MATH 3181 - Lecture Notes

Lucas Sta. Maria
stamaria.l@northeastern.edu

September 11, 2023

Contents

1 LECTURE 0

Definition 1.1 (Experiment)**.** An *experiment* is a procedure that

- could be repeated
- has a well-defined set of outcomes that can be described mathematically

Definition 1.2 (Outcome)**.** An *outcome* is a state of the universe as relevant to the experiment.

Definition 1.3 (Sample Space)**.** The *sample space* of an experiment is the set of possible outcomes for the experiment.

Definition 1.4 (Event)**.** An *event* is a collection of outcomes. Mathematically, a subset of the sample space.

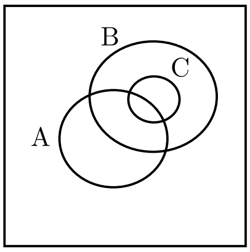
Exercise 1.5. What are the outcomes in the sample space of 3 consecutive flips of a fair coin?

$$SS = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT}\}$$

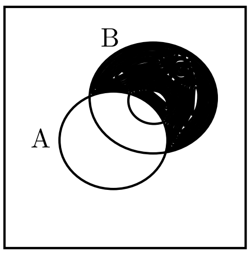
Exercise 1.6. For two coin flips, write the outcomes in the event "number of heads $\neq$ number of tails".

$$A = \{\text{HH}, \text{TT}\}$$

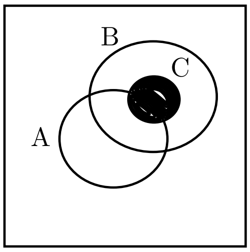
Exercise 1.7. For the following figure, shade the areas in the set.



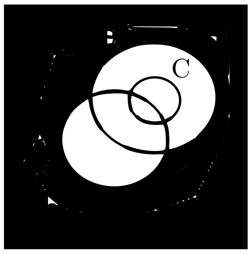
- $A^c \cap B$



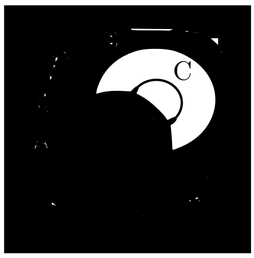
- $B \cap C$



- $(A \cup B)^c$



- $B^c \cup A$



Exercise 1.8. Let A and B be events in a sample space S. Write an expression for:

- A happened but B did not.

$$A \cap B^c$$

- Neither A or B happened.

$$(A \cup B)^c$$

- A and B both happened or both didn't happen.

$$(A \cap B) \cup (A^c \cap B^c)$$

2 LECTURE 1

Definition 2.1 (Probability Measure)**.** A *probability measure* on the sample space Ω is a function P from subsets of Ω to the real numbers that satisfies the following axioms:

1. $P(\Omega) = 1$ (*if this were not one, there are outcomes missing in* Ω)
2. If $A \subset \Omega$, then $P(A) \geq 0$ (*no notion of negative probabilities*)
3. If A_1 and A_2 are disjoint, then
$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$
More generally, if $A_1, A_2, \ldots, A_n, \ldots$ are mutually disjoint, then
$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Exercise 2.2. Prove: If A is an event, then $P(A^c) = 1 - P(A)$

- What makes a good proof?
  - Logic is clear to follow
  - Is concise
  - Not vague - no ambiguities
  - Built upon axioms or claims proven from axioms
  - Terms are well-defined - especially variables introduced
- How to come up with good proofs?
  - Start with axioms or claims proved from axioms, then find ways to combine them to reach the conclusion
  - Proofs by contradiction: assume the conclusion is wrong and then deduce that something known to be true is wrong
  - Try manipulating it algebraically to get terms that you want or can explain
  - Conceptually relate the things you are trying to prove. Why should it be true?

Proof. Suppose A is an event for the sample space Ω. By definition of the complement of an event,
$$A \cup A^c = \Omega.$$
Hence,
$$P(A \cup A^c) = P(\Omega).$$
Because A and A^c are disjoint, Axiom 3 implies $P(A \cup A^c) = P(A) + P(A^c)$. By Axiom 1, $P(\Omega) = 1$. Thus,
$$P(A) + P(A^c) = 1.$$
Thus, we conclude $P(A^c) = 1 - P(A)$. □

Exercise 2.3. If A and B are events, prove that $P(A \cap B) \leq P(A)$.

Proof. We decompose A into two disjoint regions $A = (A \cap B) \cup (A \cap B^c)$. By axiom 3
$$P(A) = P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c).$$
By axiom 2 $P(A \cap B^c) \geq 0$. Thus $P(A \cap B) \leq P(A)$. □

Exercise 2.4. If A and B are events, prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof.
$$A \cup B = (A \cap B^c) \cup (A \cap B) \cup (B \cap A^c) \tag{1}$$
$$\Rightarrow P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(B \cap A^c) \tag{2}$$
by axiom 3 and disjointness of the three regions. Similarly,
$$P(A) = P(A \cap B) + P(A \cap B^c) \tag{3}$$
$$P(B) = P(A \cap B) + P(B \cap A^c) \tag{4}$$
So,
$$RHS = P(A \cap B) + P(A \cap B^c) + P(B \cap A^c) + P(A \cap B) - P(A \cap B) \tag{5}$$
So, $LHS = RHS$. □

3 LECTURE 2

If every outcome in a sample space S is equally likely and A is an event,

$$P(A) = \frac{\#\text{ of outcomes in } A}{\#\text{ of outcomes in } S}$$

Definition 3.1 (Multiplication Principle)**.** If one experiment has m outcomes and another experiment has n outcomes, then there are mn possible outcomes for the two experiments.

Definition 3.2 (Extended Multiplication Principle)**.** If there are p experiments and the first has n_1 possible outcomes, the second $n_2, \ldots$, and the pth n_p possible outcomes, then there are a total of $n_1 \times n_2 \times \cdots \times n_p$ possible outcomes for the p experiments.

Suppose you have a pool of n objects. You select k of them. How many possible outcomes are there?

Exercise 3.3. There are 25 students in a class. 5 students will be selected to give a group presentation. Consider the group that was selected.

- Does order of selection matter? *No*
- Are the objects distinguishable? *Yes*
- Are you sampling with or without replacement? *Without*

Exercise 3.4. There are 6 people who will play three different board games in a row. To decide who goes first, a standard die is rolled. Consider the sequence of people who get to go first.

- Does order of selection matter? *Yes*
- Are the objects distinguishable? *Yes*
- Are you sampling with or without replacement? *With*

Exercise 3.5. There are 20 balloons, each makes a different sound when it pops. A cat will sequentially pop 5 balloons. Consider the sequence of sounds made.

- Does order of selection matter? *Yes*
- Are the objects distinguishable? *Yes*
- Are you sampling with or without replacement? *Without*

Exercise 3.6. Consider the rearrangements of the letters ABBA.

- Does order of selection matter? *Yes*
- Are the objects distinguishable? *No*
- Are you sampling with or without replacement? *Without*

Definition 3.7. The number of ways of rearranging k distinguishable objects is $k \cdot (k-1) \cdot (k-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1 = k!$.

Proposition 3.8. For a set of size n and a sample of size r, there are n^r different ordered samples with replacement and $n(n-1)(n-2)\cdots(n-r+1)$ different ordered samples without replacement.

Corollary 3.9. The number of orderings of n elements is $n(n-1)(n-2)\cdots 1 = n!$.

Exercise 3.10. How many ways can you rearrange the letters ABCD?

4!

Exercise 3.11. How many ways can you rearrange the letters ABCDEFG such that ABC appears in order?

5!

4 LECTURE 3

Example 4.1. If there are n people in a room, what are the odds that two share the same birthday?
Assumptions:

- 365 birthdays, each equally likely

The probability at least two people share a birthday is equal to 1 minus the probability all birthdays are distinct. What is the probability all n people share distinct birthdays (event A)?

$$P(A) = \frac{365!}{365 - n!} \cdot \frac{1}{365^n}$$

So, the probability that two share the same birthday (event B) is

$$P(B) = 1 - \frac{365!}{365 - n!} \cdot \frac{1}{365^n}$$

Proposition 4.2. The number of unordered samples of r objects selected from n objects without replacement is $\binom{n}{r}$.

$$\binom{n}{k} =_n C_k = \frac{n!}{k!(n-k)!}$$

Example 4.3. There is a room of 20 men and 30 women. A committee is formed by 2 men and 3 women. How many different committees are possible?

$$\binom{20}{2} \cdot \binom{30}{3}$$

Example 4.4. At a sandwich shop, you must select a bread, up to three toppings, and a protein. You may have at most one of each topping. There are 4 different choices for breads, 5 different choices for toppings, and 3 different choices for proteins. How many different sandwiches are possible?

$$4 \cdot (\binom{5}{3} + \binom{5}{2} + \binom{5}{1} + \binom{5}{0}) \cdot 3$$

Example 4.5. A room of 20 people is broken into 4 groups of 5. Then two of the groups are selected to prepare a presentation tomorrow. How many different possibilities are there for who presents tomorrow?
If order of presentation matters:

$$\binom{20}{5} \cdot \binom{15}{5}$$

If order of presentation does not matter:

$$\frac{\binom{20}{5} \cdot \binom{15}{5}}{2}$$

5 LECTURE 4

Proposition 5.1. The number of undordered samples of r objects selected from n objects without placement is $\binom{n}{r}$. The numbers $\binom{n}{k}$, called the *binomial coefficients*, occur in the expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

In particular,

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

The latter result can be interpreted as the number of subsets of a set of n objects. We just add the number of subsets of size 0 (with the usual convention that $0! = 1$), and the number of subsets of size 1, and the number of subsets of size 2, etc.

Proposition 5.2. The number of ways that n objects can be grouped into r classes with n_i in the ith class, $i = 1, \ldots, r$, and $\sum_{i=1}^{r} n_i = n$ is

$$\binom{n}{n_1! n_2! \cdots n_r!} = \frac{n!}{n_1!, n_2!, \cdots, n_r!}$$

Example 5.3. A committee of seven members is to be divided into three subcommittees of size three, two, and two. This can be done in

$$\binom{7}{322} = \frac{7!}{3!2!2!} = 210$$

ways.

Example 5.4.

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

$\binom{n}{k}$ is the number of subsets of size k from n items. Summing for all $k = 0, \ldots, n$ gets the total number of subsets of n things, of which there are 2^n subsets.

Example 5.5.

$$\binom{n}{k} = \binom{n}{n-k}$$

The number of ways selecting k items in the group is the same amount of ways of selecting $n-k$ members not in the group.

Example 5.6.

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

$\binom{n}{k}$ is the number of ways to select k items of n items. $\binom{n-1}{k-1}$ is the number of groups that contain item n. $\binom{n-1}{k}$ is the number groups that don't contain item n.

-	Order Matters	Order Doesn't Matter
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$
With Replacement	n^k	$\binom{k+n-1}{k}$

Exercise 5.7. In how many ways can you rearrange the symbols ******|||||?

$$\binom{11}{5} = \binom{11}{6}$$

Intuition: choosing 5 (or 6) places to put the pipes (or stars) into.

Exercise 5.8. There are n people in a room. I will draw a person at random with replacement k times. Each time, the person selected has to pay a dollar. How many different outcomes are there for the amounts paid by everyone?

(How many values of n whole numbers are there that add up to k? How many ways can you distribute k dollars to n people?)

$$\binom{k+n-1}{k}$$

Example 5.9. Roll 2 dice. What is the probability they add to 12? $\frac{1}{36}$
If you knew the first die was 5: 0
If you knew the first die was 6: $\frac{1}{6}$

Definition 5.10 (Conditional Probability)**.** The probability of some event A with the knowledge that an event B happened. Given that B happened, what is the probability of A?

Denoted $P(A|B)$.

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

Example 5.11. What in the probability that the card is an ace given that the card is a heart?

$$P(A|H) = \frac{P(A \cap H)}{P(H)} = \frac{\frac{1}{52}}{\frac{1}{4}} = \frac{1}{13}$$

Example 5.12. Roll a die twice. What is the probability the first is a 3 given the sum is 5?

$$P(3|S = 5) = \frac{P(3, 2)}{P(S = 5)} = \frac{\frac{1}{36}}{\frac{1}{9}} = \frac{1}{4}$$

Example 5.13. A family has two kids. Given that the younger kid is a girl, what is the probability the older kid is a girl.

$S = \{BB, BG, GB, GG\}$

$A = \{GB, GG\}$ event first is a girl.
$B = \{BG, GG\}$ event second is a girl.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

Example 5.14. A family has two kids. Given that one kid is a boy, what is the probability the other is a boy as well?

$S = \{BB, BG, GB, GG\}$

$A = \{BB, BG, GB\}$ event the family has a boy
$B = \{BB\}$ event the family has two boys

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

6 LECTURE 5

Definition 6.1 (Multiplication Law)**.** Let A and B be events and assume $P(B) \neq 0$. Then

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Example 6.2. There is an urn that contains 80 red balls and 20 green balls. You choose two balls. What is the probability you first get a red ball and then get a green ball?

Let A be the event that the 2nd ball is green, and B be the event that the 1st ball is red. We want to find $P(A \cap B)$.

$$\begin{aligned} P(B) &= \frac{80}{100} \\ P(A|B) &= \frac{20}{99} \\ P(A \cap B) &= \frac{80}{100} \cdot \frac{20}{99} \end{aligned}$$

Definition 6.3 (Generalized Multiplication Law)**.** For events $A_1, A_2, \ldots, A_n$, where $P(A) \neq 0$, then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_n|A_1, \ldots, A_{n-1}) \cdot P(A_{n-1}|A_1, \ldots, A_{n-2}) \cdots P(A_2|A_1) \cdot P(A_1)$$

Example 6.4. There is an urn that contains 80 red balls and 20 green balls. You choose three balls. Probability you get a red ball, then a green ball, then a red ball?

$$\begin{aligned} A_1 &= \text{ red ball first} \\ A_2 &= \text{ green ball second} \\ A_3 &= \text{ red ball third} \end{aligned}$$

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3|A_2 \cap A_1)P(A_2|A_1)P(A_1) \\ &= \frac{79}{98} \cdot \frac{20}{99} \cdot \frac{80}{100} \end{aligned}$$

Definition 6.5 (Law of Total Probability)**.** Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^n = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then, for any event A,

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$

Example 6.6. Toss a coin. If H, toss two more coins. If T, toss one more coin. What is the probability that exactly one H was tossed?

Let B_1 be the event that the first toss is heads, and B_2 be the event that first toss is tails.

$$\begin{aligned} (P(\text{ one head }) &= P(\text{ one head }|B_1)P(B_1) + P(\text{ one head }|B_2)P(B_2) \\ &= \frac{1}{4} + \frac{1}{2} \\ &= \frac{3}{4} \end{aligned}$$

Example 6.7. There is an urn that contains 80 red balls and 20 green balls. You pick a ball at random and then pick another ball at random without replacement. Intuitively, what is the probability that the sceond ball is green?

Let A be the event that the 2nd ball is green.

Let B_1 be the event that the first ball is red. Let B_2 be the event that the first ball is green.

Then,

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= \frac{20}{99} \cdot \frac{80}{100} + \frac{19}{99} \cdot \frac{20}{100} \\ &= \frac{20 \cdot 99}{99 \cdot 100} \\ &= \frac{20}{100} \end{aligned}$$

Example 6.8. There are three urns.

- Urn 1: 80 red balls and 20 green balls
- Urn 2: 50 red balls and 50 green balls
- Urn 3: 40 red balls and 60 green balls

You choose an urn at random and randomly choose a ball in it. What is the probability you choose a red ball?

Let A be the event that the ball is red.

Let B_1 be the event that the first urn is chosen, B_2 be the event that the second urn is chosen, and B_3 be the event that the third urn is chosen.

$$\begin{aligned} P(A) &= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + P(A|B_3) \cdot P(B_3) \\ &= \frac{80}{100} \cdot \frac{1}{3} + \frac{50}{100} \cdot \frac{1}{3} + \frac{40}{100} \cdot \frac{1}{3} \end{aligned}$$

7 LECTURE 6

Exercise 7.1. A company makes widgets. The company has three shifts of workers: morning, afternoon, night.

Time	Percent Widgets	Percent Mistakes
Morning	0.5	0.1
Afternoon	0.3	0.2
Night	0.2	0.05

$$\begin{aligned} P(D) &= P(D|M)P(M) + P(D|A)P(A) + P(D|N)P(N) \\ &= 0.5 \cdot 0.1 + 0.3 \cdot 0.2 + 0.2 \cdot 0.05 \\ &= 0.05 + 0.06 + 0.01 \\ &= 0.12 \end{aligned}$$

Example 7.2 (Monty Hall Problem)**.** There are three doors, with a car behind one and goats behind the others. You choose a door, then host opens one of the doors you did not choose, revealing a goat. Switch doors?

Stay strategy: Let A be the event that we win.
Let B_1 be the event we intially picked a goat.
Let B_2 be the event we initially picked a car.

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= 0 \cdot 0.6 + 1 * 0.3 \\ &= 0.3 \end{aligned}$$

Switch strategy: Let A be the event that we win.
Let B_1 be the event we intially picked a goat.
Let B_2 be the event we initially picked a car.

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \\ &= 1 \cdot 0.6 + 0 \cdot 0.3 \\ &= 0.6 \end{aligned}$$

Definition 7.3 (Bayes Theorem)**.**

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}$$

Example 7.4. A test determines if you have a disease, but it is not perfect. If you test positive, you want to know $P(\text{have disease} \mid \text{test positive})$. What is easier to measure is $P(\text{test positive} \mid \text{has disease})$.

Let P be the event we test positive.
Let D be the probability we have a disease.

$$\begin{aligned} P(P|D) &= 0.99 \\ P(P|D^c) &= 0.01 \\ P(D) &= 0.005 \end{aligned}$$

$$\begin{aligned} P(D|P) &= \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|H^c)P(H^c)} \\ &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} \\ &= \frac{0.00495}{0.0149} \end{aligned}$$

Definition 7.5 (Bayes Rule)**.** Let A and $B_1, \ldots, B_n$ be events where the B_i are disjoint, $\bigcup_{i=1}^n B_i = \Omega$, and $P(B_i) > 0$ for all i. Then,

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)}$$

Example 7.6. There are three urns.

- Urn 1: 99 red, 1 green balls
- Urn 2: 50 red, 50 green balls
- Urn 3: 1 red, 99 green balls

You choose a random urn and ball. It is red. What is the probability it came from the first urn?

$$
\begin{aligned}
P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A)} \\
&= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} \\
&= \frac{\frac{99}{100} \cdot \frac{1}{3}}{\frac{99}{100} \cdot \frac{1}{3} \cdot \frac{50}{100} \cdot \frac{1}{3} \cdot \frac{1}{100} \cdot \frac{1}{3}} \\
&= \frac{99}{150}
\end{aligned}
$$

Definition 7.7 (Independence of Random Events)**.** If A and B are events, then they are *independent* if $P(A \cap B) = P(A) \cdot P(B)$.

Example 7.8. Roll a die. Let A be the event the roll is odd. Let B be the event the roll is less than or equal to 3. Are A and B dependent?

$$
\begin{aligned}
P(A) &= \frac{1}{2} \\
P(B) &= \frac{1}{2} \\
P(A \cap B) &= \frac{2}{6} \\
\frac{2}{6} &\neq \frac{1}{2} \cdot \frac{1}{2}
\end{aligned}
$$

Therefore not independent.

Example 7.9. Roll a die twice. Let A be the event the first roll is odd. Let B be the event that the sum is odd. Are A and B independent?

$$
\begin{aligned}
P(A) &= \frac{1}{2} \\
P(B) &= \frac{1}{2} \\
P(A \cap B) &= \frac{1}{4} \\
\frac{1}{4} &= \frac{1}{2} \cdot \frac{1}{2}
\end{aligned}
$$

Therefore independent events.

Exercise 7.10. If A and B are disjoint, can they be independent?

$$
\begin{aligned}
P(A \cap B) &= P(A) \cdot P(B) \\
P(\emptyset) &= P(A) \cdot P(B) \\
0 &= P(A) \cdot P(B) \\
\iff P(A) = 0 \vee P(B) = 0
\end{aligned}
$$

Exercise 7.11. If A is a subset of B, can A and B be independent?

Exercise 7.12. Choose an integer at random from 1 to 99. Let A be the event the tens digit is a 1, and B be the event the ones digit is a 1.

$$P(A) = \frac{10}{99}$$
$$P(B) = \frac{10}{99}$$
$$P(A \cap B) = \frac{1}{99}$$
$$\frac{1}{99} \neq \frac{1}{10} \cdot \frac{10}{99}$$

Therefore not independent.

8 LECTURE 7

Exercise 8.1. Suppose you have as string of Christmas tree lights. The whole string fails when any light fails.

Suppose there are 100 lights and each independently fails with probabiilty 0.001.

Consider the following argument, and determine where it is wrong.

Let F be the event of failure, and L_i being the probability that the i_{th} light fails.

$$\begin{aligned} &= P(L_1 \cup L_2 \cup \cdots \cup L_{100}) \\ P(F) &= P(L_1) + P(L_2) + \cdots + P(L_{100}) \\ &= 0.001 \cdot 100 = 0.01 \end{aligned}$$

The false equivalence is that there is an assumption that L_i is disjoint from L_j, because the probability of a union is only the sum of their individual probabilities when they are disjoint, which they are not.

$$\begin{aligned} P(F) &= P(L_1 \cup L_2 \cup \ldots \cup L_{100}) \\ &= 1 - P(F^c) \\ &= 1 - P(A_1^c \cap A_2^c \cap \ldots \cap A_{100}^c) \end{aligned}$$

Exercise 8.2. Toss a coin until you get heads. Find:

- The probability of first heads on the first toss

$$P(H|T_1) = 0.5$$

- The probability of first heads on the i^{th} toss

$$P(H|T_i) = 0.5^i$$

- $P(\bigcup_{i=1}^{\infty} \{H_1|T_i\})$

$$P(\bigcup_{i=1}^{\infty} \{H_1|T_i\}) = \sum_{i=1}^{\infty} 0.5^i = 1$$

Important equation:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \text{ if } |r| < 1$$

9 LECTURE 8

Definition 9.1 (Random Variable)**.** In some random experiments, you want to know a particular resulting numerical value. A variable that depends on the outcome of a random experiment is a *random variable.*

Definition 9.2. Suppose that S is a finite or countably infinite sample space. Let p be a real-valued function defined for each element of S such that

1. $0 \leq p(s)$ for each $s \in S$
2. $\sum_{s \in S} p(s) = 1$

Then p is said to be a *discrete probability function.*

Definition 9.3. A function whose domain is a sample space S and whose values form a finite or countably infinite set of real numbers is called a *dicrete random variable.*

A random variable assigns a real value to every possible outcome.

Example 9.4. Flip a coin 2 times. Let $X =$ the number of heads.

We can make a table of all possible values of X and how likely they are to happen.

k	$P(X = k)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

X is a mapping

$$\begin{aligned} HH &\mapsto 2 \\ HT &\mapsto 1 \\ TH &\mapsto 1 \\ TT &\mapsto 0 \end{aligned}$$

Our discrete probability function has

$$\begin{aligned} 0 &\mapsto \frac{1}{4} \\ 1 &\mapsto \frac{1}{2} \\ 2 &\mapsto \frac{1}{4} \end{aligned}$$

Definition 9.5 (Binomial Random Variables)**.**

Exercise 9.6. There are 100 students in a class. On each day, each student independently goes to office hours with probability 0.01. The room for office hours can only hold 10 people. What is the probabilty the room cannot hold everyone that shows up?

Let X be the number of people who show up. X is a binomial random variable with probability $p = 0.01$, by assumption of the problem.

We want to find

$$P(X \geq 11) = \sum_{k=11}^{100} \binom{100}{k} 0.01^k \cdot 0.99^{100-k}$$

Definition 9.7. The *geometric distribution* is also constructed from independent Bernoulli trials, but from an infinite sequence. On each trial, a success occurs with probability p, and X is the total number of trials up to and including the first success. From the independence of the trials, this occurs with probability

$$p(k) = P(X = k) = (1 - p)^{k-1} p \text{ for } k = 1, 2, 3, \ldots$$

Note that these probabilities sum to 1.

$$\sum_{k=1}^{\infty} (1 - p)^{k-1} p = p \sum_{j=0}^{\infty} (1 - p)^j = 1$$

10 LECTURE 10

Definition 10.1. The *probability mass function* of X is $P(X) : \mathbb{R} \to \mathbb{R}$ where the range is $[0, 1]$.

Definition 10.2. The *expected value* of a random variable X is

$$E(X) = \mu_x = \sum_k kP_x(k)$$

and

$$\sum_{i=1}^{n} P_x(k) = 1.$$

If we evaluated X for many independent experiments, they would on average be $E[X]$.

Example 10.3. Toss a fair coin.

Let X be 1 if heads, otherwise 0 if tails. What is $E(X)$?

$$\begin{aligned} E(X) &= \sum_k kP(X = k) \\ &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \\ &= 1 \cdot \frac{1}{2} \\ &= \frac{1}{2} \end{aligned}$$

Example 10.4. Roll a fair die.

Let X be the value of the die. What is $E(X)$?

$$\begin{aligned} E(x) &= \sum_k kP(x = k) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} \\ &= 3.5 \end{aligned}$$

Exercise 10.5. Toss 2 fair coins. Let X be the number of heads. What is $E(X)$?

$$\begin{aligned} E(X) &= \sum_k kP(X = k) \\ &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) \\ &= 0 \cdot 0.25 + 1 \cdot 0.5 + 2 \cdot 0.25 \\ &= 1 \end{aligned}$$

Example 10.6. If X is a random variable and

$E(g(X)) = \sum_k g(k)P_X(k)$.

Let X be the value of a rolled die. Find $E(X^2)$.

$$\begin{aligned} E(X^2) &= \sum_k P(X = k) \\ &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot 1/6 \\ &= \end{aligned}$$

Theorem 10.7. If X is a random variable and a and b are constants, then

$$E[aX + b] = aE[x] + b.$$

Proof.

$$\begin{aligned}
E[aX+b] &= \sum_k (ak+b)\cdot P(X=k) \\
&= \sum_k akP(x=k) + \sum_k b\cdot P(X=k) \\
&= a\sum_k kP(x=k) + b\cdot \sum_k P(X=k) \\
&= aE[x]+b
\end{aligned}$$

□

Exercise 10.8. What is the expectation of a Binomial$[n,p]$ random variable?

$$\begin{aligned}
E(X) &= \sum_k kP(X=k) \\
&= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-P)^{n-k}
\end{aligned}$$

Definition 10.9 (Binomial Random Variable)**.** For $E_1,\ldots,E_n$ be independent events, each with probability P. Then

$$X = \sum_{i=1}^{n} I_{E_i} = I_{E_1} + I_{E_2} + \ldots + I_{E_n}$$

and

$$\begin{aligned}
E[X] &= E[I_{E_1} + \ldots + I_{E_n}] \\
&= E[I_{E_1}] + \ldots + E[I_{E_n}] \\
&= p + \ldots + p \\
&= np
\end{aligned}$$

Exercise 10.10. What is the expectation of a Geometric$[p]$ random variable?

$$\begin{aligned}
E(X) &= \sum_k kP(X=k) \\
&= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= p\sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= p\sum_{k=1}^{\infty} kr^{k-1} \\
&= p\sum_{k=1}^{\infty} \frac{d}{dr}r^k \\
&= p\frac{d}{dr}\sum_{k=1}^{\infty} r^k \\
&= p\frac{d}{dr}[\frac{1}{1-r}-1] \\
&= p(-(1-r)^{-2}(-1)) \\
&= \frac{p}{(1-r)^2} \\
&= \frac{p}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

Example 10.11 (The Umbrella Problem)**.** N people attend a party, each bringing an umbrella. Upon leaving, each person grabs a random umbrella. On average, how many people will leave with the umbrella they came with?

The indicator variable I_i is 1 if the i^{th} person grabs their own umbrella, otherwise 0.

Let $X = \sum_{i=1}^{n} I_i$ be a random variable.

$$\begin{aligned} E[X] &= E[\sum_{i=1}^{n} I_i] \\ &= \sum_{i=1}^{n} E[I_i] \\ &= \sum_{i=1}^{n} \frac{1}{n} \\ &= n \cdot \frac{1}{n} \\ &= 1 \end{aligned}$$

Definition 10.12 (Variance)**.** The *variance* of a random variable is the expected square deviation of the variable from its mean.

$$Var(X) = \sigma_x^2 = E((X - \mu)^2) = E(X^2) - \mu^2$$

Definition 10.13 (Standard Deviation)**.** $\mu(X) = \sqrt{Var(x)}$

Theorem 10.14. Claim: $Var(X) = E(X^2) - E(X)^2$

Proof. By definition $Var(X) = E[(X - E[X])^2]$

$$\begin{aligned} &= E[X^2 - 2XE[X] + E[X]^2) \\ &= E[X^2] - 2E[XE[X] + E[E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2E[1] \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - 2E[X]^2 \end{aligned}$$

□

Example 10.15. Roll a fair die. Let X be the value. Find $Var(X)$.

$$\begin{aligned} Var(X) &= E[X^2] - E[X]^2 \\ &= 3.5^2 \end{aligned}$$

Example 10.16. Flip a fair coin once. Let X be the number of heads. Find $Var(X)$.

$$\begin{aligned} Var(X) &= E[X^2] - E[X]^2 \\ &= 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} - (\frac{1}{2})^2 \\ &= \frac{1}{2} - (\frac{1}{2})^2 \\ &= \frac{1}{4} \\ \sigma(X) &= \sqrt{Var(X)} = \frac{1}{2} \end{aligned}$$

Theorem 10.17. If $Var(X)$ exists and $Y = a + bX$, then $Var(Y) = b^2Var(X)$.

Proof. Since $E(Y) = a + bE(X)$,

$$\begin{aligned} E[(Y - E(Y))^2] &= E[a + bX - a - bE(X)]^2 \\ &= Eb^2[X - E(X)]^2 \\ &= b^2E[X - E(X)]^2 \\ &= b^2Var(X) \end{aligned}$$

□

11 LECTURE 11

Exercise 11.1. What is the variance and standard deviation of a Binomial$[n, p]$ random variable?
Let $E_1, \ldots, E_n$ be independent events with probability $P(E_i) = p$.
Let I_{E_i} be the indicator variable of E_i.

$$E[X] = E\sum_{i=1}^{n}(I_{E_i}) = \sum_{i=1}^{n} E(I_{E_i}) = n \cdot p$$

Compute $E[X^2]$.

$$\begin{aligned} E[X^2] = E[(\sum_{i=1}^{n} I_{E_i})^2] &= E[(\sum_{i=1}^{n} I_{E_i})(\sum_{j=1}^{n} I_{E_j})] \\ &= E[\sum_{i=1}^{n}\sum_{j=1}^{n} I_{E_i} I_{E_j}] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} E[I_{E_i} I_{E_j}] \end{aligned}$$

We can find $E[I_{E_i} I_{E_j}] = p^2$. If $i \neq j \rightarrow E_i, E_j$ are independent. $i = j \rightarrow E[I_{E_i}, I_{E_j}] = p$, since they are dependent. So,

$$\begin{aligned} E[X^2] &= \sum_{i=1}^{n}\sum_{j=1}^{n} E[I_{E_i} I_{E_j}] \\ &= \sum_{i=1}^{n} E[I_{E_i} I_{E_j}] + \sum_{i=1}^{n}\sum_{j=1}^{n} E[I_{E_i} I_{E_j}] \\ &= n \cdot p + (n^2 - n)p^2 \\ &= np + n^2p^2 - np^2 \end{aligned}$$

$$Var(X) = E[X^2] - E[X]^2 = np - np^2 = np(1-p)$$

Definition 11.2 (Continuous Probability Functions)**.** Let S be a set of real numbers. The function $f(t)$ defines a probability function if $f(t) \geq 0$ for all t

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

The probability of the event A is $P(A) = \int_A f(t)dt$. We call f a *probability density function.*

Definition 11.3 (Uniform Random Variable)**.** A *uniform random variable* on $[a, b]$ returns a value between a and b where each point is equally likely.

$$f(t) = \{\frac{1}{b-a} \text{ if } a \leq t \leq b$$

Definition 11.4 (Normal Distribution)**.** A *normal distribution* has a mean μ and a standard deviation σ.

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(t-\mu)^2}{2\sigma^2}}$$

Exercise 11.5 (Two Numbers in a Hat)**.** In a hat are two distinct real numbers. You can reach into the hat and choose one of the numbers at random (equally likely). Your goal is to guess whether the other number is bigger or smaller than the number you picked. Come up with a way to do this such that your probability of being correct is bigger than 50%.

12 LECTURE 12

Exercise 12.1. Let $\lambda > 0$. For what value of c is $f(t) = ce^{-\lambda t}$ a probability density function over $[0, \infty)$.

$$\begin{aligned}
1 &= \int_0^\infty ce^{-\lambda t} \\
&= c\int_0^\infty e^{-\lambda t} \\
&= c\frac{e^{-\lambda t}}{-\lambda}[0, \infty) \\
&= c[0 + \frac{1}{\lambda}] \\
&= c \cdot \frac{1}{\lambda}
\end{aligned}$$

So, $c = \lambda$.

For what value of c is $f(t) = ce^{-\lambda t}$ a probability density function over $(-\infty, \infty)$.

Not possible.

Definition 12.2 (Cumulative Distribution Function)**.** If Y is a continuous random variable, it has a density $f_Y(y)$ such that

$$P(A \leq Y \leq b) = \int_a^b f_Y(y)dy$$

its *cumulative distribution function* is

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^y f_Y(y)dy$$

Definition 12.3 (Expected Value of Continuous Random Variables)**.** The expected value of a continuous random variable X is

$$E(X) = \mu_x = \int_{-\infty}^\infty x \cdot f_X(x)dx$$

13 LECTURE 13

Definition 13.1 (Joint Distribution of Two Continuous Random Variables)**.** If X and Y are jointly continuous random variables, there is a density $f_{X,Y}(x,y)$ such that for any region R in $x-y$ plane

$$P((x,y) \in R) = \int\int_R f_{X,Y}(x,y)dxdy$$

Example 13.2. Suppose (X,Y) is uniformly distributed on the square $[0,1] \times [0,1]$.

- What is the probability density function of (X,Y)?

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \le x \le 1 \wedge 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \int\int_{[0,1]\times[0,1]} f(x,y)dydx = 1 \\ = \int_0^1 \int_0^1 1dydx = 1 \end{aligned}$$

- What is $P(X \ge Y)$? Let $R_2 = \{(x,y) \text{ s.t. } x \ge y\}$.

$$\begin{aligned} P(X \ge Y) &= \int\int_{R_2} f(x,y)dydx \\ &= \int_0^1 \int_0^x 1dydx \\ &= \int_0^1 xdx \\ &= \frac{1}{2} \end{aligned}$$

Exercise 13.3. Let (X,Y) be a uniformly distributed point on the triangle given by points $(0,0),(0,1),(1,0)$. What is the probabliity density function of (X,Y)?

$$f_{X,Y}(x,y) = \begin{cases} c & 0 \le x \le 1 \wedge 0 \le y \le 1-x \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \int\int_R f(x,y)dxdy = 1 \\ c = 2 \end{aligned}$$

Definition 13.4 (Independence of Random Variables)**.** Two random variables X and Y with joint density $f_{X,Y}$ are *independent* if for any regions A and B,

$$P(X \in A \wedge Y \in B) = P(X \in A) \cdot P(Y \in B).$$

Equivalently, if

$$f_{X,Y} = g(X)h(Y)$$

for some g and h.

Exercise 13.5. Suppose X and Y are independent Uniform($[0,1]$) random variables. Compute the expected value of $min(X,Y)$.

1. What is joint density $f(x,y)$?

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(xy) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} 1 & 0 \le x \le 1 \wedge 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

2. What double integral to compute?

$$\begin{aligned}E[g(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy \\ &= \int\int_{[0,1]\times[0,1]} min(X,Y)\cdot 1dxdy \\ &= \int\int_{R_1} xdxdy + \int\int_{R_2} ydxdy \\ &= \int_0^1\int_0^1 xdydx + \int_0^1\int_0^1 ydydx \\ &= \int_0^1 x(1-x)dx + \int_0^1 \frac{1}{2}x^2dx\end{aligned}$$

Definition 13.6 (Expectation of Independent Random Variables)**.** If X and Y are independent, $E[XY] = E[X]E[Y]$ provided both $E[X]$ and $E[Y]$ exist.

$$\begin{aligned}E[XY] &= \sum_x\sum_y xyP_{X,Y}(x,y) \\ &= \sum_x\sum_y xyP_X(x)\cdot P_Y(y) \\ &= \sum_x xP_X(x)\sum_y yP_Y(y) \\ &= E[X]E[Y]\end{aligned}$$

Definition 13.7 (Covariance)**.** If X and Y are random variables, their *covariance* is

$$Cov(X,Y) = E[X\cdot Y] - E[X]\cdot E[Y]$$

Positive covariance means that larger values of X tends to happend with larger values of Y.

Negative covariance means that larger values of X tend to happend with smaller values of Y.

Zero covariance happens when variables are independent but can happen when X and Y are not independent.

Example 13.8. Let (X,Y) be a uniformly distributed point on the triangle given by points $(0,0),(0,1),(1,0)$.

$$\begin{cases}2 & (x,y)\in R \\ 0 & \text{otherwise}\end{cases}$$

$$\begin{aligned}E[X] &= \int\int_R x2dxdy \\ &= \int_0^1\int_0^{1-x} 2xdydx \\ &= \int_0^1 2x(1-x)dx \\ &= 2\int_0^1 (x-x^2)dx \\ &= \frac{1}{3} \\ E[Y] &= \frac{1}{3}\end{aligned}$$

$$
\begin{aligned}
E[XY] &= \int_0^1 \int_0^1 2xydydx \\
&= \int_0^1 2x \int_0^{1-x} ydydx \\
&= \int_0^1 2x\frac{1}{2}(1-x)^2 dx \\
&= \int_0^1 x(1-x)^2 dx \\
&= \int_0^1 x(1-2x+x^2)dx \\
&= \int_0^1 x - 2x^2 + x^3 dx \\
&= \frac{1}{12}
\end{aligned}
$$

So,

$$
\begin{aligned}
Cov(X,Y) &= E[XY] - E[X] \cdot E[Y] \\
&= \frac{1}{12} - \frac{1}{9} \\
&= \frac{-1}{36}
\end{aligned}
$$

Example 13.9. Let (X, Y) be a point chosen uniformly in the diamond $(1,0),(0,1),(-1,0),(0,-1)$.

- What is $Cov(X,Y)$?

$$
\begin{aligned}
E[X] &= 0 \text{ by symmetry} \\
E[Y] &= 0 \text{ by symmetry} \\
E[XY] &= 0 \text{ by symmetry} \\
Cov(X,Y) &= 0
\end{aligned}
$$

- Are X and Y independent?
  No.

$$
P(X \geq \frac{1}{2}) \cdot P(Y \geq \frac{1}{2}) \neq P(X \geq \frac{1}{2} \wedge Y \geq \frac{1}{2})
$$

14 LECTURE 14

Definition 14.1 (Variance of a Sum of Random Variables)**.**

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2ab\,Cov(X, Y)$$

Supposed $W_1, \ldots, W_n$ have finite variance.

$$Var(\sum_{i=1}^{n} a_iW_i) = \sum_{i=1}^{n} a_i^2Var(W_i) + 2\sum_{i<j} a_ia_jCov(W_i, W_j)$$

Definition 14.2 (Variance of Sum of Independent Random Variables)**.** If W_1 and W_2 are independent, then

$$Var(W_1 + W_2) = Var(W_1) + Var(W_2)$$

If $W_1, W_2, \ldots, W_n$ are mutually independent, then

$$Var(W_1 + W_2 + \ldots + W_n) = Var(W_1) + Var(W_2) + \ldots + Var(W_n)$$

Example 14.3. Variance of Binomial $[n, p]$.

$$\begin{aligned} E[X] &= np \\ Var[X] &= Var(X_1) + \ldots + Var(X_n) \\ &= nVar(X_1) = n(E[X_1^2] - E[X_1]^2) \\ &= n(p - p^2) \\ &= np(1 - p) \end{aligned}$$

Definition 14.4 (Standard Normal Distribution)**.** A *standard normal distribution* is a normal with mean 0 and variance 1. It has density $\frac{1}{\sqrt{2\pi}}e^{\frac{-X^2}{2}}$.

If $X\ N(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma}\ N(0, 1)$

15 LECTURE 15

Definition 15.1 (Marginal Density)**.** The *marginal density* of a random variable X given a joint distribution of X, Y with a joint density of $f_{X,Y}(x,y)$ is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy.$$

Exercise 15.2. If $X\ N(\mu_x, \sigma_x^2$ and $Y\ N(\mu_y, \sigma_y^2$ and X and Y are independent, find $X + Y\ N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

Definition 15.3 (Central Limit Theorem)**.** If $W_1, W_2, \ldots$ is an infinite sequence of indenpendent random variables with the same distribution, and $E[W_i] = \mu, Var(W_i) = \sigma^2$ are both finite, then

$$lim_{n\to\infty} P(a \leq \frac{W_1 + W_2 + \ldots + W_n}{\sqrt{n}\sigma} \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \epsilon^{\frac{-z^2}{2}}\,dz$$

So, averaging many independent samples of any random variables (subtracting off the mean and dividing by the variance), we get an approximately normal distribution.

Definition 15.4 (Moment Generating Function)**.** A *moment generating function* of a random variable X is $M(t) = E(e^{tX}$ if the expectation is defined. In the discrete case,

$$M(t) = \sum_x e^{tX} p(x)$$

and in the continuous case,

$$M(t) = \int_{-\infty}^{\infty} e^{tX} f(x)\,dx$$

If the moment-generating function exists in an open interval containing zero, then

$$M^{(r)}(0) = E(X^r).$$

Example 15.5. What is the moment generating function for a normal distribution?

$$\begin{aligned} M(t) = E(e^{tX}) &= \int_0^1 e^{tX} - 1\,dx \\ &= \frac{1}{t} e^{tX}|_0^1 \\ &= \frac{1}{t} e^t - \frac{1}{t} \\ &= \frac{1}{t}(e^t - 1) \end{aligned}$$

Example 15.6. What is the moment generating function for a standard normal distribution?

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tX} e^{\frac{-x^2}{2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2} + tX}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(X^2 - 2tX + t^2)}{2} + \frac{t^2}{2}}\,dx \\ &= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}}\,dx \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

Exercise 15.7. Let X be a geometric random variable.

- Find the moment generating function of X as an infinite sum

$$M(t) = \sum_{k=1}^{\infty} e^{tX} (1-p)^{k-1} p\,dx$$

- For what values of t is $M(t)$ defined?
  When $e^t(1-p) < 1$ it converges, so it is defined for

$$t < -\ln(1-p)$$

- Evaluate $M(t)$

16 LECTURE 16

Definition 16.1 (Parameter Estimation)**.** Given data, what probability model with what parameters is it best described by?

Example 16.2. You flip a coin 5 times. You get $HTHHT$. If the bias of the coin is p (that is, $P(H) = p$, what is the probability of $HTHHT$?

$$p^3(1-p)^2$$

What value of p maximizes this probability?

$$\begin{aligned} f(p) &= p^3(1-p)^2 \\ \ln f(p) &= = 3\ln p + 2\ln(1-p) \\ \frac{d}{dp}\ln f(p) &= \frac{3}{p} - \frac{2}{1-p} = 0 \\ \frac{3}{p} &= \frac{2}{1-p} \\ 2p &= 3(1-p) \\ 5p &= 3 \\ p &= \frac{3}{5} \end{aligned}$$

Example 16.3 (Maximum Likelihood Estimates)**.** If you have data $k_1, \ldots, k_n$ which are n independent random samples from a *pdf* with parameter θ, $P_X(k, \theta)$, then your estimate for θ is given by maximizing the likelihood of your data over all possible θ.

$$\underbrace{L(\theta)}_{\text{likelihood of data given } \theta} = \prod_{i=1}^{n} P_X(k_i, \theta)$$

and

$$\underbrace{\theta_e}_{\text{estimated } \theta} = \underbrace{argmax}_{\text{all } \theta} L(\theta)$$

Exercise 16.4. Suppose $X_1, X_2, \ldots, X_n$ are samples from a normal distribution with mean μ and standard deviation 1.

- What is the likelihood of $X_1, \ldots, X_n$?

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} f(X_i) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{\frac{-(X_i-\mu)^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}} n e^{\frac{-\sum_{i=1}^{n}(X_i-\mu)^2}{2}} \end{aligned}$$

- What is the derivative of the log of the likelihood?

$$\begin{aligned} \ln L(\mu) &= n \ln \frac{1}{\sqrt{2\pi}} - \frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^2 \\ \frac{d}{du}\ln L(\mu) &= \frac{-1}{2} \end{aligned}$$

- Set to 0 and solve.

Definition 16.5 (Maximum Likelihood with Multiple Unknown Parameters)**.** If $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ then $L(\theta) = L(\theta_1, \theta_2, \ldots, \theta_n)$. We find the maximizer of $L(\theta)$ by finding

$$\left\{ \frac{\delta L}{\delta \theta} = 0 \right.$$

17 LECTURE 17

Exercise 17.1. Find the maximum likelihood estimate for a and b if X Uniform$[a, b]$ and you have 3 independent samples $X_1 = 3$, $X_2 = 4$, $X_3 = 5$.

- What is the pdf of X?

$$f(X) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

- Find the likelihood

$$\begin{aligned} L(a, b) &= \prod_{i=1}^{3} f(x_i) \\ &= \begin{cases} (\frac{1}{b-a})^3 & a \le min(x_i) \le max(x_i) \le b \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

- Where is it maximized as a function of a and b?

$$\begin{aligned} b &= 5 = max(x_i) \\ a &= 3 = min(x_i) \end{aligned}$$

Definition 17.2 (Method of Moments)**.** If a probability model has s parameters $(\theta_1, \theta_2, \ldots, \theta_s)$, then estimate those parameters $\theta_{1e}, \theta_{2e}, \ldots, \theta_{se}$ by matching the first s moments of the random variable with the first s moments of the data.

Example 17.3. Suppose $f_Y(y; \theta) = \theta y^{\theta-1}$ for $0 \le y \le 1$, and four random samples are $y_1 = 0.42, y_2 = 0.10, y_3 = 0.65, y_4 = 0.23$. Estimate θ by the method of moments.

$$\begin{aligned} E[Y] &= \int_0^1 y \theta y^{\theta-1}\, dy \\ &= \theta \frac{y^{\theta+1}}{\theta+1}\Big|_0^1 \\ &= \frac{\theta}{\theta+1} \end{aligned}$$

Exercise 17.4. Suppose X Uniform$[a, b]$ and you have 3 independent samples $X_1 = 3, X_2 = 4, X_3 = 5$.

- Find the first moment of X

$$\begin{aligned} E[X] &= \int_a^b x \frac{1}{b-a}\, dx \\ &= \frac{b+a}{2} \end{aligned}$$

- Find the second moment of X

$$\begin{aligned} E[X^2] &= \int_a^b x^2 \frac{1}{b-a} \\ &= \frac{1}{3} x^3 \Big|_a^b \frac{1}{b-a} \\ &= \frac{b^3 - a^3}{b-a} \frac{1}{3} \\ &= \frac{b^2 + ab + a^2}{3} \end{aligned}$$

- Find the first moment of the data X_1, X_2, X_3

$$\frac{X_1 + X_2 + X_3}{3} = 4$$

- Find the second moment of the data

$$\frac{3^2 + 4^2 + 5^2}{3} = \frac{50}{3}$$

- Write two simultaneous equations that give estimates for a, b using the method of moments.

$$4 = \frac{a+b}{2}$$
$$\frac{50}{3} = \frac{b^2 + ab + a^2}{3}$$

Definition 17.5 (Estimator)**.** An *estimator* is a function of your data that provides an estimate for an unknown parameter of your probability model.

What makes an estimator good?

- Unbiased ($E\hat{\theta} = \theta$
- Efficiency/low variance (small $Var(\hat{\theta})$)

Exercise 17.6. Suppose X_1 Uniform$[0, \theta]$

- Given only X_1 what is MLE of θ?

$$\hat{\theta} = X_1$$

- Given only X_1, what is the method of moment estimate of θ?

$$\hat{\theta} = 2X_1$$
$$E[X_1] = \frac{\theta}{2} = \mu_1$$

- Which is unbiased?

$$E[\hat{\theta}_{MLE}] = E[X_1] = \frac{\theta}{2} \text{ – biased}$$
$$E[\hat{\theta}_{MOM}] = E[2X_1] = 2 \cdot \frac{\theta}{2} = \theta \text{ – unbiased}$$

18 LECTURE 18

Definition 18.1. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of θ, and if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, we say that $\hat{\theta}_1$ is more *efficient* than $\hat{\theta}_2$.

Example 18.2. Let $X_1, X_2 \quad$ Normal$(\mu, 1)$.

$$\begin{aligned} \text{Let } \hat{\mu}_1 &= X_1 \\ \hat{\mu}_2 &= 2X_1 - X_2 \\ \hat{\mu}_2 &= \frac{1}{2}X_1 + \frac{1}{2}X_2 \end{aligned}$$

- Are these biased?

$$\begin{aligned} &E[\hat{\mu}_1] = E[X_1] = \mu \text{ unbiased} \\ &E[\hat{\mu}_2] = 2E[X_1] - E[X_2] = 2\mu - \mu = \mu \text{ unbiased} \\ &E[\hat{\mu}_3] = \mu \text{ unbiased} \end{aligned}$$

- Which is most efficient?

$$\begin{aligned} &Var(\hat{\mu}_1) = 1 \\ &Var(\hat{\mu}_2) = 4 \cdot 1 + 1 = 5 \\ &Var(\hat{\mu}_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \text{ most efficient} \end{aligned}$$