Advanced Algorithms: Lecture 6 Notes

Dana Randall Spring 2024

Sarthak Mohanty

Independence and Expectation

We will cover a few fundamental topics in probability as they apply to randomized algorithms: independence and expectation.

Important: A course in probability theory was a formal prerequisite for this class. Thus, although we are reviewing some concepts, we will assume basic knowledge of probability in this course. This set of lecture notes provides a concise review of concepts you should be familiar with.

1. The **expectation** of a function $g(X)$ of a random variable is

$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx.$$

   Can be considered as “average" value. The above definition implies that $\mathbb{E}[g_1(x) + g_2(x)] = \mathbb{E}[g_1(x)] + \mathbb{E}[g_2(x)]$: this property is known as a **linearity of expectations**.

2. We call two events A and B **independent** if $\Pr(A = a, B = b) = \Pr(A = a)\Pr(B = b)$. We call a collection of events $\{X_i\}_i^n$ **k-wise independent** if every subset of k events is independent.

   Events vs random variables? Events are subsets of the real line. All these results hold for events and random variables.

Multiparty Computation

Consider the following problem. There are n students in this course who wish to determine the average score on the midterm. However, none of them want to reveal their own scores. Even if k of the n "friends" are dishonest, then they should still learn no more than the sum of the remaining $n - k$ honest ones (which they can figure out anyway from the answer).

Let's formalize the above statement. Say we have N people, each has a secret S_i, $1 \leq i \leq n$. If $0 \leq S \leq 100$, then $m = 101 \cdot n > \sum_i S_i$. uniformly at random from $\{0, \ldots, m-1\}$.

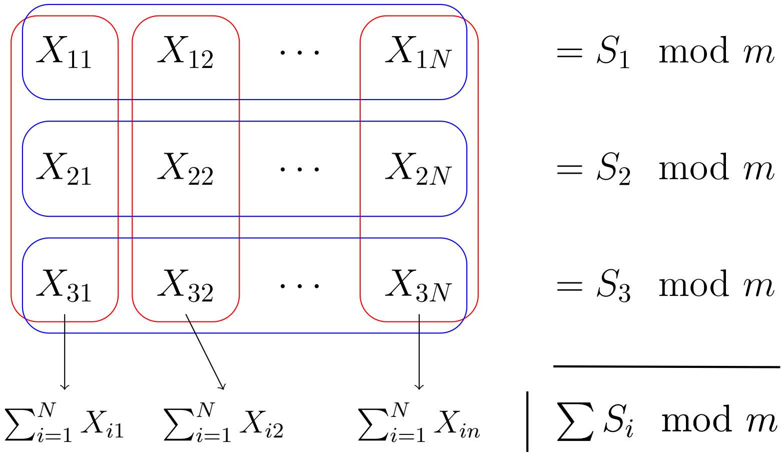
Person i:

- Pick random variables $X_{i1}, \ldots, X_{iN-1}$.

- Sets $X_{iN} = y_i = S_i - (X_{i1} + X_{i2} + \cdots + X_{iN-1}) \pmod m$.
- Distributes the $N-1$ r.vs to the others (one per person) and keeps Y.

Once this process is complete for all players, each player announces his result $(\text{mod } m)$. Then, the sum of each of the announced results is the desired sum S.

An illustration of the protocol is shown below. Hopefully by looking at the values int he table, it is clear that the sum is correct. But, it remains to determine the security of the algorithm.



$$\begin{array}{cccc|l}
X_{11} & X_{12} & \cdots & X_{1N} & = S_1 \mod m \\
X_{21} & X_{22} & \cdots & X_{2N} & = S_2 \mod m \\
X_{31} & X_{32} & \cdots & X_{3N} & = S_3 \mod m \\
\downarrow & \downarrow & & \downarrow & \\
\sum_{i=1}^{N} X_{i1} & \sum_{i=1}^{N} X_{i2} & & \sum_{i=1}^{N} X_{in} & \sum S_i \mod m
\end{array}$$

Figure 1: Diagram of each of the X_is

Some useful facts:

- Say X is a random variable uniformly distributed in $S = \{0, 1, \ldots, m-1\}$. Then, clearly for any fixed a, the variable $Y = X + a \pmod m$ is also uniform in S (though X and Y are dependent).
- If $X_1, \ldots, X_k$ are independent r.v's uniform in S and Y is defined to be $a - (X_1 + ...X_k) \pmod m$ for some fixed a, then these $k+1$ r.v.s are k-wise independent. Just check that $\Pr(Y = y | X_2 = x_2, X_3 = x_3, ..., X_k = x_k) = 1/m$.

Proof for protocol: Claim is secure in following sense: for any set of t honest players, only information that the rest of players have, even if they cheat, is the sum of the t players' values. What does this really mean? What is "information"? What we mean is that if you modify the values for these players keeping the sum the same, then the probability distribution on the sequence of outputs produced by those players doesn't change. Proof: consider the t all sent their Y value to the same honest player. (Doesn't change protocol). For any fixed sequence of coin tosses, all sets of t values with the same sum lead to exactly the same information revealed to the other players. So, all sets with the same sum have the same distribution.

Gambler's Problem

Consider a casino where are games are fair and the expected value of all games is zero, i.e. $\Pr(W) = \Pr(L) = \frac{1}{2}$. Intuitively, we know we cannot game this system. But, it's a good exercise to convert our intuition to a formal proof.

Define X_i to be the gain at time i. So $X = X_1 + \cdots + X_N$ where N is # seconds in a day. Let p_{ij} be the probability that we bet j at time i. So

$$\mathbb{E}[X_i] = \sum_j \frac{j \cdot p_{ij}}{2} - \frac{j \cdot p_{ij}}{2} = 0,$$

and by linearity of expectations, we have

$$\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{N} X_i] = \sum_{i=1}^{N} \mathbb{E}[X_i] = 0$$

Exercise. You are playing a special game of cards against Professor Randall. Each turn, she flips over a card. You are allowed, at any time, to predict the next card will be red. If it is indeed red, you get 1 pt. If it is black, Professor Randall gets one point. Is there a strategy you can use to gain an advantage (in expectation?) Answer is **No.**

Randomized 3-SAT

One fundamental application of randomized algorithms is in MAX-Exact-3-SAT.

Proposition. Given an input of MAX-Exact-3SAT, a random assignment will satisfy at least 87.5% of the clauses in expectation.

Proof. Let S be the random variable denoting the number of satisfiable clauses with a random assignment. Let S_i be the indicator random variable denoting whether or not clause i evaluates to True, for $1 \leq i \leq n$. Define S and S_i as above. Now

$$\begin{aligned}
\mathbb{E}[S_i] &= \Pr\left(\bigcup_{j=1}^{3} (\text{variable } j \text{ evaluates to True})\right) \\
&= 1 - \Pr\left(\text{no variables evaluates to True}\right) \\
&= 1 - \left(\frac{1}{2}\right)^3 \\
&= \frac{7}{8} = .875.
\end{aligned}$$

Using the linearity of expectations, we conclude

$$\mathbb{E}[S] = \sum_{i=1}^{n} E[S_i] = .875n.$$

We know that in expectation