Noise Sensitivity of Boolean Functions And Applications to Percolation

Itai Benjamini Gil Kalai Oded Schramm

December 14, 1998

Abstract

It is shown that a large class of events in a product probability space are highly sensitive to noise, in the sense that with high probability, the configuration with an arbitrary small percent of random errors gives almost no prediction whether the event occurs. On the other hand, weighted majority functions are shown to be noise-stable. Several necessary and sufficient conditions for noise sensitivity and stability are given.

Consider, for example, bond percolation on an $n+1$ by n grid. A configuration is a function that assigns to every edge the value 0 or 1. Let ω be a random configuration, selected according to the uniform measure. A crossing is a path that joins the left and right sides of the rectangle, and consists entirely of edges e with $\omega(e) = 1$. By duality, the probability for having a crossing is $1/2$. Fix an $\epsilon \in (0,1)$. For each edge e, let $\omega'(e) = \omega(e)$ with probability $1-\epsilon$, and $\omega'(e) = 1 - \omega(e)$ with probability ϵ, independently of the other edges. Let $p(\tau)$ be the probability for having a crossing in ω, conditioned on $\omega' = \tau$. Then for all n sufficiently large, $\mathbf{P}\big\{\tau : |p(\tau) - 1/2| > \epsilon\big\} < \epsilon$.

Contents

1 Introduction

1.1 Noise sensitivity — three examples

Consider the Hamming cube $\Omega_n = \{0,1\}^n$ endowed with the uniform probability measure $\mathbf{P}$. Let $\mathcal{A} \subset \Omega_n$ be some event. Given a random $x = (x_1, \dots, x_n) \in \Omega_n$, suppose that $y = (y_1, \dots, y_n)$ is a random perturbation of x; that is, for every $j \in \{1, \dots, n\}$, $y_j = x_j$ with probability $1 - \epsilon$, independently for distinct j's. Here $\epsilon \in (0,1)$ is some small fixed constant. This random perturbation of x will be denoted $N_\epsilon(x)$. We may think of $N_\epsilon(x)$ as x with some noise.

Based on the knowledge of $N_\epsilon(x)$, we would like to predict the event $x \in \mathcal{A}$. Since the joint distribution $(x, N_\epsilon(x))$ is the same as that of $(N_\epsilon(x), x)$, an equivalent problem is to predict $N_\epsilon(x) \in \mathcal{A}$ knowing x. The event $\mathcal{A}$ is **noise sensitive** if for all but a small set of x, knowing x does not significantly help in predicting the event $N_\epsilon(x) \in \mathcal{A}$. More formally, $\mathcal{A}$ is noise sensitive, if for some small $\delta > 0$,

$$\gamma(\mathcal{A}, \epsilon, \delta) := \mathbf{P}\Big\{x : \Big|\mathbf{P}\big(N_\epsilon(x) \in \mathcal{A} \mid x\big) - \mathbf{P}(\mathcal{A})\Big| > \delta\Big\} < \delta \tag{1.1}$$

Set

$$\phi(\mathcal{A}, \epsilon) = \inf\Big\{\delta > 0 : \gamma(\mathcal{A}, \epsilon, \delta) < \delta\Big\},$$

which is the infimum of all $\delta > 0$ such that (1.1) holds. This will be called the **sensitivity gauge** of $\mathcal{A}$. A sequence of events $\mathcal{A}_m \subset \Omega_{n_m}$ will be called **asymptotically noise sensitive** if

$$\lim_{m\to\infty} \phi(\mathcal{A}_m, \epsilon) = 0, \qquad \forall \epsilon \in (0, 1/2).$$

A simple example of a sequence of events which are not noise sensitive is dictatorship. The first bit dictator is the event $\mathcal{D}_n = \Big\{(x_1, \dots, x_n) \in \Omega_n : x_1 = 1\Big\}$. To verify that $\{\mathcal{D}_n\}$ is not asymptotically noise sensitive, consider some event $\mathcal{A} \subset \Omega_n$. Then for $k > n$ we may obviously consider $\mathcal{A}$ as a subset of Ω_k, by ignoring the extra variables. Note that this does not change the value of $\phi(\mathcal{A}, \epsilon)$. Consequently, $\phi(\mathcal{D}_n, \epsilon) = \phi(\mathcal{D}_1, \epsilon) \neq 0$ for all $n > 1$.

Let us examine now the example of majority. Pick some $\epsilon \in (0, 1/2)$. Let $\mathcal{M}_n \subset \Omega_n$ denote the majority event, that is,

$$\mathcal{M}_n = \left\{(x_1, \dots, x_n) \in \Omega_n : \sum_j x_j \geq n/2\right\}.$$

The probability that $\sum_j x_j - n/2 > \sqrt{n}$ is bounded from below as $n \to \infty$. Given such an x, the probability that $N_\epsilon(x) \in \mathcal{M}_n$ is greater than $\mathbf{P}[\mathcal{M}_n]+\delta_1$ for some constant $\delta_1 > 0$, depending on ϵ. We conclude that majority is not asymptotically noise sensitive as $n \to \infty$.

Majority and dictatorship are not only noise insensitive, they are actually "noise stable", in a sense defined in Subsection 1.4 below.

It turns out that the noise insensitivity of majority and dictatorship is atypical, and many natural and interesting events are asymptotically noise sensitive.

Our third example is bond percolation on an $m+1$ by m rectangle in the ordinary square grid $\mathbb{Z}^2$. A configuration is an element in $\Omega = \{0,1\}^E$, where E is the set of edges in this rectangle. Let $\omega \in \Omega$ be a random configuration, selected according to the uniform measure. A **crossing** is a path that joins the left and right sides of the rectangle, and consists entirely of edges e with $\omega(e) = 1$. Let $\mathcal{C}_m$ be the event that there is some crossing of this rectangle. By duality, it is not hard to see that $\mathbf{P}[\mathcal{C}_m] = 1/2$.

Theorem 1.1. *The crossing events $\mathcal{C}_m$ are asymptotically noise sensitive; that is, $\phi(\mathcal{C}_m, \epsilon) \to 0$ as $m \to \infty$.*

This theorem will appear as a corollary of a general result. To introduce the more general statement, we need the notion of influence.

1.2 Influences of variables

Set $[n] = \{1, \ldots, n\}$. Given $x \in \Omega$ and $j \in [n]$, let $\sigma_j x = (x'_1, \ldots, x'_n)$, where $x'_k = x_k$ when $k \neq j$ and $x'_j = 1 - x_j$. The **influence** of the k-th variable on a function $f : \Omega \to \mathbb{R}$ is defined by

$$I_k(f) = \|f(\sigma_k x) - f(x)\|_1 . \tag{1.2}$$

In other words, $I_k(f)$ is the expected absolute value of the change in f when the k'th bit x_k is flipped. We shall often not distinguish between an event $\mathcal{A}$ and its indicator function $\chi_{\mathcal{A}}$. In particular, for events $\mathcal{A}$, $I_k(\mathcal{A}) = I_k(\chi_{\mathcal{A}})$. Note that $I_k(\mathcal{A})$ is the probability that precisely one of the two elements $x, \sigma_k x$ is in $\mathcal{A}$.

This notion of influence was introduced by Ben-Or and Linial [3]. Kahn, Kalai and Linial [18] (see also, [8, 25]) showed that for every $\mathcal{A} \subset \Omega_n$ with $\mathbf{P}[\mathcal{A}] = 1/2$ there is a $j \in [n]$ with $I_j(\mathcal{A}) \geq c \log n/n$, for some constant

$c > 0$, and that there always exists a set $S \subset [n]$ with $|S| \le c(\epsilon)n/\log n$ whose cumulative influence is $> 1 - \epsilon$; that is, the measure of the set of inputs for variables in $[n] - S$ which determine the value of f is less than ϵ.

Put

$$\begin{aligned} I(f) &= \sum_k I_k(f), \\ II(f) &= \sum_k I_k(f)^2. \end{aligned}$$

Theorem 1.2. *Let $\mathcal{A}_m \subset \Omega_{n_m}$ be a sequence of events and suppose that $II(\mathcal{A}_m) \to 0$ as $m \to \infty$. Then $\{\mathcal{A}_m\}$ is asymptotically noise sensitive.*

Equivalently, there is some continuous function Φ satisfying $\Phi(0, \epsilon) = 0$ such that $\phi(\mathcal{A}, \epsilon) \le \Phi\big(II(\mathcal{A}), \epsilon\big)$ for every event $\mathcal{A}$ in some Ω_n.

On Ω_n, we use the usual lattice order: $(x_1, \dots, x_n) \le (y_1, \dots, y_n)$ iff $x_j \le y_j$ for all $j \in [n]$. A function $f : \Omega_n \to \mathbb{R}$ is **monotone** if $f(x) \le f(y)$ whenever $x \le y$. An event $\mathcal{A} \subset \Omega_n$ is **monotone** if its indicator function $\chi_{\mathcal{A}}$ is monotone.

For monotone events, Theorem 1.2 has a converse:

Proposition 1.3. *Let $\mathcal{A}_m \subset \Omega_{n_m}$ be a sequence of monotone events with*

$$\inf_m II(\mathcal{A}_m) > 0.$$

Then $\{\mathcal{A}_m\}$ is not asymptotically noise sensitive.

The assumption that the events $\mathcal{A}_m$ are monotone is necessary here. (For example, take $\mathcal{A}_m$ to be a uniform random subset of Ω_m, or parity: $\mathcal{A}_m := \{x \in \Omega_m : \|x\|_1 \text{ is odd}\}$.)

Suppose that $\mathcal{A}$ is a monotone event where the influences of all the variables are the same. The influence $I_1(\mathcal{A})$ then measures the sensitivity of $\mathcal{A}$ to flips of a single variable. Note that, quite paradoxically, $\mathcal{A}$ is *least* sensitive to noise when $I_1(\mathcal{A})$ is largest.

We now give a quantitative version of Theorem 1.2 under the assumption that $II(\mathcal{A}_m)$ goes to zero fast enough.

Theorem 1.4. *Let $\mathcal{A} \subset \Omega_n$, and suppose that $II(\mathcal{A}) \le n^{-a}$, where $a \in (0, 1/2]$. Then there exist $c_1, c_2 > 0$, depending only on a so that*

$$\phi(\mathcal{A}, \epsilon) \le c_1 n^{-c_2 \epsilon}, \qquad \forall \epsilon \in (0, 1/4).$$

Consequently, if $\mathcal{A}_m \subset \Omega_{n_m}$ *is a sequence of events satisfying* $II(\mathcal{A}_m) \le (n_m)^{-a}$ *and* ϵ_m *is a sequence in* $(0, 1/4)$ *such that* $\epsilon_m \log n_m \to \infty$*, then* $\phi(\mathcal{A}_m, \epsilon_m) \to 0$.

1.3 Weighted majority

It turns out that for monotone events noise insensitivity is also closely related to correlation with majority functions.

Let $K \subset [n]$ and define the **majority function** on K by $M_K(x) = \operatorname{sign} \sum_{j \in K}(2x_j - 1)$; that is,

$$M_K(x) = \begin{cases} -1 & \text{if } \sum_{j \in K} x_j < |K|/2\,; \\ 0 & \text{if } \sum_{j \in K} x_j = |K|/2\,; \\ 1 & \text{if } \sum_{j \in K} x_j > |K|/2\,. \end{cases} \tag{1.3}$$

For $f : \Omega_n \to \mathbb{R}$ set

$$\Lambda(f) = \max\Big\{|\mathbf{E}(f M_K)| : K \subset [n]\Big\}.$$

Theorem 1.5. *Let* $f : \Omega_n \to [0, 1]$ *be monotone. Then*

$$II(f) \le C\Lambda(f)^2 \Big(1 - \log \Lambda(f)\Big) \log n,$$

where C *is some universal constant.*

Consequently, if $\mathcal{A}_m \subset \Omega_{n_m}$ *is a sequence of monotone events with*

$$\lim_{m \to \infty} \Lambda(\mathcal{A}_m)^2 \Big(1 - \log \Lambda(\mathcal{A}_m)\Big) \log n_m = 0. \tag{1.4}$$

Then $\{\mathcal{A}_m\}$ *is asymptotically noise sensitive.*

One cannot get rid of the $\log n_m$ factor (see Remark 3.10), except by using weighted majority functions. For positive weights $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ consider a weighted majority function, which is defined by

$$M_{\mathbf{w}}(x_1, x_2, \ldots, x_n) = \operatorname{sign}\left(\sum (2x_j - 1) w_j\right).$$

Finally write

$$\widetilde{\Lambda}(A) = \max\Big\{|\mathbf{E}(\chi_A M_{\mathbf{w}})| : \mathbf{w} \in [0, 1]^n\Big\}.$$

Theorem 1.6. *Let $\mathcal{A}_m \subset \Omega_{n_m}$ be a sequence of monotone events. Then $\{\mathcal{A}_m\}$ is asymptotically noise sensitive if and only if $\lim_{m\to\infty} \widetilde{\Lambda}(A_m) = 0$.*

For a monotone event $\mathcal{A} \subset \Omega_n$, which is symmetric in the n variables, its correlation with unweighted majority is enough to determine if it is noise sensitive.

1.4 Stability

We now define the notion of **stability**, which is the opposite of noise sensitivity. Suppose $\mathcal{A} \subset \Omega_n$, and let $x \in \Omega_n$ be random-uniform. For $\epsilon > 0$, let $N_\epsilon \mathcal{A}$ denote the event $N_\epsilon(x) \in \mathcal{A}$. It is then clear that $\mathbf{P}[\mathcal{A} \triangle N_\epsilon \mathcal{A}] \to 0$ as $\epsilon \to 0$. ($\mathcal{B} \triangle \mathcal{A}$ denotes the symmetric difference, $(\mathcal{B} - \mathcal{A}) \cup (\mathcal{A} - \mathcal{B})$.) The faster $\mathbf{P}[\mathcal{A} \triangle N_\epsilon \mathcal{A}]$ tends to zero, the more noise-stable $\mathcal{A}$ is. More precisely, let $\{\mathcal{A}_i\}$ be a collection of events, where $\mathcal{A}_i \subset \Omega_{n_i}$. We say that $\{\mathcal{A}_i\}$ are **uniformly stable** if the limit $\lim_{\epsilon\to 0} \mathbf{P}[\mathcal{A}_i \triangle N_\epsilon \mathcal{A}_i] = 0$ is uniform in i.

For $\mathbf{w} \in \mathbb{R}^{\mathbf{n}}$ and $s \in \mathbb{R}$, let $\mathcal{M}_{\mathbf{w},s}$ be the (generalized) weighted majority event

$$\mathcal{M}_{\mathbf{w},s} := \left\{ x \in \mathbb{R}^n : \sum_{j=1}^{n} (2x_j - 1) w_j > s \right\} \subset \Omega_n.$$

Let $\mathfrak{M}$ denote the collection of such events:

$$\mathfrak{M} := \left\{ \mathcal{M}_{\mathbf{w},s} : n = 1, 2, \ldots,\ \mathbf{w} \in \mathbb{R}^{\mathbf{n}},\ \mathbf{s} \in \mathbb{R} \right\}.$$

In Section 3 we show that

Theorem 1.7. *$\mathfrak{M}$ is uniformly stable. Moreover, for every $\mathcal{M} \in \mathfrak{M}$*

$$\mathbf{P}[\mathcal{M} - N_\epsilon \mathcal{M}] \leq C \epsilon^{1/4},$$

where C is a universal constant independent of $\mathcal{M}$.

Note that an infinite sequence $\{\mathcal{A}_i\}$ with $\mathbf{P}[\mathcal{A}_i]$ bounded away from 0 and 1 cannot be asymptotically noise sensitive and uniformly stable. We also observe (Lemma 3.7) that when $\{\mathcal{A}_i\}$, $(\mathcal{A}_i \subset \Omega_{n_i})$, is asymptotically noise sensitive and $\{\mathcal{B}_i\}$, $(\mathcal{B}_i \subset \Omega_{n_i})$, is uniformly stable, then $\mathcal{A}_i$ and $\mathcal{B}_i$ are asymptotically uncorrelated. One can say, somewhat imprecisely, that the noise sensitive events are asymptotically in the orthocomplement of the uniformly stable events.

Stability and sensitivity are two extremes. However, there are events that are neither sensitive nor stable. For example, if $\mathcal{C}$ is the event of a percolation crossing, as described above, and $\mathcal{M}$ is the majority event, then $\mathcal{C} \cap \mathcal{M}$ is neither asymptotically noise sensitive, nor uniformly stable.

1.5 Fourier-Walsh expansion

For a boolean function f on $\{0,1\}^n$, consider the Fourier-Walsh expansion $f = \sum_{S \subset [n]} \widehat{f}(S) u_S$, where, $u_S(T) = (-1)^{|S \cap T|}$. Here and in the following, we identify any vector $x \in \Omega^n$ with the subset $\{j \in [n] : x_j = 1\}$, of $[n] = \{1, 2, \ldots, n\}$. Consequently, $|x|$ denotes the cardinality of that set; that is, $|x| = \|x\|_1$ for $x \in \Omega_n$.

Theorem 1.8. *Let $\mathcal{A}_m \subset \Omega_{n_m}$ be a sequence of events, and set $g_m = \chi_{\mathcal{A}_m}$. Then $\{\mathcal{A}_m\}$ is asymptotically noise sensitive iff for every finite k*

$$\lim_m \sum \Big\{ \widehat{g}_m(S)^2 : S \subset [n], 1 \leq |S| \leq k \Big\} = 0. \tag{1.5}$$

$\{\mathcal{A}_m\}$ is uniformly stable iff

$$\lim_{k \to \infty} \sup_m \sum \Big\{ \widehat{g}_m(S)^2 : S \subset [n], |S| \geq k \Big\} = 0. \tag{1.6}$$

It can be easily shown that for $f = \chi_{\mathcal{A}}$

$$I(f) = 4 \sum_{S \subset [n]} \widehat{f}(S)^2 |S|.$$

(This follows from (2.5) below with $p = 2$.) We will introduce another quantity

$$J(f) = \sum_{\emptyset \neq S \subset [n]} \widehat{f}(S)^2 / |S|.$$

Also set for $\mathcal{A} \subset \Omega_n$, $n > 1$,

$$\alpha(\mathcal{A}) = \log I(\mathcal{A}) / \log n,$$
$$\beta(\mathcal{A}) = -\log J(\mathcal{A}) / \log n.$$

For events $\mathcal{A}$ we clearly have $0 \leq \beta(\mathcal{A})$, and $\beta(\mathcal{A}) \leq \alpha(\mathcal{A})$, provided that $\mathbf{P}[\mathcal{A}] = 1/2$. When $\mathcal{A}$ is monotone $\alpha(\mathcal{A}) \leq 1/2$.

Perhaps some words of explanation are needed. $I(\mathcal{A})$ measures the sum of the influences of the variables. For monotone events it is maximal for majority, where $I(\mathcal{A}) \simeq \sqrt{n}$ and thus $\alpha(\mathcal{A}) \to 1/2$. In the terminology used in percolation theory, $I(\mathcal{A})$ is the expected number of pivotal edges. For the crossing events $\mathcal{C}$ of percolation (in arbitrary dimensions) it is conjectured that $I(\mathcal{C})$ behaves like a certain fractional power (a *critical exponent*) of n. It is conjectured that in dimension 2, as n tends to infinity, $\alpha(\mathcal{C})$ tends to $3/8$. Thus, this critical exponent generalizes and has a Fourier-analysis interpretation for arbitrary Boolean functions.

$\alpha(\mathcal{A})$ is large if there are substantial Fourier coefficients $\widehat{f}(S)$ for large $|S|$. In contrast, $\beta(\mathcal{A})$ is large if there are *no* substantial Fourier coefficients $\widehat{f}(S)$ for S of small positive size. We conjecture that for the crossing events for percolation, as n tends to infinity $\beta(\mathcal{C})$ tends to a positive limit which is strictly smaller than the limit for $\alpha(\mathcal{C})$.

1.6 Some related and future work

There are interesting connections between noise sensitivity and isoperimetric inequalities of the form described by Talagrand in [26]. These connections and applications for first passage percolation problems will be discussed in a subsequent paper.

Our notion of noise sensitivity is related to the study of noises by Tsilerson [28, 29]. "Noise", in Tsilerson's sense, is a type of σ-field filtration. Uniform stability seems to correspond, in the limit, to the noise being white, while asymptotic sensitivity seems to correspond to the noise being black.

1.7 The structure of this paper

Theorems 1.2, and 1.3 are proved in the next section. Our proofs combines combinatorial reasonings with applying certain inequalities for the Fourier coefficients of Bonami and Beckner which were used already in [18]. However, to get the results in the sharpest forms we have to rely on a sophisticated "bootstrap" method of [27] and on the main results of that paper which rely on this method. Talagrand's remarkable paper [27] has thus much influence on the present work.

Weighted and unweighted majority functions are considered in Section 3. An applications to percolation is described in Section 4 followed by some related open problems in Section 5. In Section 6, we will work out two examples (due to Ben-Or and Linial). In one of these $\alpha(\mathcal{A}) \to 1 - \log_2 3$ and $\beta(\mathcal{A}) \to 1 - \log_2 3$. In Section 7 we consider relations with complexity theory. A simple description of noise-sensitivity in terms of random walks is given in Section 8. In Section 9 we consider perturbations with a different sort of noise, where the number of bits that are changed is of a fixed size. The conclusions are similar to those above, but there is an amusing and slightly unexpected twist.

For simplicity we consider here the uniform measure on Ω^n. More generally, one may consider the product measure $\mathbf{P}_p$, where $\mathbf{P}_p\{x : x_j = 1\} = p$. Our results and proof apply in this setting. (All that is needed is to replace the Fourier-Walsh transform by its analog as given in Talagrand's paper [25] and the proofs go through without change.) However, the case when p itself depends on n is interesting, but will not be considered here.

Acknowledgments.

It is a pleasure to thank Noga Alon, Ehud Friedgut, Ravi Kannan, Harry Kesten, Yuval Peres, Michel Talagrand and Avi Wigderson for helpful discussions.

2 Sensitivity to noise

We now put the noise operator N_ϵ defined in the introduction into a somewhat more general framework. That will allow us to deal, for example, with the situation where the 1 bits are immune to noise but the 0 bits are noise prone.

Consider the following method for selecting a random point $x \in \Omega_n$. Let $q_1, \dots, q_n$ be independent random variables in $[0,1]$, with $\mathbf{E}q_j = 1/2$, for $j = 1, \dots, n$, and let $\omega \in [0,1]^n$ be random uniform. Set

$$x_j = \begin{cases} 1, & \text{if } 1 - \omega_j < q_j, \\ 0, & \text{otherwise.} \end{cases}$$

Then x is distributed according to the uniform measure of Ω_n; it will be denoted by $N(\omega, q)$.

Let ν be the measure on $[0,1]^n$ such that $\nu(X) = \mathbf{P}\big[(q_1, \dots, q_n) \in X\big]$. We think of x as being chosen in two stages. In the first stage, $q = (q_1, \dots, q_n)$ is selected according to ν. This q gives a product measure $\mathbf{P}_q$ on $\{0,1\}^n$ that satisfies $\mathbf{P}_q\{\tau \in \Omega_n : \tau(j) = 1\} = q_j$. Then x is chosen according to the measure $\mathbf{P}_q$.

For example, suppose $z \in \Omega_n$. Define $q = q(z) \in [0,1]^n$ by $q_j = 1 - \epsilon$ if $z_j = 1$ and $q_j = \epsilon$ if $z_j = 0$. Then for every $z \in \Omega_n$, the perturbation $N_\epsilon(z)$ has the same distribution as $N(\omega, q(z))$. The ν giving this distribution of q will be denoted ν_ϵ.

However, the construction $N(\omega, q)$ is more general than that given by the noise operator N_ϵ. As hinted above, one can create a situation where 1 bits are robust, but 0 bits are prone to noise. More precisely, take $q_j = 1$, with probability $1/2 - \epsilon$ and $q_j = \epsilon/(1/2 + \epsilon)$ with probability $1/2 + \epsilon$.

Another interesting example is obtained by taking each q_j to be 1, with probability $(1-\epsilon)/2$, 0, with probability $(1-\epsilon)/2$, and $1/2$ with probability ϵ.

Let $f : \{0,1\}^n \to \mathbb{R}$ be some function. In the following, f will be taken to be the characteristic function $\chi_{\mathcal{A}}$ of some event $\mathcal{A} \subset \{0,1\}^n$, or $f = \chi_{\mathcal{A}} - \mathbf{P}(\mathcal{A})$. What information does the first stage in the selection of $x = N(\omega, q)$, namely the selection of q, give about the value of $f(x)$? If we know that $q = z$, then our prediction for $f(x)$ would be

$$G(f, z) = \mathbf{E}\big(f(x) \mid q = z\big).$$

The expected value of $G(f, q)$ is obviously $\mathbf{E}(f)$. Let

$$Z(f, \nu) = \mathbf{E}_q G(f, q)^2 = \int G(f, z)^2 \, d\nu(z).$$

This is just the second moment of $G(f, q)$. If $Z(f, \nu) - (\mathbf{E}f)^2$ is small, then for "almost all" values of q there is no prediction for $f(x)$ that is significantly better than the a priori knowledge of $\mathbf{E}f$. We often write $G(\mathcal{A}, \cdot)$ and $Z(\mathcal{A}, \cdot)$ in place of $G(\chi_{\mathcal{A}}, \cdot)$ and $Z(\chi_{\mathcal{A}}, \cdot)$.

Lemma 2.1. *The number $Z(f, \nu)$ depends only on f and the variances ζ_j of the variables q_j. Its expression in terms of the Fourier coefficients is,*

$$Z(f, \nu) = \sum_{S \in \Omega_n} \widehat{f}(S)^2 \prod_{j \in S} 4\zeta_j.$$

PROOF:

$$\begin{aligned}
G(f,z) &= \mathbf{E}\big(f(x) \mid q = z\big) \\
&= \sum_{T\subset[n]} f(T) \prod_{j\in T} z_j \prod_{j\notin T}(1-z_j) \\
&= \sum_T \sum_S \widehat{f}(S)(-1)^{|T\cap S|} \prod_{j\in T} z_j \prod_{j\notin T}(1-z_j) \\
&= \sum_S \widehat{f}(S) \left(\sum_{T'\subset S} (-1)^{|T'|} \prod_{j\in T'} z_j \prod_{j\notin T'} (1-z_j) \right) \cdot \\
&\qquad\qquad \left(\sum_{T''\subset[n]-S} \prod_{j\in T''} z_j \prod_{j\notin T''} (1-z_j) \right) \\
&= \sum_S \widehat{f}(S) \left(\prod_{j\in S} \big((1-z_j)-z_j\big) \right) \left(\prod_{j\notin S} \big((1-z_j)+z_j\big) \right) \\
&= \sum_S \widehat{f}(S) \prod_{j\in S}(1-2z_j)
\end{aligned}$$

Therefore,

$$\begin{aligned}
Z(f,\nu) &= \mathbf{E}G(f,q)^2 \\
&= \sum_S \sum_{S'} \widehat{f}(S)\widehat{f}(S')\mathbf{E}\left(\prod_{j\in S}(1-2q_j) \prod_{j\in S'}(1-2q_j) \right) \\
&= \sum_S \sum_{S'} \widehat{f}(S)\widehat{f}(S') \prod_{j\in S\cap S'} \mathbf{E}(1-2q_j)^2 \prod_{j\in S\triangle S'} \mathbf{E}(1-2q_j).
\end{aligned}$$

Since $\mathbf{E}q_j = 1/2$, summands with $S \neq S'$ vanish. The lemma follows. □

For every $\epsilon \in [0,1]$, $x \in \Omega_m$ and $f : \Omega_n \to \mathbb{R}$ set

$$Q_\epsilon f(x) = \mathbf{E} f(N_\epsilon(x))$$

(here the expectation is only with respect to the noise). Also let

$$\mathrm{VAR}(f,\epsilon) = \mathrm{var}(Q_\epsilon f) = Z\big(f,\nu_\epsilon\big) - (\mathbf{E}f)^2\,.$$

Note that for singletons $S = \{i\} \subset [n]$, we have $Q_\epsilon u_S = (1-2\epsilon)u_S$. If $S_1, S_2 \subset [n]$ are disjoint and $x \in \Omega_n$ is fixed, then $N_\epsilon(x)\cap S_1$ and $N_\epsilon(x)\cap S_2$ are independent. Consequently, $Q_\epsilon(u_{S_1}u_{S_2}) = (Q_\epsilon u_{S_1})(Q_\epsilon u_{S_2})$. We may conclude that $Q_\epsilon u_S = (1-2\epsilon)^{|S|}u_S$ for every $S \subset [n]$, and linearity gives

$$Q_\epsilon f = \sum_{S\subset[n]} \widehat{f}(S)(1-2\epsilon)^{|S|}u_S\,. \tag{2.1}$$

One consequence of this, which can also be obtained from Lemma 2.1, is

$$\mathrm{VAR}(f,\epsilon) = \sum_{\emptyset\neq S\subset[n]} \widehat{f}(S)^2(1-2\epsilon)^{2|S|}. \tag{2.2}$$

Now we relate $\mathrm{VAR}(\mathcal{A},\epsilon)$ with the sensitivity gauge $\phi(\mathcal{A},\epsilon)$:

Proposition 2.2. *For every $\mathcal{A} \subset \Omega_n$*

$$\frac{1}{2}\mathrm{VAR}(\mathcal{A},\epsilon) \le \phi(\mathcal{A},\epsilon) \le \mathrm{VAR}(\mathcal{A},\epsilon)^{1/3}.$$

Proof: Let $\delta = \phi(\mathcal{A},\epsilon)$, and set

$$Y = \left\{ y\in\Omega_n : \Big|Q_\epsilon\chi_{\mathcal{A}}(y) - \mathbf{P}[\mathcal{A}]\Big| \ge \delta \right\}.$$

Then, by the definition of ϕ, $\mathbf{P}[Y] \ge \delta$. Consequently,

$$\mathrm{VAR}(\mathcal{A},\epsilon) \ge \delta^2\mathbf{P}[Y] \ge \delta^3 = \phi(\mathcal{A},\epsilon)^3.$$

For the other direction set

$$Y' = \left\{ y\in\Omega_n : \Big|Q_\epsilon\chi_{\mathcal{A}}(y) - \mathbf{P}[\mathcal{A}]\Big| > \delta \right\}.$$

Then $\mathbf{P}[Y'] \le \delta$. For $y \in Y'$, the trivial estimate $\Big|Q_\epsilon\chi_{\mathcal{A}} - \mathbf{P}[\mathcal{A}]\Big| \le 1$ holds. Therefore,

$$\mathrm{VAR}(\mathcal{A},\epsilon) \le \mathbf{P}[Y'] + \delta^2 \le 2\phi(\mathcal{A},\epsilon).$$

$\square$

Proof of 1.8: The first part is immediate from Prop. 2.2 and (2.2). For the proof of the second part, observe that (2.1) implies that (1.6) is equivalent to $\|g_m - Q_\epsilon g_m\|_2 \to 0$ uniformly as $\epsilon \to 0$. Since $|g_m|$ and $|Q_\epsilon g_m|$ are bounded, this is equivalent to $\|g_m - Q_\epsilon g_m\|_1 \to 0$ uniformly, which is the same as uniform stability for $\{\mathcal{A}_m\}$. $\square$

Remark 2.3. Another consequence of 2.2 and (2.2) is that for constant $\epsilon, \epsilon' \in (0, 1/2)$, we have $\phi(\mathcal{A}_m, \epsilon) \to 0$ iff $\phi(\mathcal{A}_m, \epsilon') \to 0$. Consequently, to verify that $\mathcal{A}_m$ is asymptotically noise sensitive, it is enough to prove $\mathrm{VAR}(\mathcal{A}_m, \epsilon) \to 0$ with any fixed $\epsilon \in (0, 1/2)$.

By Theorem 1.8, to establish Theorem 1.2 we need to show that the L^2 weight of the Fourier coefficients with $|S|$ small is negligible. For a function $g = \sum \widehat{g}(S) u_S$ let

$$T_\eta g = Q_{\frac{1-\eta}{2}} g = \sum \widehat{g}(S) \eta^{|S|} u_S.$$

Observe that $T_0(g) = \mathbf{E} g$ and $T_1 g = g$. Also note that

$$\|T_{1-2\epsilon} g\|_2^2 = \mathrm{VAR}(g, \epsilon) + \widehat{g}(\emptyset)^2, \tag{2.3}$$

by (2.2).

The following hyper-contractive inequality of Bonami and Beckner [5, 2], which was crucial in [18], will be useful.

Lemma 2.4 (Bonami, Beckner). $\|T_\eta f\|_2 \le \|f\|_{1+\eta^2}$.

The following is a slightly weaker version of Theorem 1.2, which is sufficient for the applications to percolation. It is presented here, since we can give an almost self-contained proof of it.

Theorem 2.5. *Suppose that $\mathcal{A}_m \subset \Omega_{n_m}$ is a sequence of events and*

$$\lim_{m \to \infty} \frac{\log II(\mathcal{A}_m)}{\log \log n_m} = -\infty\,. \tag{2.4}$$

Then $\{\mathcal{A}_m\}$ is asymptotically noise sensitive.

Proof: Abbreviate $\mathcal{A}$ for $\mathcal{A}_m$ and n for n_m, and set $f := \chi_{\mathcal{A}}$. Let $\widetilde{f} := \chi_{\mathcal{A}} - \mathbf{P}[\mathcal{A}]$. Thus, $\widehat{\widetilde{f}}(\emptyset) = 0$ and $\widehat{\widetilde{f}}(S) = \widehat{f}(S)$, when $S \ne \emptyset$.

Recall that $\sigma_j x = (x_1', \dots, x_n')$, where $x_i' = x_i$ if $i \ne j$ and $x_j' = 1 - x_j$. Let

$$f_j(x) = f(x) - f(\sigma_j x), \qquad j = 1, 2, \dots, n,$$

and note that

$$\widehat{f_j}(S) = \begin{cases} 0, & \text{if } j \notin S, \\ 2\widehat{f}(S), & \text{if } j \in S. \end{cases}$$

Since f_j takes only the values $-1, 0, 1$, equation (1.2) gives for every $p \geq 1$,

$$\|f_j\|_p = I_j(f)^{1/p} \,. \tag{2.5}$$

We set $\eta := 1 - 2\epsilon$, where $\epsilon \in (0, 1/2)$ and

$$F_{\mathcal{A}}(\eta) := \mathrm{VAR}(\mathcal{A}, \epsilon) = \|T_\eta \widetilde{f}\|_2^2 = \sum_{S \neq \emptyset} \widehat{f}(S)^2 \eta^{2|S|}.$$

By Remark 2.3 and Proposition 2.2, it is enough to prove that $F_{\mathcal{A}}(1/2) \to 0$ as $m \to \infty$. We have

$$\begin{aligned} F_{\mathcal{A}}(\eta) &\leq \sum_S \widehat{f}(S)^2 |S| \eta^{2|S|} = \frac{1}{4} \sum_{j=1}^n \|T_\eta f_j\|_2^2 \\ &\leq \frac{1}{4} \sum_{j=1}^n \|f_j\|_{1+\eta^2}^2 \qquad \text{(by Lemma 2.4)} \\ &\leq \sum_{j=1}^n I_j(\mathcal{A})^{2/(1+\eta^2)} \qquad \text{(by (2.5))} \\ &\leq n^{\eta^2/(1+\eta^2)} I\!I(\mathcal{A})^{1/(1+\eta^2)} \qquad \text{(by the means inequality).} \end{aligned} \tag{2.6}$$

Take some $\eta_1 \in (0, 1/2)$, to be later specified, and set $\lambda := \log F_{\mathcal{A}}(\eta_1)/\log \eta_1$. If $\eta \geq \eta_1$, then

$$\begin{aligned} F_{\mathcal{A}}(\eta) &\leq \sum_{1 \leq |S| \leq \lambda/2} \widehat{f}(S)^2 \eta^{2|S|} + \eta^\lambda \sum_S \widehat{f}(S)^2 \\ &\leq (\eta/\eta_1)^\lambda F_{\mathcal{A}}(\eta_1) + \eta^\lambda = 2\eta^\lambda \,. \end{aligned} \tag{2.7}$$

Assume that $I\!I(\mathcal{A}) \in (0, e^{-2})$, and let $a := \min\Big\{-\log I\!I(\mathcal{A})/\log n, 1/2\Big\}$. We may choose $\eta_1 := \sqrt{a/2}$. Then $I\!I(\mathcal{A}) \leq n^{-a}$, and therefore (2.6) and the definition of λ give

$$\lambda \geq \frac{a \log n}{3 \log(1/a)} \,. \tag{2.8}$$

The definition of a together with (2.4) and (2.8) show that $\lambda \to \infty$ as $m \to \infty$. Hence (2.7) implies $F_{\mathcal{A}_m}(1/2) \to 0$ as $m \to \infty$, which completes the proof. $\square$

PROOF OF 1.4: The above calculations together with Prop. 2.2 show that

$$\phi(\mathcal{A},\epsilon) \le \mathrm{VAR}(\mathcal{A},\epsilon)^{1/3} = F_{\mathcal{A}}(1-2\epsilon)^{1/3} \le 2^{1/3}\Big(1-2\epsilon\Big)^{\frac{a\log n}{9\log(1/a)}},$$

for $\epsilon \in (0,1/4)$, when we assume $II(\mathcal{A}) \le n^{-a}$, $a \in (0,1/2]$. The theorem follows immediately. □

For the proof of Theorem 1.2, we will need the following.

Theorem 2.6. *For each $k = 1,2,\dots$, there is a constant $C_k < \infty$ with the following property. Let $\mathcal{A} \subset \Omega_n$ be a monotone event and $f = \chi_{\mathcal{A}}$. Then*

$$\sum_{|S|=k} \widehat{f}(S)^2 \le C_k II(A)\Big(-\log II(A)\Big)^{k-1}.$$

This inequality was proved by Talagrand [27] for $k = 2$. (Talagrand considers an extension of this relation for two events, and our generalization applies for that extension as well.)

PROOF OF 2.6: To prove the theorem one can follow Talagrand's proof almost word-by-word. We will only describe the changes needed to adapt the proof. One modification required is that the inequality

$$\mathbf{P}\left\{S' : \left|\sum_{|S|=k} \alpha_S u_S(S')\right| \ge t\right\} \le e^2 \exp\left(-t^{2/k}\Big(e^2\sum \alpha_S^2\Big)^{-1/k}\right) \tag{2.9}$$

must be used in place of the sub-Gaussian estimate that appears as Prop. 2.1 in [27]. Set $q = t^{2/k}/\left(e^2\sum\alpha_S^2\right)^{1/k}$. For $q \le 2$ the inequality (2.9) is trivial, while for $q > 2$ it follows by substituting q into

$$\left\|\sum_{|S|=k} \alpha_S u_S\right\|_q \le (q-1)^{k/2}\Big(\sum \alpha_S^2\Big)^{1/2}, \qquad \forall q \ge 2, \tag{2.10}$$

which appears in [25] as (2.4) and is a consequence of the dual version of the Bonami-Beckner inequality.

Set $A_k := \{x \in \mathcal{A} : \sigma_k x \notin \mathcal{A}\}$, and note that $2\mathbf{P}[A_k] = I_k(\mathcal{A})$. In the proof for the case $k = 2$, Talagrand considers in Section 3 of [27] partitions

$I \cup J = [n]$, and estimates $\sum\{I_j(\mathcal{A})^2 : j \in L(s)\}$, where $L(s)$ is the set of $j \in J$ such that

$$\sum_{i \in I} \left(\int_{A_j} u_{\{i\}}(x) \right)^2 \geq s^2 \mathbf{P}[A_j]^2 .$$

To generalize Talagrand's argument for $k > 2$, one gives a similar estimate to $\sum\{I_j(\mathcal{A})^2 : j \in L_{k-1}(s)\}$, where $L_{k-1}(s)$ is the set of $j \in J$ such that

$$\sum \left\{ \left(\int_{A_j} u_i(x) \right)^2 : i \subset I,\ |i| = k-1 \right\} \geq s^2 \mathbf{P}[A_j]^2$$

We omit the details, since from this point on only straightforward changes are required to adapt Talagrand's beautiful (but rather mysterious) argument. $\square$

In the case of monotone events, Theorem 1.2 follows immediately from Theorems 2.6 and 1.8. In order to get rid of the monotonicity assumption, we introduce the shifting operator.

Let $j \in \{1, \ldots, n\}$, and let $f : \Omega_n \to \mathbb{R}$. For $x \in \Omega_n$, set

$$\kappa_j f(x) := \begin{cases} \max\{f(x), f(\sigma_j x)\}, & \text{if } x_j = 1, \\ \min\{f(x), f(\sigma_j x)\}, & \text{if } x_j = 0. \end{cases}$$

The operator κ_j is called the **j-shift**. The following lemma describes some useful properties of shifts.

Lemma 2.7 (Shifting). *Let $f : \Omega_n \to \mathbb{R}$, and let $j, i \in \{1, \ldots, n\}$. Then*

1. *$\kappa_1 \kappa_2 \ldots \kappa_n f$ is monotone.*
2. *$I_i(\kappa_j f) \leq I_i(f)$.*
3. *$\mathrm{VAR}(\kappa_j f, \epsilon) \geq \mathrm{VAR}(f, \epsilon)$ for each $\epsilon \in [0,1]$.*

PROOF: Suppose for the moment that $i \neq j$. For any $a, b \in \{0,1\}$ and $x \in \Omega_n$, let $x_{a,b}$ be x with the i'th coordinate set to a and the j'th coordinate set to b. Note that $\kappa_j f$ is monotone nondecreasing in the variable x_j. Hence $\kappa_i \kappa_j f(x_{1,1})$ is the maximum of f on $\{x_{0,0}, x_{0,1}, x_{1,0}, x_{1,1}\}$ and $\kappa_i \kappa_j f(x_{0,0})$ is the minimum. It follows that $\kappa_i \kappa_j f = \kappa_j \kappa_i \kappa_j f$. This relation easily implies the first claim of the lemma.

For the second part, we may assume with no loss of generality that $j \neq i$, because $I_i(\kappa_i f) = I_i(f)$. A case by case analysis shows that

$$\begin{aligned}&|f(x_{0,0}) - f(x_{1,0})| + |f(x_{0,1}) - f(x_{1,1})| \\ &\qquad \geq |\kappa_j f(x_{0,0}) - \kappa_j f(x_{1,0})| + |\kappa_j f(x_{0,1}) - \kappa_j f(x_{1,1})|,\end{aligned}$$

and the second part follows by summing over $x \in \Omega_n$.

For the last part, set

$$\begin{aligned} g(y) &= \mathbf{E}[f(N_\epsilon(x)) \mid x = y], \\ \widetilde{g}(y) &= \mathbf{E}[\kappa_j f(N_\epsilon(x)) \mid x = y]. \end{aligned}$$

Note that $g(y) + g(\sigma_j(y)) = \widetilde{g}(y) + \widetilde{g}(\sigma_j y)$, but $|g(y) - g(\sigma_j(y))| \leq |\widetilde{g}(y) - \widetilde{g}(\sigma_j y)|$. This implies $g(y)^2 + g(\sigma_j(y))^2 \leq \widetilde{g}(y)^2 + \widetilde{g}(\sigma_j y)^2$. By summing over y, we obtain $\mathbf{E}(g^2) \leq \mathbf{E}(\widetilde{g}^2)$. Since $\mathbf{E}g = \mathbf{E}\widetilde{g}$, the last claim of the lemma now follows. □

PROOF OF 1.2: Let $\mathcal{A} \subset \Omega_n$. Set $g = \kappa_1 \kappa_2 \ldots \kappa_n \chi_{\mathcal{A}}$. Then by Lemma 2.7, g is monotone, $II(g) \leq II(\mathcal{A})$ and for each $\epsilon > 0$ we have $\mathrm{VAR}(g, \epsilon) \geq \mathrm{VAR}(\mathcal{A}, \epsilon)$. Moreover, g takes only the values 0 and 1. By applying Theorem 2.6 for g, and using Proposition 2.2, Theorem 1.2 immediately follows. □

PROOF OF 1.3: Observe that for a monotone $f : \Omega_n \to \mathbb{R}$

$$I_j(f) = 2|\widehat{f}(\{j\})|, \tag{2.11}$$

and therefore

$$II(f) = 4 \sum_j \widehat{f}(\{j\})^2. \tag{2.12}$$

Hence 1.3 follows from Theorem 1.8. □

Note that (2.12) implies that

$$II(\mathcal{A}) \leq 4 \tag{2.13}$$

for monotone events $\mathcal{A}$.

Remark 2.8. It is tempting to look for a simpler proof of Theorem 1.2, along the following lines. Using (2.5) with $p = 2$, we find that

$$II(f) = \sum_{j=1}^{n} \left(\sum_{S \supset \{j\}} 4\widehat{f}(S)^2 \right)^2 = 16 \sum_{S,S'} \widehat{f}(S)^2 \widehat{f}(S')^2 |S \cap S'|, \qquad (2.14)$$

where $f = \chi_{\mathcal{A}}$ for some event $\mathcal{A} \subset \Omega_n$. This expression is more complicated than (2.12), but is still valid when $\mathcal{A}$ is not monotone. The fact that f is the indicator function of an event is summarized by the equation $f^2 = f$. In terms of the Fourier transform, this translates to a convolution equation

$$\widehat{f} * \widehat{f} = \widehat{f}. \qquad (2.15)$$

(By replacing f with $2f - 1$, this transforms to the simpler looking $\widehat{f} * \widehat{f} = \chi_{\{\emptyset\}}$.) One may suspect that there should be a direct argument that uses only (2.15) and (2.14) to prove that for every $k = 1, 2, \ldots$

$$\sum_{|S|=k} \widehat{f}(S)^2 \to 0$$

when $II(f) \to 0$. Then Theorem 1.2 would follow from Theorem 1.8.

3 Correlation with majority

3.1 Uniform weights

Fix some $n \in \mathbb{N}$. Recall the definition (1.3) of the majority function M_K, and set $M = M_n = M_{[n]}$.

Theorem 3.1. *Let $f : \Omega_n \to [0, 1]$ be monotone. Then*

$$I(f) \le C\sqrt{n}\, \mathbf{E}(fM) \left(1 + \sqrt{-\log \mathbf{E}(fM)}\right),$$

where C is some universal constant.

Proof: Write $\overline{f}(k)$ for the average of f on the set $\left\{x : \sum_j x_j = k\right\}$:

$$\overline{f}(k) = \binom{n}{k}^{-1} \sum_{|x|=k} f(x).$$

Then

$$\mathbf{E}(fM) = 2^{-n} \sum_{k>\frac{n}{2}} \binom{n}{k} \left(\overline{f}(k) - \overline{f}(n-k)\right) . \tag{3.1}$$

Recall that $s_j x = (y_1, \dots y_n)$ where $y_j = 1 - x_j$ and $y_i = x_i$ for $i \neq j$. Then

$$I(f) = 2^{-n} \sum_x \sum_j |f(x) - f(s_j x)|.$$

Since f is monotone, $f(x) - f(s_j x) \geq 0$ when $x_j = 1$ and $f(x) - f(s_j x) \leq 0$ when $x_j = 0$. Hence the expression for $I(f)$ simplifies,

$$\begin{aligned} I(f) &= 2^{-n} \sum_x f(x)\Big(2|x| - n\Big) \\ &= 2^{-n} \sum_k \binom{n}{k} \overline{f}(k)(2k-n) \\ &= 2^{-n} \sum_{k>\frac{n}{2}} \binom{n}{k} \left(\overline{f}(k) - \overline{f}(n-k)\right)(2k-n). \end{aligned} \tag{3.2}$$

For any $\lambda \geq 0$ write $k(\lambda) = (n + \lambda\sqrt{n})/2$. Since $0 \leq \overline{f}(k) \leq 1$, by comparing (3.2) and (3.1), we obtain the following estimate.

$$\begin{aligned} I(f) &\leq (2k(\lambda) - n)\, \mathbf{E}(fM) + 2^{-n} \sum_{k>k(\lambda)} \binom{n}{k} \left(\overline{f}(k) - \overline{f}(n-k)\right)(2k-n) \\ &\leq \lambda\sqrt{n}\, \mathbf{E}(fM) + 2^{-n} \sum_{k>k(\lambda)} \binom{n}{k}(2k-n). \end{aligned} \tag{3.3}$$

Because there are constants $C_1, C_2 > 0$ such that

$$2^{-n}\binom{n}{k}(2k-n) \leq C_1 \exp\left(-\frac{(2k-n)^2}{C_2 n}\right) \tag{3.4}$$

holds for every n and k, by choosing $\lambda = C_3\sqrt{-\log \mathbf{E}(fM)}$, where C_3 is a sufficiently large constant, we get

$$2^{-n} \sum_{k>k(\lambda)} \binom{n}{k}(2k-n) \leq C_4\sqrt{n}\, \mathbf{E}(fM),$$

and the theorem follows from (3.3). $\square$

Given a set $K \subset [n]$, let M_K denote the majority function on the set K; that is,

$$M_K(x) = \begin{cases} -1 & \text{if } \sum_{j\in K} x_j < |K|/2\,; \\ 0 & \text{if } \sum_{j\in K} x_j = |K|/2\,; \\ 1 & \text{if } \sum_{j\in K} x_j > |K|/2\,, \end{cases}$$

Also set,

$$I_K(f) = \sum_{k\in K} I_k(f).$$

Corollary 3.2. *Let $K \subset [n]$ and suppose that $f : \Omega_n \to [0,1]$ is monotone. Then*

$$I_K(f) \le C\sqrt{|K|}\,\mathbf{E}(fM_K)\,\left(1+\sqrt{-\log \mathbf{E}(fM_K)}\right)\,,$$

where C is some universal constant.

Proof: Set $m = |K|$, and assume, that $K = \{1,\dots,m\}$. Given $z \in \Omega_m$, set

$$f_K(z) = 2^{m-n} \sum_{y\in\Omega_{n-m}} f(z,y).$$

Then f_K is monotone and $I(f_K) = I_K(f)$. Consequently, the corollary follows from Theorem 3.1. $\square$

Proof of 1.5: Assume, with no loss of generality, that

$$I_{j+1}(f) \le I_j(f) \tag{3.5}$$

for all $j \in \{1,\dots,n-1\}$. Cor. 3.2 implies that

$$\sum_{j=1}^{k} I_j(f) \le C_1\Lambda(f)\left(1+\sqrt{-\log\Lambda(f)}\right)\sqrt{k} \tag{3.6}$$

for some constant C_1 and every $k \in [n]$. Subject to these constraints and (3.5), $II(f)$ is maximized if equality occurs in (3.6) for every k. Therefore,

$$\begin{aligned} II(f) &\le C_1^2\Lambda(f)^2\left(1+\sqrt{-\log\Lambda(f)}\right)^2 \sum_{k=1}^{n}\left(\sqrt{k}-\sqrt{k-1}\right)^2 \\ &= O(1)\Lambda(f)^2\Big(1-\log\Lambda(f)\Big)\sum_{k=1}^{n} k^{-1} \\ &= O(1)\Lambda(f)^2\Big(1-\log\Lambda(f)\Big)\log n. \end{aligned}$$

This proves the first part of Theorem 1.5. The second part now follows from Theorem 1.2. $\square$

Theorem 1.5 tells us that if $\Lambda(\mathcal{A}_m) \to 0$ fast enough for monotone events $\mathcal{A}_m$, then they are asymptotically noise sensitive. Conversely, if a sequence of (not necessarily monotone) events satisfies $\inf_m \Lambda(\mathcal{A}_m) > 0$, then it is not asymptotically noise sensitive. This can be proven directly, and also follows from Lemma 3.7 below.

It is interesting to note that

Theorem 3.3. *Majority maximizes I among monotone events $\mathcal{A} \subset \Omega_n$.*

Although this follows from [10], the explicit statement does not appear there. See also Lem. 6.1 of Friedgut and Kalai [13].

3.2 General weights

We will investigate now some relations between noise-sensitivity and weighted majority functions. Several of the properties we need for weighted majority functions are easy to establish if the distribution of weights allows us to use a normal approximation for $f(x) = \sum_j w_j x_j$. But, as it turns out, working with arbitrary weights is harder.

Our first goal is to show that weighted majority functions are uniformly noise stable. This will imply the "only if" part of Theorem 1.6. For this, the following easy (and quite standard) lemma will be needed.

Lemma 3.4. *Let $w = (w_1, \dots, w_n) \neq 0$ and $f(x) = \sum_j w_j(2x_j - 1)$. Then*

$$\mathbf{P}[|f| \geq t\|w\|_2] \leq 3t^{-4}, \tag{3.7}$$

and

$$\mathbf{P}[|f| \leq 0.3\|w\|_2] \leq 0.92. \tag{3.8}$$

A much stronger estimate than (3.7) is known (see [22]).

Proof: Without loss of generality, we assume that $\|w\|_2 = 1$. Then $\mathbf{E}[f^4] = 3\|w\|_2^4 - 2\|w\|_4^4 \leq 3$. Hence (3.7) follows:

$$\mathbf{P}[|f| \geq t] = \mathbf{P}[f^4 \geq t^4] \leq t^{-4}\mathbf{E}[f^4] = 3t^{-4}.$$

This implies

$$\mathbf{E}[\mathbf{1}_{\{f^2>t\}} f^2] = \mathbf{P}[f^2 > t] + \int_{s\ge t} \mathbf{P}[f^2 > s]\, ds \le 3t^{-2} + t^{-3}\,.$$

Hence

$$\mathbf{E}[\mathbf{1}_{\{f^2\le t\}} f^2] = \mathbf{E}[f^2] - \mathbf{E}[f^2 \mathbf{1}_{\{f^2>t\}}] \ge 1 - 3t^{-2} - t^{-3}\,.$$

We choose $t = 10$, and obtain

$$\begin{aligned} 10 - 9.9\mathbf{P}[f^2 \le 1/10] &= 10\mathbf{P}[f^2 > 1/10] + \mathbf{P}[f^2 \le 1/10]/10 \\ &\ge \mathbf{E}[\mathbf{1}_{\{f^2\le 10\}} f^2] \ge 9/10\,, \end{aligned}$$

which gives (3.8). □

Lemma 3.5. *Let $b > 0$, let $v_1, \dots, v_d \ge b$, and let $g = \sum_{j=1}^d z_j v_j$, where $\mathbf{P}[z_j = 1] = \mathbf{P}[z_j = -1] = 1/2$, and the z_j are independent. Then for every $t \ge 1$ and every $s \in \mathbb{R}$,*

$$\mathbf{P}\big[|g - s| \le tb\big] \le c \cdot t/\sqrt{d}\,, \tag{3.9}$$

where c is some universal constant.

This lemma is a consequence of Theorem 2.14 in [22], for example. However, since the proof of that theorem is arduous, we now present a simple combinatorial proof.

Proof: Let x be a random uniform element in Ω_d, and let π be a random uniform permutation of $\{1, 2, \dots, d\}$. Let C be the collection of sets S that have the form $S = \{j : \pi(j) < r\}$ for some $r \in \mathbb{R}$. Then there is a unique $y \in C$ with $|y| = |x|$. Observe that y is a random uniform element of Ω_d. Consequently, the distribution of g is the same as the distribution of $h(y) := \sum_{j=1}^d (1 - 2y_j)v_j$, where y_j is 1 or 0 when $j \in y$ or $j \notin y$, respectively. Since C is totally ordered by inclusion, there is at most one $S \in C$ such that $|h(S) - s| < b/2$. So when π is fixed, the probability that $|h(S) - s| \le b/3$ is at most $\max\Big\{\mathbf{P}\big[|x| = r\big] : r \in \mathbb{R}\Big\} = O(1)/\sqrt{d}$. This establishes (3.9) for $t = 1/3$. The result for general $t \ge 1$ follows by applying the result for $t = 1/3$ for an appropriate succession of values of s. □

Proof of 1.7: Let $w = (w_1, \dots, w_n) \ne 0$ and $s_0 \in \mathbb{R}$. Let $f(x) := \sum_{j=1}^n w_j(2x_j - 1)$, and consider the event $\mathcal{M} := \big\{x \in \Omega : f(x) > s_0\big\}$. Take

$\epsilon > 0$, and let $J \subset [n]$ be a random subset, where each $j \in [n]$ is in J with probability ϵ, independently. Set $Y(J) := \sum_{j\in J} w_j(2x_j - 1)$. Then $2Y(J)$ has the distribution of $f - N_\epsilon f$. Let $\delta \in (0,1)$ and set

$$a := \inf\Big\{t > 0 : \mathbf{P}\big[|Y(J)| \geq t\big] \leq \delta\Big\}. \tag{3.10}$$

Our goal is to give an estimate from above to $P\big[|f| < 2a\big]$ in terms of ϵ and δ, which will tend to zero when δ is positive and fixed and $\epsilon \to 0$.

Set $W(J) := \sum_{j\in J} w_j^2$. This is the variance of $Y(J)$ conditioned on J. Note that

$$\mathbf{P}\big[|Y(J)| \geq a \mid J\big] = \mathbf{P}\big[Y(J)^2 > a^2 \mid J\big] \leq \mathbf{E}\big[Y(J)^2 \mid J\big]/a^2 = W(J)/a^2.$$

Therefore,

$$\begin{aligned}\delta = \mathbf{P}\big[|Y(J)| > a\big] &= \sum_{X\subset[n]} \mathbf{P}\big[|Y(J)| > a \mid J = X\big]\mathbf{P}[J = X]\\ &\leq \sum_{X\subset[n]} \min\big\{1, a^{-2}W(X)\big\}\mathbf{P}[J = X] = \mathbf{E}\Big[\min\big\{1, a^{-2}W(J)\big\}\Big],\end{aligned} \tag{3.11}$$

and we conclude that

$$\mathbf{P}\big[W(J) \geq \delta a^2/2\big] \geq \delta/2\,. \tag{3.12}$$

Now let $z_1, z_2, \dots, z_n$ be independent variables that are uniform in $[0,1]$, and are independent from $(x_1, \dots, x_n)$. Let m be the largest integer such that $m\epsilon \leq 1$. Let $I_1, \dots, I_m$ be disjoint open intervals in $[0,1]$, each of length ϵ. Let $I_0 := [0,1] - \cup_{k=1}^m I_k$. Let J_k ($k = 0, 1, \dots, m$) be the set of $i \in [n]$ with $z_i \in I_k$. Then each J_k with $k > 0$ has the same distribution as J above. Let $\mathcal{A}_k$ be the event that $W(J_k) \geq \delta a^2/2$. Then from (3.12) with J_k in place of J we find that $\mathbf{P}[\mathcal{A}_k] \geq \delta/2$ for $k = 1, 2, \dots, m$.

We claim that for $k \neq k'$ the events $\mathcal{A}_k$ and $\mathcal{A}_{k'}$ are negatively correlated. This can be established by proving by induction on n that the events $W(J_k) \geq s_1$ and $W(J_{k'}) \geq s_2$ are negatively correlated for each $s_1, s_2 \in \mathbb{R}$ (which is intuitively obvious, since the intervals I_k and $I_{k'}$ are disjoint). Let K be the number of $k > 0$ such that the event $\mathcal{A}_k$ occurs. Then

$$\mathbf{E}[K] = \sum_{k=1}^m \mathbf{P}[\mathcal{A}_k] \geq m\delta/2\,,$$

and

$$\mathbf{E}[K^2] - \mathbf{E}[K]^2 = \sum_{k,k'} \mathbf{P}[\mathcal{A}_k \cap \mathcal{A}_{k'}] - \left(\sum_k \mathbf{P}[\mathcal{A}_k]\right)^2 \le \mathbf{E}[K],$$

because the events $\mathcal{A}_k, \mathcal{A}_{k'}$ are negatively correlated when $k \neq k'$. Therefore

$$\begin{aligned}\mathbf{P}[K < m\delta/4] &\le \mathbf{P}\left[(K - \mathbf{E}K)^2 > (m\delta/4)^2\right] \\ &\le 4(m\delta)^{-2}\mathbf{E}\left[(K - \mathbf{E}K)^2\right] = 4(m\delta)^{-2}\left(\mathbf{E}[K^2] - (\mathbf{E}K)^2\right) \\ &\le 4(m\delta)^{-2}\mathbf{E}[K] \le 4m^{-1}\delta^{-2}\,.\end{aligned} \tag{3.13}$$

Let L be the set of $k \in [m]$ such that $|Y(J_k)| > a\sqrt{\delta}/10$. By (3.8), applied to $Y(J_k)$ in place of f,

$$\mathbf{P}[k \in L \mid \mathcal{A}_k] \ge 8/100.$$

Moreover, conditioned on all the J_k, the events $\{k \in L\}$ are independent. Consequently, a calculation similar to (3.13) gives

$$\mathbf{P}\Big[|L| < m\delta/100 \mid K \ge m\delta/4\Big] \le O(1)m^{-1}\delta^{-2}.$$

When we use this and (3.13) together, we get

$$\mathbf{P}\Big[|L| < m\delta/100\Big] \le O(1)m^{-1}\delta^{-2}. \tag{3.14}$$

If we condition on L, on all $Y(J_k)$ for $k \notin L$ and on all $|Y(J_k)|$ for $k \in L$, then what remains to determine f are only the signs of $Y(J_k)$ with $k \in L$. Moreover, these signs are independent, and are $+$ or $-$ with probability $1/2$. Hence we may apply Lemma 3.5 with $b := a\sqrt{\delta}/10$, $d := |L|$, $s := s_0 - \sum_{k \notin L} Y(J_k)$, $g = \sum_{k \in L} Y(J_k)$, and take $v = (v_k)$ to be the sequence $\big(|Y(J_k)| : k \in L\big)$. The conclusion is that for $t \ge 1$

$$\mathbf{P}\left[|f - s_0| \le ta\sqrt{\delta}/10 \;\middle|\; |L| \ge m\delta/100\right] \le O(1)t/\sqrt{m\delta}\,.$$

Together with (3.14), (and choosing $t = 20/\sqrt{\delta}$) this gives

$$\mathbf{P}\Big[|f - s_0| \le 2a\Big] \le O(1)\left(\epsilon\delta^{-2} + \sqrt{\epsilon}\delta^{-1}\right). \tag{3.15}$$

We now come to analyze the effect of noise. Because $2Y(J)$ has the same distribution as $f - N_\epsilon f$, for every $a > 0$

$$\mathbf{P}\big[\mathcal{M} \triangle N_\epsilon \mathcal{M}\big] \le \mathbf{P}\big[|f - s_0| \le 2a\big] + \mathbf{P}\big[|Y(J)| \ge a\big]$$

Choose $\delta := \epsilon^{1/4}$ and, as before, use (3.10) to define a. Then $\mathbf{P}\big[|f - s_0| \le 2a\big] \le O(1)\epsilon^{1/4}$ and $\mathbf{P}\big[|Y(J)| \ge a\big] \le \epsilon^{1/4}$. Consequently,

$$\mathbf{P}\big[\mathcal{M} \triangle N_\epsilon \mathcal{M}\big] \le O(1)\epsilon^{1/4}, \tag{3.16}$$

and the theorem immediately follows. $\square$

Question 3.1. What is the best exponent possible on the right hand side of (3.16)?

Remarks 3.6. It follows from Theorems 1.7 and 1.2 that $\inf\big\{II(\mathcal{M}) : \mathcal{M} \in \mathfrak{M}\big\} > 0$. (A direct proof will follow.) We conjecture that $II(\mathcal{M})$ is minimized among $\mathcal{M}_{w,0} \subset \Omega_n$ in $\mathfrak{M}$ when all the weights are equal. It is a consequence of Theorems 1.7 and 1.8 that

$$\lim_{k\to\infty} \sup_{\mathcal{M}\in\mathfrak{M}} \sum_{|S|>k} \widehat{\chi}_{\mathcal{M}}(S)^2 = 0.$$

We actually expect that among weighted majority events in Ω_n, the one with equal weights is the least stable, and for every $k > 1$ maximizes $\sum_{|S|>k} \widehat{\chi}_{\mathcal{M}}(S)^2$.

For the proof of 1.6, the following will be needed.

Lemma 3.7. *Let $\mathcal{A}_m, \mathcal{B}_m \subset \Omega_{n_m}$ be two sequences of events. Suppose that the sequence $\{\mathcal{A}_m\}$ is noise-sensitive, while the sequence $\{\mathcal{B}_m\}$ is noise-stable. Then*

$$\lim_m \mathbf{P}[\mathcal{A}_m \cap \mathcal{B}_m] - \mathbf{P}[\mathcal{A}_m]\mathbf{P}[\mathcal{B}_m] = 0.$$

PROOF: This can be proven directly, but since $\mathbf{P}[\mathcal{A}_m \cap \mathcal{B}_m] = \mathbf{E}\big[\chi_{\mathcal{A}_m}\chi_{\mathcal{B}_m}\big]$, the lemma is immediate from 1.8. $\square$

Let the **influence vector** of an event $\mathcal{A} \subset \Omega_n$ be the vector $I^{\mathcal{A}} := \big(I_1(\mathcal{A}), \ldots, I_n(\mathcal{A})\big) \in \mathbb{R}^n$.

PROOF OF 1.6: The "only if" direction follows from Theorem 1.7 and Lemma 3.7.

For the other direction, we need to show that monotone, noise-insensitive events $\mathcal{A} \subset \Omega_n$ have a non-vanishing correlation with some weighted majority event $\mathcal{M}_w$, $w \in [0,1]^n$. Talagrand's Theorem 1.1 [27] gives a lower bound on the correlation of monotone events. This theorem asserts, in particular, that for two monotone events, if the inner product of their influence vectors is bounded away from zero, then the correlation between them is also bounded away from zero[1].

We know from Theorem 1.2 that for noise-insensitive events, $\|I^{\mathcal{A}}\|_2$ is bounded away from zero. It remains to show that for every $v \in [0,1]^n$ with $\|v\|_2 = 1$, we can find a weighted majority function $\mathcal{M} = \mathcal{M}_w$, $w \in [0,1]^n$, such that the inner product $\langle I^{\mathcal{M}}, v \rangle$ is bounded away from zero. We will prove that this holds when one chooses $w := v$.

Given any $w \in \mathbb{R}^n$, $w \neq 0$, let $I^w \in \mathbb{R}^n$ denote the influence vector of $\mathcal{M}_w$, $I^w_j := I_j(\mathcal{M}_w)$.

Proposition 3.8. *There is an absolute constant $c > 0$ such that $\langle w, I^w \rangle \geq c$ for every $n = 1, 2, \ldots$ and every $w \in \mathbb{R}^n$ with nonnegative coordinates and $\|w\|_2 = 1$.*

We first need:

Lemma 3.9. *Let $v \in \mathbb{R}^n$, $v \neq 0$, have nonegative coordinates, and let $x \in \Omega_n$ be uniform-random. Set $g(x) := \sum_{j=1}^n (2x_j - 1)v_j$, and let J be an interval of the form $[s,t)$ or $(s,t]$ with $t - s \geq 2\|v\|_\infty$ and $\max\{-s,t\} \leq \|v\|_2$. Then*

$$\mathbf{P}\big[g(x) \in J\big] \geq C(t-s)/\|v\|_2\,,$$

where $C > 0$ is an absolute constant independent of n, v and J.

Proof: We assume, as we may, that $t - s = 2\|v\|_\infty$, because J contains at least $(4\|v\|_\infty)^{-1} \cdot \mathrm{length}(J)$ disjoint intervals of length $2\|v\|_\infty$. With no loss of generality, we also assume for convenience that $v_1 \geq v_2 \geq \cdots \geq v_n$.

First, consider the case $v_{n-1} \geq (t-s)/5$. It will be shown that in this case the lemma holds even when the inequality

$$\max\{-s,t\} \leq \min\big\{2\|v\|_2, \|v\|_1\big\}, \tag{3.17}$$

replaces the requirement $\max\{-s,t\} \leq \|v\|_2$. Note that $\|v\|_1 \geq \|v\|_2$, so (3.17) is weaker.

[1] For uniformly stable events, it seems that also the converse is true: if the correlation is bounded away from zero, then so is the inner product of their influence vectors. For monotone uniformly stable events, this follows from the two-event version of Theorem 2.6.

Let π be a uniform-random permutation of $[n]$, let $h := \|x\|_1$, and let $y = y(\pi, h) \in \Omega_n$ be defined as in Lemma 3.5; that is, $y_j = 1$ iff $\pi(j) \le h$. From $\max\{-s, t\} \le \|v\|_1$ we get $g\big(y(\pi, 0)\big) \le s < t \le g\big(y(\pi, n)\big)$. As h increases from $k-1$ to k, $g(y)$ increases by $2v_{k'} \le t - s$, where $k' = \pi^{-1}(k)$. Hence, for every π there is some $k_\pi \in \{0, 1, \dots, n\}$ such that $g\big(y(\pi, k_\pi)\big) \in J$. We now estimate the probability that $|k_\pi - n/2|$ is large. Let m be the integer part of $n/2$, and let $a_\pi := g\big(y(\pi, m)\big)$. Since we are assuming that $(t-s)/2 \ge v_j$ for all j and $v_j \ge v_{n-1} \ge (t-s)/5$ for all $j \in \{1, 2, \dots, n-1\}$, it follows that the ratio $\sqrt{n}(t-s)/\|v\|_2$ is bounded away from 0 and ∞, and

$$\begin{aligned} |k_\pi - m| &\le 2 + 5\min\big\{|a_\pi - s|, |a_\pi - t|\big\}/(t-s) \\ &\le 2 + 5|a_\pi|/(t-s) + 5\max\{-s, t\}/(t-s) \\ &\le 2 + O(1)|a_\pi|/(t-s) + O(1)\|v\|_2/(t-s) \\ &\le O(1)|a_\pi|/(t-s) + O(1)\sqrt{n}. \end{aligned} \tag{3.18}$$

We now estimate the probability that a_π is large. Note that when we condition on $\|x\|_1 = h = m$, the bits x_j and x_i are negatively correlated when $i \ne j$. Consequently,

$$\mathbf{E}\big[a_\pi^2\big] = \mathbf{E}\big[g(x)^2 \mid h = m\big] = \sum_{j,i} v_j v_i \mathbf{E}\big[(2x_j - 1)(2x_i - 1) \mid h = m\big] \le \|v\|_2^2,$$

which implies

$$\mathbf{P}\big[|a_\pi| \le 2\|v\|_2\big] \ge 1/2.$$

By (3.18), when $|a_\pi| \le 2\|v\|_2$, we have $|k_\pi - n/2| \le O(1)\sqrt{n}$. Since $\mathbf{P}[h = k] = 2^{-n}\binom{n}{k}$, it follows that

$$\begin{aligned} \mathbf{P}[g(x) \in J] = \mathbf{P}[g(y) \in J] &\ge \mathbf{P}[h = k_\pi] \\ &\ge \mathbf{P}\Big[h = k_\pi \Big| |a_\pi| \le 2\|v\|_2\Big]\mathbf{P}\Big[|a_\pi| \le 2\|v\|_2\Big] \\ &\ge c_1/\sqrt{n} \ge c_2(t-s)/\|v\|_2, \end{aligned}$$

for some positive absolute constants c_1, c_2. This completes the argument for the case $v_{n-1} \ge (t-s)/5$.

We now work to reduce the case in which $\|v\|_2 \ge \max\{-s, t\}$ holds to the case where $v_{n-1} \ge (t-s)/5$ and 3.17 hold. Define a sequence v'_j as follows. Given any $j \in [n]$, let $z(j)$ be the least $j' \in \{j+1, j+2, \dots, n\}$ such that

$$\left|\sum_{i=j+1}^{j'} (2x_i - 1)v_i\right| \ge (t-s)/5,$$

and if such a j' does not exist, we set $z(j) = n$. Let $j_0 = 0$ and inductively $j_{k+1} = z(j_k)$. Let $r := \inf\{k : j_k = n\}$, and define

$$u_k := \left| \sum_{i=j_{k-1}+1}^{j_k} (2x_i - 1)v_i \right|.$$

Set $u = (u_1, u_2, \ldots, u_r)$. Note that when we condition on u, we have $g(x) = \sum_{i \le r}(2x'_i - 1)u_i$, where $x' = (x'_1, \ldots, x'_r) \in \Omega_r$ is random-uniform; that is, $g(x)$ is a sum of the u_i with signs that are independent and have probability $1/2$ to be $+$. Note also that $(t-s)/2 \ge u_j \ge (t-s)/5$, for $j = 1, \ldots, r$, except that at $j = r$ the inequality $u_j \ge (t-s)/5$ may fail.

The difficulty in applying the above analysis with u (with the entries sorted by size) in place of v is that we have to verify the inequality (3.17) for u in place of v. Let $\mathcal{A}$ be the event $\{\|u\|_2^2 \ge (9/10)\|v\|_2^2\}$. We want to show that $\mathbf{P}[\mathcal{A}]$ is bounded away from zero. Note that

$$\mathbf{E}[g(x)^2 \mid u] = \|u\|_2^2,$$

and so

$$\mathbf{E}[\|u\|_2^2] = \mathbf{E}\big[\mathbf{E}[g(x)^2 \mid u]\big] = \mathbf{E}[g(x)^2] = \|v\|_2^2. \tag{3.19}$$

As in Lemma 3.4, we use the identity $\mathbf{E}[g^4] = 3\|v\|_2^4 - 2\|v\|_4^4$ to compute

$$\mathbf{E}\Big[\|u\|_2^4\Big] \le \mathbf{E}\Big[3\|u\|_2^4 - 2\|u\|_4^4\Big] = \mathbf{E}\Big[\mathbf{E}\big[g^4 \mid u\big]\Big] = 3\|v\|_2^4 - 2\|v\|_4^4 \le 3\|v\|_2^4.$$

Together with (3.19) this easily implies that $\mathbf{P}[\mathcal{A}]$ is bounded away from zero, by an argument similar to the one used in Lemma 3.4.

Assume $\mathcal{A}$. Then in order to verify (3.17), one only needs to show that $\|u\|_1 \ge \|v\|_2$. We have,

$$(9/10)\|v\|_2^2 \le \|u\|_2^2 \le \|u\|_1\|u\|_\infty = \|u\|_1(t-s)/2.$$

So (3.17) with u in place of v follows if we have

$$(5/16)(t-s)^2 \le \|v\|_2^2.$$

Since $\mathbf{P}[\mathcal{A}]$ is bounded from below, it only remains to deal with the case

$$(5/16)(t-s)^2 > \|v\|_2^2. \tag{3.20}$$

Assume this inequality. Note that it implies that J occupies at least $2/\sqrt{5}$ of the interval $\Big[-\|v\|_2, \|v\|_2\Big]$. Recall that $v_1 = \|v\|_\infty = (t-s)/2$. Then (3.20) implies $\sum_{j>1} v_j^2 < (t-s)^2/16$. Set $g_1(x) = \sum_{j>1}(2x_j-1)v_j^2$. Then $\mathbf{E}[g_1(x)^2] < (t-s)^2/16$. Consequently,

$$\mathbf{P}\Big[|g_1(x)| < (t-s)/2\Big] \geq 1/2.$$

When we condition on this event and on $|g_1(x)|$, the sign of $g_1(x)$ and the sign of $(2x_1 - 1)$ can take all four possibilities in $(\pm, \pm)$ with probability $1/4$ each, and there is a choice of signs that gives $\pm g_1(x) \pm v_1 \in J$. Therefore $P[g(x) \in J] \geq 1/8$ when (3.20) holds. This concludes the proof. $\square$

Proof of 3.8: With no loss of generality, we assume that $w_j \geq w_{j+1}$ for all $j = 1, \ldots, n-1$. Set

$$f(x) := \sum_{k=1}^{n}(2x_k - 1)w_k, \qquad a_j := \sum_{k=1}^{j-1} w_k^2.$$

We shall show that when $a_j \leq 1/100$, the ratio I_j^w/w_j is not too small. That will imply the necessary inequality for $\langle w, I^w\rangle$.

Let

$$f_j(x) := \sum_{k=1}^{j-1}(2x_k - 1)w_k, \qquad \widetilde{f}_j(x) := \sum_{k=j+1}^{n}(2x_k - 1)w_k.$$

Note that $\mathbf{E}[f_j^2] = a_j$, and $\mathbf{E}[\widetilde{f}_j^2] = 1 - a_{j+1}$.

Consider first the case $j = 1$. Set $v := (w_2, w_3, \ldots, w_n)$ and $b := \min\big\{w_1, \|v\|_2\big\}$. Note that $b \geq \|v\|_\infty$. Hence Lemma 3.9 gives $\mathbf{P}[-b < \widetilde{f}_1 \leq b] \geq 2Cb/\|v\|_2$. Recall that I_1^w is the probability that x_1 is pivotal for the event $f > 0$. Hence

$$I_1^w = \mathbf{P}[-w_1 < \widetilde{f}_1 \leq w_1] \geq \mathbf{P}[-b < \widetilde{f}_1 \leq b] \geq 2Cb/\|v\|_2.$$

If $b = \|v\|_2$, this gives $I_1^w \geq 2C$, and otherwise it gives $I_1^w \geq 2Cw_1$. So we have verified that I_1^w/w_1 is bounded away from zero.

Now suppose that $j \in \{2, 3, \ldots, n\}$ satisfies $a_j < 1/100$. Set $v := (w_{j+1}, w_{j+2}, \ldots, w_n)$ and note that $\sqrt{a_j} \geq w_j \geq \|v\|_\infty$. Since $a_j < 1/100$ and

$a_j + w_j^2 + \|v\|_2^2 = 1$, this implies that $3\sqrt{a_j} + w_j < \|v\|_2$. Hence Lemma 3.9 implies that on the event $|f_j| \le 3\sqrt{a_j}$

$$\mathbf{P}[-w_j - f_j < \widetilde{f}_j \le w_j - f_j \mid f_j] \ge 2Cw_j/\|v\|_2 \ge 2Cw_j.$$

This gives

$$\begin{aligned} I_j^w &= \mathbf{P}[-w_j - f_j < \widetilde{f}_j \le w_j - f_j] \\ &\ge 2Cw_j \mathbf{P}\left[|f_j| \le 3\sqrt{a_j}\right] = 2Cw_j \mathbf{P}\left[f_j^2 \le 9\mathbf{E}(f_j^2)\right] \ge (16/9)Cw_j. \end{aligned}$$

We therefore conclude that $I_j^w \ge C'w_j$ for all $j \in [n]$ satisfying $a_j < 1/100$, where $C' > 0$ is some absolute constant. This gives,

$$\langle I^w, w\rangle \ge \sum \Big\{ I_j^w w_j : a_j < 1/100 \Big\} \ge \sum \Big\{ C' w_j^2 : a_j < 1/100 \Big\} \ge C'/100.$$

This completes the proof of the proposition, and the proof of Theorem 1.6. □

Remark 3.10. We now show that one cannot remove the log in Theorem 1.5. Fix some $k, n \in \mathbb{Z}$ with $n \ge k > 0$. Let $w_j = 1/\sqrt{j \log n}$ for $j = 1, \dots, n$, and let $u_j = 1/\sqrt{k}$ for $j \le k$ and $u_j = 0$ for $j > k$. Set $f_w(x) = \sum_{j=1}^n (2x_j - 1)w_j$ and $f_u(x) = \sum_{j=1}^n (2x_j - 1)u_j$, where $x \in \Omega_n$. Then the event $f_w \ge 0$ is noise stable, by 1.7. We show that $\mathbf{P}[f_w \ge 0 \mid f_u \ge 0] \to 1/2$ as $n \to \infty$, no matter how $k = k(n)$ is chosen. Indeed, given any $x \in \Omega$, let $s(x) := \sqrt{k} f_u(x) = \sum_{j \le k} (2x_j - 1)$. If $s(x) < 0$, let $\overline{x}$ be obtained from x by replacing $-s(x)$ of the 0 entries in x by 1's, where the set of entries replaced is chosen randomly and uniformly among all possibilities, and if $s(x) \ge 0$, set $\overline{x} = x$. Then $\mathbf{P}[f_w(x) \ge 0 \mid f_u(x) \ge 0] = \mathbf{P}[f_w(\overline{x}) \ge 0]$. Therefore, by Lemma 3.5 applied to w, it is enough to show that $f_w(\overline{x}) - f_w(x) \to 0$ in probability as $n \to \infty$. This follows from

$$\mathbf{E}\big|f_w(\overline{x}) - f_w(x)\big| = (2/k)\mathbf{E}\big[\max\{0, -s(x)\}\big] \sum_{j=1}^k w_j = O(1)/\sqrt{\log n}.$$

4 An application to percolation

Let R be an $(m+1)\times m$ rectangle in the square grid $\mathbb{Z}^2$, and let Ω be the set of all functions from E, the set of edges of R, to $\{0,1\}$. We identify Ω with Ω_n; where $n = n_m = |E| = 2m^2 - 1$. A point $x \in \Omega$ is called a **configuration**, and can be identified with the subgraph consisting of all vertices of R and all edges e with $x(e) = 1$. A connected component of this graph is called a **percolation cluster**.

Let $\mathcal{C} = \mathcal{C}_m \subset \Omega$ be the event that there is a left-right crossing of R; that is, $\mathcal{C}$ is the set of all configurations that contain a path joining the left and right boundaries of R. An easy and well known application of duality shows that $\mathbf{P}[\mathcal{C}] = 1/2$.

Kesten [19] gives an estimate from above for the probability that an edge near the middle of R is pivotal for $\mathcal{C}$. Similar estimates for edges near the boundary can probably be extracted from Kesten's paper. These give an inequality of the form $I_j(\mathcal{C}_m) \leq m^{-1-c}$, $c > 0$, for each j. Then Theorem 2.5 implies 1.1. However, we prefer to present another proof, based on Theorem 1.5.

The only percolation background needed to understand the proof is that in our situation the probability that a vertex in R is connected in the configuration to some vertex at Euclidean distance r is at most $Cr^{-1/\rho}$, for some constants $C, \rho > 0$. This follows from the celebrated Russo-Seymour-Welsh Theorem [23, 24] (see also [14]).

Proof of 1.1: Let E_r be the set of edges in the right half of R, with edges exactly centered included. Let $K \subset E_r$. We now estimate $\mathbf{E}(\chi_{\mathcal{C}} M_K)$.

Consider the following algorithmic method of randomly selecting a configuration. Let ω^K and $\widehat{\omega}^K$ be two independent elements of $\Omega_{|K|}$ and $\Omega_{n-|K|}$, respectively. Let V_1 be the set of vertices on the left boundary of R, and set VISITED $= \emptyset$. As long as there is some edge $[v,u] \notin$ VISITED joining a vertex $v \in V_1$ to a vertex $u \notin V_1$, choose some such edge $e = [v,u]$, and do the following. Append e to VISITED. If $e \in K$, let $y(e)$ be the first bit in the sequence ω^K that has not been previously used by the algorithm, while if $e \notin K$ let $y(e)$ be the first bit in the sequence $\widehat{\omega}^K$ that has not been previously used by the algorithm. If $y(e) = 1$, then adjoin to V_1 the vertex u.

This procedure defines y for all $e \in$ VISITED. Let $z \in \Omega$ be random, uniform, and independent of y, and let $x = y$ on VISITED while $x = z$ on

$E - \text{VISITED}$. This defines a configuration $x \in \Omega$.

The following is obvious:

Lemma 4.1. *The configuration x given by the above algorithm is uniformly distributed in Ω. The event $x \in \mathcal{C}$ is equal to the event that at the end of the algorithm V_1 intersects the right boundary and is independent from z (can be determined by y).* $\square$

Let us estimate the probability that $K \cap \text{VISITED}$ is large. An edge $e \in K$ is in VISITED iff there is in x a path joining a vertex of e to the left boundary of R. Since $K \subset E_r$, it follows from the above stated consequence of the Russo-Seymour-Welsh Theorem that the probability for the latter event is bounded by $Cm^{-1/\rho}$, for some constants $C, \rho > 0$. Consequently,

$$\mathbf{E}|K \cap \text{VISITED}| \leq C|K|m^{-1/\rho},$$

which implies

$$\mathbf{P}[\mathcal{A}_1] \leq Cm^{-1/(3\rho)},$$

where $\mathcal{A}_1$ is the event

$$\mathcal{A}_1 := \Big\{ x \in \Omega : |K \cap \text{VISITED}| \geq |K|m^{-2/(3\rho)} \Big\}.$$

Let $\mathcal{A}_2$ be the event that there is an integer j in the range $1 \leq j \leq |K|m^{-2/(3\rho)}$ such that

$$\left| \frac{j}{2} - \sum_{i=1}^{j} \omega_i^K \right| \geq \sqrt{|K|m^{-2/(3\rho)}} \log m.$$

It is easy to see that the $\mathbf{P}[\mathcal{A}_2]$ decays super-polynomially in m; in particular,

$$\mathbf{P}[\mathcal{A}_2] \leq O(1)m^{-1/\rho}.$$

As $\mathbf{P}[\mathcal{A}_1 \cup \mathcal{A}_2] \leq O(1)m^{-1/(3\rho)}$, we have

$$\mathbf{E}(\chi_{\mathcal{A}_1 \cup \mathcal{A}_2} \chi_{\mathcal{C}} M_K) \leq O(1)m^{-1/(3\rho)}. \tag{4.1}$$

Now suppose that the algorithm produced a y such that $\mathcal{A}_1 \cup \mathcal{A}_2$ does not hold. Then it follows that

$$\left| \frac{|\text{VISITED} \cap K|}{2} - \sum_{e \in \text{VISITED} \cap K} y(e) \right| \leq O(1)\sqrt{|K|m^{-2/(3\rho)}} \log m.$$

This implies that

$$\mathbf{E}[M_K(x) \mid y] \leq O(1)m^{-1/(3\rho)}\log m, \qquad \forall y \notin \mathcal{A}_1 \cup \mathcal{A}_2.$$

Since $x \in \mathcal{C}$ can be determined from y, we get

$$\mathbf{E}\Big((1 - \chi_{\mathcal{A}_1 \cup \mathcal{A}_2})\chi_{\mathcal{C}} M_K\Big) \leq O(1)m^{-1/(3\rho)}\log m.$$

In view of (4.1) this implies

$$\mathbf{E}(\chi_{\mathcal{C}} M_K) \leq O(1)m^{-1/(3\rho)}\log m,$$

and Cor. 3.2 gives

$$I_K(\mathcal{C}) \leq O(1)\sqrt{|K|}m^{-1/(3\rho)}(\log m)^{3/2} \tag{4.2}$$

for every $K \subset E_r$, since $\mathcal{C}$ is monotone. By symmetry, this would also hold for $K \subset E - E_r$, and therefore for every $K \subset E$. Consequently, by the proof of Theorem 1.5

$$I\!I(\mathcal{C}) \leq O(1)m^{-2/(3\rho)}(\log m)^4. \tag{4.3}$$

An appeal to Theorem 2.5 completes the proof. □

Remark 4.2. Since $I(\mathcal{C}) = \sum_e I_e(\mathcal{C})$ is also the expected number of pivotal edges for $\mathcal{C}$, (4.2) shows that the expected number of pivotal edges is bounded by

$$O(1)m^{1-1/(3\rho)}(\log m)^{3/2}.$$

Although this is better than the general bound of $O(1)m$ that follows from Theorem 3.3, a somewhat better bound can be extracted from Kesten's [19].

Corollary 4.3. *There is a constant $c > 0$ with the following property. If $\epsilon = c/\log m$, then for large m, with probability at least $1/4$, $\Big|\{x, N_\epsilon(x)\} \cap \mathcal{C}\Big| = 1$. That is, if each edge is switched with probability $c/\log m$, independently, then the crossing is likely to be created or destroyed.*

The corollary follows from (4.3) and Theorem 1.4. The details are left to the reader.

5 Some conjectures and problems concerning percolation

5.1 Other sensitivity conjectures

By Theorem 1.1 and Section 2, from knowing which edges are open for all but a small random set of edges, we have almost no information whether crossing occurs. This suggests that for some deterministic subsets of the rectangle $R = R_m$, knowing the configuration restricted to that configuration typically gives almost no information whether crossing occurs. It follows from the Russo-Seymour-Welsh Theorem [23, 24] that E_r, the set of edges in the right half of the rectangle, is not such a subset. Yet we believe that all the horizontal edges (or all the vertical edges) is such subset. That is, let $x, y \in \Omega$ be two independent uniform-random configurations. Let $z(e) = x(e)$ for horizontal edges e, and $z(e) = y(e)$ for vertical edges. Let $p(\omega) = \mathbf{P}[z \in \mathcal{C} | x = \omega]$.

Conjecture 5.1. For any $\epsilon > 0$, for all sufficiently large m,

$$\mathbf{P}\Big\{\omega \in \Omega : |p(\omega) - 1/2| > \epsilon\Big\} < \epsilon.$$

Here is a variant of this conjecture for **Voronoi percolation**. Fix a square in $\mathbb{R}^2$. Voronoi percolation is performed in two steps. First pick n points in the square uniformly and independently. Second each cell in the Voronoi tessellation determined by the chosen points is declared open with probability 1/2, and closed otherwise, independently of the other cells, (see Benjamini and Schramm [4] for the exact definitions and a study of Voronoi percolation). By duality, the probability of open left-right open crossing is 1/2. In the spirit of Theorem 1.1, we conjecture that typically, knowing the Voronoi tessellation (but not knowing which cells are open) gives almost no information whether an open left-right crossing exits.

5.2 Dynamical percolation

Dynamical percolation was introduced by Häggström, Peres and Steif [15]. Consider the following process. Let $\{X_e\}$ be independent Poisson point processes in $\mathbb{R}$ indexed by the edges $e \in E_R$ of the $(m+1) \times m$ rectangle $R = R_m$ in $\mathbb{Z}^2$. Let $x_0 : E_R \to \{0, 1\}$ be random-uniform. For each $t > 0$ set $x_t(e) := x_0(e)$ if the number of points in $(0, t] \cap X_e$ is even, and

$x_t(e) := 1 - x_0(e)$ if the number is odd. This gives a continuous time stationary Markov chain x_t in $\Omega = \{0,1\}^{E_R}$. Write $\widetilde{\mathbf{P}}$ for the probability measure governing this process. For each fixed t, the random variable x_t can be thought of as ordinary (Bernoulli(1/2)) percolation in $\mathbb{Z}^2$.

An interesting problem raised by [15] is weather there are (exceptional, random) times t in which there is an infinite percolation cluster in x_t. The result described below might be relevant.

As before, let $\mathcal{C}_m$ denote the set of configurations in Ω that have an open left-right crossing of R_m. For all t, $\widetilde{\mathbf{P}}[x_t \in \mathcal{C}_m] = 1/2$. Let S_m be the set of switching times; that is, S_m is the boundary of $\{t \geq 0 : x_t \in \mathcal{C}_m\}$. As a corollary of Theorem 1.1, we have,

Corollary 5.1. $\Big| S_m \cap [0,1] \Big| \to \infty$ *in probability.*

PROOF: Suppose $s > t \geq 0$. Observe that the distribution of the pair (x_t, x_s) is the same as the distribution of the pair $\Big(x_0, N_\epsilon(x_0)\Big)$, where ϵ is a function of $s - t$ and $\epsilon > 0$ when $s > t$. (Actually, $\epsilon/(s-t) \to 1$ as $s - t \to 0$.)

Let k be some positive integer, and set $\epsilon = \epsilon(1/k)$. Let $t_j := j/k$. Let $\mathcal{W}$ be the set of $\omega \in \Omega$ such that $\Big|\mathbf{P}[N_\epsilon(\omega) \in \mathcal{C}] - 1/2\Big| > 1/4$. Then $\mathbf{P}[\mathcal{W}] \to 0$ as $m \to \infty$, by Theorem 1.1. Let $\mathcal{Z}(a,b)$ be the event that $S \cap [a,b] = \emptyset$. Observe that for $\omega \notin \mathcal{W}$, we have

$$\widetilde{\mathbf{P}}\Big[\mathcal{Z}(t_j, t_{j+1}) \mid x_{t_j} = \omega\Big] \leq 3/4\,,$$

because $\mathcal{Z}(t_j, t_{j+1})$ is disjoint from the event $\Big|\{x_{t_j}, x_{t_{j+1}}\} \cap \mathcal{C}\Big| = 1$. Hence we can make the following estimate,

$$\begin{aligned}
\widetilde{\mathbf{P}}[\mathcal{Z}(0,t_{j+1})] &= \widetilde{\mathbf{P}}[\mathcal{Z}(0,t_j) \cap \mathcal{Z}(t_j,t_{j+1})] \\
&= \sum_{\omega\in\Omega} \widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j) \cap \mathcal{Z}(t_j,t_{j+1}) \mid x_{t_j} = \omega\Big]\mathbf{P}\{\omega\} \\
&= \sum_{\omega\in\Omega} \widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j) \mid x_{t_j} = \omega\Big]\widetilde{\mathbf{P}}\Big[\mathcal{Z}(t_j,t_{j+1}) \mid x_{t_j} = \omega\Big]\mathbf{P}\{\omega\} \\
&\qquad \text{(by the Markov property for } x_t\text{)} \\
&\leq \mathbf{P}[\mathcal{W}] + \sum_{\omega\in\Omega-\mathcal{W}} \widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j) \mid x_{t_j} = \omega\Big]\widetilde{\mathbf{P}}\Big[\mathcal{Z}(t_j,t_{j+1}) \mid x_{t_j} = \omega\Big]\mathbf{P}\{\omega\} \\
&\leq \mathbf{P}[\mathcal{W}] + (3/4) \sum_{\omega\in\Omega-\mathcal{W}} \widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j) \mid x_{t_j} = \omega\Big]\mathbf{P}\{\omega\}
\end{aligned}$$

$$\begin{aligned} &\leq \mathbf{P}[\mathcal{W}] + (3/4) \sum_{\omega \in \Omega} \widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j) \mid x_{t_j} = \omega\Big] \mathbf{P}\{\omega\} \\ &= \mathbf{P}[\mathcal{W}] + (3/4) \widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j)\Big]. \end{aligned}$$

Using this inequality and induction gives $\widetilde{\mathbf{P}}\Big[\mathcal{Z}(0,t_j)\Big] \leq 4\mathbf{P}[\mathcal{W}] + (3/4)^j$. By stationarity, for every $t \geq 0$, the same estimate for the probability of $\mathcal{Z}(t, t + j/k)$ holds. Since k may be chosen arbitrarily large, and $\mathbf{P}[\mathcal{W}] \to 0$ as $m \to \infty$, the corollary easily follows. □

5.3 Limits and conformal invariance

The motivating questions behind this work were the limit conjectures and conformal invariance conjectures regarding percolation. See Langlands, Pouliot and Saint-Aubin [20].

Consider a triple $\mathcal{G} = \langle G, A, B\rangle$, where $G = (V, E)$ is a finite planar graph with m edges, and $A, B \subset V$. Let $p_{\mathcal{G}}$ be the probability that there is an open crossing from A to B in a uniform-random configuration $x \in \Omega = \{0,1\}^E$.

Let $\mathcal{H} = \langle H, A', B'\rangle$ be a triple obtained from G by the following operation: for every edge e of G delete e with probability $(1-t)/2$ contract e with probability $(1-t)/2$ and leave e unchanged with probability t, independently of the other edges. $\mathcal{H}$ is a random variable which takes values in planar graphs with two distinguished vertex sets.

Suppose that we take G to be the $(m+1) \times m$ rectangle in $\mathbb{Z}^2$ and let A and B be its left and right boundaries. It then follows from Theorem 1.4 and (4.3) that $p_{\mathcal{H}} - p_{\mathcal{G}} \to 0$ in probability, provided that $t \log m \to \infty$.

This result enables one to relate the crossing probabilities of percolation on different graphs. (Note that the result continues to hold even if t depends on the edge, provided that it is bounded from below by some t_0 with $t_0 \log m \to \infty$.) We hoped that such a result will be relevant to limit conjectures and to the conformal invariance conjecture. At present, however, such applications are beyond our reach. We do not have a good understanding of planar graphs which are obtained by random deletion/contraction of the form described above starting from the graph of a rectangle in the square grid.

5.4 Fourier-Walsh coefficients of percolation

It is a natural question to try to understand the Fourier-Walsh coefficients of boolean functions given by percolation problems. Consider (for example) the event $\mathcal{C} = \mathcal{C}_m$ of a left-right crossing of an $(m+1) \times m$ rectangle $R = R_m$ of the square grid, $\mathbb{Z}^2$, which was discussed in Section 4. Let $f_m := \chi_{\mathcal{C}_m}$. The Fourier coefficients of f_m are indexed by subsets of E_R, the edges in R. Given $S \subset E_R$, we denote by $\operatorname{diam}(S)$ the diameter of S as a set in the plane.

We conjecture that $\widehat{f}^2$ is supported mostly on sets of small diameter; that is,

Conjecture 5.2. For every $\epsilon > 0$,

$$\lim_{m\to\infty} \sum \Big\{ \widehat{f}_m^2(S) : \operatorname{diam}(S) > \epsilon m \Big\} = 0.$$

It would be interesting to formulate more precise conjectures regarding the support of $\widehat{f}_m^2$.

It follows from Theorem 1.4 and our estimates for $II(\mathcal{C})$, where $\mathcal{C}$ is the event of crossing for percolation, that all but a negligible part of the L^2 weight of the Fourier coefficients $\widehat{f}(S)$, where S is non-empty, is for $|S| > c \log m$. We conjecture that, in fact, this is true for $|S| > m^\beta$ for some $\beta > 0$. Conjecture 5.1 is equivalent to the statement that for all but a negligible part of these Fourier coefficients, the number of vertical edges in S tends to infinity with m.

5.5 Other models of statistical mechanics

It would be of interest to extend the results of this paper as well as earlier results on influence ([18, 13]) to other models of statistical mechanics, such as the Ising and Potts models. Many of the results on influence and on noise sensitivity should be extendible to measures on Ω_n for which the coordinate variables are **positively associated**, namely, measures for which every two monotone real functions are positively correlated.

6 Some further examples

We will discuss now two examples, both first considered by Ben-Or and Linial [3].

6.1 Tribes

Consider n boolean variables divided into t tribes $T_1, T_2 \dots, T_t$ of size s each, and let f be the boolean function which take the value 1 if for some j, $1 \le j \le t$, all variables of T_j equal 1. If $s = \log n - \log\log n + \log\log 2$, then $\mathbf{P}[f = 1] \approx \frac{1}{2}$. Also note that $I_k(f) \sim \log n/n$ for every k. It is easy to show directly that f will be immune to ϵ-noise when $\epsilon = o(1/\log n)$ and will be devastated by ϵ-noise if $\epsilon \log n \to \infty$. Thus, $J(f) \sim \log n/n$.

6.2 Recursive majority on the ternary tree

Consider $n = 3^t$ boolean variables which form the leaves of a rooted ternary tree of height t. A boolean function f is defined as follows: Given values for the variable on the leaves compute for each other vertex its value as the majority of the values of its sons and set the value of f to be the value of the root.

Ben Or and Linial showed that $I_k(f) \sim n^{-\log 2/\log 3}$ for every k and thus $\alpha(f) \to \log 2/\log 3$ as $t \to \infty$. It is easy to see that also $\beta(f) \to \log 2/\log 3$. This follows at once from the following observation: for $t = 1$, if we switch the value of each leaf with probability p independently, then for small p the probability that the outcome will be switched is $(3/2)p + o(p)$.

Conjecture 6.1. There is an absolute constant $\beta_0 < 1/2$ (find it!) such that for every monotone Boolean function f, $\beta(f) \le \beta_0$.

7 Relations with complexity theory

There is an interesting connection between the complexity of boolean functions and the notions studied in this paper.

7.1 AC0 and influences

An important complexity class AC0 of Boolean functions are those which can be expressed by Boolean circuits of polynomial size (in the number of variables) and bounded depth. Boppana [7] proved that if f is expressed by a depth-c circuit of size N then

$$I(f) \le C_1 \log^{c-1} N. \tag{7.1}$$

Earlier Linial, Mansour and Nisan [21] proved that the Fourier coefficients of functions which can be expressed by Boolean circuits of polynomial (or quasi-polynomial) size and bounded depth in AC0 decays exponentially above poly-logarithmic "frequencies". Both these results rely on the fundamental Håstad Switching Lemma, see [16, 1].

Recall that a **monotone circuit** is one where all the gates are monotone increasing in the inputs; i.e., there are no "not" gates. The Håstad lemma for monotone boolean circuits is easier and was proved already by Boppana [6].

We conjecture that a reverse relation to 7.1 also holds.

Conjecture 7.1 (Reverse Håstad). For every $\epsilon > 0$ there is a $K = K(\epsilon) > 0$ satisfying the following. For every monotone $\mathcal{A} \subset \Omega_n$, there is a $\mathcal{B} \subset \Omega_n$ such that $\mathbf{P}[\mathcal{A} \triangle \mathcal{B}] < \epsilon$ and B can be expressed as a Boolean circuit such that

$$(\log N)^{c-1} < KI(\mathcal{A}),$$

where c and N are the depth and size of the circuit, respectively.

Monotone Boolean functions with bounded influence were characterized by Friedgut [11, 12]. The results of [9] are also relevant to this conjecture.

Ha Van Vu raised the question if there is a spectral way to distinguish between bounded depth circuits of polynomial size and bounded depth circuits of quasi-polynomial size. In particular, he was looking for a way to show that the graph property "having a clique of size $\log n$" for graphs with n vertices, cannot be expressed by a bounded depth circuit of polynomial size. (Here the set of variables correspond to the $\binom{n}{2}$ possible edges.)

Conjecture 7.2. Let $\epsilon > 0$ be a fixed real number and $c \geq 1$ be a fixed integer. Let $\mathcal{A}$ be a monotone property expressed by a depth-c circuit of size M and let $f = \chi_{\mathcal{A}}$. Then there is a set $\mathcal{S}$ of polynomial size in M (where the polynomial depends on c and ϵ) so that

$$\sum \{\hat{f}^2(S) : S \notin \mathcal{S}\} \leq \epsilon.$$

This conjecture may also apply to TC0, see below. It would be of great interest to characterize Boolean functions for which most of the weight of the Fourier coefficients is concentrated on a set of polynomial size in n.

7.2 TC0 and noise sensitivity

Noise sensitivity seems related to another class of boolean functions - threshold circuits of bounded depths see [30, 17]. In a threshold circuit each gate is a weighted majority function.

Conjecture 7.3. Let f be a boolean function given by a monotone threshold circuit of depth c and size M. Then

$$J(f) = O(1)(\log M)^{c-1}.$$

Thus, for $1/\epsilon \leq O(1)(\log M)^{c-1}$ we expect that $\mathrm{VAR}(f, \epsilon)$ is bounded away from zero. Also here it is a tempting conjecture that a reverse relation holds.

We conjecture further that all functions f that can be expressed by a depth-c monotone threshold circuit where all the threshold gates are balanced are uniformly stable. (And in particular, $J(f) = O(1)$.) Possibly, functions in this class of functions approximate arbitrary well arbitrary uniform stable monotone Boolean functions.

Conjecture 7.3 implies theorems of Yau [30] and Håstad and Goldmann [17]. They proved that the and/or tree (or equivalently the example of ternary tree of Section 6) does not belong to monotone TC0; i.e., it cannot be expressed as a monotone bounded depth circuit of polynomial size.

The results of Yau and Håstad are still open for the non-monotone case. This would follow from the following strong form of our conjecture: Let f be a monotone boolean function given by a threshold circuit of depth c and size M. Then

$$J(f) = O(1)(\log M)^{c-1}.$$

8 Random walks

For nonempty $\mathcal{A} \subset \Omega_n$, consider a random walk defined as follows: start with a point chosen at random uniformly from $\mathcal{A}$, and at each step, stay where you are with probability $1/2$, and with probability $1/(2n)$ move to any one of the neighboring vertices. Let $\mathbf{P}^t_{\mathcal{A}}$ be the measure on Ω_n given by the location of the walk after t steps, and set

$$W(\mathcal{A}, \epsilon) := \inf\big\{t : \|\mathbf{P}^t_{\mathcal{A}} - \mathbf{P}\| < \epsilon\big\}.$$

Here $\|\mathbf{P}^t_{\mathcal{A}} - \mathbf{P}\|$ is the measure (L^1) norm of the difference between $\mathbf{P}^t_{\mathcal{A}}$ and the uniform measure.

Theorem 8.1. *Suppose that $\mathcal{A}_m \subset \Omega_{n_m}$ is a sequence of events satisfying* $\inf_m \mathbf{P}[\mathcal{A}_m] > 0$.

1. *$\{\mathcal{A}_m\}$ are asymptotically noise sensitive iff $\lim_m W(\mathcal{A}, \epsilon)/n_m = 0$ for every fixed $\epsilon > 0$.*

2. *If $\beta(\mathcal{A}_m) \to \beta$, then $W(\mathcal{A}_m, \epsilon) \le n^{1-\beta-o(1)}$.*

PROOF: Set $f_t(x) := 2^{n_m} \mathbf{P}^t_{\mathcal{A}_m}[\{x\}]$. Note that

$$f_{t+1} = (1/2) f_t + (2n_m)^{-1} \sum_{j=1}^{n_m} \sigma_j f_t.$$

Consequently,

$$\widehat{f}_t(s) = \left(\frac{2n_m - |s|}{2n_m}\right)^t \widehat{f}_0(s) = \mathbf{P}[\mathcal{A}_m]^{-1} \left(\frac{2n_m - |s|}{2n_m}\right)^t \widehat{\chi}_{\mathcal{A}_m}(s).$$

This gives for every $k = 1, 2, \dots$

$$\begin{aligned}
\|\mathbf{P}^t_{\mathcal{A}_m} - \mathbf{P}\|^2 \le \|f_t - 1\|_2^2 &= \mathbf{P}[\mathcal{A}_m]^{-1} \sum_{0 \ne s \in \Omega_{n_m}} \left(\frac{2n_m - |s|}{2n_m}\right)^{2t} \widehat{\chi}_{\mathcal{A}_m}(s)^2 \\
&\le \mathbf{P}[\mathcal{A}_m]^{-1} \left(\frac{2n_m - k}{2n_m}\right)^{2t} + \sum_{0<|s|<k} \widehat{\chi}_{\mathcal{A}_m}(s)^2 \\
&\le \mathbf{P}[\mathcal{A}_m]^{-1} \exp(-tk/n_m) + \sum_{0<|s|<k} \widehat{\chi}_{\mathcal{A}_m}(s)^2.
\end{aligned}$$

The theorem follows. $\square$

9 Changing a fixed size set of bits

The noise operator N_ϵ changes every input variable independently of the others, and the expected number of bits changed is ϵn, where n is the number of variables. Understanding the effect of different types of noise may be of interest. We consider a variant where a fixed number of bits are changed. In other words, for $x \in \Omega_n$ and $q \in [n]$, let $\widetilde{N}_q(x) = x \oplus s$, where s is chosen randomly uniformly among $s \in \Omega_n$ with cardinality q, independent from x. Here $\oplus$ is addition mod 2; that is, xor.

The analysis of the noise $\widetilde{N}_q$ is similar to that of N_ϵ, but a little care is needed. Consider the following example. Let $\mathcal{P} \subset \Omega_n$ consist of those $x \in \Omega_n$ such that $|x|$ is odd. This event $\mathcal{P}$ is called **parity**. Observe that for each fixed q, the conditioned probability $\mathbf{P}[\widetilde{N}_q(x) \in \mathcal{P} | x = y]$ is either zero or 1. In other words, knowing x allows a perfect prediction for $\widetilde{N}_q(x) \in \mathcal{P}$. Note that $\widehat{\chi}_{\mathcal{P}}(S)$ is nonzero only when $S \in \{\emptyset, [n]\}$. This means that the vanishing of the weight of the lower Fourier coefficients does not imply sensitivity to $\widetilde{N}_q$, as in Theorem 1.8.

For $f : \Omega_n \to \mathbb{R}$ and $q \in [n]$ set

$$\widetilde{\mathrm{VAR}}(f, q) = \mathrm{var}_y\Big(\mathbf{E}[f(\widetilde{N}_q(x))|x = y]\Big) = \mathbf{E}_y\mathbf{E}\Big(f(\widetilde{N}_q(x))|x = y\Big)^2 - (\mathbf{E}f)^2.$$

We say that a sequence of events $\mathcal{A}_m \subset \Omega_{n_m}$ is **asymptotically noise sensitive with respect to** $\widetilde{N}$ if for every $\epsilon \in (0, 1)$ and every sequence $\{q_m\}$ with $\epsilon n_m \le q_m \le (1 - \epsilon)n_m$, we have

$$\lim_m \widetilde{\mathrm{VAR}}(\mathcal{A}_m, q_m) = 0.$$

Note that this is equivalent to the straightforward analog of the definition for asymptotic noise sensitivity to our current setting.

Theorem 9.1. *Let $\mathcal{A}_m \subset \Omega_{n_m}$ be a sequence of events, and set $g_m = \chi_{\mathcal{A}_m}$.*

1. *This sequence is asymptotically noise sensitive with respect to $\widetilde{N}$ iff for every finite k*

$$\lim_m \sum\Big\{\widehat{g}_m(S)^2 : S \subset [n], 1 \le |S| \le k \text{ or } |S| \ge n - k\Big\} = 0.$$

2. *A sufficient condition for asymptotic noise sensitivity is $II(\mathcal{A}_m) \to 0$.*

PROOF: For $f : \Omega_n \to \mathbb{R}$ set

$$\widetilde{T}_q f(y) = \mathbf{E}f(\widetilde{N}_q(y)).$$

We now compute the Fourier coefficients of $\widetilde{T}_q f$. Take $r \in \Omega_n$.

$$\begin{aligned}
\mathbf{E}(\widetilde{T}_q f \cdot u_r) &= 2^{-n} \sum_x \widetilde{T}_q f(x)(-1)^{|r \cap x|} \\
&= 2^{-n} \binom{n}{q}^{-1} \sum_x \sum_{|s|=q} f(x \oplus s)(-1)^{|r \cap x|} \\
&= 2^{-n} \binom{n}{q}^{-1} \sum_y \sum_{|s|=q} f(y)(-1)^{|r \cap y|}(-1)^{|r \cap s|} \\
&= \widehat{f}(r) \binom{n}{q}^{-1} \sum_j (-1)^j \binom{|r|}{j} \binom{n-|r|}{q-j}.
\end{aligned}$$

Consequently,

$$\widetilde{T}_q f = \sum_{r \in \Omega_n} c(n, q, |r|) \widehat{f}(r) u_r, \tag{9.1}$$

where

$$c(n, q, k) = \binom{n}{q}^{-1} \sum_j (-1)^j \binom{k}{j} \binom{n-k}{q-j}.$$

Since $c(n, q, 0) = 1$, this gives,

$$\widetilde{\mathrm{VAR}}(f, q) = \|T_q f\|_2^2 - \widehat{f}(\emptyset)^2 = \sum_{\emptyset \neq S \subset [n]} c(n, q, |S|)^2 \widehat{f}(S)^2. \tag{9.2}$$

Consequently, for 9.1.1 it is enough to understand the behavior of the coefficients $c(n, q, k)$. For this, consider the sequence

$$a_j = \binom{k}{j} \binom{n-k}{q-j} \binom{n}{q}^{-1}.$$

The sequence has a unique maximum, which occurs when j is an integer j' close to qk/n. Consequently, $c(n, q, k) \leq 2a_{j'}$. Now let $n, k, q \to \infty$, and assume that $\epsilon n \leq q \leq (1-\epsilon)n$ and $n - k \to \infty$, where $\epsilon > 0$ is fixed. Then

$$\lim c(n, q, k) = 0.$$

This gives one direction in 9.1.1.

Also note that when $q < n/(3k)$, $|c(n,q,k)|$ is approximately a_0. This gives $\liminf |c(n,q,k)| > 0$ when k is fixed, $n \to \infty$ and q is about $n/3k$. Since $c(n,q,k) = \pm c(n,q,n-k)$, we get the other direction of 1.

Now assume that $II(\mathcal{A}_m) \to 0$. From Theorem 1.8 we know that

$$\lim_m \sum \Big\{ \widehat{g}_m(S)^2 : S \subset [n], 1 \le |S| \le k \Big\} = 0$$

for every fixed k. Equation (2.14) gives,

$$\begin{aligned} II(\mathcal{A}_m) &\geq \frac{n_m}{2} \sum \Big\{ \widehat{g}_m(S)^2 \widehat{g}_m(S')^2 : S, S' \subset [n],\ |S|, |S'| \ge 3n_m/4 \Big\} \\ &\geq \left(\sum_{|S| \ge 3n_m/4} \widehat{g}_m(S)^2 \right)^2 . \end{aligned}$$

Consequently,

$$\lim_m \sum \Big\{ \widehat{g}_m(S)^2 : S \subset [n], |S| \ge n-k \Big\} = 0$$

and the proof is complete. $\square$

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, New York (1992).

[2] W. Beckner, Inequalities in Fourier analysis, *Annals of Math.* **102** (1975), 159–182.

[3] M. Ben-Or and N. Linial, Collective coin flipping, in *Randomness and Computation* (S. Micali, ed.), Academic Press, New York, (1990), pp. 91–115. Earlier version: Collective coin flipping, robust voting games, and minima of Banzhaf value, Proc. 26th IEEE Symp. on the Foundation of Computer Science, (1985), pp. 408–416.

[4] I. Benjamini and O. Schramm, Conformal invariance of Voronoi percolation, *Commun. Math. Phys.*, **197** (1998), 75–107.

[5] A. Bonami, Etude des coefficients Fourier des fonctiones de $L^p(G)$, *Ann. Inst. Fourier* **20** (1970), 335–402.

[6] R. Boppana, Threshold functions and bounded depth monotone circuits, *Proceedings of 16th Annual ACM Symposium on Theory of Computing* (1984), 475–479.

[7] R. Boppana, The average sensitivity of bounded depth circuits, *Inform. Process. Lett.* **63** (1997) 257–261.

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

[9] J. Bourgain and G. Kalai, Influences of variables and threshold intervals under group symmetries, *Geom. Funct. Anal.*, to appear.

[10] J. T. Chayes, L. Chayes D. S. Fisher and T. Spencer, Finite-size scaling and correlation length for disordered systems, *Phys. Rev. Lett.* **57** (1986), 2999–3002.

[11] E. Friedgut, Boolean functions with low average sensitivity, *Combinatorica* (1997), To appear.

[12] E. Friedgut, Necessary and sufficient conditions for sharp thresholds of graphs properties and the k-sat problem, (1997) submitted.

[13] E. Friedgut and G. Kalai, Every monotone graph property has a sharp threshold, *Proc. Amer. Math. Soc.* **124** (1996), 2993–3002.

[14] G. Grimmett, *Percolation*, Springer-Verlag, Berlin (1989).

[15] O. Häggström, Y. Peres, and J. E. Steif, Dynamical percolation, *Ann. IHP* **33** (1997), 497–528.

[16] J. Håstad, Almost optimal lower bounds for small depth circuits, in *Randomness and Computation*, **5**, ed. S. Micali, (1989), 143–170.

[17] J. Håstad and M. Goldmann, On the power of small-depth threshold circuits, *Computational Complexity*, **1** (1991), 113–129.

[18] J. Kahn, G. Kalai and N. Linial, The influence of variables on boolean functions, *Proc. 29-th Ann. Symp. on Foundations of Comp. Sci.*, (1988), 68–80.

[19] H. Kesten, Scaling relations for $2D$-percolation, *Comm. Math. Phys.* **109** (1987), 109–156.

[20] R. P. Langlands, P. Pouliot, and Y. Saint-Aubin Conformal invariance in two-dimensional percolation, *Bull. Amer. Math. Soc. (N.S.)* **30** (1994), 1–61.

[21] N. Linial, Y. Mansour and N. Nisan, Constant depth circuits, Fourier transform, and learnability, *J. Assoc. Comput. Mach.* **40** (1993), 607–620.

[22] V. V. Petrov, *Limit theorems of probability theory*, Oxford University Press, (1995).

[23] L. Russo, A note on percolation, *ZW.* **43** (1978), 39–48.

[24] P. Seymour and D. Welsh, Percolation probabilities on the square lattice. Advances in Graph Theory. *Ann. Discrete Math.* **3** (1978), 227–245.

[25] M. Talagrand, On Russo's approximate zero-one law, *Ann. of Prob.* **22** (1994), 1576–1587.

[26] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publ. I.H.E.S.*, **81** (1995), 73–205.

[27] M. Talagrand, How much are increasing sets positively correlated? *Combinatorica* **16** (1996), 243–258.

[28] B. Tsilerson, Brownian coalescence as black noise I, preprint.

[29] B. Tsilerson, The Five noises, preprint.

[30] A. Yao, Circuits and local computation, *Proceedings of 21st Annual ACM Symposium on Theory of Computing*, (1989), 186–196.

The Weizmann Institute of Science, Rehovot 76100, Israel
itai@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/∼itai/

The Hebrew University, Givat Ram, Jerusalem 91904, Israel
kalai@math.huji.ac.il
http://www.ma.huji.ac.il/∼kalai/

The Weizmann Institute of Science, Rehovot 76100, Israel
schramm@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/∼schramm/