STAT 206A: Polynomials of Random Variables	Harper's Thm; Prob. of Reading Input

Lecture 5

Lecture date: Sept 15 *Scribe: Grant Schoenebeck*

1 Continuation of Harper's Theorem

Beginning this lecture we are in the middle of proving Harper's theorem. A quick review of notation follows, see previous lecture's notes for more detail.

$A \subseteq \{0,1\}^n$

$L_m :=$ First m elements of $\{0,1\}^n$ in lexagraphical order

$\Psi(A) :=$ The size of the boundary edge set of A (see previous lecture's notes for more detail)

$C_i(A) :=$ ith compression operator applied to A (again, see previous lecture's notes for more detail)

Theorem 1 (Harper) *Among all sets of size m, Ψ is minimized only at L_m.*

We were proving the theorem using three lemmas, the first two were proven in the previous lecture.

Lemma 2 $\Psi(C_i(A)) \le \Psi(A)$

Lemma 3 *The compression operation stabilizes. That is after repeated application we will arrive at a set A so that for $0 \le i \le n$, $C_i(A) = A$.*

Proof: We look at the function $\Phi(A) = \sum_{x \in A} x$ (where we treat an elements of $\{0,1\}^n$ as integers). This function decreases with application of C_i, and strictly decreases if $C_i(A) \ne A$. Because $\Phi(A)$ only takes positive integer values, it must stabilize at some point. □

Lemma 4 *If $C_i(B) = B$ for all i, then either*

- $B = L_m$ *OR*
- $M = 2^{n-1}$ *and* $B = L_{m-1} \cup \{(1,0,0,\dots,0)\}$

Proof: In the cases that every element with $x_1 = 0$ belongs to B, we are done. For if C_1 does not change B, it must be L_m. For the same reason, in the cases that no element with $x_1 = 1$ belongs to B, we are done.

This leaves us with the cases were some element where $x_1 = 1$ belongs to B and some element with $x_1 = 0$ does not belong to B. So by the stability of B we have that that $0 < |B_1(1)| \leq |B_0(1)| < 2^{n-1}$.

We first show that $|B_1(1)|$ is less than 2. For the sake of contradiction assume that $|B_1(1)|$ contains more than 1 element. By the stability of C_1, B contains $(1, 0, 0, \dots, 0)$ and $(1, 0, 0, \dots, 1)$, but $(0, 1, 1, \dots, 1)$ does not belong to B. However, C_n would then move $(1, 0, 0, \dots, 1)$ to $(0, 1, 1, \dots, 1)$ or below. This contradicts the stability of B by actions of C_i.

The next case is where $|B_1(1)| = 1$ but $|B_0(1)| \leq 2^{n-1} - 2$. It follows similarly to the aforementioned case that if $|B_0(1)| \leq 2^{n-1} - 2$ then by applying C_n to B the element in $B_1(1)$ will fall to $B_0(1)$. This is because $(1, 0, 0, \dots, 0)$ must be the element of $B_1(1)$ and that $(0, 1, 1, \dots, 1, 0)$ must not be an element of $B_0(1)$.

The final case is where $|B_1(1)| = 1$ and $|B_0(1)| = 2^{n-1} - 1$. This is easily seen to be the exception case in the statement of the lemma.

$\square$

We note that all that is left to prove Harper's Theorem is to show that for $m = 2^{n-1}$ $\Psi(L_m) < \Psi(L_{m-1} \cup \{(1, 0, \dots, 0)\}$. This is easily done.

Corollary 5 *If A is of size m, then $\Psi(A) \geq m(n - \log_2 n)$. (Recall that $m \leq 2^{n-1}$.)*

Proof: We will show that $\Psi(L_m) \geq m(n - \log_2 n)$ by induction on n. The base case is trivial, so we proceed to the inductive step.

$m \leq 2^{n-2}$ In this case the inductive step gives us a bound for $\Psi_{n-1}(L_m)$.

We notice that if we look at L_m in $\{0, 1\}^n$ but restrict our view to edges in $\{0, 1\}^{n-1}$ the the number of edges is $\Psi_{n-1}(L_m)$. There are m edges when we look at the new coordinate. So the number of edges is $\geq \Psi_{n-1}(L_m) + m \geq m(n - 1 - \log_2 m) + m = m(n - \log_2 n)$.

$2^{n-2} + 1 \leq m \leq 2^{n-1}$ This case is only slightly trickier. Here we cannot ask about $\Psi_{n-1}(L_m)$. But we can notice that if we look at L_m in $\{0, 1\}^n$ but restrict our view to edges in $\{0, 1\}^{n-1}$ the the number of edges is $\Psi_{n-1}(L_m) = \Psi_{n-1}(L_{2^{n-1}-m})$. There are again m edges when we look at the new coordinate. So the number of edges is $m + \Psi_{n-1}(L_{2^{n-1}-m}) = m + (2^{n-1} - m)(n - 1 - \log_2(n - 1)) \geq m(n - \log_2 m)$. The last step follows because $m \geq 2^{n-1} - m$.

$\square$

2 Influence in terms of the probability of reading the input

We now look influences in terms of the probability of needing to read a particular variable in the input. We first define some notation:

- Let $f \in L_2(\prod_{i=1}^n \mu_i)$
- Let T be a randomized algorithm that evaluates f (and is always correct)
- Let $\delta_i(T) :=$ the overall probability of querying input i (over the randomness of the input and the randomness of T)
- $\Delta := \sum_{i=1}^n \delta_i(T)$

Exercise 6 *For 1 point.* $f : \{-1,1\}_0^n \to \{-1,1\}$.

- *Let $f = \prod x_i$. How many bits are needed to evaluate f? What is the minimum possible value for $\Delta(T)$?*
- *Let $f =$ Rec-Maj$_3$. Show that $2^k \leq \Delta(f) \leq (2.5)^k$.*

Theorem 7 *(Schramm, Steif; O'Donnell, Saks, Schramm, Servidio) Let $f, g \in L_2(\prod \mu_i)$, T compute f, W be an anti-chain, and $g = \sum_{S \in W} g_S$ (where g_S is computed using only coordinates in S) then*

$$(\mathbf{Cov}[f,g])^2 \leq \mathbf{Var}[f] \sum_i \delta_i(T) I_i(g)$$

We say that $W \subseteq 2^{[N]}$ is an anti-chain *W if for all $s_1, s_2 \in W$, s_1 is not a proper subset of s_2.*

Corollary 8 *$f : \{-1,1\}_0^n \to \mathbb{R}$ then*

$$\sum \langle f, U_{\{i\}} \rangle \leq \sqrt{\mathbf{Var}[f]} \sqrt{\Delta(T)}$$

in particular, if the range of f is $\{-1,1\}$ and f is monotone, then

$$\sum_{i=1}^n I_i(f) \leq \sqrt{\mathbf{Var}[f]} \sqrt{\Delta(T)}$$

Proof:[of Corollary] For the first part, just let $g = \sum_{i=1}^{n} U_{\{i\}}$. Then it is easy to see that $I_i(g) = 1$ for all i.

For the second part we simply note that for monotone functions from $\{-1,1\}_0^n \to \{-1,1\}$ we have that $I_i(f) = \langle u_{\{i\}}, f \rangle$. $\square$

This corollary improves the easily verifiable bound that $\sum I_i \leq \sum \delta_i$.

Example 9 *Let $f : \{-1,1\}_0^n \to \{-1,1\}$ and define f such that $f = 1$ if $x_1 = x_2 = \cdots = x_n = 1$ and $f = 0$ otherwise. Then let T be the obvious algorithm that looks at x_1, then looks at x_2, etc. and rejects if it ever sees in input that is 0 and accepts if they are all equal to 1.*

- $\Delta \approx 2$ *(actually* $= 2 - \frac{n-1}{2^{n-1}}$*)*
- $\sum I_i \leq \sqrt{\mathbf{Var}[f] \cdot 2} \leq c \cdot 2^{-n/2}$

2.1 Another Application

Let $y = \sum_{S:|S|=k} \hat{f}(S) U_s$ and let $\delta(T) = \max_i \delta_i(T)$. Then

$$
\begin{aligned}
(\mathbf{Cov}[f,g])^2 &\leq \mathbf{Var}[f] \cdot \sum_{i=1}^{n} \left(\delta_i(T) \cdot \sum_{S:i \in S} \hat{f}^2(S) \right) \\
(\sum_{S:|S|=k} \hat{f}^2(S))^2 &\leq \mathbf{Var}[f] \cdot k \cdot \delta(T) \cdot \sum_{S:|S|=k} \hat{f}^2(S) \\
\sum_{S:|S|=k} \hat{f}^2(S) &\leq k \cdot \mathbf{Var}[f] \cdot \delta(T)
\end{aligned}
$$

So complex functions (where $\delta(T)$ is small) have complex $\hat{f}$ (a lot of mass on the tail) because

$$\sum_{S:|S| \leq k} \hat{f}^2(S) \leq \frac{k(k-1)}{2} \mathbf{Var}[f] \cdot \delta(T)$$