STA257 Notes

Max Xu

'25 Fall

Contents

§1 Day 1: Sample Spaces and Probability (Sept 3, 2025)

Life is very random and uncertain, with interesting problems to solve (e.g. whats the probability that I win the lottery). This course uses certain mathematics to study the uncertainty of probabilities. You will be able to solve problems like this:

Problem 1.1. Which is more likely: getting at least one six when rolling a fair 6 sided die 4 times, or getting one pair of sixes when rolling two six sided dice 24 times?

5% of your grade is poll-based, more info here.

Definition 1.2 (Sample Space)**.** A non-empty set containing all possible outcomes, written S.

e.g. coin-flipping: $S = \{\text{Heads}, \text{Tails}\}$, two die: $S = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$

Definition 1.3 (Event)**.** Any subset $A \subseteq S$ is an event.

Prof says in some continuous sample spaces, there may exist some non-measurable subsets to which the probability measure defined later won't work on, but don't worry about it in this course. (yay!!)

Definition 1.4 (Probability)**.** For any event A, define probability $\mathrm{P}(A)$ that satisfies:

- For all $A \subseteq S$, $0 \leq \mathrm{P}(A) \leq 1$
- $A = S$, $\mathrm{P}(A) = \mathrm{P}(S) = 1$
- $A = \emptyset$, corresponding to no outcome, then $\mathrm{P}(A) = \mathrm{P}(\emptyset) = 0$
- **Additivity**: if $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

If $A_1, \ldots, A_n$ are disjoint[1] events, we have

- **Finite Additivity**: For some $n \in \mathbb{N}$,

$$\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathrm{P}(A_i)$$

- **Countable Additivity**:

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$$

[2]

When looking at the probability of getting heads from a coinflip, $\mathrm{P}(H)$ with $S = \{T, H\}$ is really shorthand for $\mathrm{P}(\{H\})$, since H may not be a subset of S. For uniformly picking any number between 0 and 1, denoted $\text{Uniform}[0, 1]$, we can define a probability $\mathrm{P}([a, b]) = b - a$ whenever $0 \leq a \leq b \leq 1$. (don't know what uniform means yet)

Note that by definition of probability, $\mathrm{P}(A_i)$ is positive, so the right hand side is an absolutely convergent series (prof didn't mention this).

[1] prof said this but i think he meant pairwise disjoint?

[2] we could've actually done this with just 3, see probability axioms

§2 Day 2: Properties of Probability (Sept 8, 2025)

§2.1 Additional Properties of Probability

Today we will be deriving properties of probability from the 'axioms' we stated last class. Note that most of these follow from the additivity property 1.4. Have $A, B \subseteq S$ be events.

Theorem 2.1. If A^C is the complement of A, then $P(A^C) = 1 - P(A)$.

Proof. A and A^C are by definition disjoint, and their union is S. By additivity 1.4 $P(A) + P(A^C) = P(S) = 1$. □

Theorem 2.2. $P(A) = P(A \cap B) + P(A \cap B^C)$

The set of $\{x \in A : x \in B\}$ and $\{x \in A : x \notin B\}$ are by definition disjoint, and the union of the two is A. This then follows by additivity 1.4.

Theorem 2.3. If A contains B, $P(A) = P(B) + P(A \cap B^C)$

Proof. Have 2.2, except $P(A \cup B) = P(B)$ where $A \supseteq B$. □

Theorem 2.4 (Monotonicity)**.** If $A \supseteq B$, then $P(A) \geq P(B)$

Immediately follows from 2.3, since $P(A \cap B^C)$ must be non-negative, giving the inequality.

Theorem 2.5 (Law of Total Probability)

Suppose $A_1, A_2, \ldots$ are a sequence of events which form a *partition* of S (pairwise disjoint) , with their union being the entire sample space ($\bigcup_i A_i = S$). Let B be any event. Then we have

$$P(B) = \sum_i P(A_i \cap B)$$

Theorem 2.6 (Principle of Inclusion-Exclusion)**.** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof. The events $A \cap B^C$, $B \cap A^C$, $A \cap B$ are disjoint events.

$$\begin{aligned} P(A \cup B) &= P(A \cap B) + P(A \cap B^C) + P(B \cap A^C) \\ &= P(A \cap B) + [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] \qquad \text{from 2.2} \end{aligned}$$

□

For a more generalized version of inclusion-exclusion formula, look at Challenge 1.3.10 in textbook.

Theorem 2.7 (Subadditivity)**.** For any sequence of events $A_1, A_2, \ldots$ not necessarily pairwise disjoint, have

$$P(A_1 \cup A_2 \ldots) \leq P(A_1) + P(A_2) + \ldots$$

TODO: Add proof for this

> **Remark 2.8.** This is more of a worry in grad-level courses, where you study more pathological probability spaces, but 'uncountable' subadditivity does not exist. Consider $S = \text{Uniform}([0,1])$. Have $A_x = \{x\}$ for $x \in S$. $P(\bigcup_{x\in S})A_x = P(S) = P([0,1]) = 1$. Yet for any 'singleton' x, $P(A_x) = P(\{x\}) = 0$, meaning $\sum_{x\in S} P(A_x) = 0$.

§2.2 Uniform Probabilities on Finite Spaces

Have $S = \{s_1, \ldots, s_n\}$. For all $\{s_i\}$ to have the same probability, $P(\{s_i\}) = \frac{1}{n}$, called a *discrete uniform distribution.*

Any $A \subseteq S$ with k elements, would have $P(A) = \frac{k}{n}$, meaning

$$P(A) = \frac{|A|}{|S|}$$

A problem solving technique to find P of a rather complicated event is to see if the probability of its complement can be easily found, then use 2.1.

§3 Day 3: 'Combinatorics' (Sept 10, 2025)

For the first few weeks, we will mostly be dealing with uniform spaces. Be careful, if the probability is non-uniform, meaning not all outcomes are equally likely, the counting technique from last class would not apply.

The sample space can also be a discrete infinite set, e.g. $S = \mathbb{N} = \{1, 2, \ldots\}$, with $P(\{i\}) = 2^{-i}$ for $i \in \mathbb{N}$. We can check that this is valid by checking that each $0 \leq P(\{i\})$

$$\sum_{i=1}^{\infty} 2^{-i} = 1$$

To get the probability of the even numbers, we can compute the sum

$$\sum_{i=2,4,6,\ldots}^{\infty} 2^{-i} = \frac{1}{3}$$

which is quite surprising.
On a discrete infinite space, we cannot have a uniform distribution.

§3.1 More Finite Uniform Probabilities

The number of ways to pick k distinct items *in order* out of n items total, is

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

which is also called a 'permutation', written $P(n, k)$.

There are $k!$ ways to order k distinct objects. For this reason, the number of ways to pick k distinct *unordered* objects,

$$n(n-1)\cdots(n-k+1)/k! = \frac{n!}{(n-k)!k!}$$

This formula is called 'combinations', 'choose formula', or 'binomial coefficient', written $C(n, k)$, n choose k, and $\binom{n}{k}$ respectively.

$$C(n, k) = \frac{P(n, k)}{k!}$$

> **Remark 3.1.** Regarding the lottery, my advice is to not buy a lottery ticket. But if you really wanted to, you should avoid common patterns, valid birthdays etc... so you can avoid having to share the winnings with another person. - Prof Rosenthal

In a standard deck of playing cards, there are 4 suits, with each suit having 13 ranks, making $4 \cdot 13 = 52$ cards total.

$$P(\text{Clubs or } 7) = P(\text{Clubs}) + P(7) - P(\text{Clubs and } 7)$$

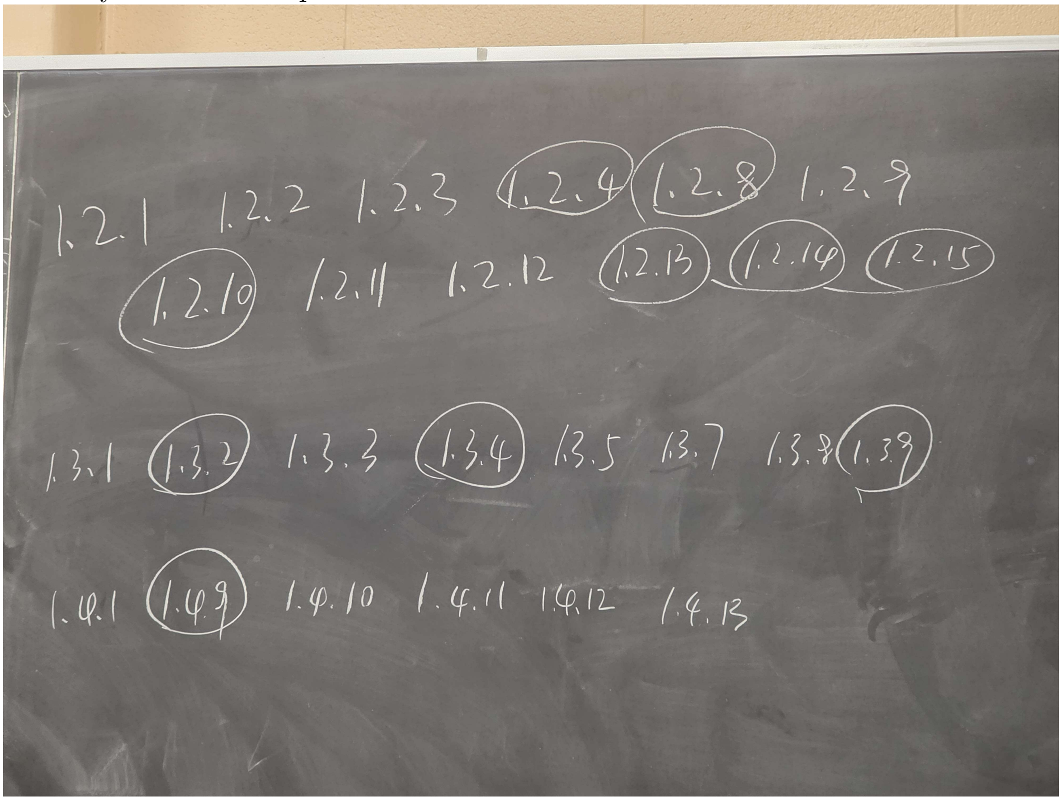
which is by inclusion-exclusion 2.6.

§4 Tutorial 1 (Sept 10, 2025)

DeMorgan's laws state that

$$(A \cap B)^C = A^C \cup B^C, (A \cup B)^C = A^C \cap B^C$$

Exercises you should complete from the textbook:



§5 Day 4: Conditional Probability (Sept 15, 2025)

From last class: choosing a subset in order is called a permutation, choosing a subset irrespective of the order is called a combination. You don't have to use R in this course, but if you wanted to there is some info here.

Problem 5.1. Suppose we flip 4 fair coins, what is $P(\text{exactly 2 heads})$?

You could solve this by writing out the entire sample space. Or by computing $\frac{\binom{4}{2}}{2^4} = \frac{3}{8}$. In general, for flipping n coins, the probability of getting exactly k heads is

$$\frac{\binom{n}{k}}{2^n} \text{for } 0 \leq k \leq n$$

§5.1 Conditional Probability

We now receive some information that restricts the sample space of interest to some subset of the original sample space S. If P was a discrete uniform distribution, P on said subset is remains a discrete uniform distribution.

Definition 5.2 (Conditional Probability)**.** If A and B are two events, where $P(B) > 0$, then the *conditional probability* of A given B is written $P(A \mid B)$ represents the fraction of the times when B occurs, in which A also occurs.

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

Theorem 5.3 (Conditional Multiplication Formula)**.**

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

Combining with law of total probability 2.5 we can get a more useful version where we replace $P(A_i \cap B)$ according to 5.3, giving

$$P(B) = \sum_i P(A_i)P(B \mid A_i)$$

Problem 5.4 (Challenge)**.** Roll n fair die. What are the odds we get more than $0 \leq k \leq n$ 5s?