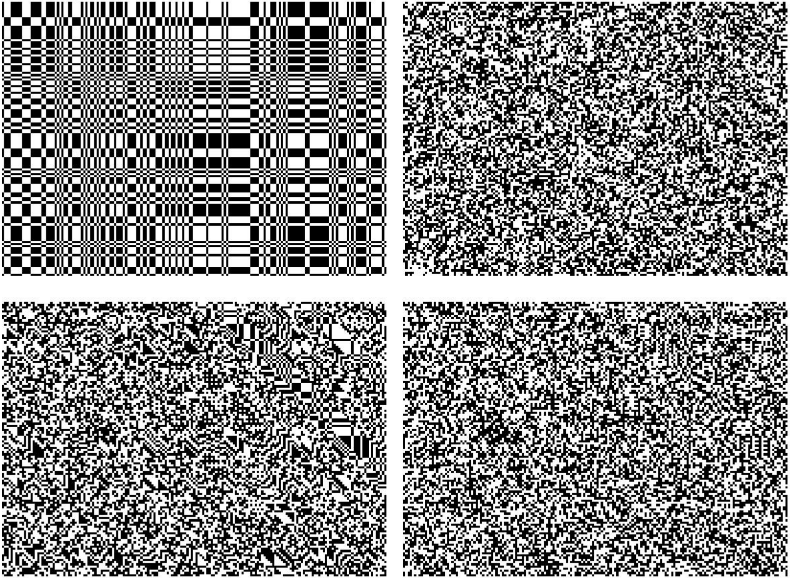
On K-wise Independent Distributions, Boolean Functions and Percolation

Ron Peled

joint work with

Itai Benjamini
and
Ori Gurel-Gurevich

Percolation pictures



K-wise independent events

- A vector $(X_1, \ldots, X_n)$ of random variables is called k-wise independent if each subset of k of the variables is independent.

K-wise independent events

- A vector $(X_1, \ldots, X_n)$ of random variables is called k-wise independent if each subset of k of the variables is independent.
- We consider the case when the variables are bits, with $\mathbb{P}(X_i = 1) = p$ for some $0 < p < 1$.

K-wise independent events

- A vector $(X_1, \ldots, X_n)$ of random variables is called k-wise independent if each subset of k of the variables is independent.
- We consider the case when the variables are bits, with $\mathbb{P}(X_i = 1) = p$ for some $0 < p < 1$.
- Define $\mathcal{A}(n, k, p)$ to be the set of all k-wise independent distributions $\mathbb{Q}$ on n bits with $\mathbb{Q}(X_i = 1) = p$ for all i.

K-wise independent events

- A vector $(X_1, \ldots, X_n)$ of random variables is called k-wise independent if each subset of k of the variables is independent.
- We consider the case when the variables are bits, with $\mathbb{P}(X_i = 1) = p$ for some $0 < p < 1$.
- Define $\mathcal{A}(n, k, p)$ to be the set of all k-wise independent distributions $\mathbb{Q}$ on n bits with $\mathbb{Q}(X_i = 1) = p$ for all i.
- In this work we try to understand for a given function $f : \{0,1\}^n \to \{0,1\}$ the quantities

$$\max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) \qquad \text{and} \qquad \min_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(f = 1)$$

CS background

- Concept of k-wise independent distributions is important in computer-science where it is used, for example, for derandomization.

CS background

- Concept of k-wise independent distributions is important in computer-science where it is used, for example, for derandomization.
- Initiated by works of Luby 85,
  Chor-Goldreich-Haståd-Friedman-Rudich-Smolensky 85,
  Karp-Wigderson 85, Alon-Babai-Itai 86
  and developed further by Schulman 92, Luby 93, Koller-Megiddo 93, Karloff-Mansour 94, Motwani-Naor-Naor 94 and others (see Luby-Wigderson 95 for a survey).

CS background

- Concept of k-wise independent distributions is important in computer-science where it is used, for example, for derandomization.
- Initiated by works of Luby 85, Chor-Goldreich-Haståd-Friedman-Rudich-Smolensky 85, Karp-Wigderson 85, Alon-Babai-Itai 86 and developed further by Schulman 92, Luby 93, Koller-Megiddo 93, Karloff-Mansour 94, Motwani-Naor-Naor 94 and others (see Luby-Wigderson 95 for a survey).
- Related concept of almost k-wise independence is also very important. It's use was pioneered for derandomization purposes by Naor-Naor 90 and developed further by Alon-Goldreich-Håstad-Peralta 90, Azar-Motwani-Naor 90, Alon-Bruck-Naor-Naor-Roth 92, Even-Goldreich-Luby-Nisan-Velićković 92, Chari-Rohathi-Srinivisan 94, Alon-Goldreich-Mansour 03 and many others.

CS background

- Concept of *k*-wise independent distributions is important in computer-science where it is used, for example, for derandomization.
- Initiated by works of Luby 85, Chor-Goldreich-Haståd-Friedman-Rudich-Smolensky 85, Karp-Wigderson 85, Alon-Babai-Itai 86 and developed further by Schulman 92, Luby 93, Koller-Megiddo 93, Karloff-Mansour 94, Motwani-Naor-Naor 94 and others (see Luby-Wigderson 95 for a survey).
- Related concept of almost *k*-wise independence is also very important. It's use was pioneered for derandomization purposes by Naor-Naor 90 and developed further by Alon-Goldreich-Håstad-Peralta 90, Azar-Motwani-Naor 90, Alon-Bruck-Naor-Naor-Roth 92, Even-Goldreich-Luby-Nisan-Velićković 92, Chari-Rohathi-Srinivisan 94, Alon-Goldreich-Mansour 03 and many others.
- In this work we will concentrate only on (perfectly) *k*-wise independent distributions. Analogous questions can be asked for the almost *k*-wise independent case but we do not address these here.

CS background contd.

- For derandomization purposes one usually checks that a certain randomized algorithm performs (about) the same on a particular k-wise independent input as in the completely independent case.

CS background contd.

- For derandomization purposes one usually checks that a certain randomized algorithm performs (about) the same on a particular k-wise independent input as in the completely independent case.
- Our questions are of a similar flavor, we ask, for a given boolean function f, how much independence is required for it to behave (about) the same on all k-wise independent inputs (including the completely independent one).

Convexity

- It is easy to see that for given n, k, p, the set $\mathcal{A}(n, k, p)$ is convex.

Convexity

- It is easy to see that for given n, k, p, the set $\mathcal{A}(n,k,p)$ is convex.
- Hence to understand

$$\max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) \qquad \text{and} \qquad \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1)$$

it is in principle enough to understand the extremal points of $\mathcal{A}(n,k,p)$. Unfortunately this appears to be very difficult and in the sequel we shall have to resort to ad-hoc methods for each function f we consider.

Convexity

- It is easy to see that for given n, k, p, the set $\mathcal{A}(n, k, p)$ is convex.
- Hence to understand

$$\max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) \qquad \text{and} \qquad \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f = 1)$$

  it is in principle enough to understand the extremal points of $\mathcal{A}(n, k, p)$. Unfortunately this appears to be very difficult and in the sequel we shall have to resort to ad-hoc methods for each function f we consider.
- For later reference, we identify the two extreme points of $\mathcal{A}(n, n-1, \frac{1}{2})$. XOR0 is the distribution on $(X_1, \ldots, X_n)$ having $\{X_i\}_{i=1}^{n-1}$ IID and $X_n \equiv \sum_{i=1}^{n-1} X_i$ mod 2, and XOR1 is the same with $X_n \equiv 1 + \sum_{i=1}^{n-1} X_i$ mod 2.

Parity

- As a first example consider the parity function $\text{Parity}_n : \{0,1\}^n \to \{0,1\}$ at $p = \frac{1}{2}$.
  Does it necessarily behave the same on a k-wise independent input as on a fully independent input?

Parity

- As a first example consider the parity function $\text{Parity}_n : \{0,1\}^n \to \{0,1\}$ at $p = \frac{1}{2}$.
  Does it necessarily behave the same on a k-wise independent input as on a fully independent input?
- No! In a very strong sense. For any $k < n$, under the XOR0 distribution the probability that parity returns 1 is 0 and under the XOR1 distribution the probability is 1.
- Hence to ensure that parity behaves normally one must take $k = n$!

Basic definitions

- Let $\varepsilon^f(k, p) := \max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) - \min_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(f = 1)$.

Basic definitions

- Let $\varepsilon^f(k,p) := \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) - \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1)$.
- Define the δ-independence sensitivity of the function f at p to be

$$K(f,\delta,p) = \min(k \mid \varepsilon^f(k,p) \le \delta)$$

or in words, how much independence of the input do we need to ensure that the probability that the function is 1 is the same as the fully independent case up to an additive error of (one half) δ.

Basic definitions

- Let $\varepsilon^f(k,p) := \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) - \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1)$.
- Define the δ-independence sensitivity of the function f at p to be

$$K(f,\delta,p) = \min(k \mid \varepsilon^f(k,p) \leq \delta)$$

  or in words, how much independence of the input do we need to ensure that the probability that the function is 1 is the same as the fully independent case up to an additive error of (one half) δ.
- For simplicity, we define arbitrarily $K(f,p) := K(f,0.01,p)$ and call this the independence sensitivity of the function f at p.

Basic definitions

- Let $\varepsilon^f(k,p) := \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) - \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1)$.
- Define the δ-independence sensitivity of the function f at p to be

$$K(f,\delta,p) = \min(k \mid \varepsilon^f(k,p) \leq \delta)$$

  or in words, how much independence of the input do we need to ensure that the probability that the function is 1 is the same as the fully independent case up to an additive error of (one half) δ.
- For simplicity, we define arbitrarily $K(f,p) := K(f,0.01,p)$ and call this the independence sensitivity of the function f at p.
- We will be mostly interested in monotone functions. We recall that a sequence $f_n : \{0,1\}^n \to \{0,1\}$ of monotone boolean functions has a sharp threshold at $p = p_c$ if ($\mathbb{P}_p$ is product distribution)

$$\lim_{n\to\infty} \mathbb{P}_p(f_n = 1) = \begin{cases} 0 & p < p_c \\ 1 & p > p_c \end{cases}$$

Majority

- A first non-trivial example is the majority function $\text{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.

Majority

- A first non-trivial example is the majority function $\text{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.
- Maj_n has a sharp threshold at $p_c = \frac{1}{2}$.

Majority

- A first non-trivial example is the majority function $\text{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.
- Maj_n has a sharp threshold at $p_c = \frac{1}{2}$.
- It is easy to see that asymptotically in n, $K(\text{Maj}_n, p) = 2$ for any $p \neq \frac{1}{2}$ $(0.01 \leq p \leq 0.99)$. Since,

Majority

- A first non-trivial example is the majority function $\text{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.
- Maj_n has a sharp threshold at $p_c = \frac{1}{2}$.
- It is easy to see that asymptotically in n, $K(\text{Maj}_n, p) = 2$ for any $p \neq \frac{1}{2}$ $(0.01 \leq p \leq 0.99)$. Since, clearly, $k = 1$ is not sufficient. However,

Majority

- A first non-trivial example is the majority function $\mathrm{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.
- Maj_n has a sharp threshold at $p_c = \frac{1}{2}$.
- It is easy to see that asymptotically in n, $K(\mathrm{Maj}_n, p) = 2$ for any $p \neq \frac{1}{2}$ $(0.01 \leq p \leq 0.99)$. Since, clearly, $k = 1$ is not sufficient. However, for $\mathbb{Q} \in \mathcal{A}(n, 2, p)$, let S_n count the number of 1's under $\mathbb{Q}$. We have $\mathbb{E}_{\mathbb{Q}}(S_n) = np$ and $\mathrm{Var}_{\mathbb{Q}}(S_n) = np(1-p)$. If, WLOG, $p < 1/2$ then by Chebyshev's inequality

Majority

- A first non-trivial example is the majority function $\text{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.
- Maj_n has a sharp threshold at $p_c = \frac{1}{2}$.
- It is easy to see that asymptotically in n, $K(\text{Maj}_n, p) = 2$ for any $p \neq \frac{1}{2}$ $(0.01 \leq p \leq 0.99)$. Since, clearly, $k = 1$ is not sufficient. However, for $\mathbb{Q} \in \mathcal{A}(n, 2, p)$, let S_n count the number of 1's under $\mathbb{Q}$. We have $\mathbb{E}_{\mathbb{Q}}(S_n) = np$ and $\text{Var}_{\mathbb{Q}}(S_n) = np(1-p)$. If, WLOG, $p < 1/2$ then by Chebyshev's inequality

$$\mathbb{Q}(S_n > n/2) \leq \mathbb{Q}((S_n - np) > n(1/2 - p)) \leq$$
$$\leq \frac{np(1-p)}{(n(1/2-p))^2} = O(\frac{1}{n}) \to 0$$

Majority

- A first non-trivial example is the majority function $\text{Maj}_n : \{0,1\}^n \to \{0,1\}$ defined for odd n.
- Maj_n has a sharp threshold at $p_c = \frac{1}{2}$.
- It is easy to see that asymptotically in n, $K(\text{Maj}_n, p) = 2$ for any $p \neq \frac{1}{2}$ $(0.01 \leq p \leq 0.99)$. Since, clearly, $k = 1$ is not sufficient. However, for $\mathbb{Q} \in \mathcal{A}(n, 2, p)$, let S_n count the number of 1's under $\mathbb{Q}$. We have $\mathbb{E}_{\mathbb{Q}}(S_n) = np$ and $\text{Var}_{\mathbb{Q}}(S_n) = np(1-p)$. If, WLOG, $p < 1/2$ then by Chebyshev's inequality

$$\mathbb{Q}(S_n > n/2) \leq \mathbb{Q}((S_n - np) > n(1/2 - p)) \leq$$
$$\leq \frac{np(1-p)}{(n(1/2-p))^2} = O(\frac{1}{n}) \to 0$$

- What about $p = \frac{1}{2}$? (can voters bias an election by using a voting scheme which is close to fully independent? can they do it if the scheme is only a little independent?)

Majority contd.

- The following simple argument shows $K(\mathrm{Maj}_n, \frac{1}{2}) \leq \omega(1)$.

Majority contd.

- The following simple argument shows $K(\text{Maj}_n, \frac{1}{2}) \leq \omega(1)$.
- Consider the distribution of S_n under some $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$. Let also $\overline{S_n} = (S_n - n/2)/\sqrt{n/4}$.

Majority contd.

- The following simple argument shows $K(\mathrm{Maj}_n, \frac{1}{2}) \leq \omega(1)$.
- Consider the distribution of S_n under some $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$. Let also $\overline{S_n} = (S_n - n/2)/\sqrt{n/4}$.
  Obviously, $E_{\mathbb{Q}}(S_n^l) = E_{\mathbb{P}_{1/2}}(S_n^l)$ for any $l \leq k$. The same holds for $\overline{S_n}$ as it is a linear function of S_n.

Majority contd.

- The following simple argument shows $K(\mathrm{Maj}_n, \frac{1}{2}) \leq \omega(1)$.
- Consider the distribution of S_n under some $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$. Let also $\overline{S_n} = (S_n - n/2)/\sqrt{n/4}$.
  Obviously, $E_{\mathbb{Q}}(S_n^l) = E_{\mathbb{P}_{1/2}}(S_n^l)$ for any $l \leq k$. The same holds for $\overline{S_n}$ as it is a linear function of S_n.
  Therefore, $E_{\mathbb{Q}_n}(\overline{S_n}^l) \to s_l$ where $s_l = \mathbb{E}(N(0,1)^l)$ is the l-th moment of a standard normal distribution.

Majority contd.

- The following simple argument shows $K(\text{Maj}_n, \frac{1}{2}) \leq \omega(1)$.
- Consider the distribution of S_n under some $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$. Let also $\overline{S_n} = (S_n - n/2)/\sqrt{n/4}$.
  Obviously, $E_{\mathbb{Q}}(S_n^l) = E_{\mathbb{P}_{1/2}}(S_n^l)$ for any $l \leq k$. The same holds for $\overline{S_n}$ as it is a linear function of S_n.
  Therefore, $E_{\mathbb{Q}_n}(\overline{S_n}^l) \to s_l$ where $s_l = \mathbb{E}(N(0,1)^l)$ is the l-th moment of a standard normal distribution.
  The normal distribution is determined by its moments. Hence, if $k(n) \in \omega(1)$ and $\mathbb{Q}_n \in \mathcal{A}(n, k(n), 1/2)$ then $\overline{S_n} \to N(0,1)$ weakly. In particular, $\mathbb{Q}_n(\text{Maj}_n = 1) = \mathbb{Q}_n(\overline{S_n} > 0) \to 1/2$.

Majority contd. II

- In fact, for Majority we know more precise results.

Majority contd. II

- In fact, for Majority we know more precise results.

- Theorem

  There exists a $C > 0$ *such that for any even* $2 \leq k < n$

$$\frac{C}{\sqrt{k \log k}} \leq \max_{\mathbb{Q} \in \mathcal{A}(n,k,\frac{1}{2})} \mathbb{Q}(\mathrm{Maj}_n = 1) - \frac{1}{2} \leq \frac{2\sqrt{2}}{\sqrt{k}}$$

  And when $\mathbb{Q}_0 \in \mathcal{A}(n, n-1, \frac{1}{2})$ *is the XOR0 distribution we have* $|\mathbb{Q}_0(\mathrm{Maj}_n = 1) - \frac{1}{2}| \geq \frac{1}{3\sqrt{n}}$.

Majority contd. II

- In fact, for Majority we know more precise results.

- Theorem

There exists a $C > 0$ such that for any even $2 \leq k < n$

$$\frac{C}{\sqrt{k \log k}} \leq \max_{\mathbb{Q} \in \mathcal{A}(n,k,\frac{1}{2})} \mathbb{Q}(\mathrm{Maj}_n = 1) - \frac{1}{2} \leq \frac{2\sqrt{2}}{\sqrt{k}}$$

And when $\mathbb{Q}_0 \in \mathcal{A}(n, n-1, \frac{1}{2})$ is the XOR0 distribution we have $|\mathbb{Q}_0(\mathrm{Maj}_n = 1) - \frac{1}{2}| \geq \frac{1}{3\sqrt{n}}$.

- The theorem shows that $K(\mathrm{Maj}_n, \frac{1}{2})$ is constant for all n (which tends to infinity with the arbitrary threshold 0.01). Voters cannot significantly bias the election even when only finite independence is required.

Majority contd. II

- In fact, for Majority we know more precise results.

- Theorem

  There exists a $C > 0$ *such that for any even* $2 \leq k < n$

$$\frac{C}{\sqrt{k \log k}} \leq \max_{\mathbb{Q} \in \mathcal{A}(n,k,\frac{1}{2})} \mathbb{Q}(\mathrm{Maj}_n = 1) - \frac{1}{2} \leq \frac{2\sqrt{2}}{\sqrt{k}}$$

And when $\mathbb{Q}_0 \in \mathcal{A}(n, n-1, \frac{1}{2})$ *is the XOR0 distribution we have* $|\mathbb{Q}_0(\mathrm{Maj}_n = 1) - \frac{1}{2}| \geq \frac{1}{3\sqrt{n}}$.

  - The theorem shows that $K(\mathrm{Maj}_n, \frac{1}{2})$ is constant for all n (which tends to infinity with the arbitrary threshold 0.01). Voters cannot significantly bias the election even when only finite independence is required.
  - Upper bound (with worse constant) was known in coding theory (Sidel'nikov's theorem, see Macwilliams and Sloane). But our proof seems much simpler, it uses the theory of the classical moment problem (Akhiezer 65, Kreĭn-Nudel'man 77).

Noise sensitivity

- Recall the notion of noise sensitivity of a sequence f_n of boolean functions.

Noise sensitivity

- Recall the notion of noise sensitivity of a sequence f_n of boolean functions.
  The sequence is called *noise sensitive* if when you sample it on a uniform input ($p = \frac{1}{2}$) and when you sample it on the same input where you flip each bit with probability ε then the outputs are asymptotically independent (this does not depend on ε).

Noise sensitivity

- Recall the notion of noise sensitivity of a sequence f_n of boolean functions.
  The sequence is called *noise sensitive* if when you sample it on a uniform input ($p = \frac{1}{2}$) and when you sample it on the same input where you flip each bit with probability ε then the outputs are asymptotically independent (this does not depend on ε).
  If the outputs are asymptotically the same the sequence is called *noise stable*.

Noise sensitivity

- Recall the notion of noise sensitivity of a sequence f_n of boolean functions.
  The sequence is called *noise sensitive* if when you sample it on a uniform input ($p = \frac{1}{2}$) and when you sample it on the same input where you flip each bit with probability ε then the outputs are asymptotically independent (this does not depend on ε).
  If the outputs are asymptotically the same the sequence is called *noise stable*.
- It is well known that a sequence is noise sensitive if for any k, the fraction of its fourier mass at frequencies of weight less than k tends to 0.
  It is noise stable if for any ε there exists k such that the fraction of the Fourier mass below weight k is at least $1 - \varepsilon$.

Noise sensitivity

- Recall the notion of noise sensitivity of a sequence f_n of boolean functions.
  The sequence is called *noise sensitive* if when you sample it on a uniform input ($p = \frac{1}{2}$) and when you sample it on the same input where you flip each bit with probability ε then the outputs are asymptotically independent (this does not depend on ε).
  If the outputs are asymptotically the same the sequence is called *noise stable*.
- It is well known that a sequence is noise sensitive if for any k, the fraction of its fourier mass at frequencies of weight less than k tends to 0.
  It is noise stable if for any ε there exists k such that the fraction of the Fourier mass below weight k is at least $1 - \varepsilon$.
- One may also define a quantitative version by setting $K_{\mathrm{NS}}(f)$ to be the minimal weight such that the fraction of Fourier mass above it is less than 0.01, say.

Iterated majority

- Since for parity $K_{\text{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.

Iterated majority

- Since for parity $K_{\mathrm{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!

Iterated majority

- Since for parity $K_{\text{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!
- Consider (n^a, n^{1-a}) iterated majority of height 2. That is, group the bits to groups of size n^a and perform majority on each group, then take the majority of the results. Call this function Maj_a^2.

Iterated majority

- Since for parity $K_{\mathrm{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!
- Consider (n^a, n^{1-a}) iterated majority of height 2. That is, group the bits to groups of size n^a and perform majority on each group, then take the majority of the results. Call this function Maj_a^2.
- This function is noise stable for any $0 < a < 1$.

Iterated majority

- Since for parity $K_{\text{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!
- Consider (n^a, n^{1-a}) iterated majority of height 2. That is, group the bits to groups of size n^a and perform majority on each group, then take the majority of the results. Call this function Maj_a^2.
- This function is noise stable for any $0 < a < 1$. But we show $K(\text{Maj}_a^2, \frac{1}{2}) \sim n^{\min(a,1-a)}$.

Iterated majority

- Since for parity $K_{\text{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!
- Consider (n^a, n^{1-a}) iterated majority of height 2. That is, group the bits to groups of size n^a and perform majority on each group, then take the majority of the results. Call this function Maj_a^2.
- This function is noise stable for any $0 < a < 1$. But we show $K(\text{Maj}_a^2, \frac{1}{2}) \sim n^{\min(a,1-a)}$.
- Proof uses a theorem about independence sensitivity of compositions of Majority with other functions.

Iterated majority

- Since for parity $K_{\text{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!
- Consider (n^a, n^{1-a}) iterated majority of height 2. That is, group the bits to groups of size n^a and perform majority on each group, then take the majority of the results. Call this function Maj_a^2.
- This function is noise stable for any $0 < a < 1$. But we show $K(\text{Maj}_a^2, \frac{1}{2}) \sim n^{\min(a,1-a)}$.
- Proof uses a theorem about independence sensitivity of compositions of Majority with other functions.
- It utilizes the duality of the problem to the problem of approximating the function by real polynomials.

Iterated majority

- Since for parity $K_{\text{NS}} = n$ and for Majority K_{NS} is constant, the previous examples may suggest that noise sensitivity and independence sensitivity (at $p = \frac{1}{2}$) are very similar, perhaps even the same concept up to constants.
- However this is not the case!
- Consider (n^a, n^{1-a}) iterated majority of height 2. That is, group the bits to groups of size n^a and perform majority on each group, then take the majority of the results. Call this function Maj_a^2.
- This function is noise stable for any $0 < a < 1$. But we show $K(\text{Maj}_a^2, \frac{1}{2}) \sim n^{\min(a,1-a)}$.
- Proof uses a theorem about independence sensitivity of compositions of Majority with other functions.
- It utilizes the duality of the problem to the problem of approximating the function by real polynomials.
- One of our main open questions is whether we can have $K_{\text{NS}}(f_n) = \omega(K(f_n, \frac{1}{2}))$.

Duality - approximation by polynomials

- For a given function $f : \{0,1\}^n \to \{0,1\}$, let $P_k^+(f)$ be all real polynomials $P : \mathbb{R}^n \to \mathbb{R}$ of degree at most k satisfying $P(x) \geq f(x)$ on $\{0,1\}^n$.

Duality - approximation by polynomials

- For a given function $f : \{0,1\}^n \to \{0,1\}$, let $P_k^+(f)$ be all real polynomials $P : \mathbb{R}^n \to \mathbb{R}$ of degree at most k satisfying $P(x) \geq f(x)$ on $\{0,1\}^n$.
- Similarly define $P_k^-(f)$ with $P(x) \leq f(x)$ on $\{0,1\}^n$.

Duality - approximation by polynomials

- For a given function $f : \{0,1\}^n \to \{0,1\}$, let $P_k^+(f)$ be all real polynomials $P : \mathbb{R}^n \to \mathbb{R}$ of degree at most k satisfying $P(x) \geq f(x)$ on $\{0,1\}^n$.
- Similarly define $P_k^-(f)$ with $P(x) \leq f(x)$ on $\{0,1\}^n$.
- By linear programming duality

$$\max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) = \min_{P\in P_k^+(f)} \mathbb{E}_{\mathbb{P}_p} P(X_1,\ldots,X_n)$$

$$\min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) = \max_{P\in P_k^-(f)} \mathbb{E}_{\mathbb{P}_p} P(X_1,\ldots,X_n)$$

Duality - approximation by polynomials

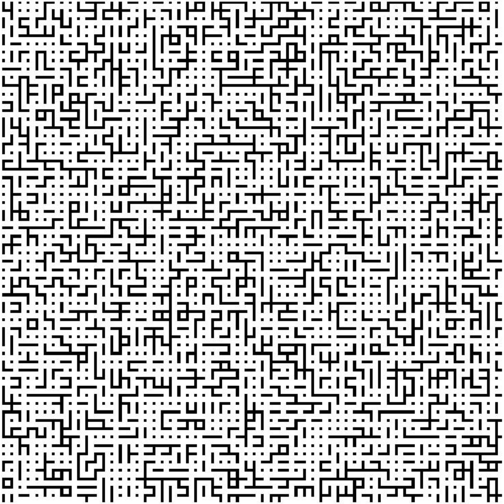
- For a given function $f : \{0,1\}^n \to \{0,1\}$, let $P_k^+(f)$ be all real polynomials $P : \mathbb{R}^n \to \mathbb{R}$ of degree at most k satisfying $P(x) \geq f(x)$ on $\{0,1\}^n$.
- Similarly define $P_k^-(f)$ with $P(x) \leq f(x)$ on $\{0,1\}^n$.
- By linear programming duality

$$\max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) = \min_{P\in P_k^+(f)} \mathbb{E}_{\mathbb{P}_p} P(X_1,\ldots,X_n)$$

$$\min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) = \max_{P\in P_k^-(f)} \mathbb{E}_{\mathbb{P}_p} P(X_1,\ldots,X_n)$$

- Hence for f to behave the same under all k-wise independent distributions is equivalent to f having a "sandwich L_1" approximation by real polynomials of degree k.

Percolation with $p = \frac{1}{3}$



Percolation crossing

- Our main result is about the percolation crossing function. We consider percolation in a finite box in d dimensions ($d \geq 2$) with side length n. Consider the function f which says if there is a crossing from left to right. Recall that f has a sharp threshold at $0 < p_c < 1$.

Percolation crossing

- Our main result is about the percolation crossing function. We consider percolation in a finite box in d dimensions ($d \geq 2$) with side length n. Consider the function f which says if there is a crossing from left to right. Recall that f has a sharp threshold at $0 < p_c < 1$.
- What is $K(f, p)$? for example, how much independence is needed to have that for any $p > p_c$ the probability of crossing tends to 1 (with n) and for any $p < p_c$ the probability of crossing tends to 0?
  Is it possible that 1% of the edges are present and any 100 are independent, yet there is a crossing with high probability?

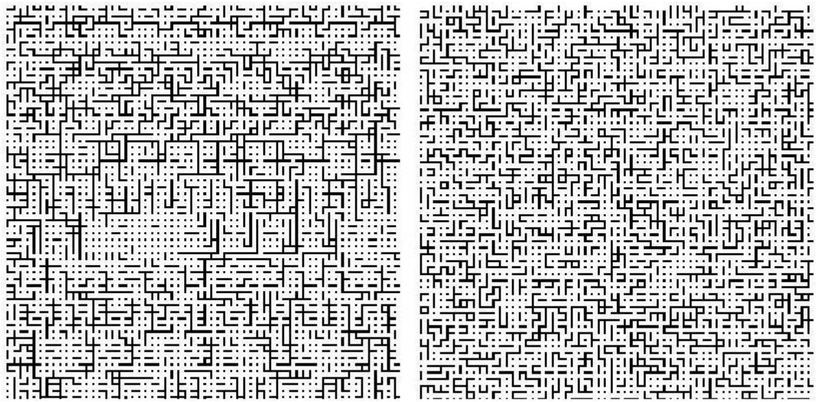
Percolation crossing

- Our main result is about the percolation crossing function. We consider percolation in a finite box in d dimensions ($d \geq 2$) with side length n. Consider the function f which says if there is a crossing from left to right. Recall that f has a sharp threshold at $0 < p_c < 1$.
- What is $K(f, p)$? for example, how much independence is needed to have that for any $p > p_c$ the probability of crossing tends to 1 (with n) and for any $p < p_c$ the probability of crossing tends to 0?
  Is it possible that 1% of the edges are present and any 100 are independent, yet there is a crossing with high probability?
- We find that $(\log n)^{c_p/\sqrt{\log\log n}} \leq K(f, p) \leq C_p \log n$ asymptotically for $p \neq p_c$.

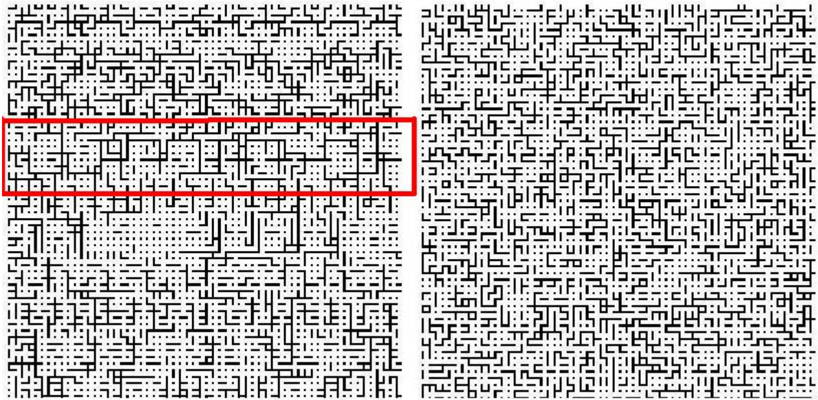
Percolation crossing

- Our main result is about the percolation crossing function. We consider percolation in a finite box in d dimensions ($d \geq 2$) with side length n. Consider the function f which says if there is a crossing from left to right. Recall that f has a sharp threshold at $0 < p_c < 1$.
- What is $K(f, p)$? for example, how much independence is needed to have that for any $p > p_c$ the probability of crossing tends to 1 (with n) and for any $p < p_c$ the probability of crossing tends to 0? Is it possible that 1% of the edges are present and any 100 are independent, yet there is a crossing with high probability?
- We find that $(\log n)^{c_p/\sqrt{\log \log n}} \leq K(f, p) \leq C_p \log n$ asymptotically for $p \neq p_c$.
- When $d > 2$ we only know the upper bound for $p < p_c$.
- A similar result holds on the d-ary tree.
- This answers a question of Benjamini, Kozma and Romik.

2-wise percolation at $p = \frac{1}{3}$



2-wise percolation at $p = \frac{1}{3}$ with box



Orthogonal arrays

- In applications in CS, one usually generates a k-wise independent distribution with $p = \frac{1}{2}$ by finding a space $\Omega \subseteq \{0,1\}^n$ such that sampling a uniform string from Ω one obtains a k-wise independent distribution.

Orthogonal arrays

- In applications in CS, one usually generates a k-wise independent distribution with $p = \frac{1}{2}$ by finding a space $\Omega \subseteq \{0,1\}^n$ such that sampling a uniform string from Ω one obtains a k-wise independent distribution.
- An important question one often has to deal with is how small can Ω be taken to be? (this is highly related to the question of how large can a linear error correcting code with given distance be).

Orthogonal arrays

- In applications in CS, one usually generates a k-wise independent distribution with $p = \frac{1}{2}$ by finding a space $\Omega \subseteq \{0,1\}^n$ such that sampling a uniform string from Ω one obtains a k-wise independent distribution.
- An important question one often has to deal with is how small can Ω be taken to be? (this is highly related to the question of how large can a linear error correcting code with given distance be).
- What is the analogue of this question for general k-wise independent distributions?

Orthogonal arrays

- In applications in CS, one usually generates a k-wise independent distribution with $p = \frac{1}{2}$ by finding a space $\Omega \subseteq \{0,1\}^n$ such that sampling a uniform string from Ω one obtains a k-wise independent distribution.
- An important question one often has to deal with is how small can Ω be taken to be? (this is highly related to the question of how large can a linear error correcting code with given distance be).
- What is the analogue of this question for general k-wise independent distributions?
- Denote $L := |\Omega|$. If we suppose WLOG that Ω contains the all zeros string we have that the probability to sample this string is $\frac{1}{L}$.

Orthogonal arrays

- In applications in CS, one usually generates a k-wise independent distribution with $p = \frac{1}{2}$ by finding a space $\Omega \subseteq \{0,1\}^n$ such that sampling a uniform string from Ω one obtains a k-wise independent distribution.
- An important question one often has to deal with is how small can Ω be taken to be? (this is highly related to the question of how large can a linear error correcting code with given distance be).
- What is the analogue of this question for general k-wise independent distributions?
- Denote $L := |\Omega|$. If we suppose WLOG that Ω contains the all zeros string we have that the probability to sample this string is $\frac{1}{L}$.
- Hence one analogue of the question in our case is: How high can the probability of the all zeros string be?

Orthogonal arrays

- In applications in CS, one usually generates a k-wise independent distribution with $p = \frac{1}{2}$ by finding a space $\Omega \subseteq \{0,1\}^n$ such that sampling a uniform string from Ω one obtains a k-wise independent distribution.
- An important question one often has to deal with is how small can Ω be taken to be? (this is highly related to the question of how large can a linear error correcting code with given distance be).
- What is the analogue of this question for general k-wise independent distributions?
- Denote $L := |\Omega|$. If we suppose WLOG that Ω contains the all zeros string we have that the probability to sample this string is $\frac{1}{L}$.
- Hence one analogue of the question in our case is: How high can the probability of the all zeros string be?
- An upper bound on this quantity implies a lower bound on L, but it is more general since it also implies a bound on the atom at the all zeros string for any k-wise independent distribution.

Probability of all bits 1

- Another of our main results considers the following quantity

$$M(n, k, p) := \max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$$

Probability of all bits 1

- Another of our main results considers the following quantity

$$M(n,k,p) := \max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$$

- Using the theory of the classical moment problem we show for even k

$$M(n,k,p) \le \frac{p^n}{\mathbb{P}(\text{Bin}(n, 1-p) \le \frac{k}{2})}$$

Probability of all bits 1

- Another of our main results considers the following quantity

$$M(n, k, p) := \max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$$

- Using the theory of the classical moment problem we show for even k

$$M(n, k, p) \le \frac{p^n}{\mathbb{P}(\text{Bin}(n, 1-p) \le \frac{k}{2})}$$

- Bound believed to be sharp in all ranges of the parameters.

Probability of all bits 1

- Another of our main results considers the following quantity

$$M(n,k,p) := \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$$

- Using the theory of the classical moment problem we show for even k

$$M(n,k,p) \le \frac{p^n}{\mathbb{P}(\text{Bin}(n,1-p) \le \frac{k}{2})}$$

- Bound believed to be sharp in all ranges of the parameters.For example, gives for k even

$$M(n,k,p) \le 2\sqrt{k}\left(\frac{kp}{2e(1-p)(n-\frac{k}{2})}\right)^{\frac{k}{2}} \le \left(\frac{Ckp}{(1-p)n}\right)^{\frac{k}{2}}$$

$$M(n,k,p) \le 2p^n \qquad \text{When } n(1-p) \le \frac{k}{2}$$

(second corollary uses a result of Jogdeo-Samuels 68).

Probability of all bits 1 contd.

- For $p = \frac{1}{2}$ the bound is a generalization of the bound of Alon-Babai-Itai 86 for the minimal size of an orthogonal array of bits. In this case we get an upper bound on the size of any atom of the distribution (by xoring a constant string).

Probability of all bits 1 contd.

- For $p = \frac{1}{2}$ the bound is a generalization of the bound of Alon-Babai-Itai 86 for the minimal size of an orthogonal array of bits. In this case we get an upper bound on the size of any atom of the distribution (by xoring a constant string).
- For $p = \frac{1}{k}$ for an integer k, the bound is a generalization of the Rao bound 47 for the minimal size of an orthogonal array over $\mathbb{Z}_k$.

Proof of percolation theorem

- Our main lemma for the percolation result is inspired by the $(u \mid u+v)$ lemma of error-correcting codes (Macwilliams-Sloane 77). It allows to "amplify" independence.

Proof of percolation theorem

- Our main lemma for the percolation result is inspired by the $(u \mid u+v)$ lemma of error-correcting codes (Macwilliams-Sloane 77). It allows to "amplify" independence.

- Lemma

Fix $m \geq 1$. Let $X := (X_1, \ldots, X_n) \in \mathcal{A}^r(n,k)$. Let $X^i := (X^i_j)_{j=1}^n$ be m IID copies of X. Let also $Y := (Y_1, \ldots, Y_n) \in \mathcal{A}^r(n, 2k+1)$ be a vector independent of all the X's. Then the vector with the following coordinates

$$\begin{array}{cccc} X_1^1 + Y_1, & X_2^1 + Y_2, & \ldots, & X_n^1 + Y_n, \\ X_1^2 + Y_1, & X_2^2 + Y_2, & \ldots, & X_n^2 + Y_n, \\ \vdots, & \vdots, & \vdots, & \vdots, \\ X_1^m + Y_1, & X_2^m + Y_2, & \ldots, & X_n^m + Y_n \end{array} \tag{1}$$

is in $\mathcal{A}^r(mn, 2k+1)$

Proof of percolation theorem

- Our main lemma for the percolation result is inspired by the $(u \mid u+v)$ lemma of error-correcting codes (Macwilliams-Sloane 77). It allows to "amplify" independence.

- Lemma

  Fix $m \geq 1$. Let $X := (X_1, \ldots, X_n) \in \mathcal{A}^r(n,k)$. Let $X^i := (X^i_j)^n_{j=1}$ be m IID copies of X. Let also $Y := (Y_1, \ldots, Y_n) \in \mathcal{A}^r(n, 2k+1)$ be a vector independent of all the X's. Then the vector with the following coordinates

$$\begin{array}{cccc} X^1_1 + Y_1, & X^1_2 + Y_2, & \ldots, & X^1_n + Y_n, \\ X^2_1 + Y_1, & X^2_2 + Y_2, & \ldots, & X^2_n + Y_n, \\ \vdots, & \vdots, & \vdots, & \vdots, \\ X^m_1 + Y_1, & X^m_2 + Y_2, & \ldots, & X^m_n + Y_n \end{array} \qquad (1)$$

  is in $\mathcal{A}^r(mn, 2k+1)$

  - We also have generalizations of this lemma which we do not present here.

Some open questions

Open questions

- Say anything non-trivial about the extremal points of $\mathcal{A}(n, k, p)$.
- What is K for percolation crossing in the plane at $p = p_c = \frac{1}{2}$?
- What is K at $p = \frac{1}{2}$ for iterated majority of height 3, for recursive majority of 3's?
- Can we have a boolean function whose Fourier spectrum is concentrated on high levels, but its K at $p = \frac{1}{2}$ is small? i.e. that $K_{\text{NS}}(f_n) = \omega(K(f_n, \frac{1}{2}))$.
- Conjecture of Linial-Nisan 90, about the independence sensitivity at $p = \frac{1}{2}$ for AC0 circuits.