The Probabilistic Method, Contd.

Recall, we started our probability discussion with the following problem on Error Correcting Codes:

Qn: Does there exist error correcting codes $(\mathcal{E}, \mathcal{D})$ s.t. $\mathcal{E} : \{0,1\}^n \to \{0,1\}^\ell$,
$\mathcal{D} : \{0,1\}^\ell \to \{0,1\}^n$,
s.t. it can correct errors introduced by an $(n/4)$-noisy channel, and $\ell = \Theta(n)$?

Equivalently, does there exist ECCs $(\mathcal{E}, \mathcal{D})$ s.t. $\mathcal{E} : \{0,1\}^n \to \{0,1\}^\ell$, $\ell = \Theta(n)$, and for all $x, x' \in \{0,1\}^n$, $x \neq x'$,

$$\underbrace{\text{Dist}(\mathcal{E}(x), \mathcal{E}(x'))}_{\text{number of bits where } \mathcal{E}(x) \text{ and } \mathcal{E}(x') \text{ differ}} > n/2 + 1$$

Thm: There exists $\mathcal{E} : \{0,1\}^n \to \{0,1\}^{100n}$ s.t. for all $x \neq x'$, $\text{Dist}(\mathcal{E}(x), \mathcal{E}(x')) > n/2 + 1$.

Proof: via probabilistic method.

Consider a uniformly random fn.

$$f : \{0,1\}^n \to \{0,1\}^{100n}.$$

We will show that

$$(*) \quad \Pr\left[\exists\, x \neq x' \text{ s.t. } \mathrm{Dist}\left(f(x), f(x')\right) \leq n/2\right] < 1.$$

As a result, since the above prob. is strictly less than 1, there exists some fn. $f : \{0,1\}^n \to \{0,1\}^{100n}$ s.t. for all $x \neq x'$, $f(x)$ and $f(x')$ are at least $(n/2 + 1)$ far apart, and therefore errors in at most $n/4$ locations can be corrected.

Using union bound, we get

$$\Pr\left[\exists\, x \neq x' \quad \mathrm{Dist}\left(f(x), f(x')\right) \leq n/2\right]$$

$$\leq \sum_{x \neq x'} \Pr\left[\mathrm{Dist}\left(f(x), f(x')\right) \leq n/2\right]$$

There are $\binom{2^n}{2}$ pairs of distinct x, x' in $\{0,1\}^n$. Therefore, it suffices to show that for all $x \neq x'$,

$$\Pr\left[\mathrm{Dist}\left(f(x), f(x')\right) \leq n/2\right] < 1/\binom{2^n}{2}$$

Fix any x, x', and let $\rho = \Pr\left[\text{Dist}(f(x), f(x')) \leq n/2\right]$.

How to prove that $\rho < \frac{1}{2^{2n}}$?

Approach 1: compute ρ.

$$\rho = \sum_{t=0}^{n/2} \binom{100n}{t} 2^{-100n}$$

This sum is a bit ugly to compute, but not too complicated.

$$2^{-100n} \sum_{t=0}^{n/2} \binom{100n}{t} < 2^{-100n} \cdot \frac{n}{2} \cdot \binom{100n}{n/2}$$

From here, we need some bound on $\binom{100n}{n/2}$.

One option is to use Stirling's approximation for computing $k!$. As you can see, it gets a bit messy.

Approach 2 : Use "concentration inequalities" which show that the random variable sampled is close to the expected value, with high probability.

This is a very useful tool in the analysis of randomized algorithms / processes, and therefore we will spend some time with this approach.

CONCENTRATION INEQUALITIES :

In the last two lectures, we discussed random variables and their expected value. The expected value is a good starting point to understand a random variable. For instance, if X denotes the running time of a randomized algorithm, and you show that $E[X] = O(n)$, then it gives you some indication of the 'average running time' of the algorithm. However, are you guaranteed that the running time will be close to $O(n)$? Are you even guaranteed that the running time is finite?

For instance, consider an algorithm that terminates in exactly $n \cdot 2^{i/2}$ steps, with prob. $1/2^i$, $i \geq 1$.

$$E[\text{running time}] = \sum_{i=1}^{\infty} \frac{(n \cdot 2^{i/2})}{2^i}$$

$$= n \cdot \left(\sum_{i=1}^{\infty} 1/2^{i/2} \right)$$

$$= \frac{n}{\sqrt{2}-1}$$

But the algorithm isn't even guaranteed to terminate.

a. What guarantees can we get from the expected value of a random variable?

b. If we have more information about the random variable, can we get better guarantees?

For (a), the most basic result is called Markov's inequality.

Markov's inequality :

X : any non-neg. valued random variable

$$Pr[X \geq a] \leq \frac{E[X]}{a}.$$

- Applicable only if X takes non-neg. values.
- Cannot be used to prove an upper bound on $Pr[X \leq b]$

Let us go back to our error correcting codes problem.

Let X be a r.v. denoting $Dist(f(x), f(x'))$ where $f : \{0,1\}^n \to \{0,1\}^{100n}$ is unif. rand.

$$E[X] = ?$$

Let $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ bit of } f(x) \text{ and } f(x') \text{ differ} \\ 0 & \text{otherwise} \end{cases}$

$$X = \sum_{i=1}^{100n} X_i$$

$E[X_i] = 1/2$, and therefore $E[X] = 50n$.

Therefore, we want to know

$$\Pr\left[X \leq n/2 \right],$$

given $E[X] = 50n$.

However, Markov's inequality cannot be used for showing a bound on $\Pr[X \leq \dots]$.

What if we have some more info about the random variable?

Similar to expectation of a r.v., another useful quantity is the variance of a random variable, denoted by $\mathrm{Var}[X]$.

Def: $\mathrm{Var}[X] = E[(X - E[X])^2]$

Using the variance of a random variable, we can get a better concentration bound.

Chebyshev's Inequality:

For any r.v. with finite expectation $\mu = E[X]$ and finite variance $\text{Var}[X] = \sigma^2$,

$$Pr\left[|X - \mu| \geq k \right] \leq \frac{\sigma^2}{k^2}$$

Chebyshev's inequality is just Markov's inequality, but in terms of $\text{Var}[X]$.

Let us see if Chebyshev's ineq. is sufficient for our ECC problem. Recall, $\mu = E[X] = 50n$.

$$Pr\left[X \leq n/2 \right] \leq Pr\left[|X - \mu| \geq 49.5n \right]$$

$$\leq \frac{\text{Var}[X]}{(49.5n)^2}$$

To compute $\text{Var}[X]$, note that

$$X = X_1 + X_2 + \dots + X_{100n},$$

where all the random variables are pairwise - independent.

Def: Random variables Y, Z defined over (Ω, p) are independent if for all possible $a \in Supp(Y)$, $b \in Supp(Z)$,

$$Pr[Y = a \wedge Z = b] = Pr[Y = a] \cdot Pr[Z = b].$$

If Y and Z are independent r.v.s, then $E[Y \cdot Z] = E[Y] \cdot E[Z]$.

As a result, $Var[Y + Z] = Var[Y] + Var[Z]$.

Hence, in our ECC problem,

$$Var[X] = \sum_{i=1}^{100n} Var[X_i] = (100n)(1/2 - 1/4)$$
$$= 25n.$$

$$\therefore Pr[X \leq n/2] \leq \frac{(25n)}{(49.5n)^2}$$

This bound is not good enough, since we want to show that

$$\binom{2^{2n}}{2} \cdot \Pr\left[X \leq n/2 \right] < 1.$$

This brings us to the main focus of this lecture : Chernoff bounds.

Chernoff Bounds (Simplified Version) :

Let $X = X_1 + X_2 + \dots + X_t$, where

- each $X_i : \Omega \to \{0, 1\}$
- $\Pr[X_i = 1] = p_i$
- all X_is are independent

$$\mu = E[X] = \sum_{i=1}^{t} p_i$$

For any $\delta \in (0, 1)$,

$$\Pr\left[X \leq (1-\delta)\mu \right] < e^{-\mu\delta^2/2}$$

For any $\delta \geq 0$,

$$\Pr\left[X \geq (1+\delta)\mu \right] < e^{-\mu\delta^2/(2+\delta)}$$

You will prove this in Tutorial 7. The proof involves Markov's inequality, plus high school mathematics.

Let us use Chernoff bounds for our ECC problem.

$$\Pr\left[X \leq n/2\right] = \Pr\left[X \leq \left(1 - \underbrace{\frac{49.5}{50}}_{\delta}\right)\mu\right]$$

$$< e^{-25n\left(\frac{49.5}{50}\right)^2}$$

$$< e^{-25n\cdot\left(\frac{4}{5}\right)^2} = e^{-16n}$$

$$\therefore \Pr\left[\exists\, x, x' \text{ s.t. } \text{Dist}(f(x), f(x')) \leq n/2\right]$$

$$\leq \binom{2^n}{2} \cdot e^{-16n} < \frac{2^{2n}}{e^{16n}} \ll 1.$$

This probability is very small. As a result, most functions mapping n bits to $100n$ bits are good ECCs. However, giving an explicit const. is very challenging, and finding explicit optimal ECCs is an active area of research.

Chernoff bounds are extremely useful in the analysis of randomized algorithms/processes. In this lecture and the next, we will see a few applications.

Amplifying error gap in randomized algorithms

Consider a (Yes, No) problem like primality testing. You are given an algo. A s.t.

- For all Yes instances x, $\Pr[A(x) \to \text{Yes}] = 1$
- For all No instances x, $\Pr[A(x) \to \text{No}] \geq 1/2$.

Want: an algorithm B s.t.

- For all Yes instances x, $\Pr[B(x) \to \text{Yes}] \geq 1 - 1/2^n$
- For all No instances x, $\Pr[B(x) \to \text{No}] \geq 1 - 1/2^n$.

$B(x)$:

For $i = 1$ to n

if $A(x) = \text{No}$, then output No

Output Yes.

In each run of A, you must run it with fresh randomness

- For all Yes instances x, $\Pr[B(x) \to \text{Yes}] = 1$
- For all No instances x, $\Pr[B(x) \to \text{No}] \geq 1 - 1/2^n$.

Same problem, but now A can make mistakes for both Yes and No instances.

Consider a (Yes, No) problem like primality testing. You are given an algo. A s.t.

- For all Yes instances x, $\Pr[A(x) \to \text{Yes}] \geq 2/3$
- For all No instances x, $\Pr[A(x) \to \text{No}] \geq 2/3$

Want: an algorithm B s.t.

- For all Yes instances x, $\Pr[B(x) \to \text{Yes}] \geq 1 - 1/2^n$
- For all No instances x, $\Pr[B(x) \to \text{No}] \geq 1 - 1/2^n$.

$B(x)$:

For $i = 1$ to $t = \text{poly}(n)$

set $z_i = A(x)$

Output Majority $(\{z_1, \ldots, z_t\})$.

In each run of A, you must run it with fresh randomness

Analysis of B :

Suppose x is a 'yes' instance.

To show : $Pr[B(x) \to \text{'no'}] \leq 1/2^n$.

X : number of iterations where $A(x) \to$ 'no'.

$$X_i = \begin{cases} 1 & \text{if, in } i^{th} \text{ iteration, } A(x) \to \text{'no'} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{i=1}^{t} X_i \qquad E[X] = \sum_{i=1}^{t} E[X_i]$$

$$E[X_i] = Pr[A(x) \to \text{'no'}] \leq 1/3$$

x is a 'yes' instance

$Pr[A(x) \to \text{'yes'}] \geq 2/3$

$$\mu = E[X] \leq t/3$$

$$Pr[B(x) \to \text{'no'}] = Pr[X > t/2]$$

$$\leq Pr[X > \mu(1 + \underbrace{t/6\mu}_{\delta})]$$

$Pr[X > t/2] = Pr[X > t/3 + t/6] \leq Pr[X > \mu + t/6]$

$= Pr[X > \mu(1 + t/6\mu)]$

$$< e^{-\mu\delta^2/2+\delta}$$

$$\frac{\mu\delta^2}{2+\delta} = \frac{t^2}{72\mu + 6t} \geq \frac{t^2}{72(t/3) + 6t} = \frac{t}{30}$$

$$\Rightarrow \quad -\mu\delta^2/2+\delta \leq -t/30$$

$$\therefore \quad Pr[B(x) \rightsquigarrow \text{'no'}] \leq e^{-\mu\delta^2/2+\delta} \leq e^{-t/30}$$

Set $t = 30n$, we get

$$Pr[B(x) \text{ gives wrong output}] \leq 1/e^n < 1/2^n$$

A similar analysis for the case where x is a 'no' instance but B outputs 'yes'.

■