RSA Revisited

Please read through the following articles from Duo and answer the proceeding questions:

Duo:SHA-1 ’Fully and Practically Broken by New Collision
Duo:Flaw in Crypto library causes revocation of SSH keys for GIT services

Task 1: Using an example from one of the readings, what are the potential vulnerabilities of using broken encryption schemes like SHA-1?

Now read this article from Microsoft about their decisions to remove windows updates that use SHA-1. While you read think about the responsibility technology companies have to their users to ensure their privacy.

Task 2: What have tech companies done to prevent more vulnerabilities from SHA-1? Reference one strategy used by tech company’s mentioned in the article.

Counting Arguments

A counting argument shows that the LHS (lefthand side) and the RHS (righthand side) of some equation count the same thing. Instead of using algebraic manipulation, we explain why both sides enumerate the elements of some set, just in different ways.

For instance, consider the following identity.

\[
\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}
\]

Let \( S \) be a set with \( n \) elements.

- The LHS counts the number of ways to form a subset of \( S \) size \( k \).
- The RHS also counts the number of subsets of size \( k \). It partitions it into two parts:
  - \( k \)-sized subsets with a fixed element \( x \). Since we already know \( x \) is in the subset, so we just want to pick \( k - 1 \) more elements from the \( n - 1 \) remaining elements.
- $k$-sized subsets without $x$. We know we can’t pick $x$ for our subsets, so we just choose $k$ elements from $n - 1$ other elements in $S$.

**Task 3**

Below is a table of counting problems involving putting $k$ balls in $n$ distinct bins, and seven expressions. Match each problem with its solution.

<table>
<thead>
<tr>
<th></th>
<th>No restrictions</th>
<th>At most 1 ball per bin</th>
<th>At least 1 ball per bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical balls</td>
<td>1</td>
<td>3</td>
<td>(5) Optional</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distinct balls</td>
<td>2</td>
<td>4</td>
<td>(6) Optional</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A ( \binom{n}{k} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B ( n^k )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C ( \binom{k-1}{n-1} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D ( \frac{n!}{(n-k)!} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E ( \binom{k+n-1}{k} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F ( n^k - \sum_{i=1}^{n-1} \binom{n}{i} (n-i)^k (-1)^{i+1} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G ( \frac{n^k}{n!} \binom{n^k}{n} )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Checkpoint 1 - Call a TA over!

2
Inclusion/Exclusion

Say we have two sets, $A$ and $B$, and we want to know how many elements there are in their union, $A \cup B$. If $A$ and $B$ have no elements in common, we can compute this just by adding the cardinality of each, $|A| + |B|$. However, this approach will not work in general. For instance, if $A = \{1, 2\}$ and $B = \{1, 3\}$, then $A \cup B = \{1, 2, 3\}$. $|A \cup B| = 3$, but $|A| + |B| = 2 + 2 = 4$. The problem is that we’ve double counted the elements that are in both $A$ and $B$, that is, element $1$. To fix this problem, we should subtract $|A \cap B|$.

The resulting Inclusion/Exclusion property for two sets is:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$ 

Let’s try generalizing this idea to three sets, $A$, $B$, and $C$!
If we tried adding $|A| + |B| + |C|$, which regions would we double count? Which regions would we triple count?

As it turns out, the final formula for $|A \cup B \cup C|$ is the following. Convince yourself that it works!

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| + |A \cap B \cap C|.$$ 

Going further, we can repeat this process for any number of sets, alternating between adding and subtracting the sizes of sets.

**Task 4**

a. Let $S = \{1, 2, 3, 4, 5\}$.

i. How many permutations of $S$ contain the sequence 24?

ii. How many permutations of $S$ contain the sequence 52?

iii. How many permutations of $S$ contain the sequence 24 or 52?

b. Let $X$ and $Y$ be sets such that $|X| = 8$ and $Y = \{a, b, c\}$.

i. How many functions from $X \to Y$ do not map to $a$?

ii. How many functions from $X \to Y$ do not map to $a$ and also do not map to $b$?
iii. How many functions from $X \to Y$ are not surjective?

**Hint:** A function is not surjective if nothing maps to $a$, nothing maps to $b$, or nothing maps to $c$.

---

**The Pigeonhole Principle**

Let’s say we have $n$ pigeons who are trying to fit in $n-1$ holes.

It’s not possible for each pigeon to get its own hole: at least two of them are going to have to share. It could be the case that they are all in the same hole, or, like the picture above, all but 2 pigeons get their own hole, or anything in between.
This is the Pigeonhole Principle: in general, if we are assigning \( n \) objects to \( m \) categories, where \( n > m \), there is at least one category that has more than one object assigned to it.

Solve the following problems using the Pigeonhole Principle:

**Task 5**

a. Celeste is pulling socks out of her drawer. She only has four types of socks: solid, striped, polka-dotted, and ones with coffee mugs on them. What is the minimum number of socks Celeste should pull out to ensure she has a pair?

b. There are \( n > 2 \) astronauts having a party on Mars. Throughout the night, they dance with each other in pairs.

i. The minimum number of total dance partners someone can have is 0 (they didn’t dance with anyone). What is the maximum number of dance partners one can have?

ii. Prove that at least 2 astronauts have the same number of dance partners by the end of the night. (Come back to this question later if you get stuck.)

c. *Optional:* Given any 5 points inside a square with side length 2, there is always a pair whose distance apart is at most square root of 2.
Checkpoint 2 - Call over a TA
Probability

A finite discrete probability space is a pair \((S, \Pr)\) for some finite set \(S\) and some function \(\Pr : S \to \mathbb{R}\), where, for all \(x \in S\), \(\Pr(x) > 0\), and \(\sum_{x \in S} \Pr(x) = 1\). The set \(S\) is called the sample space, and the elements of \(S\) are called outcomes. The function \(\Pr\) is called the probability distribution function.

Where do these definitions come from? It’s all about predictions. The sample space \(S\) is supposed to represent a set of atomic outcomes—that is, the set of all things that can happen. All these outcomes are mutually exclusive.

An event is a subset of the sample space \(S\), containing some atomic outcomes. For example, if I flip a coin 3 times, the outcomes are HHH, TTT, \ldots, so the sample space is \(S = \{\text{HHH}, \text{TTT}, \text{TTH}, \ldots\}\), with \(2^3 = 8\) outcomes. An event can be: We see exactly 1 H, which we view as a set of outcomes \(\{\text{HTT}, \text{THT}, \text{TTH}\}\). An event can also be the empty set: for example, the event We see 4 Hs is the empty set, since it is not a possible outcome (it does not correspond to anything in the sample space). Similarly, the event We see an elephant is the empty set for this probability space.

How many different events can the above sample space have? Since each event is a subset of the sample space, the number of events is equal to the number of possible subsets, which here is \(2^8\) (since the sample space has 8 elements).

To understand the probability distribution function, let’s think about our sample space as a box of area 1. Then, each outcome occupies a certain amount of area within the sample space. That is, the area is distributed among the outcomes according to the probability distribution function! The larger the area, the more likely the outcome is.

<table>
<thead>
<tr>
<th>It rains</th>
<th>It doesn’t rain</th>
</tr>
</thead>
</table>

In this example, we have that each outcome occupies the same amount of area in the box—each is equally likely. Because we have 2 outcomes, we have that the probability of each outcome is \(\frac{1}{2}\).
Let’s take a look at another way we could divide up the area of the box among the outcomes.

Here, the probability of it raining is $\frac{3}{4}$, so this outcome takes up $\frac{3}{4}$ of the area within the box. The remaining amount of area is $\frac{1}{4}$, which becomes the probability that it doesn’t rain.

**Task 6**

a. If we have $n$ outcomes in $S$, and Pr assigns each outcome an equal amount of area within our box, what is the probability of a particular outcome?

**Note**: Such a division of the space is called a *uniform distribution*!

b. Consider $S = \{\text{It’s sunny, It’s raining}\}$, where the probability distribution is 0.8 and 0.2, respectively. Under what constraints on the world is this sample space reasonable?

c. Is the following probability distribution valid? Experiment: flipping a biased coin. $S = \{H, T\}$, $\Pr(H) = 1/3$, $\Pr(T) = 1/3$. 
The Relationship Between Counting and Probability

Questions about probability are often closely tied to questions about counting. Often, we need to count two things—the set of outcomes in our event, and the set of total possible outcomes. In a uniform distribution, the probability of getting an event $E$ is simply $\Pr(E) = |E|/|S|$.

Task 7

a. Suppose $S$ is the set of all binary strings of length $n$, and $\Pr$ is a uniform distribution. What is the probability of getting a string with $k$ ones (where $0 \leq k \leq n$)?

b. For what value(s) of $k$ is the probability of getting a string with $k$ ones the largest? Try it out for some values and see if you find a pattern.

c. Explain how we can use $n$-ary (not just binary) strings to help us with other problems that involve, say, flipping a fair coin $m$ times, or rolling a die $m$ times.

Conditional Probability and Independence

We sometimes want to know what the probability of something is, given that we know an outcome from certain subset of outcomes has happened (that is, no outcome outside that subset can happen).

Prof. Lewis has 2 coins: one fair, and one double heads. He picks one uniformly at random (that is, each coin has probability of $1/2$ of being picked), but he doesn’t tell us which one he picks. Prof. Lewis then flips the coin he picks, and he then shouts out the result.
Suppose Prof. Lewis shouts out “Heads.” What is the probability he flipped the fair coin, given that we know the coin flip resulted in heads (H)?

*Hint:* How many ways can the coin flip result in heads, and how many of them start with the fair coin?

We can generalize our work here. Given two events $A, B \subseteq S$ (the sample space) where $\Pr(B) > 0$, we define the *conditional probability of $A$ given $B$* as

$$
\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}.
$$

We know that the only possible outcomes are those in $B$ (since $B$ is supposed to have happened). We therefore make $\Pr(B)$ our denominator to renormalize our universe. Looking at this formula, we also get a general definition for $\Pr(A \cap B)$:

$$
\Pr(A \cap B) = \Pr(A) \Pr(B \mid A) = \Pr(B) \Pr(A \mid B)
$$

This is a derivation of *Bayes’ Theorem*, and it will be your friend in the probability unit.

There are certain special situations where $\Pr(B \mid A) = \Pr(B)$; that is, where the fact that $A$ happened does not affect the probability of $B$ happening. We have a name for this situation.

**Definition:** Two events $A, B \subseteq S$ are *(pairwise) independent* if

$$
\Pr(A \mid B) = \Pr(A)
$$

If $\Pr(B) = 0$, $B$ and any other event are independent.

Equivalently, $A$ and $B$ are independent if

$$
\Pr(A \cap B) = \Pr(A) \Pr(B).
$$

Note that these two definitions are equivalent—try to convince yourself by substituting the definition of conditional probability!

**Task 8**

a. If two events $A$ and $B$ have an empty intersection, what is the probability of $A$ AND $B$?
b. For each of the following pairs of events $A$ and $B$, identify whether they are independent, and justify why or why not.

i. Suppose we roll a fair die, and suppose $A =$ rolling an even number, and $B =$ rolling a number greater than three.

ii. *Optional:* Suppose we flip a fair coin three times, and suppose $A =$ the last coin is a tails, and $B =$ there is a run of exactly two tails (that is, two, but not three, tails are flipped in a row).

c. *Optional:* If two events $A$ and $B$ are *mutually exclusive*, are they *independent*? You can assume $\Pr(A)$ and $\Pr(B)$ are nonzero.

Checkoff — call over a TA!