Sums of independent variables approximating a boolean function *†

Jakub Onufry Wojtaszczyk

13.IX.2010

Abstract

Friedgut, Kalai and Naor have shown that if the absolute value of a sum of weighted Rademacher variables has small variance, then one of the variables has to dominate the sum. We provide a simpler proof of this result, and generalize it to symmetric random variables with a moment comparison condition.

1 Introduction and notation

Consider a family of independent symmetric random variables $X_i$, with $\sum \text{Var} X_i = 1$. We are interested in learning how close their sum can approximate a boolean function (by which we shall mean a function into $\{-1, 1\}$). We shall prove that for this approximation to be good, one of the variables has to dominate the sum as follows:

Theorem 1.1. If $(X_i)_{i=1}^N$ is a sequence of independent, symmetric random variables with $\mathbb{E}X_i^4 \leq \tau(\mathbb{E}X_i^2)^2$ for some universal constant $\tau$ and $\text{Var} \left| \sum X_i \right| = \delta$, then for some $i$ and we have $\text{Var} X_i > 1 - K(\tau) \delta$.

This result for Rademacher variables was shown in [4] by E. Friedgut et al., and was a part of the proof of their theorem on Boolean functions on the discrete cube with Fourier coefficients concentrated at the first two levels. They gave two proofs of the result concerning Rademacher variables. One was a direct application of a theorem of König et al. ([5]). The other used a more elementary approach (using Chernoff’s inequality), but contained an omission — it worked only under the assumption that we already know

*Keywords: Rademacher variables, absolute value variation, Fourier coefficients, Irit Dinur PCP proof
†2010 Mathematical Subject Classification: 60E15, 42C10
that $\text{Var} X_i > C$ for some $C$ close to 1. This can be fixed (one of the authors suggested in private communication using an argument based on the Berry–Esseen theorem, which indeed can be done).

The Fourier coefficient theorem (which was the main goal of [4] and will be stated in Section 3) is a direct application of the above. It was proved with application in discrete combinatorics and social science in mind, but turned out to be useful also in computer science. In particular the theorem is used in analyzing a Long Code Test in the celebrated expander proof of the PCP theorem by Irit Dinur ([3]). With that in mind it is unfortunate that the proof of this fact is not self-contained — anyone looking for a complete understanding of the PCP theorem would have to either grasp the intricacies of the results of [5] (or [2] and [1], as a proof using the Bonami–Beckner inequality is also given), or find their own way to close the gap in the Chernoff argument.

With that in mind, I decided to give an elementary, self-contained and simple proof of Theorem 1.1.

I think it is also interesting (although not very surprising) that the inequality still holds if we replace the Rademacher variables by symmetric variables with a second–fourth moment comparison condition. Note that we do not require the random variables to be identically distributed, in particular the result still holds if we take a Rademacher (or a slightly perturbed Rademacher) to be the “big” variable, while trying to fill in the gaps with other variables.

My contribution: I would like to emphasize that the Theorem 3.1 and Theorem 1.1 for Rademacher variables (which are the case with the most interesting applications) were already proved in [4]. This paper generalizes theorem 1.1 to the case of symmetric random variables with a moment condition. However I perceive the main value of this paper to lie in providing a simple and self-contained proof of both these facts.

2 Proof of Theorem 1.1

We begin by proving two Lemmata, after which the proof of the Theorem will be mainly cleaning up.

Lemma 2.1. Let $X$, $Y$ be symmetric, independent variables. Then for any $\beta, \gamma \in \mathbb{R}$ we have

$$\mathbb{P}\left(\left|X + Y - \beta\right| \geq \gamma\right) \geq \frac{1}{2}\mathbb{P}(\left|X\right| \geq \gamma)\mathbb{P}(\left|Y\right| \geq \gamma).$$
Proof. Let $\bar{X}$ be $X$ conditioned upon $|X| \geq \gamma$ and $\bar{Y}$ by $Y$ conditioned upon $|Y| \geq \gamma$ (if any of these events have probability zero, the thesis is trivial). Then

$$P\left(\left|X + Y - \beta\right| \geq \gamma\right) \geq P(|X| \geq \gamma)P(|Y| \geq \gamma)P\left(\left|\bar{X} + \bar{Y} - \beta\right| \geq \gamma\right).$$

Let $U = \max\{|\bar{X}|, |\bar{Y}|\}$ and $V = \min\{|\bar{X}|, |\bar{Y}|\}$, and let $\varepsilon$ be an independent Rademacher variable. Then $U + \varepsilon V$ has the same distribution as $|X + Y|$. Notice, however, that as $V \geq \gamma$, the events $\{|U - \beta + V| < \gamma\}$ and $\{|U - \beta - V| < \gamma\}$ are disjoint. Thus

$$P(|U + \varepsilon V - \beta| \geq \gamma) \geq \frac{1}{2},$$

which proves the thesis. \hfill \Box

Lemma 2.2. Let $X, Y$ be symmetric, independent variables, and assume $EX^4 \leq \tau(EX^2)^2$, $EY^4 \leq \tau(EY^2)^2$. Let $\alpha = \min\{\text{Var } X, \text{Var } Y\}$. Then

$$\text{Var } |X + Y| \geq C(\tau)\alpha.$$

Proof. Let $\gamma$ be any positive constant. Let $\beta = E|X + Y|$. Then

$$\text{Var } |X + Y| = E(|X + Y| - \beta)^2 \geq \frac{1}{2}\gamma^2 P(|X| \geq \gamma)P(|Y| \geq \gamma)$$

by Lemma 2.1. Now we use a Paley–Zygmund type inequality to estimate $P(|X| \geq \gamma)$:

$$P(|X| \geq \gamma) = \frac{E(X^2)^2}{EX^4}E1_{\{|X| \geq \gamma\}}^2 \geq \frac{(EX^21_{\{|X| \geq \gamma\}})^2}{EX^4}$$

by the Cauchy-Schwarz inequality. Now put $\gamma^2 = EX^2/2$. Then $E(X^21_{|X| \leq \gamma} \leq \gamma^2$, and so $EX^21_{|X| \geq \gamma} \geq EX^2/2$. At the same time $EX^4 \leq \tau(EX^2)^2$, so we get

$$P(|X| > \alpha/2) \geq P(|X| \geq \gamma) \geq \frac{(EX^2)^2/4}{\tau(EX^2)^2} = 1/4\tau,$$

which gives

$$\text{Var } |X + Y| \geq \frac{\alpha}{128\tau^2}.$$

\hfill \Box

Theorem 2.3 (Theorem 1.1). If $(X_i)_{i=1}^N$ is a sequence of independent, symmetric random variables with $EX_i^4 \leq \tau(EX_i^2)^2$ for some universal constant $\tau$ and $\text{Var } \sum X_i = \delta$, then for some $i$ and we have $\text{Var } X_i > 1 - K(\tau)\delta$.\hfill 3
Proof of Theorem 1.1. If $\max \text{Var } X_i < 1/3$, we can find such a subset $A$ of indices that $1/3 < \sum_{i \in A} \text{Var } X_i < 2/3$. Indeed, it is enough to start with an empty set, and then add indices (in any order) until the sum of variances exceeds $1/3$. Due to the bound on the maximum variance, the sum of the variances at this moment will be smaller than $2/3$.

Let $X = \sum_{i \in A} X_i$ and $Y = \sum_{i \notin A} X_i$. By Lemma 2.2 we have $\delta = \text{Var } |X + Y| \geq C(\tau)/3$. By choosing $K(\tau) = 3/C(\tau)$ we can fulfill the thesis in this case.

Otherwise we have $\text{Var } X_{i_0} > 1/3$ for some $i_0$. Let $m = \text{Var } X_{i_0}$. Take $Y = \sum_{i \neq i_0} X_i$. As $\frac{1}{2} \text{Var } Y \leq \text{Var } X_{i_0}$, Lemma 2.2 gives us

$$\delta = \text{Var } |X_{i_0} + Y| \geq \frac{C(\tau) \text{Var } Y}{2} = C(\tau)(1 - m)/2,$$

which gives $m \geq 1 - \frac{2}{C(\tau)}\delta$. \qed

As stated, the proof gives $K = 384\tau^2$. This constant can easily be improved, which we ignored to preserve simplicity.

3 The Fourier coefficients

Consider real functions on the discrete cube $\{-1,1\}^N$. We treat this cube as a probability space with the probability of each point being $2^{-N}$. We consider the scalar product $\langle f, g \rangle = \mathbb{E} f(X) g(X)$, where $X$ is a random $\pm 1$ vector. For a set $S \subset \{1, \ldots, N\}$ we put $W_S(x_1, x_2, \ldots, x_n) = \prod_{i \in S} x_i$. It is trivial to check these functions form an orthonormal basis in our space. Let $\hat{f}(S) = \langle f, W_S \rangle$ be the Fourier coefficients of $f$. The following theorem is the main result of [4]:

**Theorem 3.1.** Let $f : \{-1,1\}^N \rightarrow \{-1,1\}$. Assume $\sum_{|S| \geq 2} \hat{f}^2(S) < \delta$. Then there exists a set $S \subset \{1, \ldots, N\}$, $|S| \leq 1$, such that $|\hat{f}(S)| > 1 - K\delta$.

For the sake of completeness we supply a proof of this theorem, somewhat simplifying the argument of [4].

**Proof.** First we will reduce the problem to balanced functions $f$ (that is functions with $\mathbb{E} f = 0$). This elegant argument is attributed in [4] to Guy Kindler.

Let $g(x, x_{n+1}) = x_{n+1} f(x_{n+1} x)$ for any $x \in \{-1,1\}^n$, $x_{n+1} \in \{-1,1\}$. Then

$$g(x, x_{n+1}) = x_{n+1} \sum_S \hat{f}(S) W_S(x_{n+1} x) = \sum_{2 \mid |S|} \hat{f}(S) W_{S \cup \{n+1\}}(x) + \sum_{2 \nmid |S|} \hat{f}(S) W_S(x).$$
Thus, in particular, $\hat{g}(\emptyset) = 0$, and the Fourier coefficients of $g$ corresponding to sets of size 1 are exactly the Fourier coefficients of $f$ corresponding to sets of size at most 1. Thus we can restrict ourselves to proving the theorem for functions with $\mathbb{E}f = 0$.

Now let $X = \sum f(\{i\})W_{\{i\}}$ and $Y = \sum_{|S|>1} \hat{f}(S)W_S$, where we treat the functions $W_S$ as random variables on the probability space $\{-1, 1\}^n$. Note that $W_{\{i\}}$ are a sequence of independent Rademacher variables. We have

$$\text{Var} |X| = \mathbb{E}(|X| - \mathbb{E}|X|)^2 \leq \mathbb{E}(|X| - 1)^2 \leq \mathbb{E}Y^2 = \delta,$$

where we use $|X + Y| = 1$, so $|X| \leq |Y| + 1$. We also have

$$\text{Var} X = \mathbb{E}X^2 = \sum \hat{f}_{\{i\}}^2 \geq 1 - \delta.$$

Let $\tilde{X} = X/\sqrt{\text{Var} X}$. Then $\tilde{X}$ is a sum of independent Rademacher variables with $\text{Var} \tilde{X} = 1$ and $\text{Var} |\tilde{X}| = \text{Var} |X|/\text{Var} X \leq \delta/(1 - \delta)$. Thus by Theorem 1.1 we get that one of the coefficients $\tilde{a}$ of $\tilde{X}$ is at least $1 - K\delta/(1 - \delta)$. Thus, the corresponding coefficient of $X$ satisfies

$$a \geq \sqrt{1 - \delta(1 - K\delta/(1 - \delta))} \geq (1 - \delta)(1 - K\delta/(1 - \delta)) = 1 - (K + 1)\delta.$$

\[ \square \]

4 Closing remarks

In our proofs we did not attempt to make the constants sharp. In view of PCP theorem applications this is no problem — the constants involved in the PCP theory tend to be uncontrollable anyway. If one would wish to obtain better results, a Chernoff type argument (for instance as given in [4]) has to be applied, the second–fourth moment comparison is too weak to provide good constants. As the Chernoff argument of [4] works once you prove $\max |a_i| \geq 0.99999$, which (for sufficiently small $\delta$) is given by our proof, we still get an elementary argument giving good constants. This alleviates the need to use outside theory (such as the results of [5], [1],[2] or the Berry-Esseen theorem) to get $\max |a_i| \geq 0.99999$.

References


