CAS CS 599 B: Mathematical Methods for TCS

Lecturer: Mark Bun — Spring 2022

Lecture Notes 6:

Intro to Hypercontractivity

Reading.

- O'Donnell, Analysis of Boolean Functions §9.1-9.5

Here are some examples of random variables. Discussion question: Which ones would you say are "nice" to work with? Which ones are not so nice? Why?

1. $x \sim \{-1, 1\}$
2. $u \sim U[0, 1]$
3. t an r.v. with probability density $\propto \frac{1}{1+z^2}$
4. $g \sim \mathcal{N}(0, 1)$
5. $y = \begin{cases} 1 & \text{w.p. } 2^{-n} \\ 0 & \text{w.p. } 1 - 2^{-n} \end{cases}$

Definition 1. For a parameter $B \geq 1$, we say that a random variable X is B-reasonable if $\mathbb{E}[X^4] \leq B\mathbb{E}[X^2]^2$. (Equivalently, if $\|X\|_4 \leq B^{1/4}\|X\|_2$.)

In statistics, the parameter B is an upper bound on the "kurtosis" of X, which is a measure of how heavy its tails are.

For the examples above, x is 1-reasonable, u is $(9/5)$-reasonable, t is not reasonable for any value of B, g is 3-reasonable, and y is 2^n-reasonable.

Why should we like reasonable random variables? First off, they have small tails:

Proposition 2. *Let nonzero X be B-reasonable. Then*

$$\Pr[|X| \geq t \cdot \|X\|_2] \leq B/t^4.$$

This should be compared with Chebyshev's inequality, which gives the weaker bound $\Pr[|X| \geq t \cdot \|X\|_2] \leq 1/t^2$.

Proof. This follow from Markov's inequality on the fourth moment:

$$\begin{aligned}
\Pr[|X| \geq t \cdot \|X\|_2] &= \Pr[X^4 \geq t^4 \|X\|_2^4] \\
&\leq \frac{\mathbb{E}[X^4]}{t^4 \mathbb{E}[X^2]^2} \\
&\leq \frac{B}{t^4}.
\end{aligned}$$

□

On the other hand, reasonable random variables are also *anti-concentrated*, meaning they don't put "too much" of their probability mass around 0. Our tool for deriving this is the Paley-Zygmund inequality, which you can think of as a reverse Chebyshev inequality.

Proposition 3 (Paley-Zygmund)**.** *Let $Z \geq 0$ be a random variable and let $t \in [0,1]$. Then*

$$\Pr[Z > t\mathbb{E}[Z]] \geq (1-t)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}.$$

Proof. By linearity, we can write

$$\mathbb{E}[Z] = \mathbb{E}[Z \cdot \mathbf{1}_{\{Z \leq t\mathbb{E}[Z]\}}] + \mathbb{E}[Z \cdot \mathbf{1}_{\{Z > t\mathbb{E}[Z]\}}].$$

By construction, the first term is at most $t\mathbb{E}[Z]$. Meanwhile, by Cauchy-Schwarz, we can bound the second by

$$\mathbb{E}[Z \cdot \mathbf{1}_{\{Z > t\mathbb{E}[Z]\}}] \leq \sqrt{\mathbb{E}[Z^2] \cdot \mathbb{E}[\mathbf{1}^2_{\{Z > t\mathbb{E}[Z]\}}]} = \sqrt{\mathbb{E}[Z^2]} \cdot \sqrt{\Pr[Z \geq t\mathbb{E}[Z]]}.$$

Rearranging gives the stated inequality. □

We can now obtain an anti-concentration result for reasonable random variables.

Proposition 4. *Let nonzero X be B-reasonable. Then for all $t \in [0,1]$,*

$$\Pr[|X| > t \cdot \|X\|_2] \geq \frac{(1-t^2)^2}{B}.$$

Proof.

$$\begin{aligned}\Pr[|X| > t \cdot \|X\|_2] &= \Pr[X^2 > t^2 \cdot \mathbb{E}[X^2]] \\ &\geq (1-t^2)^2 \frac{\mathbb{E}[X^2]^2}{\mathbb{E}[X^4]} \\ &\geq \frac{(1-t^2)^2}{B}.\end{aligned}$$

□

1 Bonami's Lemma

Bonami's Lemma says that if f is a low-degree polynomial, then $f(x)$ for $x \sim \{-1,1\}^n$ is reasonable.

Lemma 5 (Bonami's Lemma)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ have degree at most k. Then $f(x)$ for $x \sim \{-1,1\}^n$ is 9^k-reasonable, i.e., $\mathbb{E}[f^4] \leq 9^k \mathbb{E}[f^2]^2$.*

Before proving the lemma, here's a reminder of some useful technical tools. Recall the *derivative* operator

$$(D_i f)(x) = \frac{f(x^{(i \to 1)}) - f(x^{(i \to -1)})}{2}.$$

The derivative has a counterpart, the "expectation" operator, which captures the average value of f at x in the direction of i.

$$(E_i f)(x) = \frac{f(x^{(i \to 1)}) + f(x^{(i \to -1)})}{2}.$$

Observe that for any function $f : \{-1, 1\}^n \to \mathbb{R}$, we have

$$f(x) = x_n(D_n f)(x) + (E_n f)(x).$$

The utility of this expression is that the functions $D_n f$ and $E_n f$ depend only on the variables $x_1, \ldots, x_{n-1}$. In particular, for uniform x, the random variables $D_n f(x)$ and $E_n f(x)$ are independent of x_n. So this decomposition facilitates proofs by induction on the number of variables n.

Proof. We prove the claim by induction on n.

Base case: If $n = 0$, then f is constant and the statement is true.

Inductive case: Assume the statement is true for $n-1$ variables. Write $f(x) = x_n(D_n f)(x) + (E_n f)(x)$. Define the random variables $f = f(x), d = (D_n f)(x), e = (E_n f)(x)$ for $x \sim \{-1, 1\}^n$. Then

$$\begin{aligned}
\mathbb{E}[f^4] &= \mathbb{E}[(x_n \cdot d + e)^4] \\
&= \mathbb{E}[x_n^4 d^4] + 4\mathbb{E}[x_n^3 d^3 e] + 6\mathbb{E}[x_n^2 d^2 e^2] + 4\mathbb{E}[x_n d e^3] + \mathbb{E}[e^4] \\
&= \mathbb{E}[x_n^4]\mathbb{E}[d^4] + 4\mathbb{E}[x_n^3]\mathbb{E}[d^3 e] + 6\mathbb{E}[x_n^2]\mathbb{E}[d^2 e^2] + 4\mathbb{E}[x_n]\mathbb{E}[d e^3] + \mathbb{E}[e^4] \\
&= \mathbb{E}[d^4] + 6\mathbb{E}[d^2 e^2] + \mathbb{E}[e^4].
\end{aligned}$$

The second-to-last inequality uses the fact that x_n is independent of d and e. The final equality uses the fact that $\mathbb{E}[x_n^4] = \mathbb{E}[x_n^2] = 1$ and $\mathbb{E}[x_n^3] = \mathbb{E}[x_n] = 0$.

Similarly,

$$\mathbb{E}[f^2] = \mathbb{E}[(x_n \cdot d + e)^2] = \mathbb{E}[x_n^2]\mathbb{E}[d^2] + 2\mathbb{E}[x_n]\mathbb{E}[de] + \mathbb{E}[e^2] = \mathbb{E}[d^2] + \mathbb{E}[e^2].$$

To bound $\mathbb{E}[f^4]$ in terms of $\mathbb{E}[f^2]$, we analyze the summands on the left. Recall that if f is a degree-k polynomial, then its derivative d is a degree $k-1$ polynomial. So by the inductive hypothesis

$$\mathbb{E}[d^4] \leq 9^{k-1}\mathbb{E}[d^2]^2.$$

We don't get a reduction in degree for the expectation term, but the inductive hypothesis still tells us that $\mathbb{E}[e^4] \leq 9^k \mathbb{E}[e^2]^2$. For the middle term, we apply Cauchy-Schwarz and the inductive hypothesis again to get

$$\mathbb{E}[d^2 e^2] \leq \sqrt{\mathbb{E}[d^4]\mathbb{E}[e^4]} \leq \sqrt{9^{k-1}\mathbb{E}[d^2]^2 \cdot 9^k \mathbb{E}[e^2]^2} = \frac{1}{3} \cdot 9^k \mathbb{E}[d^2]\mathbb{E}[e^2].$$

Putting everything together,

$$\begin{aligned}
\mathbb{E}[f^4] &\leq 9^{k-1}\mathbb{E}[d^2]^2 + 2 \cdot 9^k \mathbb{E}[d^2]\mathbb{E}[e^2] + 9^k \mathbb{E}[e^2]^2 \\
&\leq 9^k (\mathbb{E}[d^2] + \mathbb{E}[e^2])^2 \\
&= 9^k \mathbb{E}[f^2]^2.
\end{aligned}$$

$\square$

2 An Application: The FKN Theorem

Theorem 6 (Friedgut-Kalai-Naor)**.** *Suppose $f : \{-1,1\}^n \to \{-1,1\}$ has $\mathbf{W}^1[f] = 1-\delta$. Then f is $O(\delta)$ close to $\pm\chi_i$ for some $i \in [n]$.*

Proof. Let $g = f^{=1}$ be the degree-1 part of f. By Parseval, $\mathbb{E}[g^2] = \mathbf{W}^1[f] = 1-\delta$. Applying last week's exercise to g,

$$\begin{aligned}\frac{1}{2}\operatorname{Var}\left[g^2\right] &= \sum_{i\neq j}\hat{g}(i)^2\hat{g}(j)^2\\ &= \left(\sum_{i=1}^n \hat{g}(i)^2\right)^2 - \sum_{i=1}^n \hat{g}(i)^4\\ &= (1-\delta)^2 - \sum_{i=1}^n \hat{f}(i)^4\\ &\geq (1-2\delta) - \sum_{i=1}^n \hat{f}(i)^4.\end{aligned}$$

Rearranging, we get

$$\begin{aligned}(1-2\delta) - \frac{1}{2}\operatorname{Var}\left[g^2\right] &\leq \sum_{i=1}^n \hat{f}(i)^4\\ &\leq \max_i\{\hat{f}(i)^2\}\sum_{i=1}^n \hat{f}(i)^2\\ &\leq \max_i\{\hat{f}(i)^2\}\\ &\leq \max_i\{|\hat{f}(i)|\}.\end{aligned}$$

Noting that $\hat{f}(i)$ is the correlation of f with χ_i, to prove the theorem, it suffices to show that $\operatorname{Var}\left[g^2\right] = O(\delta)$.

By Bonami's Lemma, the degree-2 polynomial g^2 is 81-reasonable. So by anticoncentration Proposition 4, setting $X = g^2 - \mathbb{E}[g^2] = g^2 - (1-\delta)$ and $t = 1/2$, we get

$$\Pr\left[|g^2 - (1-\delta)| \geq \frac{1}{2}\sqrt{\operatorname{Var}[g^2]}\right] \geq \frac{(3/4)^2}{81} = \frac{1}{144}.$$

By the triangle inequality,

$$\Pr\left[|g^2 - 1| \geq \frac{1}{2}\sqrt{\operatorname{Var}[g^2]} - \delta\right] \geq \frac{1}{144}.$$

Intuitively, this says that if $\operatorname{Var}\left[g^2\right]$ is large, then g^2 (and hence g) deviates from 1 with large probability. But since $|f| = 1$ always, this means we must have that g deviates from f with large probability. But this can't happen, because $\mathbb{E}[(f-g)^2] = \delta$ is small, so we must conclude that $\operatorname{Var}\left[g^2\right]$ is small.

More precisely, one can show that if $\operatorname{Var}\left[g^2\right] > 6400\delta$, then $|g^2 - 1| > 39\sqrt{\delta} \implies (f-g)^2 \geq 169\delta$. So

$$\Pr[|g^2 - 1| > 39\sqrt{\delta}] \geq \frac{1}{144} \implies \mathbb{E}[(f-g)^2] \geq \frac{1}{144}\cdot 169\delta > \delta,$$

a contradiction.

□

3 Hypercontractivity

Let $f : \{-1,1\}^n \to \mathbb{R}$, and let $f^{=k}$ be its degree-k part. The noise operator has a very clean effect on the homogeneous polynomial $f^{=k}$:

$$T_\rho f^{=k} = \sum_{|S|=k} \rho^{|S|} \hat{f}(S) = \rho^k f^{=k}.$$

Setting $\rho = 1/\sqrt{3}$ and applying Bonami's Lemma,

$$\begin{aligned} \|T_{1/\sqrt{3}} f^{=k}\|_4 &= \frac{1}{\sqrt{3}^k} \|f^{=k}\|_4 \\ &= \|f^{=k}\|_2. \end{aligned}$$

It turns out that we can relax the condition that f is a homogeneous polynomial:

Theorem 7 ($(4,2)$-Hypercontractivity Theorem)**.** *For every $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\|T_{1/\sqrt{3}} f\|_4 \le \|f\|_2.$$

One can conclude this from the calculation above, plus a few nice analytic tricks. Another way to prove it is by just repeating the induction underlying Bonami's Lemma.

One can also conclude Bonami's Lemma from the hypercontractivity theorem. Extending the definition of the noise operator to $\rho > 1$ via $T_\rho f = \sum \rho^{|S|} \hat{f}(S) \chi_S$, we have for all f of degree k that

$$\|f\|_4 = \|T_{1/\sqrt{3}} T_{\sqrt{3}} f\|_4 \le \|T_{\sqrt{3}} f\|_2 \le \sqrt{3}^k \|f\|_2.$$

Raising both sides to the 4th power gives Bonami's Lemma.

So the hypercontractivity theorem is basically just a reformulation of Bonami's Lemma, but it says something different. It quantifies the extent to which T_ρ is a "smoothing" operator, i.e., one which mollifies peaks in the distribution of $f(x)$. You saw in the exercises that T_ρ is a contractive map: $\|T_\rho f\|_2 \le \|f\|_2$, for instance. The hypercontractivity theorem says that T_ρ is actually a *hyper*contractive map, in that it smooths out f even when one measures $T_\rho f$ in a higher norm.

With some more machinery (see Section 10.1 of O'Donnell) one can prove the general hypercontractivity theorem.

Theorem 8 (Hypercontractivity Theorem)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $1 \le p \le q \le \infty$. Then $\|T_\rho f\|_q \le \|f\|_p$ for $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$.*

We won't prove this, but there is a nice trick to getting another special case "for free" as a consequence of the $(4,2)$-hypercontractivity theorem. It relies on the observation that the noise operator is self-adjoint:

Claim 9. *For $\rho \in \mathbb{R}$ and $f, g : \{-1,1\}^n \to \mathbb{R}$, we have $\langle T_\rho f, g\rangle = \sum \rho^{|S|} \hat{f}(S)\hat{g}(S) = \langle f, T_\rho g\rangle$.*

Corollary 10. *For all* $f : \{-1, 1\}^n \to \mathbb{R}$,

$$\|T_{1/\sqrt{3}} f\|_2 \le \|f\|_{4/3}.$$

Proof. Writing $T = T_{1/\sqrt{3}}$, we have

$$\begin{aligned}
\|Tf\|_2^2 &= \langle Tf, Tf\rangle \\
&= \langle f, TTf\rangle \\
&\le \|f\|_{4/3}\|TTf\|_4 && \text{by Hölder's inequality} \\
&\le \|f\|_{4/3}\|Tf\|_2 && \text{by (2, 4)-hypercontractivity.}
\end{aligned}$$

Dividing both sides by $\|Tf\|_2$ gives the claim. $\square$