Local tail bounds for polynomials on the discrete cube

Bo'az Klartag[1] and Sasha Sodin[2]

Abstract

Let P be a polynomial of degree d in independent Bernoulli random variables which has zero mean and unit variance. The Bonami hypercontractivity bound implies that the probability that $|P| > t$ decays exponentially in $t^{2/d}$. Confirming a conjecture of Keller and Klein, we prove a local version of this bound, providing an upper bound on the difference between the e^{-r} and the e^{-r-1} quantiles of P.

1

This note is concerned with concentration inequalities for polynomials on the discrete cube. Concentration inequalities, i.e. tail bounds on the distribution of functions on high-dimensional spaces belonging to certain classes, were put forth by Vitali Milman in the 1970-s and have since found numerous applications; see e.g. [2, 3] and references therein.

Let $X_1, \ldots, X_n$ be independent, identically distributed symmetric Bernoulli variables, so that $X = (X_1, \ldots, X_n)$ is distributed uniformly on the discrete cube $\{-1, 1\}^n$. The starting point for this work is the concentration inequality for polynomials in X (see e.g. [3, Theorem 9.23]), which we now recall. Let $d \geq 1$, and consider a polynomial of the form

$$P_d(x) = \sum_{\#(S)=d} a_S \cdot \left(\prod_{i \in S} x_i \right) \tag{1}$$

where the sum runs over all subsets $S \subseteq \{1, \ldots, n\}$ of cardinality d, and the coefficients (a_S) are arbitrary real numbers. In other words, P_d is a d-homogeneous, square-free polynomial in $\mathbb{R}^n$. The Bonami hypercontractivity theorem [3, Chapter 9] tells us that for any $1 < p \leq q$,

$$\|P_d(X)\|_q \leq \left(\frac{q-1}{p-1} \right)^{d/2} \|P_d(X)\|_p. \tag{2}$$

[1]Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. Email: boaz.klartag@weizmann.ac.il. Supported by a grant from the Israel Science Foundation (ISF).

[2]School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. Email: a.sodin@qmul.ac.uk. Supported in part by a Royal Society Wolfson Research Merit Award (WM170012), and a Philip Leverhulme Prize of the Leverhulme Trust (PLP-2020-064).

A general polynomial P of degree at most d on $\{-1,1\}^n$ takes the form

$$P(x) = \sum_{k=0}^{d} P_k(x) \tag{3}$$

where P_k is a k-homogeneous, square-free polynomial. Thanks to orthogonality relations we have

$$\|P(X)\|_2^2 = \sum_{k=0}^{d} \|P_k(X)\|_2^2.$$

Hence, by the Bonami bound (2) and the Cauchy-Schwarz inequality, for any polynomial P of degree at most d and any $q \geq 3$,

$$\begin{aligned} \|P(X)\|_q &\leq \sum_{k=0}^{d} \|P_k(X)\|_q \leq \sum_{k=0}^{d} (q-1)^{k/2} \|P_k(X)\|_2 \leq \sqrt{\sum_{k=0}^{d} (q-1)^k} \cdot \sqrt{\sum_{k=0}^{d} \|P_k(X)\|_2^2} \\ &\leq \sqrt{2} \cdot (q-1)^{d/2} \cdot \|P(X)\|_2 \leq \sqrt{2} q^{d/2} \|P(X)\|_2. \end{aligned} \tag{4}$$

For $r > 0$ (not necessarily integer), write a_r for a e^{-r}-quantile of $P(X)$, i.e. a number satisfying

$$\mathbb{P}(P(X) \geq a_r) \geq \frac{1}{e^r} \qquad \text{and also} \qquad \mathbb{P}(P(X) \leq a_r) \geq 1 - \frac{1}{e^r}.$$

Assume the normalization $\|P(X)\|_2 = 1$. It follows from (4) that if $q \geq 3$ then

$$\frac{1}{e^r} \leq \mathbb{P}(P(X) \geq a_r) \leq \frac{\mathbb{E}|P(X)|^q}{a_r^q} \leq \left(\sqrt{2} \cdot \frac{q^{d/2}}{a_r} \right)^q.$$

Substituting $q = 2r/d$ (when $r \geq 3d/2$), we get

$$a_r \leq \sqrt{2} \cdot (2er/d)^{d/2} \leq (Cr/d)^{d/2} \quad (r \geq 3d/2), \tag{5}$$

with a universal constant $C = 4$. Without assuming any normalisation, we obtain

$$a_r - a_1 \leq C^d \left(\frac{r}{d} + 1 \right)^{d/2} \|P(X)\|_2 \,. \tag{6}$$

(with a different numerical constant $C > 0$), which is valid for all $r \geq 1$.

The estimate (6) is a a tail bound for the distribution of $P(X)$, i.e. concentration inequality. We refer to [2] and references therein for background on concentration inequalities, particularly, for polynomials, and to [3] for applications of (6) .

In some applications, it is important to have bounds on $a_s - a_r$ when $s \geq r$ are close to one another, e.g. $s = r + 1$. Such bounds are called *local* tail bounds; see [1] and references therein. The following proposition, confirming a conjecture of Nathan Keller and Ohad Klein, provides a local version of (6). In the case $d = 1$, it follows from the results in the aforementioned work [1].

Proposition 1. *Let P be a polynomial of degree at most d on $\{-1,1\}^n$. Then for all $r \geq 1$,*

$$a_{r+1} - a_r \leq C^d \left(\frac{r}{d} + 1\right)^{\frac{d}{2}-1} \|P(X)\|_2 \, , \tag{7}$$

where $C > 0$ is a universal constant.

Clearly, (7) implies (6). The estimate (7) gives the right magnitute of $a_r - a_{r+1}$, say, for

$$P(X) = (X_1 + \cdots + X_n)^d \, , \quad n \gg 1 \, . \tag{8}$$

2

We now turn to the proof of Proposition 1. Write $\partial_i P$ for the partial derivative of P with respect to the i^{th} variable. Thus

$$\partial_i P(x) = \frac{P(T_i^1 x) - P(T_i^{-1} x)}{2} \qquad \text{for } x \in \{-1,1\}^n,$$

where T_i^j is the map that sets the i^{th}-coordinate of x to the value j, and keeps the other coordinates intact. Observe that $\partial_i P$ is a polynomial of degree at most $d-1$ if P is of degree d. We denote by ∇P the vector function with coordinates $\partial_i P$. The first step in the proof of Proposition 1 is to sharpen the quantile bound (5).

Lemma 2. *Let P be a polynomial of degree at most d with $\mathbb{E}|P(X)|^2 = 1$. Then for any non-empty subset $A \subseteq \{-1,1\}^n$ of relative size $\varepsilon = \#(A)/2^n$ we have*

$$\frac{1}{\#(A)} \sum_{x \in A} |P(x)|^2 \leq C^d \cdot \max\left\{1, \left(\frac{|\log \varepsilon|}{d}\right)^d\right\}, \tag{9}$$

and

$$\frac{1}{\#(A)} \sum_{x \in A} |\nabla P(x)|^2 \leq C^d \cdot \max\left\{1, \left(\frac{|\log \varepsilon|}{d}\right)^{d-1}\right\} \, , \tag{10}$$

for a universal constant $C > 0$.

Proof. Let $q \geq 3$. By Hölder's inequality followed by an application of (4),

$$\begin{aligned} \sum_{x \in A} |P(x)|^2 &\leq (\#(A))^{1-2/q} \cdot \left(\sum_{x \in A} |P(x)|^q\right)^{2/q} = (\#(A))^{1-2/q} \cdot 2^{2n/q} \cdot \|P(X)\|_q^2 \\ &\leq (\#(A))^{1-2/q} \cdot 2^{2n/q} \cdot 2q^d \, , \end{aligned}$$

whence

$$\frac{1}{\#(A)} \sum_{x \in A} |P(x)|^2 \leq 2\varepsilon^{-2/q} q^d \, .$$

The estimate (9) clearly holds for $\varepsilon \geq e^{-\frac{3d}{2}}$, therefore we assume that $\varepsilon < e^{-\frac{3d}{2}}$. Set

$$q = 2|\log \varepsilon|/d \geq 3$$

and obtain

$$\frac{1}{\#(A)} \sum_{x \in A} |P(x)|^2 \leq \left(\frac{C}{d}\right)^d |\log \varepsilon|^d.$$

This proves (9). Since $\partial_i P$ is a polynomial of degree at most $d-1$, from (9),

$$\frac{1}{\#(A)} \sum_{x \in A} |(\partial_i P)(x)|^2 \leq C^d \cdot \max\left\{1, \left(\frac{|\log \varepsilon|}{d}\right)^{d-1}\right\} \cdot \mathbb{E}|(\partial_i P)(X)|^2 ,$$

whence

$$\frac{1}{\#(A)} \sum_{x \in A} |(\nabla P)(x)|^2 \leq C^d \cdot \max\left\{1, \left(\frac{|\log \varepsilon|}{d}\right)^{d-1}\right\} \cdot \mathbb{E}|(\nabla P)(X)|^2 .$$

We decompose $P(X) = \sum_{k=0}^d P_k(X)$ as in (3), and use the orthogonality relations

$$\mathbb{E}|\nabla P(X)|^2 = \sum_{k=0}^d \mathbb{E}|\nabla P_k(X)|^2 = \sum_{k=0}^d k \cdot \mathbb{E}|P_k(X)|^2 \leq d \cdot \mathbb{E}|P(X)|^2 = d.$$

This proves (10). □

Note that for any $f : \{-1,1\}^n \to \mathbb{R}$,

$$\sum_{x \in \{-1,1\}^n} |\nabla f(x)|^2 \leq 2 \cdot \sum_{x \in \{-1,1\}^n} |\nabla f(x)|^2 \cdot 1_{\{f(x) \neq 0\}}. \tag{11}$$

Indeed, the expression on the left-hand side of (11) is the sum over all oriented edges $(x,y) \in E$ in the Hamming cube of the squared difference $|f(x) - f(y)|^2/4$. This is clearly at most twice the sum over all oriented edges $(x,y) \in E$ of the quantity $|f(x) - f(y)|^2 \cdot 1_{\{f(x) \neq 0\}}/4$.

Recall the log-Sobolev inequality (e.g. [3, Chapter 10]) which states that for any function $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbb{E}f^2(X) \log f^2(X) - \mathbb{E}f^2(X) \cdot \log \mathbb{E}f^2(X) \leq 2\mathbb{E}|\nabla f(X)|^2. \tag{12}$$

Moreover, let $A \subseteq \{-1,1\}^n$ be a non-empty set and denote $\varepsilon = \#(A)/2^n$. If the function f is supported in A and is not identicaly zero, then denoting $g = f/\sqrt{\mathbb{E}f^2(X)}$,

$$\begin{aligned} \mathbb{E}f^2(X) \log f^2(X) - \mathbb{E}f^2(X) \cdot \log \mathbb{E}f^2(X) &= \mathbb{E}f^2(X) \cdot \mathbb{E}g^2(X) \log g^2(X) \\ &\geq \mathbb{E}f^2(X) \cdot |\log \varepsilon|, \end{aligned} \tag{13}$$

because g^2 is supported in A, and among all probability distributions supported in A, the maximal entropy is attained for the uniform distribution.

Proof of Proposition 1. Without loss of generality $\|P(X)\|_2 = 1$. We may assume that $a_{r+1} > a_r$, as otherwise there is nothing to prove. Let $U = \{x \in \{-1,1\}^n \,;\, f(x) > a_r\}$ and set $\varepsilon = \#(U)/2^n$. Then $e^{-(r+1)} \le \varepsilon \le e^{-r}$, by the definition of the quantiles a_r and a_{r+1}. Denote $\chi(t) = \max(t - a_r, 0)$; this is a 1-Lipschitz function on the real line. Applying the log-Sobolev inequality (12) to the function $h = \chi \circ P : \{-1,1\}^n \to \mathbb{R}$ we get

$$\mathbb{E}h^2(X)\log h^2(X) - \mathbb{E}h^2(X) \cdot \log \mathbb{E}h^2(X) \le 2\mathbb{E}|\nabla h|^2(X). \tag{14}$$

Since h is supported in U, with $\varepsilon = \#(U)/2^n$, by (13) and (14),

$$\mathbb{E}h^2(X) \cdot |\log\varepsilon| \le 2\mathbb{E}|\nabla h|^2(X) \le 4\mathbb{E}|\nabla h(X)|^2 \cdot 1_{\{h(X)>0\}}.$$

The last passage is the content of (11). Since χ is 1-Lipschitz, we know that $|\nabla h|^2 \le |\nabla P|^2$. Hence, by (10),

$$\mathbb{E}|\nabla h(X)|^2 \cdot 1_{\{h(X)>0\}} \le \mathbb{E}|\nabla P(X)|^2 1_{\{X\in U\}} \le \varepsilon \cdot C^d \cdot \max\left\{1, \left(\frac{|\log\varepsilon|}{d}\right)^{d-1}\right\}.$$

To summarize,

$$\mathbb{E}h^2(X) \cdot |\log\varepsilon| \le \varepsilon \cdot C_1^d \cdot \max\left\{1, \left(\frac{|\log\varepsilon|}{d}\right)^{d-1}\right\}, \tag{15}$$

for a universal constant $C_1 > 0$. Recall that $e^{-(r+1)} \le \varepsilon \le e^{-r}$. By the definition of a_{r+1}, we know that $h(X) \ge a_{r+1} - a_r$ with probability at least $e^{-(r+1)}$. Therefore, from (15),

$$e^{-(r+1)} \cdot (a_{r+1} - a_r)^2 \cdot \frac{r}{2} \le e^{-r} \cdot C_1^d \cdot \max\left\{1, \left(\frac{2r}{d}\right)^{d-1}\right\}$$

or

$$a_{r+1} - a_r \le C_2^d \cdot \max\left\{\frac{1}{\sqrt{r}}, \left(\frac{r}{d}\right)^{d/2-1}\right\} \le C_3^d \left(\frac{r}{d} + 1\right)^{\frac{d}{2}}. \qquad \square$$

3

We remark that Proposition 1 implies the following corollary which holds true without the normalization by $\|P(X)\|_2$.

Corollary 3. *There exists $C > 0$ such that the following holds. Let P be a polynomial of degree at most d with $\mathbb{E}P(X) = 0$. Then for $r \ge Cd$,*

$$a_{r+1} \le a_r \left[1 + C^d \left(\frac{r}{d} + 1\right)^{\frac{d}{2}-1}\right]. \tag{16}$$

Remark 4. We conjecture that (16) also holds with the power -1 in place of $\frac{d}{2} - 1$. Such an estimate would give the right order of magnitude for the polynomial (8).

Proof of Corollary 3. Write $\sigma^2 = \mathbb{E}|P(X)|^2$. We shall prove that $\sigma \leq C_1^d a_r$. Once this inequality is established, we deduce from Proposition 1 that

$$\frac{a_{r+1} - a_r}{\sigma} \leq C^d \left(\frac{r}{d} + 1\right)^{\frac{d}{2}-1} ,$$

whence

$$a_{r+1} \leq a_r + \sigma \cdot C^d \left(\frac{r}{d} + 1\right)^{\frac{d}{2}-1} \leq a_r \left(1 + (CC_1)^d \left(\frac{r}{d} + 1\right)^{\frac{d}{2}-1}\right) ,$$

as claimed.

Let $\sigma_\pm = \sqrt{\mathbb{E}(P(X)_\pm)^2}$. First, we claim that $\sigma_+ \geq C_2^{-d}\sigma$. Indeed, if $\sigma_+ \geq \sigma_-$ then $\sigma_+ \geq \sigma/\sqrt{2}$. If $\sigma_+ < \sigma_-$, then, using (4),

$$\begin{aligned} \sigma_+ \geq \mathbb{E}P(X)_+ = \mathbb{E}P(X)_- &\geq \frac{(\mathbb{E}P(X)_-^2)^{3/2}}{(\mathbb{E}P(X)_-^4)^{1/2}} \\ &\geq \frac{\sigma_-^3}{2 \cdot 3^d \cdot (\sigma_+^2 + \sigma_-^2)} \geq \frac{1}{4 \cdot 3^d}\sigma_- . \end{aligned}$$

Second, another application of (4) yields

$$\mathbb{E}P(X)_+^4 \leq \mathbb{E}P(X)^4 \leq 4 \cdot 3^d \sigma^4 \leq C_3^d \sigma_+^4 ,$$

thus by the Paley–Zygmund inequality

$$e^{-Cd} \geq e^{-r} \geq \mathbb{P}\left\{P(X) > a_r\right\} \geq \frac{(1 - a_r^2/\sigma_+^2)_+^2}{C_3^d} ,$$

whence $\sigma_+ \leq 2a_r$ if we ensure that, say, $e^C \geq 2C_3$. This concludes the proof. □

Finally, we remark that both Proposition 1 and Corollary 3 can be generalised in several directions. For example, instead of the Hamming cube, one can consider a general measure which is invariant under a Markov diffusion satisfying the Bakry–Émery CD(R, ∞) condition; in this setting, linear combinations of eigenfunctions of the generator play the rôle of polynomials. The proof requires only notational modifications.

Acknowledgement. We are grateful to Nathan Keller for helpful correspondence.

References

[1] Devroye, L., Lugosi, G., *Local tail bounds for functions of independent random variables.* Ann. Probab. 36 (2008), no. 1, 143–159.

[2] Giannopoulos, A. A., Milman, V. D., *Concentration property on probability spaces.* Adv. Math. 156 (2000), no. 1, 77–106.

[3] O'Donnell, R., *Analysis of Boolean Functions*. Cambridge University Press, New York, 2014.