LECTURE 4

Large Deviations and Nonprobabilistic Algorithms

Large deviation. Let us start with a beautiful proof of a result used in Lecture 1.

THEOREM. *Let* $S_n = X_1 + \cdots + X_n$ *where*

$$\Pr[X_i = +1] = \Pr[X_i = -1] = \tfrac{1}{2}$$

and the X_i *are mutually independent. Then for any* $\lambda > 0$

$$\Pr[S_n > \lambda] < e^{-\lambda^2/2n}.$$

Proof. For any $\alpha > 0$

$$E[e^{\alpha X_i}] = \tfrac{1}{2}[e^{\alpha} + e^{-\alpha}] = \cosh(\alpha) \le e^{\alpha^2/2}.$$

(The inequality can be shown by comparing the Taylor Series.) As the X_i are mutually independent,

$$E[e^{\alpha S_n}] = E\left[\prod_{i=1}^{n} e^{\alpha X_i}\right] = \prod_{i=1}^{n} E[e^{\alpha X_i}] < \prod_{i=1}^{n} e^{\alpha^2/2} = e^{\alpha^2 n/2}.$$

Thus

$$\Pr[S_n > \lambda] = \Pr[e^{\alpha S_n} > e^{\alpha\lambda}] \le E[e^{\alpha S_n}]\, e^{-\alpha\lambda} < e^{\alpha^2 n/2 - \alpha\lambda}.$$

We now choose $\alpha = \lambda/n$, optimizing the inequality

$$\Pr[S_n > \lambda] < e^{-\lambda^2/2n}. \qquad \square$$

More generally, let $Y_1, \cdots, Y_n$ be independent with

$$\Pr[Y_i = 1] = p_i, \qquad \Pr[Y_i = 0] = 1 - p_i$$

and normalize by setting $X_i = Y_i - p_i$. Set $p = (p_1 + \cdots + p_n)/n$ and $X = X_1 + \cdots + X_n$. We give the following without proof:

$$\Pr[X > a] < e^{-2a^2/n},$$

$$\Pr[X < -a] < e^{-a^2/2pn},$$

$$\Pr[X > a] < e^{-a^2/2pn + a^3/2(pn)^3}.$$

The last two bounds are useful when $p \ll 1$. The $a^3/2(pn)^3$ term is usually small in applications. When all $p_i = p$, $X = B(n, p) - np$ is roughly normal with zero mean and variance $np(1-p) \sim np$, explaining somewhat the latter two inequalities.

Discrepancy. Let $\mathcal{A} \subseteq 2^\Omega$ be an arbitrary family of finite sets. Let $\chi : \Omega \to \{+1, -1\}$ be a two-coloring of the underlying points. Define

$$\chi(A) = \sum_{a \in A} \chi(a),$$

$$\text{disc}\,(\chi) = \max_{A \in \mathcal{A}} |\chi(A)|,$$

$$\text{disc}\,(\mathcal{A}) = \min_{\chi} \text{disc}\,(\chi).$$

Note that $\text{disc}\,(\chi) \le K$ means that there is a two-coloring of Ω so that every $A \in \mathcal{A}$ has $|\chi(A)| \le K$.

THEOREM. *If* $|\mathcal{A}| = |\Omega| = n$, *then*

$$\text{disc}\,(\mathcal{A}) \le \sqrt{2n \ln (2n)}.$$

Proof. With χ random and $|A| = r$, $\chi(A)$ has distribution S_r. As all $A \subseteq \Omega$ have $|A| \le |\Omega| = n$

$$\Pr[|\chi(A)| > \lambda] < 2\, e^{-\lambda^2/2n}.$$

Thus

$$\Pr[\text{disc}\,(\chi) > \lambda] < \sum_{A \in \mathcal{A}} \Pr[|\chi(A)| > \lambda] < 2n\, e^{-\lambda^2/2n} = 1$$

by taking $\lambda = \sqrt{2n \ln (2n)}$. Thus $\Pr[\text{disc}\,(\chi) \le \lambda] > 0$ and so there exists a χ with $\text{disc}\,(\chi) \le \lambda$.

We can also express this result in vector form.

THEOREM. *Let* $u_j \in R^n$, $|u_j|_\infty \le 1$, $1 \le j \le n$. *Then there exist* $\varepsilon_j \in \{-1, +1\}$ *so that, setting* $u = \varepsilon_1 u_1 + \cdots + \varepsilon_n u_n$, $|u|_\infty \le \sqrt{2n \ln (2n)}$. (Note: With $u = (L_1, \cdots, L_n)$, $|u|_\infty = \max |L_i|$, the L^∞ norm.)

Here is the translation between the formulations. Given $\mathcal{A} \subseteq 2^\Omega$, number the elements $1, \cdots, n$ and the sets $A_1, \cdots, A_n$ and define the incidence matrix $A = [a_{ij}]$ by $a_{ij} = 1$ if $j \in A_i$; 0 otherwise. Let u_j be the column vectors. A two-coloring $\chi : \Omega \to \{+1, -1\}$ corresponds to $\varepsilon_j = \chi(j)$, $\text{disc}\,(A_i)$ to the ith coordinate L_i of u and $\text{disc}\,(\chi)$ to $|u|_\infty$. Now, however, we allow $a_{ij} \in [-1, +1]$ to be arbitrary. The proof of our first theorem can be easily modified to show that with ε_j random $\Pr[|L_i| > \lambda] < 2\, e^{-\lambda^2/2n}$ and the rest of the proof follows as before.

$$\begin{array}{c} \\ S_1 \\ \vdots \\ S_n \\ \\ \\ \end{array}\!\!\begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & n \end{array} \\ \left[\begin{array}{cccc} & & & \\ & & & \\ & & & \end{array}\right] \\ \begin{array}{cccc} u_1 & & & u_n \\ + & - & & - \end{array} \end{array} \qquad \begin{array}{l} L_{11} = \chi(S_1) \\ \quad \vdots \\ L_{1n} = \chi(S_n) \\ \\ u = \pm u_1 \pm \cdots \pm u_n = (L_{11}, \cdots, L_{1n}) \end{array}$$

LECTURE 10

Six Standard Deviations Suffice

I suppose every mathematician has a result he is most pleased with. This is mine.

THEOREM. *Let* $S_1, \cdots, S_n \subset [n]$. *Then there exists* $\chi: [n] \to \{-1, +1\}$ *such that*

$$|\chi(S_i)| < 6n^{1/2}$$

for all i, $1 \le i \le n$.

The elementary methods of Lecture 1 already give χ with all $|\chi(S_i)| < cn^{1/2}(\ln n)^{1/2}$. From our vantage point $n^{1/2}$ is one standard deviation. With χ random $|\chi(S)| > 6n^{1/2}$ occurs with a small but positive probability $\varepsilon < e^{-6^2/2} = e^{-18}$. There are n sets, so the expected number of i with $|\chi(S_i)| > 6n^{1/2}$ is εn, which goes to infinity with n. A random χ will not work; the key is to meld probabilistic ideas with the Pigeonhole Principle.

The constant 6 is simply the result of calculation, the essential point is that it is an absolute constant. In the original paper it is 5.32, in our proof here "6" = 11.

Proof. Let C be the set of $\chi: [n] \to \{-1, +1\}$. Call $\chi \in C$ *realistic* if

(1) $|\chi(S_i)| > 10n^{1/2}$ for at most $4(2e^{-50})n$ i's,
(2) $|\chi(S_i)| > 30n^{1/2}$ for at most $8(2e^{-450})n$ i's,
(3) $|\chi(S_i)| > 50n^{1/2}$ for at most $16(2e^{-1250})n$ i's,

and, in general,

(s) $|\chi(S_i)| > 10(2s-1)n^{1/2}$ for at most $2^{s+1}(2e^{-50(2s-1)^2})n$ i's.

CLAIM. At least half the $\chi \in C$ are realistic.

Pick $\chi \in C$ at random. Let Y_i be the indicator random variable for $|\chi(S_i)| > 10n^{1/2}$. Set $Y = \sum_{i=1}^{n} Y_i$. Then

$$E[Y_i] = \Pr[|\chi(S_i)| > 10n^{1/2}] < 2e^{-50}$$

by the bound of Lecture 4. By linearity of expectation

$$E[Y] = \sum_{i=1}^{n} E[Y_i] = (2e^{-50})n.$$

(We do not know much about the distribution of Y since the dependence of the Y_i may be complex, reflecting the intersection pattern of the S_i. Fortunately,

linearity of expectation ignores dependency.) As $Y \geq 0$

$$\Pr[Y > 4E[Y]] < \tfrac{1}{4}.$$

That is,

$$\Pr[\chi \text{ fails (1)}] < \tfrac{1}{4}.$$

Apply the same argument to (2), letting Y_i be one if $|\chi(S_i)| > 30n^{1/2}$. The $10^2/2 = 50$ becomes $30^2/2 = 450$. Everything is identical except that 4 was changed to 8 so that

$$\Pr[\chi \text{ fails (2)}] < \tfrac{1}{8}.$$

In general,

$$\Pr[\chi \text{ fails (s)}] < 2^{-s-1}.$$

The probability of a disjunction is at most the sum of the probabilities.

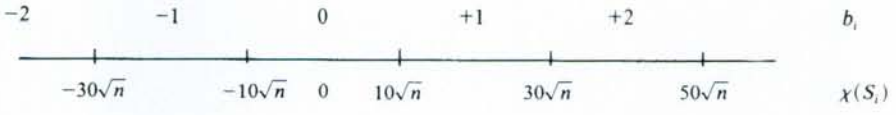
$$\Pr[\chi \text{ not realistic}] \leq \sum_{s=1}^{\infty} 2^{-s-1} = \tfrac{1}{2}.$$

A random χ has probability at least $\frac{1}{2}$ of being realistic. Each χ was given equal weight 2^{-n} in the probability space so at least 2^{n-1} χ are realistic, completing the claim.

Now we define a map

$$T(\chi) = (b_1, \cdots, b_n)$$

where b_i is the nearest integer to $\chi(S_i)/20n^{1/2}$. That is,



$b_i = 0$ means $|\chi(S_i)| \leq 10n^{1/2}$,

$b_i = 0, \pm 1$ means $|\chi(S_i)| < 30n^{1/2}$, etc.

Let B be the set of possible $T(\chi)$ with χ realistic. That is, B is all $(b_1, \cdots, b_n)$ such that

(1) $b_i \neq 0$ for at most $4(2e^{-50})n$ i's,

(2) $b_i \neq 0, \pm 1$ for at most $8(2e^{-450})n$ i's,

and, in general,

(s) $|b_i| \geq s$ for at most $2^{s+1}(2e^{-50(2s-1)^2})n$ i's.

(Let us pause for a geometric view. We are actually showing that if $A = [a_{ij}]$ is an $n \times n$ matrix with all $|a_{ij}| \leq 1$, then there exists $x \in C = \{-1, +1\}^n$ so that $|Ax|_\infty \leq Kn^{1/2}$. We consider $A: R^n \to R^n$ given by $x \to Ax$. We split the range into

cubes of size $20n^{1/2}$, centered about the origin. We want to prove that some $x \in C$ is mapped into the central cube. We cannot do this directly, but we plan to find $x, y \in C$ mapped into the same cube and examine $(x-y)/2$. We want to employ the Pigeonhole Principle, with the 2^n $x \in C$ mapped into the cubes. To do this we restrict attention to the 2^{n-1} realistic x as they are mapped (shown by the next claim) into a much smaller number of cubes.)

CLAIM. $|B| < (1.0000000000000001)^n$.

We use the inequality, valid for all n, all $a \in [0, 1]$,

$$\sum_{i<na} \binom{n}{i} < 2^{nH(a)}$$

where $H(a) = -a \lg a - (1-a) \lg (1-a)$ is the entropy function. We can choose $\{i: b_i \neq 0\}$ in at most $2^{nH(8e^{-50})}$ ways. Then we can choose the signs of the nonzero b_i in at most $2^{8e^{-50}n}$ ways. For each s there are at most $2^{**}[nH(2^{s+1}e^{-50(2s+1)^2})]$ ways to choose $\{i: |b_i| > s\}$. These choices determine $(b_1, \cdots, b_n)$. Thus $|B| < 2^{\beta n}$ where

$$\beta = 8e^{-50} + H(8e^{-50}) + H(16e^{-450}) + H(32e^{-1250}) + \cdots.$$

This series clearly converges and the claim follows from a calculation.

Now apply the Pigeonhole Principle to the map T from the (at least) 2^{n-1} realistic χ to the (at most) $(1+10^{-16})^n$ pigeonholes B. There is a set C' of at least $2^{n-1}/(1+10^{-16})^n$ realistic χ mapping into the same $(b_1, \cdots, b_n)$.

Let us think of C as the Hamming Cube $\{-1, +1\}^n$ with the metric

$$\rho(\chi_1, \chi_2) = |\{i: \chi_1(i) \neq \chi_2(i)\}|.$$

D. Kleitman has proved that if $C' \subset C$ and $C' \geq \sum_{i \leq r} \binom{n}{i}$ with $r \leq n$ then C' has diameter at least $2r$. That is, the set of a given size with minimal diameter is the ball. In our case $|C'| > 2^{n-1}/(1+10^{-16})^n$ so we may take $r = \frac{1}{2}n(1-10^{-6})$ with room to spare and C' has diameter at least $n(1-10^{-6})$. Let $\chi_1, \chi_2 \in C'$ be at maximal distance. (Kleitman's Theorem is not really necessary. The elementary argument at the end of Lecture 9 gives that C' has diameter at least $.4999n$. We could use this value and finish the proof with a much worse, but still absolute constant, value of "6".) Now set

$$\chi = (\chi_1 - \chi_2)/2;$$

then χ is a partial coloring of $[n]$. As $T(\chi_1) = T(\chi_2)$ both $\chi_1(S_i)$ and $\chi_2(S_i)$ lie in a common interval $[(20b_i - 10)n^{1/2}, (20b_i + 10)n^{1/2}]$. Then

$$(*) \qquad |\chi(S_i)| = |(\chi_1(S_i) - \chi_2(S_i))/2| \leq 10n^{1/2}.$$

Also

$$(**) \qquad |\{i: \chi(i) \neq 0\}| < 10^{-6}n.$$

Now iterate. We now have n sets on $10^{-6}n$ points. If we had only $10^{-6}n$ sets we could partially color all but a millionth of the points, giving all sets discrepancy of at most $10(10^{-6}n)^{1/2} = .01n^{1/2}$. Things are not quite so simple as we still have n sets. We actually need the following result: Given n sets on r points, $r \leq n$,

there is a partial coloring of all but at most a millionth of the points so that all

$$|\chi(S)| < 10r^{1/2}[\ln(2n/r)]^{1/2}.$$

The argument is basically that given when $r = n$ but the formulae are a bit more fierce. Let us assume the result (read the original paper!) and now iterate. On the second iteration

$$|\chi(S)| < 10(n10^{-6})^{1/2}[\ln(2\times 10^6)]^{1/2} < .4n.$$

The future terms decrease even faster. The logarithmic term, while annoying, does not affect the convergence. At the end, with all points colored,

$$\begin{aligned}|\chi(S)| &\le 10n^{1/2} + 10(n10^{-6})^{1/2}[\ln(2\times 10^6)]^{1/2} \\ &\quad + 10(n10^{-12})^{1/2}[\ln(2\times 10^{12})]^{1/2} \\ &\quad + \cdots \\ &\le 11n^{1/2},\end{aligned}$$

completing the proof for "6" = 11.

From the reductions of Lecture 5 we derive the following.

COROLLARY. $\operatorname{disc}(\mathcal{A}) \le K|\mathcal{A}|^{1/2}$. *That is, given any n finite sets there is a two-coloring of the underlying points so that each set has discrepancy at most* $Kn^{1/2}$.

This result is best possible up to the constant. Here are two proofs. First, take an $n \times n$ Hadamard Matrix $H = (h_{ij})$ with the first row all ones. Set $A = (a_{ij}) = (H + J)/2$ so that $a_{ij} = 1$ when $h_{ij} = 1$ and $a_{ij} = 0$ when $h_{ij} = -1$. Let $\vec{1} = v_1, v_2, \cdots, v_n$ be the columns of H and $\vec{1} = w_1, w_2, \cdots, w_n$ be the columns of A, so that $w_i = (v_i + \vec{1})/2$. For any choice of signs

$$u = \pm w_1 \pm w_2 \pm \cdots \pm w_n = \tfrac{1}{2}v + s\vec{1}$$

where $v = \pm v_2 \pm \cdots \pm v_n$. As the v_i are orthogonal and $|v_i|_2 = n^{1/2}$, $|v|_2 = [n(n-1)]^{1/2}$. Also $v \cdot \vec{1} = 0$ so $|u|_2 \ge \frac{1}{2}|v|_2 = [n(n-1)]^{1/2}/2$ and thus

$$|u|_\infty \ge (n-1)^{1/2}/2.$$

The second proof involves turning the probabilistic method on its head. Let $T_1, \cdots, T_n$ be randomly chosen subsets of $[n]$. That is, for all i, j $\Pr[j \in T_i] = \frac{1}{2}$ and these events are mutually independent. Let $\chi: [n] \to \{-1, +1\}$ be arbitrary but fixed. Let $P = \{j: \chi(j) = +1\}$, $N = \{j: \chi(j) = -1\}$, $a = |P|$ so $n - a = |N|$. Then $|T_i \cap P|$ has binomial distribution $B(a, \frac{1}{2})$ while $|T_i \cap N|$ has $B(n-a, \frac{1}{2})$ and thus $\chi(T_i)$ has distribution $B(a, \frac{1}{2}) - B(n-a, \frac{1}{2})$. When $a = n/2$ $\chi(T_i)$ is roughly normal with zero mean and standard deviation $\frac{1}{2}n^{1/2}$. Then

$$\lim_n \Pr[|\chi(T_i)| \le \tfrac{1}{2}c\sqrt{n}] = \int_{-c}^{+c} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

One can show that this probability is maximized essentially when $a = n/2$. Pick $c \sim .67$ so that the above integral is .5. Decrease c slightly so that the inequality is strict:

$$\Pr[|\chi(T_i)| < .33\sqrt{n}] < .499.$$

Now since the T_i are chosen independently the events $|\chi(T_i)| > .33n^{1/2}$ are mutually independent so that

$$\Pr[|\chi(T_i)| < .33n^{1/2}, 1 \le i \le n] < .499^n.$$

Let Y_χ be one if $|\chi(T_i)| < .33n^{1/2}$ for all i and let

$$Y = \sum Y_\chi$$

the sum over all 2^n colorings χ. Then

$$E[Y] = \sum_\chi E[Y_\chi] < 2^n(.499)^n \ll 1.$$

Thus the event $Y = 0$ has positive probability, actually probability nearly one. There is a point in the probability space (i.e. actual sets $T_1, \cdots, T_n$) so that $Y = 0$ which, when we unravel the definitions, means that the family $\mathcal{A} = \{T_1, \cdots, T_n\}$ has disc $(\mathcal{A}) > .33n^{1/2}$.

Let us restate our basic theorem of this Lecture in vector form.

THEOREM. *Let $u_j \in R^n$, $1 \le j \le n$, $|u_j|_\infty \le 1$. Then for some choice of signs*

$$|\pm u_1 \pm \cdots \pm u_n|_\infty \le Kn^{1/2}.$$

To prove this we set $u = (L_1, \cdots, L_n) = \pm u_1 \pm \cdots \pm u_n$. Then each L_i has distribution $L_i = \pm a_{i1} \pm \cdots \pm a_{in}$ with all $|a_{ij}| \le 1$. From the arguments of Lecture 4, $\Pr[|L_i| > 10n^{1/2}] < e^{-50}$, etc. and the proof goes through as before.

The methods of Lecture 5 also allow us to re-express our result in terms of simultaneous approximation. Given data a_{ij}, $1 \le i \le m$, $1 \le j \le n$ with all $a_{ij} \le 1$. Given initial values x_j, $1 \le j \le n$. A simultaneous round-off is a set of integers y_j, each y_j either the "round-up" or "round-down" of x_j. Let E_i be the error

$$E_i = \sum_{j=1}^{n} a_{ij}(x_j - y_j)$$

and E the maximal error, $E = \max |E_i|$.

COROLLARY. *There exists a simultaneous round-off with $E \le Km^{1/2}$.*

Is there a polynomial time algorithm that gives the simultaneous round-off guaranteed by this corollary? Given $u_1, \cdots, u_n \in R^n$ with all $|u_j|_\infty \le 1$, is there a polynomial time algorithm to find signs such that $|\pm u_1 \pm \cdots \pm u_n|_\infty < Kn^{1/2}$? The difficulties in converting these theorems to algorithms go back to the basic theorem of this Lecture and lie, I feel, in the use of the Pigeonhole Principle. In Lecture 4 we saw that there is a rapid algorithm giving $|\pm u_1 \pm \cdots \pm u_n|_\infty < cn^{1/2}(\ln n)^{1/2}$. We also saw that no nonanticipative algorithm could do better. That is, a better algorithm could not determine the sign of u_j simply on the basis of $u_1, \cdots, u_{j-1}, u_j$ but would have to look ahead. Also we can show, standing the probabilistic method on its head, that there are $u_1, \cdots, u_n$ so that the number of choices of signs with $|\pm u_1 \pm \cdots \pm u_n|_\infty < Kn^{1/2}$ is less than $(2-c)^n$ of the 2^n possible choices. Hence a randomly selected choice of signs will not work. Let us rephrase back into the language of sets and conclude the Lectures with the following problem.

Open Problem. Is there a polynomial time algorithm which, given input $S_1, \cdots, S_n \subset [n]$, outputs a two-coloring $\chi: [n] \to \{-1, +1\}$ such that

$$|\chi(S_i)| \leq Kn^{1/2}$$

for *all* i, $1 \leq i \leq n$?

REFERENCE

J. SPENCER, *Six standard deviations suffice*, Trans. Amer. Math. Soc., 289 (1985), pp. 679–706.