Gaussian Bounds for Noise Correlation of Functions and Tight Analysis of Long Codes

Elchanan Mossel
U.C. Berkeley and Weizmann Institute
mossel@stat.berkeley.edu

Abstract

We derive tight bounds on the expected value of products of low influence *functions defined on correlated probability spaces. The proofs are based on extending Fourier theory to an arbitrary number of correlated probability spaces, on a generalization of an invariance principle recently obtained with O'Donnell and Oleszkiewicz for multilinear polynomials with low influences and bounded degree and on properties of multi-dimensional Gaussian distributions.*

Let $(X_i^j : 1 \le i \le k, 1 \le j \le n)$ *be a matrix of random variables whose columns* $X^1, \dots, X^n$ *are independent and identically distributed and such that any two rows* X_i, X_j *for* $1 \le i \ne j \le k$ *are independent. Assume further that the values that row* X_i *takes with non-zero probability are the same no matter how one conditions on the remaining rows* $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k$*. Our results show that given* k *functions* $f_1, \dots, f_k$ *taking values in* $[0,1]$ *it holds that* $|\mathbf{E}[\prod_{i=1}^k f_i(X_i)] - \prod_{i=1}^k \mathbf{E}[f_i(X_i)]| < \epsilon$ *if all influences of the functions* f_i *are smaller than* $\tau(\epsilon, k)$ *which is independent of* n*. In words: low influence functions of pairwise independent rows behave like independent random variables. The general statement of our result applies when the rows are not pairwise independent and when (some) of the variables do not have low influences for (some) functions.*

The results obtained here allow analyzing hyper-graph long-code tests. A number of applications in hardness of approximation assuming the Unique Games Conjecture were obtained using the results derived here in subsequent work by Raghavendra and jointly by Austrin and the author. Our results imply new results on voting schemes in social choice and in additive number theory. In particular we show that among all low influence functions, Majority is asymptotically the most predictable and is (almost) optimal in the context of Condorcet voting.

1 Introduction

1.1 Rough Statement of the Main Result

This paper studies low influence functions $f : \Omega^n \to [0,1]$, where (Ω^n, μ^n) is a product probability space and where the *influence of the* i*th coordinate on* f, denoted by $\mathrm{Inf}_i(f)$ is defined by

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_{x}[\mathop{\mathbf{Var}}_{x_i}[f(x)]]. \tag{1}$$

The study of low influence functions is motivated by applications from the theory of social choice in mathematical economics, from applications in the theory of hardness of approximation in theoretical computer science and from problems in additive number theory. We refer the reader to some recent papers [13, 14, 15, 6, 18, 8] for motivation and general background. The main theorems established here provide tight bounds on the expected value of the product of functions defined on correlated probability spaces. These in turn imply some new results in the theory of social choice and additive number theory type results. Application to hardness of approximation in computer science were derived in subsequent work in [3] and [17].

In our main result we consider a probability measure $\mathbf{P}$ defined on a space $\prod_{i=1}^k \Omega^{(i)}$. Letting $f_i : (\Omega^{(i)})^n \to [0,1], 1 \le i \le k$ be a collection of low influence functions we derive tight bounds on $\mathbf{E}[f_1 \dots f_k]$ in terms of $\mathbf{E}[f_1], \dots, \mathbf{E}[f_k]$ and a measure of correlation between the spaces $\Omega^{(1)}, \dots, \Omega^{(k)}$. The bounds are expressed in terms of extremal probabilities in Gaussian space, that can be calculated in the case $k = 2$. When $k \ge 2$ and $\mathbf{P}$ is a *pairwise independent distribution* our bounds show that $\mathbf{E}[f_1 \dots f_k]$ is close to $\prod_{i=1}^k \mathbf{E}[f_i]$. More formally, for pairwise independent distributions, we show that for every $\epsilon > 0$ there exists a $\tau = \tau(\epsilon) > 0$ such that if $\sup_{i,j} I_j(f_i) \le \tau(\epsilon)$ then $|\mathbf{E}[f_1 \dots f_k] - \prod_{i=1}^k \mathbf{E}[f_i]| < \epsilon$. For the results to hold we need that a measure of correlation between the spaces $\Omega^{(1)}, \dots, \Omega^{(k)}$ denoted $\rho(\Omega^{(1)}, \dots, \Omega^{(k)}; \mathbf{P})$ to be strictly less than 1. The later condition holds whenever the support of $\mathbf{P}$ is *connected*: for every $x, y \in \prod_{i=1}^{k} \Omega^{(i)}$ such that $\mathbf{P}[x] > 0$ and $\mathbf{P}[y] > 0$ there exists a path $x = x_0, x_1, \ldots, x_r = y \in \prod_{i=1}^{k} \Omega^{(i)}$ such that $\mathbf{P}[x_i] > 0$ for all $1 \leq i \leq r$ and x_i, x_{i+1} differ at one coordinate only for all i. In particular, it holds when $\mathbf{P}$ has full support.

We also apply a simple recursive argument in order to obtain results for general functions not necessarily of low influences. The results show that the bounds for low influence functions hold for general functions after the functions have been "modified" in a bounded number of coordinates.

The rest of the abstract is devoted to discussing motivation and related results from hardness of approximation and from social choice followed by definitions and basic properties of correlated spaces and the statement of the main technical results. The main steps of the proof are briefly outlined while the full proof can be found on the Arxiv at http://arxiv.org/abs/math/0703683.

1.2 Analysis of Long Codes Under Unique Games

A common feature of a number of hardness of approximation results proven under the Unique Games Conjecture (UCG) [12] is their long-code analysis. For 2-query reductions, what is often needed in the long-code analysis is to show that if f and g are two long codes, and $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$ are two correlated inputs (so that x_i and y_i are correlated but different pairs (x_i, y_i) are independent) then "if $\mathbf{E}[f(x)g(y)]$ is large then f and g have a common variable with high influence". See e.g. [13, 2].

The main results of the current paper allow to perform a similar type of analysis for k-query PCPs. Now we look at k correlated vectors $x_1 = (x_1^1, \ldots, x_1^n), \ldots, x_k = (x_k^1, \ldots, x_k^n)$ and the main results show that if $\mathbf{E}[f_1(x_1) \ldots\ldots f_k(x_k)]$ is "large", then there are at least 2 functions that share a large influence variable. The results have the nicest formulation when the vectors x_i are pairwise independent in which case the meaning of "large" is that $\mathbf{E}[\prod_{i=1}^{k} f_i(x_i)]$ is significantly larger/smaller than $\prod_{i=1}^{k} \mathbf{E}[f_i(x_i)]$.

When the first draft of this paper was circulated, we did not know of any concrete application of the results to hardness of approximation. However, since then the results have been already applied at least twice. The two applications are discussed below.

1.3 Subsequent Work And Applications in Hardness of Approximation

Subsequently to posting a draft of this paper on the Arxiv, two applications of our results to hardness of approximation have been established. Both results are in the context of the Unique Games conjecture in computational complexity [12]. Furthermore, both results consider the problem of solving *constraint satisfaction problems (CSP)*.

Given a predicate $P : [q]^k \to \{0, 1\}$, where $[q] = \{1, \ldots, q\}$ for some integer q, we define MAX CSP(P) to be the algorithmic problem where we are given a set of variables $x_1, \ldots, x_n$ taking values in $[q]$ and a set of constraints of the form $P(l_1, \ldots, l_k)$, where each $l_i = x_j + a$, where x_j is one of the variables and $a \in [q]$ is a constant (addition is mod q). More generally, in the problem of MAX k-CSP$_q$ we are given set of constraints each involving k of the variables $x_1, \ldots, x_n$. The most well studied case is the case of $q = 2$ denoted MAX k-CSP.

The objective is to find an assignment to the variables satisfying as many of the constraints as possible. The problem of MAX k-CSP$_q$ is NP-hard for any $k \geq 2, q \geq 2$, and as a consequence, a large body of research was devoted to studying how well the problem can be approximated. We say that a (randomized) algorithm has *approximation ratio* α if, for all instances, the algorithm is guaranteed to find an assignment which (in expectation) satisfies at least $\alpha \cdot \mathrm{Opt}$ of the constraints, where Opt is the maximum number of simultaneously satisfied constraints, over any assignment.

The results of [3] assume the Unique Games Conjecture and consider any predicate P for which there exists a pairwise independent distribution over $[q]^k$ with uniform marginals whose support is contained in $P^{-1}(1)$. The results prove that such a predicate is approximation resilient. In other words, there is no polynomial time algorithm which achieves a better approximation factor than assigning the variables at random. This result imply in turn that for general $k \geq 3$ and $q \geq 2$, the MAX k-CSP$_q$ problem is UG-hard to approximate within $\mathcal{O}(kq^2)/q^k + \epsilon$. Moreover, for the special case of $q = 2$, i.e., boolean variables, it gives hardness of $(k + \mathcal{O}(k^{0.525}))/2^k + \epsilon$, improving upon the best previous bound [18] of $2k/2^k + \epsilon$ by essentially a factor 2. Finally, again for $q = 2$, assuming that the famous Hadamard Conjecture is true, the results are further improved, and the $\mathcal{O}(k^{0.525})$ term can be replaced by the constant 4.

These results should be compared to prior work by Samordnitsky and Trevisan [18] who using the Gowers norm, proved that the MAX k-CSP problem has a hardness factor of $2^{\lceil \log_2 k+1 \rceil}/2^k$, which is $(k+1)/2^k$ for $k = 2^r - 1$, but can be as large as $2k/2^k$ for general k.

From the quantitative point of view [3] give stronger stronger hardness than [18] for MAX k-CSP$_q$, even in the already thoroughly explored $q = 2$ case. These improvements may seem small, being an improvement only by a multiplicative factor 2. However, it is well known that it is impossible to get non-approximability results which are better than $(k+1)/2^k$, and thus, in this respect, the hardness of $(k+4)/2^k$ assuming the Hadamard Conjecture is in fact optimal to within a very small *additive* factor. Also, the results of [3] give approximation resistance of MAX CSP(P) for a much larger variety of predicates (any P containing a balanced pairwise independent distribution).

From a qualitative point of view, the analysis of [3] is very direct. Furthermore, it is general enough to accommodate any domain $[q]$ with virtually no extra effort. Also, their proof using the main result of the current paper, i.e., bounds on expectations of products under certain types of correlation, putting it in the same general framework as many other UGC-based hardness results, in particular those for 2-CSPs.

In a second beautiful result by Raghavendra [17] the results of the current paper were used to obtained very general hardness results for MAX CSP(P). In [17] it is shown that for every predicate P and for every approximation factor which is smaller than the UG-hardness of the problem, there exists a polynomial time algorithm which achieves this approximation ratio. Thus for every P the UG-hardness of MAX CSP(P) is sharp. The proof of the results uses the results obtained here in order to define and analyze the reduction from UG given the integrality gap of the corresponding convex optimization problem. We note that for most predicates the UG hardness of MAX CSP(P) is unknown and therefore the results of [17] complement those of [3].

1.4 Majority is Most Predictable

We now consider the first social choice application. Suppose n voters are to make a binary decision. Assume that the outcome of the vote is determined by a *social choice* function $f : \{-1,1\}^n \to \{-1,1\}$, so that the outcome of the vote is $f(x_1,\ldots,x_n)$ where $x_i \in \{-1,1\}$ is the vote of voter i. We assume that the votes are independent, each ± 1 with probability $\frac{1}{2}$. It is natural to assume that the function f satisfies $f(-x) = -f(x)$, i.e., it does not discriminate between the two candidates. Note that this implies that $\mathbf{E}[f] = 0$ under the uniform distribution. A natural way to try and predict the outcome of the vote is to sample a subset of the voters, by sampling each voter independently with probability ρ. Conditioned on a vector X of votes the distribution of Y, the sampled votes, is i.i.d. where $Y_i = X_i$ with probability ρ and $Y_i = *$ (for unknown) otherwise.

Conditioned on $Y = y$, the vector of sampled votes, the optimal prediction of the outcome of the vote is given by $\mathrm{sgn}((Tf)(y))$ where

$$(Tf)(y) = \mathbf{E}[f(X)|Y = y]. \quad (2)$$

This implies that the probability of correct prediction (also called predictability) is given by

$$\mathbf{P}[f = \mathrm{sgn}(Tf)] = \frac{1}{2}(1 + \mathbf{E}[f\,\mathrm{sgn}(Tf)]).$$

For example, when $f(x) = x_1$ is the dictator function, we have $\mathbf{E}[f\,\mathrm{sgn}(Tf)] = \rho$ corresponding to the trivial fact that the outcome of the election is known when voter 1 is sampled and are ± 1 with probability $1/2$ otherwise. The notion of predictability is natural in statistical contexts. It was also studied in [16].

In the first application presented here we show that

Theorem 1.1 ("Majority Is Most Predictable") *Let $0 \le \rho \le 1$ and $\epsilon > 0$ be given. Then there exists $\tau > 0$ such that if $f : \{-1,1\}^n \to [-1,1]$ satisfies $\mathbf{E}[f] = 0$ and $\mathrm{Inf}_i(f) \le \tau$ for all i, then*

$$\mathbf{E}[f\,\mathrm{sgn}(Tf)] \le \tfrac{2}{\pi}\arcsin\sqrt{\rho} + \epsilon, \quad (3)$$

where T is defined in (2).

Moreover, it follows from the central limit theorem that if $\mathrm{Maj}_n(x_1,\ldots,x_n) = \mathrm{sgn}(\sum_{i=1}^n x_i)$, then

$$\lim_{n\to\infty} \mathbf{E}[\mathrm{Maj}_n \mathrm{sgn}(T\mathrm{Maj}_n)] = \tfrac{2}{\pi}\arcsin\sqrt{\rho}.$$

Remark 1.2 *Note that Theorem 1.1 proves a weaker statement than showing that Majority is the most predictable function. The statement only asserts that if a function has low enough influences than its predictability cannot be more than ϵ larger than the asymptotic predictability value achieved by the majority function when the number of voters $n \to \infty$. This slightly inaccurate title of the theorem is inline with previous result such as the "Majority is Stablest Theorem" (see below). Similar language may be used later when informally discussing statements of various theorems.*

Remark 1.3 *One may wonder if for a finite n, among* all *functions $f : \{-1,1\}^n \to \{-1,1\}$ with $\mathbf{E}[f] = 0$, majority is the most predictable function. Note that the predictability of the dictator function $f(x) = x_1$ is given by ρ, and $\frac{2}{\pi}\arcsin\sqrt{\rho} > \rho$ for $\rho \to 0$. Therefore when ρ is small and n is large the majority function is more predictable than the dictator function. However, note that when $\rho \to 1$ we have $\rho > \frac{2}{\pi}\arcsin\sqrt{\rho}$ and therefore for values of ρ close to 1 and large n the dictator function is more predictable than the majority function.*

We note that the bound obtained in Theorem 1.1 is a reminiscent of the Majority is Stablest theorem [14, 15] as both involve the arcsin function. However, the two theorems are quite different. The Majority is Stablest theorem asserts that under the same condition as in Theorem 1.1 it holds that

$$\mathbf{E}[f(X)f(Y)] \leq \frac{2}{\pi}\arcsin\rho + \epsilon.$$

where $(X_i, Y_i) \in \{-1,1\}^2$ are i.i.d. with $\mathbf{E}[X_i] = \mathbf{E}[Y_i] = 0$ and $\mathbf{E}[X_i Y_i] = \rho$. Thus "Majority is Stablest" considers two correlated voting vectors, while "Majority is Most Predictable" considers a sample of one voting vector. In fact, both results follow from the more general invariance principle presented here. We note a further difference between stability and predictability: It is well known that in the context of "Majority is Stablest", for *all* $0 < \rho < 1$, among all boolean functions with $\mathbf{E}[f] = 0$ the maximum of $\mathbf{E}[f(x)f(y)]$ is obtained for dictator functions of the form $f(x) = x_i$. As discussed above, for ρ close to 0 and large n, the dictator is less predictable than the majority function.

We also note that the "Ain't over until it's over" Theorem [14, 15] provides a bound under the same conditions on

$$P[Tf > 1-\delta],$$

for small δ. However, this bound is not tight and does not imply Theorem 1.1. Similarly, Theorem 1.1 does not imply the "Ain't over until it's over" theorem. The bounds in "Ain't Over Until It's Over" were derived using invariance of Tf while the bound (3) requires the joint invariance of f and Tf.

1.5 Condorcet Paradoxes

The second social choice application has to do with Condorcet voting. Suppose n voters rank k candidates. It is assumed that each voter i has a linear order $\sigma_i \in S(k)$ on the candidates. In *Condorcet voting*, the rankings are aggregated by deciding for each pair of candidates which one is superior among the n voters. The question of properties of Condorcet aggregation was first discussed by Marie Jean Antoine Nicolas de Caritat, marquis de Condorcet in the 18th century. Since then the problem was studied extensively in economics with a major contribution by Arrow [1]. The quantitative study of the problem was initiated by Kalai [10] who was the first to state (a special case of) the Majority is Stablest conjecture.

More formally, the aggregation results in a tournament G_k on the set $[k]$. Recall that G_k is a *tournament* on $[k]$ if it is a directed graph on the vertex set $[k]$ such that for all $a, b \in [k]$ either $(a > b) \in G_k$ or $(b > a) \in G_k$. Given individual rankings $(\sigma_i)_{i=1}^n$ the tournament G_k is defined as follows.

Let $x^{a>b}(i) = 1$, if $\sigma_i(a) > \sigma_i(b)$, and $x^{a>b}(i) = -1$ if $\sigma_i(a) < \sigma_i(b)$. Note that $x^{b>a} = -x^{a>b}$.

The binary decision between each pair of candidates is performed via a anti-symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ so that $f(-x) = -f(x)$ for all $x \in \{-1,1\}$. The tournament $G_k = G_k(\sigma; f)$ is then defined by letting $(a > b) \in G_k$ if and only if $f(x^{a>b}) = 1$.

Note that there are $2^{\binom{k}{2}}$ tournaments while there are only $k! = 2^{\Theta(k \log k)}$ linear rankings. For the purposes of social choice, some tournaments make more sense than others.

Definition 1.4 *We say that a tournament G_k is* linear *if it is acyclic. We will write* $\mathrm{Acyc}(G_k)$ *for the logical statement that G_k is acyclic. Non-linear tournaments are often referred to as non-rational in economics as they represent an order where there are 3 candidates a, b and c such that a is preferred to b, b is preferred to c and c is preferred to a.*

We say that the tournament G_k is a unique max tournament *if there is a candidate $a \in [k]$ such that for all $b \neq a$ it holds that $(a > b) \in G_k$. We write* $\mathrm{UniqMax}(G_k)$ *for the logical statement that G_k has a unique max. Note that the unique max property is weaker the linearity. It corresponds to the fact that there is a candidate that dominates all other candidates.*

Following [11, 10], we consider the probability distribution over n voters, where the voters have independent preferences and each one chooses a ranking uniformly at random among all $k!$ orderings. Note that the marginal distributions on vectors $x^{a>b}$ is the uniform distribution over $\{-1,1\}^n$ and that if $f : \{-1,1\}^n \to \{-1,1\}$ is anti-symmetric then $\mathbf{E}[f] = 0$.

The case that is now understood is $k = 3$. Note that in this case G_3 is unique max if and only if it is linear. Kalai [10] studied the *probability* of a rational outcome given that the n voters vote independently and at random from the 6 possible rational rankings. He showed that the probability of a rational outcome in this case may be expressed as $\frac{3}{4}(1 + \mathbf{E}[fTf])$ where T is the *Bonami-Beckner* operator with parameter $\rho = 1/3$. The Bonami-Beckner operator may be defined as follows. Let $(X_i, Y_i) \in \{-1,1\}^2$ be i.i.d. with $\mathbf{E}[X_i] = \mathbf{E}[Y_i] = 0$ and $\mathbf{E}[X_i Y_i] = \rho$ for $1 \leq i \leq n$. For $f : \{-1,1\}^n \to \mathbb{R}$, and $x \in \{-1,1\}^n$, the Bonami-Beckner operator T applied to f is defined via

$$(Tf)(x) = \mathbf{E}[f(Y)|X = x],$$

where $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$.

It is natural to ask which function f with small influences is most likely to produce a rational outcome. Instead of considering small influences, Kalai considered the essentially stronger assumption that f is monotone and "transitive-symmetric"; i.e., that for all $1 \le i < j \le n$ there exists a permutation σ on $[n]$ with $\sigma(i) = j$ such that $f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for all $(x_1, \ldots, x_n)$. Kalai conjectured that as $n \to \infty$ the maximum of $\frac{3}{4}(1 + \mathbf{E}[fTf])$ among all transitive-symmetric functions approaches the same limit as $\lim_{n\to\infty} \frac{3}{4}(1 + \mathbf{E}[\mathrm{Maj}_n T \mathrm{Maj}_n])$. This was proven using the Majority is Stablest Theorem [14, 15]. Here we obtain similar results for any value of k. Our result is not tight, but almost tight. More specifically we show that:

Theorem 1.5 ("Majority is best for Condorcet") *Consider Condorcet voting on k candidates. Then for all $\epsilon > 0$ there exists $\tau = \tau(k, \epsilon) > 0$ such that if $f : \{-1,1\}^n \to \{-1,1\}$ is anti-symmetric and $\mathrm{Inf}_i(f) \le \tau$ for all i, then*

$$\mathbf{P}[\mathrm{UniqMax}(G_k(\sigma; f))] \le k^{-1+o_k(1)} + \epsilon. \quad (4)$$

Moreover for $f = \mathrm{Maj}_n$ we have $\mathrm{Inf}_i(f) \le O(n^{-1/2})$ and it holds that

$$\mathbf{P}[\mathrm{UniqMax}(G_k(\sigma; f))] \ge k^{-1-o_k(1)} - o_n(1). \quad (5)$$

Interestingly, we are not able to derive similar results for Acyc. We do calculate the probability that Acyc holds for majority.

Proposition 1.6 *We have*

$$\lim_{n\to\infty} \mathbf{P}[\mathrm{Acyc}(G_k(\sigma; \mathrm{Maj}_n))] = \exp(-\Theta(k^{5/3})). \quad (6)$$

We note that results in economics [4] have shown that for majority vote the probability that the outcome will contain a Hamiltonian cycle when the number of voters goes to infinity is $1 - o_k(1)$.

1.6 Hyper Graph and Additive Applications

We briefly discuss some applications to Hyper Graphs and Additive Combinatorics. For details and the formal statement of the result see the full paper. We let Ω be a finite set equipped with the uniform probability measure. We let $R \subset \Omega^k$ denote a k-wise relation. For sets $A_1, \ldots, A_k \subset \Omega^n$ we will be interested in the number of k-tuples $x_1 \in A_1, \ldots, x_k \in A_k$ satisfying the relation R in all coordinates, i.e. $(x_1^i, \ldots, x_k^i) \in R$ for all i.

An application of the main results of the paper shows that if the uniform measure on the support of R defines a pairwise independent distribution, the support of R is connected (in the same meaning as before) and if the influences of the sets A_i are all at most $\tau = \tau(\epsilon)$ then the number of k-tuples $x_1 \in A_1, \ldots, x_k \in A_k$ satisfying the relation R in all coordinates is

$$|R|^n \prod_{i=1}^{k} \frac{|A_i|}{|\Omega^n|} \quad (7)$$

up to an error of

$$\epsilon |R|^n. \quad (8)$$

In other words, up to error terms, the number of times the relation is satisfied for low influence sets is the same as the expected value of the number of times it is satisfied for random sets of the same size.

Moreover, we also obtain results for general sets of arbitrary influences. Let $A \subset \Omega^n$ and $S \subset [n]$.Let

$$\overline{A}_S = \{y : \exists x \in A, x_{[n]\setminus S} = y_{[n]\setminus S}\},$$

Then we show that for arbitrary sets and for all $\epsilon > 0$, there exists a set $S \subset [n]$ of size at most $O(1/\tau(\epsilon))$ such that (7) and (8) hold for $\overline{A}_1^S, \ldots, \overline{A}_k^S$. More details and quantitative bounds can be found at the full version of the paper.

1.7 Basic Setup: Correlated Spaces

A central concept that is extensively studied and repeatedly used in the paper is that of correlated probability spaces. We begin by defining notions of correlation between probability spaces. We will later show how to relate these notion to noise operators.

Definition 1.7 *Given a probability measure $\mathbf{P}$ defined on $\prod_{i=1}^k \Omega^{(i)}$, we say that $\Omega^{(1)}, \ldots, \Omega^{(k)}$ are* correlated spaces. *For $A \subset \Omega^{(i)}$ we let*

$$\mathbf{P}[A] = \mathbf{P}[(\omega_1, \ldots, \omega_k) \in \prod_{j=1}^{k} \Omega^{(j)} : \omega_i \in A],$$

and similarly $\mathbf{E}[f]$ for $f : \Omega^{(i)} \to \mathbb{R}$. We will abuse notation by writing $\mathbf{P}[A]$ for $\mathbf{P}^n[A]$ for $A \subset (\prod_{i=1}^k \Omega^{(i)})^n$ or $A \subset (\Omega^{(i)})^n$ and similarly for $\mathbf{E}$.

Definition 1.8 *Given two linear subspaces A and B of $L^2(\mathbf{P})$ we define the* correlation *$\rho(A, B; \mathbf{P}) = \rho(A, B)$ between A and B by*

$$\rho(A, B) = \sup\{\frac{\mathbf{Cov}[f, g]}{\sqrt{\mathbf{Var}[f]\mathbf{Var}[g]}} : f \in A, g \in B\}. \quad (9)$$

Let $\Omega = (\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$. We define the correlation *$\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ by letting:*

$$\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P}) = \rho(L^2(\Omega^{(1)}, \mathbf{P}), L^2(\Omega^{(2)}, \mathbf{P}); \mathbf{P}). \quad (10)$$

More generally, let $\Omega = (\prod_{i=1}^k \Omega^{(i)}, \mathbf{P})$ and for a subset $S \subset [k]$, write $\Omega^{(S)} = \prod_{i\in S} \Omega^{(i)}$. The correlation vector

$\underline{\rho}(\Omega^{(1)}, \ldots, \Omega^{(k)})$ *is a length* $k-1$ *vector whose* i*'th coordinate is given by*

$$\underline{\rho}(i) = \rho(\prod_{j=1}^{i} \Omega^{(j)}, \prod_{j=i+1}^{k} \Omega^{(j)}),$$

for $1 \leq i \leq k-1$. *The* correlation $\rho(\Omega^{(1)}, \ldots, \Omega^{(k)})$ *is defined by letting:*

$$\rho(\Omega^{(1)}, \ldots, \Omega^{(k)}) = \max_{1 \leq i \leq k} \rho(\prod_{j=1}^{i-1} \Omega^{(j)} \times \prod_{j=i+1}^{k} \Omega^{(j)}, \Omega^{(i)}). \tag{11}$$

Remark 1.9 *It is easy to see that* $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ *is the* second singular value *of the conditional expectation operator mapping* $f \in L^2(\Omega^{(2)}, \mathbf{P})$ *to* $g(x) = \mathbf{E}[f(Y)|X = x] \in L^2(\Omega^{(1)}, \mathbf{P})$. *Thus* $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ *is the second singular value of the matrix corresponding to the operator* T *with respect to orthonormal basis of* $L^2(\Omega^{(1)}, \mathbf{P})$ *and* $L^2(\Omega^{(2)}, \mathbf{P})$.

Definition 1.10 *Given* $(\prod_{i=1}^{k} \Omega^{(i)}, \mathbf{P})$, *we say that* $\Omega^{(1)}, \ldots, \Omega^{(k)}$ *are* r-wise independent *if for all* $S \subset [k]$ *with* $|S| \leq r$ *and for all* $\prod_{i \in S} A_i \subset \prod_{i \in S} \Omega^{(i)}$ *it holds that*

$$\mathbf{P}[\prod_{i \in S} A_i] = \prod_{i \in S} \mathbf{P}[A_i].$$

1.8 Gaussian Stability

Our main result states bounds in terms of Gaussian stability measures which we discuss next. Let γ be the one dimensional Gaussian measure.

Definition 1.11 *Given* $\mu \in [0, 1]$, *define* $\chi_\mu : \mathbb{R} \to \{0, 1\}$ *to be the indicator function of the interval* $(-\infty, t]$, *where* t *is chosen so that* $\mathbf{E}_\gamma[\chi_\mu] = \mu$. *Explicitly,* $t = \Phi^{-1}(\mu)$, *where* Φ *denotes the distribution function of a standard Gaussian. Furthermore, define*

$$\overline{\Gamma}_\rho(\mu, \nu) = \mathbf{P}[X \leq \Phi^{-1}(\mu)\, ,\, Y \leq \Phi^{-1}(\nu)],$$

$$\underline{\Gamma}_\rho(\mu, \nu) = \mathbf{P}[X \leq \Phi^{-1}(\mu)\, ,\, Y \geq \Phi^{-1}(1-\nu)],$$

where (X, Y) *is a two dimensional Gaussian vector with covariance matrix* $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$

Given $(\rho_1, \ldots, \rho_{k-1}) \in [0, 1]^{k-1}$ *and* $(\mu_1, \ldots, \mu_k) \in [0, 1]^k$ *for* $k \geq 3$ *we define by induction*

$$\overline{\Gamma}_{\rho_1, \ldots, \rho_{k-1}}(\mu_1, \ldots, \mu_k) = \overline{\Gamma}_{\rho_1}(\mu_1, \overline{\Gamma}_{\rho_2, \ldots, \rho_{k-1}}(\mu_2, \ldots, \mu_k)),$$

and similarly $\underline{\Gamma}()$.

1.9 Statements of main results

We now state our main results. We state the results both for low influence functions and for general functions. For general functions it is useful to define the following notions:

Definition 1.12 *Let* $f : \Omega^n \to \mathbb{R}$ *and* $S \subset [n]$. *We define*

$$\overline{f}^S(x) = \sup(f(y) : y_{[n]\setminus S} = x_{[n]\setminus S}),$$

$$\underline{f}^S(x) = \inf(f(y) : y_{[n]\setminus S} = x_{[n]\setminus S}).$$

Theorem 1.13 *Let* $(\prod_{j=1}^{k} \Omega^{(j)}, \mathbf{P})$ *be a finite probability space such that the minimum probability of any atom in* $\prod_{j=1}^{k} \Omega_i^{(j)}$ *is at least* α. *Assume furthermore that there exists* $\underline{\rho} \in [0, 1]^{k-1}$ *and* $0 \leq \rho < 1$ *such that for all* j:

$$\rho(\Omega^{(1)}, \ldots, \Omega^{(k)}; \mathbf{P}) \leq \rho,$$
$$\rho(\Omega^{(\{1,\ldots,j\})}, \Omega^{(\{j+1,\ldots,k\})}; \mathbf{P}) \leq \underline{\rho}(j) \tag{12}$$

Then for all $\epsilon > 0$ *there exists* $\tau > 0$ *such that if*

$$f_j : \left(\Omega^{(j)}\right)^n \to [0, 1].$$

for $1 \leq j \leq k$ *satisfy*

$$\max_{i,j}(\mathrm{Inf}_i(f_j)) \leq \tau \tag{13}$$

then

$$\mathbf{E}[\prod_{j=1}^{k} f_j] \geq \underline{\Gamma}_{\underline{\rho}}(\mathbf{E}[f_1], \ldots, \mathbf{E}[f_k]) - \epsilon \tag{14}$$

$$\mathbf{E}[\prod_{j=1}^{k} f_j] \leq \overline{\Gamma}_{\underline{\rho}}(\mathbf{E}[f_1], \ldots, \mathbf{E}[f_k]) + \epsilon. \tag{15}$$

If we instead of (12) we assume that for all $j \neq j'$:

$$\rho(\Omega^{(j)}, \Omega^{(j')}; \mathbf{P}) = 0, \tag{16}$$

then

$$\prod_{j=1}^{k} \mathbf{E}[f_j] - \epsilon \leq \mathbf{E}[\prod_{j=1}^{k} f_j] \leq \prod_{j=1}^{k} \mathbf{E}[f_j] + \epsilon. \tag{17}$$

One may take

$$\tau = \epsilon^{O(\frac{\log(1/\epsilon)\log(1/\alpha)}{(1-\rho)\epsilon})}.$$

The result above also holds for functions with low-degree low influences as is often needed in hardness of approximation.

A truncation argument allows to relax the conditions on the influences.

Proposition 1.14 *For statement (14) to hold in the case where $k = 2$ it suffices to require that*

$$\max_i(\min(\mathrm{Inf}_i(f_1), \mathrm{Inf}_i(f_2))) \le \tau \qquad (18)$$

instead of (13).

In the case where for each i the spaces $\Omega_i^{(1)}, \ldots, \Omega_i^{(k)}$ are s-wise independent, for statement (17) to hold it suffices to require that for all i

$$|\{j : \mathrm{Inf}_i(f_j) > \tau\}| \le s. \qquad (19)$$

An easy recursive argument allows to conclude the following result that does not require low influences (13).

Proposition 1.15 *Consider the setting of Theorem 1.13* without *the assumptions on low influences (13).*

Assuming (12), there exists a set S of size $O(1/\tau)$ such that the functions $\overline{f}_j^S$ satisfy

$$\mathbf{E}[\prod_{j=1}^{k} \overline{f}_j^S] \ge \overline{\Gamma}_{\underline{\rho}}(\mathbf{E}[\overline{f}_1^S], \ldots, \mathbf{E}[\overline{f}_k^S]) - \epsilon$$

and similarly for $\underline{f}_j^S$. Assuming (16), we have

$$\mathbf{E}[\prod_{j=1}^{k} \overline{f}_j^S] \ge \prod_{j=1}^{k} \mathbf{E}[\overline{f}_j^S] - \epsilon \ge \prod_{j=1}^{k} \mathbf{E}[f_j] - \epsilon,$$

and similarly for $\underline{f}$.

1.10 Road Map

Let us review some of the main techniques we use in this paper.

- We develop a Fourier theory on correlated spaces in Section 2. Previous work considered Fourier theory on one product space and reversible operators with respect to that space [6]. Our results here allow to study non-reversible operators which in turn allows to study products of k correlated spaces. An important fact we prove that is used repeatedly is that general noise operators respect "Efron-Stein" decomposition. This fact in particular allows to "truncate" functions to their low degree parts when considering the expected value of the product of functions on correlated spaces.

- In order to derive an invariance principle we need to extend the approach of [14, 15] to prove the joint invariance of a number of multi-linear polynomials. The proof of the extension appears in the full version of the paper. The proof follows the same main steps as in [14, 15] but requires a number of adaptations.

- In the Gaussian realm, we need to extend Borell's isoperimetric result [5] both in the case of two collections of Gaussians and in the case of $k > 2$ collections. This is done in the full version of the paper.

- The proof of the main result, Theorem 1.13 follows in the full paper. The proof of the extensions given in Proposition 1.14 uses a truncation argument for which s-wise independence plays a crucial role. The proof of Proposition 1.15 is based on a simple "weak regularity" argument.

- In the full version what follows is an application of the noise bounds in order to derive the social choice results and the hyper-graph and additive results. In Section 3 we prove the easiest among the application, i.e., "Majority is Most Predictable".

1.11 Acknowledgments

The research reported here is supported by a Sloan fellowship in Mathematics, by BSF grant 2004105, NSF Career Award (DMS 054829) and by ONR award N00014-07-1-0506. Part of this work was carried out while the author was visiting IPAM.

I would like to thank Noam Nisan for suggesting that generalization of the invariance principle should be useful in the social choice context and Gil Kalai for pointing out some helpful references. I would also like to thanks Terence Tao for helpful discussions and references on additive number theory and Szemeredi regularity. Thanks to Per Austrin, Nick Crawford, Marcus Issacson and Ryan O'Donnell for a careful reading of a draft of this paper. Finally, many thanks to an anonymous referee for many helpful comments.

2 Correlated Spaces and Noise

In this section we define and study the notion of correlated spaces and noise operators in a general setting. The results derived here extend Fourier theory to general correlated spaces and may be useful in a number of problems.

2.1 Correlated Probability Spaces and Markov Operators

We begin by defining noise operators and giving some basic examples.

Definition 2.1 *Let $(\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$ be two correlated spaces. The* Markov Operator *associated with $(\Omega^{(1)}, \Omega^{(2)})$ is the operator mapping $f \in L^p(\Omega^{(2)}, \mathbf{P})$ to $Tf \in L^p(\Omega^{(1)}, \mathbf{P})$ by:*

$$(Tf)(x) = \mathbf{E}[f(Y)|X = x],$$

for $x \in \Omega^{(1)}$ and where $(X,Y) \in \Omega^{(1)} \times \Omega^{(2)}$ is distributed according to $\mathbf{P}$.

Example 2.2 *In order to define* Bonami-Beckner *operator $T = T_\rho$ on a space (Ω, μ), consider the space $(\Omega \times \Omega, \nu)$ where $\nu(x,y) = (1-\rho)\mu(x)\mu(y) + \rho\delta(x=y)\mu(x)$, where $\delta(x=y)$ is the function on $\Omega \times \Omega$ which takes the value 1 when $x = y$, and 0 otherwise. In this case, the operator T satisfies:*

$$(Tf)(x) = \mathbf{E}[f(Y)|X = x], \tag{20}$$

where the conditional distribution of Y given $X = x$ is $\rho\delta_x + (1-\rho)\mu$, where δ_x is the delta measure on x.

Remark 2.3 *The construction above may be generalized as follows. Given any Markov chain on Ω that is reversible with respect to μ, we may look at the measure ν on $\Omega \times \Omega$ defined by the Markov chain. In this case T is the Markov operator determined by the chain. The same construction applies under the weaker condition that T has μ as its stationary distribution.*

It is straightforward to verify that:

Proposition 2.4 *Suppose that for each $1 \leq i \leq n$, $(\Omega_i^{(1)} \times \Omega_i^{(2)}, \mu_i)$ are correlated spaces and T_i is the Markov operator associated with $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$. Then $(\prod_{i=1}^n \Omega_i^{(1)}, \prod_{i=1}^n \Omega_i^{(2)}, \prod_{i=1}^n \mu_i)$ defines two correlated spaces and the Markov operator T associated with them is given by $T = \otimes_{i=1}^n T_i$.*

Example 2.5 *For product spaces $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mu_i)$, the Bonami-Beckner operator $T = T_\rho$ is defined by*

$$T = \otimes_{i=1}^n T_\rho^i, \tag{21}$$

where T^i is the Bonami-Beckner operator on $(\Omega_i \times \Omega_i, \mu_i)$. This Markov operator is the one most commonly discussed in previous work, see e.g. [9, 13, 15]. In a more recent work [6] the case of $\Omega_i \times \Omega_i$ with T_i a reversible Markov operator with respect to a measure μ_i on Ω_i was studied.

Example 2.6 *In the context of the Majority is most Predictable Theorem 1.1, the underlying space is $\Omega = \{\pm 1\} \times \{0, \pm 1\}$ where element $(x,y) \in \Omega$ corresponds to a voter with vote x and a sampled vote y where either $y = x$ if the vote is queried or $y = 0$ otherwise.*

Note that the marginal distributions on $\Omega_S = \{0, \pm 1\}$ and $\Omega_V = \{\pm 1\}$ are given by

$$\mu = (1-\rho)\delta_0 + \frac{\rho}{2}(\delta_{-1} + \delta_1), \;\; \nu = \frac{1}{2}(\delta_{-1} + \delta_1),$$

and

$$\nu(\cdot | \pm 1) = \delta_{\pm 1}, \;\; \nu(\cdot | 0) = \frac{1}{2}(\delta_1 + \delta_{-1}).$$

Given independent copies μ_i of μ and ν_i of ν, the measure $\mu = \otimes_{i=1}^n \mu_i$ corresponds to the distribution of a sample of voters where each voter is sampled independently with probability ρ and the distribution of the voters is given by $\nu = \otimes_{i=1}^n \nu_i$.

Example 2.7 *The second non-reversible example is natural in the context of Condorcet voting. For simplicity, we first discuss the case of 3 possible outcomes.*

Let τ denote the uniform measure on the set permutations on the set $[3]$ denoted $S_{[3]}$. Note that each element $\sigma \in S_{[3]}$ defines an element $f \in \{-1,1\}^{\binom{3}{2}}$ by letting $f(i,j) = \mathrm{sgn}(\sigma(i) - \sigma(j))$. The measure so defined, defines 3 correlated probability spaces $(\{\pm 1\}^{\binom{3}{2}}, \mathbf{P})$.

Note that the projection of $\mathbf{P}$ to each coordinate is uniform and

$$\mathbf{P}(f(3,1) = -1 | f(1,2) = f(2,3) = 1) = 0,$$

$$\mathbf{P}(f(3,1) = 1 | f(1,2) = f(2,3) = -1) = 0,$$

and

$$\mathbf{P}(f(3,1) = \pm 1 | f(1,2) \neq f(2,3)) = 1/2.$$

2.2 Properties of Correlated Spaces and Markov Operators

Here we derive properties of correlated spaces and Markov operators that will be repeatedly used below. For the proof see the full version of the paper.

Lemma 2.8 *Let $(\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$ be two correlated spaces. Let f be a $\Omega^{(2)}$ measurable function with $\mathbf{E}[f] = 0$, and $\mathbf{E}[f^2] = 1$. Then among all g that are $\Omega^{(1)}$ measurable satisfying $\mathbf{E}[g^2] = 1$, a maximizer of $|\mathbf{E}[fg]|$ is given by*

$$g = \frac{Tf}{\sqrt{\mathbf{E}[(Tf)^2]}}, \tag{22}$$

where T is the Markov operator associated with $\Omega^{(1)}, \Omega^{(2)}$. Moreover,

$$|\mathbf{E}[gf]| = \frac{|\mathbf{E}[fTf]|}{\sqrt{\mathbf{E}[(Tf)^2]}} = \sqrt{\mathbf{E}[(Tf)^2]}. \tag{23}$$

The following lemma is useful in bounding $\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P})$ from Definition 1.8 in generic situations. Roughly speaking, it shows that connectivity of the support of $\mathbf{P}$ on correlated spaces $\Omega^{(1)} \times \Omega^{(2)}$ implies that $\rho < 1$.

Lemma 2.9 *Let $(\Omega^{(1)} \times \Omega^{(2)}, \mathbf{P})$ be two correlated spaces such that the probability of the smallest atom in $\Omega^{(1)} \times \Omega^{(2)}$ is at least $\alpha > 0$. Define a bi-partite graph $G = (\Omega^{(1)}, \Omega^{(2)}, E)$ where $(a,b) \in \Omega^{(1)} \times \Omega^{(2)}$ satisfies $(a,b) \in E$ if $\mathbf{P}(a,b) > 0$. Then if G is connected then*

$$\rho(\Omega^{(1)}, \Omega^{(2)}; \mathbf{P}) \leq 1 - \alpha^2/2.$$

One nice property of Markov operators that will be repeatedly used below is that they respect the Efron-Stein decomposition. Given a vector x in an n dimensional product space and $S \subset [n]$ we write x_S for the vector $(x_i : i \in S)$. Given probability spaces $\Omega_1, \ldots, \Omega_n$, we use the convention of writing X_i for a random variable that is distributed according to the measure of Ω_i and x_i for an element of Ω_i. We will also write X_S for $(X_i : i \in S)$.

Definition 2.10 *Let $(\Omega_1, \mu_1), \ldots, (\Omega_n, \mu_n)$ be discrete probability spaces $(\Omega, \mu) = \prod_{i=1}^n (\Omega_i, \mu_i)$. The Efron-Stein decomposition of $f : \Omega \to \mathbb{R}$ is given by*

$$f(x) = \sum_{S \subseteq [n]} f_S(x_S), \tag{24}$$

where the functions f_S satisfy that:

- *f_S depends only on x_S.*
- *For all $S \not\subseteq S'$ and all $x_{S'}$ it holds that:*

$$\mathbf{E}[f_S | X_{S'} = x_{S'}] = 0.$$

It is well known that the Efron-Stein decomposition exists and that it is unique [7]. The function f_S is given by:

$$f_S(x) = \sum_{S' \subseteq S} (-1)^{|S \setminus S'|} \mathbf{E}[f(X) | X_{S'} = x_{S'}].$$

An important property is that the Efron-Stein decomposition "commutes" with Markov operators.

Proposition 2.11 *Let $(\Omega_i^{(1)} \times \Omega_i^{(2)}, \mathbf{P}_i)$ be correlated spaces and let T_i the Markov operator associated with $\Omega_i^{(1)}$ and $\Omega_i^{(2)}$ for $1 \le i \le n$. Let*

$$\Omega^{(1)} = \prod_{i=1}^n \Omega_i^{(1)}, \quad \Omega^{(2)} = \prod_{i=1}^n \Omega_i^{(2)},$$

$$\mathbf{P} = \prod_{i=1}^n \mathbf{P}_i, \quad T = \otimes_{i=1}^n T_i.$$

Suppose $f \in L^2(\Omega^{(2)})$ has Efron-Stein decomposition (24). Then the Efron-Stein decomposition of Tf satisfies:

$$(Tf)_S = T(f_S).$$

Finally we derive a useful bound showing that in the setting above if $\rho(\Omega_i^{(1)} \times \Omega_i^{(2)}; \mathbf{P}) < 1$ for all i then Tf depends on the "low degree expansion" of f.

Proposition 2.12 *Assume the setting of Proposition 2.11 and that further for all i it holds that $\rho(\Omega_i^{(1)}, \Omega_i^{(2)}; \mathbf{P}_i) \le \rho_i$. Then for all f it holds that*

$$\|T(f_S)\|_2 \le \left(\prod_{i \in S} \rho_i \right) \|f_S\|_2.$$

Proposition 2.13 *Assume the setting of Proposition 2.11. Then*

$$\rho(\prod_{i=1}^n \Omega_i^{(1)}, \prod_{i=1}^n \Omega_i^{(2)}; \prod_{i=1}^n \mathbf{P}_i) = \max \rho(\Omega_i^{(1)}, \Omega_i^{(2)}).$$

3 Majority is Most Predictable

We conclude the abstract with a proof of the easiest application of the main results established here by proving Theorem 1.1.

3.1 ρ for samples of votes

In the first social choice example we consider example 2.6. The correlated probability spaces are the ones given by

$$\Omega_V = \{\{x = 1\}, \{x = -1\}\},$$

representing the intended vote and

$$\Omega_S = \{\{(x = 1, y = 1)\}, \{(x = -1, y = 1)\}, \{y = 0\}\}$$

representing the sampled status.

In order to calculate $\rho(\Omega_1, \Omega_2)$ it suffices by lemma 2.8 to calculate $\sqrt{E[(Tf)^2]}$ where $f(x, y) = x$ is the (only) Ω_V measurable with $\mathbf{E}[f] = 0$ and $\mathbf{E}[f^2] = 1$. We see that $Tf(x, y) = 0$ if $y = 0$ and $Tf(x, y) = x$ when $y \neq 0$. Therefore

$$\sqrt{\mathbf{E}[(Tf)^2]} = \rho^{1/2}.$$

Lemma 3.1

$$\rho(\Omega_V, \Omega_S) = \rho^{1/2}.$$

3.2 Predictability of Binary Vote

Here we prove Theorem 1.1.

Proof: The proof follows directly from Proposition 1.14 and Lemma 3.1 as

$$\Gamma_{\sqrt{\rho}}(1/2, 1/2) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \sqrt{\rho},$$

by Sheppard formula [19]. □

References

[1] K. Arrow. A difficulty in the theory of social welfare. *J. of Political Economy*, 58:328–346, 1950.

[2] P. Austrin. Towards Sharp Inapproximability For Any 2-CSP. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 307–317, 2007.

[3] P. Austrin and E. Mossel. Approximation resistant predicates from pairwise independence. In *23rd Annual IEEE Conference on Computational Complexity*, pages 249–258, Los Alamitos, CA, USA, 2008. IEEE Computer Society.

[4] C. E. Bell. A random voting graph almost surely has a hamiltonian cycle when the number of alternatives is large. *Econometrica*, 49(6):1597–1603, 1981.

[5] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985.

[6] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing (STOC 2006)*, pages 344–353, 2006.

[7] B. Efron and C. Stein. The jacknife estimate of variance. *Annals of Statistics*, 9(3), 1996.

[8] B. Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.

[9] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[10] G. Kalai. A Fourier-theoretic perspective on the Concordet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29(3):412–426, 2002.

[11] G. Kalai. Social Indeterminacy. *Econometrica*, 72:1565–1581, 2004.

[12] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th Ann. STOC*, pages 767–775, 2002.

[13] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37:319–357, 2007.

[14] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality (extended abstract). In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 21–30. IEEE Computer Society, 2005.

[15] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. To appear in Ann. Math., 2008.

[16] R. O'Donnell. Hardness amplification within np. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, 2002.

[17] P. Raghavendra. Optimal Algorithms and Inapproximability Results For Every CSP? In *Proceedings of the 40th Annual Symposium on Theory of Computing (STOC)*, pages 245–254, 2008.

[18] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 11–20, New York, NY, USA, 2006. ACM Press.

[19] W. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London*, 192:101–168, 1899.