Note 7
(RSA)

- Def: The public key is (N, e) where $N = pq$ such that p, q are two very large primes. e is (typically a relatively small $\mathbb{Z}^+$) such that $e, (p-1)(q-1)$ are coprime. The private key is d where $d \equiv e^{-1} \pmod{(p-1)(q-1)}$. Finally, x is the original message, which will be encrypted into $y \equiv E(x) = x^e \pmod{N}$, which could be decrypted by having $D(y) = D(E(x)) \equiv y^d = x \pmod{N}$.
- Theorem 7.1: Under the above definitions of the encryption and decryption functions E and D, we have $D(E(x)) = x \pmod{N}$ for every possible message $x \in \{0, 1, ..., N-1\}$.
- Theorem 7.2 (Fermat's Little Theorem, FLT): For any prime p and any $a \in \{1, 2, ..., p-1\}$, we have $a^{p-1} \equiv 1 \pmod{p}$ (Extra: $a^y = a^{y \pmod{p-1}} \pmod{p}$ since $a \neq 0$; for any x, $x^p \equiv x \pmod{p}$).
- Theorem 7.3 (Prime Number Theorem): Let $\pi(n)$ denote the number of primes that are $\leq n$. Then for all $n \geq 17$, we have $\pi(n) \geq \frac{n}{\ln n}$ (And in fact $\lim_{n\to\infty} \frac{\pi(n)}{n/\ln n} = 1$).
- HW: $a^{p(p-1)} \equiv 1 \pmod{p^2}$ where p is prime and a, p are coprime
- Extra (need proof by ???): If $ax + bm = d$, $\gcd(x, m) = d, xu \equiv v \pmod{m}$, then it has a solution $\iff d \mid v$; if so, one such solution is $u \equiv \frac{va}{d} \pmod{\frac{m}{d}}$, (in essence, $a = (\frac{x}{d})^{-1} \pmod{\frac{m}{d}}$,) and there are exactly d-many solutions (of the form $u = \frac{va}{d} + i \cdot \frac{m}{d} \pmod{m}$).
- Extra (Proved in HW): If $a \equiv b \pmod{m_1}, a \equiv b \pmod{m_2}$ and $\gcd(m_1, m_2) = 1$, then $a \equiv b \pmod{m_1 m_2}$.
- Extra (Proved in HW, FLT entended to composite): If $n = p_1 p_2 \cdots p_k$ where p_i are distinct primes and $(p_i - 1) \mid (n-1)\ \forall i$, then $a^{n-1} \equiv 1 \pmod{n}\ \forall a \in \{i \mid 1 \leq i \leq n \wedge \gcd(n, i) = 1\}$.
- Extra: $(p-1)! \equiv (p-1) \pmod{p}$: Proof by the fact that $2, ..., p-2$ will pair up with their own inverse, and only ones that map back to themselves are $1, -1$.

Note 8
(Polynomials)

- Fundamental properties of polynomials
  - Property 1: A non-zero polynomial of degree d has at most d roots.
  - Property 2: Given $d+1$ pairs $(x_1, y_1), \ldots, (x_{d+1}, y_{d+1})$, with all the x_i distinct, then there is a unique polynomial $p(x)$ of degree (at most) d such that $p(x_i) = y_i$ for $1 \leq i \leq d+1$.
- Lagrange interpolation
  - Task: Construct the unique polynomial $p(x)$ of degree (at most) d from the given $d+1$ pairs
  - Step 0: Def: Let $\Delta_i(x)$ be the degree d polynomial that goes through these $d+1$ points. Then $\Delta_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}$
  - Step 1: Given the $d+1$ pairs, we first construct $\Delta_1(x), \ldots, \Delta_{d+1}(x)$ as described above.
  - Step 2: Construct $p(x) = \sum_{i=1}^{d+1} y_i \Delta_i(x)$
- Polynomial division: degree d polynomial $p(x)$ divided by degree $\leq d$ polynomial $q(x)$: $p(x) = q'(x)q(x) + r(x)$ where $\deg(r(x)) < \deg(q(x))$

- Finite field: $GF(p)$ or $F(p)$ when we work with numbers modulo a prime p
- In general, we can specify a degree d polynomial $p(x)$ by specifying its values at $d+1$ points, say $x = 0, 1, \ldots, d$. These $d+1$ values, $(y_0, y_1, \ldots, y_d)$, are called the *value representation* of $p(x)$. The *coefficient representation* can be converted to the value representation by evaluating the polynomial at $0, 1, ..., d$.
- Note that the reason that we can count the number of polynomials in this setting is because we are working over a finite field.

Polynomials of degree $\leq d$ over F_m	
# of points	# of polynomials
$d+1$	1
d	m
$d-1$	m^2
$\vdots$	$\vdots$
$d-k$	m^{k+1}
$\vdots$	$\vdots$
0	m^{d+1}

- Secret Sharing utilizes Lagrange interpolation and works under $GF(p)$. It is important to note that, two people have no more information about the secret than one person does.
- HW: Min/Max number of roots for non-zero real polynomials f and g
  - $f + g$: min = 0 if sum is even degree, 1 if odd degree; max = max(deg(f), deg(g)) if sum is non-zero, and infinity if $f = -g$
  - $f \cdot g$: min = 0; max = deg(f) + deg(g)
  - f/g (assume polynomial): min = 0; max = deg(f) - deg(g)
- HW: Consider $f = x^{p-1} - 1, g = x$ be polynomials over $GF(p)$. Both are non-zero polynomials, but their product have a zero on all points in $GF(p)$.
- HW: If f is under $GF(p)$ such that deg(f) $\geq p$, then there exists a polynomial h with deg(h) $< p$ such that $f(x) = h(x)$ for all $x \in \{0, 1, \ldots, p-1\}$.
- HW: There are $(p-1) \cdot p^{d-1}$ many f of degree *exactly* $d < p$ over $GF(p)$ such that $f(0) = a$ for some fixed $a \in \{0, 1, \ldots, p-1\}$
- Extra: There are p^p polynomials in $GF(p)$ for some prime p.
- Extra: In $GF(p)$, $\exists$ $p(x)$ of degree d and $q(x)$ of degree $d-1$ such that a degree 1 polynomial $y(x) = p(x)$ satisfies $p(-y(0)) = 0$, where $d < p-1$ ($\equiv p, q$ share $d-1$ roots).

Note 9
(Error Correction)

- In a modular setting, division is equivalent to multiplying by the inverse, i.e. $\frac{a}{x} \equiv ax^{-1} \pmod{p}$
- If we want to send k packets in total, we need to work under $GF(p)$ where $p > k$.
- Erasure Errors: We should transmit at least $(n+k)$ packets to guard against k erasure errors. We can uniquely reconstruct $P(x)$ from its values at *any* n distinct points via Lagrange interpolation, since it has degree $n-1$.
- General Errors: To guard against k general/corruption errors, we should transmit at least $(n+2k)$ packets.
  - Error-locator polynomial (deg k): $E(x) = (x-e_1)(x-e_2)\dots(x-e_k)$
  - $P(i)E(i) = r_i E(i)$ for $1 \le i \le n+2k$
  - Define $Q(x) := P(x)E(x)$ $(= r_x E(x))$, which is a polynomial of degree $(n+k-1)$, and is therefore described by $n+k$ coefficients $a_0, a_1, \dots, a_{n+k-1}$
  - The error-locator polynomial $E(x) = (x-e_1)(x-e_2)\dots(x-e_k)$ has degree k and is described by $k+1$ coefficients $b_0, b_1, \dots, b_k$ but the leading coefficient (the coefficient b_k of x_k) is always 1, which leads to:

$$Q(x) = a_{n+k-1}x^{n+k-1} + \dots + a_1 x + a_0$$

$$E(x) = x^k + b_{k-1}x^{k-1} + \dots + b_1 x + b_0$$

  - We thus have $(n+2k)$ linear equations, one for each value of i, and $(n+2k)$ unknowns. We can solve these equations and get $E(x)$ and $Q(x)$. We can then compute the ratio $\frac{Q(x)}{E(x)}$ to obtain $P(x)$.
- Error Detection (HW): We should transmit at least $(n+k)$ packets to detect any errors, assuming that a maximum of k errors exists.

Note 10
(Infinity and Countability)

- Notes: f is bijection $\iff$ it is both one-to-one and onto.
- Def: A set S is countable if there is a bijection between S and $\mathbb{N}$ or some subset of $\mathbb{N}$. (Extra: The latter part implies that a set S is countable if there's an injection from S to $\mathbb{N}$)
- The Cantor-Bernstein theorem (?): $|B| \le |A| \le |B| \iff A, B$ have the same cardinality. (one-to-one + onto (reversed 1-to-1) $\to$ bijection)
- Notes: The set of all (finite-degree) polynomials with $\mathbb{N}$ coefficients, which we denote $N(x)$, is countable.
- Theorem: The real interval $\mathbb{R}[0,1]$ (and hence also $\mathbb{R}$) is uncountable. (Proof by Diagonalization)
- Theorem: $|\mathscr{P}(\mathbb{N})| > |\mathbb{N}|$ (and thus, the power set of an infinite set is uncountable).
- Disc: The countable union of countable sets is countable.
- Disc: The subset of any countable set is countable, which also implies that the intersection of (a countable or uncountable number of) countable sets is countable.
- Disc: The functions from $\mathbb{N}$ to $\mathbb{N}$ is uncountable. (Diagonalization)

- Disc: The total number of programs is countably infinite, since each is a string of characters, so halting programs, a subset of all programs, is either countable. Since there are an infinite number of halting programs, for example for each number i the program that just prints i, so the total number of halting programs is countably infinite. The two results implies that not every function from $\mathbb{N}$ to $\mathbb{N}$ can be written as a program.
- Disc: Given A is a countable, non-empty set. For all $i \in A$, S_i is an uncountable set; B is an uncountable set. For all $i \in B$, Q_i is a countable set.
  - $A \cap B$ countable; $A \cup B$ uncountable
  - $\cup_{i \in A} S_i$ uncountable; $\cap_{i \in A} S_i$ both could happen (disjoint vs. identical)
  - $\cup_{i \in B} Q_i$ both could happen (identical vs. $B = \mathbb{R}, Q_i = \{i\}$); $\cap_{i \in B} Q_i$ countable
- Disc: It's not possible to determine if we ever execute a specific line because this depends on the logic of the program, but the number of computer instructions can be counted.
- Extra: Finite graphs (defined as any set of vertices and edges) is countable infinite (defined isomorphically).

Note 11
(Self-Reference and Computability)

- Theorem: The Halting Problem is uncomputable; i.e., there does not exist a computer program *TestHalt* that can determine, on all inputs (P, x), whether the program P halts on input x.
- Notes: The Easy Halting Problem is uncomputable (testing whether a program P halts on input 0). Proof by Reduction.
- Godel's Incompleteness Theorem: Any formal system that is sufficiently rich to formalize arithmetic is either inconsistent (there are false statements that can be proved) or incomplete (there are true statements that cannot be proved).
- HW
  - $f(x) = p \pmod{x}$, where $p > 2$ is a prime, then $f : \{\frac{p+1}{2}, \ldots, p\} \to \{0, \ldots, \frac{p-1}{2}\}$ is a bijection.
  - Countable: $A \times B$, where A and B are both countable.
  - Countable: $\cup_{i \in A} B_i$ where A, B_i are all countable
  - Uncountable: The set of all functions $f : \mathbb{N} \to \mathbb{N}$ such that f is non-decreasing, i.e. $f(x) \leq f(y)$ whenever $x \leq y$. (Proof by Diagonalization)
  - Countable: The set of all functions $f : \mathbb{N} \to \mathbb{N}$ such that f is non-increasing, i.e. $f(x) \geq f(y)$ whenever $x \geq y$.
  - Uncountable: The set of all bijective functions from $\mathbb{N}$ to $\mathbb{N}$.
- Disc: We can determine whether a program halts/outputs sth in k^{th} steps, but not (1) before k^{th} line or (2) on any input x.
- Extra: We can determine whether a program contains a loop, given finite memory (since we can list all possible k states and run the program $k + 1$ steps). We can't check any program as a whole without the finite memory restriction.

Note 12
(Counting)

- First Rule of Counting: $n_1 n_2 \cdots n_k$ (order matters)
- Second Rule of Counting: $\binom{n}{k}$ (order doesn't matter)

	Replacement	No Replacement
Order	n^k	$\frac{n!}{(n-k)!}$
No Order	$\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$	? $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

- Combinatorial Proofs
  - $\binom{n}{k+1} = \binom{n-1}{k} + \binom{n-2}{k} + \cdots + \binom{k}{k}$
  - $2^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}$
  - $\sum_{i=1}^{n} i\binom{n}{i} = n2^{n-1}$
  - $\sum_{i=1}^{n} i\binom{n}{i}^2 = n\binom{2n-1}{n-1}$
- HW: If there are exactly 20 students currently enrolled in a class, then there are $19 \cdot 17 \cdots 3 \cdot 1 = \frac{20!}{10! \cdot 2^{10}} = \frac{\binom{20}{2}\binom{18}{2}\cdots\binom{2}{2}}{10!}$ different ways to pair up the 20 students.
- Disc: $\sum_{k=3}^{n} \frac{n!}{(n-k)! \cdot 2k}$ distinct cycles are there in a complete graph with n vertices. Assuming no duplicated edges and two cycles are considered the same if they are rotations or inversions of each other.

Note 13
(Discrete Probability)

- Def: The outcome of a random experiment is called a sample point, and the sample space (often denoted by Ω) is the set of all possible outcomes of the experiment.
- Def: A **probability space** is a **sample space** Ω, together with a probability $\mathbb{P}[\omega]$ (aka $\Pr[\omega]$) for each sample point ω, such that
  - (Non-negativity): $0 \leq \mathbb{P}[\omega] \leq 1$ for all $\omega \in \Omega$
  - (Total one): $\sum_{\omega \in \Omega} \mathbb{P}[\omega] = 1$, i.e., the sum of the probabilities over all outcomes is 1.
- If the coin has $\mathbb{P}(H) = p$, and if we consider any sequence of n coin flips with exactly r H's, then the probability of this sequence is $p^r(1-p)^{n-r}$. Now consider the event C that we get exactly r H's when we flip the coin n times, so $\mathbb{P}[C] = \binom{n}{r} p^r (1-p)^{n-r}$
- If we throw m labeled balls into n labeled bins, we have a sample space of size n^m (generalized).

Note 14
(Conditional Probability)

- Def 14.1 (Conditional Probability): For events $A, B \subseteq \Omega$ in the same probability space such that $\mathbb{P}[B] > 0$, the conditional probability of A *given* B is $\mathbb{P}[A|B] = \sum_{\omega \in A \cap B} \mathbb{P}[\omega|B] = \sum_{\omega \in A \cap B} \frac{\mathbb{P}[\omega]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$
- $\mathbb{P}[A \cap B] = \mathbb{P}[A|B] \cdot \mathbb{P}[B] = \mathbb{P}[B|A] \cdot \mathbb{P}[A] \longrightarrow \mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}$ Bayes' Rule

- $\mathbb{P}[B] = \mathbb{P}[A \cap B] + \mathbb{P}[\overline{A} \cap B] = \mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\overline{A}]\mathbb{P}[A] = \mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\overline{A}](1 - \mathbb{P}[A])$ Total Probability Rule
- Thus, combining the two equations above gives: $\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\overline{A}](1-\mathbb{P}[A])}$
- Def 14.2 (Partition of an event): We say that an event A is partitioned into n events $A_1, \ldots, A_n$ if
  1. $A = A_1 \cup A_2 \cup \cdots \cup A_n$, and
  2. $A_i \cap A_j = \emptyset$ for all $i \neq j$ (mutually exclusive)
- Let $A_1, \ldots, A_n$ be a partition of the sample space Ω. Then, the Total Probability Rule for any event B is $\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \cap A_i] = \sum_{i=1}^{n} \mathbb{P}[B|A_i]\mathbb{P}[A_i]$, and the Bayes' Rule (assuming $\mathbb{P}[B] \neq 0$) is $\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]\mathbb{P}[A_i]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\mathbb{P}[A_i]}{\sum_{j=1}^{n} \mathbb{P}[B|A_j]\mathbb{P}[A_j]}$
- Def 14.3 (Independence): Two events A, B in the same probability space are said to be independent if $\mathbb{P}[A \cap B] = \mathbb{P}[A] \times \mathbb{P}[B]$. For events A, B such that $\mathbb{P}[B] > 0$, the condition $\mathbb{P}[A|B] = \mathbb{P}[A]$ is actually equivalent to the definition of independence.
- Def 14.4/5 (Mutual independence): Events $A_1, \ldots, A_n$ are said to be mutually independent if (two equivalent definitions):
  - for EVERY subset $I \subseteq \{1, \ldots, n\}$ with size $|I| \geq 2$, we have $\mathbb{P}[\cap_{i \in I} A_i] = \prod_{i \in I} \mathbb{P}[A_i]$
  - for all $B_i \in \{A_i, \overline{A_i}\}, i = 1, \ldots, n$, we have $\mathbb{P}[B_1 \cap \cdots \cap B_n] = \prod_{i=1}^{n} \mathbb{P}[B_i]$
- The independence of every pair of events (so-called pairwise independence) does not necessarily imply mutual independence.
- Theorem 14.1 (Product Rule): $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B|A]$. More generally, for any events $A_1, \ldots, A_n$, we have $\mathbb{P}[\cap_{i=1}^{n} A_i] = \mathbb{P}[A_1] \cdot \mathbb{P}[A_2|A_1] \cdots \mathbb{P}[A_3|A_1 \cap A_2] \cdots \mathbb{P}[A_n| \cap_{i=1}^{n-1} A_i]$ (Proof by Induction)
- Theorem 14.2 (Inclusion-Exclusion): Let $A_1, \ldots, A_n$ be events in some probability space, where $n \geq 2$. Then, we have $\mathbb{P}[\cup_{i=1}^{n} A_i] = \sum_{i=1} n\mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k] - \cdots + (-1)^{n-1}\mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n]$ (Proof by Induction)
- (Mutually exclusive events) If the events $A_1, \ldots, A_n$ are mutually exclusive, then $\mathbb{P}[\cup_{i=1}^{n} A_i] = \sum_{i=1} n\mathbb{P}[A_i]$
- (Union bound) Upper bound always due to overestimating: $\mathbb{P}[\cup_{i=1}^{n} A_i] \leq \sum_{i=1} n\mathbb{P}[A_i]$

Note 15
(Random Variables, Expectation)

- Def 15.1 (Random Variable): A random variable X on a sample space Ω is a function $X : \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.
- Def 15.2 (Distribution): The distribution of a discrete random variable X is the collection of values $\{(a, \mathbb{P}[X = a]) : a \in \mathscr{A}\}$, where $\mathscr{A}$ is the set of all possible values taken by X (Note that the collection of events form a partition of the sample space Ω).

- Bernoulli Distribution: $X \sim \text{Bernoulli}(p)$ which takes value in $\{0, 1\}$ such that $p = \mathbb{P}[X = 1] = 1 - \mathbb{P}[X = 0]$ where $0 \leq p \leq 1$.
- Binomial Distribution: $X \sim \text{Bin}(n, p)$ such that $\mathbb{P}[X = i] = \binom{n}{i} \cdot p^i(1-p)^{n-i}$ for $i = 0, 1, \ldots, n$.
- The $\text{Bin}(n, p)$ gives probabilistic proof for the Binomial Theorem since by properties of X we have $\sum_{i=0}^{n} \mathbb{P}[X = i] = 1 \longrightarrow \sum_{i=0}^{n} \binom{n}{i} \cdot p^i(1-p)^{n-i} = 1$
- Relation to Error Correction: If we model each packet getting lost with probability p and the losses are independent, then if we transmit $n + k$ packets, the number of packets received is a random variable X with binomial distribution: $X \sim \text{Bin}(n + k, 1 - p)$, so the probability of successfully decoding the original data is: $\mathbb{P}[X \geq n] = \sum_{i=n}^{n+k} \mathbb{P}[X = i] = \sum_{i=n}^{n+k} \binom{n+k}{i} \cdot (1-p)^i p^{n+k-i}$
- Def 15.3 (Expectation): The expectation of a discrete random variable X is defined as $\mathbb{E}[X] = \sum_{a \in \mathscr{A}} a \cdot \mathbb{P}[X = a]$ where the sum is over all possible values taken by the r.v. (Expectation is well defined provided that the sum on the RHS is absolutely convergent, and there are random variables for which expectations do not exist (haven't encountered yet but keep in mind).
- The expectation can be seen in some sense as a "typical" value of the r.v. (though note that $\mathbb{E}[X]$ may not actually be a value that X can take, like a regular die).
- Theorem 15.1 (Linearity of Expectation): For any two random variables X, Y on the same probability space and any constant c, we have $\mathbb{E}[X+Y] = \mathbb{E}[X]+\mathbb{E}[Y], \mathbb{E}[cX] = c\mathbb{E}[X]$ regardless of whether or not X and Y are independent. If X, Y independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- To take advantage of Theorem 15.1, we need to write X_n as a sum of simpler r.v.'s. Since X_n counts the number of times something happens, we can write it as a sum using the following useful trick: $X_n = I_1 + \cdots + I_n$ where $I_i = 1$ if student i gets their own HW and 0 otherwise. So, we see that the expected number of students who get their own homeworks in a class of size n is 1. That is, the expected number of fixed points in a random permutation of n items is always 1, regardless of n.
- HW: If A and B are integer-valued random variables such that for every integer i, $P(A = i) = P(B = i)$, then $P(A = B) > 0$ is not necessarily true. (Counterexample: Let A be 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$. Let $B = 1 - A$. Then A and B are never equal but B takes the values 0, 1 with probability $\frac{1}{2}$ each as well.)
- Extra: $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$ since $\mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \mathbb{E}\big[(Z - \mathbb{E}[Z])^2\big] \geq 0$ with $Z \cdot \mathbb{E}[Z] = \mathbb{E}[Z^2]$

Extra Sanity Check

- All polynomials that Lagrange Interpolation can output $\neq$ all polynomial solutions (e.g. polynomial with the coordinates $(-1, 1), (0, 0), (2, 4)$ exist in $GF(8)$).
- Prove strategy for uncountable: Diagonalization; for uncomputable: Reduction to the Halting Problem OR Use self-referencing algorithms
- There are two ways to show undecidability: Use your program as a subroutine to solve a problem we know is undecidable OR Proof by Diagonalization.
- Calculating probability with these steps:
  - What is the sample space (i.e., the experiment and its set of possible outcomes)?

  - What is the probability of each outcome (sample point)?
  - What is the event we are interested in (i.e., which subset of the sample space)?
  - Finally, compute the probability of the event by adding up the probabilities of the sample points contained in it.

- The independence of every pair of events (so-called pairwise independence) does not necessarily imply mutual independence (counterexample: flip first coin, flip second coin, both flips are the same).
- When solving r.v. problems of intuition, always double-check the edge cases of $\mathbb{P}[A] = 0$ or 1.
- A random experiment must define the sample space (the set of possible outcomes) AND a set of probabilities.