Probability Theory Lecture Notes

Phanuel Mariano

Contents

CHAPTER 1

Combinatorics

1.1. Counting Principle

- We need a way to help us count faster rather than counting by hand one by one.

FACT. (**Basic Counting Principle**) *Suppose 2 experiments are to be performed.*
If one experiement can result in m possibilities
Second experiment can result in n possibilities
Then together there are mn possibilities

- I like to use the box method. For example. Each box represent the number of possibilities in that experiement.
- **Example1:** There are 20 teachers and 100 students in a school. How many ways can we pick a teacher and student of the year?
  - **Solution:** Use the box Method: $20 \times 100 = 2000$.
- The counting principle can be generalized to any amount of experiments: $n_1 \cdots n_r$ possibilities
- **Example2:**
  - A college planning committee consists of 3 freshmen, 4 sophomores, 5 juniors, and 2 seniors.
  - A subcomittee of 4 consists 1 person from each class. How many?
  - **Solution:** Box method $3 \times 4 \times 5 \times 2 = 120$.
- **Example3:** How many differen $6-$place license plates are possible if the first 3 places are to be occupied by letters and the finals 3 by numbers?
  - **Solution:** $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 =$?
  - Question: What if no repetition is allowed?
  - **Solution:**$26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8$
- **Example4:** How many functions defined on n points are possible if each functional value is either 0 or 1.
  - **Solution:** Box method on the $1, \ldots, n$ points gives us 2^n possible functions.

1.2. Permutations

- How many different ordered arrangements of the letters a, b, c are possible?
  - abc, acb, bac, bca, cab Each arrangement is a permutation
  - Can also use the box method to figure this out: $3 \cdot 2 \cdot 1 = 6$.

FACT. *With n objects. There are*

$$n(n-1)\cdots 3 \cdot 2 \cdot 1 = n!$$

different permutations of the n objects.

($\star$) *Note that ORDER matters when it comes to Permutations*

- **Example1:** What is the numnber of possible batting order with 9 players?
  - **Answer:** 9!(Box Method or permutations)
- **Example2:** How many ways can one arrange 4 math books, 3 chemistry books, 2 physics books, and 1 biology book on a bookshelf so that all the math books are together, all the chemistry books are together, and all the physics books are together.
  - **Answer:** We can arrange the math books in 4! ways, the chemistry in 3! ways, the physics in 2! ways, and B in 1! $= 1$ way.
  - But we also have to decide which set of books go on the left, which next, and so on. That is the same as the number of ways of arranging the letters M,C, P,B, and there are 4! ways of doing that. $MCPB$, $PBPB$ ect..
  - So 4! (4!3!2!1!) ways.
- **Example3:** Repetitions: How many ways can one arrange the letters a, a, b, c?
  - Let us label them A, a, b, c. There are 4!, or 24, ways to arrange these letters. But we have repeats: we could have Aa or aA. So we have a repeat for each possibility (**so divide!!!**), and so the answer should be $4!/2! = 12$.
  - If there were 3 a's, 4 b's, and 2c's, we would have

$$\frac{9!}{3!4!2!}$$

- **Example4:** How many different letter arrangements can be formed from the word PEPPER?
  - **Answer:** There 3 P's 2 E's and one R. So $\frac{6!}{3!2!1!} = 30$.

FACT. *There are*

$$\frac{n!}{n_1! \cdots n_r!}$$

different permutations of n objects of which n_1 are alike, n_2 are alike, n_r are alike.

- **Example4:** Suppose there are 4 Czech tennis players, 4 U.S. players, and 3 Russian players, in how many ways could they be arranged?
  - Answer: $\frac{11!}{4!4!3!}$.

1.3. Combinations

- We are often interested in selecting r objects from a total of n objects.
- How many ways can we choose 3 letters out of 5? (Does order matter here? NO)If the letters are a, b, c, d, e then there would be 5 for the first position, 4 for the second, and 3 for the third, for a total of $5 \times 4 \times 3$. But order doesn't matter here. So we're over counting here....
  - But suppose the letters selected were a, b, c. If order doesn't matter, we will have the letters a, b, c $3! = 6$ times, because there are $3!$ ways of arranging a group of 3. The same is true for any choice of three letters. So we should have
    $$\frac{5 \cdot 4 \cdot 3}{3!} = \frac{5!}{3!2!} = 10.$$
    Or what we did was $5 \cdot 4$, or $n(n-1) \cdots (n-r+1)$ then divided by the repeats $3!$.
  - This is often written $\begin{pmatrix} 5 \\ 3 \end{pmatrix}$, read "5 choose 3". More generally..

FACT. *If $r \leq n$, then*
$$\begin{pmatrix} n \\ r \end{pmatrix} = \frac{n!}{(n-r)!r!}$$
*and say n choose r, represents the number of possible **combinations** of objects taken r at a time.*
($\star$) Order DOES NOT Matter here

- Recall in Permutations order did matter.
- **Example1:** How many ways can one choose a committee of 3 out of 10 people?
  - Answer: $\begin{pmatrix} 10 \\ 3 \end{pmatrix} = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2} = 10 \cdot 3 \cdot 4 = 120.$
- **Example2:** Suppose there are 9 men and 8 women. How many ways can we choose a committee that has 2 men and 3 women?
  - **Answer:** We can choose 2 men in $\begin{pmatrix} 9 \\ 2 \end{pmatrix}$ ways and 3 women in $\begin{pmatrix} 8 \\ 3 \end{pmatrix}$ ways. The number of committees is then the product $\begin{pmatrix} 9 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 8 \\ 3 \end{pmatrix}$.
- **Example3:** A person has 8 friends, of whom 5 will be invited to a party. (We've all been through this)
  - **(a)** How many choices are there if 2 of the friends are feuding and will not attend together?
    * Box it: $[none] + [$ one of them$][others]$
    * $\begin{pmatrix} 6 \\ 5 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 4 \end{pmatrix}$ (recall that when we have OR, use +)
  - **(b)** How many choices if 2 of the friends will only attend together?
    * Box it: $[none] + [$with both$]$
    * $\begin{pmatrix} 6 \\ 5 \end{pmatrix} + 1 \cdot 1 \cdot \begin{pmatrix} 6 \\ 3 \end{pmatrix}$
- The value of $\begin{pmatrix} n \\ r \end{pmatrix}$ are called binomials coefficients because of their prominence in the binomial theorem.

THEOREM. *(The Binomial Theorem)*
$$(x+y)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k y^{n-k}.$$

PROOF. To see this, the left hand side is $(x+y)(x+y)\cdots(x+y)$. This will be the sum of 2^n terms, and each term will have n factors. How many terms have k x's and $n-k$ y's? This is the same as asking in a sequence of n positions, how many ways can one choose k of them in which to put x's? (Box it) The answer is $\begin{pmatrix} n \\ k \end{pmatrix}$, so the coefficient of $x^k y^{n-k}$ should be $\begin{pmatrix} n \\ k \end{pmatrix}$. $\square$

- **Example:** Expand $(x+y)^3$.
    - **Solution:** $(x+y)^3 = y^3 + 3xy^2 + 3x^2y + x^3$.
- **Problem:** Using Combinatorics: Let's prove

$$\begin{pmatrix} 10 \\ 4 \end{pmatrix} = \begin{pmatrix} 9 \\ 3 \end{pmatrix} + \begin{pmatrix} 9 \\ 4 \end{pmatrix}$$

  with no algebra:
    - The LHS represents the number of committees having 4 people out of the 10.
    - Let's say the President of the university will be in one of these committees and he's special, so we want to know when he'll be there or not.
    - When he's there, then there are $1 \cdot \begin{pmatrix} 9 \\ 3 \end{pmatrix}$ is the number of ways that contain the President while $\begin{pmatrix} 9 \\ 4 \end{pmatrix}$ is the number of comittees that do not contain the President and contain 4 out of the remaining people.
- The more general equation is

$$\begin{pmatrix} n \\ r \end{pmatrix} = \begin{pmatrix} n-1 \\ r-1 \end{pmatrix} + \begin{pmatrix} n-1 \\ r \end{pmatrix}$$

1.4. Multinomial Coefficients

- **Example:** Suppose one has 9 people and one wants to divide them into one committee of 3, one of 4, and a last of 2. How many different ways are there?
    - **Solution:** (Box it) There are $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$ ways of choosing the first committee. Once that is done, there are 6 people left and there are $\begin{pmatrix} 6 \\ 4 \end{pmatrix}$ ways of choosing the second committee. Once that is done, the remainder must go in the third committee. So there is 1 one to choose that. So the answer is
    $$\frac{9!}{3!6!} \frac{6!}{4!2!} = \frac{9!}{3!4!2!}.$$
- **In general:** Divide n objects into one group of n_1, one group of n_2, ... and a kth group of n_k, where $n = n_1 + \cdots + n_k$, the answer is there are
$$\frac{n!}{n_1! n_2! \cdots n_k!} \text{ ways.}$$
- These are known as multinomial coefficients. We write them as
$$\begin{pmatrix} n \\ n_1, n_2, \ldots, n_k \end{pmatrix} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$
- **Example:** Suppose we are to assign Police officers their duties . Out of 10 officers: 6 patrols, 2 in station, 2 in schools.
    - **Answer**: $\frac{10!}{6!2!2!}$.
- **Example:** There are 10 flags:5 indistinguishable Blue flags, 3 indistinguishable Red flags, and 2 indistinguishable Yellow flags. How may different ways can we order them on a flag pole?
    - **Answer:** $\frac{10!}{5!3!2!}$.
- **Example:** Suppose one has 8 indistinguishable balls. How many ways can one put them in 3 boxes?
    - **Solution1:** Let us make sequences of o's and $|$'s; any such sequence that has $|$ at each side, 2 other $|$'s, and 8 o's represents a way of arranging balls into boxes. For example, if one has
    $$| \, oo \, | \, ooo \, | \, ooo \, | \, .$$
    - How many different ways can we arrange this where we have start with $|$ and end with $|$. In between, we are only arranging $8 + 2 = 10$ symbols, of which only 8 are o's
    - So the question is: How many ways out of **10 spaces** can one pick **8 of them** into which to put an o?
   - $\begin{pmatrix} 10 \\ 8 \end{pmatrix}$.
    - **Solution2:** Look at spaces between. There are 9 spaces. So $\begin{pmatrix} 9 \\ 2 \end{pmatrix} + 9$.

CHAPTER 2

Axioms of Probability

2.1. Sample Space and Events

- We will have a **sample space**, denoted S (sometimes Ω, or $\mathcal{U}$) that consists of all possible outcomes from an experiment.
  - **Example1:**
    - ∗ <u>Experiment</u>: Roll two dice,
    - ∗ <u>Sample Space</u>: $S =$ would be all possible pairs made up of the numbers one through six. List it here.$\{(i,j) : i, j = 1, \ldots 6\}$. 36 points.
  - **Example 2:**
    - ∗ <u>Experiment</u>: Toss a coin twice
    - ∗ $S = \{HH, HT, TH, TT\}\}$
  - **Example3:**
    - ∗ <u>Experiment</u>: Measuring the number of accidents of a random person before they had turn 18.
      - · $S = \{0, 1, 2, \ldots\}$
  - **Others:**
    - ∗ Let S be the possible orders in which 5 horses finish in a horse race;
    - ∗ Let S be the possible price of some stock at closing time today; or $S = [0, \infty)$;
    - ∗ The age at which someone dies, $S = [0, \infty)$.
- **Events:** An event A is a subset of S. In this case we use the notation $A \subset S$, to mean A is a subset of S.
  - $A \cup B$: points in S such that is in A OR B OR BOTH.
  - $A \cap B$, points in A AND B. (you may also see AB)
  - A^c is the compliment of A, the points NOT in A. (you may also see A')
  - Can extend to $A_1, \ldots, A_n$ events. $\bigcup_{i=1}^{n} A_i$ and $\bigcap_{i=1}^{n} A_i$.
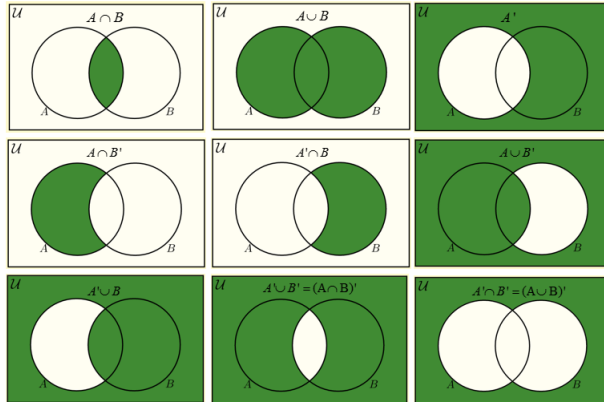
- **Example1**: Roll two dice.
  - Example of an Events
  - $E =$ the two dies come up even and equal $\{(2,2),(4,4),(6,6)\}$
  - $F =$ the sum of the two dice is 8. $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$.
  - $E \cup F = \{(2,2),(2,6),(3,5),(4,4),(5,3),(6,2),(6,6)\}$
  - $E \cap F = \{(4,4)\}$.
  - F^c all the 31 other ways that does not include $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$.
- **Example2:** $S = [0,\infty)$ age someone dies.
  - Event $A =$ person dies before they reached 30.
    * $A = [0,30)$.
  - Interpret $A^c = [30,\infty)$
    * The person dies after they turned 30.
  - $B = (15,45)$. Do $A \cup B, A \cap B$ and so on.
- **Properties:** Events also have commutative and associate and Distributive laws.
- What is $A \cup A^c$? $= S$.
- **DeMorgan's Law:**
  - $(A \cup B)^c = A^c \cap B^c$ .Try to draw a picture
  - $(A \cap B)^c = A^c \cup B^c$.
  - This works for general $A_1, \ldots, A_n$: $\left(\cup_{i=1}^n A_i\right)^c = \cap_{i=1}^n A_i^c$ and $\left(\cap_{i=1}^n A_i\right)^c = \cup_{i=1}^n A_i^c$.
- The empty set $\emptyset = \{\}$ is the set that has nothing in it.
- A and B are **disjoint** if $A \cap B = \emptyset$.
  - In Probability we may say that events A and B are "mututally exclusive" if they are disjoint.
  - mutually exclusive means the same thing as disjoint

2.2. Axioms of Probability

- Let E be an event. How do we defined the probability of an event?
  - We can attempt to define a probability by the relative frequency,
  - Perform an experiment (e.g. Flipping a coin)
  - Perform that experiment n times and let $n(E) =$ the number of times the event occured in n repetitions
    * (e.g. Flip a coin $n = 1000$ times, and let's say that $n(\{Tails\}) = 551$) Then it's reasonanble to think $\mathbb{P}(\{Tails\}) \approx \frac{551}{1000}$
  - So maybe we can define the probability of an event as $\mathbb{P}(E) = \lim_{n \to \infty} \frac{n(E)}{n}$. But we don't know if this limit exists, or if $n(E)$ is even well defined!!!
  - So we need a new approach.
- Probability will be a rule given by the following <u>Axioms</u> (Laws that we all agree on)
  - A probability will be a function $\mathbb{P}(E)$ where the input is a set/event such that
  - **Axiom 1:** $0 \leq \mathbb{P}(E) \leq 1$ for all events E.
  - **Axiom 2:** $\mathbb{P}(S) = 1$.
  - **Axiom 3:** (disjoint property) If the events $E_1, E_2, \ldots$ are pairwise disjoint/mutually exclusive then
  $$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$
    * <u>Mutually exclusive</u> means that $E_i \cap E_j = \emptyset$ when $i \neq j$.
- **Remark:** Note that you take a probability of a subset of S, not of points of S. However it is common to write $P(x)$ for $P(\{x\})$.
  - Say if the experiment is tossing a xoin. Then $S = \{H, T\}$. The probability of heads should be written as $\mathbb{P}(\{H\})$, but it is common to see $\mathbb{P}(H)$.
- **Example1:**
  - (a) Suppose we toss a coin and they are equally likely then $S = \{H, T\}$ and
    * $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$. We may write $\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$.
  - (b) If biased coin is tosse then one could have a different assignment of probability $\mathbb{P}(H) = \frac{2}{3}, \mathbb{P}(T) = \frac{1}{3}$.
- **Example2:**
  - Rolling a fair die, the probability space consists of $S = 1, 2, 3, 4, 5, 6$, each point having probability $\frac{1}{6}$.
  - We can compute the probability of rolling an even number by
  $$\begin{aligned} \mathbb{P}(\{even\}) &= \mathbb{P}(\{2, 4, 6\}) \\ &= \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = \frac{1}{2} \end{aligned}$$
  where we used the rules of probability by breaking it down into a sum.

PROPOSITION 1. *(a)* $\mathbb{P}(\emptyset) = 0$
(b) If $A_1, \ldots, A_n$ are pairwise disjoint, $\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.
(c) $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
(d) If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.
(e) $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

- It helps to draw diagrams to prove these.

• Try to prove at least some of these yourself.

PROOF. (a) Let $A_i = \emptyset$ for each i which are disjoint. So

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

since this would be infinite sum so that $\mathbb{P}\left(\emptyset\right) = 0$ since $0 \leq \mathbb{P}\left(\emptyset\right) \leq 1$.

(b) Let $A_{n+1} = A_{n+2} = \cdots = \emptyset$ so that $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{n} A_i$ hence

$$\mathbb{P}\left(\cup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right)$$
$$= \sum_{i=1}^{n} \mathbb{P}\left(A_i\right) + \sum_{n=1}^{\infty} \mathbb{P}\left(\emptyset\right)$$
$$== \sum_{i=1}^{n} \mathbb{P}\left(A_i\right) + \sum_{n=1}^{\infty} 0$$
$$= \sum_{i=1}^{n} \mathbb{P}\left(A_i\right)$$

(c) Use $S = E \cup E^c$. By Axiom (2) we have

$$1 = \mathbb{P}\left(S\right) = \mathbb{P}\left(E\right) + \mathbb{P}\left(E^c\right),$$

hence $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.

(d) If $E \subset F$, then write $F = E \cup (F \cap E^c)$ thus since this is disjoint

$$\mathbb{P}\left(F\right) = \mathbb{P}\left(E \cup (F \cap E^c)\right) = \mathbb{P}\left(E\right) + \mathbb{P}\left(F \cap E^c\right) \geq \mathbb{P}\left(E\right) + 0 = \mathbb{P}\left(E\right).$$

(e) Write $E \cup F = E \cup (E^c \cap F)$, (Picture of venn diagram of both)hence by disjointness again

$$\mathbb{P}\left(E \cup F\right) = \mathbb{P}\left(E\right) + \mathbb{P}\left(E^c \cap F\right).$$

Now write F(with picture) as $F = (E \cap F) \cup (E^c \cap F)$ and using disjointness

$$\mathbb{P}\left(F\right) = \mathbb{P}\left(E \cap F\right) + \mathbb{P}\left(E^c \cap F\right) \implies \mathbb{P}\left(E^c \cap F\right) = \mathbb{P}\left(F\right) - \mathbb{P}\left(E \cap F\right),$$

substitute into first equation to get

$$\begin{aligned}
\mathbb{P}\left(E \cup F\right) &= \mathbb{P}\left(E\right) + \mathbb{P}\left(E^c \cap F\right) \\
&= \mathbb{P}\left(E\right) + \mathbb{P}\left(F\right) - \mathbb{P}\left(E \cap F\right),
\end{aligned}$$

as needed.   □

• **Example:** Uconn Basketball is playing Kentucky this year.
    – Home game has .5 chance of winning
    – Away game has .4 chance of winning.
    – .3 that uconn wins both games.
    – What's the probability that Uconn loses both games?
    – **Answer.**
        * Let $\mathbb{P}\left(A_1\right) = .5$, $\mathbb{P}\left(A_2\right) = .4$ and $\mathbb{P}\left(A_1 \cap A_2\right) = .3$.
        * We want to find $\mathbb{P}\left(A_1^c \cap A_2^c\right)$. Simplify as much as we can:

$$\begin{aligned}
\mathbb{P}\left(A_1^c \cap A_2^c\right) &= \mathbb{P}\left((A_1 \cup A_2)^c\right) \text{ by DeMorgan's Law} \\
&= 1 - \mathbb{P}\left(A_1 \cup A_2\right), \text{ by Proposition 1c}
\end{aligned}$$

∗ Using Proposition 1e, we have
$$\mathbb{P}\left(A_1 \cup A_2\right) = .5 + .4 - .3 = .6,$$
Hence $\mathbb{P}\left(A_1^c \cap A_2^c\right) = 1 - .6 = .4$ as needed.

2.3. Equally Likely Outcomes

- In many experiments, a probability space consists of finitely many points, all with equally likely probabilities.
    - Basic example was a tossing a coin $P(H) = P(T) = \frac{1}{2}$
    - Fair die: $P(i) = \frac{1}{6}$ for $i = 1, \dots, 6$.
- In this case from Axiom 3 we have that
$$P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}.$$
- **Example1:** What is the probability that if we roll 2 dice, the sum is 7?
    - **Answer:** There are 36 total outcomes , of which 6 have a sum of 7:
        * $E =$ "sum is 7" $= \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$. Since they are all equally likely, the probability is $\mathbb{P}(E) = \frac{6}{6 \cdot 6} = \frac{1}{6}$.
- **Example 2:** If 3 balls are "randomly drawn" from a bowl containing 6 white and 5 black balls, what is the probability that one ball is white and the other two are black?
    - **Method 1:** (regard as a ordered selection)
$$
\begin{aligned}
P(E) &= \frac{WBB + BWB + BBW}{11 \cdot 10 \cdot 9} \\
&= \frac{6 \cdot 5 \cdot 4 + 5 \cdot 6 \cdot 4 + 5 \cdot 4 \cdot 6}{990} = \frac{120 + 120 + 120}{990} = \frac{4}{11}.
\end{aligned}
$$
    - **Method2:** (Regard as unordered set of drawn balls)
$$
P(E) = \frac{(1 \text{ white}) (2 \text{ black})}{\binom{11}{3}} = \frac{\binom{6}{1} \binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}.
$$

- We can always choose which way to regard our experiements.
- **Example 3** A committee of 5 is to selected from a group of 6 men and 9 women. What is probability consistsd of 3 men and 2 women
    - Answer: Easy $\frac{men \cdot women}{all} = \frac{\binom{6}{3} \binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}$.

- **Example 4:** Seven balls are randomly withdrawn from an urn that contains 12 red, 16 blue, and 18 green.
    - (b) Find probability that "**at least** 2 red balls are withdrawn;"
    - **Ans:** Let E be this event then $P(E) = 1 - P(E^c)$, $P(\text{at least 2 red}) = 1 - \mathbb{P}(\text{drawing 0 or 1 balls})$. Now
$$
\mathbb{P}(\text{drawing 0 or 1 red balls}) = \frac{\binom{16 + 18 = 34}{7}}{\binom{46}{7}} + \frac{\binom{12}{1} \binom{34}{6}}{\binom{46}{7}}.
$$

- Explanation of Poker/Playing cards : Ranks and suits,etc!
    - There are 52 cards in a standard deck of playing cards. The *poker hand* is consists of five cards. There are 4 *suits*: heats, spades, diamonds, and clubs ($\heartsuit \spadesuit \diamondsuit \clubsuit$). The suits diamonds

and hearts are red while clubs and spades are black. In each suit there are 13 *ranks*: the numbers $2, 3 \ldots, 10$, the face cards, Jack, Queen, King, and the Ace(not a face card).

- **Example 5**: What is the probability that in a poker hand (5 cards out of 52) we get exactly 4 of a kind?
  - **Answer:** Consider 4 aces and 1 king: $AAAK = \begin{pmatrix} 4 \\ 4 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix}$. But $JJJJ3$ is the same probability.
    * Thus there are 13 ways to pick the first rank, and 12 ways to pick the second rank

$$
\begin{aligned}
P\left(4 \text{ of a kind}\right) &= \frac{[\text{choice of ranks}]\,[\text{given rank how to choose a hand}]}{\begin{pmatrix} 52 \\ 5 \end{pmatrix}} \\
&= \frac{\left(13 \cdot 12 \cdot \begin{pmatrix} 4 \\ 4 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix}\right)}{\begin{pmatrix} 52 \\ 5 \end{pmatrix}} \approx .0000139
\end{aligned}
$$

- **Example 6**: What is the probability that in a poker hand (5 cards out of 52) we get a straight. (no straight flushes, can't be of the same suit)
  - **Answer:** Consider: A-2-3-4-5-6-7-8-9-10-J-Q-K-A- There are 10 possible straights.
    * Given a straight Say A2345: There are $4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 - (\text{of the same suit}) = 4^5 - 4$.

$$
\begin{aligned}
P\left(\text{Straight}\right) &= \frac{[\text{choice of straight}]\,[\text{given striaght how to choose a hand}]}{\begin{pmatrix} 52 \\ 5 \end{pmatrix}} \\
&= \frac{10 \cdot \left(4^5 - 4\right)}{\begin{pmatrix} 52 \\ 5 \end{pmatrix}} \approx .0039
\end{aligned}
$$

- **Example 7**: What is the probability that in a poker hand (5 cards out of 52) we get a Full House. (3 and a 2 of a kind)
  - **Answer:** It would be [3 of a kind][2 of a kind]. AAAKK or KKAAA are different!Choose suit: $13 \cdot 12$.
  - Then once we choose within each group there

$$
\begin{aligned}
P\left(\text{Full House}\right) &= \frac{[\text{choice of rank}]\,[3 \text{ of a kind}]\,[2 \text{ of a kind}]}{\begin{pmatrix} 52 \\ 5 \end{pmatrix}} \\
&= \frac{13 \cdot 12 \begin{pmatrix} 4 \\ 3 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \end{pmatrix}}{\begin{pmatrix} 52 \\ 5 \end{pmatrix}} \approx .0014.
\end{aligned}
$$

- **Example 8:** (Birthday Problem) In a class of 32 people, what is the probability that at least two people have the same birthdays? (We assume each day is equally likely.)
  - Answer: Let the first person have a birthday on some day. The probability that the second person has a different birthday will be $\frac{364}{365}$. The probability that the third person has a

different birthday from the first two people is $\frac{363}{365}$. So the answer is

$$
\begin{aligned}
\mathbb{P}\left(\text{at least 2 people}\right) &= 1 - \mathbb{P}\left(\text{Everyone different birthday}\right) \\
&= 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{(365-31)}{365} \\
&= 1 - 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{334}{365} \approx 0.752374.
\end{aligned}
$$

– Really High!!!

CHAPTER 3

Independence

3.1. Independent Events

DEFINITION. We say E and F are **independent events** if
$$\mathbb{P}\left(E \cap F\right) = \mathbb{P}\left(E\right)\mathbb{P}\left(F\right).$$

- **Example1:** Suppose you flip two coins.
  - The event that you get heads on the second coin is independent of the event that you get tails on the first.
  - This is why: Let A_t be the event of getting is tails for the first coin and B_h is the event of getting heads for the second coin, and we assume we have fair coins (although this is not necessary), then
  $$\mathbb{P}\left(A_t \cap B_h\right) = \frac{1}{4}, \text{ list out all outcomes}$$
  $$\mathbb{P}\left(A_t\right)\mathbb{P}\left(B_h\right) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}.$$
- **Example2:** Experiment: Draw a card from an ordinary deck of cards
  - Let $A =$ draw ace, $S=$ draw a spade.
    * These are independent events since you're taking one at a time, so one doesn't effect the other. To see this using the definition we have compute
    * $\mathbb{P}\left(A\right)\mathbb{P}\left(S\right) = \frac{1}{13}\frac{1}{4}.$
    * White $\mathbb{P}\left(A \cap S\right) = \frac{1}{52}$ since there is only 1 Ace of spades.

PROPOSITION 2. *If E and F are independent, then E and F^c are independent.*

PROOF. Draw a Venn Diagram to help with the computation, but note that
$$\begin{aligned}
\mathbb{P}\left(E \cap F^c\right) &= \mathbb{P}\left(E\right) - \mathbb{P}\left(E \cap F\right) \\
&= \mathbb{P}\left(E\right) - \mathbb{P}\left(E\right)\mathbb{P}\left(F\right) \\
&= \mathbb{P}\left(E\right)\left(1 - \mathbb{P}\left(F\right)\right) \\
&= \mathbb{P}\left(E\right)\mathbb{P}\left(F^c\right).
\end{aligned}$$

□

- **Remark**: Independence and mutually exclusive, are two different things!

DEFINITION. We say E, F, G are **independent** if E, F are independent, E, G are independent, F, G are independent, and $\mathbb{P}\left(E \cap F \cap G\right) = \mathbb{P}\overline{\left(E\right)\mathbb{P}\left(F\right)\mathbb{P}\left(G\right)}$.

- **Example:** Experiment is you roll two dice:
  - Define the following events:

- $S_7 = \{\text{sum is 7}\}$
- $A_4 = \{\text{first die is a 4}\}$
- $B_3 = \{\text{second die is a 3}\}$
- Are the events S_7, A_4, B_3 independent?
    * Compute

$$\mathbb{P}\left(S_7 \cap A_4 \cap B_3\right) = \mathbb{P}\left(\{(4,3)\}\right) = \frac{1}{36}$$

  but

$$\mathbb{P}\left(S_7\right)\mathbb{P}\left(A_4\right)\mathbb{P}\left(B_3\right) = \frac{6}{36}\frac{1}{6}\frac{1}{6} = \frac{1}{36 \cdot 6}.$$

- **Remark:** This generalizes to events $A_1, \ldots, A_n$. We say events $A_1, \ldots, A_n$ are independent if for all subcollections $i_1, \ldots, i_r \in \{1, \ldots, n\}$ we have that $\mathbb{P}\left(\bigcap_{j=1}^{r} A_{i_j}\right) = \prod_{j=1}^{r} \mathbb{P}\left(A_{i_j}\right)$.
- **Example:**
    - An urn contains 10 balls: 4 red and 6 blue.
    - A second urn contains 16 red balls and an unknown number of blue balls.
    - A single ball is drawn from each urn. The probability that both balls are the same color is 0.44.
    - **Question:** Calculate the number of blue balls in the second urn.
    - **Solution:** Let R_i = even that a red ball is drawn from urn i and let B_i = event that a blue ball is drawn from urn i.
        * Let x be the number of blue balls in urn 2,
        * Note that drawing from urn 1 and independent from drawing from urn 2. They are completely different urns! They shouldn't effect the other.
        * Then

$$\begin{aligned}
.44 &= \mathbb{P}\left((R_1 \cap R_2)\bigcup(B_1 \cap B_2)\right) = \mathbb{P}\left(R_1 \cap R_2\right) + \mathbb{P}\left(B_1 \cap B_2\right) \\
&= \mathbb{P}\left(R_1\right)\mathbb{P}\left(R_2\right) + \mathbb{P}\left(B_1\right)\mathbb{P}\left(B_2\right), \text{ by independence} \\
&= \frac{4}{10}\frac{16}{x+16} + \frac{6}{10}\frac{x}{x+16}.
\end{aligned}$$

        * Solve for x! You will get $x = 4$.
- **Example** (Gambler's Ruin)(Used in Finance or Actuarial Science)
    - Experiment: Suppose you toss a fair coin repeatedly and independently. If it comes up heads, you win a dollar, and if it comes up tails, you lose a dollar. Suppose you start with \$50. What's the probability you will get to \$200 before you go broke?
    - Answer: It's actuallly easier if we generalize the problem.
        * Let $p(x)$ be the probability you get 200 before 0 if you start with x dollars.
        * We know $p(0) = 0$ and $p(200) = 1$. So by the law of total probability

$$\begin{aligned}
p(x) &= \mathbb{P}\left(\text{Win 200 before 0}\right) \\
&= \mathbb{P}\left(H\right)\mathbb{P}\left(\text{Win 200 before 0} \mid H\right) + \mathbb{P}\left(H^c\right)\mathbb{P}\left(\text{Win 200 before 0} \mid H^c\right) \\
&= \frac{1}{2}p\left(x+1\right) + \frac{1}{2}p\left(x-1\right).
\end{aligned}$$

∗ Rearrange the function to get

$$
\begin{aligned}
2p(x) = p\,(x-1) + p\,(x+1) \quad &\Longleftrightarrow \quad p(x) + p(x) = p\,(x-1) + p\,(x+1) \\
&\Longleftrightarrow \quad p(x) - p\,(x-1) = p\,(x+1) - p(x) \\
&\Longleftrightarrow \quad \frac{p(x) - p\,(x-1)}{x - (x-1)} = \frac{p\,(x+1) - p(x)}{(x+1) - x}.
\end{aligned}
$$

∗ This tellls you that the slows are constant. What does that tell you about $p(x)$? It's a line!
  · Thus we must have $p(x) = \frac{x}{200}$.
∗ Thus $p(50) = \frac{1}{4}$.

- **Example (A variation of Gambler's ruin)**
  - **Problem:** Suppose we are in the same situation, but you are allowed to go arbitrarily far in debt. Let $p(x)$be the probability you ever get to \$200. What is a formula for $p(x)$?
    ∗ Answer: Just as before $p(x) = \frac{1}{2}p(x+1) + \frac{1}{2}p(x-1)$. So that $p(x)$ is linear.
    ∗ But now all we have is that $p(200) = 1$ and linear and domain is $(-\infty, 200)$.
    ∗ **Draw a graph:** Now the slope, or $p'(x)$ can't be negative, or else we would have it that $p(x) > 1$ for $x \in (-\infty, 200)$.
      · The slope can't be positive or else we would get $p(x) < 0$ for $x \in (-\infty, 200)$.
    ∗ Thus we must have that $p(x) \equiv$ constant. Hence $p(x) = 1$ for all $x \in (-\infty.200)$.
    ∗ Sol: So we are certain to get \$200 if we cna get into debt.
  - Method2:
    ∗ Just compute There is nothing special about the figure 200. Another way of seeing this is to compute as above the probability of getting to 200 before $-M$ and then letting $M \to \infty$.
      · We would get $p(x)$ is a line with $p(-M) = 0$ and $p(200) = 1$ so that

$$
p(x) - 0 = \frac{1-0}{200 - (-M)}\,(x - (-M))
$$

and letting $M \to \infty$ wee see that $p(x) = \frac{x+M}{200+M} \to 1$.

- **Example:** Experiment: Roll 10 dice.
  - What is the probability that exactly <u>4 twos</u> will show if you roll <u>10 dice</u>?
  - **Answer:** These are independent. The probability that the 1st, 2nd, 3rd, and 10th dice will show a three and the other 6 will not is $\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7$.
  - Independence is used here: the probability is $\frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{1}{6}$. Note that the probability that the 10th, 9th, 8th, and 7th dice will show a *two* and the other 6 will not has the same probability.
  - So to answer our original question, we take $\left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6$ and multiply it by the number of ways of choosing 4 dice out of 10 to be the ones showing the *twos*. There are $\binom{10}{3}$ ways to do this $\binom{10}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^6$.
- This is an example of Bernoulli trials, or the Binomial distribution.

– If we have n independent trials, where the probability of success if p. The probability that there are k successes in n trials is

$$\binom{n}{k} p^k \left(1 - p\right)^{n-k}.$$

CHAPTER 4

Conditional Probability and Independence

4.1. Conditional Probabilities

- Suppose there are
  - 200 men, of which 100 are smokers,
  - 100 women, of which 20 are smokers.
  - Question1: What is the probability that a person chosen at random will be a smoker? $\frac{120}{300}$
  - Question2: Now, let us ask, what is the probability that a person chosen at random is a smoker given that the person is a women? $\frac{20}{100}$ right?
    * Note this is
    $$\frac{\#\,(\text{women smokers})}{\#\,(\text{women})} = \frac{P\,(\text{women and a smoker})}{P\,(\text{woman})}.$$
- Thus we make the following definition:

DEFINITION. If $\mathbb{P}\,(F) > 0$, we define
$$\mathbb{P}\,(E \mid F) = \frac{\mathbb{P}\,(E \cap F)}{\mathbb{P}\,(F)}.$$

Now $\mathbb{P}\,(E \mid F)$ is read "the probability of E given F."

- Note that $\mathbb{P}\,(E \cap F) = \mathbb{P}\,(E \mid F)\,\mathbb{P}\,(F)$!
- This is the *conditional probability* that E occurs given that F has already occured!
- **Remark:** Suppose $\mathbb{P}\,(E \mid F) = \mathbb{P}(E)$, i.e. knowing F doesn't help predict E. Then this implies that E and F are independent of each other. Rearranging $\mathbb{P}\,(E \mid F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \mathbb{P}\,(E)$ we see that $\mathbb{P}\,(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.
- **Example1:** Experiment: Roll two dice.
  - (a) What is the probability the sum is 8?
    * Solution: Note that $A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$ so we know $\mathbb{P}\,(A) = \frac{5}{36}$.
  - (b) What is the probability that the sum is 8 given that the first die shows a 3? (In other words, find $\mathbb{P}\,(A \mid B)$)
    * Solution: Let $B = \{\text{first die shows three}\}$.
    * $\mathbb{P}\,(A \cap B) = \mathbb{P}\,(\{(3,5)\}) = \frac{1}{36}$ is probability that the first die shows a 3 **and** the sum is 8
    * Finally we can compute
    $$\mathbb{P}\,(A \mid B) = \mathbb{P}\,(\text{sum is 8} \mid \text{1st is a 3}) = \frac{1/36}{1/6} = \frac{1}{6}.$$
- Remark: When computing $\mathbb{P}\,(E \mid F)$, Sometime its easier to work with the reduced sample space $F \subset S$.

22

– Note in the previous example when we computed

$$\mathbb{P}\left(\text{sum is 8} \mid \text{1st is a 3}\right)$$

we could have worked in the smaller sample space of $\{\text{1st is a 3}\} = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$. Since only $(3,5)$ begins with a 3 and has the sum of 8, then the probability is

$$\frac{\text{total number of outcomes in the event}}{\text{total number of outcomes in \textbf{new} sample space}} = \frac{1}{6}.$$

- **Example2:** Experiment: Suppose a box has 3 red marbles and 2 black ones. We select 2 marbles.
  – Question: What is the probability that second marble is red given that the first one is red?
    ∗ Answer:
      · $R_1 = \{\text{First marble is is red }\}$,
      · $R_2 = \{\text{Second marble is red}\}$, then

$$
\begin{aligned}
\mathbb{P}\left(R_2 \mid R_1\right) &= \frac{\mathbb{P}\left(R_1 \cap R_2\right)}{\mathbb{P}\left(R_1\right)} \\
&= \frac{(2 \text{ red})(0 \text{ black}) / \binom{5}{2}}{3/5} \\
&= \frac{\binom{3}{2}\binom{2}{0} / \binom{5}{2}}{3/5} \\
&= \frac{3/10}{3/5} = \frac{1}{2}.
\end{aligned}
$$

    ∗ Solution 2:
      · We could have done the same example more easiely if we look at the new sample space $S' = \{R,R,B,B\}$ thus $\mathbb{P}\left(R_2 \mid R_1\right) = \mathbb{P}'\left(\{\text{drawing red}\}\right) = \frac{2}{4} = \frac{1}{2}$.
- **Example3:** Landon is 80% sure he forgot his textbook at the Union or Monteith buildings. 40% sure that it is at the union, and 40% sure that it is at Monteith. Given that Landon already went to Monteith and noticed his textbook not there, what is the probability that it's at the Union?
  – Solution:

$$
\begin{aligned}
\mathbb{P}\left(Union \mid Not \text{ Monteith}\right) &= \frac{\mathbb{P}\left(U \cap M^c\right)}{\mathbb{P}\left(M^c\right)} \\
&= \frac{\mathbb{P}\left(U\right)}{1 - \mathbb{P}\left(M\right)}, \text{ since } U \subset M^c \\
&= \frac{4/10}{6/10} = \frac{2}{3}.
\end{aligned}
$$

- **Example4:** Suppose that Annabelle and Bobby each draw 13 cards from a standard deck of 52. Given that Sarah has exactly two aces, what is the probability that Bobby has exactly one ace?
  – Solution: Let A be the event "Annabelle has two aces," and let B be the event "Bobby has exactly one ace." Again, we want $\mathbb{P}\left(B \mid A\right)$, so we calculate $\mathbb{P}(A)$ and $\mathbb{P}(A \cap B)$. Annabelle could have any of $\binom{52}{13}$ possible hands. Of these hands, $\binom{4}{2} \cdot \binom{48}{11}$ will have exactly

two aces, so

$$\mathbb{P}(A) = \frac{\binom{4}{2} \cdot \binom{48}{11}}{\binom{52}{13}}.$$

Now the number of ways in which Annabelle can have a certain hand and Bobby can have a certain hand is $\binom{52}{13} \cdot \binom{39}{13}$, and the number of ways in which A and B can both occur is $\binom{4}{2} \cdot \binom{48}{11} \cdot \binom{2}{1} \cdot \binom{37}{12}$. so

$$\mathbb{P}(A \cap B) = \frac{\binom{4}{2} \cdot \binom{48}{11} \cdot \binom{2}{1} \cdot \binom{37}{12}}{\binom{52}{13} \cdot \binom{39}{13}}.$$

Therefore,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\frac{\binom{4}{2} \cdot \binom{48}{11} \cdot \binom{2}{1} \cdot \binom{37}{12}}{\binom{52}{13} \cdot \binom{39}{13}}}{\frac{\binom{4}{2} \cdot \binom{48}{11}}{\binom{52}{13}}}$$

$$= \frac{\binom{2}{1} \cdot \binom{37}{12}}{\binom{39}{13}}.$$

- Note that since $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$.
    - **In general:** If $E_1, \ldots, E_n$ are events then

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1)\mathbb{P}(E_2 \mid E_1)\mathbb{P}(E_3 \mid E_1 \cap E_2) \cdots \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

- **Example5:**
    - Experiment: Suppose an urn has 5 White balls and 7 Black balls. Each ball that is selected is returned to the urn along with an additional ball of the same color. Suppose draw 3 balls.
    - **Part (a):** What is the probability that you get 3 white balls.
        * Then

$$\mathbb{P}(3 \text{ white balls}) = \mathbb{P}(1\text{st W})\mathbb{P}(2\text{nd W} \mid 1\text{st W})\mathbb{P}(3\text{nd W} \mid 1\text{st \& 2nd W})$$

$$= \frac{5}{12}\frac{6}{13}\frac{7}{14}$$

– **Part (b):** What is the probability of getting 1 white ball.

$$\begin{aligned}
\mathbb{P}\left(1 \text{ white ball}\right) &= \mathbb{P}\left(WBB\right) + \mathbb{P}\left(BWB\right) + \mathbb{P}\left(BBW\right) \\
&= 3 \frac{5 \cdot 7 \cdot 8}{12 \cdot 13 \cdot 14}.
\end{aligned}$$

- Note that

$$\mathbb{P}\left(E \cap F\right) = \mathbb{P}\left(E \mid F\right) \mathbb{P}\left(F\right)$$

- **Example 6:** Phan wants to take a Biology course or a Chemistry course. Given that the students take Biology, the probability that they get an A is is $\frac{4}{5}$. While the probability of getting an A given that the student took Chemistry is $\frac{1}{7}$. If Phan makes a decision on the course to take randomly, what's probability of "getting an A in Chem"?

  – **Solution:** Let $B = \{\text{Takes Biology}\}$ and $C = \{\text{Takes Chemistry}\}$ and $A = \{"\text{gets an A}"\}$, then

$$\begin{aligned}
\mathbb{P}\left(A \cap C\right) &= \mathbb{P}\left(C\right) \mathbb{P}\left(A \mid C\right) \\
&= \frac{1}{2} \cdot \frac{1}{7} = \frac{1}{14}.
\end{aligned}$$

- **Example 7:** A total of 500 married couples are poled about salaries:

Wife	Husband makes less than 25,000	Husband makes more than 25,000
Less than $25,000	212	198
More than $25,000	36	54

  – **Part (a):** Find the probability that a Husband earns less than 25,000?

    * Answer: $\frac{212+36}{500}$

  – **Part (b):** Find $\mathbb{P}\left(\text{wife makes } > 25,000 \mid \text{Husband makes } > 25,000\right)$

    * Answer: $\frac{54/500}{(198+54)/500} = \frac{54}{252} = .214$

  – **Part (c):** Find $\mathbb{P}\left(\text{wife makes } > 25,000 \mid \text{Husband makes } < 25,000\right)$

    * Answer: $\frac{36/500}{(248)/500} = .145$.

4.2. Bayes's Formula

- Sometimes it's easier to compute a probability once we know something has or has not happened.
- Note that we can compute,

$$
\begin{aligned}
\mathbb{P}(E) &= \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) \\
&= \mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)\mathbb{P}(F^c) \\
&= \mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)(1 - \mathbb{P}(F)).
\end{aligned}
$$

- This formula is called: **The Law of Total Probability:**

$$
\mathbb{P}(E) = \mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)(1 - \mathbb{P}(F))
$$

- The following problem will describe the types of problems of this section.
- **Example1:** Insurance company believes
  - The probability that "an accident prone person" has an accident within a year is .4.
  - The probability that "Non-accident prone person" has an accident with year is .2.
  - 30% of the population is "accident prone".
  - <u>Part (a):</u> Find $\mathbb{P}(A_1)$ where $A_1 =$ new policy holder will have <u>an accident within a year</u>?
    * Let $A = \{$Policy holder IS accident prone.$\}$

$$
\begin{aligned}
\mathbb{P}(A_1) &= \mathbb{P}(A_1 \mid A)\mathbb{P}(A) + \mathbb{P}(A_1 \mid A^c)(1 - \mathbb{P}(A)) \\
&= .4(.3) + .2(1 - .3) \\
&= .26
\end{aligned}
$$

  - <u>Part (b):</u> Suppose new policyholder has accident with one year. What's probability that he or she is accident prone?

$$
\begin{aligned}
\mathbb{P}(A \mid A_1) &= \frac{\mathbb{P}(A \cap A_1)}{\mathbb{P}(A_1)} \\
&= \frac{\mathbb{P}(A)\mathbb{P}(A_1 \mid A)}{.26} \\
&= \frac{(.3)(.4)}{.26} = \frac{6}{13}.
\end{aligned}
$$

- **In general:**
  - So in Part (a) we had to break a probability into two cases: If $F_1, \ldots, F_n$ are mutually exclusive events such that they make up everythinn $S = \bigcup_{i=1}^{n} F_i$ then

$$
\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E \mid F_i)\mathbb{P}(F_i).
$$

    * This is called **Law of Total Probability.**
  - In Part (b), we wanted to find a probability of a separate conditional event: then

$$
\mathbb{P}(F_j \mid E) = \frac{\mathbb{P}(E \mid F_j)\mathbb{P}(F_j)}{\sum_{i=1}^{n} \mathbb{P}(E \mid F_i)\mathbb{P}(F_i)}.
$$

    * This is known as **Baye's Formula**
    * Note that the denominator of the Bayes's formula is the Law of total probability.
- **Example2:** Suppose the test for HIV is
  - 98% accurate in both directions
  - 0.5% of the population is HIV positive.

- Question: If someone tests positive, what is the probability they actually are HIV positive?
- Solution: Let $T_+ = \{$tests positive$\}$, $T_- = \{$tests negative$\}$, while $+ = \{$actually HIV positive,$\}$ $- = \{$actually negative$\}$.
  * Want

$$\begin{aligned}
\mathbb{P}\left(+ \mid T_+\right) &= \frac{\mathbb{P}\left(+ \cap T_+\right)}{\mathbb{P}\left(T_+\right)} \\
&= \frac{\mathbb{P}\left(T_+ \mid +\right)\mathbb{P}\left(+\right)}{\mathbb{P}\left(T_+ \mid +\right)\mathbb{P}\left(+\right) + \mathbb{P}\left(T_+ \mid -\right)\mathbb{P}\left(-\right)} \\
&= \frac{(.98)(.005)}{(.98)(.005) + .02(.995)} \\
&= 19.8\%.
\end{aligned}$$

- **Example3:** Suppose
  - 30% of the women in a class received an A on the test
  - 25% of the men/or else received an A.
  - 60% of the class are women.
  - Question: Given that a person chosen at random received an A, what is the probability this person is a women?
    * Solution: Let A the event that a students receives an A. Let $W =$being a women, $M =$not a women. Want

$$\begin{aligned}
\mathbb{P}\left(W \mid A\right) &= \frac{\mathbb{P}\left(A \mid W\right)\mathbb{P}\left(W\right)}{\mathbb{P}\left(A \mid W\right)\mathbb{P}\left(W\right) + \mathbb{P}\left(A \mid M\right)\mathbb{P}\left(M\right)}, \text{ by Bayes's} \\
&= \frac{.3(.6)}{.3(.6) + .25(.4)} = \frac{.18}{.28} \approx .64.
\end{aligned}$$

- (General Baye's Theorem) Here's one with more than 3 possibilities:
- **Example4:** Suppose in Factory with Machines I,II,III producing Iphones
  - Machines I,II,III produce 2%,1%, and 3% defective iphones, respectively.
  - Out of total production, Machines I makes 35% of all Iphones, II-25%, III- 40%.
  - If one Iphone is selected at random from the factory,
  - Part (a): what is probability that one Iphone selected is defective?

$$\begin{aligned}
\mathbb{P}\left(D\right) &= P\left(I\right)\mathbb{P}\left(D \mid I\right) + P\left(II\right)\mathbb{P}\left(D \mid II\right) + P\left(III\right)\mathbb{P}\left(D \mid III\right) \\
&= (.35)(.02) + (.25)(.01) + (.4)(.03) \\
&= \frac{215}{10,000}.
\end{aligned}$$

  - Part (b): What is the conditional prob that if an Iphone is defective, that it was produced by machine III?

$$\begin{aligned}
\mathbb{P}\left(III \mid D\right) &= \frac{\mathbb{P}\left(III\right)\mathbb{P}\left(D \mid III\right)}{\mathbb{P}\left(D\right)} \\
&= \frac{(.4)(.03)}{215/10,000} = \frac{120}{215}.
\end{aligned}$$

- **Example5:** In a Multiple Choice Test, students either knows the answer or randomly guesses the answer to a question.
  - Let $m =$number of choices in a question.

- Let $p =$ the probability that the students knows the answer to a question.
- **Question:** What is the probability that the student actually knew the answer, given that the student answers correctly.
- Solution:
- Let $K = \{$Knows the answer$\}$ and $C = \{$Answer's correctly$\}$. Then

$$
\begin{aligned}
\mathbb{P}\left(K \mid C\right) &= \frac{\mathbb{P}\left(C \mid K\right)\mathbb{P}\left(K\right)}{\mathbb{P}\left(C \mid K\right)\mathbb{P}\left(K\right) + \mathbb{P}\left(C \mid K^c\right)\mathbb{P}\left(K^c\right)} \\
&= \frac{1 \cdot p}{1 \cdot p + \frac{1}{m}\left(1 - p\right)} = \frac{mp}{1 + (m-1)p}.
\end{aligned}
$$

CHAPTER 5

Random Variables

5.1. Random Variables

- When we perform an experiment, we are interested in some function of the outcomes, instead of the actual outcome.
    - We want to attach for each outcome, a numerial value.
- **Definition:** A random variable is a function $X : S \to \mathbb{R}$ or write $X : \Omega \to \mathbb{R}$. (Use capital letters to denote r.v)
    - We can think of X as a numerical value that is random, like as if X is a random number.
- **Example:** Toss a coin
    - Let X be 1 if heads and $X = 0$ if tails
    - Then $X(H) = 1$ and $X(T) = 0$.
    - We can do calculus on real numbers but not on $\Omega = S = \{H, T\}$.
- **Example:** Roll a die
    - Let X denote the outcome, so $X = 1, 2, 3, 4, 5, 6$ (its random)
    - That is $X(1) = 1, X(2) = 2, \ldots$.
- **Example:** Roll a die, define

$$Y = \begin{cases} 1 & \text{outomce= odd} \\ 0 & \text{outomce= even} \end{cases}$$

    - Can be thought of as

$$Y(s) = \begin{cases} 1 & s = \text{odd} \\ 0 & s = \text{even} \end{cases}.$$

- A common question we'll have is **"What values can X attain ?"**
    - In other words, what is the range of X? Since $X : S \to$?
- **Example:** Toss a coin 10 times
    - Let X be the number of heads showing
    - What random values can X be? $0, 1, 2, \ldots, 10$.
- **Example:** In general in n trials, X is the number of successes
- **Example1:** Let X be the amount of liability(damages) a driver incurs in a year.
    - $X : S \to [0, \infty)$.
- **Example2:** Toss a coin 3 times
    - Let X be the number of heads that appear, so $X = 0, 1, 2, 3$.
    - In other words, $X : S \to \{0, 1, 2, 3\}$

– We may assign probabiliies to the different values of the random variable:

$$
\begin{aligned}
\mathbb{P}\left(X=0\right) &= \mathbb{P}\left((T,T,T)\right)=\frac{1}{2^{3}}=\frac{1}{8}\\[2mm]
\mathbb{P}\left(X=1\right) &= \mathbb{P}\left((T,T,H),(T,H,T),(H,T,T)\right)=\frac{3}{8}\\[2mm]
\mathbb{P}\left(X=2\right) &= \mathbb{P}\left((T,H,H),(H,H,T),(H,T,H)\right)=\frac{3}{8}\\[2mm]
\mathbb{P}\left(X=3\right) &= \mathbb{P}\left((H,H,H)\right)=\frac{1}{8}.
\end{aligned}
$$

– Note that since X must take the values of 0 through 3 then

$$
1=\mathbb{P}\left(\bigcup_{i=0}^{3}\{X=i\}\right)=\sum_{i=0}^{3}\mathbb{P}\left(X=i\right),
$$

which makes sense from our previous calculation.

5.2. Discrete Random Variables

DEFINITION. A random variable that can take on at most countable number of possible values is said to be a **discrete r.v.**

DEFINITION. For a discrete random variable, we can define the **probability mass function** (pmf), or the **density** function of X by $p(x) = \mathbb{P}(X = x)$. Note that $p : \mathbb{R} \to [0, 1]$.

- Note that $(X = x) = (\omega \in \Omega \mid X(\omega) = x)$ is an abbreviation.
- Let X assume only the values $x_1, x_2, x_3 \ldots$
  - In other words, $X : S \to \{x_1, x_2, \ldots\}$
  - **Properties of a pmf $p(x)$:**
    * Note that we must have $0 < p(x_i) \leq 1$ for $,i = 1, 2, \ldots$. and $p(x) = 0$ for all other values of x can't attain.
    * Also must have
    $$\sum_{i=1}^{\infty} p(x_i) = 1.$$
- We often draw bar graphs for discrete r.v.
- **Example:** If we toss a coin
  - $X = 1$ if we have H and $X = 0$ if we have T.
  - Then draw a **BAR graph**
  $$p_X(x) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2} & x = 1, \\ 0 & \text{otherwise} \end{cases}$$
- Oftentimes someone has already found the pmf for you, and you can use to compute probabilities.
- **Example:** The pmf of X is given by $p(i) = e^{-\lambda} \frac{\lambda^i}{i!}$ for $i = 0, 1, 2, \ldots$ where λ is a parameter(what is this?) that is any positive number
  - **Part (a)** What values can the random variable X attain? In other words, what is the range of X?
    * **Sol:** By definition we have $\mathbb{P}(X = 0) = p(0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$
  - **Part (b)** Find $\mathbb{P}(X = 0)$
    * **Sol:** By definition we have $\mathbb{P}(X = 0) = p(0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$
  - **Part (c)** Find $\mathbb{P}(X > 2)$
    * **Sol:** Note that
    $$\begin{aligned} \mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \leq 2) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - p(0) - p(1) - p(2) \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}. \end{aligned}$$

5.3. Expected Value

- One of the most important concepts in probability is that of expectation. If X is a random variable that what is the average value of X, that is what is the **expected value** of X.

DEFINITION. Let X have a pmf $p(x)$. We define the **expectation**, or **expected value** of X to be

$$\mathbb{E}[X] = \sum_{x:p(x)>0} xp(x).$$

- Notation $\mathbb{E}X$, or EX.
- **Example1:** Let $X(H) = 0$ and $X(T) = 1$. What is $\mathbb{E}X$?

$$\begin{aligned} \mathbb{E}X &= 0 \cdot p(0) + 1 \cdot p(1) \\ &= 0\frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}. \end{aligned}$$

- **Example2:** Let X be the outcome when we roll a fair die. What is $\mathbb{E}X$?

$$\begin{aligned} \mathbb{E}X &= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \cdots + 6\frac{1}{6} \\ &= \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2} = 3.5 \end{aligned}$$

  - Note that X can never be 3.5 , so expectation is to give you an idea, what an exact.
- **Recall infinite series:** If $0 \le x < 1$ then a geometric series is

$$\begin{aligned} \sum_{n=0}^{\infty} x^n &= 1 + x + x^2 + x^3 + \cdots \\ &= \frac{1}{1-x}. \end{aligned}$$

  - One thing you can do with series is differentiate them and integrate them: So if

$$1 + x + x^2 + x^3 + \cdots + = \frac{1}{1-x}$$

  then

$$0 + 1 + 2x + 3x^2 + \cdots + = \frac{1}{(1-x)^2}$$

- **Example3:** Let X be the number or tornados in Connecticut per year. Meaning that the random variable X can be any number $X = 0, 1, 2, 3, \ldots$. Suppose the state of Connecticut did some analysis and found out that

$$\mathbb{P}(X = i) = \frac{1}{2^{i+1}}.$$

  - Question: What is $\mathbb{E}X$? That is, what is the expected number of tornados per year in Connecticut.
  - **Solution:** Note that X is infinite, but still countable, hence still discrete.

– Note that

$$p(i) = \begin{cases} \frac{1}{2} & i = 0, \\ \frac{1}{4} & i = 1, \\ \frac{1}{8} & i = 2, \\ \vdots & \vdots \\ \frac{1}{2^{n+1}} & i = n. \end{cases}$$

– We have that

$$
\begin{aligned}
\mathbb{E}X &= 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + \cdots \\
&= 0 \cdot \frac{1}{2} + 1\frac{1}{2^2} + 2\frac{1}{2^3} + 3\frac{1}{2^4} + \cdots \\
&= \frac{1}{2^2}\left(1 + 2\frac{1}{2} + 3\frac{1}{2^2} + \cdots\right) \\
&= \frac{1}{4}\left(1 + 2x + 3x^2 + \cdots\right), \text{ with} x = \frac{1}{2} \\
&= \frac{1}{4}\frac{1}{(1-x)^2} = \frac{1}{4\left(1 - \frac{1}{2}\right)^2} = 1.
\end{aligned}
$$

5.4. The C.D.F.

DEFINITION. Define $F : \mathbb{R} \to [0, 1]$ to be the function

$$F(x) = \mathbb{P}\left(X \leq x\right), \quad \text{for any } -\infty < x < \infty$$

to be the cumulative distribution function, or the distribution function of X., or CDF of X, or c.d.f

- Note that when X is discrete,

$$F(x_0) = \mathbb{P}\left(X \leq x_0\right) = \sum_{x \leq x_0} p(x).$$

- We sometimes use the notation $F_X(x)$ to highlight that F_X is the CDF of the random variable X.
- **Example:** Suppose X is equals to the number of heads in 3 coin flips. From Section 5.1, we calculated the p.m.f to be.:
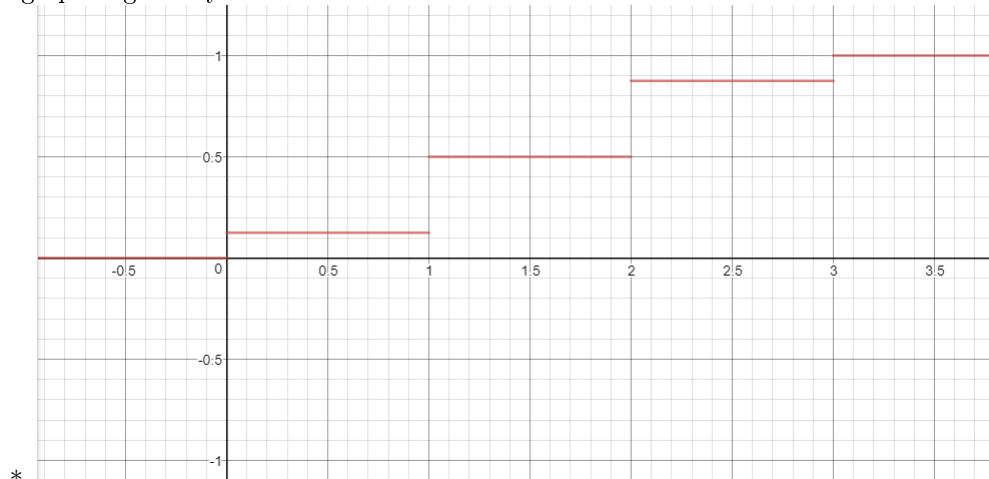
$$
\begin{aligned}
p(0) = \mathbb{P}\left(X = 0\right) &= \frac{1}{8} \\
p(1) = \mathbb{P}\left(X = 1\right) &= \frac{3}{8} \\
p(2) = \mathbb{P}\left(X = 2\right) &= \frac{3}{8} \\
p(3) = \mathbb{P}\left(X = 3\right) &= \frac{1}{8}.
\end{aligned}
$$

Question: Find the c.d.f of X. Plot the graph of the c.d.f.

  − **Solution:** Summing up the probabilities up to that value of x we get the following:

$$
F(x) = \begin{cases}
0 & -\infty < x < 0 \\
\frac{1}{8} & 0 \leq x < 1 \\
\frac{4}{8} & 1 \leq x < 2 \\
\frac{7}{8} & 2 \leq x < 3 \\
1 & 3 \leq x < \infty
\end{cases}.
$$

  − The graph is given by

*

- – Note that this is a step function.
- – This function has jumps, and not continouous everywhere.
- – But it looks like it never decreases.
- **Properties of the CDF:**
  - – 1. F is nondecreasing , that is
    - * if $x < y$ then $F(x) \leq F(y)$
  - – 2. $\lim_{x \to \infty} F(x) = 1$.
  - – 3. $\lim_{x \to -\infty} F(x) = 0$.
  - – 4. F is right continuous. That are two ways that you can think of right continuity:
    - * $\lim_{x \to 1^+} F(x_n) = F(x)$, meaning "the limit from the right equals where the function is defined"
    - * If $x_n \downarrow x$ is a decreasing sequence then $\lim_{n \to \infty} F(x_n) = F(x)$.
- We take these properties as facts, though one would normally have to prove these.
- The following proposition does not have to be proved in class, and can be highlighted with the following example. But we include it here for completeness.

PROPOSITION 3. *Let $F_X(x)$ be the CDF for some random variable X. Then the following holds:*
(a) For any $a \in \mathbb{R}$, we have $\mathbb{P}(X < a) = \lim_{x \to a^-} F_X(x)$
(b) For any $a \in \mathbb{R}$, we have $\mathbb{P}(X = a) = F_X(a) - \lim_{x \to a^-} F_X(x)$

PROOF. For part (a).
We first write

$$(X < a) = \bigcup_{n=1}^{\infty} \left(X \leq a - \frac{1}{n} \right)$$

$$= (X \leq a - 1) \bigcup \left[\bigcup_{n=1}^{\infty} \left(a - \frac{1}{n} < X \leq a - \frac{1}{n+1} \right) \right]$$

and since the events $E_n = \left(a - \frac{1}{n} \leq X \leq a - \frac{1}{n+1} \right)$ are disjoint then we can use Axiom 3 so prove that

$$\mathbb{P}(X < a) = \mathbb{P}(X \leq a - 1) + \sum_{n=1}^{\infty} \mathbb{P}\left(a - \frac{1}{n} < X \leq a - \frac{1}{n+1} \right)$$

$$= \mathbb{P}(X \leq a - 1) + \lim_{k \to \infty} \sum_{n=1}^{k} \left[\mathbb{P}\left(X \leq a - \frac{1}{n+1} \right) - \mathbb{P}\left(X \leq a - \frac{1}{n} \right) \right]$$

$$= \mathbb{P}(X \leq a - 1) + \lim_{k \to \infty} \left[\mathbb{P}\left(X \leq a - \frac{1}{k+1} \right) - \mathbb{P}(X \leq a - 1) \right], \text{ by telescoping}$$

$$= \lim_{k \to \infty} \mathbb{P}\left(X \leq a - \frac{1}{k+1} \right) + \mathbb{P}(X \leq a - 1) - \mathbb{P}(X \leq a - 1)$$

$$= \lim_{n \to \infty} F_X\left(a - \frac{1}{n} \right).$$

Now you can replace the sequence $a_n = a - \frac{1}{n}$ with any sequence a_n that is increasing towards a, and we get the similar result,

$$\lim_{n \to \infty} F_X(a_n) = \mathbb{P}(X < a),$$

since this holds for all increasing sequences a_n towards a, then we've shown that
$$\lim_{x \to a^-} F_X(x) = \mathbb{P}(X < a).$$

For part (b). We use part (a) and get
$$\mathbb{P}(X = a) = \mathbb{P}(X \leq a) - \mathbb{P}(X < a)$$
$$= F_X(a) - \lim_{x \to a^-} F_X(x).$$

$\square$

- **Example:** Let X have distribution
$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x < 1 \\ \frac{2}{3} & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x. \end{cases}$$

Graph this and answer the following:
  - **Part (a):** Compute $\mathbb{P}(2 < X \leq 4)$. We have that
$$\begin{aligned} \mathbb{P}(2 < X \leq 4) &= \mathbb{P}(X \leq 4) - \mathbb{P}(X \leq 2) \\ &= F(4) - F(2) \\ &= \frac{1}{12}. \end{aligned}$$
  - **Part (b):** Compute $\mathbb{P}(X < 3)$.
    * We have that
$$\begin{aligned} \mathbb{P}(X < 3) &= \lim_{n \to \infty} \mathbb{P}\left(X \leq 3 - \frac{1}{n}\right) \\ &= \lim_{x \to 3^-} F_X(x) \\ &= \frac{11}{12} \end{aligned}$$
  - **Part (c):** Compute $\mathbb{P}(X = 1)$.
    * We have that
$$\begin{aligned} \mathbb{P}(X = 1) &= \mathbb{P}(X \leq 1) - \mathbb{P}(X < 1) \\ &= F_X(1) - \lim_{x \to 1^-} F_X(x) \\ &= \frac{2}{3} - \lim_{x \to 1} \frac{x}{2} \\ &= \frac{2}{3} - \frac{1}{2} = \frac{1}{6}. \end{aligned}$$

5.5. Expectated Value of Sums of Random Variables

- Recall our current definition of $\mathbb{E}X$
    - List out $X = x_1, x_2, \ldots$ and let $p(x_i)$ be the density of X
    - Then $\mathbb{E}X = \sum_{i=1}^{\infty} x_i p(x_i)$.
- We need a <u>new definition</u> that will help the linearity of expectation.
    - Goal: If $Z = X + Y$ then $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$.
- **<u>Definition 2:</u>** Let S (or Ω) be the sample space then define

$$\mathbb{E}X = \sum_{\omega \in S} X(\omega)\mathbb{P}\left(\{\omega\}\right).$$

  - **<u>Example:</u>** Let $S = \{1, 2, 3, 4, 5, 6\}$ and $X(1) = X(2) = 1$ and $X(3) = X(4) = 3$ and $X(5) = X(6) = 5$
      * <u>Def1:</u> We know $X = 1, 3, 5$ with $p(1) = p(3) = p(5) = \frac{1}{3}$
      * Then $\mathbb{E}X = 1 \cdot \frac{1}{3} + 3\frac{1}{3} + 5\frac{1}{3} = \frac{9}{3} = 3$.
      * <u>Def2:</u> We list all of $S = \{1, 2, 3, 4, 5, 6\}$ and
      * Then

$$\begin{aligned} \mathbb{E}X &= X(1)\mathbb{P}\left(\{1\}\right) + \cdots + X(6) \cdot \mathbb{P}\left(\{6\}\right) \\ &= 1\frac{1}{6} + 1\frac{1}{6} + 3\frac{1}{6} + 3\frac{1}{6} + 5\frac{1}{6} + 5\frac{1}{6} = 3. \end{aligned}$$

- Difference
    - Def1: We list all the values that X can attain and only care about those. (Range)
    - Def2: List all possible outcomes. (Domain)

PROPOSITION 4. *If X is a discrete random variable and S is countable, then the two definitions are equivalent*

- **NOTE:** No need to prove in lecture. But here for completeness.

PROOF. We start with the first definition. Let $X = x_1, x_2, \ldots$

$$\begin{aligned} \mathbb{E}X &= \sum_{x_i} x_i p(x_i) \\ &= \sum_{x_i} x_i \mathbb{P}\left(X = x_i\right) \\ &= \sum_{x_i} x_i \sum_{\omega \in \{\omega : X(\omega) = x_i\}} \mathbb{P}\left(\omega\right) \\ &= \sum_{x_i} \sum_{\omega \in \{\omega : X(\omega) = x_i\}} x_i \mathbb{P}\left(\omega\right) \\ &= \sum_{x_i} \sum_{\omega \in \{\omega : X(\omega) = x_i\}} X(\omega)\mathbb{P}\left(\omega\right) \\ &= \sum_{\omega \in S} X(\omega)\mathbb{P}\left(\omega\right), \end{aligned}$$

where I used that each $S_i = \{\omega : X(\omega) = x_i\}$ are mutually exclusinve events that union up to S. $\qquad\square$

- Using this definition, we can prove linearity of the expectation.

THEOREM 5. *(Linearity) If X and Y are discrete random variables and $a \in \mathbb{R}$ then*
(a) $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$.
(b) $\mathbb{E}[aX] = a\mathbb{E}X$.

PROOF. We have that

$$
\begin{aligned}
\mathbb{E}[X + Y] &= \sum_{\omega \in S} (X(\omega) + Y(\omega)) \, \mathbb{P}(\omega) \\
&= \sum_{\omega \in S} (X(\omega)\mathbb{P}(\omega) + Y(\omega)\mathbb{P}(\omega)) \\
&= \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega) + \sum_{\omega \in S} Y(\omega)\mathbb{P}(\omega) \\
&= \mathbb{E}X + \mathbb{E}Y.
\end{aligned}
$$

If $a \in \mathbb{R}$ then

$$
\begin{aligned}
\mathbb{E}[aX] &= \sum_{\omega \in S} (aX(\omega)) \, \mathbb{P}(\omega) \\
&= a \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega) \\
&= a\mathbb{E}X.
\end{aligned}
$$

$\square$

- **Generality:** Linearity is true for general random variable $X_1, X_2, \ldots, X_n$.

5.6. Expectation of a Function of a Random Variable

- Let X be a random variable.
  - Can we find the expected value of things like $X^2, e^X, \sin X$ etc?
- **Example1:** Let X denote a random variable such that

$$
\begin{aligned}
\mathbb{P}\left(X=-1\right) &= .2, \\
\mathbb{P}\left(X=0\right) &= .5 \\
\mathbb{P}\left(X=1\right) &= .3
\end{aligned}
$$

  Let $Y = X^2$. Find $\mathbb{E}Y$.
  - **Solution:** Note that $Y = \left\{0^2, \left(-1\right)^2, \left(1\right)^2\right\} = \{0,1\}$.
  - Note that $p_Y(1) = .2 + .3 = .5$ and $p_Y(0) = .5$.
  - Thus $\mathbb{E}Y = 0 \cdot .5 + 1 \cdot .5 = .5$.
- IMPORTANT:
  - Note that $\mathbb{E}X^2 = .5$.
  - While $\left(\mathbb{E}X\right)^2 = .01$. Not equal!
    * Since $\mathbb{E}X = .3 - .2 = .1$. Thus

$$
\mathbb{E}X^2 \neq \left(\mathbb{E}X\right)^2.
$$

- In general, there is a formula for $g(X)$ where g is function. That use the fact that $g(X)$ will be $g(x)$ for some x such that $X = x$.

THEOREM 6. *If X is a discrete random varianle that takes values $X \in \{x_1, x_2, x_3, \dots\}$ with respective probability mass function $p(x_i)$, then for any real valued function $g : \mathbb{R} \to \mathbb{R}$ we have that*

$$
\mathbb{E}\left[g\left(X\right)\right] = \sum_{i=1}^{\infty} g\left(x_i\right) p(x_i).
$$

- **NOTE:** No need to prove in lecture. But here for completeness.

PROOF. The random variable $Y = g(X)$ can take on values, say $Y = y_1, y_2, \dots$. But we know that

$$
y_j = g(x_i)
$$

and as we see there could be more than one value x_i such that $y_j = g(x_i)$. Thus we will group this sum into this fashion: Using the definition of expectation we have that

$$
\begin{aligned}
\mathbb{E}\left[Y\right] &= \sum_j y_j \mathbb{P}\left(Y = y_j\right) \\
&= \sum_j y_j \mathbb{P}\left(g(X) = y_j\right) \\
&= (\star).
\end{aligned}
$$

Now

$$
\begin{aligned}
\mathbb{P}\left(g(X) = y_j\right) &= \mathbb{P}\left(\bigcup_{i:g(x_i)=y_j} \left(g(x_i) = y_j\right)\right) \\
&= \sum_{i:g(x_i)=y_j} p(x_i).
\end{aligned}
$$

Thus plugging this back into $(\star)$ we have that

$$
\begin{aligned}
\mathbb{E}\left[Y\right] &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\
&= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\
&= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
&= \sum_{i=1}^{\infty} g(x_i)p(x_i),
\end{aligned}
$$

as needed. $\qquad\qquad\qquad\square$

- Remark: $\mathbb{E}X^2 = \sum x_i^2 p(x_i)$.
- **Example1 (Revisted):** Let X denote a random variable such that

$$
\begin{aligned}
\mathbb{P}\left(X = -1\right) &= .2, \\
\mathbb{P}\left(X = 0\right) &= .5 \\
\mathbb{P}\left(X = 1\right) &= .3
\end{aligned}
$$

Let $Y = X^2$. Find $\mathbb{E}Y$.
  - **Sol:** We have that $\mathbb{E}X^2 = \sum x_i^2 p(x_i) = (-1)^2 (.2) + 0^2(.5) + 1^2(.3) = .5$.

DEFINITION. We call $\mu = \mathbb{E}X$ to be the **mean**, or the **first moment** of X.
The quantity $\mathbb{E}X^n$ for $n \geq 1$, is called the n**th moment** of X.

- From out theorem we know that the nth moments can be calculated a

$$
\mathbb{E}X^n = \sum_{x:p(x)>0} x^n p(x).
$$

5.7. Variance

- The variance of a r.v. is a measure of how spread out the values of X are.
- The expectation of a r.v. is quantity that help us differentiate different r.v.'s, but it doesn't tell us how spread out values are.
  - For example, take

$$
\begin{aligned}
X &= 0 \text{ with probability } 1 \\
Y &= \begin{cases} -1 & p = \frac{1}{2} \\ 1 & p = \frac{1}{2} \end{cases} \\
Z &= \begin{cases} -100 & p = \frac{1}{2} \\ 100 & p = \frac{1}{2} \end{cases}.
\end{aligned}
$$

  - What are the expected values? $0, 0, 0$.
  - But there is much greater spread in Z than Y and Y than X. Thus expectation is not enough to detect spread, or variation.

DEFINITION. If X is a r.v with mean $\mu = \mathbb{E}X$, then the **variance** of X, denoted by $\mathrm{Var}(X)$, is defined by

$$
\mathrm{Var}\,(X) = \mathbb{E}\left[(X-\mu)^2\right].
$$

- **Remark:** $\mathbb{E}c = c$.
- We prove an alternate formula for the variance. (The technique of using linearity is important here!!! Hint Hint)

$$
\begin{aligned}
\mathrm{Var}\,(X) &= \mathbb{E}\left[(X-\mu)^2\right] \\
&= \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mu\mathbb{E}\left[X\right] + \mathbb{E}\left[\mu^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}\left[X^2\right] - \mu^2.
\end{aligned}
$$

THEOREM. *We have that*

$$
Var\,(X) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2.
$$

- **Example1:** Calculate $\mathrm{Var}(X)$ if X represents the outcome when a fair die is rolled.
  - Solution: Previously we calculated that $\mathbb{E}X = \frac{7}{2}$.
  - Thus we only need to calculate the second moment:

$$
\begin{aligned}
\mathbb{E}X^2 &= 1^2\left(\frac{1}{6}\right) + \cdots + 6^2\frac{1}{6} \\
&= \frac{91}{6}.
\end{aligned}
$$

– Using our formula we have that

$$
\begin{aligned}
\mathrm{Var}\,(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}\,[X])^2 \\
&= \frac{91}{6} - \left(\frac{7}{2}\right)^2 \\
&= \frac{35}{12}.
\end{aligned}
$$

• Here is a useful formula:

PROPOSITION 7. *For constants a,b we have that $Var\,(aX+b) = a^2\,Var\,(X)$.*

PROOF. We compute

$$
\begin{aligned}
\mathrm{Var}\,(aX+b) &= \mathbb{E}\left[(aX+b-\mathbb{E}\,[aX+b])^2\right] \\
&= \mathbb{E}\left[(aX+b-a\mu-b)^2\right] \\
&= \mathbb{E}\left[a^2\,(X-\mu)^2\right] \\
&= a^2\mathbb{E}\left[(X-\mu)^2\right] \\
&= a^2\mathrm{Var}\,(X).
\end{aligned}
$$

$\square$

DEFINITION. We define

$$
SD\,(X) = \sqrt{\mathrm{Var}(X)}
$$

to be the **standard deviation** of X.

CHAPTER 6

Some Discrete Distributions

6.1. Bernouli and Binomial Random Variables

- **Bernoulli Distribution**
    - Suppose that a trial or experiment takes place, whose outcome is either *success* or *failure*.
    - Let $X = 1$ when the outcome is a success and $X = 0$ if it is a failure.
    - The pmf of X is given by

$$
\begin{aligned}
p(0) &= \mathbb{P}\left(X = 0\right) = 1 - p \\
p(1) &= \mathbb{P}\left(X = 1\right) = p
\end{aligned}
$$

    where $0 \leq p \leq 1$.
    - For this X, X is said to be a Bernoulli random variable with parameter p,
        * We wrtie this as $X \sim Bernoulli(p)$,
        * Properties:
            · $\mathbb{E}X = p \cdot 1 + (1 - p) \cdot 0 = p$
            · $\mathbb{E}X^2 = 1^2 \cdot p + 0^2(1 - p) = p$.
            · So $\operatorname{Var}X = p - p^2 = p(1 - p)$.
- **Binomial Distribution:**
- We say X has a binomial distribution with parameters n and p if

$$
p_X(k) = \mathbb{P}\left(X = k\right) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k \left(1 - p\right)^{n-k}.
$$

    - Interpret: $X =$ the number of successes in n *indepedent* trials.
        * Let's take this as given.
    - We say $X \sim Binomial(n, p)$ or $X \sim bin(n, p)$.
- Properties of the Binomial
    - **Check that probabilities sums to 1:** Not really a property but more of a check that X is indeed a random variable:
        * We need to check two things:
            (1) That $p_X(k) \geq 0$, and this is obvious from the fomula
            (2) Need to check that $\sum_{k=0}^{n} p_X(k) = 1$.
        * First recall the Binomial Theorem: $\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^k y^{n-k} = (x + y)^n$.
    - Then

$$
\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k \left(1 - p\right)^{n-k} = (p + (1 - p))^n = 1^n = 1.
$$

    - **Mean:** Easiest way to compute $\mathbb{E}X$ is by recognizing that $X = Y_1 + \cdots + Y_n$ where Y_i are independent Bernoulli's.

43

* Thus $\mathbb{E}X = \mathbb{E}Y_1 + \cdots + \mathbb{E}Y_n = p + \cdots + p = np$.
* We can do this directly too, but this would involve proving that

$$\mathbb{E}X = \sum_{k=0}^{n} kp(k) = np,$$

meaning we would have to prove

$$\sum_{k=0}^{n} k \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} = np.$$

- **Variance:** We first compute the second moment. As before write $X = Y_1 + \cdots + Y_n$ where Y_i are bernoulli's.

$$
\begin{aligned}
\mathbb{E}X^2 &= \mathbb{E}(Y_1 + \cdots + Y_n)^2 \\
&= \sum_{k=1}^{n} \mathbb{E}Y_k^2 + \sum_{i \neq j} \mathbb{E}[Y_i Y_j] \\
&= \sum_{k=1}^{n} p + \sum_{i \neq j} \mathbb{E}[Y_i Y_j] \\
&= np + \sum_{i \neq j} \mathbb{E}[Y_i Y_j] \\
&= (\star)
\end{aligned}
$$

* Now each term $\mathbb{E}[Y_i Y_j]$ for fixed i, j can be computed as $=$

$$
\begin{aligned}
\mathbb{E}[Y_i Y_j] &= 1 \cdot \mathbb{P}(Y_i Y_j = 1) + 0 \cdot \mathbb{P}(Y_i Y_j = 0) \\
&= \mathbb{P}((Y_i = 1) \cap (Y_j = 1)) \\
&= \mathbb{P}(Y_i = 1) \mathbb{P}(Y_j = 1), \text{ by independence} \\
&= p^2.
\end{aligned}
$$

* Now there are a total of n^2 terms is $(Y_1 + \cdots + Y_n)^2$, n of which are of form Y_k^2. Thus there are $n^2 - n$ terms of the form $Y_i Y_j$ with $i \neq j$.
* Hence using $(\star)$ we have $\mathbb{E}X^2 = np + (n^2 - n) p^2$.
* Thus

$$\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = np + (n^2 - n) p^2 - (np)^2 = np(1-p).$$

- Sumarize: $\mathbb{E}X = np$ and $\text{Var}X = np(1-p)$.
- **Moments:** We can also prove $\mathbb{E}X^k = np\mathbb{E}\left[(Y+1)^{k-1}\right]$.
* **Calculator(TI-84):**
  - 2ndDistri>binomialpdf(n, p, x)=$\mathbb{P}(X = x)$.
  - same with cdf.
* **Example1:** A company prices its hurricane insurance using the following assumptions:
  - (i) In any calendar year, there can be at most one hurricane.
  - (ii) In any calendar year, the probability of a hurricane is 0.05.
  - (iii) The numbers of hurricanes in different calendar years are mutually independent. Using the company's assumptions, calculate the probability that there are fewer than 3 hurricanes in a 20-year period

- **Solution:** We have that $X \sim bin(20, .05)$ then

$$
\begin{aligned}
\mathbb{P}\left(X < 3\right) &= \mathbb{P}\left(X \leq 2\right) \\
&= \binom{20}{0}(.05)^0 (.95)^{20} + \binom{20}{1}(.05)^1 (.95)^{19} + \binom{20}{2}(.05)^2 (.95)^{12} \\
&= .9245.
\end{aligned}
$$

- **Example2:** Phan has a .6 probability of making a free throw. Suppose each free throw is independent of the other. If he attempts 10 free throws, what is the probability that he makes at least 2 of them?

  - **Solution:** Let $X \sim bin(10, .6)$ then

$$
\begin{aligned}
\mathbb{P}\left(X \geq 2\right) &= 1 - \mathbb{P}\left(X = 0\right) - \mathbb{P}\left(X = 1\right) \\
&= 1 - \binom{10}{0}(.6)^0 (.4)^{10} - \binom{10}{1}(.6)^1 (.4)^9 \\
&= .998.
\end{aligned}
$$

6.2. The Poisson Distribution

- We say that $X = 0, 1, 2, \ldots$ is Poisson with parameter $\lambda > 0$ if

$$p_X(i) = \mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!} \ \text{ for } i = 0, 1, 2, 3, \ldots.$$

  - Or $X \sim$ Poisson(λ).
- In general Poisson random variables are of the following form
  - Suppose success happens λ times on average in a given period (per year, per month etc). Then $X =$ number of times sucess happens in that given period.
  - Possion is like binomial, excpect, X is infinitely countable!
- Examples that obey Poisson R.V
  - 1. The number of misprints on a page ogf a book
  - 2. # of people in community that survive to age 100
  - 3. # of telephone numbers that are dialed in a day.
  - 4. # of customers entering post office on a day.
- **Calc2:** Recall that $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$.
- **Properties of Poisson:** Let $X \sim Poisson(\lambda)$
  - First we check that $p_X(i)$ is indeed a pmf: First it is obvious that $p_X(i) \geq 0$ since $\lambda > 0$. We to need to check that all the probabilities add up to one:

$$\sum_{i=0}^{\infty} p_X(i) = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1.$$

  - **Mean:** We have

$$\begin{aligned}
\mathbb{E}X &= \sum_{i=0}^{\infty} i e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\lambda\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!} \\
&= e^{-\lambda}\lambda e^{\lambda} = \lambda.
\end{aligned}$$

  - **Variance:** We first have

$$\begin{aligned}
\mathbb{E}X^2 &= \sum_{i=0}^{\infty} i^2\frac{e^{-\lambda}\lambda^i}{i!} \\
&= \lambda\sum_{i=0}^{\infty} i\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
&= \lambda\sum_{j=0}^{\infty}(j+1)\frac{e^{-\lambda}\lambda^j}{j!}, \text{let } j = i-1 \\
&= \lambda\left[\sum_{j=0}^{\infty} j\frac{e^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^j}{j!}\right] \\
&= \lambda\left[\lambda + e^{-\lambda}e^{\lambda}\right] \\
&= \lambda(\lambda+1).
\end{aligned}$$

    Thus

$$\text{Var}X = \lambda(\lambda+1) - \lambda^2 = \lambda.$$

- **Example1:** Suppose on average there are 5 homicides per month in Hartford, CT. What is the probability there will be at most 1 in a certain month?
  - **Answer:** If X is the number of homicides, we are given that $EX = 5$. Since the expectation for a Poisson is $\lambda = 5$. Therefore $P(X = 0) + P(X = 1) = e^{-5} + 5e^{-5}$.
- **Example2:** Suppose on average there is one large earthquake per year in Mexico. What's the probability that next year there will be exactly 2 large earthquakes?
  - **Answer:** $\lambda = EX = 1$, so $P(X = 2) = \frac{e^{-1}}{2}$.
- **Example3:** Phan receives texts on the average of two every 3 minutes. Assume Poisson.
  - Question: What is the probability of five or more texts arriving in a 9−minute period.
  - **Answer:** Let X number of calls in a 9−minute period. Let $n =$ number of periods, $\lambda_1 = 2$ Thus $\lambda = 3 \cdot 2 = 6$. Thus

$$
\begin{aligned}
\mathbb{P}\left(X \geq 5\right) &= 1 - \mathbb{P}\left(X \leq 4\right) \\
&= 1 - \sum_{n=0}^{4} \frac{e^{-6}6^n}{n!} \\
&= 1 - .285 = .715.
\end{aligned}
$$

- **Important:** Poisson is similar to Binomial in the following way
  - FACT: Poisson approximates $Bin(n, p)$ when n is **large** and p is **small enough** so that np is of moderate size.

THEOREM 8. *If X_n is binomial with parameters n and p_n and $np_n \to \lambda$, then*

$$\mathbb{P}\left(X_n = i\right) \to \mathbb{P}\left(Y = i\right)$$

where $Y \sim Poisson(\lambda)$.

PROOF. See class textbook. □

- **Summary of Theorem:** This theorem says that suppose n is large and p is small, Thus
  - If $X \sim Bin(n, p)$ then we approximate X with a possion by letting let $\lambda = np$ so that

$$\mathbb{P}\left(X = i\right) \approx e^{-np}\frac{(np)^i}{i!}.$$

- **When can we assume X is Poisson:** Another consequence of this theorem says that when $Y =$"the number of successes in a given period". And if the number possible of trials n is large, and if the probability p of success is small, then Y can be treated as a Poisson random variable.
- NOTE:
  - (1)Why is number of misprints on a page will be approximately Poisson with $\lambda = np$
    * Let $X =$ number of misprints on a page of a book.
    * Since prob of error, say $p = .01$ is usually small, and number of letters on a page is usually large, say $n = 1000$. Then the average is $\lambda = np$.
    * Then because p is small and n is large, then X can be approximated by a Poisson.
  - (2) Let X number of accidents in a year
    * X is Poisson because the probability of an accident p in a given periord is usually small and while the number n of times someone drives in a given period is high.
- Example: Here is an example showing this.

– If X is number of times you get heads on a biased coin where $\mathbb{P}\left(H\right) = \frac{1}{100}$. Suppose you you toss 1000 times. Then $np = 10$

$$\mathbb{P}\left(X = 5\right) \approx e^{-10}\frac{10^5}{5!} = .0378$$

while the actual value is

$$
\begin{aligned}
\mathbb{P}\left(X = 5\right) &= \left(\begin{array}{c} 1000 \\ 5 \end{array}\right)\left(.01\right)^5\left(.99\right)^{995} \\
&= \frac{1000!}{995!5!}\left(.01\right)^5\left(.99\right)^{995} \\
&= .0375.
\end{aligned}
$$

6.3. Other Discrete Distributions

- **Uniform Distribution:**
  - We say X is *uniform*, and write this as $X \sim uniform(n)$, if $X \in \{1, 2, \ldots, n\}$ and

$$p_X(i) = \mathbb{P}(X = i) = \frac{1}{n} \quad \text{for } i = 1, 2, \ldots, n.$$

  - **Exercise:** $\mathbb{E}X = \sum_{i=1}^{n} i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \frac{n(n-1)}{2} = \frac{n-1}{2}$ and find $\text{Var}X$.

- **Geometric Distribution:**
  - <u>Experiment:</u> Suppose that independent trials are held until success occurs. Trials are stopped once success happens. Let p be the probabiliy of having a success in each trial.
  - Let $X =$ "number of trials required until first success occurs". Thus $X \in \{1, 2, 3, 4, \ldots\}$Here we have

$$p_X(i) = \mathbb{P}(X = i) = (1-p)^{i-1} p \quad \text{for} i = 1, 2, 3, 4 \ldots.$$

  - We say $X \sim \text{geometric}(p)$.
  - <u>Properties:</u>
    * We first double check is indeed a discrete random variable: This follows from what we know about geometric series:

$$\sum_{i=1}^{\infty} \mathbb{P}(X = i) = \sum_{i=1}^{\infty} (1-p)^{i-1} p = \frac{p}{1 - (1-p)} = 1.$$

    * **Mean:** Recall that by differentiation of the geometric series, we came up with the formula $\sum_{n=0}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}$, so that

$$\begin{aligned}
\mathbb{E}X &= \sum_{i=1}^{\infty} i \mathbb{P}(X = i) \\
&= \sum_{i=1}^{\infty} i (1-p)^{i-1} p \\
&= \frac{p}{(1 - (1-p))^2} = \frac{1}{p}.
\end{aligned}$$

    * **Variance:(Leave as Exercise for student)** Note that

$$\mathbb{E}X^2 = \sum_{i=1}^{\infty} i^2 (1-p)^{i-1} p. \quad (\star)$$

Thus we can differentiate $\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}$ again to get $\sum_{n=2}^{\infty} n(n-1) x^{n-2} = \frac{2}{(1-x)^3}$.

$*$ From this we will attempt to get $\mathbb{E}X^2$ in $(\star)$ by splitting the sum up:

$$
\begin{aligned}
\sum_{n=2}^{\infty} n\left(n-1\right)\left(1-p\right)^{n-2} &= \frac{2}{\left(1-\left(1-p\right)\right)^3} = \frac{2}{p^3}, \\
\sum_{n=2}^{\infty} n\left(n-1\right)\left(1-p\right)^{n-2}p &= \frac{2}{p^2}, \text{ now split}, \\
\sum_{n=2}^{\infty} n^2\left(1-p\right)^{n-2}p &= \frac{2}{p^2} + \sum_{n=2}^{\infty} n\left(1-p\right)^{n-2}p \\
\left(1-p\right)^{-1}\sum_{n=1}^{\infty} n^2\left(1-p\right)^{n-1}p &= \frac{2}{p^2} + \sum_{n=2}^{\infty} n\left(1-p\right)^{n-2}p + \left(1-p\right)^{-1}p \\
\left(1-p\right)^{-1}\mathbb{E}X^2 &= \frac{2}{p^2} + \left(1-p\right)^{-1}\sum_{n=1}^{\infty} n\left(1-p\right)^{n-1}p \\
&= \frac{2}{p^2} + \left(1-p\right)^{-1}\frac{1}{p}
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathbb{E}X^2 &= \frac{2\left(1-p\right)}{p^2} + \frac{1}{p} \\
&= \frac{2-2p+p}{p^2} = \frac{2-p}{p^2}
\end{aligned}
$$

$*$ So Thus

$$
\begin{aligned}
\mathrm{Var}X &= \mathbb{E}X^2 - \left(\mathbb{E}X\right)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} \\
&= \frac{\left(1-p\right)}{p^2}
\end{aligned}
$$

- **Example1:** An urn contains 10 white balls and 15 black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability.
    - **Part (a):** Exactly 6 draws are needed?
        * $X =$number of draws needed to select a black ball, the probability of sucess is
        $$p = \frac{15}{10+15} = \frac{15}{25} = .6.$$
        * Thus
        $$\mathbb{P}\left(X=6\right) = \left(.4\right)^{6-1}\left(.6\right) = .006144$$
    - **Part (a):** What is the expected number of draws in this game?
        * Since $X \sim$ geometric$(.6)$ then
        $$\mathbb{E}X = \frac{1}{p} = \frac{10}{6} = 1.\bar{6}$$
    - **Part (c)(Extra Problem to be done at home)** Find exactly that probability at least k draws are needed?

∗ We have that

$$
\begin{aligned}
\mathbb{P}\left(X \geq k\right) &= \sum_{n=k}^{\infty} \mathbb{P}\left(X = k\right) \\
&= \sum_{n=k}^{\infty} (.4)^{n-1}(.6) \\
&= (.6)(.4)^{-1}\sum_{n=k}^{\infty}(.4)^{n} \\
&= (.6)(.4)^{-1}(.4)^{k}\sum_{n=0}^{\infty}(.4)^{n} \\
&= (.6)(.4)^{k-1}\frac{1}{1-.4} \\
&= (.4)^{k-1}.
\end{aligned}
$$

- Note: This could have been done for a general p. Thus
$$
\mathbb{P}\left(X \geq k\right) = (1-p)^{k-1}.
$$

- **Negative Binomial(Need to know for Actuarial Exam):**
  - Experiment: Suppose that independent trials are held with probability p of having a success. The trials are perfomed until a total of r sucesses are accumulated.
    ∗ Let X equal the number of trials required to obtain r succeses. Here we have
    $$
    \mathbb{P}\left(X = n\right) = \binom{n-1}{r-1} p^{r}(1-p)^{n-r} \ \text{for} n = r, r+1, \ldots.
    $$
  - We say $X \sim NegativeBinomial(r, p)$.
  - Properties:
    ∗ This is a probability mass function. Can check that $\sum_{n=r}^{\infty}\mathbb{P}\left(X = n\right) = 1$.
    ∗ **Mean:**
    $$
    \mathbb{E}X = \frac{r}{p}.
    $$
    ∗ **Variance:**
    $$
    \mathrm{Var}(X) = \frac{r(1-p)}{p^{2}}.
    $$
  - Note that $Geometric(p) = NegativeBinomial\left(1, p\right)$.
- **Example:** Find the expected value of the number of times one must throw a die until the outcome 1 has occured 4 times.
  - **Solution:** $X \sim NegativeBinomial\left(4, \frac{1}{6}\right)$. So
    $$
    \mathbb{E}X = \frac{4}{\frac{1}{6}} = 24.
    $$

- **Hypergeometric Distribution(Need to know for Actuarial Exam):**
  - Experiment: Suppose that a sample of size n is to be chosen randomly (without replacement) from an urn containing N balls, of which m are white and $N-m$ are black.

∗ Let X equal the number of white balls selected. Then

$$\mathbb{P}\left(X = i\right) = \frac{\left(\begin{array}{c} m \\ i \end{array}\right)\left(\begin{array}{c} N - m \\ n - i \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)} \text{ for } n = 0, 1, \ldots, n.$$

− We say $X \sim Hypergeometric(n, N, m)$.
− Properties:
  ∗ **Mean:**
$$\mathbb{E}X = \frac{nm}{N}.$$
  ∗ **Variance:**
$$\text{Var}(X) = n\frac{m}{N}\left(1 - \frac{m}{N}\right)\left(1 - \frac{n-1}{N-1}\right).$$

CHAPTER 7

Continuous Random Variables

7.1. Intro to continuous R.V

DEFINITION. A random variable X is said to have a **continuous distribution** if there exists a non-negative function f such that

$$\mathbb{P}\left(a \leq X \leq b\right) = \int_a^b f(x)dx$$

for every a and b. [Sometimes we write that for nice sets $B \subset \mathbb{R}$ we have $\mathbb{P}\left(X \in B\right) = \int_B f(x)dx$.]

We call f the **pdf (probability density function)** for X. Sometime we we the notation f_X to signify f_X correponds to the pdf of X. We sometimes call f_X the **density** of X.

- In fact, any function f satisfying the following two properties is called a density, and could be considered a pdf of some random variable X:
  (1) $f(x) \geq 0$ for all x
  (2) $\int_{-\infty}^{\infty} f(x)dx = 1$.
- **Important Note!**
  − (1) In this case $X : S \to \mathbb{R}$ and the could attain uncountably many values (doesn't have to discrete)
  − (2) $\int_{-\infty}^{\infty} f(x)dx = \mathbb{P}\left(-\infty < X < \infty\right) = 1$.
  − (3) $\mathbb{P}\left(X = a\right) = \int_a^a f(x)dx = 0$.
  − (4) $\mathbb{P}\left(X < a\right) = \mathbb{P}\left(X \leq a\right) = F(a) = \int_{-\infty}^a f(x)dx$.
    * Recall that F is the cdf of X.
  − (5) Draw a pdf of X
    * Note that $\mathbb{P}\left(a < X < b\right)$ is just the area under the curve.
- **Remark:** What are some random variables that are considered continuous?
  − Let X be the *time* it takes it take for a student to finish a probability exam. $X \in (0, \infty)$.
  − Let X be the *value* of a Apple's stock price at the end of the day. Again $X \in [0, \infty)$.
  − Let X be the *height* of a college student.
  − Any sort of continuous measurement can be considered a continuous random variable.
- **Example1:** Suppose we are given

$$f(x) = \begin{cases} \frac{c}{x^3} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

  is the pdf of X. What must the value of c be?
  − **Solution:** We would need

$$1 = \int_{-\infty}^{\infty} f(x)dx = c \int_1^{\infty} \frac{1}{x^3}dx = \frac{c}{2},$$

thus $c = 2$.

- **Example2:** Suppose we are given

$$f_X(x) = \begin{cases} \frac{2}{x^3} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

is the pdf of X from Example1.

  - Part (a): Find the c.d.f, $F_X(x)$.
    * **Solution**: First we check thast if $x < 1$ then

$$F_x(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(y)dy = \int_{-\infty}^{x} 0dy = 0.$$

Now when $x \geq 1$ we hav e

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(y)dy \\ &= \int_{-\infty}^{1} 0dy + \int_{1}^{x} \frac{2}{y^3}dy \\ &= \int_{1}^{x} \frac{2}{y^3}dy \\ &= 1 - \frac{1}{x^2}. \end{aligned}$$

thus

$$F_X(x) = \begin{cases} 1 - \frac{1}{x^2} & x \geq 1 \\ x & x < 1 \end{cases}$$

  - Part (b): Use the cdf in Part (a) to help you find $\mathbb{P}(3 \leq X \leq 4)$.
    * **Solution:** We have

$$\begin{aligned} \mathbb{P}(3 \leq X \leq 4) &= \mathbb{P}(X \leq 4) - \mathbb{P}(X < 3) \\ &= F_X(4) - F_X(3) \\ &= \left(1 - \frac{1}{4^2}\right) - \left(1 - \frac{1}{3^2}\right) = \frac{7}{144}. \end{aligned}$$

- **Fact:** For continuous R.V we have the following useful relationship
  - Since $F(x) = \int_{-\infty}^{x} f(y)dy$ then by the fundamentat theorem of calculus(Do you remenber this form Calculus 1 or 2?)

$$F'(x) = f(x).$$

  - **This means that for continuous random variables, the derivative of the CDF is the PDF!**
- **Example3:** Let

$$f(x) = \begin{cases} ce^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find c.

  - **Solution:** $c = 2$.

7.2. Expectation and Variance

- Recall that if $p(x)$ is the pmf (density) of a discrete random variable, we had

$$\mathbb{E}X = \sum_{i=1}^{\infty} x_i p(x_i).$$

DEFINITION. If X is continuous with density $f(x)$ then

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x) dx.$$

- **Example1:** Suppose X has density

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

  Find $\mathbb{E}X$.
  - **Solution:** We have that

$$\begin{aligned} \mathbb{E}\left[X\right] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{0}^{1} x \cdot 2x \, dx \\ &= \frac{2}{3}. \end{aligned}$$

THEOREM 9. *If X and Y are continuous random variable then*
(a) $\mathbb{E}\left[X + Y\right] = \mathbb{E}X + \mathbb{E}Y$.
(b) $\mathbb{E}\left[aX\right] = a\mathbb{E}X$ where $a \in \mathbb{R}$.

PROOF. See textbook. It will be shown later. $\square$

PROPOSITION. *If X is a continuous R.V. with pdf $f(x)$, then for any real valued function g,*

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

- **Example2:** The density of X is given by

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}.$$

  Find $\mathbb{E}\left[e^X\right]$.
  - **Solution:** From the previous proposition we have that $g(x) = e^x$ in this case thus

$$\mathbb{E}e^X = \int_{0}^{2} e^x \cdot \frac{1}{2} dx = \frac{1}{2}\left[e^2 - 1\right].$$

LEMMA 10. *For nonnegative random variable $Y \geq 0$ we have*

$$\mathbb{E}Y = \int_{0}^{\infty} \mathbb{P}\left(Y > y\right) dy.$$

- **Bonus:**
  - This proof is a good practice with interchanging order of integrals in Multivariable Calculus.

PROOF. Recall that $dxdy$ means Right-Left and $dydx$ means Top-Bottom.

$$
\begin{aligned}
\int_0^\infty \mathbb{P}\left(Y > y\right) dy &= \int_0^\infty \int_y^\infty f_Y(x) dx dy \\
&= \int\int_D f_Y(x) dy dx, \text{ interchange order in Calc III} \\
&= \int_0^\infty \int_0^x f_Y(x) dy dx \text{ draw the region to do this} \\
&= \int_0^\infty x f_Y(x) dx \\
&= \mathbb{E}X.
\end{aligned}
$$

$\square$

- **Variance:**
  - Will be define in the same way as we did with discrete random variable:
  $$
  \begin{aligned}
  \text{Var}(X) &= \mathbb{E}\left[(X - \mu)^2\right] \\
  \text{Var}(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2.
  \end{aligned}
  $$
  - As before
  $$
  \text{Var}\left(aX + b\right) = a^2 \text{Var}(X).
  $$
- **Example3:** (Example 1 continued) Suppose X has density
$$
f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}.
$$
Find $\text{Var}(X)$.
  - **Solution:** From Example 1 we found $\mathbb{E}\left[X\right] = \frac{2}{3}$. Now
  $$
  \begin{aligned}
  \mathbb{E}\left[X^2\right] &= \int_0^1 x^2 \cdot 2x dx = 2\int_0^1 x^3 dx \\
  &= \frac{1}{2}.
  \end{aligned}
  $$
Thus
$$
\text{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.
$$
- **Example4:** Suppose X has density
$$
f(x) = \begin{cases} ax + b & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}.
$$
and that $\mathbb{E}\left[X^2\right] = \frac{1}{6}$. Find the values of a and b.
  - **Solution:** We need to use the fact that $\int_{-\infty}^\infty f(x) dx = 1$ and $\mathbb{E}\left[X^2\right] = \frac{1}{6}$. The first one gives us,
  $$
  1 = \int_0^1 \left(ax + b\right) dx = \frac{a}{2} + b
  $$

and the second one give us

$$\frac{1}{6} = \int_0^1 x^2 \left(ax + b\right) dx = \frac{a}{4} + \frac{b}{3}.$$

Solving these equations gives us

$$a = -2, \text{ and } b = 2.$$

7.3. The uniform Random Variable

- A continuous random variable is said to be **uniformly distributed** on the interval $[a, b]$ if

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}.$$

  - So X can only attain values in $X \in [a, b]$.
  - We say $X \sim Uniform(a, b)$.
  - The cdf is

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}.$$

- **Example1:** Suppose $X \sim Uniform(a, b)$ Part (a) Find the mean of X. Part (b) Find the variance of X.

  - **Part (a):** We compute

$$\begin{aligned} \mathbb{E}X &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}. \end{aligned}$$

    * Which makes sense right? It should be the midpoint of the interval $[a, b]$.

  - **Part(b):** We compute first the second moment

$$\begin{aligned} \mathbb{E}X^2 &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) \\ &= \frac{1}{3} \frac{1}{b-a} (b-a) \left(a^2 + ab + b^2 \right) \\ &= \frac{a^2 + ab + b^2}{3}. \end{aligned}$$

  Thus after some algebra

$$\text{Var}X = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}.$$

7.4. More practice

- Suppose we are given the pd.f.

$$f(x) = \begin{cases} 9e^{-9x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

  - **Part (a):** Set up integral to find $F_X(x)$:
    * We have for $x > 0$, that

    $$F_X(x) = \int_0^x 9e^{-9y} dy = 1 - e^{-9x},$$

    so that

    $$F_X(x) = \begin{cases} 1 - e^{-9x} & x \geq 0 \\ 0 & x < 0 \end{cases} /$$

  - **Part (b):** Set up integral to find $\mathbb{P}(1 < X < 5)$
    * $\int_1^5 9e^{-9x} dx$
  - **Part (c):** Set up integral to find $\mathbb{P}(X > 3)$
    * $\int_3^\infty 9e^{-9x} dx$.
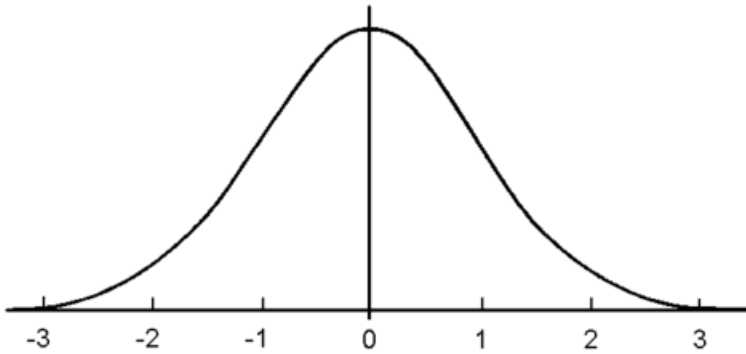 - **Part (s):** Set up integral to find $\mathbb{P}(X < 2)$
    * $\int_0^2 9e^{-9x} dx$.

CHAPTER 8

Normal Distributions

8.1. The normal distribution

- We say that X is a normal (Gaussian) random variable, or X is normally distributed with parameters μ and σ^2 if the density of X is given by

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

- We'll usually write $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.
  - Turns out that in practice, many random variable overy the normal distribution
    * Grades
    * Height of a man or a women
- Note the following:
  - If $X \sim \mathcal{N}(0,1)$ then

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$$

  - To show this we use polar coordinates. Let $I = \int_{-\infty}^{\infty} e^{-x^2/2} dx = 2\int_{0}^{\infty} e^{-x^2/2} dx$ The trick is to write

$$\begin{aligned} I^2 &= 4\int_{0}^{\infty}\int_{0}^{\infty} e^{-x^2/2} e^{-y^2/2} dxy \\ &= 4\int_{0}^{\pi/2}\int_{0}^{\infty} re^{-r^2/2} dr = 4 \cdot \frac{\pi}{2} = 2\pi, \end{aligned}$$

Thus $I = \sqrt{2\pi}$ as needed.

THEOREM 11. *To help us compute the mean and variance of X its not too hard to show $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ if and only if*

$$\frac{X - \mu}{\sigma} = Z \text{ where } Z \sim \mathcal{N}(0, 1).$$

PROOF. We only show the ($\Longleftarrow$) direction. Note that

$$
\begin{aligned}
F_X(x) &= \mathbb{P}\left(X \le x\right) = \mathbb{P}\left(\sigma Z + \mu \le x\right) \\
&= \mathbb{P}\left(Z \le \frac{x - \mu}{\sigma}\right) \\
&= F_Y\left(\frac{x - \mu}{\sigma}\right)
\end{aligned}
$$

for $\sigma > 0$. Similar for $\sigma < 0$. By the chain rule

$$
\begin{aligned}
f_X(x) &= F_X'(x) \\
&= F_Y'\left(\frac{x - \mu}{\sigma}\right)\frac{1}{\sigma} \\
&= \frac{f_Z\left(\frac{x-\mu}{\sigma}\right)}{\sigma} \\
&= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}.
\end{aligned}
$$

$\square$

- **Summary of the normal distribution:**
    - If $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ then X is normally distributed with

    $$
    \begin{aligned}
    \mathbb{E}X &= \mu, \\
    \text{Var}(X) &= \sigma^2.
    \end{aligned}
    $$

    - If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $X = \sigma Z + \mu$ where $Z \sim \mathcal{N}(0, 1)$. We call Z a **standard normal random variable**.
        * A Table of probabilities for Z will be given!!!
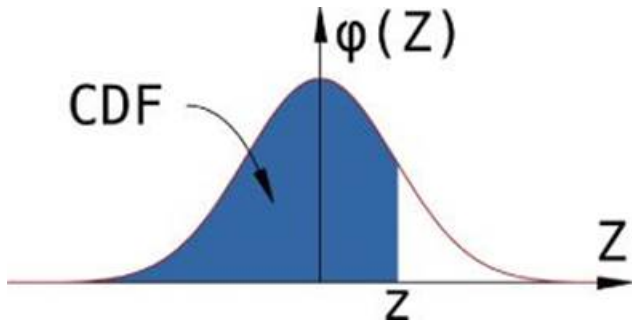   * This will be called a z-score table.
- Z scores:
    - Because $Z \sim \mathcal{N}(0, 1)$ is so important we give it's cumulative distribution function (cdf) a name. The distribution $F_Z(x)$ of Z is

    $$\Phi(x) = \mathbb{P}(Z \le x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2} dy.$$

    - **NOTE:** A table of $\Phi(x)$ will be given but only for values of $x > 0$
    - Note this is symmetric[**DRAW this**] thus here is an **important fact: $\Phi(-x) = 1 - \Phi(x)$**

THEOREM 12. *If $X \sim \mathcal{N}(\mu, \sigma)$ then*

$$\mathbb{P}\left(a < X < b\right) = \mathbb{P}\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right).$$

- **Example1:** Find $\mathbb{P}\left(1 \le X \le 4\right)$ if $X \sim \mathcal{N}\left(2, 25\right)$.
  - **Answer:** Then $\mu = 2$ and $\sigma^2 = 25$ thus $\frac{X-2}{5} = Z$ so that

$$
\begin{aligned}
\mathbb{P}\left(1 \le X \le 4\right) &= \mathbb{P}\left(\frac{1-2}{5} \le \frac{X-2}{5} \le \frac{4-2}{5}\right) \\
&= \mathbb{P}\left(-.2 \le Z \le .4\right) \\
&= \mathbb{P}\left(X \le .4\right) - \mathbb{P}\left(X \le -.2\right) \\
&= \Phi(.4) - \Phi(-2) \\
&= .6554 - \left(1 - \Phi\left(.2\right)\right) \\
&= .6554 - \left(1 - .5793\right).
\end{aligned}
$$

- **Example2:** Suppose X is normal with mean 6. If $\mathbb{P}\left(X > 16\right) = .0228$, then what is the standard deviation of X?
  - **Answer:** We apply our Theorem that says $\frac{X-\mu}{\sigma} = Z$ is $\mathcal{N}(0, 1)$ and get

$$
\begin{aligned}
\mathbb{P}\left(X > 16\right) = .0228 \quad &\Longleftrightarrow \quad \mathbb{P}\left(\frac{X-6}{\sigma} > \frac{16-6}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad \mathbb{P}\left(Z > \frac{10}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad 1 - \mathbb{P}\left(Z \le \frac{10}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad 1 - \Phi\left(\frac{10}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad \Phi\left(\frac{10}{\sigma}\right) = .9772.
\end{aligned}
$$

  Using the standard normal table we see that $\Phi\left(2\right) = .9772$, thus we must have that

$$2 = \frac{10}{\sigma}$$

  and hence $\sigma = 5$.
- **Example (Extra):** Suppose $X \sim \mathcal{N}\left(3, 9\right)$ find $\mathbb{P}\left(|X - 3| > 6\right)$.

– **Answer:** Get

$$
\begin{aligned}
\mathbb{P}\left(|X - 3| > 6\right) &= \mathbb{P}\left(X - 3 > 6\right) + \mathbb{P}\left(-\left(X - 3\right) > 6\right) \\
&= \mathbb{P}\left(X > 9\right) + \mathbb{P}\left(X < -3\right) \\
&= \mathbb{P}\left(Z > 2\right) + \mathbb{P}\left(Z < -2\right) \\
&= 1 - \Phi(2) + \Phi(-2) \\
&= 2\left(1 - \Phi(2)\right) \\
&\approx .0456.
\end{aligned}
$$

- **FACT:** The $68 - 95 - 99.7$ Rule
  - About 68% of all area is contained within 1 standard deviation of the mean
  - About 95% of all area is contained within 2 standard deviation of the mean
  - About 99.7% of all area is contained within 3 standard deviation of the mean
  - This can be explained by the following graph:

CHAPTER 9
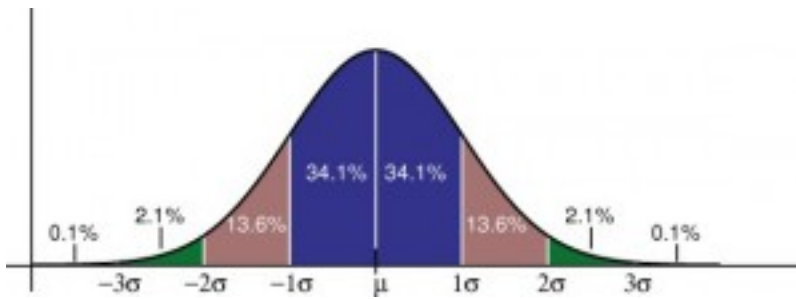
Normal approximations to the binomial

9.1. The normal approximates Binomial

THEOREM 13. *If S_n is a binomial with parameter n and p, then*

$$\mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np\,(1-p)}} \le b\right) \to \mathbb{P}\left(a \le Z \le b\right)$$

as $n \to \infty$ where Z is a $\mathcal{N}(0,1)$.

- Recall that if $S_n \sim Bin\,(n,p)$ then its mean is $\mu = np$ and standard deviation is $\sigma = \sqrt{np(1-p)}$.
  - So what this theorem says is that if you want to compute $\mathbb{P}\,(c \le S_n \le d)$ then using the fact that
    $$\frac{S_n - np}{\sqrt{np\,(1-p)}} \approx Z$$
    or
    $$\frac{S_n - \mu}{\sigma} \approx Z,$$
    then
    $$\mathbb{P}\,(c \le S_n \le d) = \mathbb{P}\left(\frac{c - \mu}{\sigma} \le \frac{S_n - \mu}{\sigma} \le \frac{d - \mu}{\sigma}\right)$$
    $$\approx \mathbb{P}\left(\frac{c - \mu}{\sigma} \le Z \le \frac{d - \mu}{\sigma}\right).$$

- Note that S_n is really **discrete**. In fact $S_n \in \{0, 1, 2, \ldots, 100\}$, while the normal distribution is **continuous**!
  - Note that if I tried to estimate an *equality*: The ***wrong way*** to do it would be:
    $$\mathbb{P}\,(S_n = i) = \mathbb{P}\left(\frac{S_n - \mu}{\sigma} = \frac{i - \mu}{\sigma}\right)$$
    $$\approx \mathbb{P}\left(Z = \frac{i - \mu}{\sigma}\right) = 0$$
    as we know that for continuous random variables X we always have $\mathbb{P}\,(X = a) = 0$!
  - Hence we need inequalities if we want to estimate a discrete random variable using a continuous random variable.
    * So we use the following convention. $\mathbb{P}(S_n = i) = \mathbb{P}\left(i - \frac{1}{2} < S_n < i + \frac{1}{2}\right)$.
    * We have no problem here, because S_n can only be integers, so we'r not hurting anything by saying " $i - \frac{1}{2} < S_n < i + \frac{1}{2}$" as we know that S_n can only be i in that interval anyways.
- **<u>Example:</u>** Suppose a fair coin is tossed 100 times.

- (a) What is the probability there will be more than 60 heads?
- **Answer:** Let $S_{100} \sim Bin\left(100, \frac{1}{2}\right)$ so that S_{100} represents the numbers of heads in 100 coin tosses.
  * The **actual answer** would be
  $$
  \begin{aligned}
  \mathbb{P}\left(S_{100} > 60\right) &= \sum_{i=61}^{100} \mathbb{P}\left(S_{100} = i\right) \\
  &= \sum_{i=61}^{100} \binom{100}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{100-i} \\
  &= \sum_{i=61}^{100} \binom{100}{i} \left(\frac{1}{2}\right)^{100}.
  \end{aligned}
  $$
  * But this would be almost impossible to do this the long way by hand.
  * So we will give an approximate answer using the normal distribution:
    · So here take $\mu = np = 50$ and $\sigma = \sqrt{np(1-p)} = \sqrt{50\frac{1}{2}} = 5$. We want more than 60, so approximate using $60 + \frac{1}{5}$:
  $$
  \begin{aligned}
  \mathbb{P}\left(S_{100} > 60\right) = \mathbb{P}\left(S_{100} \geq 60.5\right) &= \mathbb{P}\left(\frac{S_{100} - 50}{5} \geq \frac{60.5 - 50}{5}\right) \\
  &\approx \mathbb{P}\left(Z \geq 2.1\right) \\
  &\approx 1 - \Phi(2.1) \\
  &= .0179
  \end{aligned}
  $$
- (b) Estimate the probability of getting exactly 60 heads?
  $$
  \begin{aligned}
  \mathbb{P}\left(S_n = 60\right) &= \mathbb{P}\left(59.5 \leq S_n \leq 60.5\right) \\
  &\approx \mathbb{P}\left(1.9 \leq Z \leq 2.1\right) \\
  &\approx \Phi(2.1) - \Phi(1.9).
  \end{aligned}
  $$

CHAPTER 10

Some continuous distributions

10.1. Exponential Random Variables

- A continuous R.V. is said to be **exponential** with parameter λ if its pdf is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

  - We write $X \sim exponential\,(\lambda)$.
- **Summary:**
  - **CDF:** Let $a > 0$. Note that the cdf is

$$F_X(a) = \mathbb{P}\,(X \leq a) = \int_0^a \lambda e^{-\lambda y} dy = -e^{\lambda y} \mid_0^a = 1 - e^{-\lambda a}.$$

    * Thus
$$\mathbb{P}\,(X > a) = 1 - \mathbb{P}\,(X \leq a) = e^{-\lambda a}.$$

  - **Mean:** $\mathbb{E}X = \frac{1}{\lambda}$ Thus $\lambda = \frac{1}{\mu}$.
  - **Variance:** We have $\text{Var}(X) = \frac{1}{\lambda^2}$.
- How to interpret X
  - $X = $ The amount of time until some specific event occurs.
  - **Example:**
    * Time until earthquake occurs
    * Length of a phone call
    * Time until an accident happens
- **Example1:** Suppose that the length of a phone call in minutes is an exponential r.v with average length 10 minutes.
  - **Part (a)** What's probability of your phone call being more than 10 minutes?
    * **Answer:** Here $\lambda = \frac{1}{10}$ thus

$$\mathbb{P}(X > 10) = e^{-\left(\frac{1}{10}\right)10} = e^{-1} \approx .368.$$

  - **Part (b)** Between 10 and 20 minutes?
    * **Answer:** We have that

$$\mathbb{P}(10 < X < 20) = F(20) - F(10) = e^{-1} - e^{-2} \approx .233.$$

- **Exponential distribution is Memoryless (Markov)**
- **Example2:** Suppose the life of an iphone has exponential distribution with mean life of 4 years.
  - **Part(a):** What is the probability the phone lasts more than 5 years?

– **Answer:** Let X denote the life of an iphone (or time until it dies). Note that $X \sim$ $exponential(\frac{1}{4})$ since $\lambda = \frac{1}{\mu} = \frac{1}{4}$. Then

$$\mathbb{P}\left(X > 5\right) = e^{-\frac{1}{4} \cdot 5}.$$

– **Part(b):** Given that the iphone has already lasted 3 years, what is the probability that it will last another 5 more years?

– **Answer:** We compute

$$
\begin{aligned}
\mathbb{P}\left(X > 5 + 3 \mid X > 3\right) &= \frac{\mathbb{P}\left((X > 8) \cap (X > 3)\right)}{\mathbb{P}\left(X > 3\right)} \\
&= \frac{\mathbb{P}\left(X > 8\right)}{\mathbb{P}\left(X > 3\right)} \\
&= \frac{e^{-\frac{1}{4} \cdot 8}}{e^{-\frac{1}{4} \cdot 3}} \\
&= e^{-\frac{1}{4} \cdot 5}.
\end{aligned}
$$

– **Memoryless:** Note that the probability of lasting 5 more years, is the same as if it started 5 years from anew!!

• In general the **memoryless** property says that if $t, s > 0$ then

$$\mathbb{P}\left(X > t + s \mid X > t\right) = \mathbb{P}\left(X > s\right).$$

THEOREM 14. *If X is an exponential random variable, then X is memoryless.*

PROOF. To show this we have

$$
\begin{aligned}
\mathbb{P}\left(X > t + s \mid X > t\right) &= \frac{\mathbb{P}\left((X > t + s) \cap (X > t)\right)}{\mathbb{P}\left(X > t\right)} \\
&= \frac{\mathbb{P}\left(X > t + s\right)}{\mathbb{P}\left(X > t\right)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\
&= e^{-\lambda s} \\
&= \mathbb{P}\left(X > s\right),
\end{aligned}
$$

as needed. □

• **Example3:(Exam P Q29)**
  – The # of days from beginning of a calendar year until accident for a BAD driver is exponentially distributed
  – An insurance company expects 30% of BAD drivers will have an accident during first 50 days.
  – **Q:** Whats prob that a BAD driver will have Accident during first 80 days?
  – **Answer:**
  – **Step1:** Let $X \sim exp(\lambda)$ number of days until accident. We know

$$.3 = \mathbb{P}\left(X \leq 50\right) = \int_0^{50} \lambda e^{-\lambda x} dx = -e^{-\lambda t} \mid_0^{50} = 1 - e^{-50\lambda}.$$

  – Solve for λ and get $\lambda = -\frac{1}{50} \ln .7$.

– **Step2:** Then compute

$$
\begin{aligned}
\mathbb{P}\left(X \le 80\right) &= \int_0^{80} \lambda e^{-\lambda x} dx = -e^{-\lambda t} \mid_0^{80} = 1 - e^{-80\lambda} \\
&= 1 - e^{\left(\frac{80}{50}\right)\ln .7} = .435.
\end{aligned}
$$

10.2. Other Continuous Distributions

- **Gamma Distribution:**
  - We say $X \sim Gamma\,(\alpha, \lambda)$ has density
  $$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$
  where $\Gamma(\alpha)$ is the Gamma function
  $$\Gamma(\alpha) = \int_0^\infty e^{-y}y^{\alpha-1}dy.$$
  - If $Y \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \chi_n^2$, this is called the **Chi-Squared distribution**.
  - The chi-sqaure distribution is used a lot in statistics.
    * Its mean is $\mathbb{E}X = \frac{\alpha}{\lambda}$ and $\mathrm{Var}X = \frac{\alpha}{\lambda^2}$.
- **Weibull Distribution:**
  - Usefull in engineering: Look in the book for its pdf.
  - $X =$. If there is an object consisting many parts, and suppose that the object experiences death once any of tis parts fails. $X =$ lifetime of this object.
- **Cauchy Distribution:**
  - We say X is cauchy with parameter $-\infty < \theta < \infty$ if
  $$f(x) = \frac{1}{\pi}\frac{1}{1 + (x - \theta)^2}.$$
  - Importance: It does not have finite mean: That is $\mathbb{E}X = \infty$.
  - To see this, We compute for $\theta = 0$
  $$\begin{aligned} \mathbb{E}X &= \frac{1}{\pi}\int_{-\infty}^\infty \frac{x}{1 + x^2}dx \\ &\sim \frac{1}{\pi}\int_{-\infty}^\infty \frac{1}{x}dx \\ &\sim \lim_{x\to\infty}\ln|x| - \ln\lim_{x\to-\infty}|x| \end{aligned}$$
  which is not defined.

10.3. The distribution function of a Random variable

- **Fact:** For continuous R.V we have the following usefull relationship
  - Since $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy$ then by the fundamentat theorem of calculus we have

$$F'(x) = f(x).$$

- **Example1:** If X is continuous with distribution function F_X and density function f_X, find a fomula for the density function of the random varianle $Y = 2X$.
  - **Solution:** First you start with the distribution of Y:
  - **Step1:** First start by writing the cdf of Y and in terms of F_X:

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \leq x) \\ &= \mathbb{P}(2X \leq x) \\ &= \mathbb{P}\left(X \leq \frac{x}{2}\right) \\ &= F_X\left(\frac{x}{2}\right). \end{aligned}$$

  - **Step2:** Then use the relation $f_Y(y) = F_Y'(y)$ and take a derivative of both sides to get

$$\begin{aligned} F_Y'(x) &= \frac{d}{dx}\left[F_X\left(\frac{x}{2}\right)\right], \\ F_Y'(x) &= F_X'\left(\frac{x}{2}\right) \cdot \left(\frac{x}{2}\right)', \text{ by chain rule on RHS} \\ f_Y(x) &= f_X\left(\frac{x}{2}\right)\frac{1}{2}. \end{aligned}$$

- **Goal:** To be able to compute the cdf and pdf of $Y = g(X)$ where $g : \mathbb{R} \to \mathbb{R}$ is a function given that we know the cdf and pdf of X.
  - Why is this useful?
    * For example suppose X represent the income for a random US worker. And let $Y = g(X)$ be the amount of taxes a US worker pays per year. Note that taxes Y is dependent on the random variable X. So if we only care about the random varibale Y then finding its PDF and CDF can help us find out everything we need to know about Y given we can find the PDF. Recall that any probability and expected value can be found using the pdf.
- **Example2:** Let $X \sim Uniform((0, 10))$ and $Y = e^{3X}$. Find the pdf f_Y of Y.
  - **Solution:** Recall that since $X \sim Uniform((0, 1))$ then

$$f_X(x) = \begin{cases} \frac{1}{10} & 0 < x < 10 \\ 0 & \text{otherwise} \end{cases}.$$

- **Step1:** First start by writing the cdf of Y and in terms of F_X:

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) \\
&= \mathbb{P}\left(e^{3X} \le y\right), \text{ then solve for } X \\
&= \mathbb{P}(3X \le \ln y) \\
&= \mathbb{P}\left(X \le \frac{1}{3}\ln y\right) \\
&= F_X\left(\frac{1}{3}\ln y\right).
\end{aligned}
$$

- **Step2:** Then use the relation $f_Y(y) = F_Y'(y)$ and take a derivative

$$
\begin{aligned}
f_Y(y) &= F_Y'(y) \\
&= \frac{d}{dy}\left[F_X\left(\frac{1}{3}\ln y\right)\right], \text{ use chain rule} \\
&= F_X'\left(\frac{1}{3}\ln y\right)\frac{1}{3y} \\
&= f_X\left(\frac{1}{3}\ln y\right)\frac{1}{3y}, \text{ since } F_X' = f_X \\
&= \begin{cases} \frac{1}{10}\cdot\frac{1}{3y} & 0 < \frac{1}{3}\ln y < 10 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- but since

$$
\begin{aligned}
0 < \frac{1}{3}\ln y < 10 \quad &\Longleftrightarrow \quad 0 < \ln y < 30 \\
&\Longleftrightarrow \quad e^0 < y < e^{30} \\
&\Longleftrightarrow \quad 1 < y < e^{30}.
\end{aligned}
$$

- then

$$
f_Y(y) = \begin{cases} \frac{1}{30y} & 1 < y < e^{30} \\ 0 & \text{otherwise} \end{cases}.
$$

- **Example3:** Let $X \sim Uniform\left((0,1]\right)$ and $Y = -\ln X$. Find the pdf of Y? What distribution is it?
  - **Solution:** Recall that

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

  - **Step1:** First start with the cdf and write it terms of F_X

$$
\begin{aligned}
F_Y(x) &= \mathbb{P}(Y \le x) \\
&= \mathbb{P}(-\ln X \le x) \\
&= \mathbb{P}(\ln X > -x) \\
&= \mathbb{P}\left(X > e^{-x}\right) \\
&= 1 - \mathbb{P}\left(X \le e^{-x}\right) \\
&= 1 - F_X\left(e^{-x}\right).
\end{aligned}
$$

– **Step2:** Then take a derivative

$$
\begin{aligned}
f_Y(x) &= F_Y'(x) \\
&= 1 - \frac{d}{dx} F_X\left(e^{-x}\right) \\
&= -F_X'(e^{-x}) \cdot \left(-e^{-x}\right) \\
&= -f_X(e^{-x}) \cdot \left(-e^{-x}\right) \\
&= f_X(e^{-x}) \cdot e^{-x} \\
&= \begin{cases} 1 \cdot e^{-x} & 0 < e^{-x} < 1 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} e^{-x} & -\infty < -x < 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

– Thus $Y \sim exponential\,(1)$.

• **Example4:** Suppose X is uniform on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $Y = \tan X$. Find the density of Y and what known distribution is it?

– **Solution:**

– **Step1:** Find the cdf and write in terms of F_X

$$
\begin{aligned}
F_Y(x) &= \mathbb{P}\left(\tan X \leq x\right) \\
&= \mathbb{P}\left(X \leq \tan^{-1} x\right) \\
&= F_X(\tan^{-1} x)
\end{aligned}
$$

– **Step2:** Take a derivative and recall that since $\frac{1}{\frac{\pi}{2} + \frac{\pi}{2}} = \frac{1}{\pi}$ then

$$
f_X(x) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} < x < \frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases}
$$

Thus

$$
\begin{aligned}
f_Y(x) &= F_Y'(x) \\
&= \frac{d}{dx} F_X(\tan^{-1} x) \\
&= F_X'\left(\tan^{-1} x\right) \left(\tan^{-1} x\right)' \\
&= F_X'\left(\tan^{-1} x\right) \frac{1}{1+x^2} \\
&= \begin{cases} \frac{1}{\pi} \frac{1}{1+x^2} & -\frac{\pi}{2} < \tan^{-1} x < \frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} \frac{1}{\pi} \frac{1}{1+x^2} & -\infty < x < \infty \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

– Thus Y is $Cauchy(0)$.

- **Exercise:** Show that if $Z \sim \mathcal{N}(0,1)$ then $Y = Z^2$ is a Gamma with parameter $\frac{1}{2}$ and $\frac{1}{2}$.
- **Example5:**(Actuarial Exam type question) The time, T, that a manufacturing system is out of operation has cumulative distribution function

$$F(t) = \begin{cases} 1 - \left(\frac{2}{t}\right)^2 & ,t > 2 \\ 0 & \text{otherwise} \end{cases}.$$

  The resulting cost to the company is $Y = T^2$. Let f_Y be the density function for Y. Determine $f_Y(y)$, for $y > 4$.

  – **Answer:**
  – **Step1:** Find the cdf of Y is and

$$\begin{aligned} F_Y(y) &= \mathbb{P}\left(T^2 \le y\right) \\ &= \mathbb{P}\left(T \le \sqrt{y}\right) \\ &= F\left(\sqrt{y}\right) \\ &= 1 - \frac{4}{y} \end{aligned}$$

  for $y > 4$.
  – **Step2:** Take a derivative

$$\begin{aligned} f_Y(y) &= F_Y'(y) \\ &= \frac{4}{y^2}. \end{aligned}$$

- One thing to note, is that we've been using the following useful property:

PROPOSITION 15. *Suppose $g : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function, then the inverse g^{-1} exists and*

$$g(x) \le y \text{ implies } x \le g^{-1}(y).$$

CHAPTER 11

Multivariate distributions

11.1. Joint distribution functions

- We discuss the collection of random variables $(X_1, \ldots, X_n)$.
- <u>Discrete:</u>
  - For random variables X, Y we let $p(x, y)$ be the **joint probability mass(discrete density) function**

  $$p(x, y) = \mathbb{P}(X = x, Y = y).$$

    * Properties of joint pmf:
      · 1) $0 \le p \le 1$
      · 2) $\sum_i \sum_j p(x_i, y_j) = 1$
  - We also have the **<u>multivariate cdf:($\star\star$)</u>** defined by

  $$F_{X,Y}(x, y) = \mathbb{P}(X \le x, Y \le y).$$

- **<u>Example1:</u>** Experiment: Suppose you roll two 3-sided die.
  - Let X be the largest value obtained on any of the two dice. Let Y be the sum of the two dice. Find the joint pmf of X and Y.
  - **Solution:** First need to find the values of $X = 1, 2, 3$ and $Y = 2, 3, 4, 5, 6$.
  - The table for possible outcomes and their associated values (X, Y):

outcome	1	2	3
1	$(X = 1, Y = 2) = (1, 2)$	$(2, 3)$	$(3, 4)$
2	$(2, 3)$	$(2, 4)$	$(3, 5)$
3	$(3, 4)$	$(3, 5)$	$(3, 6)$

  - Using this table we have that the p.m.f. is given by:

$X\backslash Y$	2	3	4	5	6
1	$\mathbb{P}(X = 1, Y = 2) = \frac{1}{9}$	0	0	0	0
2	0	$\frac{2}{9}$	$\frac{1}{9}$	0	0
3	0	0	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

  - **Question:** Find $\mathbb{P}(X = 2 \mid Y = 4)$?
    * **<u>Answer:</u>** $\mathbb{P}(X = 2 \mid Y = 4) = \frac{1/9}{3/9} = \frac{1}{3}$.
- <u>Continuous</u>
  - For random variables X, Y we let $f(x, y)$ be the **joint probability density function**, if

  $$\mathbb{P}(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x, y) dy dx.$$

74

which is equivalent to saying that for any set $D \subset \mathbb{R}^2$ then

$$\mathbb{P}\left((X,Y) \in D\right) = \int \int_D f(x,y)dA.$$

* Properties:
   · 1) $f(x,y) \geq 0$
   · 2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1$.
– We also have the **multivariate cdf:($\star\star$)** defined by

$$F_{X,Y}(x,y) = \mathbb{P}\left(X \leq x, Y \leq y\right).$$

* Note that $F_{X,Y}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)dydx$.
– Thus note that

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

– Marginal Density: If $f_{X,Y}$ is the joint density of X, Y. We recover the **marginal densities** of X, Y respectively by the following
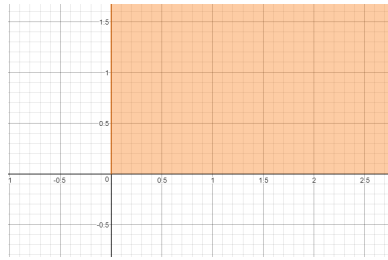
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$

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

• **Example2:** Let X, Y have joint pdf

$$f(x,y) = \begin{cases} ce^{-x}e^{-2y} & , 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}.$$

– **Part(a):** Find c that makes this a joint pdf:
   * Sol: **Step1: Draw region of Domain first!!!**
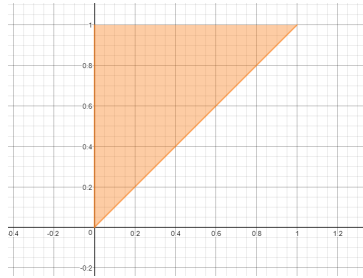

.
   * Thus

$$1 = \int_0^{\infty} \int_0^{\infty} ce^{-x}e^{-2y}dxdy = c\int_0^{\infty} e^{-2y}\left[-e^{-x}\right]_{x=0}^{x=\infty} dy$$

$$= c\int_0^{\infty} e^{-2y}dy = c\left[-\frac{1}{2}e^{-2y}\right]_0^{\infty} = c\frac{1}{2}.$$

   Then $c = 2$.
– **Part(b):** Find $\mathbb{P}\left(X < Y\right)$.
   * **Sol:** Need to draw the region (Recall Calc III!!) Let $D = \{(x,y) \mid 0 < x < y, 0 < y < \infty\}$

·

· There are two ways to set up this integral:

· Method1: To set up $dA = dydx$. We use the Top-Bottom Method:

· Where the region is bounded by

$$\text{Top Function:} y = \infty$$
$$\text{Bottom Function} y = x$$
$$\text{Range of Values} 0 \leq x \leq \infty$$

· Hence we use this information to set up

$$
\begin{aligned}
\mathbb{P}\left(X < Y\right) &= \int \int_D f(x,y)dA \\
&= \int_0^\infty \int_x^\infty 2e^{-x}e^{-2y}dydx \\
&= \int_0^\infty 2e^{-x}\frac{1}{2}\left[-e^{-2y}\right]_{y=x}^{y=\infty}dx \\
&= \int_0^\infty e^{-x}e^{-2x}dx = \int_0^\infty e^{-3x}x \\
&= \frac{1}{3}.
\end{aligned}
$$

· Method2: To set up $dA = dxdy$. We use the Right-Left Method:

· Where the region is bounded by

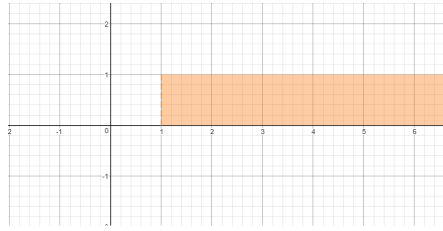$$\text{Right Function:} x = y$$
$$\text{Left Function} x = 0$$
$$\text{Range of Values} 0 \leq y \leq \infty$$

· Hence we use this information to set up

$$
\begin{aligned}
\mathbb{P}\left(X < Y\right) &= \int \int_D f(x,y)dA \\
&= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}dxdy \\
&= \text{do some work} \\
&= \frac{1}{3},
\end{aligned}
$$

which matches the answer from before.

− **Part(c):** Set up $\mathbb{P}\left(X > 1, Y < 1\right)$

* The region is given by
* Setting this up we have

$$\mathbb{P}\left(X > 1, Y < 1\right) = \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}dxdy.$$

– **Part(d):** Find the marginal $f_X(x)$:
  * **Sol:**
  * Then

$$
\begin{aligned}
f_X(x) &= \int_0^\infty f(x,y)dy = \int_0^\infty 2e^{-x}e^{-2y}dy \\
&= 2e^{-x}\left[\frac{-1}{2}e^{-2y}\right]_0^\infty = 2e^{-x}\left[0 + \frac{1}{2}\right] \\
&= e^{-x}.
\end{aligned}
$$

– **Part(e):** Find $\mathbb{E}X$ We have

$$\mathbb{E}X = \int_0^\infty xe^{-x}dx = 1$$

11.2. Independent Random Variables

- Discrete: We say discrete r.v. X, Y are **independent** if
$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y),$$
  for every x, y in the range of X and Y.
    - This is the same as saying that X, Y ar independent if the joint pmf splits into the marginal pmfs: $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$
- Continuous: We say continuous r.v. X, Y are independent if
$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$
  for any set A, B
    - This equivalent: $\mathbb{P}(X \le a, Y \le b) = \mathbb{P}(X \le a)\mathbb{P}(Y \le b)$.
    - Equivalent to $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.
- Random variables that are not independent, are said to be **dependent**.
- How can we check independence?

THEOREM 16. *Continuous (discrete) r.v. X, Y are independent if and only if their joint pdf (pmf) can be expressed as*
$$f_{X,Y}(x, y) = f_X(x)f_Y(y). \ (Continuous \ Case),$$
$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \ (Discrete \ Case).$$

PROOF. See textbook. □

- **Example1:** Let X, Y be r.v. with joint pdf
$$f(x, y) = 6e^{-2x}e^{-3y} \ \ 0 < x < \infty, 0 < y < \infty.$$
  Are X, Y independent?
    - **Solution:** Find the marginals f_X and f_Y and see if $f = f_X f_Y$. First
$$
\begin{aligned}
f_X(x) &= \int_0^\infty 6e^{-2x}e^{-3y}dy = 2e^{-2x}, \\
f_Y(y) &= \int_0^\infty 6e^{-2x}e^{-3y}dx = 3e^{-2y}.
\end{aligned}
$$
    which are both exponential. Since $f = f_X f_y$ then yes they are independent!
- **Example2:** Let X, Y have
$$f_{X,Y}(x, y) = x + y, \ \ 0 < x < 1, 0 < y < 1$$
  Are X, Y independent?
    - **Solution:** Note that there is no way to factor $x + y = f_X(x)f_Y(y)$, hence they can't be independent.
- **Example3:** Let X, Y have
$$f_{X,Y}(x, y) = 2, \ \ 0 < x < y < 1$$
- Are X, Y independent?
    - **Solution:**
        * We cannot use the previous argument to claim $f_{X,Y}$ can't split, because for example, maybe hypothetically speaking $2 = 1 \cdot 1$, so hypothetically it could split.
        * So we must find the marginal pdfs and then check if $f_{X,Y} = f_X \cdot f_Y$.

* **Important!** But whenever the domain of f is not a rectangle, you MUST draw the region of domain for $f_{X,Y}$. And here the region is $D = \{(x,y) \mid 0 < x < y < 1\}$. (Please try drawing this region on your own. If you struggle with this region, go to https://www.wolframalpha.com/ and type in $0 < x < y < 1$)
    * Note that $f_X(x) = \int_x^1 2dy = 2(1-x)$ for $0 < x < 1$
    * Then $f_Y(y) = \int_0^y 2dx = 2y$ for $0 < y < 1$.
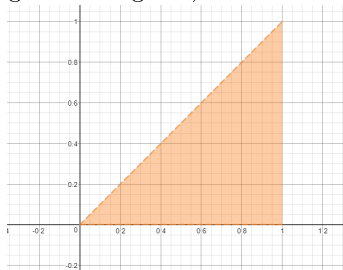   * But $f_{X,Y}(x,y) = 2 \neq f_X(x)f_Y(y) = 2(1-x)2y$!! Therefore X,Y are NOT independent.
* **Example4:** Suppose X,Y are independent uniformly distributed over $(0,1)$. Find $\mathbb{P}(Y < X)$.
    - **Solution:** Since X,Y are independent then using the Theorem form this section we have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = 1 \cdot 1,$$

for $0 \le x \le 1$ and $0 \le y \le 1$. **Draw region** ('What do you think probability will be by looking at the region?)

    *
    * and get

$$
\begin{aligned}
\mathbb{P}(Y < X) &= \int_0^1 \int_0^x f(x,y)dydx \\
&= \int_0^1 \int_0^x 1dydx = \int_0^1 xdx \\
&= \frac{1}{2}.
\end{aligned}
$$

11.3. Sums of independent Random Variables($\star$)

- **Fact:** If X, Y are independent, its not too hard to show that the **cdf** of $Z = X + Y$ is

$$
\begin{aligned}
F_{X+Y}(a) &= \mathbb{P}(X + Y \leq a) \\
&= \iint_{\{x+y\leq a\}} f_X(x) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) dx f_Y(y) dy \\
&= \int_{-\infty}^{\infty} F_X(a - y) f_Y(y) dy.
\end{aligned}
$$

   – By differentiating we have that

$$
f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y) f_Y(y) dy.
$$

- ($\star$)Here are some **interesting cases:**
- **Fact 1(Only thing I'll test you on):** If $X_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$ for $1 \leq i \leq n$ and are all independent then $Y = X_1 + \cdots + X_n \sim \mathcal{N}\left(\mu_1 + \cdots \mu_n, \sigma_1^2 + \cdots + \sigma_n^2\right)$.
   – In particular if $X \sim \mathcal{N}\left(\mu_x, \sigma_x^2\right)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ are independent then $X + Y \sim \mathcal{N}\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$ and $X - Y \sim \mathcal{N}\left(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2\right)$.
   – In general $aX \pm bY \sim \mathcal{N}\left(a\mu_x \pm b\mu_y, a^2 \sigma_x^2 + b^2 \sigma_y^2\right)$.
- **Example1:** Suppose $T \sim \mathcal{N}(95, 25)$ and $H \sim \mathcal{N}(65, 36)$ represents the grades of Tyler and Habib. Asssume their grades are independent.
   – **Part(a):** What is the probability that their average grades will be less than 90?
   – **Solution:** $T + H \sim \mathcal{N}(160, 61)$. Thus

$$
\begin{aligned}
\mathbb{P}\left(\frac{T + H}{2} \leq 90\right) &= \mathbb{P}(T + H \leq 180) \\
&= \mathbb{P}\left(Z \leq \frac{180 - 160}{\sqrt{61}}\right) = \Phi\left(\frac{180 - 160}{\sqrt{61}}\right) \\
&= \Phi(2.56) = .9961
\end{aligned}
$$

   – **Part (b):** What is the probability that Habib will have scored higher than Tyler?
   – **Solution:** Using $H - T \sim \mathcal{N}(-30, 61)$ we compute

$$
\begin{aligned}
\mathbb{P}(H > T) &= \mathbb{P}(H - T > 0) \\
&= 1 - \mathbb{P}(H - T < 0) \\
&= 1 - \mathbb{P}\left(Z \leq \frac{0 - (-30)}{\sqrt{61}}\right) \\
&= 1 - \Phi(3.84) = 1 - 1 = 0.
\end{aligned}
$$

- Other facts.
- **Fact 2:** Let $Z \sim \mathcal{N}(0, 1)$ then $Z^2 \sim \chi_1^2$.
   – If $Z_1, \ldots, Z_n$ are indepedent $\mathcal{N}(0, 1)$ then $Y = Z_1^2 + \cdots Z_n^2 \sim \chi_n^2$.

- **Fact 3:** If $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$, and they are independent, then $X + Y \sim Poisson(\lambda + \mu)$.
- List out stuff and then stop.

11.4. Conditional Distributions- Discrete(⋆)

- The **conditional pmf** for a discrete R.V. is

$$
\begin{aligned}
p_{X|Y}\left(x \mid y\right) &= \mathbb{P}\left(X = x \mid Y = y\right) \\
&= \frac{p(x,y)}{p_Y(y)}.
\end{aligned}
$$

- We also have the **condition cdf:** $F_{X|Y}\left(x \mid y\right) = \mathbb{P}\left(X \le x \mid Y = u\right)$
- Fact:
  - If X, Y are indepedent then

$$
p_{X|Y}(x \mid y) = p_X(x)
$$

- **Example1:** Suppose the joint pmf of (X, Y) is

$x\backslash y$	0	1
0	.4	.2
1	.1	.3

  - **Compute some conditional pmf:** Then the second column is

$$
p_{X|Y}\left(0 \mid 1\right) = \frac{.2}{.5} = \frac{2}{5} \text{ and } p_{X|Y}\left(1 \mid 1\right) = \frac{.3}{.5} = \frac{3}{5}.
$$

  - **Are they independent?** Note that $p_X(0) = .4 + .2 = .6 \ne p_{X|Y}\left(0 \mid 1\right)$, so no!

11.5. Conditional Distributions- Continuous($\star$)

- **Def:** If X, Y are continuous with joint pdf $f(x, y)$ then the **conditional pdf of X given $Y = y$** is defined as
$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}.$$
  - defined only when $f_Y(y) > 0$.
- **Def:** The **conditional cdf of X** given $Y = y$ is
$$
\begin{aligned}
F_{X|Y}(a \mid y) &= \mathbb{P}(X \le a \mid Y = y) \\
&= \int_{-\infty}^{a} f_{X|Y}(x \mid y)\, dx.
\end{aligned}
$$

- **Fact:** If X, Y are indepedent then
$$f_{X|Y}(x \mid y) = f_X(x).$$

- **Example1:** The joint pdf of X, Y is given by
$$f_X(x, y) = \begin{cases} \frac{12}{5} x\,(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

  Commpute the conditional pdf of X givne that $Y = y$ where $0 < y < 1$.
  - Solution: We have
$$
\begin{aligned}
f_{X|Y}(x \mid y) &= \frac{f(x, y)}{f_Y(y)} = \frac{x\,(2 - x - y)}{\int_0^1 x\,(2 - x - y)\, dy} \\
&= \frac{x\,(2 - x - y)}{\frac{2}{3} - \frac{y}{2}}.
\end{aligned}
$$

11.6. Joint PDF of functions

- **Goal:**
  - Recall that from section 5.7 we can find the pdf of a new random variable $Y = g(X)$.
  - Suppose we know the distributions of X_1, X_2 then what is the distribution of $g_1(X_1, X_2)$ and $g_2(X_1, Y_1)$
    - * For example if we know X_1, X_2 what is the distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1^2 - e^{X_1 X_2}$.
- **Steps to finding the joint cdf of new R.V. made from old ones.:**
  - Suppose X_1, X_2 are jointly distributed with pdf f_{X_1, X_2}. Let $g_2(x_1, x_2), g_2(x_2, x_2)$ be multi-variable functions.
  - **Goal:** Find the joint pdf of $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_1(X_2, X_2)$
  - **Step1:** Find the Jacobian:

$$J(x_1, x_2) = \left| \begin{array}{c} \nabla g_1 \\ \nabla g_2 \end{array} \right| = \left| \begin{array}{cc} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{array} \right| = \frac{\partial g_1}{\partial x_1} \frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2} \frac{\partial g_2}{\partial x_1} \neq 0.$$

  at all points (x_1, x_2)
  - **Step2:** Find the unique solutions of equationf $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ in terms of

$$\begin{aligned} x_1 &= h_1(y_1, y_2), \\ x_2 &= h_2(y_1, y_2). \end{aligned}$$

  - **Step3:** The joint pdf of Y_1, Y_2 is

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= f_{X_1, X_2}(x_1, x_2) \left| J(x_1, x_2) \right|^{-1} \\ &= f_{X_1, X_2}(h_1(y_1, y_2), h_2(y_1, y_2)) \left| J(x_1, x_2) \right|^{-1}. \end{aligned}$$

- **Example1:** Suppose X_1, X_2 have joint distribution

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} 2x_1 x_2 & 0 \leq x_1, x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Question: Find the joint pdf of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.
  - **Step1:** Find the Jacobian: Note that

$$\begin{aligned} y_1 &= g_1(x_1, x_2) = x_1 + x_2, \\ y_2 &= g_2(x_1, x_2) = x_1 - x_2. \end{aligned}$$

So

$$J(x_1, x_2) = \left| \begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right| = -2.$$

  - **Step2:** Solve for x_1, x_2 and get

$$\begin{aligned} x_1 &= \frac{1}{2}(y_1 + y_2), \\ x_2 &= \frac{1}{2}(y_1 - y_2). \end{aligned}$$

- **Step3:** The joint pdf of Y_1, Y_2 is given by the formula:

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2)\,|J(x_1,x_2)|^{-1} \\
&= f_{X_1,X_2}\left(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)\right)\frac{1}{|-2|} \\
&= \begin{cases} \frac{1}{2}(y_1 + y_2)(y_1 - y_2) & 0 \le \frac{1}{2}(y_1 + y_2) \le 1, \\ & 0 \le \frac{1}{2}(y_1 - y_2) \le 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- **Example2:** Suppose $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,4)$ and independent.
  - Let $Y_1 = 2X_1 + X_2$ and $Y_2 = X_1 - 3X_2$.
  - **Question:** Find the joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ of Y_1 and Y_2.
  - **Step1:** Find the Jacobian: Note that

$$
\begin{aligned}
y_1 &= g_1(x_1, x_2) = 2x_1 + x_2, \\
y_2 &= g_2(x_1, x_2) = x_1 - 3x_2.
\end{aligned}
$$

So

$$
J(x_1, x_2) = \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} = -7.
$$

  - **Step2:** Solve for x_1, x_2 and get

$$
\begin{aligned}
x_1 &= \frac{3}{7}y_1 + \frac{1}{y}y_2 \\
x_2 &= \frac{1}{7}y_1 - \frac{2}{7}y_2
\end{aligned}
$$

  - **Step3:** The joint pdf of Y_1, Y_2 is given by the formula:

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2)\,|J(x_1,x_2)|^{-1} \\
&= f_{X_1,X_2}\left(\frac{3}{7}y_1 + \frac{1}{y}y_2, \frac{1}{7}y_1 - \frac{2}{7}y_2\right)\frac{1}{7}.
\end{aligned}
$$

So we need to find the joint pdf of X_1 and X_2.

  * But since $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,4)$ and indepedent Then

$$
f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ and } f_{X_2}(x_2) = \frac{1}{\sqrt{2 \cdot 4\pi}}e^{-x^2/(2\cdot4)}.
$$

Thus by inpedence

$$
\begin{aligned}
f_{X_1,X_2}(x_1, x_2) &= f_{X_1}(x_1)f_{X_2}(x_2) \\
&= \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2 \cdot 4\pi}}e^{-x^2/(2\cdot4)}.
\end{aligned}
$$

  * Thus we have

$$
f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{\sqrt{2\pi}}e^{-\left(\frac{3}{7}y_1 + \frac{1}{y}y_2\right)^2/2}\frac{1}{\sqrt{8\pi}}e^{-\left(\frac{1}{7}y_1 - \frac{2}{7}y_2\right)^2/8}\frac{1}{7}.
$$

- **Example3(if time):** Suppose X_1, X_2 have joint distribution

$$
f_{X_1,X_2}(x_1, x_2) = \begin{cases} x_1 + \frac{3}{2}(x_2)^2 & 0 \le x_1 \le 1, 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}.
$$

Question: Find the joint pdf of $Y_1 = X_1 + X_2$ and $Y_2 = X_1^2$.

- **Step1:** Find the Jacobian: Note that

$$
\begin{aligned}
y_1 &= g_1(x_1, x_2) = x_1 + x_2, \\
y_2 &= g_2(x_1, x_2) = x_1^2.
\end{aligned}
$$

So

$$
J(x_1, x_2) = \begin{vmatrix} 1 & 2 \\ 2x_1 & 0 \end{vmatrix} = -4x_2
$$

- **Step2:** Solve for x_1, x_2 and get

$$
\begin{aligned}
x_1 &= \sqrt{y_2}, \\
x_2 &= y_1 - \sqrt{y_2}.
\end{aligned}
$$

- **Step3:** The joint pdf of Y_1, Y_2 is given by the formula:

$$
\begin{aligned}
f_{Y_1, Y_2}(y_1, y_2) &= f_{X_1, X_2}(x_1, x_2) \left| J(x_1, x_2) \right|^{-1} \\
&= f_{X_1, X_2}\left(\sqrt{y_2}, y_1 - \sqrt{y_2}\right) \frac{1}{|4x_2|} \\
&= \begin{cases} \frac{1}{|4x_2|} \left[\sqrt{y_2} + \frac{3}{2}\left(y_1 - \sqrt{y_2}\right)^2 \right] & \begin{array}{l} 0 \leq \sqrt{y_2} \leq 1, \\ 0 \leq y_1 - \sqrt{y_2} \leq 1 \end{array} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

CHAPTER 12

Expectations

12.1. Expectation of Sums of R.V.

THEOREM 17. *Let $g : \mathbb{R}^2 \to \mathbb{R}$. If X, Y have joint pmf $p(x, y)$ then*

$$\mathbb{E}\left[g\left(X, Y\right)\right] = \sum_y \sum_x g(x, y)p(x, y).$$

If X, Y have joint pdf $f(x, y)$ then

$$\mathbb{E}\left[g(X, Y)\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy.$$

- **Example1:** Suppose the joint p.m.f of X and Y is given by

$X\backslash Y$	0	2
0	.2	.7
1	0	.1

  . Find $\mathbb{E}\left[XY\right]$.

  - **Solution:** Using the formula we have with the function $g(x, y) = xy$:

$$
\begin{aligned}
\mathbb{E}\left[XY\right] &= \sum_{i,j} x_i y_j p(x_i, y_j) \\
&= 0 \cdot 0 p(0, 0) + 1 \cdot 0 p(1, 0) + 0 \cdot 2 p(0, 2) + 1 \cdot 2 p(1, 2) \\
&= 0 \cdot 0 \cdot .2 + 1 \cdot 0 \cdot 0 + 0 \cdot 2 \cdot .7 + 1 \cdot 2 \cdot .1 \\
&= .2
\end{aligned}
$$

- **Example2:** Suppose X, Y are independent exponential r.v. with parameter $\lambda = 1$. Set up a double integral that represents

$$\mathbb{E}\left[X^2 Y\right].$$

  - **Solution:** Since X, Y are independent then

$$f_{X,Y}(x, y) = e^{-1x}e^{-1y} = e^{-(x+y)}. \ \ 0 < x, y < \infty.$$

  - Then **DRAW FIRST** then

$$\mathbb{E}\left[X^2 Y\right] = \int_0^\infty \int_0^\infty x^2 y e^{-(x+y)} dy dx.$$

- **Example3:** Suppose the joint pdf of X, Y is

$$f(x, y) = \begin{cases} 10xy^2 & 0 < x < y, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Find $\mathbb{E}XY$ and $\text{Var}\left(Y\right)$.

  - **Solution:**

87

– We **First DRAW** and then set up

$$\mathbb{E}XY = \int_0^1 \int_0^y xy \left(10xy^2\right) dxdy = 10 \int_0^1 \int_0^y x^2y^3 dxdy$$
$$= \frac{10}{3} \int_0^1 y^3y^3 dy = \frac{10}{3}\frac{1}{7} = \frac{10}{21}.$$

– First note that $\operatorname{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$.
– Then

$$\mathbb{E}Y^2 = \int_0^1 \int_0^y y^2 \left(10xy^2\right) dxdy = 10 \int_0^1 \int_0^y y^4 x dxdy$$
$$= 5 \int_0^1 y^4y^2 dy = \frac{5}{7}.$$

and

$$\mathbb{E}Y = \int_0^1 \int_0^y y \left(10xy^2\right) dxdy = 10 \int_0^1 \int_0^y y^3 x dxdy$$
$$= 5 \int_0^1 y^3y^2 dy = \frac{5}{6}.$$

So that $\operatorname{Var}(Y) = \frac{5}{7} - \left(\frac{5}{6}\right)^2 = \frac{5}{252}$.

THEOREM 18. *(Properties of Expectation)*
(a) $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$
(b) If $X \leq Y$ then $\mathbb{E}X \leq \mathbb{E}Y$.

PROOF. Part (a) was proved for the discrete case. So we only need to show the continuous case:

$$\mathbb{E}[X + Y] = \int\int (x + y) f(x,y)dydx$$
$$= \int\int xf(x,y)dydx + \int\int yf(x,y)dydx$$
$$= \int xf_X(x)dx + \int yf_Y(y)dy$$
$$= \mathbb{E}X + \mathbb{E}Y.$$

□

- **Example4:** Let $X_1, \ldots, X_n$ be independent and **identically distributed random (i.i.d.)** random variables. Suppose $\mathbb{E}X_i = \mu$. We call the quantity

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

the **sample mean.** Compute $\mathbb{E}[\bar{X}]$.

   – **Solution:** We use the properties of expectation

$$
\begin{aligned}
\mathbb{E}\left[\bar{X}\right] &= \mathbb{E}\left[\sum_{i=1}^{n}\frac{X_i}{n}\right] \\
&= \frac{1}{n}\mathbb{E}\left[X_1+\cdots+X_n\right] \\
&= \frac{1}{n}\left(\mathbb{E}\left[X_1\right]+\cdots+\mathbb{E}\left[X_n\right]\right) \\
&= \frac{1}{n}\left(\mu+\cdots+\mu\right)=\frac{n\mu}{n} \\
&= \mu.
\end{aligned}
$$

   – **In statistics,** the sample mean is used to estimate the actual mean of a distribution.

THEOREM 19. *If X,Y are independent then*

$$\mathbb{E}\left[XY\right]=\left(\mathbb{E}X\right)\left(\mathbb{E}Y\right).$$

PROOF. In the continuosuc case we have

$$
\begin{aligned}
\mathbb{E}\left[XY\right] &= \int\int xy f_{X,Y}(x,y)dydx \\
&= \int\int xy f_X(x)f_Y(y)dydx \\
&= \left(\int x f_X(x)dx\right)\left(\int y f_Y(y)dy\right) \\
&= \left(\mathbb{E}X\right)\left(\mathbb{E}Y\right).
\end{aligned}
$$

The discrete case is the same, except replace integrals with summations. $\qquad\square$

   • In general, the following is true:

THEOREM 20. *If X,Y are indepedent and $g,h:\mathbb{R}\to\mathbb{R}$ then*

$$\mathbb{E}\left[g\left(X\right)h\left(Y\right)\right]=\mathbb{E}\left[g\left(X\right)\right]\mathbb{E}\left[h\left(Y\right)\right].$$

   • **Mixed Discrete/Continuous Random Variables:**

THEOREM 21. *Let $g:\mathbb{R}^2\to\mathbb{R}$. Let X be a discrete random variable and Y be a continuous random variable. If X has pmf $p_X(x)$, Y has joint pdf $f_Y(y)$, and X,Y are independent then*

$$\mathbb{E}\left[g\left(X,Y\right)\right]=\sum_{i=1}^{\infty}\int_{-\infty}^{\infty}g(x_i,y)p_X(x)f_Y(y)dy,$$

provided the right handside exists.

12.2. Covariance and Correlations.

- Note that $\mathbb{E}X$ and $\mathrm{Var}X$ give information about a single random variable.
- What statistic can give us information about how X effects Y, or vice versa?

DEFINITION. The **covariance** between X and Y, is defined by

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right].$$

- After some algebra one can show that

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}\left[XY\right] - \mathbb{E}X\mathbb{E}Y.$$

- The covariance between two random variables give us information about relationship between the random variables.
    - Covariance is a measure of how much two random variables change together.
    - If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, i.e., the variables tend to show similar behavior, the covariance is positive.
        * Thus convariance measures if there is a linear relationship betwen X and Y.
    - The sign of the covariance therefore shows the tendency in the **linear** relationship between the variables.
    - For example, the following plots shows a *positive* linear relationship between X and Y:
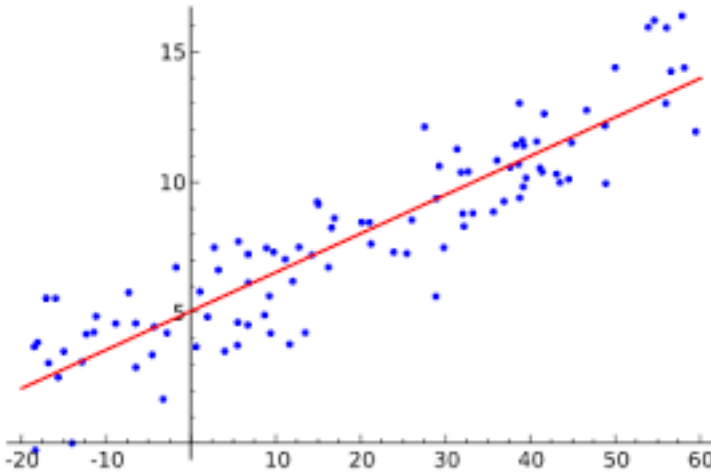
In this case $\mathrm{Cov}\,(X,Y) > 0$.

- Note: If X, Y are independent then $\mathrm{Cov}\,(X,Y) = 0$. (This is not true in the other direction. Meaning $\mathrm{Cov}\,(X,Y) = 0$ does not imply that X, Y are independent!)
    - So $\mathrm{Cov}\,(X,Y) = 0$ means they are **uncorrelated**.
- **Properties:**
    - (i) $\mathrm{Cov}\,(X,Y) = \mathrm{Cov}\,(Y,X)$
    - (ii) $\mathrm{Cov}\,(X,X) = \mathrm{Var}\,(X)$
    - (iii) $\mathrm{Cov}\,(aX,Y) = a\mathrm{Cov}\,(X,Y)$
    - (iv) $\mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \mathrm{Cov}\,(X_i, Y_j)$.

THEOREM 22. $(\star)$ *Formula for Sum of Variation:*

$$Var\,(X + Y) = Var\,(X) + Var\,(Y) + 2Cov\,(X,Y).$$

Gives us a formula for variation of $X_1, \ldots, X_n$:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var\left(X_i\right) + 2\sum\sum_{i<j} Cov\left(X_i, X_j\right).$$

- **Fact:** Note that if X, Y are independent then

$$\mathrm{Var}\left(X + Y\right) = \mathrm{Var}(X) + \mathrm{Var}\left(Y\right).$$

- Finally we have the following: Its standardized way to know how correlated two random variables are:

DEFINITION. The **correlation coefficient** of two random variables X and Y, denoted by $\rho(X, Y)$ is defined by

$$\rho\left(X, Y\right) = \frac{\mathrm{Cov}\left(X, Y\right)}{\sqrt{\mathrm{Var}\left(X\right)\mathrm{Var}\left(Y\right)}}.$$

- **Fact:**
  - (1) $-1 \leq \rho(X, Y) \leq 1$
  - (2) If $\rho(X, Y) = 1$ then $Y = a + bX$ where $b = \frac{\sigma_y}{\sigma_x} > 0$ (Straight positive sloped line)
  - (3) If $\rho(X, Y) = -1$ then $Y = a + bX$ where $b = -\frac{\sigma_y}{\sigma_x} < 0$ (Straight negatively sloped line)
  - (4) This ρ is a measure of linearity between Y and X.
    * $\rho > 0$ positive linearity: Meaning that if you were to **draw a line of best fit**, then it must be a positive sloped line
      · The closer ρ gets to 1, the more (X, Y) seems to be in a positive sloped straight line
    * $\rho < 0$ negative linearity: Meaning that if you were to draw a line of best fit, then it must be a negative sloped line
      · The closer ρ gets to -1, the more (X, Y) seems to be in a negative sloped straight line
  - (5) If $\rho\left(X, Y\right) = 0$, then X and Y are uncorrelated.
- **Warning:**
  - $\rho\left(X, Y\right)$ does not pick up any other relationship, such as quadratic, or cubic
  - $\rho(X, Y)$ is not the slope of the line of best fit. It is simply tell us if it's positive, or negative relationship, and the strength of relationship.
- **Example1:**Suppose X, Y are random variables whose joint pdf is given by

$$f(x, y) = \begin{cases} \frac{1}{y} & 0 < y < 1, 0 < x < y \\ 0 & \text{otherwise} \end{cases}.$$

  - Part (a): Find the covariance of X and Y.
  - Part (b) Compute $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$.
  - Part (c) Calculate $\rho(X, Y)$.
  - **Solution:**
  - **Part (a):** Find the covariance of X and Y.

– Recall that $\mathrm{Cov}\,(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y.$ So

$$
\begin{aligned}
\mathbb{E}XY &= \int_0^1 \int_0^y xy\frac{1}{y}dxdy = \int_0^1 \frac{y^2}{2}dy = \frac{1}{6} \\
\mathbb{E}X &= \int_0^1 \int_0^y x\frac{1}{y}dxdy = \int_0^1 \frac{y}{2}dy = \frac{1}{4}. \\
\mathbb{E}Y &= \int_0^1 \int_0^y y\frac{1}{y}dxdy = \int_0^1 ydy = \frac{1}{2}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathrm{Cov}\,(X,Y) &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y \\
&= \frac{1}{6} - \frac{1}{4}\frac{1}{2} \\
&= \frac{1}{24}.
\end{aligned}
$$

– **Part (b):** Compute Var(X) and Var(Y).
– We have that

$$
\begin{aligned}
\mathbb{E}X^2 &= \int_0^1 \int_0^y x^2\frac{1}{y}dxdy = \int_0^1 \frac{y^2}{3}dy = \frac{1}{9}. \\
\mathbb{E}Y^2 &= \int_0^1 \int_0^y y^2\frac{1}{y}dxdy = \int_0^1 y^2dy = \frac{1}{3}.
\end{aligned}
$$

– Thus recall that

$$
\begin{aligned}
\mathrm{Var}\,(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\
&= \frac{1}{9} - \left(\frac{1}{4}\right)^2 = \frac{7}{144}
\end{aligned}
$$

Also

$$
\begin{aligned}
\mathrm{Var}\,(Y) &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \\
&= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.
\end{aligned}
$$

– **Part (c):** Calculate $\rho(X,Y)$.
– We now use

$$
\begin{aligned}
\rho\,(X,Y) &= \frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}} \\
&= \frac{\frac{1}{24}}{\sqrt{\left(\frac{7}{144}\right)\left(\frac{1}{12}\right)}} \approx .6547.
\end{aligned}
$$

CHAPTER 13

Moment generating functions

13.1. Moment Generating Functions

- For each random variable X, we can define its **moment generating function** $m_X(t)$ by

$$
\begin{aligned}
m_X(t) &= \mathbb{E}\left[e^{tX}\right] \\
&= \begin{cases} \sum_{x_i} e^{tx_i} p(x_i) & \text{,if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) ds & \text{,if } X \text{ is continuous} \end{cases}.
\end{aligned}
$$

- $m_X(t)$ is called the moment generating function (m.g.f.) because we can find all the moments of X by differentiating $m(t)$ and then evaluating at $t = 0$.
- Note that

$$
\begin{aligned}
m'(t) &= \frac{d}{dt}\mathbb{E}\left[e^{tX}\right] \\
&= \mathbb{E}\left[\frac{d}{dt}e^{tX}\right] \\
&= \mathbb{E}\left[Xe^{tX}\right].
\end{aligned}
$$

Now evaluate at $t = 0$ and get

$$
m'(0) = \mathbb{E}\left[Xe^{0\cdot X}\right] = \mathbb{E}\left[X\right].
$$

- Similarly,

$$
\begin{aligned}
m''(t) &= \frac{d}{dt}\mathbb{E}\left[Xe^{tX}\right] \\
&= \mathbb{E}\left[X^2 e^{tX}\right]
\end{aligned}
$$

so that

$$
m''(0) = \mathbb{E}\left[X^2 e^0\right] = \mathbb{E}\left[X^2\right].
$$

THEOREM 23. *For all $n \geq 0$ we have*

$$
\mathbb{E}\left[X^n\right] = m^{(n)}(0).
$$

- Examples of Moment generating Functions
- **Bernoulli:** Recall that $p(1) = p$ and $p(0) = 1 - p$. Thus

$$
\begin{aligned}
m_X(t) &= \mathbb{E}e^{tX} = e^{t\cdot 0}p(0) + e^{t\cdot 1}p(1) \\
&= pe^t + (1 - p).
\end{aligned}
$$

- **Binomial:** Recall that $X \sim Bin(n,p)$ if $X = \sum_{i=0}^{n} Y_i$ where $Y_i \sim Bern(p)$ thus

$$
\begin{aligned}
m_X(t) &= \mathbb{E}e^{tX} = \mathbb{E}e^{tX \sum_{i=0}^{n} Y_i} \\
&= \mathbb{E}\left[\left(e^{tY_1}\right) \cdots \left(e^{tY_n}\right) \right] \\
&= \mathbb{E}\left[e^{tY_1} \right] \cdots \mathbb{E}\left[e^{tY_n} \right], \text{ by independence} \\
&= \left(pe^t + (1-p)\right)^n
\end{aligned}
$$

- **Poisson:** If $X \sim Poisson(\lambda)$ then

$$
m_X(t) = \mathbb{E}e^{tX} = \sum_{n=0}^{\infty} e^{tn} e^{-\lambda} \frac{\lambda^n}{n!}
$$

$$
= e^{-\lambda} \sum_{n=0}^{\infty} e^{tn} \frac{\lambda^n}{n!}
$$

$$
= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^t \lambda)^n}{n!}
$$

now recall from Calculus 2 that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ so that

$$
m_X(t) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{x^n}{n!}, \text{ with } x = e^t \lambda
$$

$$
= e^{-\lambda} e^{e^t \lambda}
$$

$$
= e^{e^t \lambda - \lambda}
$$

$$
= \exp\left(\lambda \left(e^t - 1 \right) \right)
$$

- **Exponential:** If $X \sim exp(\lambda)$ then

$$
\begin{aligned}
m_X(t) &= \mathbb{E}e^{tX} \\
&= \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx \\
&= \frac{\lambda}{\lambda - t},
\end{aligned}
$$

which is valied whenever $t > \lambda$.

- **Standard Normal:** If $X \sim \mathcal{N}(0,1)$ then

$$
\begin{aligned}
m_X(t) &= \mathbb{E}e^{tX} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx \\
&= e^{t^2/2}.
\end{aligned}
$$

How did we obtain that $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = e^{t^2/2}$? The trick is to use $1 = e^{t^2/2} e^{-t^2/2}$, hence

$$
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} e^{tx} e^{-x^2/2} dx
$$

$$
= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t-x)^2/2} dx
$$

$$
= e^{t^2/2} \cdot 1,
$$

and use the fact that since the function $f(x) = \frac{1}{\sqrt{2\pi}}e^{-(t-x)^2/2}$ is the pdf of a normal random variable with mean t and variance 1, then $\int_{-\infty}^{\infty} f(x)dx = 1$.

- **Normal:** If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $X = \mu + \sigma Z$ so that

$$
\begin{aligned}
m_X(t) &= \mathbb{E}e^{tX} \\
&= \mathbb{E}e^{t\mu}e^{t\sigma Z} = e^{t\mu}\mathbb{E}e^{(t\sigma)Z} \\
&= e^{t\mu}m_X(t\sigma) = e^{t\mu}e^{(t\sigma)^2/2} \\
&= \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right).
\end{aligned}
$$

- **Property:** Suppose X, Y are independent then what is that m.g.f. of $X + Y$?
  - Let's try to figure out:

$$
\begin{aligned}
m_{X+Y}(t) &= \mathbb{E}e^{t(X+Y)} = \mathbb{E}\left(e^{tX}e^{tY}\right) \\
&= \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right), \text{ by independence} \\
&= m_X(t)m_Y(t).
\end{aligned}
$$

  - Thus we know that

$$
m_{X+Y}(t) = m_X(t)m_Y(t).
$$

- **Note:** Note that if $f_X(x)$ is the pdf of a r.v. then it's m.g.f is

$$
\mathbb{E}e^{tX} = \int e^{tx}f_X(x)dx.
$$

This is similar to the **laplace transform** of $f_X(x)$. $[\mathcal{L}\left[f\right](s) = \int e^{-sx}f_X(x)dx]$.
  - Recall that there is one to one correspondence of laplace transforms that completely determines a function.

THEOREM 24. *If $m_X(t) = m_Y(t) < \infty$ for all t in an interval, then X and Y have the same distribution. That is, m.g.fs completely determines the distribution.*

- **Example1:** Suppose that m.g.f of X is given by $m(t) = e^{3(e^t-1)}$. Find $\mathbb{P}\left(X = 0\right)$.
  - **Solution:** (We want to work backwards). Match this m.g.f to a known m.g.f in our table. Looks like

$$
m(t) = e^{3(e^t-1)} = e^{\lambda(e^t-1)} \quad \text{where } \lambda = 3.
$$

Thus $X \sim Poisson(3)$. Thus

$$
\mathbb{P}\left(X = 0\right) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-3}.
$$

- **Summary:**
  (1) $m(t) = \mathbb{E}e^{tX}$. We have a table of mgf of distributions:
  (2) The m.g.f helps us find moments: $\mathbb{E}\left[X^n\right] = m^{(n)}\left(0\right)$
  (3) If X, Y are independent then $m_{X+Y}(t) = m_X(t)m_Y(t)$.
  (4) The m.g.f. helps us determine the distribution of random variables. If $m_X(t) = m_Y(t)$ then X and Y have the same distribution.
- Recall we had a section on sums of independent random variables.
- **Example2:** Recall $X \sim \mathcal{N}\left(\mu_x, \sigma_x^2\right)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, independent. Then what is

$$
X + Y \sim \mathcal{N}(?, ?)
$$

– **Sol:** Note that

$$
\begin{aligned}
m_{X+Y}(t) &= m_X(t)m_Y(t) \\
&= \exp\left(t\mu_x + \frac{t^2\sigma_x^2}{2}\right)\exp\left(t\mu_y + \frac{t^2\sigma_y^2}{2}\right) \\
&= \exp\left(t\left(\mu_x + \mu_y\right) + \frac{t^2\left(\sigma_x^2 + \sigma_y^2\right)}{2}\right).
\end{aligned}
$$

So then you look at our table and check which distribution has this mg.f. with $\mu = \mu_x + \mu_y$ and $\sigma^2 = \sigma_x^2 + \sigma_y^2$. so that $X + Y \sim \mathcal{N}\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$

• **Example3:** Suppose $X \sim bin(n,p)$ and $Y \sim bin(m,p)$, independent, then what is the distribution of $X + Y$?

– **Solution:** We use

$$
\begin{aligned}
m_{X+Y}(t) &= m_X(t)m_Y(t) \\
&= \left(pe^t + (1-p)\right)^n \left(pe^t + (1-p)\right)^m \\
&= \left(pe^t + (1-p)\right)^{n+m}.
\end{aligned}
$$

Look at the table and see what distribution has this m.g.f. Thus

$$X + Y \sim bin(n+m,p).$$

• **Example4:** Suppose X is a **discrete random variable** and has the m.g.f.

$$m_X(t) = \frac{1}{7}e^{2t} + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{8t}.$$

Question: What is the p.m.f of X? Find $\mathbb{E}X$.

– **Solution(a):** This doesn't match any of the known mg.f.s. Thus we can read off from the mgf that since

$$\frac{1}{7}e^{2t} + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{8t} = \sum_{i=1}^{4} e^{tx_i}p(x_i)$$

then $p(2) = \frac{1}{7}, p(3) = \frac{3}{7}, p(5) = \frac{2}{7}$ and $p(8) = \frac{1}{7}$.

– **Solution(b):** First

$$m'(t) = \frac{2}{7}e^{2t} + \frac{9}{7}e^{3t} + \frac{10}{7}e^{5t} + \frac{8}{7}e^{8t},$$

so that

$$\mathbb{E}\left[X\right] = m'(0) = \frac{2}{7} + \frac{9}{7} + \frac{10}{7} + \frac{8}{7} = \frac{29}{7}.$$

• **Example5:** Suppose X has m.g.f

$$m_X(t) = (1 - 2t)^{-\frac{1}{2}} \text{ for } t < \frac{1}{2}.$$

Find the first and second moments of X.

– **Solution:** We have

$$
\begin{aligned}
m'_X(t) &= -\frac{1}{2}\left(1 - 2t\right)^{-\frac{3}{2}}(-2) = \left(1 - 2t\right)^{-\frac{3}{2}}, \\
m''_X(t) &= -\frac{3}{2}\left(1 - 2t\right)^{-\frac{5}{2}}(-2) = 3\left(1 - 2t\right)^{-\frac{5}{2}}.
\end{aligned}
$$

So that

$$\begin{aligned}
\mathbb{E}X &= m'_X(0) = (1 - 2 \cdot 0)^{-\frac{3}{2}} = 1, \\
\mathbb{E}X^2 &= m''_X(0) = 3\,(1 - 2 \cdot 0)^{-\frac{5}{2}} = 3.
\end{aligned}$$

CHAPTER 14

Limit Laws

14.1. The Central Limit Theorem

- The CLT is one of the most remarkable theorems in Probability.
  - It helps us understand why the emperical frequencies of so many natural populations exihibit bell-shaped (normal) curves.
- Recall that *i.i.d.* means independent and identically distributed random variables.

THEOREM 25. *(CLT) Let $X_1, X_2, X_3 \ldots$ be i.i.d. each with mean μ and variance σ^2. Then the distribution of*

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal Z as $n \to \infty$. This means

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \mathbb{P}\left(Z \le b\right).$$

- **How we use CLT:** That is,

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le b\right) \approx \mathbb{P}\left(Z \le b\right) = \Phi(b).$$

  when n is large.
- The CLT helps us approximate the probability of **anything** involving $X_1 + \cdots + X_n$ where X_i are independent and identically distributed.
- When approximating discrete distributions: USE the $\pm.5$ continuity correction:
- **Example1:** If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive.
  - **Solution:** Let X_i denote the value of the ith die. Recall that

$$\mathbb{E}\left(X_i\right) = \frac{7}{2} \quad \mathrm{Var}(X_i) = \frac{35}{12}.$$

  Take

$$X = X_1 + \cdots + X_{10}$$

  to be their sum.
  - Using the CLT we need with $n = 10$,

$$
\begin{aligned}
n\mu &= 10 \cdot \frac{7}{2} = 35 \\
\sigma\sqrt{n} &= \sqrt{\frac{350}{12}}
\end{aligned}
$$

thus using the continuity correction, then

$$
\begin{aligned}
\mathbb{P}\left(29.5 \le X \le 40.5\right) &= \mathbb{P}\left(\frac{29.5 - 35}{\sqrt{\frac{350}{12}}} \le \frac{X - 35}{\sqrt{\frac{350}{12}}} \le \frac{40.5 - 35}{\sqrt{\frac{350}{12}}}\right) \\
&\approx \mathbb{P}\left(-1.0184 \le Z \le 1.0184\right) \\
&= \Phi\left(1.0184\right) - \Phi\left(-1.0184\right) \\
&= 2\Phi\left(1.0184\right) - 1 = .692.
\end{aligned}
$$

- **Example2:** An instructor has 1000 exams that will be graded in sequence.
  - The times required to grade exam exam are i.i.d. with $\mu = 20$ minutes and $\sigma = 4$ minutes.
  - Approximate prob that the intructor will grade at least 25 exams in the first 450 minutes of work.
  - **Solution:**
  - Let X_i be the time it takes to grade exam i. Then
  $$
  X = X_1 + \cdots + X_{25}
  $$
  is the time it takes to grade the first 25 exams. We want $\mathbb{P}\left(X \le 450\right)$.
  - Use CLT,
  $$
  \begin{aligned}
  n\mu &= 25 \cdot 20 = 500 \\
  \sigma\sqrt{n} &= 4\sqrt{25} = 20.
  \end{aligned}
  $$
  - Thus
  $$
  \begin{aligned}
  \mathbb{P}\left(X \le 450\right) &= \mathbb{P}\left(\frac{X - 500}{20} \le \frac{450 - 500}{20}\right) \\
  &\approx \mathbb{P}\left(Z \le -2.5\right) \\
  &= 1 - \Phi(2.5) \\
  &= .006.
  \end{aligned}
  $$