CSG399: Gems of Theoretical Computer Science. Lecture 1. Jan. 9, 2009.
Instructor: Emanuele Viola Scribe: Laura Poplawski

Preliminaries: probability, correlation, circuits, and generators

1 Preliminaries on Probability

In this class we will need to do some counting. Probability is "just" a convenient language for counting, and in this section we review some basics about it.

To start, we have an experiment in mind. For example, tossing a die. The *sample space* (S) is the set of possible outcomes of the experiment. For example, in the tossing a die experiment, the sample space $S = \{1, 2, 3, 4, 5, 6\}$.

A *probability distribution* is a map from $S \to \mathbb{R}^+$ (so $\forall s \in S$, $\Pr[s] \geq 0$) such that $\sum_{s \in S} Pr[s] = 1$. An *event* is a subset $E \subseteq S$.

For example, the event "the die is even" is given by $E = \{2, 4, 6\}$.

The probability of an event is defined as

$$Pr[E] = \sum_{s \in E} Pr[s].$$

Fact 1 (Union bound)**.** $\forall E_1, E_2$, $Pr[E_1 \cup E_2] \leq Pr[E_1] + Pr[E_2]$. *If* $E_1 \cap E_2 = \emptyset$*, then this is tight equality.*

Two events E_1 and E_2 are *independent* if $Pr[E_1 \cap E_2] = Pr[E_1] \cdot Pr[E_2]$. *Conditional Probability*: if E_1 and E_2 are events (and $Pr[E_2] > 0$), then the probability of E_1 conditional to E_2 is given by

$$Pr[E_1|E_2] = \frac{Pr[E_1 \cap E_2]}{Pr[E_2]}$$

For example, in the die tossing experiment, if $E_1 = \{2\}$ and $E_2 = \{2, 4, 6\}$, then $Pr[E_1|E_2] = \frac{1/6}{3/6} = \frac{1}{3}$.

A useful way to define events if via the concept of a *random variable*, which is a mappings $X : S \to \mathbb{R}$. If X is a random variable, we have the expected value of X: $E[X] := \sum_{a \in \mathbb{R}} a \cdot Pr[X = a]$. For random variable X and event A, $X|A$ means X conditioned to event A. We again require that $Pr[A] > 0$.

$$Pr[(X|A) = a] = Pr[X = a|A] = \frac{Pr[X = a \cap A]}{Pr[A]}$$

Similary, $E[X|A]$ is the expectation of random variable X conditioned to event A.

1.1 Probabilistic Method

Suppose that your task it prove that there is an object x that satisfies property A. One way to do that is to prove that $Pr_x[x \in A] > 0$, since this implies that $\exists \bar{x}$ such that $\bar{x} \in A$.

The following is a slightly more involved example of the probabilistic method, which we will use often.

Claim 1. *Let X, Y be independent random variables, and f a function $f : X, Y \to \mathbb{R}$. If $E_{X,Y}[f(X,Y)] \geq \epsilon$, then $\exists$ fixed $\bar{x}$ such that $E_Y[f(\bar{x},Y)] \geq \epsilon$.*

Proof. Suppose for the sake of contradiction that $\forall \bar{x}$, $E_Y[f(\bar{x},y)] < \epsilon$. Then,

$$E_{X,Y}[f(X,Y)] = \sum_{\bar{x}} E_{X,Y}[f(X,Y)|X = \bar{x}] \cdot Pr[X = \bar{x}].$$

But X and Y are independent, so this is equal to

$$\sum_{\bar{x}} E_Y[f(\bar{x},Y)] \cdot Pr[X = \bar{x}] < \epsilon \cdot \sum_{\bar{x}} Pr[X = \bar{x}] = \epsilon$$

This contradicts our hypothesis on f in the claim. □

2 Correlation

Suppose f and g are boolean functions, $\{0,1\}^n \to \{0,1\}$. Let D be a distribution on $\{0,1\}^n$ (inputs). For example, D may be uniform over $\{0,1\}^n$.

Definition 1 (Correlation, a.k.a. approximability, a.k.a. average-case hardness)**.** *The* correlation *between f and g with respect to D:*

$$COR_D(f,g) := |Pr_{x\sim D}[f(x) = g(x)] - Pr_{x\sim D}[f(x) \neq g(x)]| \in [0,1].$$

Also, if G is a set of functions, $COR_D(f,G) = max_{g\in G} COR_D(f,g)$.

$COR_D(f,G)$ captures how well functions from G can compute f on a random input from D.

Note that the correlation equals $COR_D(f,g) = |1 - 2Pr[f(x) \neq g(x)]| = 2\left|\frac{1}{2} - Pr[f(x) \neq g(x)]\right|$.

We now present a third way to write correlation. This will use the notation $e(z) = (-1)^z$. Note this means that $e(z) = 1$ if z is even, $e(z) = -1$ if z is odd. The third way is the following:

Proposition 2. $COR_D(f,g) = |E_{x\sim D} e[f(x) + g(x)]|$.

Proof.

$$\begin{aligned}
COR_D(f,g) &= |Pr_{x\sim D}[f(x)=g(x)] - Pr_{x\sim D}[f(x)\neq g(x)]| \\
&= |1\cdot Pr_{x\sim D}[f(x)=g(x)] + (-1)\cdot Pr_{x\sim D}[f(x)\neq g(x)]| \\
&= |e(\text{positive number})\cdot Pr_{x\sim D}[f(x)=g(x)] \\
&\quad +e(\text{negative number})\cdot Pr_{x\sim D}[f(x)\neq g(x)]\,| \\
&= |Ee[f(x)+g(x)|f(x)=g(x)]\cdot Pr_{x\sim D}[f(x)=g(x)] \\
&\quad +Ee[f(x)+g(x)|f(x)\neq g(x)]\cdot Pr_{x\sim D}[f(x)\neq g(x)]\,| \\
&= |E_{x\sim D}e[f(x)+g(x)]|
\end{aligned}$$

□

This definition of correlation differs from the definition given in statistics in multiple ways. One is that we are taking absolute values. In fact, the absolute values can be removed because we will be interested in correlations with classes of functions $COR_D(f,G)$: typically, the class of functions G will be closed under complement, which means that $COR_D(f,G)\geq 0$ (exercise).

Note that $f=g \Rightarrow COR_D(f,g)=1$, $\forall D$. Conversely, for any fixed f it can be shown that with high probability over the choice of a random function g, the correlation between f and g is nearly 0 (it would be exactly 0 had we not defined correlation using absolute values).

3 Circuits

A circuit is made up of inputs and gates (AND, OR, and NOT). The *fan-in* is the number of inputs per gate. For now, we will assume unbounded fan-in. We will define the size of a circuit as the number of wires. Notice that the number of wires is at least the number of gates, since each gate must have at least one wire exiting (Otherwise, we could remove the gate without affecting the circuit).

Fact 2. *Any function $f:\{0,1\}^l\to\{0,1\}$ can be computed by a circuit of size $<2\cdot 2^l\cdot l$.*

Proof. Brute force circuit. Note that $f(x)=1$ if and only if $\exists a$ such that $f(a)=1$ and $x=a$. We can write the circuit:

$$\vee_{a\in\{0,1\}^l, f(a)=1}\wedge_{i=1}^{l}(x_i=a_i)$$

In other words, create an AND gate for each input that return 1, computing the AND of all the digits to identify the input. Then, create an OR of all of these inputs that return 1. There will be one wire from each AND gate into the OR gate, which is at most 2^l wires (since there are a total of 2^l inputs, and at most all of these can return 1). There will be at most two wires to get to each AND gate: one from the correct value to the and gate, one more from the input digit to a NOT if necessary. This makes at most $2l$ inputs per or gate, for a total of $2\cdot 2^l\cdot l$ wires. □

Definition 3. *$G : \{0,1\}^s \to \{0,1\}^n$ is a generator that fools circuits of size w (w stands for "wire," since we are measuring circuit sizes as the number of wires) with error ϵ and seed length s if $\forall$ circuits C of size w:*

$$\left|E_{s\in\{0,1\}^s} e[C(G(s))] - E_{u\in uniform} e[C(u)]\right| \leq \epsilon.$$

Note that the inequality in the definition of a generator can be rewritten as follows:

$$\begin{aligned} & |Pr\left[C(G(s)) = 0\right] - Pr\left[C(G(s)) = 1\right] - Pr\left[C(u) = 0\right] + Pr\left[C(u) = 1\right]| \\ = \; & |1 - 2Pr\left[C(G(s)) = 1\right] - 1 + 2Pr\left[C(u) = 1\right]| \\ = \; & 2\,|Pr\left[C(u) = 1\right] - Pr\left[C(G(s)) = 1\right]| \end{aligned}$$

If ϵ is very small, we can essentially ignore the factor 2.

Definition 4. *G is* explicit *if, given $x \in \{0,1\}^s$ and $i \leq n$ (an index to the output), we can compute $G(x)_i$ in (uniform) time* $\mathrm{poly}(s, \log n)$ *(polynomial in the input length, since $\log n$ is the size of the index i).*

Theorem 5 (Non-trivial generators from functions with low correlation)**.** *Let $f : \{0,1\}^l \to \{0,1\}$, $COR_{uniform}(f, \text{circuits of size } w) \leq \epsilon$. Then, $G : \{0,1\}^l \to \{0,1\}^{l+1}$ (defined as $G(x) = x \circ f(x)$, x concatenated with $f(x)$) fools circuits of size $w - O(1)$ with error ϵ.*

The complete proof will be given next lecture, but the basic idea is as follows. If $\exists C$ that distinguishes G from random, then $\exists C'$ that correlates with f. How is C' computing f? Given x, it will guess a random $b = f(x)$. Run $C(x \circ b)$ (Think of $C = 1$ if the input is random, $C = 0$ otherwise.) If $C(x \circ b) = 0$, assume we guessed correctly at $f(x)$ and output b. If $C(x \circ b) = 1$, output $\bar{b}$. This can be written succinctly and conveniently as follows:

On input x, $C'(x)$ tosses a coin b and outputs $C(x \circ b) + b \in \{0,1\}$.

Instead of having C' toss a coin, we can think of the uniform distribution on the two circuits $C(x \circ 0) + 0$ and $C(x \circ 1) + 1$.