Chapter 1 — Combinatorial Analysis
(The mathematical theory of counting)

Ex: A communication system consists of a row of n antennas. m of them are defective, but the system is still functional as long as no two consecutive antennas are defective.

Q: If $n=4$ and $m=2$, what is the probability that the system will function?

A: The possible configurations are
$$(0\,1\,1\,0)$$
$$(0\,1\,0\,1)$$
$$(1\,0\,1\,0)$$
$$(0\,0\,1\,1)$$
$$(1\,0\,0\,1)$$
$$(1\,1\,0\,0).$$

Because 3 of the 6 configurations are functional, it seems reasonable to say the probability is $\frac{3}{6} = 50\%$.

Many problems in probability can be solved by counting the number of ways an event can occur.

Ex: A certain community consists of 10 women each of which has exactly 3 children.

Q: If one woman and one of her children are to be chosen as mother & daughter of the year, how many choices are possible?

A: We will denote a possible choice by a pair such as (3,2), which represents the $3^{\underline{rd}}$ mother and her $2^{\underline{nd}}$ child. Then, we can list the possibilities as

$$\begin{array}{ccc} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \\ \vdots & \vdots & \vdots \\ (10,1) & (10,2) & (10,3) \end{array}$$

We can now see that there are $10 \times 3 = 30$ different choices.

This technique is called "The basic principle of counting." In general, if there are m choices for the first element of the pair and if, for each of these there are n choices for the second element, then there are mn different <u>ordered</u> pairs. Note that for <u>ordered</u> pairs, $(1,2) \neq (2,1)$.

Ex: A 7-place license plate uses the first 3 places for letters and the last 4 places for numbers

Q: How many different plates are possible?

A: We can generalize the basic principle of counting to 3-tuples, 4-tuples, or in this case 7-tuples. One possible plate is (AAF 127). There are $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10$ (about 176 million) different possible plates

Ex: A store owner has 3 shelves for chocolate. He wants to put Milka, Cadbury, and Lindt

Q: on separate shelves. How many different arrangements (permutations) are possible?

A: One possibility is (M, C, L). To find the number of possible permutations, we use the basic principle of counting. There are 3 choices for what to put on the first shelf. Then, after placing one type on the first shelf, there are 2 choices for what to put on the second shelf. Then, after placing another type on the second shelf, there is only one choice for what to put on the last shelf. So, there are $3 \cdot 2 \cdot 1 = 6$ different arrangements (permutations).

Note that there are $n!$ permutations possible from n objects. In the previous example n was 3, giving $3! = 6$ permutations.

Ex: A class consists of 10 graduate students and 20 undergraduate students. These students will be ranked according to their performance. Assume there are no ties.

Q: How many different rankings are possible?
A: $30!$

Q: If Graduate students are ranked only among themselves and undergrads are only ranked among themselves, how many different rankings are possible?
A: $(10!)(30!)$ This should be 10! 20! because there are 10! ways to permute the grads and then 20! ways to permute the undergrads. The result follows from the basic principle of couting.

Ex: I have 4 math books, 3 chemistry books, 2 history books, and 1 fantasy novel. I want to have books of the same subject together on the shelf.

Q: How many arrangements are possible?
A: $(4!)(4!)(3!)(2!)(1!)$

Note that this $4!$ comes from the fact that there are $4!$ possible orderings of subjects.

Ex: How many different letter arrangements can be formed from the letters "PEPPER?"

A: There are 6! permutations of the letters, but many of them are the same because the two E's are indistinguishable from each other. The 3 P's are also indistinguishable from each other.

Let's label the E's: E_1, E_2
We label the P's: P_1, P_2, P_3

One possible permutation is $P_2 P_1 E_1 R P_3 E_2$.
All permutations which were previously indistinguishable from $P_2 P_1 E_1 R P_3 E_2$ are:

$P_2 P_1 E_1 R P_3 E_2$	$P_2 P_1 E_2 R P_3 E_1$
$P_1 P_2 E_1 R P_3 E_2$	$P_1 P_2 E_2 R P_3 E_1$
$P_1 P_3 E_1 R P_2 E_2$	$P_1 P_3 E_2 R P_2 E_1$
$P_2 P_3 E_1 R P_1 E_2$	$P_2 P_3 E_2 R P_1 E_1$
$P_3 P_1 E_1 R P_2 E_2$	$P_3 P_1 E_2 R P_2 E_1$
$P_3 P_2 E_1 R P_1 E_2$	$P_3 P_2 E_2 R P_1 E_1$

Notice that the difference between columns is that we permuted E_1 and E_2. Since There are 2 E's, we have 2! columns.

The difference between rows is that we permuted P_1, P_2, and P_3. Since there are 3 P's, we have 3! rows.

Thus for each "word" there are $3!\,2! = 12$ indistinguishable permutations. Thus only $\frac{1}{12}$ of the $6!$ permutations actually correspond to distinguishable "words." Therefore, there are $\frac{6!}{3!\,2!}$ different arrangements (words) which can be formed from the letters "PEPPER."

Ex: In a chess tournament, there are 4 Russian, 3 US, 2 Brittish, and 1 Brazilian players.

Q: If tournament results show only the nationalities of the players in the order they placed, how many different outcomes are possible?

A: There are $10!$ permutations of the players, but each outcome has $4!\,3!\,2!\,1!$ permutations which are indistinguishable. Thus there are $\frac{10!}{4!\,3!\,2!\,1!}$ different outcomes.

Ex: The Kuna Cross Country team consists of 13 runners. Only 7 of them can

Q: run the varsity race. How many different varsity teams are possible?

A: There are $13!$ different permutations of runners, but each "team" has $7!(13-7)!$ permutations which are indistinguishable. Thus there are only $\frac{13!}{7!(13-7)!}$ possible varsity teams.

Because choosing a group from a larger set of objects or individuals is so common, we have special notation for the number of possible groups.

$\binom{13}{7} = \frac{13!}{7!(13-7)!}$ is the number of groups of size 7 that can be selected from 13 individuals.

In general, $\binom{n}{r}$ is the number of groups of size r that can be selected from n individuals.

Note: $\binom{n}{r} = \frac{n!}{n!(n-r)!}$.

Ex: A committee is to be created from a group of 5 women and 7 men. The committee should have 2 women and 3 men.

Q: How many different committee could be formed?

A: $\binom{5}{2}\binom{7}{3}$ This follows from the basic principle of counting since there are $\binom{5}{2}$ possible choices for women and $\binom{7}{3}$ possible choices for men.

Q: How many committees can be formed if 2 of the men refuse to work together?

A: $\binom{5}{2} * \left[\binom{7}{3} - \binom{5}{1}\right]$ because the number of choices for men is reduced by $\binom{5}{1}$ since there are $\binom{5}{1}$ choices for men in which the two men who don't get along are together.

Ex: A system consists of a row of n antennas of which m are defective. The system is functional if no two defectives are consecutive. Assume all functional antennas are indistinguishable and all defective antennas are indistinguishable.

Q: How many different configurations are possible?

A: $\binom{n}{m}$

Q: How many functional configurations are possible?

A: Imagine that the n-m working antennas are lined up by themselves. Now, the task is to place the defectives among them in a way that produces a functional system.

$$\wedge\ 1\ \wedge\ 1\ \wedge\ 1\ \ldots\ 1\ \wedge$$

1's represent working antennas and $\wedge$'s represent places where we can place at most 1 defective antenna.

There are $n-m+1$ places ($\wedge$'s). We must select a group of size m from among them. There are $\binom{n-m+1}{m}$ possible choices.

Let's now apply our bag of tricks to prove the binomial theorem.

The Binomial Theorem:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \quad \text{for } n = 0, 1, 2, \ldots$$

proof:

$(x+y)^n = (x+y)(x+y) \ldots (x+y)$. There will be 2^n terms when this is multiplied out, but many will be the same. Each term is obtained by selecting one object (either x or y) from each set of parentheses. If we select x from exactly k sets of parentheses and y from the remaining $n-k$ sets, then we get the term $x^k y^{n-k}$. The next question is how many such terms do we get. Well, to select one of these terms is to select a group of size k from the n objects. Thus there are $\binom{n}{k}$ terms equal to $x^k y^{n-k}$.

Therefore $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$, since

k can be $0, 1, 2, \ldots, n$.

Ex: How many subsets of $\{5, 7, 23, 100\}$ are there?

A: There are $\binom{4}{0}$ subsets of size 0,
$\binom{4}{1}$ of size 1,
$\binom{4}{2}$ of size 2,
$\binom{4}{3}$ of size 3,
$\binom{4}{4}$ of size 4.

Thus there are $\sum_{i=0}^{4} \binom{4}{i}$ subsets in all.

Alternative solution: Each subset can be described by an ordered 4-tuple of 0's and 1's. For example, (1,0,0,1) represents the subset obtained by including the 1st and 4th elements but dropping the 2nd & 3rd elements.

Thus the selection of a subset is simply selecting whether to place a 0 or a 1 in each spot. By the generalized principle of counting, There are $2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$ different subsets.

Chapter 2 — Axioms of Probability

Def: The set of all possible outcomes of an experiment is called the "sample space."

Def: Any subset of the sample space is called an "event."

Ex: If the outcome of an experiment consists in the determination of the sex of a newborn child, then $S = \{M, F\}$

$\{M\}$ is the event that the baby is a boy.

Ex: If the outcome of an experiment is the order of finish in a horse race among the 7 horses having positions $1, 2, \ldots, 7$, then S is the set of all permutations of $(1, 2, 3, 4, 5, 6, 7)$. So S has $7!$ elements.

One possible event is the event that horse 1 wins the race. There are $6!$ different outcomes / elements of this event.

Ex: If an experiment consists of flipping a coin twice, then $S = \{HH, HT, TH, TT\}$, where HT means the first toss was "heads" and the second resulted in "tails."

The event that at least one "heads" gets tossed is $\{HH, HT, TH\}$.

The event that the first toss results in tails is $\{TH, TT\}$.

Because we are using sets, it will be useful to use the notation and terminology associated with sets. Let A, B be events. For example, you can think of A as the event that the first toss results in heads and B as the event that the second toss results in tails.

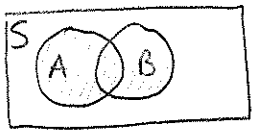
$A \cup B$ is the "union" of A and B. The event $A \cup B$ consists of everything that is in A <u>or</u> in B. Thus $A \cup B = \{HH, HT, TT\}$.

$A \cap B$ is the "intersection" of A and B and consists of those outcomes which are in A <u>and</u> B. Thus $A \cap B = \{HT\}$.

A^c is the "compliment" of A and consists of those outcomes/elements in the sample space which are <u>not</u> in A. Thus $A = \{TH, TT\}$.

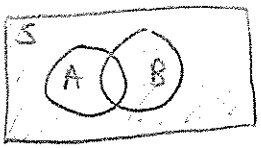
Venn diagrams are useful to illustrate these ideas:

$A \cup B$



$A \cap B$

A^c



We can now begin to define the probability of an event.

Intuitively, the probability of an event occuring (such as tossing heads) can be thought of as the proportion of times the event would occur if we were to repeat the experiment many many times.

For a fair coin, this would be $\frac{1}{2}$.

There are a number of things we should expect any reasonable definition of probability to satisfy. These will be called axioms.

Axiom 1: $0 \leq \mathbb{P}(E) \leq 1$

Axiom 2: $\mathbb{P}(S) = 1$

Axiom 3: If $E_1, E_2, \ldots$ are mutually exlusive, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$.

Note: We say $E_1, E_2 \ldots$ are mutually exlusive or disjoint if $E_i \cap E_j = \emptyset$ for $i \neq j$. $\emptyset = \{\ \}$ is called the empty set or null event.

There are several instances in which we can determine the probability of an event using only these axioms.

Ex: $\mathbb{P}(\emptyset)=0$

proof: Let $E_1 = S$ and $E_2, E_3, \ldots = \emptyset$

Then $1 = \mathbb{P}(S) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$

$$= \mathbb{P}(S) + \sum_{i=2}^{\infty} \mathbb{P}(E_i)$$

$$= 1 + \sum_{i=2}^{\infty} \mathbb{P}(E_i)$$

$$\Rightarrow \mathbb{P}(E_i) = 0 \quad \text{for} \quad i \geq 2. \quad \square$$

Ex: For a fair coin, heads and tails are equally likely outcomes: $\mathbb{P}(H) = \mathbb{P}(T)$.

Let $E_1 = \{H\}$, $E_2 = \{T\}$, $E_3 = E_4 = \ldots = \emptyset$

Then $1 = \mathbb{P}(S) = \mathbb{P}(\bigcup E_i) = \sum \mathbb{P}(E_i) = \mathbb{P}(H) + \mathbb{P}(T) + 0 + \ldots$

$= 2\mathbb{P}(H)$

$\Rightarrow \mathbb{P}(H) = \frac{1}{2}$ and $\mathbb{P}(T) = \frac{1}{2}$.

Ex: For a die, the outcomes $\{1,2,3,4,5,6\}$ are all equally likely. Find $\mathbb{P}(\{1,2,3\})$.

$E_1 = \{1\}$ $E_2 = \{2\}, \ldots, E_6 = \{6\}$, $E_7 = E_8 = \ldots \emptyset$.

$1 = \mathbb{P}(S) = \mathbb{P}(\bigcup E_i) = \sum \mathbb{P}(E_i) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \ldots + \mathbb{P}(E_6) + 0$

$= 6\mathbb{P}(E_1)$

$\Rightarrow \mathbb{P}(E_1) = \mathbb{P}(E_2) = \ldots = \mathbb{P}(E_6) = \frac{1}{6}$.

Now let $A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{3\}$, $A_4 = A_5 = \ldots = \emptyset$.

Then $\mathbb{P}(\{1,2,3\}) = \mathbb{P}(\bigcup A_i) = \sum \mathbb{P}(A_i) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + 0$

$= \frac{3}{6} = \frac{1}{2}$.

Proposition: $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$

proof:

$$1 = \mathbb{P}(S) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c)$$

$$\Rightarrow \mathbb{P}(E^c) = 1 - \mathbb{P}(E). \quad \square$$

Ex: I have a weighted coin such that $\mathbb{P}(\{H\}) = \frac{2}{3}$.
Thus $\mathbb{P}(\{H\}^c) = \mathbb{P}(\{T\}) = \frac{1}{3}$.

Proposition: If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.

proof: Let $A_1 = E$, $A_2 = F \cap E^c$.

Then
$$\begin{aligned}\mathbb{P}(F) &= \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) \\ &= \mathbb{P}(E) + \mathbb{P}(F \cap E^c) \\ &\geq \mathbb{P}(E). \quad \square\end{aligned}$$

Ex: when rolling a die, $\mathbb{P}(\{1\}) \leq \mathbb{P}(\{1, 3, 5\})$.

Proposition: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

proof: First, note that
$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}[(B \cap A) \cup (B \cap A^c)] \\ &= \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)\end{aligned}$$

$$\Rightarrow \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B \cap A^c)$$

Similarly, $\mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A \cap B^c)$.

Thus
$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B) \\ &= \mathbb{P}(A \cap B) + \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \quad \square\end{aligned}$$

This is easy to visualize with a Venn diagram.

Ex: Ellie will take 2 books on vacation. She will like the first with probability $\frac{1}{2}$ and the second with probability $\frac{2}{5}$. She will like both with probability $3/10$.

Q: What is the probability she likes at least one of them?

A: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
$$= \frac{1}{2} + \frac{2}{5} - \frac{3}{10} = \frac{3}{5}$$

Q: What is the probability she dislikes both?

A: $\frac{2}{5}$

Consider a sample space, S, in which all outcomes are equally likely. Label these outcomes $1, 2, \ldots, N(S)$.
Then
$$1 = \mathbb{P}(S) = \mathbb{P}(\{1\} \cup \{2\} \cup \ldots \cup \{N(S)\})$$
$$= \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) + \ldots + \mathbb{P}(\{N(S)\})$$
$$= N(S)\,\mathbb{P}(\{1\})$$
Thus $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \ldots = \mathbb{P}(\{N(S)\}) = \frac{1}{N(S)}$.

Now consider and event, E, and let $N(E)$ be the number of outcomes in E. Let E_1 be the event consisting of only the first outcome in E. Let E_2 be the second, etc.

Then $\mathbb{P}(E_i) = \left(\frac{1}{N(S)}\right)$ and

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i=1}^{N(E)} E_i\right) = \sum_{i=1}^{N(E)} \mathbb{P}(E_i) = N(E)\mathbb{P}(E_1) = \frac{N(E)}{N(S)}$$

$$= \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S}$$

Ex: A die is rolled twice.

Q: What is the probability that the sum of the two rolls is 7?

A: Using the basic principle of counting, there are $6 \cdot 6 = 36$ outcomes in S. The event of interest is $E = \{(1,6), (2\ 5), (3\ 4), (4\ 3), (5\ 2), (6, 1)\}$. Since E has 6 outcomes, $\mathbb{P}(E) = \frac{6}{36} = \frac{1}{6}$.

Ex: A bowl contains 6 white and 5 black balls. Cameron mixes them up and then reaches in and scoops out 3 balls.

Q: What is the probability that exactly 2 of the 3 are black?

A: There are $\binom{11}{3}$ different groups of balls that could be drawn. $\binom{5}{2}\binom{6}{1}$ of these groups have exactly 2 black balls. So, the probability is $\frac{\binom{5}{2}\binom{6}{1}}{\binom{11}{3}} = \frac{4}{11}$.

Alternative solution:
The probability of getting exactly two black balls is the same if Cameron draws them out one at a time. If we then think of the outcome as an <u>ordered</u> triple, Then there are $11 \cdot 10 \cdot 9$ outcomes in the sample space.

There are $5 \cdot 4 \cdot 6$ outcomes in which only the first two are black, $6 \cdot 5 \cdot 4$ in which only the second two are black, and $5 \cdot 6 \cdot 4$ in which only the first and third are black.

Thus there are $3 \cdot (6 \cdot 5 \cdot 4)$ outcomes with exactly two black balls. Thus the probability of this event is $\dfrac{3(6 \cdot 5 \cdot 4)}{11 \cdot 10 \cdot 9} = \dfrac{4}{11}$.

Ex: A committee of 5 is to be selected from a group of
Q: 6 men and 9 women. What is the probability it will consist of 3 men and 2 women?

A: $\dfrac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}}$

Ex: n white balls are in a bowl. One of
Q: these has my lucky number on it. If I randomly select k balls one at a time, what is the probability of selecting my lucky number?

A: $$\frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$

Ex: What is the probability of being dealt a straight in poker? (5-card hand) (consecutive ranks but <u>not</u> all the same suit.)

A: Cards in each suit: A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K.
There are 4 suits. One possible straight is 4, 5, 6, 7 in hearts and 8 in clubs.
It is important to know that the Ace can be counted as lowest or highest in the ranking system, so A, 2, 3, 4, 5 can be a straight as well as 10, J, Q, K, A as long as they aren't all in the same suit.

There are $\binom{52}{5}$ possible hands. 4^5 of these contain A, 2, 3, 4, 5, but 4 of them are all of a single suit. Thus there are $4^5 - 4$ "5-high" straights possible

Similarly, there are 4^5-4 "6-high" straights possible, etc. The high card can be anywhere from 5 to Ace. So there are $10(4^5-4)$ hands possible containing a straight.

Thus the probability of getting a straight is

$$\frac{10(4^5-4)}{\binom{52}{5}} \approx .0039.$$

Ex: In the game of bridge, all 52 cards are dealt out to 4 players.

Q: What is the probability each player gets an ace?

A: $$\frac{\left[\binom{4}{1}\binom{48}{12}\right]\left[\binom{3}{1}\binom{36}{12}\right]\left[\binom{2}{1}\binom{24}{12}\right]\left[\binom{1}{1}\binom{12}{12}\right]}{\binom{52}{13}\quad\binom{39}{13}\quad\binom{26}{13}\quad\binom{13}{13}}$$

Ex: n people are in a room

Q: What is the probability that no two have the same birthday?

A: $\dfrac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365-n+1)}{(365)^n} = \dfrac{365!}{(365-n)!(365)^n}$

If $n \geq 23$, this probability is less than $\frac{1}{2}$.

Ex: A deck of cards is shuffled. Then cards are turned over one by one.

Q: What is the probability that the Ace of hearts is directly after the first Ace?

A: First, there are $51!$ permutations or arrangements of all other cards. In order to have the ace of hearts directly after the first Ace, the is only one place to put it. Thus there are $(51!)(1)$ permutations of the deck such that the ace of hearts immediately follows the first ace.

Thus the probability is $\dfrac{51!}{52!}$

Chapter 3 — Conditional Probability

The idea of conditional probability is that having some partial information may affect the likelihood of a particular event occuring

Ex: A person from SLC is randomly selected. What is the probability the person is taller than 6 ft (call this event A).

Suppose I know that the person is an NBA player (call this event B). Now what is the probability that he is taller than 6ft?

We see that $\mathbb{P}(A|B) > \mathbb{P}(A)$, where $\mathbb{P}(A|B)$ is the probability that A occurs given that B occurs.

The two probabilities are related according to $\mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(A \cap B)$.

If $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Ex: A coin is tossed twice. Let A be the event that both result in heads and let B be the event that at least one results in heads.

Q: Find $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B \mid A)$.

A: $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \dfrac{\frac{1}{4}}{3/4} = \frac{1}{3}.$

$\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 1.$

Note that $\mathbb{P}(A \mid B) > \mathbb{P}(A)$.

Ex: Let L_x denote the event that a student finishes an exam in less than x hours.

$$\mathbb{P}(L_x) = \begin{cases} \frac{x}{2} & 0 < x < 2 \\ 0 & x \leq 0 \\ 1 & x \geq 2 \end{cases}$$

Madelyn is still working after .75 hours, what is the conditional probability that she will take more than an hour?

A: $\mathbb{P}(L_1^c \mid L_{.75}^c) = \dfrac{\mathbb{P}(L_1^c \cap L_{.75}^c)}{\mathbb{P}(L_{.75}^c)} = \dfrac{\mathbb{P}(L_1^c)}{\mathbb{P}(L_{.75}^c)} = \dfrac{1 - \frac{1}{2}}{1 - \frac{3}{8}} = \frac{4}{5}$

Ex: In the card game "Bridge," 52 cards are completely dealt out to 4 players Ron, Brett, Josh, and Jeremy.

Q: If Ron & Brett have 8 spades between them, What is the probability that Josh has exactly 3 of the remaining 5 spades?

A: Let A be the event that Ron & Brett have 8 spades and let B denote the event that Josh has 3.

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\text{Josh gets 3 \& Jeremy 2})}{\mathbb{P}(A)}$$

$$= \frac{\left[\dfrac{\binom{13}{2}\binom{39}{11} \cdot \binom{11}{3}\binom{28}{10}}{\binom{52}{13} \cdot \binom{39}{13}}\right]}{\left[\dfrac{\binom{13}{8}\binom{39}{18}}{\binom{52}{26}}\right]}$$

There is an easier way. Imagine Ron & Brett have already received their cards. The probability that Josh gets 3 spades is now $\dfrac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}}$ since you know Ron & Brett have 8 of the spades.

Ex: An urn contains r red and b blue balls. n balls are sequential selected of which k are blue.

Q: What is the conditional probability that the 1st ball was blue given that k of the first n were blue?

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\binom{b}{1}\binom{b-1}{k-1}\binom{r}{n-k} \Big/ \left[\binom{r+b}{1}\binom{r+b-1}{n-1}\right]}{\binom{b}{k}\binom{r}{n-k} \Big/ \binom{r+b}{n}}$$

$$= \frac{b\left[\dfrac{(b-1)!}{(k-1)!\,(b-k)!}\right]\dfrac{1}{n}}{\left[\dfrac{b!}{(b-k)!\,k!}\right]}$$

$$= \frac{k}{n}$$

Alternative solution: Another way to solve this is to note that $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$

I prefer this second approach.

Ex: Ron isn't sure if he should take art history or music history. He estimates the probability of getting an A in art is $\frac{1}{2}$, and the probability of getting an A is music is $\frac{2}{3}$. He decides to base his decision on the toss of a fair coin.

Q: What is the probability he gets an A in <u>art</u>?

A: Let A be the event that he gets an A, and let B denote the event he takes art.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Ex: There are 8 red and 4 white balls in an urn. Two balls are selected—one at a time.

Q: What is the probability both are red?

A: Let R_1, R_2 be the events that the first, second ball is red respectively.

Then, $\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_2 \mid R_1)\,\mathbb{P}(R_1) = \left(\frac{7}{11} \cdot \frac{8}{12}\right).$

Note: This could have been computed as $\dfrac{\binom{8}{2}\binom{4}{0}}{\binom{12}{2}}$

Ex: A shuffled deck is separated into 4 piles
Q: of 13. What is the probability that each pile has an ace?

Note: We answered this before using a different trick.

A: Note that $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1)\,\mathbb{P}(E_2 \mid E_1)\,\mathbb{P}(E_3 \mid E_1 \cap E_2)$

proof: $RHS = \cancel{\mathbb{P}(E_1)\,\frac{\mathbb{P}(E_2 \cap E_1)}{\mathbb{P}(E_1)}\,\frac{\mathbb{P}(E_3 \cap E_1 \cap E_2)}{\mathbb{P}(E_1 \cap E_2)}}$

$= RHS.$

Let A_i denote the event that the ith pile has exactly 1 ace.

$$\mathbb{P}\left(\bigcap_{i=1}^{4} A_i\right) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_2 \cap A_1)\,\mathbb{P}(A_4 \mid A_1 \cap A_2 \cap A_3)$$

$$= \left(\frac{\left[\binom{4}{1}\binom{48}{12}\right]}{\binom{52}{13}}\right)\left(\frac{\binom{3}{1}\binom{36}{12}}{\binom{39}{13}}\right)\left(\frac{\binom{2}{1}\binom{24}{12}}{\binom{26}{13}}\right)\left(\frac{\binom{1}{1}\binom{12}{12}}{\binom{13}{13}}\right)$$

$\approx .105$

Bayes' Formula:

$$\mathbb{P}(A) = \mathbb{P}[(A \cap B) \cup (A \cap B^c)] = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$$
$$= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

There are many problems in which computing a probability seems quite difficult to do directly but is relatively easy to compute by conditioning on another event, B.

Ex: An accident prone person will crash in a given year with probability .4 while a safe drive will crash with probability .2. 30% of the population consists of accident prone drivers.

Q: What is the probability that a randomly selected driver will have an accident within the next year?

A: Let $B = \{\text{person is accident prone}\}$
$C = \{\text{person will crash}\}$

$$\mathbb{P}(C) = \mathbb{P}(C \mid B)\mathbb{P}(B) + \mathbb{P}(C \mid B^c)\mathbb{P}(B^c)$$
$$= (.4)(.3) + (.2)(.7)$$
$$= .26$$

Q: Compute $\mathbb{P}(B|C)$.

A: $\mathbb{P}(B|C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(C|B)\mathbb{P}(B)}{\mathbb{P}(C)} = \frac{(.4)(.3)}{(.26)} = \frac{6}{13}$

Ex: Consider the following game:
A deck of cards is shuffled. The one card is turned over at a time. At any time the play can guess that the next card will be the Ace of spades. If he is correct, he wins. If only one card remains unflipped and you haven't seen the "special card," you win.

Q: What is a good strategy?

A: All strategies are equal. $\mathbb{P}(\text{Winning}) = \frac{1}{52}$

We will prove that it is true for any size deck. For a deck of size n, $\mathbb{P}(\text{Win}) = \frac{1}{n}$.

proof: If $n=1$, you win — no strategy needed.
So, $\mathbb{P}(\text{Win}) = \frac{1}{n}$ for $n=1$.
Now suppose it works for a deck of size k. (This is certainly true for $k=1$)

Let W be the event that you win and G the event that your strategy was to guess the <u>first</u> card.

Let's now play with a $(k+1)$-card deck.

let $p = \mathbb{P}(G)$.

Let $A = \{\text{first card is the special card}\}$

$$\begin{aligned}
\mathbb{P}(W) &= \mathbb{P}(W \mid A)\,\mathbb{P}(A) + \mathbb{P}(W \cap A^c) \\
&= \mathbb{P}(W \mid A)\,\mathbb{P}(A) + \mathbb{P}(W \cap A^c \cap G^c) \\
&= \mathbb{P}(W \mid A)\,\mathbb{P}(A) + \mathbb{P}(G^c)\,\mathbb{P}(A^c \mid G^c)\,\mathbb{P}(W \mid A^c \cap G^c) \\
&= p\left(\frac{1}{k+1}\right) + (1-p)\left(\frac{k}{k+1}\right)\frac{1}{k} \\
&= = \frac{1}{k+1}.
\end{aligned}$$

Note that $\mathbb{P}(W \mid A^c \cap G^c)$ is the same as the probability of winning with a k-card deck.

We have shown that the probability of winning with any strategy is $\frac{1}{1}$ with a deck of size one and also that if the probability is $\frac{1}{k}$ for a deck of size k, then it is $\frac{1}{k+1}$ for a deck of size $k+1$.

Using $k=1$, we see that $\mathbb{P}(W) = \frac{1}{2}$ for a 2-card deck. Then, using $k=2$, we see that $\mathbb{P}(W) = \frac{1}{3}$ for a 3-card deck, etc.

$\square$

On a multiple choice quiz with responses a, b, c:

Ex: Let p be the probability that a student knows the answer. If the don't know, they guess. Let $C = \{\text{they get it correct}\}$, $K = \{\text{they knew the answer}\}$.

Q: $$\mathbb{P}(K|C) = \frac{\mathbb{P}(C|K)\mathbb{P}(K)}{\mathbb{P}(C)} = \frac{\mathbb{P}(K)}{\mathbb{P}(C)} = \frac{p}{\mathbb{P}(C|K)\mathbb{P}(K) + \mathbb{P}(C|K^c)\mathbb{P}(K^c)}$$

$$= \frac{p}{p + \frac{1}{3}(1-p)}.$$

Ex: A blood test not perfect. A false positive (negative) is when the test <u>incorrectly</u> concludes that the disease is present (absent).

If the disease is present, the test has a 95 % chance of detecting it. (Good, right?)

However, a healthy person will get a false positive 1 % of the time.

Half of the population has the disease.

Q: What are the chances I have the disease given that I tested positive?

A: Let $P = \{\text{positive result}\}$
$D = \{\text{disease present}\}$

Reviewing what we know: $\mathbb{P}(P \mid D) = .95$
$\mathbb{P}(P \mid D^c) = .01$
$\mathbb{P}(D) = .50$

$$\mathbb{P}(D \mid P) = \frac{\mathbb{P}(D \cap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P \mid D)\,\mathbb{P}(D)}{\mathbb{P}(P)}$$

$$= \frac{\mathbb{P}(P \mid D)\,\mathbb{P}(D)}{\mathbb{P}[(P \cap D) \cup (P \cap D^c)]} = \frac{(.95)(.50)}{\mathbb{P}(P \cap D) + \mathbb{P}(P \cap D^c)}$$

$$= \frac{(.95)(.50)}{\mathbb{P}(P \mid D)\,\mathbb{P}(D) + \mathbb{P}(P \mid D^c)\,\mathbb{P}(D^c)}$$

$$= \frac{(.95)(.50)}{(.95)(.50) + (.01)(.50)}$$

$$= .98958$$

Q: Repeat the calculation when $\mathbb{P}(D) = .005$

A: $$\mathbb{P}(D \mid P) = \frac{\mathbb{P}(P \mid D)\,\mathbb{P}(D)}{\mathbb{P}(P \mid D)\,\mathbb{P}(D) + \mathbb{P}(P \mid D^c)\,\mathbb{P}(D^c)} = \frac{(.95)(.005)}{(.95)(.005) + (.01)(.995)}$$

$= .323$ (Surprising, right?)

Def: We say that two events are independent if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. In other words, knowing that B has occured doesn't help you determine the likelihood of A.

Note: If A, B are independent, then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

proof: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)$ $\square$

Note: If $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, then A, B are independent.

Proof: $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$ $\square$

Therefore, we can (and will) use $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ to be the definition of two independent events.

Ex: A card is drawn from a shuffled deck.

Let $E = \{\text{Ace}\}$

$F = \{\text{Spade}\}$

Q: Are E, F independent?

A: Yes. $\mathbb{P}(E \cap F) = \frac{1}{52} = \left(\frac{4}{52}\right)\left(\frac{1}{4}\right) = \mathbb{P}(E)\,\mathbb{P}(F)$.

Ex: Toss a fair coin twice. $E = \{\text{first toss results in heads}\}$
$F = \{\text{second toss results in heads}\}$.

Q: Are E, F independent?

A: $\mathbb{P}(A \cap B) = \mathbb{P}\{HH\} = \frac{1}{4} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \mathbb{P}\{HH, HT\}\,\mathbb{P}\{TH, HH\} = \mathbb{P}(E)\mathbb{P}(F)$

Because this is one of 4 equally likely outcomes.

Thus A, B are independent.

Ex: Roll a die twice. $E = \{\text{sum is } 6\}$, $F = \{\text{first roll is } 4\}$

Q: Are E, F independent?

A: $\mathbb{P}(E) = \frac{5}{36}$, $\mathbb{P}(F) = \frac{1}{6}$, $\mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F)$
$= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)$

Therefore, $\mathbb{P}(E \cap F) \neq \mathbb{P}(E)\mathbb{P}(F)$ and E, F are "<u>dependent</u>." In other words, they are <u>not</u> independent.

Def: Three or more events, A_i, are independent if $\mathbb{P}\left(\bigcap_k A_{i_k}\right) = \prod_k A_{i_k}$ for any subset of the events.

Note: Events can be "pairwise independent" but not independent.

Ex: A sequence of n independent trials are performed. A trial is an expirement with only two possible outcomes — labeled as sucess or failure. Each trial results in sucess with probability p.

Q: What is the probability that at least 1 success occurs?

A: $E = \{1 \text{ or more success}\}$

$E^c = \{\text{All failures}\}$

$$\mathbb{P}(E^c) = \mathbb{P}(S_1^c \cap S_2^c \cap S_3^c \cap \ldots \cap S_n^c) = \prod_{i=1}^{n} \mathbb{P}(S_i^c) = [\mathbb{P}(S_1^c)]^n = [1-p]^n$$

Thus $\mathbb{P}(E) = 1 - [1-p]^n$.

Q: What is the probability of getting exactly k successes? $(0 \le k \le n)$

A: $B = \{\text{exactly } k \text{ successes}\}$

$C = \{\text{First } k \text{ are successes, others are failures}\}$

Note that $C \subseteq B$. In particular, B contains $\binom{n}{k}$ outcomes which each have the same probability as C.

$$\begin{aligned}\mathbb{P}(C) &= \mathbb{P}(S_1 \cap S_2 \cap \ldots \cap S_k \cap S_{k+1}^c \cap \ldots \cap S_n^c) \\ &= [\mathbb{P}(S_1)]^k [\mathbb{P}(S_1^c)]^{n-k} \\ &= p^k (1-p)^{n-k}\end{aligned}$$

Other sequences consisting of exactly k successes have the same probability — why?

How many configurations of exactly k successes are there? This is equivalent to selecting k of the n trials to be successes. There are $\binom{n}{k}$ ways of choosing.

Therefore $\mathbb{P}(B) = \binom{n}{k}\mathbb{P}(C) = \binom{n}{k}p^k(1-p)^{n-k}$

Ex: A system of n components is in parallel so that it will work as long as one of the components works. Each component works with probability p, independent of all other components.

Q: What is the probability that the system works?

A: $1-(1-p)^n$

Ex: two dice are rolled at a time. What is the probability that they sum to 5 before they sum to 7? (The game ends in either case)

A: Let $E_n = \left\{ \begin{array}{l} \text{no 5 or 7 rolled in first } n-1 \text{ rolls} \\ \text{and the } n^{\text{th}} \text{ roll is a 5} \end{array} \right\}$.

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n) = \sum_{n=1}^{\infty}\left(\frac{26}{36}\right)^{n-1}\left(\frac{4}{36}\right) = \frac{2}{5}.$$

Ex: Two Gamblers, A and B, bet on successive tosses of a coin. If Heads, A collects a dollar from B. Otherwise B collects from A. The game continues until someone runs out of money.

Let γ_i be the probability that A wins if A started with i and B started with $N-i$ dollars. (We assume that N is their combined wealth.)

Let H, T be the event that the <u>first</u> toss resulted in heads, tails respectively.

$$\gamma_i = \mathbb{P}(A \text{ wins starting from } i \mid H)\,\mathbb{P}(H) + \mathbb{P}(A \text{ wins starting from } i \mid T)\,\mathbb{P}(T)$$

$$= \gamma_{i+1}\, p + \gamma_{i-1}\,(1-p) \qquad (\text{Let } q = 1-p)$$

$$\Rightarrow p\gamma_i + q\gamma_i = p\gamma_{i+1} + q\gamma_{i-1}$$

$$\Rightarrow \frac{q}{p}(\gamma_i - \gamma_{i-1}) = (\gamma_{i+1} - \gamma_i) \qquad (*)$$

Note: $\gamma_0 = 0$, $\gamma_N = 1$.

From (*) we obtain: $(\gamma_i - \gamma_{i-1}) = \left(\frac{q}{p}\right)^{i-1}(\gamma_1 - \gamma_0)$

$$= \left(\frac{q}{p}\right)^{i-1}\gamma_1 .$$

$$\gamma_n = \sum_{i=1}^{n}(\gamma_i - \gamma_{i-1}) = \gamma_1 \sum_{i=1}^{n}\left(\frac{q}{p}\right)^{i-1} = \left[\frac{\gamma_1\left(1-\left(\frac{q}{p}\right)^n\right)}{1-\frac{q}{p}}\right]$$

Because $\gamma_0 = 0$.

This is where it Breaks down when $p = \frac{1}{2}$.

Note: $1 = \gamma_N = \dfrac{\gamma_1\left(1-\left(\frac{q}{p}\right)^N\right)}{\left(1-\frac{q}{p}\right)}$

$$\Rightarrow \gamma_1 = \left[\frac{\left(1-\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N}\right]$$

Thus $\gamma_n = \left[\dfrac{1-\left(\frac{q}{p}\right)^n}{1-\left(\frac{q}{p}\right)^N}\right]$ for $0 < n < N$ and $p \neq \frac{1}{2}$

Q: What do we do when $p = q$?

A: $\gamma_n = n\gamma_1$. Since $\gamma_N = 1$, $\gamma_n = \frac{n}{N}$ for $0 < n < N$ and $p = \frac{1}{2}$.

Chapter 4 — Random Variables

Def: A "random variable" is a real-valued function whose domain is the sample space.

Ex: Toss a coin 3 times.

$S = \{(HHH)\ (HHT)\ (HTH)\ (HTT)\ (THH)\ (THT)\ (TTH)\ (TTT)\}$

Let X = # of heads tossed. X is a r.v..

$X(HHH) = 3$, $X(HTT) = 1$, etc.

$$\mathbb{P}(X=1) = \mathbb{P}\left(\{(HTT),(THT),(TTH)\}\right) = \frac{3}{8},$$

$$\mathbb{P}(X=3) = \mathbb{P}(\{HHH\}) = \frac{1}{8}, \quad \text{etc.}$$

Ex: 20 balls (numbered 1 – 20) are in an urn.

Q: If you draw 3 balls, what is the probability that the highest number is 17 or larger?

A: Let X = highest number

$$\mathbb{P}(X=20) = \frac{\binom{1}{1}\binom{19}{2}}{\binom{20}{3}}$$

$$\mathbb{P}(X=19) = \frac{\binom{1}{1}\binom{18}{2}}{\binom{20}{3}}$$

$$\mathbb{P}(X=x) = \frac{\binom{1}{1}\binom{x-1}{2}}{\binom{20}{3}} \qquad \text{for } 3 \le x \le 20$$

Thus the probability we are looking for is:

$$\sum_{x=17}^{20} \binom{x-1}{2} \Big/ \binom{20}{3} \approx .508$$

Ex: A coin is tossed until it lands on Heads
Let X = # of tosses.

$\mathbb{P}(X=1) = p$
$\mathbb{P}(X=2) = (1-p)p$
$\mathbb{P}(X=3) = (1-p)^2 p$
$\mathbb{P}(X=x) = (1-p)^{x-1} p$

Def: A r.v. is called "discrete" if the set of possible outcomes is finite or countably infinite.

Def: For a discrete r.v., the "probability mass function," $p(x)$, is given by

$$p(x) = \mathbb{P}(X=x)$$

Ex: In the previous example, the mass function is

$$p(x) = (1-p)^{x-1} p \qquad \text{for } x \in \{1, 2, 3, \ldots\}$$

Fact: $\sum_{\{x \mid p(x)>0\}} p(x) = 1$, where the sum is over all possible outcomes of the r.v..

Q: Why?

Ex: The probability mass function (pmf) of a Poisson(λ) r.v. is $p(x) = \begin{cases} \frac{c\,\lambda^x}{x!} & \text{for } x \in \{0, 1, 2, \dots\} \\ 0 & \text{o/w.} \end{cases}$

Q: What is c?

A: $1 = c\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = ce^{\lambda}$. Thus $c = e^{-\lambda}$.

Q: Find $\mathbb{P}(X=0)$

A: $\mathbb{P}(X=0) = e^{-\lambda}$

Q: Find $\mathbb{P}(X>2)$

A: $\mathbb{P}(X>2) = 1 - \mathbb{P}(X \le 2)$
$= 1 - \sum_{x=0}^{2} p(x) = 1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right)$

Def: The "cummulative distribution function," (cdf) F, of a r.v. is given by

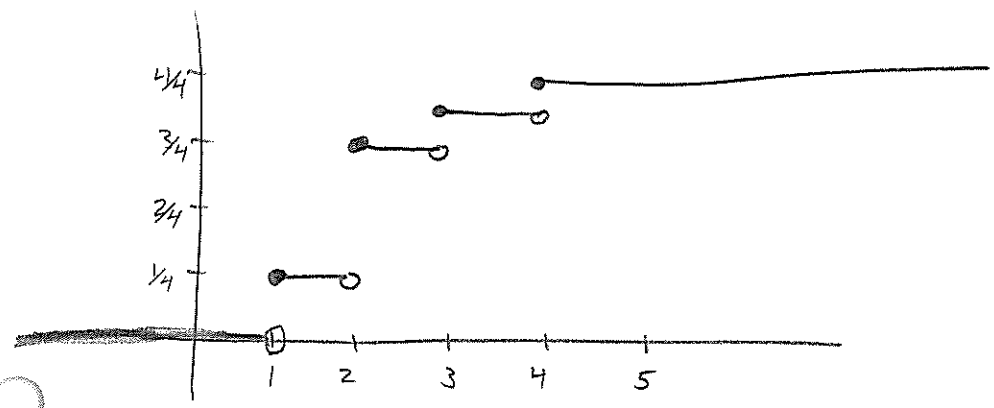
$$F(x) = \mathbb{P}(X \le x).$$

Ex: Consider a r.v. with mass function

$$p(x) = \begin{cases} \frac{1}{4} & x = 1, \\ \frac{1}{2} & x = 2, \\ \frac{1}{8} & x \in \{3, 4\}, \\ 0 & \text{o/w.} \end{cases}$$

Q: Find the cdf.

A: $F(x) = \begin{cases} 0 & x < 1, \\ \frac{1}{4} & 1 \le x < 2, \\ \frac{3}{4} & 2 \le x < 3, \\ \frac{7}{8} & 3 \le x < 4, \\ 1 & 4 \le x. \end{cases}$



Def: The "expected value," $E(X)$, of a <u>discrete</u> r.v. is given by

$$E(X) = \sum_{\{x \mid p(x) > 0\}} x\, p(x)$$

One interpretation of $E(X)$: If X is the amount of money you win in a game and you play this game many many times, your average winnings will be close to $E(X)$, your "expected winnings."

Another way to think about $E(X)$ is a "best" guess or prediction of the outcome of X.

Ex: $p(2) = p(1) = \frac{1}{2}$

Then $E(X) = 1(\frac{1}{2}) + 2(\frac{1}{2}) = \frac{3}{2}$

Ex: Rolling a die. $X = \#$ on die

$$E(X) = 1(\tfrac{1}{6}) + 2(\tfrac{1}{6}) + 3(\tfrac{1}{6}) + 4(\tfrac{1}{6}) + 5(\tfrac{1}{6}) + 6(\tfrac{1}{6})$$
$$= \frac{21}{6} = \frac{7}{2} = 3.5.$$

Ex: An "indicator" is a random variable whose outcome is 1 if a given event occurs and 0 otherwise.

The notation we use is $\mathbb{1}\{A\}$.

Thus $\mathbb{1}\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{o/w.} \end{cases}$

Q: Find $E(\mathbb{1}\{A\})$.

A: $E(\mathbb{1}\{A\}) = 1\,\mathbb{P}(A) + 0\,\mathbb{P}(A^c)$
$= \mathbb{P}(A)$.

Ex: If $\mathbb{P}(X = -1) = \frac{1}{5}$
$\mathbb{P}(X = 0)\ \frac{1}{2}$
$\mathbb{P}(X = 1) = 3/10$

Q: Compute $E(X^2)$ by first computing the mass function of $Y = X^2$.

Important note: $E(X^2) = \frac{1}{2} \neq (EX)^2 = .01$

Proposition: If X is a discrete r.v., then
$$E[g(X)] = \sum_{\{x \mid p(x) > 0\}} g(x)\, p(x)$$

Ex: Recompute $E(X^2)$ using the above proposition.

Ex: A certain store sells Christmas lights. For each string of lights they sell, they profit \$3. For each unsold string of lights they lose \$2. Let X be the number of customers who will want to buy lights with p.m.f. $p(x) = \begin{cases} \frac{1}{10} & \text{for } x = 11, 12, \ldots, 20, \\ 0 & \text{o/w.} \end{cases}$

Let Y_ℓ be the profit earned by the company if they order ℓ strings of lights at the beginning of the season. Note: $Y_\ell = 3\min\{X, \ell\} - 2(\max\{X, \ell\} - X)$

Equivalently, $Y_\ell = \begin{cases} 3\ell & X \geq \ell, \\ 3X - 2(\ell - X) & X < \ell. \end{cases}$

$$= \begin{cases} 3\ell & X \geq \ell \\ 5X - 2\ell & X < \ell \end{cases}$$

Q: Find $E(Y_{15})$.

A: $$E(Y_{15}) = \sum_{\{x \mid p(x)>0\}} g(x)\,p(x) = \sum_{\{x \geq 15 \ (x)>0\}} 3(15)p(x) + \sum_{\{x<15 \mid p(x)>0\}} (5x-2(15))\,p(x)$$

$$= \sum_{x=15}^{20} 45/10 + \sum_{x=11}^{14} (5x-30)\tfrac{1}{10}$$

$$= 6\left(\tfrac{45}{10}\right) - 3(4) + \tfrac{1}{2}(11+12+13+14)$$

$$= 40$$

We could find a general expression for $E(Y_\ell)$. The next step would be to choose ℓ to maximize the expected profits.

Def: $\operatorname{var}(X) = E[X - E(X)]^2$

While the mean or expected value of a r.v. tells us about the center of the distribution, the variance is a measure of the spread.

If, for example, X has a high probability of being far from the center, the variance (and the spread) will be large.

Ex: Consider r.v.'s X, Y with mass functions given by

$$p_X(x) = \begin{cases} \frac{1}{2} & x = 1, -1, \\ 0 & o/w. \end{cases}$$

$$p_Y(y) = \begin{cases} \frac{1}{2} & x = 100, -100, \\ 0 & o/w. \end{cases}$$

Q: Compute $E(X)$, $E(Y)$, $var(X)$, $var(Y)$.

A: $E(X) = E(Y) = 0$

$var(X) = 1$

$var(Y) = 10{,}000$

Proposition: $var(X) = E(X^2) - (EX)^2$

proof: $var(X) = E(X - EX)^2 = E(X - \mu)^2$

$$= E(X^2 - 2\mu X + \mu^2)$$

$$= \sum_{\{x \mid p(x) > 0\}} (x^2 - 2\mu x + \mu^2) p(x)$$

$$= \sum_{\{x \mid p(x) > 0\}} x^2 p(x) - 2\mu \sum_{\{x \mid p(x) > 0\}} (x) p(x) + \mu^2 \sum_{\{x \mid p(x) > 0\}} p(x)$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - (EX)^2.$$

Ex: Let X = the number resulting from a die roll.

Q: Find $var(X)$.

A: $\frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$

Proposition: $E(aX+b) = a\,E(X) + b$

proof: $E(aX+b) = \sum (ax+b)\,p(x) = a\sum x\,p(x) + b\sum p(x)$

$= a\,E(X) + b$

Proposition: $Var(aX+b) = a^2\,var(X)$.

proof: $var(aX+b) = E\left[(aX+b) - E(aX+b)\right]^2$

$= E\left[aX+b - (aE(X)+b)\right]^2$

$= E\left[a(X-E(X))\right]^2$

$= a^2\,E\left[X-E(X)\right]^2$

$= a^2\,var(X)$

Def: $SD(X) = \sqrt{var(X)}$ (Another measure of spread)

Ex: A p-coin lands on heads with probability p. Let's call an outcome of Heads a success and an outcome of Tails a failure.

Define a r.v. $X\{H\} = 1$ and $X\{T\} = 0$

Def: X is called a "Bernoulli r.v." and it has mass function $p(x) = \begin{cases} p & x = 1, \\ 1-p & x = 0. \end{cases}$

Ex: Now suppose I were to toss the p-coin n times. Let $Y = \#$ of successes. If $n=1$, then Y is a BERNoulli r.v. (BER(p)).

Def: If n is any positive integer, then Y is called a "binomial r.v." (BIN(n,p)). You'll see why when you see the mass function.

Suppose that $n = 5$.

$$\mathbb{P}(Y=0) = (1-p)^5$$

$$\mathbb{P}(Y=1) = p(1-p)^4 + (1-p)p(1-p)^3 + (1-p)^2 p(1-p)^2 + (1-p)^3 p(1-p) + (1-p)^4 p$$

$$= 5p(1-p)^4$$

$$= \binom{5}{1} p(1-p)^4$$

$$\mathbb{P}(Y=2) = \binom{5}{2} p^2 (1-p)^3$$

$$\mathbb{P}(Y=y) = \begin{cases} \binom{5}{y} p^y (1-p)^{5-y} & y \in \{0,1,\dots,5\}, \\ 0 & \text{o/w.} \end{cases}$$

For general n, $\mathbb{P}(Y=y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & y = 0,1,\dots, \\ 0 & \text{o/w.} \end{cases}$

Note that a BINOMIAL r.v. is a sum of independent BERNOULLIs. In particular,

$$BIN(n,p) = \sum_{i=1}^{n} X_i \qquad \text{where } X_i \sim \text{iid } BER(p)$$

Note: i.i.d. means independent and identically distributed.

Ex: ACME sells screws in packages of 10. Each screw is defective with probability .01 independent of the other screws.

ACME will refund your money if 2 or more are defective.

What proportion of sales must be refunded (ie what is the probability that a randomly selected package has 2+ defectives)

Let X = # of defectives (successes)

The mass function for X is

$$p(x) = \begin{cases} \binom{10}{x}(.01)^x(1-.01)^{10-x} & x \in \{0,1,\dots,10\}, \\ 0 & o/w. \end{cases}$$

Thus the probability of 2+ defectives is

$$1 - \sum_{x=0}^{1} \binom{10}{x} (.01)^x (.99)^{10-x} \approx .0042662$$

Ex: Find $E(X)$ where $X \sim BIN(n,p)$.

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x}$$

$$= \sum_{x-1=0}^{n-1} n \binom{n-1}{x-1} p^{x-1} p (1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y}$$

$= np$ since $\binom{n-1}{y} p^y (1-p)^{n-1-y}$ is the mass function of $BIN(n-1, p)$.

Ex: Find $E(X^2)$ where $X \sim BIN(n, p)$

$$E(X^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x-1=0}^{n-1} x \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x}$$

$$np \sum_{x-1=0}^{n-1} \left[(x-1)+1\right] \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{x-1=0}^{n-1} \left[(x-1)+1\right] \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{y=0}^{n-1} (y+1) \underbrace{\binom{n-1}{y} p^y (1-p)^{(n-1)-y}}$$

$$= np\left((n-1)p\right) + np$$

because this is the mass function for $BIN(n-1, p)$.

$$= np\left((n-1)p + 1\right)$$

Ex: Find $Var(X) = np(np - p + 1) - (np)^2$

$$= np(1-p)$$

Def: A r.v., X, is "poisson" if it has mass function $p(x) = P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$ for some $\lambda > 0$.

Ex: Find the mean and variance of $X \sim POI(\lambda)$.

$$E(X) = \sum_{x=0}^{\infty}\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)x = e^{-\lambda}\lambda \sum_{x-1=0}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = \lambda$$

$$E(X^2) = \sum_{x=0}^{\infty}\frac{x^2 e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\lambda \sum_{x-1=0}^{\infty}\frac{x\,\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda}\lambda \sum_{x-1=0}^{\infty}\frac{[(x-1)+1]\,\lambda^{x-1}}{(x-1)!}$$

$$= \lambda(\lambda+1) = \lambda^2+\lambda$$

$$\mathrm{var}(X) = E(X^2)-(EX)^2 = \lambda^2+\lambda-(\lambda)^2 = \lambda.$$

Ex: (Negative Binomial) Suppose a couple wishes to have r daughters Each child born will be a girl with probability p — independent of other births. They will stop having children once they have r girls. Let X = total number of children they will have.

Then $\mathbb{P}(X=x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}$ for $x = r, r+1, \ldots$

Ex: (Hypergeometric) A bowl contains m black and $N-m$ white balls. $n < N$ balls are scooped out. Let X = # of black balls scooped out.

Then $$\mathbb{P}(X=x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$ for $x \geq 0$, $m \geq x$, $N-m \geq n-x$

Chapter 5 — Continuous Random Variables

In the previous chapter, we considered only r.v.'s whose set of possible outcomes was finite or countably infinite (ie discrete r.v.'s).

Def: In this chapter we consider "continuous" r.v.'s. A r.v., X, is called "continuous" if there exists a non-negative function, f, such that

$$\mathbb{P}(X \in A) = \int_A f(x)\,dx$$

for any set $A \subseteq \mathbb{R}$.

Def: The function, f, above is called the "probability density function" of X.

Note: $1 = \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x)\,dx$

Note: $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$ (for $a \leq b$)

Note: $\mathbb{P}(X = a) = \int_a^a f(x)\,dx = 0$

Q: What is $\mathbb{P}(a \leq X < b)$?

Ex: Suppose X is a continuous r.v. with density function

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2, \\ 0 & o/w. \end{cases}$$

Q: What is C?

A: $1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^2 C(4x - 2x^2)\,dx = C\left(\frac{4x^2}{2} - \frac{2x^3}{3}\right)\Big|_0^2$

$= C\left(8 - \frac{16}{3}\right) = C\left(\frac{8}{3}\right)$

Therefore $C = \frac{3}{8}$.

Q: Find $\mathbb{P}(X > 1)$.

A: $\mathbb{P}(X > 1) = \int_1^{\infty} f(x)\,dx = \int_1^2 \left(\frac{3}{8}\right)(4x - 2x^2)\,dx = \frac{3}{8}\left(\frac{4x^2}{2} - \frac{2x^3}{3}\right)\Big|_1^2$

$= \frac{3}{8}\left[\left(8 - \frac{16}{3}\right) - \left(2 - \frac{2}{3}\right)\right] = \frac{3}{8}\left[\frac{8}{3} - \frac{4}{3}\right] = \frac{3}{8}\,\frac{4}{3} = \frac{4}{8} = \frac{1}{2}.$

Ex: (Exponential r.v.) An "exponential" r.v. has density function

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & x > 0, \\ 0 & o/w. \end{cases}$$

for some $\lambda > 0$

Suppose $\lambda = 100$ and X represents the lifetime of a light bulb (in hours).

Q: What is the probability that the bulb fails in the first 100 hours?

A: $\mathbb{P}(X \leq 100) = \int_{-\infty}^{100} f(x)\,dx = \int_0^{100} \left(\frac{1}{100}\right) e^{-x/100}\,dx$

$$= -e^{-\frac{x}{100}}\Big|_0^{100} = -\left(e^{-1} - 1\right)$$

$$= \left(1 - \frac{1}{e}\right)$$

Q: Find the c.d.f., $F(x)$, for X.

A: $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} \int_0^x \left(\frac{1}{100}\right) e^{-t/100}\,dt & x \geq 0, \\ 0 & x < 0. \end{cases}$

$$= \begin{cases} 1 - e^{-x/100} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Q: $\mathbb{P}(X > 300)$

A: $\mathbb{P}(X > 300) = 1 - F(300) = 1 - \left(1 - e^{-300/100}\right)$

$$= e^{-3}.$$

Q: Find $\frac{d}{dx} F(x)$.

A: $F'(x) = \begin{cases} \frac{1}{100} e^{-x/100} & x > 0 \\ 0 & x < 0 \end{cases}$

Note: Notice that $F'(x) = f(x)$ except at $x = 0$ where $F'(x)$ does not exist.

Proposition: $F' = f$ at all points where F' is defined.

Note: F' is always a density function for X — even if it doesn't exist at some points. The reason for this is that the definition of f at a single point (or countable set of points) is of no consequence. Remember $\mathbb{P}(X = a) = 0$ for all $a \in \mathbb{R}$. The integral $\int_A f(x)\,dx$ is not changed by redefining the value of f at a single point.

Ex: Let F_X, f_X be the cdf and density for X respectively. Suppose $Y = 2X$.

Q: Find F_Y and f_Y

A: $F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(2X \le y) = \mathbb{P}\left(X \le \frac{y}{2}\right) = F_X\left(\frac{y}{2}\right)$

Thus $f_Y(y) = f_X\left(\frac{y}{2}\right)\frac{1}{2}$.

Def: The "expected value" of a continuous r.v., X, is given by $E(X) = \int_{\mathbb{R}} x f(x)\,dx$

Def: The "variance" of a continuous r.v., X, is given by $E\left[X - E(X)\right]^2$.

Ex: Suppose X has cdf $F(x) = \begin{cases} x^2 & x \in (0,1) \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$

Q: Find $\text{var}(X)$.

A: $f(x) = 2x\, 1_{(0,1)}(x) = \begin{cases} 2x & \text{if } x \in (0,1), \\ 0 & \text{o/w.} \end{cases}$

$$E(X) = \int_0^1 x\, 2x\, dx = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$$

$$E(X^2) = \int_0^1 x^2\, 2x\, dx = \frac{2x^4}{4}\Big|_0^1 = \frac{1}{2}$$

$$\text{var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

Ex: Suppose X has density $f(x) = \begin{cases} 1 & x \in (0,1), \\ 0 & o/w. \end{cases}$

Q: Find $E(e^X)$

A: Let $Y = e^X$. $F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(e^X \le y)$

$$= \mathbb{P}(X \le \log(y)) = F_X(\log(y))$$

Therefore, $f_Y(y) = f_X(\log(y))\frac{1}{y} = \begin{cases} 1/y & \text{if } \log(y) \in (0,1), \\ 0 & o/w. \end{cases}$

$$= \begin{cases} \frac{1}{y} & 1 < y < e \\ 0 & o/w. \end{cases}$$

Now, $E(Y) = \int_{\mathbb{R}} y f_Y(y)\, dy$

$$= \int_1^e y \frac{1}{y}\, dy = e - 1.$$

Proposition: For a continuous r.v., X, and real-valued function, g,

$$E[g(X)] = \int_{\mathbb{R}} g(x) f(x)\, dx.$$

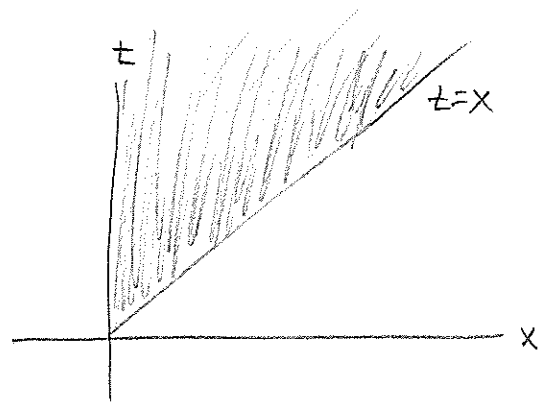
Ex: Redo the previous problem using the proposition above.

$$E(e^X) = \int_0^1 e^x \, dx = e^x \Big|_0^1 = e - 1.$$

Lemma: If X is non-negative and a continuous r.v., then $E(X) = \int_0^\infty \mathbb{P}(X > x)\,dx$

proof: $RHS = \int_0^\infty \mathbb{P}(X > x)\,dx = \int_0^\infty \int_x^\infty f(t)\,dt\,dx$

$$= \int_0^\infty \int_0^t f(t)\,dx\,dt$$



$$= \int_0^\infty f(t)\left(\int_0^t dx\right) dt$$

$$= \int_0^\infty t\, f(t)\,dt$$

$$= \int_{\mathbb{R}} t\, f(t)\,dt = E(X). \quad \square$$

Proposition: $E(aX+b) = a\,E(X) + b$ (where X is a continuous r.

proof: $E(aX+b) = \int_{\mathbb{R}} (ax+b)\, f(x)\,dx = a\int_{\mathbb{R}} x\, f(x)\,dx + b\int_{\mathbb{R}} f(x)\,dx$

$$= a\,E(X) + b. \quad \square$$

Proposition: For a continuous r.v., X, $\text{var}(X) = E(X^2) - [E(X)]^2$

proof: $\text{var}(X) = E[X - E(X)]^2 = E[X - \mu]^2 = \int_{\mathbb{R}} (x-\mu)^2 f(x)\, dx$

$$= \int_{\mathbb{R}} (x^2 - 2\mu x + \mu^2) f(x)\, dx$$

$$= \int x^2 f(x)\, dx - 2\mu \int x f(x)\, dx + \mu^2 \int f(x)\, dx$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

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

$$= E(X^2) - [E(X)]^2. \qquad \square$$

Def: A "uniform" r.v. has density function

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{o/w}. \end{cases}$$

for some $a < b$

Ex: Find $F(x)$, $E(X)$, and $\text{var}(X)$ when $X \sim \text{UNIF}(2,5)$

Ex: Find $\mathbb{P}(X < 2)$ when $X \sim \text{UNIF}(1,4)$.

Ex: My arrival time, Y, at a bus stop is uniformly distributed from 7AM to 7:30 AM.
Buses come every 15 minutes: 7, 7:15, 7:30.
Let X = minutes I wait for the bus.

Q: Find $f_X(x)$, $\mathbb{P}(X > 5)$, and $\text{var}(X)$

A: $$F_X(x) = \mathbb{P}(X \le x) = \begin{cases} 0 & x \le 0 \\ \int_{15-x}^{15} \frac{1}{30}\,dt + \int_{30-x}^{30} \frac{1}{30}\,dt & 0 < x < 15 \\ 1 & x \ge 15 \end{cases} = \begin{cases} 0 & x \le 0 \\ \frac{2x}{30} & x \in (0,15) \\ 1 & x \ge 15 \end{cases}$$

Thus $$f_X(x) = \begin{cases} \frac{1}{15} & x \in (0,15), \\ 0 & \text{o/w}. \end{cases}$$

Note: $X \sim \text{UNIF}(0,15)$.

Then, $\mathbb{P}(X > 5) = \frac{2}{3}$

$E(X) = 7.5$

$E(X^2) = \int_0^{15} x^2 \frac{1}{15}\,dx = \frac{x^3}{3(15)}\Big|_0^{15} = \frac{(15)^2}{3}$

$\text{var}(X) = \frac{15^2}{3} - \frac{15^2}{4}$.

Def: A continuous r.v. is called "normal" or "Gaussian" if it has density function

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

for some μ, σ.

Q: Prove that $f(x)$ is a density function

A: Let $I = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

Let $y = \frac{x-\mu}{\sigma}$. Then $\sigma dy = dx$ and

$$I = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Now $I^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dx\,dy$$

In polar coordinates, we have

$$I^2 = \int_0^{2\pi}\int_0^{\infty} \frac{1}{2\pi} e^{-r^2/2} r\,dr\,d\theta = \int_0^{2\pi} \left(\frac{-1}{2\pi} \cdot e^{-r^2/2} \Big|_{r=0}^{r=\infty} \right) d\theta$$

$$= \int_0^{2\pi} \frac{1}{2\pi} d\theta = 1.$$

Thus $I = 1$ and $f(x)$ is a density function.

Q: Find $E(X)$ and $var(X)$. Hint: Integration by parts is needed to find $var(X)$.

$$E(X) = \int x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

$$= \mu + \int (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \mu + \int y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \qquad \text{(by substitution)}$$

$$= \mu \quad \text{since} \quad \frac{y}{\sqrt{2\pi}} e^{-y^2/2} \quad \text{is an odd function.}$$

$$var(X) = \int (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \sigma^2 \int y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \qquad \text{(by substitution)}$$

$$= \sigma^2 \left(-\frac{y}{\sqrt{2\pi}} e^{-y^2/2} \Big|_{-\infty}^{\infty} \right) + \sigma^2 \int \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= \sigma^2$$

Proposition: If $X \sim N(\mu, \sigma^2)$, then $aX+b$ is normally distributed.

Proof: Let $Y = aX+b$. Then $F_Y(y) = \mathbb{P}(Y \le y)$

$$= \mathbb{P}(aX+b \le y) = \begin{cases} \mathbb{P}\left(X \le \frac{y-b}{a}\right) & a > 0, \\ \mathbb{P}\left(X \ge \frac{y-b}{a}\right) & a < 0, \\ \mathbb{1}\{b \le y\} & a = 0. \end{cases}$$

$$= \begin{cases} F_X\left(\frac{y-b}{a}\right) & a > 0, \\ 1 - F\left(\frac{y-b}{a}\right) & a < 0, \\ \mathbb{1}\{b \le y\} & a = 0. \end{cases}$$

Thus $f_Y(y) = \begin{cases} f_X\left(\frac{y-b}{a}\right)\frac{1}{a} & a > 0, \\ -f_X\left(\frac{y-b}{a}\right)\frac{1}{a} & a < 0, \\ \text{DNE} & a = 0. \end{cases}$

So, if $a \neq 0$, then $f_Y(y) = \frac{1}{|a|}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y-b}{a} - \mu\right)^2 / 2\sigma^2}$

$$= \left(\frac{1}{\sqrt{2\pi(a\sigma)^2}}\right) e^{-(y - [\mu a + b])^2 / 2(a\sigma)^2}$$

Thus $Y \sim N(a\mu + b, a^2\sigma^2)$

If $a = 0$, then Y is a constant (non-random) which is a normal random variable with zero variance. This case is degenerate and is <u>not</u> a continuous random variable. It is discrete. $\square$

Ex: Scores in math 5010 are approximately normally distributed with mean μ and standard deviation σ.

Grading scale:

A	$(\mu+\sigma, \infty)$
B	$(\mu, \mu+\sigma]$
C	$(\mu-\sigma, \mu]$
D	$(\mu-2\sigma, \mu-\sigma]$
E	$(-\infty, \mu-2\sigma]$

Q: Find the proportion of the class who will get an A (B, C, D, E).

A: $\mathbb{P}(X > \mu+\sigma) = \mathbb{P}\left(\frac{X-\mu}{\sigma} > 1\right) = 1 - \Phi(1) \approx .1587$

since $\frac{X-\mu}{\sigma} \sim N(0,1)$ (Prove it!)

and $\Phi(x)$ is the cdf of $N(0,1)$.

etc.

Def: An "exponential" r.v. has density function

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

for some $\lambda > 0$.

Optional Question: Find mean and variance.

Ex: Suppose $a \geq 0$. Your sister is on the phone again. She has been on it for a minutes already. Let X = length of her call. $X \sim EXP(\frac{1}{10})$.

Q: Find $\mathbb{P}(X > a+b \mid X \geq a)$: (Assume $b > 0$.)

A: $$\mathbb{P}(X > a+b \mid X \geq a) = \frac{\mathbb{P}(X > a+b \text{ and } X \geq a)}{\mathbb{P}(X \geq a)} = \frac{\mathbb{P}(X > a+b)}{\mathbb{P}(X \geq a)}$$

$$= \frac{\int_{a+b}^{\infty} \frac{1}{10} e^{-x/10}\, dx}{\int_{a}^{\infty} \frac{1}{10} e^{-x/10}\, dx} = \frac{-e^{-x/10}\Big|_{a+b}^{\infty}}{-e^{-x/10}\Big|_{a}^{\infty}} = \frac{e^{-(a+b)/10}}{e^{-a/10}}$$

$$= e^{-b/10}.$$ Note this doesn't depend on a.

This means that no matter how long she has been on, the probability that you will wait 5 more minutes is the same.

Chapter 6 — Jointly Distributed Random Variables

We are often interested in probabilities involving two (or more) r.v.'s such as $\mathbb{P}(X \leq 6 \text{ and } Y > 2)$. These types of probabilities cannot be computed if we only know the individual density (or mass) functions of X and Y. We need to know the joint distribution of X and Y.

jointly Discrete Case: $\mathbb{P}((X,Y) \in A) = \sum_{(x,y) \in A} p(x,y)$

where $p(x,y) = \mathbb{P}(X = x \text{ and } Y = y)$ is the "joint mass function."

jointly Continuous Case: $\mathbb{P}((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$

where $f(x,y)$ is the "joint density function."

Both Cases: We define the "joint c.d.f." of X and Y as $F(x,y) = \mathbb{P}(X \leq x \text{ \underline{and} } Y \leq y)$

In the continuous case this yields $\longrightarrow$ $= \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\,dx\,dy$

Def: The individual density (or mass) functions will be refered to as the "marginal" density (or mass) functions (eg $f_X(x)$, $f_Y(y)$, $p_X(x)$, $p_Y(y)$)

Ex: $f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & x>0,\ y>0, \\ 0 & o/w. \end{cases}$

Find the marginals.

Find the joint cdf..

Ex: 3 balls are scooped out of a bowl. The bowl contains 3 red, 4 white, and 5 blue balls. Let X = # of red and Y = # of white balls scooped out.

Find the joint mass function.

Find the marginals

Find the joint cdf..

Ex: Find the cdf and then the density of $W = \frac{X}{Y}$ when $f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & x>0 \quad y>0, \\ 0 & o/w. \end{cases}$

Def: Two r.v.'s are independent if $\mathbb{P}(X \in A \text{ and } Y \in B)$

$$= \mathbb{P}(X \in A)\,\mathbb{P}(X \in B)$$

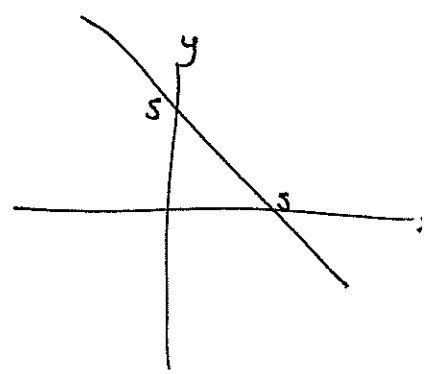
for all A, B.

Equivalently, $F(x,y) = F_X(x)\, F_Y(y)$.

Since $f(x,y) = \dfrac{\partial^2}{\partial x \partial y} F(x,y)$ independence for continuous r.v.'s can be expressed as $f(x,y) = f_X(x) f_Y(y)$.

Suppose X, Y are jointly continuous & independent. Then

$$F_{X+Y}(s) = \mathbb{P}(X+Y \le s) = \int_{-\infty}^{\infty} \int_{-\infty}^{s-x} f_X(x) f_Y(y)\, dy\, dx$$



$$= \int_{-\infty}^{\infty} F_Y(s-x)\, f_X(x)\, dx$$

Thus $f_{X+Y}(s) = \underbrace{\int_{-\infty}^{\infty} f_Y(s-x)\, f_X(x)\, dx}$

This is called the "convolution" of $f_X(x)$ and $f_Y(y)$.

Ex: Suppose X, Y are independent UNIF$(0,1)$ r.v.'s.

Find the joint density.

Find the density of $X+Y$.

Note: The convolution formula can be used to show that $X+Y$ is Normal if X, Y are independent normals. (Do this only if there is time to spare) [Use $X \sim N(0,1)$ $Y \sim N(0,\sigma^2)$]

Def: For two jointly discrete r.v.'s the "conditional mass function" of X given Y is

$$p_{X|Y}(x|y) = \mathbb{P}(X=x \mid Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} = \frac{p(x,y)}{p_Y(y)}.$$

Def: For two jointly continuous r.v.'s the "conditional density" of X given Y is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Now, we can define conditional expectations:

$$E(X \mid Y=y) = \begin{cases} \sum x \, p_{X|Y}(x|y) & \text{discrete case} \\ \int x \, f_{X|Y}(x|y) & \text{continuous case} \end{cases}$$

Joint distribution of functions of r.v.'s

Ex: $X, Y \sim$ independent UNIF$(0,1)$

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

Now Let $U = X + Y$
$V = X - Y$

Q: What is the joint density, $f_{UV}(u,v)$, for U, V?

Theorem: If (1) we can uniquely solve for X, Y in terms of U, V

(2) The "Jacobian" of the transformation is continuous and never zero

Then $f_{UV}(u,v) = f_{XY}\big(x(u,v), y(u,v)\big) \dfrac{1}{|J|}$,

where $J = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$ is the "Jacobian."

In our example, $x = \frac{(u+v)}{2}$

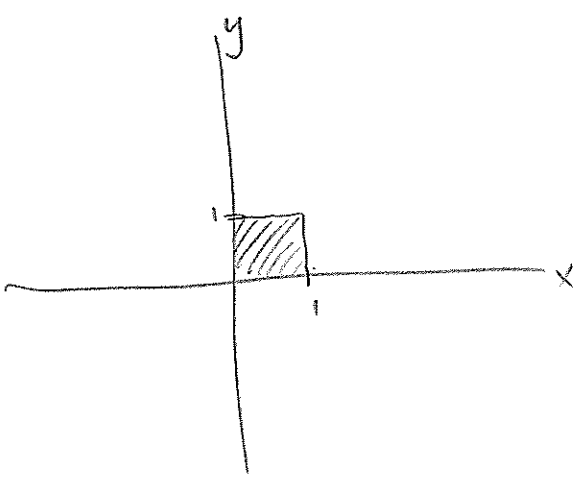
$$y = \left(\frac{u-v}{2}\right)$$

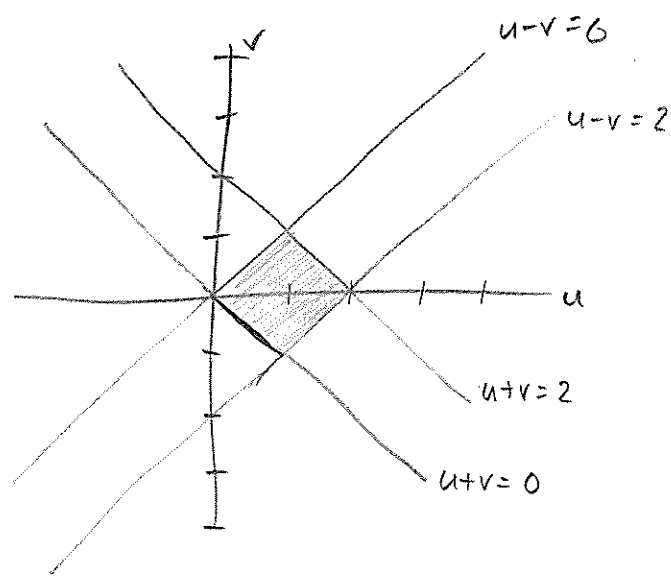
$$J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

Thus $f_{UV}(u,v) = f_{XY}\big(x(u,v), y(u,v)\big) \frac{1}{|J|}$

$$= f_{XY}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \frac{1}{2}$$

$$= \begin{cases} \frac{1}{2} & \frac{u+v}{2}, \frac{u-v}{2} \in (0,1), \\ 0 & \text{o/w}. \end{cases}$$





Ex: R, Θ independent, $R \sim EXP(1)$, $\Theta \sim UNIF(0, 2\pi)$

$$R = \sqrt{X^2 + Y^2} \qquad X = R\cos(\Theta)$$

$$\Theta = \tan^{-1}\left(Y/X\right) \qquad Y = R\sin(\Theta)$$

Find $f_{XY}(x, y)$.

$$J = \begin{vmatrix} \cos(\Theta) & -r\sin(\Theta) \\ \sin(\Theta) & r\cos(\Theta) \end{vmatrix} = r\cos^2(\Theta) + r\sin^2(\Theta)$$

$$= r.$$

$$f_{XY}(x,y) = f_{R\Theta}\left(r(x,y), \Theta(x,y)\right) \frac{1}{J}$$

$$= f_{R\Theta}\left(\sqrt{x^2+y^2}, \tan^{-1}\left(\frac{y}{x}\right)\right) \frac{1}{\sqrt{x^2+y^2}}$$

$$= e^{-\sqrt{x^2+y^2}} \frac{1}{\sqrt{x^2+y^2}}$$

Def: covariance (intuition)

(1) $\text{cov}(X,Y) = \text{cov}(Y,X)$

(2) $\text{cov}(X,X) = \text{var}(X)$

(3) $\text{cov}(aX,Y) = a\,\text{cov}(X,Y)$

(4) $\text{cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \text{cov}(X_i, Y_j)$

Def: Correlation (scaling)

(1) $\alpha \le \rho \le 1$

proposition: $\text{var}\left(\sum_{i=1}^{n} X_i\right) = \text{cov}\left(\sum X_i, \sum X_j\right)$

$$= \sum_i \sum_j \text{cov}(X_i, X_j)$$

$$= \sum \text{var}(X_i) + \sum\sum_{i \neq j} \text{cov}(X_i, X_j)$$

proposition: if $X_i \sim$ independent

Then $\text{var}\left(\sum X_i\right) = \sum \text{var}(X_i)$

Def: "Moment Generating Function"

$$M(t) = E\left(e^{tX}\right) = \begin{cases} \int e^{tx} f(x)\,dx & \text{continuous case} \end{cases}$$

Differentiate and eval @ zero

$$M'(t) = \frac{d}{dt} E\left(e^{tX}\right) = E\left[\frac{d}{dt}\left(e^{tX}\right)\right]$$

$$= E\left(Xe^{tX}\right)$$

$= E(X)$ when $t = 0$

$M''(0) = E(X^2)$, etc.

Ex: $X \sim BIN(n,p)$

$$M(t) = E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left(e^t p\right)^x (1-p)^{n-x}$$

$$= \left(e^t p + 1 - p\right)^n$$

Thus $E(X) = M'(0) = n\left(e^t p + 1 - p\right)^{n-1} pe^t \Big|_{t=0}$

$$= np$$

Ex: $POI(\lambda)$ Find $M(t)$ and $E(X)$

Prop: If $X_1, \ldots, X_n$ ~~iid~~ indep
Then $M_{\Sigma X_i}(t) = \prod M_{X_i}(t)$

perhaps prove only for two r.v.s $X+Y$

Chapter 8 — Limit Theorems

Markov's inequality: If $P(X > 0) = 1$ and $a > 0$, then $P(X \geq a) \leq \frac{E(X)}{a}$.

Proof: $\frac{X}{a} \geq 1\{\frac{X}{a} \geq 1\}$

$$\Rightarrow E\left(\frac{X}{a}\right) \geq E\, 1\{X/a \geq 1\}$$

$$\Rightarrow \frac{E(X)}{a} \geq P\left(\frac{X}{a} \geq 1\right) = P(X \geq a)$$

Chebyshev's inequality:

If X has mean μ and variance σ^2, then

$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$ for all $k > 0$.

Proof: $P(|X-\mu| \geq k) = P((X-\mu)^2 \geq k^2) \leq \frac{E(X-\mu)^2}{k^2} = \frac{\sigma^2}{k^2}$

Example: If $var(X) = 0$, then by the above inequality $P(|X-\mu| > k) = 0$ for all $k > 0$. Therefore $P(|X-\mu| > 0) = 0$ since

$$P(|X-\mu| > 0) = P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) = 0$$

where $A_1 = [1, \infty)$ and $A_i = [\frac{1}{i}, \frac{1}{i-1})$ for $i = 2, 3, \ldots$

Therefore $P(|X-\mu| = 0) = 1$, i.e $X = \mu$ with probability 1.

Law of large numbers (weak)
Suppose that $X_i \sim$ iid with mean μ. Then,

$$P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) \to 0 \quad \text{as } n \to \infty \quad \text{for all } \varepsilon > 0.$$

Proof: (under the restriction that variance exist (i.e. $E(X - EX)^2 < \infty$)).

Note that $\operatorname{var}(\bar{X}_n) = \operatorname{var}\left(\frac{X_1 + \dots + X_n}{n}\right)$

$$= \frac{1}{n^2} \operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot n \operatorname{var}(X_1) = \frac{\sigma^2}{n}.$$

By Chebyshev's inequality,

$$P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\left(\frac{\sigma^2}{n}\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

Law of Large numbers (strong)
If $X_i \sim$ iid with mean μ, then

$$P(\bar{X} \to \mu) = 1.$$

Central Limit Theorem

If $X_i \sim \text{iid}(\mu, \sigma^2)$, then $P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) \to \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt$

as $n \to \infty$.

That is, $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is approximately $N(0,1)$.

The proof is done by showing that the [illegible] of MGF [illegible]

The proof of this involves finding the moment generating function and then showing that it converges the the MGF of a standard normal.

Another nice approximation is that for large n and small p

$$BIN(n,p) \overset{approx}{\sim} POI(np)$$

A natural example of this is study enrollmen in which n and p might be unknown, but we can still estimate $\lambda = np$.

Example (Self-test 8.5)

$X_i \sim$ iid with density $2x\,\mathbb{1}\{x \in (0,1)\}$

X_i represents the time to failure of the ith light bulb in my shower.

$r = \lim_{n \to \infty} \frac{n}{\sum_{i=1}^{n} X_i}$ is the long run failure rate (in failures per year).

$$r = \lim \frac{1}{\bar{X}} \to \frac{1}{\mu} = \frac{3}{2} \quad \text{since}$$

$$\mu = E(X) = \int_0^1 2x^2\,dx = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$$