 Hao Zhang nhz@eecs.berkeley.edu

Disclaimer: *These notes have not been subjected to the usual scrutiny for formal publications. They are to be used only for the class.*

Outline:

1. Dart Throwing (for homework 1)
2. Mutual vs. Pairwise Independence
3. Conditional Independence

1 Dart Throwing

First, recall the following useful fact: given n mutually exclusive events $A_1, \ldots, A_n$ such that $\Pr[A_1] + \ldots + \Pr[A_n] = 1$. Then,

$$\Pr[B] = \Pr[B|A_1]\Pr[A_1] + \ldots + \Pr[B|A_n]\Pr[A_n]$$

This is so called the "total probability theorem". Notice that the only criterion on the sets A_i's is that they form a partition of the event space. A simple special case of the theorem is when $n = 2$, so that it reduces to

$$\Pr[B] = \Pr[B|A]\Pr[A] + \Pr[B|\bar{A}]\Pr[\bar{A}]$$

that is, to compute the probability of B, break it up into two cases depending on whether A happens or not, and then add up the conditional probabilities multiplied by the marginal probability.

Now, consider the game of throwing a dart uniform at random (u.a.r.) at a sequence of numbers

$$1, 2, 3, \ldots, i, \ldots, j, \ldots, n-1, n$$

The game terminates when the dart falls into the range $[i, j]$, otherwise, simply throw the dart again. Upon termination, if the dart lands on element i or j, then the game is declared to be a "success", otherwise a "failure".

First, you can see that with probability 1, the game terminates. Why? What's the probability of termination at the k'th trial? Recall geometric distribution, which expresses the probability of seeing the first "head" in a sequence of coin tosses. Let $p = \frac{j-i+1}{n}$, then the game terminates at the k'th trial iff the dart misses the range $[i, j]$ in the first $k-1$ trials and hit the range in the k'th trial. That gives $\Pr[\text{termination at } k\text{'th trial}] = (1-p)^{k-1}p$. Sum up the probability for all $k = 1, 2, \ldots$, and it yields 1 by definition of probability distribution.

So, we can assume that the game always terminates. Then what's the probability of a "success" upon termination? It's easier to answer this question provided that the game terminates at the k'th trial. Because we know that the game terminates, the dart must have landed in the range $[i, j]$ in the last trial. Then, the probability of hitting i or j in the last trial is simply $\frac{2}{j-i+1}$. To formalize this notion of "breaking it up into cases", we invoke the total probability theorem

$$\Pr[\text{success of the game}] = \sum_{k=1}^{\infty} \Pr[\text{success}|\text{termination at round } k]\Pr[\text{termination at round } k]$$

That is,

$$\sum_{k=1}^{\infty} \frac{2}{j-i+1}(1-p)^{k-1} p = \frac{2}{j-i+1}$$

Notice the redundancy in the preceding summation. You are encouraged to think of simpler ways to explain the same result.

Now, armed with this knowledge, we can analyze the RANDOMIZED-FIND-RANK(RandMed) in the same way as what we did with RANDOMIZED-QUICK-SORT(RandQS) in the first lecture. For RandQS, the analysis boils down to computing p_{ij}, the probability of $S_{(i)}$ and $S_{(j)}$ being compared during the execution of the algorithm (where $S_{(i)}$ denotes the i'th smallest element). Think of the sequence of numbers

$$S_{(1)} \ldots S_{(i)} \ldots S_{(j)} \ldots S(n)$$

Every time RandQS selects a pivot randomly, it is equivalent to throw a dart u.a.r. to these numbers and split the set of numbers into more pieces. Notice that $S(i)$ and $S(j)$ have the chance to be compared as long as the pivot (i.e., the dart) doesn't fall in between them. Otherwise, if the pivot happens to be $S(i)$ or $S(j)$, then the two elements are compared and it corresponds to a "success" of the game. If the pivot falls outside the range of $S(i)$ to $S(j)$, then the game continues. Of course, in this case the game won't continue indefinitely. But, no matter what is the probability of termination at the k'th trial, they all have to sum up to 1 and the preceding result of the dart throwing game still holds. It follows that the answer of p_{ij} is the probability of "success" of the game: $\frac{2}{j-i+1}$. The same reasoning can be carried over in analyzing RandMed. The details will show up in the solutions to homework 1.

2 Mutual vs. Pairwise Independence

Recall that a set of n events $A_1, \ldots, A_n$ are said to be *mutually independent* iff for any $I \subseteq \{1, 2, \ldots, n\}$,

$$\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$$

They are said to be *pairwise independent* if the I above is only those of size 2, that is, for any i, j,

$$\Pr[A_i \cap A_j] = \Pr[A_i] \Pr[A_j]$$

Question: By definition, mutual independence is a stronger criterion than pairwise independence. To verify this intuition, can you think of three random variables that are pairwise independent but not mutually independent?

...

Answer: One possible construction is X, Y and $|X - Y|$ where X and Y are independent and $\Pr[X] = \Pr[Y] = \frac{1}{2}$. Clearly, there's only two free variables so the three random variables are not mutually independent. However, pick any two of them, say, X and $|X - Y|$, their joint distribution is the product of the marginal probabilities (**Exercise**: Verify this).

3 Conditional Independence

The following question is adapted from a problem I encountered in research, to illustrate the usefulness of conditional independence. Let $f_X(x)$ denote the probability mass function (pmf), that is $f_X(x) = \Pr[X = x]$; and let $F_X(x)$ denote the cumulative distribution function (cdf), that is $F_X(x) = \Pr[X \leq x]$. Assume three independent random variables X_1, X_2 and X_3 for which the pmf and cdf are known. The goal is to compute the probability that X_1 is the maximum among the three.

One way, by brute force, is to sum up the probability mass on the sample points that satisfy the criterion that X_1 is maximum. This leads to the summation:

$$\begin{aligned} \Pr[X_1 \text{is max}] &= \sum_{i \geq j \wedge i \geq k} \Pr[X_1 = i, X_2 = j, X_3 = k] \\ &= \sum_{i \geq j \wedge i \geq k} \Pr[X_1 = i] \Pr[X_2 = j] \Pr[X_3 = k] \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{i} \sum_{k=-\infty}^{i} f_{X_1}(i) f_{X_2}(j) f_{X_3}(k) \end{aligned}$$

Notice that this is a triple summation. Even if the pmf f has finite support on N points, this is an $O(N^3)$ operation.

On the other hand, notice that when X_1 is fixed, say, $X_1 = c$, then the random variables $X_1 - X_2$ and $X_1 - X_3$ are independent because they are simply $c - X_2$ and $c - X_3$, and X_2 and X_3 are known to be independent. To formalize this notion, we define the concept of conditional independence: Events A and B are said to be *conditionally independent* on C if

$$\Pr[A|C] \Pr[B|C] = \Pr[A \cap B|C]$$

That is to say, when C happens, A and B are independent. Equivalently, $\Pr[A|C] = \Pr[A|B,C]$ (**Exercise**: Verify that this is an equivalent definition)

Now, recall total probability theorem, which enables us to break up analysis into cases. Combine it with conditional independence, we reach the following:

$$\begin{aligned} \Pr[X_1 \text{ is max}] &= \sum_{c=-\infty}^{\infty} \Pr[X_1 = c] \Pr[X_1 \text{ is max} | X_1 = c] \\ &= \sum_{c=-\infty}^{\infty} \Pr[X_1 = c] \Pr[X_2 \leq c \wedge X_3 \leq c] \\ &= \sum_{c=-\infty}^{\infty} \Pr[X_1 = c] \Pr[X_2 \leq c] \Pr[X_3 \leq c] \\ &= \sum_{c=-\infty}^{\infty} f_{X_1}(c) F_{X_2}(c) F_{X_3}(c) \end{aligned}$$

Notice that this is no longer a nested summation and therefore an $O(N)$ operation. **Exercise**: Can you tell in which steps of the derivation above did I use total probability theorem, and in which steps conditional independence?