequality is for finite distributive lattices. Since in this paper we deal only with functions on the discrete cube, we present the inequality in this partial case.

Theorem 3 (Harris, Kleitman) *Let A, B be monotone subsets of $\{0,1\}^n$ endowed with the uniform measure μ. Then*

$$\mu(A \cap B) \geq \mu(A)\mu(B), \tag{3}$$

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

Clearly, the inequality in the Harris-Kleitman lemma is tight, since the correlation between independent monotone subsets of the discrete cube is zero. However, if A and B are not independent, 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 [9], where the measure of dependence is how much the two sets depend simultaneously on the same coordinates.

Definition 4 *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 5 (Talagrand) *Let A, B be monotone subsets of $\{0,1\}^n$ endowed with the uniform measure μ. Then*

$$\mu(A \cap B) - \mu(A)\mu(B) \geq K\varphi\Big(\sum_{i \leq n} \mu(A_i)\mu(B_i)\Big), \tag{4}$$

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

In this paper we generalize the result of Talagrand to monotone functions on the discrete cube endowed with any product measure $\mu = \mu_{p_1} \otimes \ldots \otimes \mu_{p_n}$. The influences of the coordinates are replaced by the first-level Fourier-Walsh coefficients with respect to an appropriate orthonormal basis:

Definition 6 *Consider the discrete cube $\{0,1\}^n$ endowed with the product measure $\mu = \mu_{p_1} \otimes \ldots \otimes \mu_{p_n}$, and the functions:*

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

Let $f : \{0,1\}^n \to \mathbb{R}$. The first-level Fourier-Walsh coefficients of f with respect to the measure μ are

$$\hat{f}(\{i\}) = \int f s_i d\mu,$$

for $1 \leq i \leq n$.

Our main result is the following:

Theorem 7 *Let f, g be monotone functions on the discrete cube endowed with the product measure $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$, that satisfy $|f(x)| \leq 1$ and $|g(x)| \leq 1$ for all $x \in \{0,1\}^n$. Then*

$$\int fg d\mu - \int f d\mu \int g d\mu \geq K \min_{i \leq n} H(p_i) \varphi\Big(\sum_{i \leq n} \hat{f}(\{i\})\hat{g}(\{i\})\Big), \tag{5}$$

where $\varphi(x) = \frac{x}{\log(e/x)}$, K is a universal constant, and $H(p) = -p\log p - (1-p)\log(1-p)$ is the entropy function.

Theorem 7 generalizes Talagrand's result in two directions. First, the measure is generalized from the uniform measure to any product measure on the discrete cube. Second, the examined functions are generalized from Boolean functions (which are equivalent to subsets of the discrete cube) to general functions. As a result, our result provides an improvement of the FKG inequality in the partial case of product measures on the discrete cube. We note that when μ is the uniform measure, then for every monotone subset A of the discrete cube and for all $1 \leq i \leq n$ we have $\hat{1}_A(\{i\}) = \mu(A_i)$, and hence Talagrand's result is indeed a partial case of Theorem 7. However, unlike Talagrand's result, we weren't able to prove the tightness of Theorem 7, and the best example we found "misses" the lower bound asserted by the theorem by a multiplicative factor of $1/p$.

The proof of Theorem 7 is based on the proof of Talagrand's result, but also uses two additional ingredients. The first is the Fourier-Walsh expansion of functions on the discrete cube endowed with a general product measure (developed in [8]) and its properties. The second is a Subgaussian inequality bounding the large deviations of a weighted sum of independent Bernoulli random variables. While in the uniform case discussed in [9], a simple bound using the Azuma martingale inequality is sufficient, in the general case we use a more precise estimate due to Bobkov, Houdré, and Tetali [1], adapted to our case. The other ingredients of the proof are the same used by Talagrand, but the proof is much more complicated due to the more general framework.

The paper is organized as follows: In Section 2 we present the two main tools used in the proof of Theorem 7. The proof of Theorem 7 is presented in Section 3. Finally, in Section 4 we study the tightness of the results.

2 Preliminaries

In this section we present the main tools used in the proof of Theorem 7.

2.1 Fourier-Walsh Expansion of Functions on the Discrete Cube Endowed with a General Product Measure

Consider the discrete cube $\{0,1\}^n$ endowed with a product measure $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$, i.e.,

$$\mu(x) = \mu\Big((x_1, \ldots, x_n)\Big) = \prod_{i=1}^{n} p_i^{x_i}(1-p_i)^{1-x_i}.$$

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

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

This inner product induces a norm on Y:

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

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

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

As was observed in [8], these functions constitute an orthonormal system in Y (with respect to the measure μ). Moreover, this system can be completed to an orthonormal basis in Y by defining

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

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

$$f = \sum_{T \subset \{1,\ldots,n\}} \langle f, s_T \rangle s_T.$$

The coefficients in this expansion are denoted

$$\hat{f}(T) = \langle f, s_T \rangle.$$

A coefficient $\hat{f}(T)$ is called *k-th level coefficient* if $|T| = k$. By the Parseval identity, for all $f \in Y$ we have

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

In the proof of Theorem 7 we use the following two simple claims concerning functions on the discrete cube with the product measure μ:

Claim 8 *Let $f : \{0,1\}^n \to \mathbb{R}$ be monotone. Then for all $1 \leq i \leq n$, $\hat{f}(\{i\}) \geq 0$.*

Proof Without loss of generality, we prove the claim for $i = n$. For all $(x_1, \ldots, x_{n-1}) \in \{0,1\}^{n-1}$, let

$$f^0(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, 0), \qquad f^1(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, 1).$$

Denote by μ' the measure induced by μ on $\{0,1\}^{n-1}$. Since f is monotone, we have $f^1(x) \geq f^0(x)$ for all $x \in \{0,1\}^{n-1}$. Therefore,

$$\hat{f}(\{n\}) = \int_{\{x \in \{0,1\}^n : x_n = 1\}} \sqrt{\frac{1-p_n}{p_n}} f(x) d\mu(x) + \int_{\{x \in \{0,1\}^n : x_n = 0\}} \Big(- \sqrt{\frac{p_n}{1-p_n}} \Big) f(x) d\mu(x) =$$

$$= \sqrt{p_n(1-p_n)} \Big(\int_{\{0,1\}^{n-1}} f^1 d\mu' - \int_{\{0,1\}^{n-1}} f^0 d\mu' \Big) = \sqrt{p_n(1-p_n)} \int_{\{0,1\}^{n-1}} (f^1 - f^0) d\mu' \geq 0,$$

as asserted. ■

Claim 9 *Let $f, g : \{0,1\}^n \to [-1,1]$. Then*

$$\sum_{i=1}^{n} \hat{f}(\{i\})\hat{g}(\{i\}) \leq 1.$$

Proof By the Cauchy-Schwarz inequality and Parseval's identity,

$$\sum_{i=1}^{n} \hat{f}(\{i\})\hat{g}(\{i\}) \leq \Big(\sum_{i=1}^{n} \hat{f}(\{i\})^2\Big)^{1/2}\Big(\sum_{i=1}^{n} \hat{g}(\{i\})^2\Big)^{1/2} \leq ||f||_2||g||_2 \leq 1,$$

as asserted. ■

2.2 Sharp Subgaussian Bounds on the Large Deviations of a Weighted Sum of Independent Bernoulli Random Variables

Another component in the proof of Theorem 7 is a bound on the large deviations of the random variable

$$X = \sum_{i \in I} \alpha_i s_i,$$

where $I \subset \{1,\ldots,n\}$, $\{s_i\}_{i=1}^n$ are the functions defined in Section 2.1, and $\sum_{i\in I}\alpha_i^2 = 1$. To be more precise, we are interested in the minimal value of $\gamma = \gamma(p_1,\ldots,p_n)$, such that for all $I \subset \{1,\ldots,n\}$ and all the sequences $\{\alpha_i\}_{i\in I}$ such that $\sum_{i\in I}\alpha_i^2 = 1$, the inequality

$$Prob\Big(|\sum_{i\in I}\alpha_i s_i| > a\Big) < 2\exp\Big(-\frac{a^2}{2\gamma}\Big) \tag{6}$$

holds for all $a > 0$.

We start with examining the case $p_1 = p_2 = \ldots = p_n = p$. In the case of uniform measure (i.e., p=1/2), discussed in [9], the Azuma martingale inequality (see [7], Lemma 1.5) yields the value $\gamma = 1$, which is optimal up to a constant factor. For a general value of p, Azuma's inequality yields the value $\gamma = \max\Big(\frac{p}{1-p}, \frac{1-p}{p}\Big)$, but this value is not optimal. We show the better bound $\gamma = K/H(p)$, where $H(p) = -p\log p - (1-p)\log(1-p)$ is the entropy function, and K is a universal constant.

We use the following Subgaussian bound for functions on the two-point space endowed with the measure μ_p ([1], Proposition 2.1):

Proposition 10 (Bobkov, Houdré, and Tetali) *Consider the two-point space $\{0,1\}$ endowed with the measure μ_p (i.e., $\mu_p(1) = p$, and $\mu_p(0) = 1-p$). Given a function $f : \{0,1\} \to \mathbb{R}$, the optimal value σ^2 in the inequality*

$$E\Big(e^{t(f-Ef)}\Big) \leq e^{\sigma^2 t^2/2},$$

where $t \in \mathbb{R}$ is arbitrary, is given by

$$2\sigma^2 = \frac{p-(1-p)}{\log p - \log(1-p)}(f(1)-f(0))^2.$$

Theorem 11 *Let $\{X_i\}_{i\in I}$ be i.i.d. random variables with the distribution:*

$$X_i = \begin{cases} \sqrt{\frac{1-p}{p}}, & Prob.\ p, \\ -\sqrt{\frac{p}{1-p}}, & Prob.\ (1-p). \end{cases}$$

Then for all $a > 0$ and all $\{\alpha_i\}_{i\in I}$ such that $\sum_{i\in I}\alpha_i^2 = 1$,

$$Prob\Big(|\sum_{i\in I}\alpha_i X_i| > a\Big) \le 2\exp\Big(Ka^2(p\log p + (1-p)\log(1-p))\Big), \tag{7}$$

where K is a universal constant.

Proof The theorem follows from Proposition 10 by the standard Chernoff-type argument described below.

Denote

$$C(p) = \frac{p-(1-p)}{4(\log p - \log(1-p))}.$$

For all $i \in I$, we apply Proposition 10 to the function $f_i = \alpha_i X_i$, and get

$$E\Big(e^{t\alpha_i X_i}\Big) \le \exp\Big(\frac{\alpha_i^2}{p(1-p)}C(p)t^2\Big).$$

Since the X_i-s are independent and since $\sum_{i\in I}\alpha_i^2 = 1$,

$$E\Big(e^{t\sum_{i\in I}\alpha_i X_i}\Big) = \prod_{i\in I} E\Big(e^{t\alpha_i X_i}\Big) \le \prod_{i\in I}\exp\Big(\frac{\alpha_i^2}{p(1-p)}C(p)t^2\Big) = \exp\Big(\frac{C(p)}{p(1-p)}t^2\Big).$$

Thus, by the Markov inequality,

$$Prob\Big(\sum_{i\in I}\alpha_i X_i > a\Big) = Prob\Big(e^{t\sum_{i\in I}\alpha_i X_i} > e^{ta}\Big) \le \exp\Big(\frac{C(p)}{p(1-p)}t^2 - ta\Big).$$

It is easy to see that the optimal bound is obtained for

$$t = \frac{ap(1-p)}{2C(p)},$$

and the bound is

$$Prob\Big(\sum_{i\in I}\alpha_i X_i > a\Big) \le \exp\Big(-\frac{a^2p(1-p)}{4C(p)}\Big).$$

Similar considerations for the random variable $(-\sum_{i\in I}\alpha_i X_i)$ imply that

$$Prob\Big(|\sum_{i\in I}\alpha_i X_i| > a\Big) \le 2\exp\Big(-\frac{a^2p(1-p)}{4C(p)}\Big).$$

Finally, since for a small value of p, $C(p)$ is close to $(-1/4\log p)$, and for a small value of $1-p$, $C(p)$ is close to $(-1/4\log(1-p))$, it is easy to see that there exists a universal constant K such that

$$\frac{p(1-p)}{4C(p)} \geq K\Big(-p\log p-(1-p)\log(1-p)\Big),$$

and hence,

$$Prob\Big(|\sum_{i\in I}\alpha_i X_i|>a\Big) \leq 2\exp\Big(Ka^2(p\log p+(1-p)\log(1-p))\Big).$$

This completes the proof of the theorem. ■

In order to find the optimal bound for general (not necessarily equal) probabilities $(p_1,\ldots,p_n)$, we use the following lemma:

Lemma 12 *The function*

$$F(p)=\frac{C(p)}{p(1-p)}=\frac{2p-1}{4p(1-p)\log\frac{p}{1-p}}$$

is monotone decreasing in p for $0<p<1/2$, and monotone increasing in p for $1/2<p<1$.

Proof Consider the derivative of $F(p)$.

$$F'(p)=\frac{1}{4\Big(p(1-p)\log\frac{p}{1-p}\Big)^2}\Big(2p(1-p)\log\frac{p}{1-p}-(2p-1)((1-2p)\log\frac{p}{1-p}+1)\Big)=$$

$$=\frac{1}{4\Big(p(1-p)\log\frac{p}{1-p}\Big)^2}\Big((1-2p+2p^2)\log\frac{p}{1-p}+(1-2p)\Big).$$

Hence, $F'(p)<0$ if and only if

$$\log\frac{p}{1-p}<-\frac{1-2p}{1-2p+2p^2}.$$

Let

$$G(p)=\log\frac{p}{1-p}, \qquad \text{and} \qquad H(p)=-\frac{1-2p}{1-2p+2p^2},$$

for $0<p<1$. Since $G(1/2)=H(1/2)$, it is sufficient to prove that $G'(p)>H'(p)$ for all $p\neq 1/2$. We have

$$G'(p)=\frac{1}{p(1-p)},$$

and

$$H'(p)=\frac{2(1-2p+2p^2)-(2p-1)(-2+4p)}{(1-2p+2p^2)^2}=\frac{4p(1-p)}{(1-2p+2p^2)^2}.$$

Thus, $G'(p)>H'(p)$ holds if and only if

$$\frac{1}{p(1-p)}>\frac{4p(1-p)}{(1-2p+2p^2)^2},$$

or equivalently,

$$\Big(1-2p+2p^2\Big)^2 > \Big(2p(1-p)\Big)^2,$$

which indeed holds for all $p \neq 1/2$. This completes the proof of the lemma. ■

Theorem 13 *Let $\{X_i\}_{i\in I}$ be independent random variables with the distribution:*

$$X_i = \begin{cases} \sqrt{\frac{1-p_i}{p_i}}, & \textit{Prob. } p_i, \\ -\sqrt{\frac{p_i}{1-p_i}}, & \textit{Prob. } (1-p_i). \end{cases}$$

Then for all $a>0$ and all $\{\alpha_i\}_{i\in I}$ such that $\sum_{i\in I}\alpha_i^2 = 1$,

$$Prob\Big(|\sum_{i\in I}\alpha_i X_i| > a\Big) \leq 2\exp\Big(Ka^2(p\log p + (1-p)\log(1-p))\Big), \tag{8}$$

where $p = \max_{i\in I}\max(p_i, 1-p_i)$, and K is a universal constant.

Proof For all $i \in I$, since the function $F(x) = C(x)/x(1-x)$ is symmetric around $x = 1/2$, and since $p \geq \max(p_i, 1-p_i)$, it follows from Lemma 12 that

$$\frac{C(p)}{p(1-p)} \geq \frac{C(p_i)}{p_i(1-p_i)}.$$

Hence,

$$E\Big(e^{t\alpha_i X_i}\Big) \leq \exp\Big(\frac{\alpha_i^2}{p_i(1-p_i)}C(p_i)t^2\Big) \leq \exp\Big(\frac{\alpha_i^2}{p(1-p)}C(p)t^2\Big),$$

and thus,

$$Exp\Big(e^{t\sum_{i\in I}\alpha_i X_i}\Big) \leq \prod_{i\in I}\exp\Big(\frac{\alpha_i^2}{p(1-p)}C(p)t^2\Big) = \exp\Big(\frac{C(p)}{p(1-p)}t^2\Big).$$

The rest of the proof is the same as the proof of Theorem 11. ■

In the general case, i.e., if we are seeking for a value of γ such that Inequality (6) will hold for every sequence $\{\alpha_i\}_{i\in I}$ with $\sum_{i\in I}\alpha_i^2 = 1$, the assertion of Theorem 13 is tight. Indeed, if $1-p_j = \max_{i\in I}\max(p_i, 1-p_i)$, and we consider the sequence consisting of $\alpha_j = 1$ and zeros, we have

$$Prob\Big(\sum_{i\in I}\alpha_i X_i > \sqrt{\frac{1-p_j}{2p_j}}\Big) = Prob\Big(X_j > \sqrt{\frac{1-p_j}{2p_j}}\Big) = p_j \geq$$

$$\geq \exp\Big(4\frac{1-p_j}{2p_j}(p_j\log p_j + (1-p_j)\log(1-p_j))\Big).$$

Hence, Theorem 13 cannot be improved in the general setting in which it is used in the proof of Theorem 7.

However, if it is sufficient that Inequality (6) will be valid only for sequences of equal α-s (i.e., $\alpha_i = 1/\sqrt{|I|}$ for all $i \in I$), the assertion of Theorem 13 can be easily improved to get the following:

Proposition 14 *Let* $\{X_i\}_{i\in I}$ *be independent random variables with the distribution:*

$$X_i = \begin{cases} \sqrt{\frac{1-p_i}{p_i}}, & \textit{Prob. } p_i, \\ -\sqrt{\frac{p_i}{1-p_i}}, & \textit{Prob. } (1-p_i). \end{cases}$$

Then for all $a > 0$*,*

$$Prob\Big(|\sum_{i\in I} \frac{1}{\sqrt{|I|}} X_i| > a\Big) \leq 2\exp\Big(-a^2 \frac{|I|}{4\sum_{i\in I}\frac{C(p_i)}{p_i(1-p_i)}}\Big). \tag{9}$$

The proof of Proposition 14 follows the lines of the proof of Theorem 11. In Section 4 we show that Proposition 14 can be used to improve the assertion of Theorem 7 in a certain partial case.

3 Proof of Theorem 7

In this section we present the proof of Theorem 7. We start with an outline of the proof.

3.1 Proof Outline

The main ingredient in the proof in the following lemma:

Lemma 15 *Let* f, g *be monotone functions on the discrete cube endowed with the product measure* $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$ *that satisfy* $|f(x)| \leq 1$ *and* $|g(x)| \leq 1$ *for all* $x \in \{0,1\}^n$*. Then*

$$\sum_{k\leq n}\sum_{i\neq k} |\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \leq K\gamma \sum_{k\leq n} \hat{f}(\{k\})\hat{g}(\{k\}) \log \frac{e}{\sum_{k\leq n}\hat{f}(\{k\})\hat{g}(\{k\})}, \tag{10}$$

where $\gamma = \max_{i\leq n} \frac{1}{-p_i\log p_i - (1-p_i)\log(1-p_i)}$ *, and* K *is a universal constant.*

This lemma, establishing a relation between the first-level and the second-level Fourier-Walsh coefficients of monotone functions on the discrete cube, is of independent interest. In the partial case of Boolean functions on the discrete cube endowed with a uniform measure, this lemma (proved in this case in [9]) was used in [10] to establish a relation between the influences of coordinates on a subset of the discrete cube and the size of its boundary.

The proof of Lemma 15 is, in turn, based on the following lemma:

Lemma 16 *Let*

$$\gamma' = \max_{1\leq i\leq n} \frac{1}{2K_1(-p_i\log p_i - (1-p_i)\log(1-p_i))},$$

where K_1 *is the optimal constant for which the assertion of Theorem 13 holds. Let* f *be a monotone function on the discrete cube endowed with the product measure* $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$ *that satisfies* $|f(x)| \leq 1$ *for all* $x \in \{0,1\}^n$*. Let* $\{I, J\}$ *be a partition of* $\{1,\ldots,n\}$*, and let* $S > 0$*. Denote*

$$L = \{k \in J : \Big(\sum_{i\in I}(\hat{f}(\{i,k\}))^2\Big)^{1/2} \geq \sqrt{\gamma'} S \hat{f}(\{k\})\}.$$

Then
$$\sum_{k \in L} \hat{f}(\{k\})^2 \leq K \exp\Big(-\frac{S^2}{K}\Big), \tag{11}$$
where K is a universal constant.

The rest of this section is organized as follows: In Section 3.2 we present the proof of Lemma 16. In Section 3.3 we prove Lemma 15. Finally, we derive Theorem 7 from Lemma 15 in Section 3.4. For the convenience of reading, we divide each proof to several steps, and start each step with a one-sentence description of its content.

3.2 Proof of Lemma 16

Step 1. For each $k \in L$, we define a function f_k related to the condition defining L. The special structure of f_k will allow us to use the bounds on large deviations presented in Section 2.2 to estimate its value.

For each $k \in L$, denote
$$\beta_{i,k} = \hat{f}(\{i,k\}) = \int_{\{0,1\}^n} f s_i s_k d\mu,$$
and
$$\alpha_{i,k} = \frac{\beta_{i,k}}{\sqrt{\sum_{i \in I} \beta_{i,k}^2}},$$
where $\{s_i\}_{i=1}^n$ are the functions defined in Section 2.1. Define
$$f_k = \sum_{i \in I} \alpha_{i,k} s_i.$$
We have
$$\int_{\{0,1\}^n} f_k s_k f d\mu = \int_{\{0,1\}^n} \sum_{i \in I} \alpha_{i,k} s_i s_k f d\mu = \sum_{i \in I} \alpha_{i,k} \int_{\{0,1\}^n} s_i s_k f d\mu = \sum_{i \in I} \alpha_{i,k} \hat{f}(\{i,k\}) =$$
$$= \frac{1}{\sqrt{\sum_{i \in I} \beta_{i,k}^2}} \sum_{i \in I} \beta_{i,k} \hat{f}(\{i,k\}) = \frac{1}{\sqrt{\sum_{i \in I} \hat{f}(\{i,k\})^2}} \sum_{i \in I} \hat{f}(\{i,k\}) \cdot \hat{f}(\{i,k\}) =$$
$$= \Big(\sum_{i \in I} (\hat{f}(\{i,k\}))^2\Big)^{1/2}.$$
Hence, by the definition of L, for each $k \in L$,
$$\int_{\{0,1\}^n} f_k s_k f d\mu \geq \sqrt{\gamma'} S \hat{f}(\{k\}). \tag{12}$$

Step 2. We represent the discrete cube as a product of two spaces corresponding to the partition $\{I,J\}$. This will allow us to use Fubini's theorem to compute $\int_{\{0,1\}^n} f_k s_k f d\mu$.

For a partition $\{I,J\}$ of $\{0,1\}^n$, the discrete cube can be represented as a product $\{0,1\}^n = \Omega_1 \times \Omega_2$, where $\Omega_1 = \{0,1\}^I$ and $\Omega_2 = \{0,1\}^J$. As a result, for each $\omega \in \{0,1\}^n$, we can write

$\omega = (\omega_1, \omega_2)$ where $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. We denote by μ' and μ'' the induced measures on Ω_1 and Ω_2, respectively.

Given $\omega_1 \in \Omega_1$ and $k \in L$, the function $f_k(\omega_1, \omega_2)$ does not depend on ω_2. Hence, we denote its value by $f_k(\omega_1)$. For each $\omega_1 \in \Omega_1$,

$$\int_{\Omega_2} f(\omega_1,\omega_2) f_k(\omega_1,\omega_2) s_k(\omega_1,\omega_2) d\mu''(\omega_2) = f_k(\omega_1) \int_{\Omega_2} f(\omega_1,\omega_2) s_k(\omega_1,\omega_2) d\mu''(\omega_2).$$

Integrating the both sides of the equation with respect to ω_1 and using Fubini's theorem, we get

$$\int_{\{0,1\}^n} f_k s_k f d\mu = \int_{\Omega_1} f_k(\omega_1) \Big(\int_{\Omega_2} f(\omega_1,\omega_2) s_k(\omega_1,\omega_2) d\mu''(\omega_2) \Big) d\mu'(\omega_1). \tag{13}$$

Step 3. We divide the discrete cube to subsets according to the value of f_k and use the bounds on large deviations presented in Section 2.2 to establish an upper bound on the integral $\int f_k s_k f d\mu$ over each of the subsets.

For all $k \in L$ and every integer $q \geq 1$, denote

$$\Omega_{k,0} = \{\omega_1 \in \Omega_1 : |f_k(\omega_1)| \leq 2\sqrt{\gamma'}\},$$

and

$$\Omega_{k,q} = \{\omega_1 \in \Omega_1 : 2^q \sqrt{\gamma'} < |f_k(\omega_1)| \leq 2^{q+1}\sqrt{\gamma'}\}.$$

Let

$$C_{k,q} = \int_{\Omega_{k,q}} f_k(\omega_1) \Big(\int_{\Omega_2} f(\omega_1,\omega_2) s_k(\omega_1,\omega_2) d\mu''(\omega_2) \Big) d\mu'(\omega_1).$$

Note that by Equation (13),

$$\int_{\{0,1\}^n} f_k s_k f d\mu = \sum_{q \geq 0} C_{k,q}. \tag{14}$$

By the Cauchy-Schwarz inequality, for every $q \geq 0$,

$$C_{k,q}^2 \leq \mu'(\Omega_{k,q}) \int_{\Omega_{k,q}} \Big(f_k(\omega_1) \Big(\int_{\Omega_2} f(\omega_1,\omega_2) s_k(\omega_1,\omega_2) d\mu''(\omega_2) \Big) \Big)^2 d\mu'(\omega_1) \leq$$

$$\leq \mu'(\Omega_{k,q}) \gamma' 2^{2q+2} \int_{\Omega_{k,q}} \Big(\int_{\Omega_2} f(\omega_1,\omega_2) s_k(\omega_1,\omega_2) d\mu''(\omega_2) \Big)^2 d\mu'(\omega_1). \tag{15}$$

Consider the function f_k. Note that for each $k \in L$,

$$\sum_{i \in I} \alpha_{i,k}^2 = \frac{1}{\sum_{i \in I} \beta_{i,k}^2} \sum_{i \in I} \beta_{i,k}^2 = 1.$$

Hence, by Theorem 13,

$$\mu'(\Omega_{k,q}) \leq Pr\Big(|f_k| > 2^q \sqrt{\gamma'} \Big) \leq 2 \exp\Big(-\frac{\gamma' 2^{2q}}{2\gamma'} \Big) = 2\exp(-2^{2q-1}).$$

Note that we are able to use Theorem 13 for the space Ω_1 with the value γ' since

$$\max_{i\in I} \frac{1}{2K_1(-p_i\log p_i-(1-p_i)\log(1-p_i))} \le \max_{1\le i\le n} \frac{1}{2K_1(-p_i\log p_i-(1-p_i)\log(1-p_i))} = \gamma'.$$

Substituting into Inequality (15) we get

$$C_{k,q}^2 \le \gamma' 2^{2q+3}\exp(-2^{2q-1})\int_{\Omega_{k,q}}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)s_k(\omega_1,\omega_2)d\mu''(\omega_2)\Big)^2 d\mu'(\omega_1) \le$$

$$\le \gamma' 2^{2q+3}\exp(-2^{2q-1})\int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)s_k(\omega_1,\omega_2)d\mu''(\omega_2)\Big)^2 d\mu'(\omega_1). \tag{16}$$

Finally, for each $\omega_1\in\Omega_1$ we consider the Fourier-Walsh expansion (with respect to the measure μ'') of the function $f_{\omega_1}:\Omega_2\to\mathbb{R}$ defined by $f_{\omega_1}(\omega_2)=f(\omega_1,\omega_2)$. We have

$$\int_{\Omega_2} f(\omega_1,\omega_2)s_k(\omega_1,\omega_2)d\mu'' = \int_{\Omega_2} f_{\omega_1}(\omega_2)s_k(\omega_2)d\mu'' = \hat{f}_{\omega_1}(\{k\}),$$

and hence by the Parseval identity,

$$\sum_{k\in L}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)s_k(\omega_1,\omega_2)d\mu''(\omega_2)\Big)^2 = \sum_{k\in L}\hat{f}_{\omega_1}(\{k\})^2 \le \sum_{T\in\Omega_2}\hat{f}_{\omega_1}(T)^2 = ||f_{\omega_1}||_2^2 \le 1,$$

since $|f|$ (and thus, $|f_{\omega_1}|$) is bounded from above by 1. Thus,

$$\sum_{k\in L}\int_{\Omega_1}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)s_k(\omega_1,\omega_2)d\mu''(\omega_2)\Big)^2 d\mu'(\omega_1) =$$

$$= \int_{\Omega_1}\Big(\sum_{k\in L}\Big(\int_{\Omega_2} f(\omega_1,\omega_2)s_k(\omega_1,\omega_2)d\mu''(\omega_2)\Big)^2\Big)d\mu'(\omega_1) \le 1.$$

Substituting into Inequality (16) we get

$$\sum_{k\in L} C_{k,q}^2 \le \gamma' 2^{2q+3}\exp(-2^{2q-1}). \tag{17}$$

Step 4. We divide $\sum_{q\ge 0} C_{k,q}$ to the sum over "high" values of q and the sum over "low" values of q.

We use the elementary inequality

$$(\sum_{m=1}^{\infty} a_m)^2 \le \sum_{m=1}^{\infty} 2^m a_m^2, \tag{18}$$

which holds for all $\{a_m\}_{m=1}^{\infty}\in\ell^1(\mathbb{N})$. The inequality follows from the convexity of the function $h(x)=x^2$, using the convex combination

$$\sum_{m=1}^{\infty} a_m = \sum_{m=1}^{\infty} 2^{-m}(2^m a_m).$$

By Equations (12) and (14), for each $k \in L$,

$$\sqrt{\gamma'}S\hat{f}(\{k\}) \leq \sum_{q \geq 0} C_{k,q}.$$

Consider an integer q_0 to be determined later, and denote $C_k = \sum_{q \leq q_0} C_{k,q}$. We have

$$\sqrt{\gamma'}S\hat{f}(\{k\}) \leq C_k + \sum_{q > q_0} C_{k,q}.$$

Hence, by Inequality (18),

$$\gamma' S^2 \hat{f}(\{k\})^2 \leq (C_k + \sum_{q > q_0} C_{k,q})^2 \leq 2C_k^2 + \sum_{q > q_0} 2^q C_{k,q}^2.$$

Summing over all $k \in L$ and using Inequality (17) we get

$$\gamma' S^2 \sum_{k \in L} \hat{f}(\{k\})^2 \leq 2 \sum_{k \in L} C_k^2 + \sum_{q > q_0} \gamma' 2^{3q+3} \exp(-2^{2q-1}). \tag{19}$$

Step 5. We use the monotonicity of the function f to establish an upper bound on C_k.

By the definition of C_k and the triangle inequality,

$$\begin{aligned}
|C_k| &= \Big| \int_{\{\omega_1 : |f_k(\omega_1)| \leq \sqrt{\gamma'} 2^{q_0+1}\}} f_k(\omega_1) \Big(\int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) \Big) d\mu'(\omega_1) \Big| \leq \\
&\leq \int_{\{\omega_1 : |f_k(\omega_1)| \leq \sqrt{\gamma'} 2^{q_0+1}\}} \Big| f_k(\omega_1) \Big| \Big| \int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) \Big| d\mu'(\omega_1) \leq \\
&\leq \sqrt{\gamma'} 2^{q_0+1} \int_{\{\omega_1 : |f_k(\omega_1)| \leq \sqrt{\gamma'} 2^{q_0+1}\}} \Big| \int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) \Big| d\mu'(\omega_1) \leq \\
&\leq \sqrt{\gamma'} 2^{q_0+1} \int_{\Omega_1} \Big| \int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) \Big| d\mu'(\omega_1).
\end{aligned} \tag{20}$$

For all $\omega_2 \in \Omega_2$, denote the k-th coordinate of ω_2 by $\omega_2(k)$, and let $\omega_2 \oplus e_k$ be the coordinate-wise sum modulo 2 of ω_2 with the unit vector e_k. Then for all $\omega_1 \in \Omega_1$,

$$\begin{aligned}
&\int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) = \sum_{\omega_2 \in \Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) \mu''(\omega_2) = \\
&= \sum_{\{\omega_2 : \omega_2(k) = 0\}} \Big(f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) \mu''(\omega_2) + f(\omega_1, \omega_2 \oplus e_k) s_k(\omega_1, \omega_2 \oplus e_k) \mu''(\omega_2 \oplus e_k) \Big) = \\
&= \sum_{\{\omega_2 : \omega_2(k) = 0\}} \Big(f(\omega_1, \omega_2) \Big(-\sqrt{\frac{p_k}{1 - p_k}} \Big) \mu''(\omega_2) + f(\omega_1, \omega_2 \oplus e_k) \Big(\sqrt{\frac{1 - p_k}{p_k}} \Big) \Big(\frac{p_k}{1 - p_k} \Big) \mu''(\omega_2) \Big) = \\
&= \sum_{\{\omega_2 : \omega_2(k) = 0\}} \sqrt{\frac{p_k}{1 - p_k}} \mu''(\omega_2) \Big(f(\omega_1, \omega_2 \oplus e_k) - f(\omega_1, \omega_2) \Big).
\end{aligned}$$

Since f is monotone, each of the terms $(f(\omega_1, \omega_2 \oplus e_k) - f(\omega_1, \omega_2))$ is non-negative, and hence for all $\omega_1 \in \Omega_1$,

$$\int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) \geq 0.$$

Thus, by Fubini's theorem,

$$\int_{\Omega_1} \Big| \int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) \Big| d\mu'(\omega_1) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu''(\omega_2) d\mu'(\omega_1) =$$

$$= \int_{\Omega} f(\omega_1, \omega_2) s_k(\omega_1, \omega_2) d\mu(\omega_1, \omega_2) = \hat{f}(\{k\}),$$

and therefore, by Inequality (20),

$$|C_k| \leq \sqrt{\gamma'} 2^{q_0+1} \hat{f}(\{k\}). \tag{21}$$

Step 6. We complete the proof by choosing an appropriate value of q_0.

Combining Inequalities (19) and (21), we get

$$\gamma' S^2 \sum_{k \in L} \hat{f}(\{k\})^2 \leq \gamma' 2^{2q_0+3} \sum_{k \in L} \hat{f}(\{k\})^2 + \sum_{q > q_0} \gamma' 2^{3q+3} \exp(-2^{2q-1}). \tag{22}$$

We choose q_0 to be the largest integer such that $2^{2q_0+3} \leq S^2/2$. Note that $q_0 \geq 0$ whenever $S \geq 4$. Substituting the chosen value of q_0 into Inequality (22), we get

$$S^2 \sum_{k \in L} \hat{f}(\{k\})^2 \leq \sum_{q > q_0} 2^{3q+4} \exp(-2^{2q-1}). \tag{23}$$

As was noted in ([9], p. 251), there exists a universal constant K' such that for every q_0,

$$\sum_{q > q_0} 2^{3q+4} \exp(-2^{2q-1}) \leq K' \exp(-2^{2q_0}).$$

Thus, by Inequality (23),

$$S^2 \sum_{k \in L} \hat{f}(\{k\})^2 \leq K' \exp(-2^{2q_0}).$$

By the choice of q_0 we have $S^2 < 2^{2(q_0+1)+4}$, and hence $-2^{2q_0} < -\frac{S^2}{64}$. Therefore, for all $S \geq 4$,

$$\sum_{k \in L} \hat{f}(\{k\})^2 \leq S^2 \sum_{k \in L} \hat{f}(\{k\})^2 \leq K' \exp\Big(-\frac{S^2}{64}\Big) \leq K'' \exp\Big(-\frac{S^2}{K''}\Big), \tag{24}$$

where $K'' = max(K', 64)$. This proves the assertion of Lemma 16 for $S \geq 4$.

For $S < 4$ and for every $K \geq 16$,

$$K \exp\Big(-\frac{S^2}{K}\Big) \geq K \exp\Big(-\frac{16}{K}\Big) \geq 1.$$

Since $|f|$ is bounded from above by 1, by the Parseval identity we have

$$\sum_{k\in L}\hat{f}(\{k\})^2 \le ||f||_2^2 \le 1 \le K\exp\Big(-\frac{S^2}{K}\Big), \tag{25}$$

for all $K \ge 16$, and in particular, for K'' defined above. Therefore, combining Inequalities (24) and (25) we get that for every $S > 0$,

$$\sum_{k\in L}\hat{f}(\{k\})^2 \le K''\exp\Big(-\frac{S^2}{K''}\Big). \tag{26}$$

This completes the proof of Lemma 16.

3.3 Proof of Lemma 15

Step 1. We reduce the assertion of the lemma to a similar assertion concerning partitions $\{I, J\}$ of $\{1,\dots,n\}$.

We note that it is sufficient to prove that for every partition $\{I, J\}$ of $\{1,\dots,n\}$,

$$\sum_{i\in I}\sum_{k\in J}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le K'\gamma\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\})\log\frac{e}{\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\})} \tag{27}$$

for some universal constant K'. The assertion of Lemma 15 (with the weaker constant $K = 2K'$) follows from Inequality (27) by taking the average over all the possible partitions $\{I, J\}$ such that $|I| = |J|$ if n is even or $|I| = |J| + 1$ if n is odd, since for each pair (i, k) the probability that it is included in the cut of a random partition of this type is greater than $1/2$.

Step 2. We extend the result of Lemma 16 to the multiplication of two monotone functions.

Let

$$\gamma' = \max_{1\le i\le n}\frac{1}{2K_1(-p_i\log p_i-(1-p_i)\log(1-p_i))},$$

as defined in Lemma 16. For $S > 0$, let

$$L_S = \{k\in J : \sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \ge \gamma' S^2\hat{f}(\{k\})\hat{g}(\{k\})\}.$$

We want to show that there exists a universal constant K_0 such that for all $S > 0$,

$$\sum_{k\in L_S}\hat{f}(\{k\})\hat{g}(\{k\}) \le K_0\exp\Big(\frac{-S^2}{K_0}\Big). \tag{28}$$

Note that if $f = g$, this claim is reduced to the statement of Lemma 16.

Define

$$L^1 = \{k\in J : \Big(\sum_{i\in I}(\hat{f}(\{i,k\}))^2\Big)^{1/2} \ge \sqrt{\gamma'}S\hat{f}(\{k\})\},$$

and

$$L^2 = \{k \in J : \Big(\sum_{i\in I}(\hat{g}(\{i,k\}))^2\Big)^{1/2} \geq \sqrt{\gamma'}S\hat{g}(\{k\})\}.$$

By the Cauchy-Schwarz inequality,

$$\sum_{k\in L^1} \hat{f}(\{k\})\hat{g}(\{k\}) \leq \Big(\sum_{k\in L^1} \hat{f}(\{k\})^2\Big)^{1/2}\Big(\sum_{k\in L^1} \hat{g}(\{k\})^2\Big)^{1/2}.$$

By Lemma 16, there exists a constant K_2 such that

$$\Big(\sum_{k\in L^1} \hat{f}(\{k\})^2\Big)^{1/2} \leq \Big(K_2 \exp\Big(-\frac{S^2}{K_2}\Big)\Big)^{1/2} = \sqrt{K_2}\exp\Big(-\frac{S^2}{2K_2}\Big).$$

By Parseval's identity,

$$\Big(\sum_{k\in L^1} \hat{g}(\{k\})^2\Big)^{1/2} \leq ||g||_2 \leq 1.$$

Hence,

$$\sum_{k\in L^1} \hat{f}(\{k\})\hat{g}(\{k\}) \leq \Big(\sum_{k\in L^1} \hat{f}(\{k\})^2\Big)^{1/2}\Big(\sum_{k\in L^1} \hat{g}(\{k\})^2\Big)^{1/2} \leq \sqrt{K_2}\exp\Big(-\frac{S^2}{2K_2}\Big).$$

Similarly,

$$\sum_{k\in L^2} \hat{f}(\{k\})\hat{g}(\{k\}) \leq \sqrt{K_2}\exp\Big(-\frac{S^2}{2K_2}\Big).$$

Finally, by the Cauchy-Schwarz inequality, $L_S \subset (L^1 \cup L^2)$, and thus, taking $K_0 = \max(2\sqrt{K_2}, 2K_2)$ we get

$$\sum_{k\in L_S} \hat{f}(\{k\})\hat{g}(\{k\}) \leq \sum_{k\in L^1} \hat{f}(\{k\})\hat{g}(\{k\}) + \sum_{k\in L^2} \hat{f}(\{k\})\hat{g}(\{k\}) \leq 2\sqrt{K_2}\exp\Big(-\frac{S^2}{2K_2}\Big) \leq K_0\exp\Big(-\frac{S^2}{K_0}\Big),$$

as asserted.

Step 3. We divide J to subsets according to the value of the sum $\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})|$, and establish an upper bound on this sum over each of the subsets.

Let

$$L_0 = \{k \in J : \sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \leq 4\gamma'\hat{f}(\{k\})\hat{g}(\{k\})\}$$

and

$$L_q = \{k \in J : 2^{2q}\gamma'\hat{f}(\{k\})\hat{g}(\{k\}) < \sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \leq 2^{2q+2}\gamma'\hat{f}(\{k\})\hat{g}(\{k\})\},$$

for every integer $q \geq 1$. By Inequality (28), for every $q > 0$,

$$\sum_{k\in L_q} \hat{f}(\{k\})\hat{g}(\{k\}) \leq K_0\exp\Big(-\frac{2^{2q}}{K_0}\Big).$$

Hence, by the definition of L_q,

$$\sum_{k\in L_q}\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le \sum_{k\in L_q} 2^{2q+2}\gamma'\hat{f}(\{k\})\hat{g}(\{k\}) \le 2^{2q+2}\gamma' K_0 \exp\Big(-\frac{2^{2q}}{K_0}\Big).$$

As noted in ([9], p. 253), there exists a universal constant K_3 such that for every $q_0 > 0$,

$$\sum_{q>q_0} 2^{2q+2}\exp\Big(-\frac{2^{2q}}{K_0}\Big) \le K_3 \exp\Big(-\frac{2^{2q_0}}{K_3}\Big).$$

Therefore, for every $q_0 > 0$,

$$\sum_{q>q_0}\sum_{k\in L_q}\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le K_0K_3\gamma' \exp\Big(-\frac{2^{2q_0}}{K_3}\Big). \tag{29}$$

On the other hand, by the definition of L_q,

$$\sum_{q\le q_0}\sum_{k\in L_q}\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le \sum_{q\le q_0}\sum_{k\in L_q}\gamma' 2^{2q+2}\hat{f}(\{k\})\hat{g}(\{k\}) \le \gamma' 2^{2q_0+2}\sum_{k\in J}\hat{f}(\{k\})\hat{g}(\{k\}). \tag{30}$$

Combining Inequalities (29) and (30), we get

$$\sum_{k\in J}\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le K_0K_3\gamma' \exp\Big(-\frac{2^{2q_0}}{K_2}\Big) + \gamma' 2^{2q_0+2}\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\}). \tag{31}$$

Step 4. We complete the proof by choosing an appropriate value of q_0.

Let q_0 be the smallest integer such that

$$\exp\Big(-\frac{2^{2q_0}}{K_3}\Big) \le \sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\}).$$

Substituting q_0 into Inequality (31) we get

$$\sum_{k\in J}\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le \gamma'\Big(\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\})\Big)(K_0K_3 + 2^{2q_0+2}). \tag{32}$$

Note that by the choice of q_0,

$$\exp\Big(-\frac{2^{2(q_0-1)}}{K_3}\Big) \ge \sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\}),$$

and thus,

$$2^{2q_0+2} \le 16K_3 \log\Big(\frac{1}{\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\})}\Big).$$

Substituting into Inequality (32) we get

$$\sum_{k\in J}\sum_{i\in I}|\hat{f}(\{i,k\})||\hat{g}(\{i,k\})| \le \gamma'\Big(\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\})\Big)\Big(K_0K_3 + 16K_3\log\Big(\frac{1}{\sum_{k\le n}\hat{f}(\{k\})\hat{g}(\{k\})}\Big)\Big).$$

By Claim 9,

$$\log\Big(\frac{e}{\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})}\Big)\ge 1,$$

and hence for $K''=\max(16K_3,K_0K_3)$ we have

$$\sum_{k\in J}\sum_{i\in I}|\hat f(\{i,k\})||\hat g(\{i,k\})|\le \gamma'\Big(\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})\Big)\Big(K_0K_3+16K_3\log\Big(\frac{e}{\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})}\Big)\Big)\le$$

$$\le K''\gamma'\Big(\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})\Big)\log\Big(\frac{e}{\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})}\Big).$$

Finally, since γ' and γ differ only by the multiplicative constant $2K_1$, we denote $K'=K''/2K_1$, and get

$$\sum_{k\in J}\sum_{i\in I}|\hat f(\{i,k\})||\hat g(\{i,k\})|\le K'\gamma\Big(\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})\Big)\log\Big(\frac{e}{\sum_{k\le n}\hat f(\{k\})\hat g(\{k\})}\Big). \tag{33}$$

This concludes the proof of Lemma 15.

3.4 Proof of Theorem 7

In the proof of the theorem we use the following simple proposition from [9] concerning properties of the function $\varphi(x)=\frac{x}{\log(e/x)}$ for $0<x\le 1$:

Proposition 17 *The function $\varphi(x)=\frac{x}{\log(e/x)}$ is monotone increasing and convex in $(0,1]$, and for all $0<u\le v\le 1$,*

$$\varphi(v)\le\varphi(u)+\frac{2(v-u)}{\log(e/v)}.$$

The proof of the theorem is by induction on n. Since the case $n=1$ is trivial, we present only the induction step.

Step 1. We use Lemma 15 to choose an appropriate coordinate to perform the inductive step on.

We note that by Lemma 15, there exists a universal constant K' and a coordinate $k\le n$ such that

$$\sum_{i\ne k}|\hat f(\{i,k\})||\hat g(\{i,k\})|\le K'\gamma\hat f(\{k\})\hat g(\{k\})\log\frac{e}{\sum_{j=1}^n\hat f(\{j\})\hat g(\{j\})}. \tag{34}$$

Indeed, if for all the coordinates Inequality (34) does not hold then by summing over all the coordinates we get a contradiction to the assertion of Lemma 15. We assume, without loss of generality, that the coordinate k for which Inequality (34) holds is the n-th coordinate, and perform the induction step on this coordinate.

Step 2. We establish a relation between the correlation of f and g and the correlation of the restrictions of f and g to the first $n-1$ coordinates.

We start with several notations. As usual, we denote:

$$Cov(f,g) = \int_{\{0,1\}^n} fg d\mu - \int_{\{0,1\}^n} f d\mu \int_{\{0,1\}^n} g d\mu.$$

We define $f^0, f^1, g^0, g^1 : \{0,1\}^{n-1} \to \mathbb{R}$ by:

$$f^0(x_1,\ldots,x_{n-1}) = f(x_1,\ldots,x_{n-1},0), \qquad f^1(x_1,\ldots,x_{n-1}) = f(x_1,\ldots,x_{n-1},1),$$

for all $(x_1,\ldots,x_{n-1}) \in \{0,1\}^{n-1}$, and similarly for g^0, g^1. Finally, we denote by $\mu' = \mu_{p_1} \otimes \cdots \otimes \mu_{p_{n-1}}$ the measure induced by μ on $\{0,1\}^{n-1}$.

By the definition of f^0, f^1, g^0, and g^1,

$$\int_{\{0,1\}^n} fg d\mu = \int_{\{x\in\{0,1\}^n : x_n=0\}} fg(x) d\mu(x) + \int_{\{x\in\{0,1\}^n : x_n=1\}} fg(x) d\mu(x) =$$

$$= (1-p_n) \int_{\{0,1\}^{n-1}} f^0 g^0 d\mu' + p_n \int_{\{0,1\}^{n-1}} f^1 g^1 d\mu',$$

and thus,

$$\int_{\{0,1\}^n} fg d\mu - (1-p_n) \int_{\{0,1\}^{n-1}} f^0 g^0 d\mu' - p_n \int_{\{0,1\}^{n-1}} f^1 g^1 d\mu' = 0. \tag{35}$$

Similarly, since

$$\int_{\{0,1\}^n} f d\mu = (1-p_n) \int_{\{0,1\}^{n-1}} f^0 d\mu' + p_n \int_{\{0,1\}^{n-1}} f^1 d\mu',$$

and

$$\int_{\{0,1\}^n} g d\mu = (1-p_n) \int_{\{0,1\}^{n-1}} g^0 d\mu' + p_n \int_{\{0,1\}^{n-1}} g^1 d\mu',$$

a direct computation shows that

$$\Big(\int_{\{0,1\}^n} f d\mu \int_{\{0,1\}^n} g d\mu\Big) - \Big((1-p_n) \int_{\{0,1\}^{n-1}} f^0 d\mu' \int_{\{0,1\}^{n-1}} g^0 d\mu'\Big) - \Big(p_n \int_{\{0,1\}^{n-1}} f^1 d\mu' \int_{\{0,1\}^{n-1}} g^1 d\mu'\Big) =$$

$$= -p_n(1-p_n)\Big(\int_{\{0,1\}^{n-1}} f^1 d\mu' - \int_{\{0,1\}^{n-1}} f^0 d\mu'\Big)\Big(\int_{\{0,1\}^{n-1}} g^1 d\mu' - \int_{\{0,1\}^{n-1}} g^0 d\mu'\Big). \tag{36}$$

Subtracting Equation (36) from Equation (35) we get:

$$Cov(f,g) - (1-p_n) Cov(f^0,g^0) - p_n Cov(f^1,g^1) = p_n(1-p_n) \int_{\{0,1\}^{n-1}} (f^1 - f^0) d\mu' \int_{\{0,1\}^{n-1}} (g^1 - g^0) d\mu', \tag{37}$$

where the correlation of f and g is with respect to the measure μ and the correlations of (f^0, g^0) and (f^1, g^1) are with respect to the measure μ'.

Step 3. We apply the induction hypothesis to the restrictions of f and g to the first $n-1$ coordinates.

Denote the constant in the assertion of Theorem 7 by K_0. By the induction hypothesis,

$$Cov(f^0,g^0) \geq \frac{1}{\gamma K_0}\varphi\Big(\sum_{i<n}\widehat{f^0}(\{i\})\widehat{g^0}(\{i\})\Big),$$

and

$$Cov(f^1,g^1) \geq \frac{1}{\gamma K_0}\varphi\Big(\sum_{i<n}\widehat{f^1}(\{i\})\widehat{g^1}(\{i\})\Big).$$

Note that while the value of γ asserted by the induction hypothesis is

$$\gamma(p_1,\ldots,p_{n-1}) = \max_{i<n}\frac{1}{-p_i\log p_i-(1-p_i)\log(1-p_i)},$$

we can replace it by the greater value

$$\gamma(p_1,\ldots,p_n) = \max_{i\leq n}\frac{1}{-p_i\log p_i-(1-p_i)\log(1-p_i)},$$

which is the value asserted in the induction step. Hence, we use the same value of γ for both cases.

By Proposition 17, the function $\varphi(x)$ is convex on $(0,1]$, and hence,

$$(1-p_n)Cov(f^0,g^0)+p_nCov(f^1,g^1) \geq \frac{1}{\gamma K_0}\varphi\Big(\sum_{i<n}(1-p_n)\widehat{f^0}(\{i\})\widehat{g^0}(\{i\})+p_n\widehat{f^1}(\{i\})\widehat{g^1}(\{i\})\Big).$$

Substituting into Equation (37) we get

$$Cov(f,g) \geq \frac{1}{\gamma K_0}\varphi\Big(\sum_{i<n}(1-p_n)\widehat{f^0}(\{i\})\widehat{g^0}(\{i\})+p_n\widehat{f^1}(\{i\})\widehat{g^1}(\{i\})\Big)+$$
$$+p_n(1-p_n)\int_{\{0,1\}^{n-1}}(f^1-f^0)d\mu'\int_{\{0,1\}^{n-1}}(g^1-g^0)d\mu'.$$

Therefore, in order to prove Theorem 7, it is sufficient to prove the following inequality:

$$\frac{1}{\gamma K_0}\varphi\Big(\sum_{i\leq n}\hat{f}(\{i\})\hat{g}(\{i\})\Big) \leq \frac{1}{\gamma K_0}\varphi\Big(\sum_{i<n}(1-p_n)\widehat{f^0}(\{i\})\widehat{g^0}(\{i\})+p_n\widehat{f^1}(\{i\})\widehat{g^1}(\{i\})\Big)+$$
$$+p_n(1-p_n)\int_{\{0,1\}^{n-1}}(f^1-f^0)d\mu'\int_{\{0,1\}^{n-1}}(g^1-g^0)d\mu'. \tag{38}$$

We will deduce this inequality from Proposition 17, with

$$u = \sum_{i<n}(1-p_n)\widehat{f^0}(\{i\})\widehat{g^0}(\{i\})+p_n\widehat{f^1}(\{i\})\widehat{g^1}(\{i\}),$$

and

$$v = \sum_{i\leq n}\hat{f}(\{i\})\hat{g}(\{i\}).$$

Step 4. We use the specially chosen induction coordinate to establish an upper bound on $|u - v|$.

Since

$$\hat{f}(\{i\}) = \int_{\{x\in\{0,1\}^n : x_n=0\}} f s_i(x) d\mu(x) + \int_{\{x\in\{0,1\}^n : x_n=1\}} f s_i(x) d\mu(x) =$$

$$= (1-p_n) \int_{\{0,1\}^{n-1}} f^0 s_i d\mu' + p_n \int_{\{0,1\}^{n-1}} f^1 s_i d\mu' = (1-p_n)\widehat{f^0}(\{i\}) + p_n \widehat{f^1}(\{i\}),$$

and similarly for g, a direct computation shows that for all $1 \le i \le n-1$,

$$\hat{f}(\{i\})\hat{g}(\{i\}) - (1-p_n)\widehat{f^0}(\{i\})\widehat{g^0}(\{i\}) - p_n\widehat{f^1}(\{i\})\widehat{g^1}(\{i\}) =$$

$$= \Big((1-p_n)\widehat{f^0}(\{i\}) + p_n\widehat{f^1}(\{i\})\Big)\Big((1-p_n)\widehat{g^0}(\{i\}) + p_n\widehat{g^1}(\{i\})\Big) - (1-p_n)\widehat{f^0}(\{i\})\widehat{g^0}(\{i\}) - p_n\widehat{f^1}(\{i\})\widehat{g^1}(\{i\}) =$$

$$= -p_n(1-p_n)(\widehat{f^1}(\{i\}) - \widehat{f^0}(\{i\}))(\widehat{g^1}(\{i\}) - \widehat{g^0}(\{i\})).$$

Therefore,

$$v - u = \hat{f}(\{n\})\hat{g}(\{n\}) - p_n(1-p_n)\sum_{i<n}\Big(\widehat{f^1}(\{i\}) - \widehat{f^0}(\{i\})\Big)\Big(\widehat{g^1}(\{i\}) - \widehat{g^0}(\{i\})\Big),$$

and thus, by the triangle inequality,

$$|v - u| \le \hat{f}(\{n\})\hat{g}(\{n\}) + p_n(1-p_n)\sum_{i<n}|\widehat{f^1}(\{i\}) - \widehat{f^0}(\{i\})||\widehat{g^1}(\{i\}) - \widehat{g^0}(\{i\})|. \tag{39}$$

(Note that $\hat{f}(\{n\})\hat{g}(\{n\}) \ge 0$ by Claim 8). We observe that for every $i < n$,

$$\widehat{f^1}(\{i\}) - \widehat{f^0}(\{i\}) = \int_{\{0,1\}^{n-1}} f^1 s_i d\mu' - \int_{\{0,1\}^{n-1}} f^0 s_i d\mu' =$$

$$= \frac{1}{p_n}\int_{\{x\in\{0,1\}^n : x_n=1\}} f s_i(x) d\mu(x) - \frac{1}{1-p_n}\int_{\{x\in\{0,1\}^n : x_n=0\}} f s_i(x) d\mu(x) =$$

$$= \frac{1}{\sqrt{p_n(1-p_n)}}\Big(\int_{\{x\in\{0,1\}^n : x_n=1\}} \sqrt{\frac{1-p_n}{p_n}} f s_i(x) d\mu(x) - \int_{\{x\in\{0,1\}^n : x_n=0\}} \sqrt{\frac{p_n}{1-p_n}} f s_i(x) d\mu(x)\Big) =$$

$$= \frac{1}{\sqrt{p_n(1-p_n)}}\Big(\int_{\{x\in\{0,1\}^n : x_n=1\}} f s_i s_n(x) d\mu(x) + \int_{\{x\in\{0,1\}^n : x_n=0\}} f s_i s_n(x) d\mu(x)\Big) =$$

$$= \frac{1}{\sqrt{p_n(1-p_n)}}\int_{\{0,1\}^n} f s_i s_n d\mu = \frac{1}{\sqrt{p_n(1-p_n)}}\hat{f}(\{i,n\}), \tag{40}$$

and similarly for g, and hence

$$|\widehat{f^1}(\{i\}) - \widehat{f^0}(\{i\})||\widehat{g^1}(\{i\}) - \widehat{g^0}(\{i\})| = \frac{1}{p_n(1-p_n)}|\hat{f}(\{i,n\})||\hat{g}(\{i,n\})|.$$

Substituting into Inequality (39), we get

$$|v - u| \le \hat{f}(\{n\})\hat{g}(\{n\}) + \sum_{i<n}|\hat{f}(\{i,n\})||\hat{g}(\{i,n\})|. \tag{41}$$

Finally, by Inequality (34),

$$|v-u| \le \hat{f}(\{n\})\hat{g}(\{n\}) + K'\gamma \hat{f}(\{n\})\hat{g}(\{n\}) \log \frac{e}{\sum_{j=1}^{n} \hat{f}(\{k\})\hat{g}(\{k\})}. \tag{42}$$

Step 5. We apply Proposition 17 and conclude the proof.

By Claim 9,

$$\log \frac{e}{\sum_{j=1}^{n} \hat{f}(\{k\})\hat{g}(\{k\})} \ge 1.$$

In addition, by the definition of γ we have $\gamma \ge 1$, and we may assume that $K' \ge 1$. Thus,

$$K'\gamma \hat{f}(\{n\})\hat{g}(\{n\}) \log \frac{e}{\sum_{j=1}^{n} \hat{f}(\{k\})\hat{g}(\{k\})} \ge \hat{f}(\{n\})\hat{g}(\{n\}).$$

Substituting into Inequality (42) we get

$$|v-u| \le 2K'\gamma \hat{f}(\{n\})\hat{g}(\{n\}) \log \frac{e}{\sum_{j=1}^{n} \hat{f}(\{k\})\hat{g}(\{k\})}, \tag{43}$$

and hence, by Proposition 17 (with u and v as defined above),

$$\varphi(v) \le \varphi(u) + 4K'\gamma \hat{f}(\{n\})\hat{g}(\{n\}). \tag{44}$$

Therefore, by Inequality (38), in order to prove Theorem 7, it is sufficient to prove the following inequality:

$$4K'\gamma \hat{f}(\{n\})\hat{g}(\{n\}) \le K_0 \gamma p_n(1-p_n) \int_{\{0,1\}^{n-1}} (f^1 - f^0) d\mu' \int_{\{0,1\}^{n-1}} (g^1 - g^0) d\mu'. \tag{45}$$

Finally, by a computation similar to the computation in Inequality (40),

$$\int_{\{0,1\}^{n-1}} (f^1 - f^0) d\mu' = \sqrt{\frac{1}{p_n(1-p_n)}} \hat{f}(\{n\})$$

and similarly for g, and hence Inequality (45) holds for $K_0 = 4K'$. This concludes the proof of Theorem 7.

4 Tightness of Results

In this section we discuss the tightness of Theorems 7 and 5.

4.1 Tightness of Theorem 7

As was shown in [9], the assertion of Theorem 5 is tight for the sets:

$$A = \{x : \sum_{i=1}^{n} x_i > k\}, \qquad \text{and} \qquad B = \{x : \sum_{i=1}^{n} x_i \ge n-k\}.$$

We check the tightness of Theorem 7 by examining a similar example in our more general framework. Since the assertion of Theorem 7 contains a universal multiplicative constant, all the computations in the example are approximations up to a constant factor. As a result, the symbol "=" is replaced by "≈", standing for equivalence up to a constant multiplicative factor.

Example Let p be "small" (i.e., tend to zero when $n \to \infty$). Consider the discrete cube endowed with the product measure $\mu = \mu_p^{\otimes n} = \mu_p \otimes \cdots \otimes \mu_p$, and let

$$A = \{x : \sum_{i=1}^{n} x_i \geq k\}, \qquad \text{and} \qquad B = \{x : \sum_{i=1}^{n} x_i \geq l\},$$

where $k \leq np/2$ and $l \geq n/2$. (These values are chosen in order to simplify the computations). We have:

$$\int_{\{0,1\}^n} 1_A 1_B d\mu - \int_{\{0,1\}^n} 1_A d\mu \int_{\{0,1\}^n} 1_B d\mu = \mu(A \cap B) - \mu(A)\mu(B) = \mu(B)(1 - \mu(A)) =$$

$$= \Big(\sum_{i=l}^{n} \binom{n}{i} p^i (1-p)^{n-i} \Big) \Big(\sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} \Big).$$

For $k \leq np/2$,

$$\sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} \approx \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}.$$

Similarly, for $l \geq n/2$,

$$\sum_{i=l}^{n} \binom{n}{i} p^i (1-p)^{n-i} \approx \binom{n}{l} p^l (1-p)^{n-l}.$$

Hence,

$$\int_{\{0,1\}^n} 1_A 1_B d\mu - \int_{\{0,1\}^n} 1_A d\mu \int_{\{0,1\}^n} 1_B d\mu \approx \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} \binom{n}{l} p^l (1-p)^{n-l} =$$

$$= \frac{n^2}{(n-k+1)l} \binom{n-1}{k-1} \binom{n-1}{l-1} p^{k+l-1} (1-p)^{2n-k-l+1} \approx$$

$$\approx \binom{n-1}{k-1} \binom{n-1}{l-1} p^{k+l-1} (1-p)^{2n-k-l+1}. \tag{46}$$

On the other hand, for all $1 \leq j \leq n$,

$$\hat{1}_A(\{j\}) = \int_{\{0,1\}^n} 1_A s_j d\mu = \sqrt{\frac{1-p}{p}} \mu(A_j) =$$

$$= \sqrt{\frac{1-p}{p}} \binom{n-1}{k-1} p^k (1-p)^{n-k} = \binom{n-1}{k-1} p^{k-1/2} (1-p)^{n-k+1/2},$$

and similarly,

$$\hat{1}_B(\{j\}) = \sqrt{\frac{1-p}{p}}\mu(B_j) = \binom{n-1}{l-1}p^{l-1/2}(1-p)^{n-l+1/2}.$$

Thus,

$$\sum_{j=1}^{n} \hat{1}_A(\{j\})\hat{1}_B(\{j\}) = n\binom{n-1}{k-1}\binom{n-1}{l-1}p^{k+l-1}(1-p)^{2n-k-l+1}.$$

It is easy to see that the dominant term in the product in the right hand side is p^{k+l} (i.e., the other terms are much closer to 1). Hence,

$$\log \frac{e}{\sum_{j=1}^{n} \hat{1}_A(\{j\})\hat{1}_B(\{j\})} \approx \log \frac{1}{p^{k+l}} \approx -n\log p,$$

and thus,

$$\varphi\Big(\sum_{j=1}^{n} \hat{1}_A(\{j\})\hat{1}_B(\{j\})\Big) \approx \binom{n-1}{k-1}\binom{n-1}{l-1}p^{k+l-1}(1-p)^{2n-k-l+1}/(-\log p). \tag{47}$$

Dividing Equation (46) by Equation (47) we get:

$$\frac{\int_{\{0,1\}^n} 1_A 1_B d\mu - \int_{\{0,1\}^n} 1_A d\mu \int_{\{0,1\}^n} 1_B d\mu}{\varphi\Big(\sum_{j=1}^{n} \hat{1}_A(\{j\})\hat{1}_B(\{j\})\Big)} \approx (-\log p). \tag{48}$$

By Theorem 7, the lower bound on the left hand side in our case is $K\gamma \approx -p\log p$, and hence the example misses the lower bound by a factor of $1/p$.

It seems challenging to further examine the tightness of Theorem 7 by checking different examples. A natural candidate to be considered is the "tribes" function (see [2]) and its dual function, for which it was shown in [5] that the assertion of Theorem 5 is tight. However, the calculation of the correlation between these functions with respect to a general product measure seems complicated.

4.2 Improvement of Theorem 7 in a Partial Case

As was shown in Section 2.2, the bound on large deviations asserted in Theorem 13 can be improved, if only random variables of the type $\sum_{i\in I}(1/\sqrt{|I|})X_i$ are considered (see Proposition 14). Since Theorem 13 is used in the proof of Theorem 7 as a "black-box", such improvement would immediately imply an improvement of the assertion of Theorem 7 if only random variables of the prescribed type were used in the proof. We show that this is indeed the case in some specific class of functions.

We observe that the values $\alpha_{i,k}$ used in the proof of Theorem 7 are "normalized" second-level Fourier-Walsh coefficients of the functions f and g. Hence, if all the second-level Fourier-Walsh coefficients of f and g are equal, then the bound on large deviations is indeed applied only for equal α-s, and thus, the application of Theorem 13 can be replaced by application of the stronger Proposition 14. We note that since the sets I considered in the proof are of size either $n/2$ or $(n+1)/2$ (see Step 1 of the proof of Lemma 15), and since the assertion of Theorem 7 contains a constant multiplicative factor, the value $|I|$ in Proposition 14 can be replaced by n. Substituting into the proof of Theorem 7, we get the following:

Proposition 18 *Let f, g be monotone symmetric functions on the discrete cube endowed with the product measure $\mu = \mu_{p_1} \otimes \cdots \otimes \mu_{p_n}$, that satisfy $|f(x)| \leq 1$ and $|g(x)| \leq 1$ for all $x \in \{0,1\}^n$. Assume further that all the second-level Fourier-Walsh coefficients of each of the functions f and g are equal to each other. Then*

$$\int fg d\mu - \int f d\mu \int g d\mu \geq K \frac{n}{\sum_{i \leq n} \frac{2p_i - 1}{p_i(1-p_i)\log \frac{p_i}{1-p_i}}} \varphi\Big(\sum_{i \leq n} \hat{f}(\{i\}) \hat{g}(\{i\}) \Big),$$

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

For unequal probabilities $p_1, \ldots, p_n$, Proposition 18 is stronger than Theorem 7. On the other hand, the result seems a bit messy, and it is not clear which functions satisfy the property of second-level Fourier-Walsh coefficients required in the proposition (for example, symmetric functions do not satisfy it).

4.3 Tightness of Theorem 5

In this subsection we consider the discrete cube endowed with the uniform measure μ. The lower bound on the correlation between monotone subsets presented by Talagrand (Theorem 5) is tight in several important cases: As was shown in [9], it is tight for certain pairs of Hamming balls, and as was shown in [5], it is tight for a "tribes" subset (see [2]) and its dual. However, we show that the assertion can be tight only if the correlation is relatively small. We use the following claim ([9], Proposition 2.2):

Proposition 19 (Talagrand) *There exists a universal constant K_1 such that for all $A \subset \{0,1\}^n$,*

$$\sum_{i=1}^{n} \hat{1}_A(\{i\})^2 \leq K_1 \mu(A)^2 \log \frac{e}{\mu(A)}.$$

Note that for small values of $\mu(A)$, this upper bound is much stronger than the bound $\mu(A)(1-\mu(A))$ which follows from the Parseval identity.

Proposition 20 *There exists a universal constant K such that for all $A, B \subset \{0,1\}^n$,*

$$\varphi\Big(\sum_{i=1}^{n} \hat{1}_A(\{i\}) \hat{1}_B(\{i\}) \Big) \leq K \mu(A) \mu(B),$$

where $\varphi(x) = x / \log(e/x)$.

Proof By the Cauchy-Schwarz inequality and Proposition 19,

$$\sum_{i=1}^{n} \hat{1}_A(\{i\}) \hat{1}_B(\{i\}) \leq \Big(\sum_{i=1}^{n} \hat{1}_A(\{i\})^2 \Big)^{1/2} \Big(\sum_{i=1}^{n} \hat{1}_B(\{i\})^2 \Big)^{1/2} \leq$$

$$\leq K_1 \mu(A) \mu(B) \Big(\log \frac{e}{\mu(A)} \Big)^{1/2} \Big(\log \frac{e}{\mu(B)} \Big)^{1/2}.$$

By Proposition 17, the function $\varphi(x)$ is monotone increasing in $(0,1]$, and hence

$$\varphi\Big(\sum_{i=1}^{n}\hat{1}_A(\{i\})\hat{1}_B(\{i\})\Big) \leq \varphi\Big(K_1\mu(A)\mu(B)\Big(\log\frac{e}{\mu(A)}\Big)^{1/2}\Big(\log\frac{e}{\mu(B)}\Big)^{1/2}\Big).$$

By the inequality between the arithmetic and the geometric means,

$$K_1\mu(A)\mu(B)\Big(\log\frac{e}{\mu(A)}\Big)^{1/2}\Big(\log\frac{e}{\mu(B)}\Big)^{1/2} \leq \frac{K_1}{2}\mu(A)\mu(B)\Big(\log\frac{e}{\mu(A)}+\log\frac{e}{\mu(B)}\Big). \tag{49}$$

On the other hand, for all $0 < t \leq 1$,

$$\log\frac{e}{t} = 1+\log\frac{1}{t} \leq \frac{1}{t},$$

and hence,

$$\mu(A)\Big(\log\frac{e}{\mu(A)}\Big)^{1/2} \leq (\mu(A))^{1/2},$$

and similarly for B. Therefore,

$$K_1\mu(A)\mu(B)\Big(\log\frac{e}{\mu(A)}\Big)^{1/2}\Big(\log\frac{e}{\mu(B)}\Big)^{1/2} \leq K_1\Big(\mu(A)\mu(B)\Big)^{1/2},$$

and thus,

$$\log\frac{e}{K_1\mu(A)\mu(B)(\log\frac{e}{\mu(A)})^{1/2}(\log\frac{e}{\mu(B)})^{1/2}} \geq \log\frac{e}{K_1(\mu(A)\mu(B))^{1/2}} =$$

$$= \log\frac{1}{K_1}+\frac{1}{2}\log\frac{e}{\mu(A)}+\frac{1}{2}\log\frac{e}{\mu(B)} \geq \frac{1}{4}\Big(\log\frac{e}{\mu(A)}+\log\frac{e}{\mu(B)}\Big). \tag{50}$$

The last inequality holds if $\mu(A)\mu(B)$ is bigger than some constant depending on K_1. This assumption is possible since the assertion of Proposition **??** holds trivially if $\mu(A)$ and $\mu(B)$ are bounded away from zero. Combining Inequalities (49) and (50), we get

$$\varphi\Big(K_1\mu(A)\mu(B)\Big(\log\frac{e}{\mu(A)}\Big)^{1/2}\Big(\log\frac{e}{\mu(B)}\Big)^{1/2}\Big) \leq \frac{\frac{K_1}{2}\mu(A)\mu(B)(\log\frac{e}{\mu(A)}+\log\frac{e}{\mu(B)})}{\frac{1}{4}(\log\frac{e}{\mu(A)}+\log\frac{e}{\mu(B)})} = 2K_1\mu(A)\mu(B).$$

This completes the proof of Proposition 20. ■

Using Proposition 20 we obtain the following:

Corollary 21 *Theorem 5 can be tight only if the correlation between the subsets satisfies*

$$\mu(A\cap B)-\mu(A)\mu(B) \leq K\mu(A)\mu(B),$$

where K is an absolute constant.

It seems challenging to find further improvements of the Harris-Kleitman lemma and of the FKG inequality, which will be tight even if the correlation between the subsets is substantially bigger.

5 Acknowledgements

We are grateful to Ofer Zeitouni for fruitful discussions regarding the improved bounds on large deviations presented in Section 2.2, and to Sergey Bobkov for suggesting that the results from [1] can be useful (as indeed they were). We are also thankful to Dan Romik, Noam Berger, and Noga Alon for helpful comments. Finally, we thank Gil Kalai for numerous useful suggestions.

References

[1] S. G. Bobkov, C. Houdré, and P. Tetali, The Subgaussian Constant and Concentration Inequalities, *Israel J. of Math.*, **156** (2006), pp. 255–283.

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

[3] C.M. Fortuin, P.W. Kasteleyn, and J. Ginibre, Correlation Inequalities on Some Partially Ordered Sets, *Comm. Math. Phys.* **22** (1971), pp. 89–103.

[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] N. Keller, Lower Bound on the Correlation Between Monotone Families in the Average Case, submitted.

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

[7] M. Ledoux and M. Talagrand, Probability in Banach Spaces, Springer-Verlag, 1991.

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

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

[10] M. Talagrand, On Boundaries and Influences, *Combinatorica* **17** (1997), no. 2, pp. 275–285.