STA247

Probability with Computer Applications

Prof. K. H. Wong
Sinan Li

PUBLISHED BY PUBLISHER

BOOK-WEBSITE.COM

First printing, March 2013

Contents

I Contents

Contents

Introduction

This is an introductory course to probability, where our main focus will be developing an understanding of probability and the concept of probability distributions, both for discrete and continuous quantities. This includes developing the intuition for how probabilities 'behave' and the situations in which it is valid to describe randomness using probability, as well as relying on simulations in R to help us visualize these properties.

By the end of the course, students should be able to...

- Describe random quantities in various ways, such as by: their features, density functions, distribution functions, graphs
- Select an appropriate probability model based on their unique properties to quantify the randomness of random quantities
- Compute and interpret the various features of a random quantity: expected value, variance, standard deviation, correlation, covariance, event probabilities (either exactly, through approximations, or simulation)
- Select the most appropriate model to represent randomness and compute probabilities
- Use simulation in R for estimation purposes
- Explain the relationship of transforming a random variable and its effect on its distribution
- Use bivariate distributions to describe the association between two random quantities

Textbooks

We will be referencing these two textbooks regularly:

1. *Probability with Applications and R*, 2nd ed. by *Wagaman and Dobrow* through the library here and with student companion site here.
2. *Modern Mathematical Statistics with Applications*, 3rd ed. by *Devore and Berk* available through the library here.

We will be using R throughout the course to help us understand probability distributions, and to simulate probabilities for quantities that are harder to compute by hand. R Markdown files will be provided for you with necessary starter code. You'll also gain some experience with using LaTeXto produce documents with well-presented math notation.

For new users to R, you may find chapters 3-7 from R for Data Science to be helpful as reference material in data visualization and data manipulation tools in R.

Course Information

All course-related materials can be found on our Quercus page:

- **Lectures** are held in MC 102 for both sections. We begin 10 minutes after the hour, and no recordings will be provided.
- **Weekly materials** such as slides, suggested problems for practice, and reminders of upcoming due dates are posted each week's page on our course home page.
- **Tutorial materials** will be distributed at the beginning of your tutorial.
- **Announcements**: this is the primary channel to distribute important information to everyone! You're expected to check and read announcements regularly to ensure you don't miss any important communication!
- **Assignments** and documents will be distributed and submitted through Crowdmark.

Discussion Board

Throughout the term, beginning September 14, there will be pinned weekly discussions on our discussion board page. They will remain pinned from Wednesday 8 AM to the following Tuesday 8 PM.

- **General Q&A page**: general questions, clarifications, request for additional explanations, share your thoughts/understanding of topics
- **Grouped practice problem discussions**: post your solutions, thoughts, approaches, questions here for that collection of textbook questions.
  - Each discussion thread (i.e. reply to one original comment) is dedicated to ONE question, labeled in bold in the original comment

You earn your discussion board credits by contributing at least five (5) times during this course to any of the discussions during the pinned period (Wednesday to Tuesday) in any of the following ways:

- Posting a question to a problem you tried, with a clear explanation of your process, and if you got stuck: what you tried to do, and where you need help moving forward
- Responding to a question with a thoughtful explanation to help your peer by sharing your own understanding of the problem
- Posting to the general Q&A page with your own question, about a course topic that is still unclear and being specific in describing what you do not understand (and perhaps what you did understand!)
- Providing a detailed and clear response to a question in the general Q&A page

To ensure boards are easy to reference, posting similar content is discouraged and these would be ineligible for earning credits.

- Only contributions during the pinned period will be counted. The discussions will remain open the rest of term for students who come across new problems or would like to continue the discussions.
- You are encouraged to keep the discussion going, but in terms of credit, it will be capped at 5 points.
- While there is no weekly cap on points, the maximum points you can earn in the last two weeks is 2 points, with no more than 2 points per week (i.e. don't wait to the last minute to participate in the class discussion).
- The discussion boards exist to facilitate peer-to-peer collaboration and learning, while also encouraging regular active engagement with course content. The course offers many

opportunities that most students shouldn't find themselves unable to contribute in a unique way.

Why a Discussion Board?

- There are records and studies that have shown the process of explaining and teaching to others is an effective way to learn, consolidate, and retain what has been taught. See here and here.
- It's a space for students to come together to work collaboratively, receive and provide peer support.
- It's also a space to get feedback and guidance from TAs and myself.
- It's a good opportunity to self-assess ('how comfortable am I explaining this to another student?', 'how often do I need to refer back to my notes to explain this concept clearly?') – an important component of good study skills!
- It's valuable information to us! Common questions/misunderstandings that pop up in the weekly discussions can be addressed during our weekly lecture meetings.

Discussion Board Rubric

Points	1 point	0.5 points	0 points
Quality of contribution	Student has made a substantial and unique contribution with detailed explanations and/or clearly outlined process of the approach to a problem.	Student has made a contribution to the discussion that is dismissive, lacking in detail, or not completely unique. Unable to further the discussion in a way that fosters a collaborative learning environment.	Student has not contributed to the weekly discussion topic thread, or whose posts are off-topic/irrelevant/do not contribute to the thread or is not unique to what has already been discussed in the thread.
	Student was involved in follow-up discussions and worked collaboratively with their peers to develop a better understanding of the concepts involved.	e.g. responses such as 'you just need to integrate this and solve for it' or 'I got the same answer doing... (reiterates OPs process)'	e.g. 'I got the same answer', 'How did you get that number?'

Tutorial

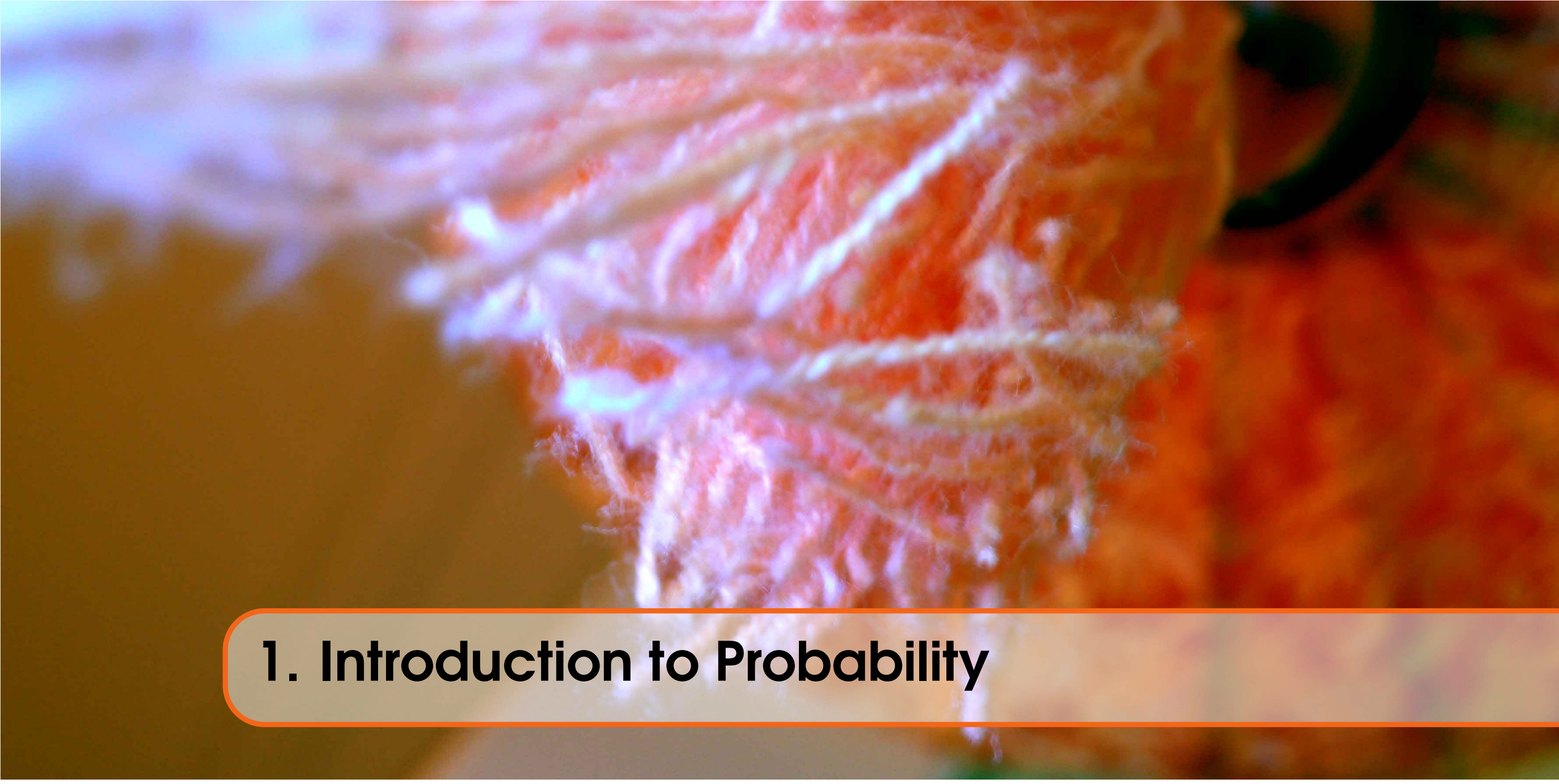
- A mix of R labs and collaborative pair work
- R labs: These labs have guided exercises to practice the R tools covered in class or learn new R skills. Labs will be TA-guided.
- R labs require individual .rmd and knitted document submissions at the end of tutorial (Note that the labs are guided and meant to be completed within the tutorial time)
- Pair work: In your tutorial section, you will with a partner of your choice work on more challenging but guided problems together. Discussion and sharing your ideas is a great way to learn from one another, and consider different approaches to problem solving! TAs will be there to support and help answer clarification questions.

Habits for Success

- Attend and participate in lecture. Try to work along with the problems presented. Ask questions and interact with your peers during the open work periods!
- Make sure you focus on *understanding* the concepts and how they relate and build upon each other. This one is hard!
- Regularly attempt as many of the suggested problems as you can and work towards being able to work on the problems closed book. Use this to gauge your familiarity with the material – if it takes you an hour to work through two problems, then it's a good indication to seek out advice and support from the teaching team. The earlier, the better!
- Drop by during the office hours or post on discussion board if you get stuck. Work through practice problems with classmates. Take turns *explaining* your thinking and problem solving process.
- **Create a schedule and stick with it**. This course covers many topics to ensure you have a good foundation for latter courses. Many topics build on top of each other so falling behind can quickly snowball and make it difficult to catch up.

Some Suggestions from Previous Students

1. Will the final be cumulative?
   Yes.
2. I find that I *really* struggle with keeping my mind focused on the question I'm presenting working on, and not think about how a similar question was solved previously. I thought I was making some progress, but then I realized this wasn't the case at all after today when I spent 30 minutes trying to recall the solution I wrote for A2 (I don't remember even reading the test question..I just caught a few phrases and began regurgitating my assignment solution incorrectly).
   I realize the most helpful thing I can do for myself now is to practice regularly, but I can't say I've been doing that well. For the last week, I've been attempting the textbook questions (open book)in preparation for the test. For the final, I plan to solve questions independently, which was something I didn't do until last minute. What else can I do to improve my problem-solving/critical thinking skills?
   Definitely do textbook / slide questions closed book. The only thing you should be looking at is the sample aid sheet. Otherwise you will never know if you actually understand the material.
3. You need to know when you don't know something. Meaning that when you see a question, you need to recognize if you actually know how to solve it or not. And if you don't know how to solve it you need to skip it and go to the next question. I'm sure you know that spending 30 min on any question is not an efficient use of the time.
4. Go to office hours. There is only like 2 other people that I've ever seen at office hours.
5. What does data have to look like for an exponential distribution to describe it?
   Why do we use different distributions? What's the importance of justifying a decision to use one distribution over another? (the textbook often tells you the distribution so image all of the questions posed to you without that information. Would you still be confident?) Answering these questions thoroughly will make you a lot better at figuring out how to approach a question when you see it.
   Understanding the conceptual side in depth is the most important part of studying I would think. Computational skills come with practice, eventually you get it. But usually the problem with stats or courses like this is interpreting the question - is figuring out how to even approach the question in the first place. Knowing how to solve all of the normal distribution questions that confront you is great, but that won't help you if you can't even recognize when and when not to use it.



1. Introduction to Probability

1.1 Useful Terminologies

1.1.1 Random Experiments

Definition 1.1.1 — Random Experiment. A process that allows us to gather data or observations. Experiment can be repeated multiple times under the same conditions. The set of possible outcomes of the experiment are known, but the outcome of a specific experiment is not known.

■ **Example 1.1 — Random Experiments.** Below are some examples of random experiments.

- Rolling a die and observing the top-facing number
- Rolling a pair of dice and observing the sum of top-facing numbers
- A patient being administered a painkiller and observing the amount of time in minutes before relief is felt

■

Definition 1.1.2 — Sample Space. The *sample space* is the set of all possible outcomes/results from a random experiment, usually denoted by Ω or S. The elements in the sample space are determined by the outcome of interest. Elements and are often denoted by ω.

■ **Example 1.2 — Sample Space.** Below are some examples of the sample space of random experiments.

- Experiment: Rolling a 20-sided die and observing the top-facing result.
  - $\Omega = \{1, 2, 3, \ldots, 20\}$ OR
  - $S = \{1 \leq x \leq 20, x \in \mathbb{Z}\}$
- Experiment: Selecting a random student and observing whether they are a CS student
  - $\Omega = \{\text{CS Student}, \text{Non-CS Student}\}$ OR
  - $S = \{0, 1\}$ where $\begin{cases} 0 = \text{Non-CS Student} \\ 1 = \text{CS Student} \end{cases}$
- Experiment: Select a random student and record the amount of liquids consumed that day $\Omega = \{L \geq 0, L \in \mathbb{R}\}$

■

Definition 1.1.3 — Event. An *event* is a subset of the sample space, usually represented by a capital letter near the beginning of the alphabet. A *simple event* has exactly one element of the sample space, while a *compound event* consists of multiple elements.

Definition 1.1.4 — Complement Event. The *complement event* is the set of outcomes in Ω that are not in A. Can be denoted as one of: A^c (preferred), $\overline{A}$, or A'.

■ **Example 1.3** Consider the following example.

- $A = B^c =$ roll an even number $= \{2,4,6\}$
- $B = A^c =$ roll an odd number $= \{1,3,5\}$

Here A and B are complement events. ■

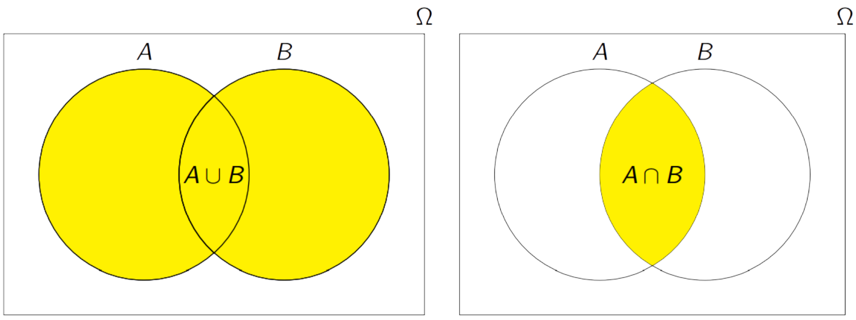
1.1.2 Operations on Sets

Definition 1.1.5 — Union. The *union* of two events A and B is the set of outcomes that are elements of A, B, or both. This is denoted as $A \cup B$.

- Union of events is usually described as A *or* B
- Note that $A \cup A^c = \Omega$

Definition 1.1.6 — Intersection. The *intersection* of two events A and B is the set of outcomes that are common to both A and B. This is denoted as $A \cap B$, or AB in the textbook.

- Intersection of event is usually described as A *and* B.
- Note that $A \cap A^c = \emptyset$



■ **Example 1.4 — Examples of Event Relations.** Let's keep it simple: suppose we roll two differently coloured[1] dice and record the paired outcomes. List out the following events:

a) Outcomes are doubles.
Let D be the event we roll doubles.
$D = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$

b) Outcomes sum to 8.
Let E be the event where rolls sum to 8.
$E = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

c) Outcomes where one dice has twice the face value as the other.
Let T be the event where one die outcome is twice the other.
$T = \{(1,2),(2,4),(3,6),(6,3),(4,2),(2,1)\}$

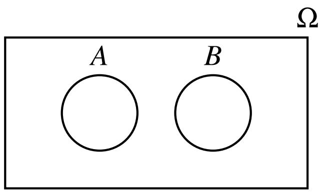
Note that here $E \cap T = \emptyset$ but $E \cup T \neq \Omega$. E and T are *disjoint* or *mutually exclusive*. E is **not** the complement of T. ■

subsectionEvent Relations

[1] Note that these two dice are distinguishable, so we would get $\omega = (i, j)$

Definition 1.1.7 — Mutually Exclusive. Two events A and B are *mutually exclusive* if the events cannot both occur or occur simultaneously as an outcome of the experiment. This means that the **intersection** of A and B is **empty** and they have no overlapping elements. A and B are also called *disjoint* events.

In the following Venn Diagram, A and B would be disjoint.



Definition 1.1.8 — Independent. Two events A and B are *independent* if the occurrence of one event does not alter the probability of occurrence of the other in any way.

Example 1.5 Classify the following paris of events / variables as mutually exclusive, independent, or dependent.

a) A student surveyed at random: $A = \{\text{Studies CS}\}$, $B = \{\text{Studies Stats}\}$
Dependent.
b) A biased coin (80% Head) is tossed twice: $C = \{\text{Head on first toss}\}$ and $D = \{\text{Tail on second toss}\}$
Independent.
c) An individual surveyed at random: $E = \{\text{Dislikes hiking}\}$ and $F = \{\text{Likes or indifferent to hiking}\}$
Mutually Exclusive.
d) A playing card is drawn: $G = \{\text{Card is red}\}$ and $H = \{\text{Card is Queen of Hearts}\}$
Dependent.

■

1.1.3 Important Laws

The following laws are useful relationships between unions and intersections of events that can help re-express events in simpler forms.

Theorem 1.1.1 — Commutative Law.

$$A \cup B = B \cup A$$

Theorem 1.1.2 — Associative Law.

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

Theorem 1.1.3 — Distributive Law.

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

In addition to the commutative, associative, and distributive laws, **DeMorgan's Laws** present some interesting relationships between union and intersection of events.

Theorem 1.1.4 — DeMorgan's Laws. For two events A and B,

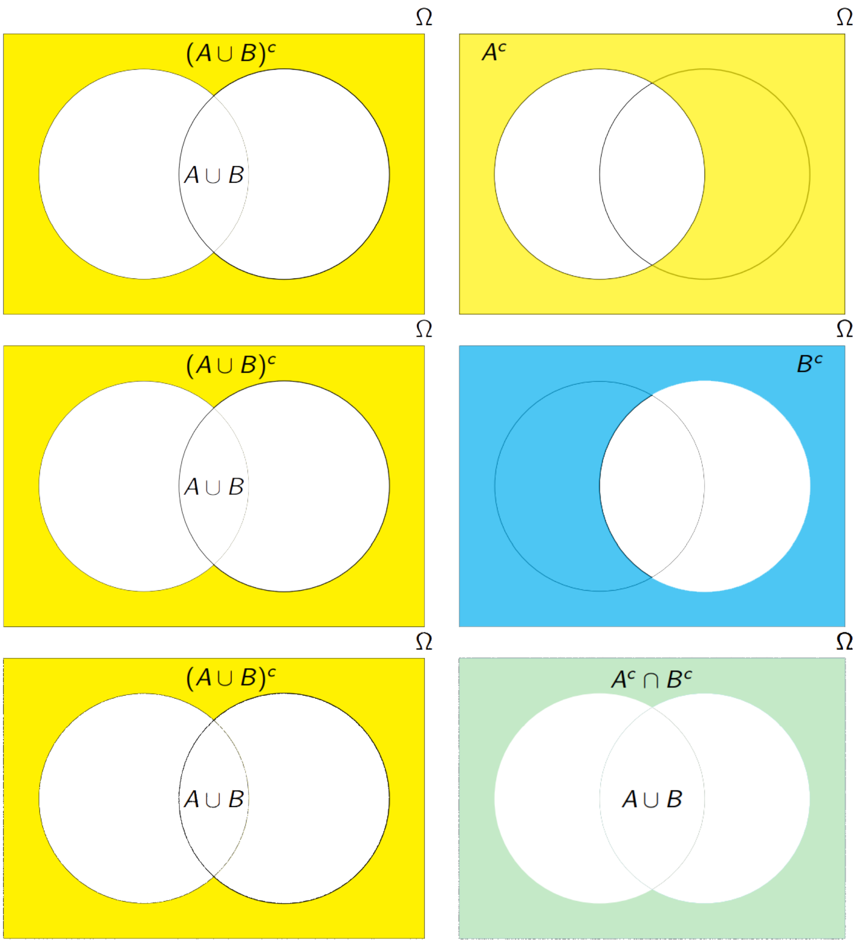
$$(A \cup B)^c = A^c \cap B^c$$

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

Or more generally, for the set of events $\{A_1, A_2, \dots, A_n\}$,

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i{}^c$$

$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i{}^c$$



1.2 Probability Functions

1.2.1 Probability Functions

Definition 1.2.1 — Probability. In a random experiment with sample space Ω, the *probability* of an event A, denoted as $P(A)$ is a function that assigns to event A a **numerical value** ($P(A) \in [0,1]$) that measures the chance that event A will occur.

There are certain axioms that must hold for probability functions.

axiom 1.2.1 — Axiom 1.

$$P(A) \geq 0$$

axiom 1.2.2 — Axiom 2.

$$P(\Omega) = 1$$

axiom 1.2.3 — Axiom 3. For a set of **disjoint** (mutually exclusive) elements $A_1, A_2, \ldots, A_n$ in Ω,

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

Probabilities for outcomes of a random experiment can be represented in many ways. When we have outcomes that can be represented discretely, we can express the associated probability as a function, called a *probability function*. A valid probability function must satisfy all the probability axioms.

Definition 1.2.2 — Probability Function. Suppose the sample space Ω can be represented with a finite number of elements: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ or a countably infinitely elements $\Omega = \{\omega_1, \omega_2, \ldots\}$, then the *probability function* P is a function on Ω with the following properties:

1. $P(\omega) \geq 0$ for all $\omega \in \Omega$.
2. $\sum_{\omega \in \Omega} P(\omega) = 1$
3. For all events $A \subseteq \Omega$, $P(A) = \sum_{\omega \in A} P(\omega)$

1.2.2 Probability and Event Relations

Using the definition of probability, as well as the event relations learned last week, we can relate probabilities of event relations.

Proposition 1.2.4 — Complement Probability. $A \cup A^c = \Omega$ and A is disjoint from A^c. Then,

$$P(\Omega) = P(A \cup A^c)$$

We can further derive that

$$\begin{aligned} 1 &= P(A \cup A^c) && \text{Axiom 2} \\ 1 &= P(A) + P(A^c) && \text{Axiom 3} \\ P(A^c) &= 1 - P(A) \end{aligned}$$

■ **Example 1.6** Suppose we have a box of 20 dice. Eight of the dice are custom made, and 15 of them are 10-sided dice. Assuming that each die belongs in either of these two categories, can we determine how many are custom made and are 10-sided?

Let $n(X)$ be the number of elements in X.
Let C be the custom made die
Let T be the 10-sided die
We know that $n(\Omega) = 20$, $n(C) = 8$, $n(T) = 15$, and $n((C \cup T)^c) = 0$.

$$\begin{aligned} 8 + 15 &= 23 && \text{notice } n(C \cup T) \text{ was counted twice} \\ \text{excess} &= \text{over count} \\ n(C \cap T) &= 3 \\ n(c \cup T) &= n(\Omega) \end{aligned}$$

■

■ **Example 1.7 — Union and Intersection.** Notice event B (or A) can be broken down into the following mutually exclusive events:

$$\begin{aligned} B &= (B\cap A)\cup(B\cap A^c) \\ P(B) &= P(A\cap B)+P(A^c\cap B) \quad \text{Axiom 3} \\ P(A^c\cap B) &= P(B)-P(A\cap B) \end{aligned}$$

Note also that $A\cup B$ can also be broken down in a similar way:

$$\begin{aligned} P(A\cup B) &= P(A\cup(A\cap B)\cup(B\cap A^c)) & \\ &= P(A\cup(B\cap A^c)) & A\cap B\subseteq A \\ &= P(A)+P(A^c\cap B) & \text{Axiom 3, since disjoint} \\ P(A\cup B) &= P(A)+P(B)-P(A\cap B) & \text{from above} \end{aligned}$$

If A and B are disjoint (mutually exclusive), then $P(A\cup B)=P(A)+P(B)$ since $P(A\cap B)=P(\emptyset)=0$. ■

■ **Example 1.8** In a class of 50 students, 23 could not roll their tongue, 15 had attached earlobes, and 10 could roll their tongues and had attached earlobes. A student is randomly selected from the class. Let T denote the event that the student can roll their tongue, and E denote the event that they have attached earlobes. Symbolically denote the following events and identify the number of students in each.

a) The student can roll his or her tongue.
$$\begin{aligned} n(T) &= n(\Omega)-n(T^c) \\ &= 50-23 \\ &= 27 \end{aligned}$$
b) The student can neither roll his or her tongue nor has attached earlobes.
$$\begin{aligned} n((T\cup E)^c) &= n(\Omega)-n(T\cup E) \\ &= n(\Omega)-(n(T)+n(E)-n(T\cap E)) \\ &= 50-(27+15-10) \\ &= 18 \end{aligned}$$

■

In general, for the set of events $\{A_1,A_2,\ldots,A_n\}$, we have the **Includion-Exclusion Principle**

Theorem 1.2.5 — Includion-Exclusion Principle.

$$\begin{aligned} P\left(\bigcup_{i=2}^{n}A_i\right) &= \sum_{r=1}^{n}\left((-1)^{r+1}\sum_{i_1<i_2<\cdots<i_r}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_r})\right) \\ &= \sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_i\cap A_j)+\cdots+(-1)^{r+1}\sum_{i_1<i_2<\cdots<i_r}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_r})+\cdots \\ &\quad +(-1)^{n+1}P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_n}) \end{aligned}$$

If the set of events $\{A_1,A_2,\ldots,A_n\}$ are **disjoint**, then the probability of the union is simply the sum of the probabilities of each set.

■ **Example 1.9** For three events A_1, A_2, and A_3,

$$\begin{aligned}
P(A_1 \cup A_2 \cup A_3) &= \sum_{r=1}^{n} \left((-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) \right) \\
&= (-1)^{1+1} (P(A_1) + P(A_2) + P(A_3)) + \\
&\quad (-1)^{2+1} (P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_3 \cap A_1)) + (-1)^{3+1} P(A_1 \cap A_2 \cap A_3) \\
&= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_3 \cap A_1) + \\
&\quad P(A_1 \cap A_2 \cap A_3)
\end{aligned}$$

■

■ **Example 1.10 — MMSA ex. 15.** A consulting firm presently has bids out on three projects. Let $A_i = \{\text{Award project } i\}$, for $i = 1, 2, 3$, and suppose that $P(A_1) = 0.22$, $P(A_2) = 0.25$, $P(A_3) = 0.28$, $P(A_1 \cup A_2) = 0.11$, $P(A_1 \cap A_3) = 0.05$, $P(A_2 \cap A_3) = 0.07$, and $P(A_1 \cap A_2 \cap A_3) = 0.01$.

Express in words each of the following events, and compute the probability of each event.

a) ${A_1}^C \cap {A_2}^C$

Event where they get neither project 1 nor project 2.

$$\begin{aligned}
P({A_1}^C \cap {A_2}^C) &= P((A_1 \cup A_2)^C) && \text{DeMorgan's Law} \\
&= 1 - P(A_1 \cup A_2) \\
&= 1 - (A(A_1) + P(A_2) - P(A_1 \cap A_2)) \\
&= 1 - (0.22 + 0.25 - 0.11) \\
&= 0.64
\end{aligned}$$

b) ${A_1}^C \cap {A_2}^C \cap A_3$

Event where they only win project 3.

${A_1}^C \cap {A_2}^C \cap A_3 = (A_1 \cup A_2)^C \cap A_3$ by DeMorgan's Law.

Note that $A_3 = ((A_1 \cup A_2)^C \cap A_3) \cup ((A_1 \cup A_2) \cap A_3)$.

By Axiom 3, $P(A_3) = P(((A_1 \cup A_2)^C \cap A_3)) + P(((A_1 \cup A_2) \cap A_3))$.

That is,
$$\begin{aligned}
P((A_1 \cup A_2)^C \cap A_3) &= P(A_3) - P((A_1 \cup A_2) \cap A_3) \\
&= P(A_3) - P((A_1 \cap A_3) \cup (A_2 \cap A_3)) \\
&= P(A_3) - (P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3)) \\
&= 0.28 - (0.07 + 0.05 - 0.01) \\
&= 0.17
\end{aligned}$$

c) What event can be user to represent the outcome the firm is awarded at least two project? Find this probability.

$(A_1 \cap A_2) \cup (A_2 \cap A_3) \cup (A_3 \cap A_1)$

Let $X = A_1 \cap A_2$, $Y = A_2 \cap A_3$ and $Z = A_3 \cap A_1$.

Then, by the inclusion-exclusion principle, we have

$$\begin{aligned}
P(X \cup Y \cup Z) &= P(X) + P(Y) + P(Z) - P(X \cap Y) - P(Y \cap Z) - P(Z \cap X) + P(X \cap Y \cap Z) \\
&= 0.11 + 0.07 + 0.05 - 0.01 - 0.01 - 0.01 + 0.01 \\
&= 0.21
\end{aligned}$$

■

2. Counting

2.1 Counting and Probability

Proposition 2.1.1 — Probability as Relative Frequency. The long-run relative frequency of an event A will approach the probability of event A. In cases where the sample space consists of **equally likely** elements, we can find this probability by calculating the relative frequency of A in Ω by

$$P(A) = \frac{\text{Number of outcomes in } A}{\text{Total number of possible outcomes in the random experiment}} = \frac{n(A)}{n(\Omega)}$$

This is valid **only if** each element in Ω is **equally likely**.

2.2 Fundamental Principle of Counting

Experiments that involve equally likely discrete outcomes make calculating probabilities much easier when we have methods to count outcomes.

- For an experiment with two events of interest, A and B, you can use the Inclusion-Exclusion Principle to count the number of outcomes in event $A \cup B$: $n(A \cup B) = n(A) + n(B) - n(A \cap B)$, where $n(X)$ denotes the number of elements in event X.
- For experiments that involve *multiple* ***ordered*** stages, we can use the **Fundamental Principle of Counting** (FPC) to count the number of unique outcomes from this multi-stage experiment.

■ **Example 2.1 — Toy Example.** A new sandwich shop seems to only have limited customization options. They offer three types of greens (lettuce, spinach, mixed greens), five types of deli meat, and four types of cheese. Sandwiches are built by layering with greens, followed by deli meat, and topping it off with cheese. How many unique sandwiches can be created if a customer randomly chooses one of each item to include in their sandwich?

This experiment involves 3 ordered stages:

- Stage 1: Pick a green - 3 choices
- Stage 2: Pick a deli meat - 5 choices
- Stage 3: Pick a cheese - 4 choices

Each option is equally likely to be chosen since we're considering all possible combinations There are in total $3 \times 5 \times 4 = 60$ unique combinations. ■

When counting the number of (ordered) outcomes from a multistage experiment, we can use the Fundamental Principle of Counting.

Theorem 2.2.1 If an experiment consists of m (**ordered**) stages with n_1 possible outcomes in stage 1, n_2 possible outcomes in stage 2, $\ldots$, n_m possible outcomes in stage m, then the total number of possible outcomes is

$$\prod_{i=1}^{m} n_i$$

2.3 Permutation

■ **Example 2.2** FPC counts specifically **ordered stages**. In the toy example, the number of sandwiches only include those that are layered in a specific order: with greens on the bottom, then deli meat, then cheese on top. However, customers might have a preference for how the ingredients are layered.

a) How many different ways can the ingredients (greens, deli meat, cheese) be layered/permuted?
   - Stage 1: 3 choices
   - Stage 2: 2 choices
   - Stage 3: 1 choice

   There are in total $3 \times 2 \times 1 = 6$ ways.
b) If the choice and order of ingredients each result in a 'different' sandwich, how many sandwich choices are there?
   6 choices to order ingredients $\times 60$ ingredient combinations $= 360$ 'different' sandwiches.

■

Definition 2.3.1 — Permutation - ${}_nP_n$. The number of ways to order n **distinct** item is

$$n! = n \times (n-1) \times \cdots \times 2 \times 1$$

Definition 2.3.2 — Permutation - ${}_nP_k$. The number of ways to select *ordered subset* of k elements from a group of n **distinct** items is

$${}_nP_k = \frac{n!}{(n-k)!}$$

The intuition behind this formula is to count all possible arrangements ($n!$) and group together all arrangements that have the same objects in the first k stages. The resulting number of 'groups' is the number of unique ordered subset. The number of elements in each group is equivalent to the number of ways to arrange the remaining $(n-k)$ objects.

3. Conditional Probability

4. Discrete Distributions

4.1 Random Variables

4.1.1 Introduction to Random Variables

Definition 4.1.1 — Random Variable. A *random variable* is a real-valued function that assigns a numerical value to each event in the sample space Ω arising from a random experiment. A random variable X is a **real-valued function** $X : \Omega \to \mathbb{R}$ such that for every $\omega \in \Omega, X(\omega) = x \in \mathbb{R}$. It is a mapping from the sample space to the real numbers.

■ **Example 4.1** Consider the random experiment of tossing a coin.

- $\Omega = \{H, T\}$
- Let X be the RV denoting the outcome of a toss. We ca define X such that $X(H) = 1$, $X(T) = 0$ essentially converting each outcome into a number.
- By convention, we will denote random variables with capital letters, and a particular (unknown value) of a random variable with its lower case equivalent. i.e. for a random variable X, a particular value of this RV would be denoted by x.

■

Definition 4.1.2 — Discrete Random Variable. A *discrete* of a random variable X is one tha can take on only a finite number of a countably infinite number of possible values x. A random variable X is *continuous* if its domain is an interval of real numbers.

Definition 4.1.3 — Probability Mass Function. A *probability mass function* (PMF) of a discrete random variable is one that assigns a probability to each value $x \in C$ such that

- $0 \leq P(X = x) \leq 1$
- $\sum_{x \in X} P(X = x) = 1$

■ **Example 4.2** Below are some examples of random variables.

Discrete RV Examples

- The number of defects in a day's production of car parts
- The number of new arrivals in a queue

- The status of your internet service: online or offline
- The number of students online at a particular time

Continuous RV Examples

- The weight of a randomly selected individual
- The time it takes to load a video
- The temperature in the morning of a random day

■

■ **Example 4.3** Determine the value of k such that $f(x) = \dfrac{kx^2 - x + 2}{4}$ will be a valid probability mass function for $X = \{0,1,2,3,4\}$.

X	0	1	2	3	4
$P(X=x)$	$\frac{2}{4}$	$\frac{k+1}{4}$	$\frac{4k}{4}$	$\frac{9k-1}{4}$	$\frac{16k-2}{4}$

We need $\sum_{x=0}^{4} P(X=x) = 1$. That is, $\dfrac{2+(k+1)+(4k)+(9k-1)+(16k-2)}{4} = 1$

$$30k = 4$$

$$k = \frac{2}{15}$$

Thus, k must be $\frac{2}{15}$ for $\sum_{x \in X} P(X=x) = 1$.

X	0	1	2	3	4
$P(X=x)$	$\frac{1}{2}$	$\frac{17}{60}$	$\frac{2}{15}$	$\frac{1}{20}$	$\frac{1}{30}$

■

■ **Example 4.4** A factory producing computer parts sends out a shipment of 10 parts of which 3 are defective. Find the probability mass function for the number of defectives a customer will get if the first customer randomly purchases 4 computer parts.

Let D be the number of defectives purchased.

$D = \{0,1,2,3\}$.

- $P(D=0) = \dfrac{{}_7C_4}{{}_{10}C_4}$
- $P(D=1) = \dfrac{{}_3C_1 \times {}_7C_3}{{}_{10}C_4}$
- $P(D=1) = \dfrac{{}_3C_2 \times {}_7C_2}{{}_{10}C_4}$
- $P(D=1) = \dfrac{{}_3C_3 \times {}_7C_1}{{}_{10}C_4}$

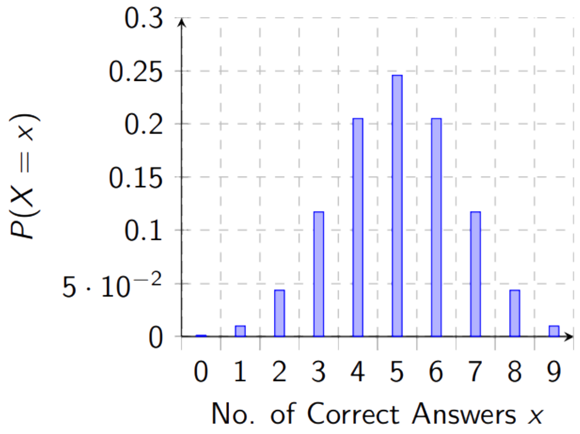
We may conclude that $P(D=d) = \dfrac{{}_3C_d \times {}_7C_{4-d}}{{}_{10}C_4}$. ■

■ **Example 4.5** A quiz consists of 10 true/false problems. A student takes the quiz by randomly selecting the answers. Examine the graph of the probability mass function and describe the behaviour of the random number of correct answers X.

For discrete distribution, we use bar charts. The bins are $0,1,2,3,\ldots,10$, which are discrete. Histograms, on the other hand, has bins that span a continuous interval, for example, $[0,1),[1,2),[2,3),\ldots$.

- $4,5,6$ have the highest probability mass, meaning they are the most *likely*;
- $0,1,2,8,9,10$ have the lowest probability mass, meaning that they are the most *unlikely*;

- 3 and 7 lie between *likely* and *unlikely*.



■

4.1.2 Characteristics of Random Variables

Both the PMF and CDF describe the exact distribution of a random quantity. Distributions have two useful characteristics that are often used in statistics:

- *Expected Value*: The long run/theoretical average. If a random experiment were to be conducted n times, then as $n \to \infty$ then the average of outcomes converges to the expected value. This is often denoted as μ.
- *Variance* or *Standard Deviation*: Are measures of the spread and variability of a random variable. Standard deviation is the square root of variance. Variance is often denoted as σ^2 and SD as σ.

Definition 4.1.4 — Expected Value. The *expected value* of a discrete random quantity X is defined to be

$$\mu = E(X) = \sum_{x \in X} x \cdot P(X = x)$$

For a given transformation of X, $g(X)$, $E(g(x))$ can be found by

$$E(g(x)) = \sum_{x \in X} g(x) \cdot P(X = x)$$

Note that $E(g(x)) \neq g(E(X))$ except when $g(X)$ is a linear transformation.

Review of Summation Rules

- $\sum_{i=1}^{n} c = n \cdot c$ for a constant c
- $\sum_{i=1}^{n} c \cdot x_i = c \cdot \left(\sum_{i=1}^{n} x_i \right)$
- $\sum_{i=1}^{n} i = \frac{n \cdot (n+1)}{2}$

■ **Example 4.6** A mining company needs a type of drill bit for a project. It is known through historical data that these drill bits for similar projects will last 2, 4, or 7 hours with probabilities 0.1, 0.7, and 0.2. How long do they expect each drill bit to last, on average?

Let L be the longevity of a random drill bit.

l	2	4	7
$P(L = l)$	0.1	0.7	0.2

$$
\begin{aligned}
E(L) &= \sum_{l \in L} l \cdot P(L = l) \\
&= 2 \cdot 0.1 + 4 \cdot 0.7 + 7 \cdot 0.2 \\
&= 4.4 \text{ hours}
\end{aligned}
$$

$\therefore$ on average, the drill bits will last 4.4 hours. ■

Properties of Expectation

For any constants a, b, and discrete variables X, the following are true.

- $E(a) = 0$
- $E(X + a) = E(X) + a = \mu + a$.
  Increase in $x \in X$ will shift the centre / average by the same amount.
- $E(aX) = a \cdot E(X) = a \cdot \mu$

  Proof. Take $g(X) = aX$. Then,
  $$
  \begin{aligned}
  E(g(x)) &= \sum_{x \in X} g(x) \cdot P(X = x) \\
  &= \sum_{x \in X} ax \cdot P(X \in x) \\
  &= a \sum_{x \in X} x \cdot P(X = x) \\
  &= a \cdot E(x) \\
  &= a \cdot \mu
  \end{aligned}
  $$
  That is, $E(aX) = a \cdot E(X) = a \cdot \mu$. ■

- $E(aX + b) = a \cdot E(X) + b = a \cdot \mu + b$

  Proof. Take $g(X) = aX + b$. $g(x)$ is a linear transformation of X.
  $E(g(x)) = g(E(x))$ if $g(X)$ is a linear transformation of X.
  Thus, $E(aX + b) = a \cdot E(X) + b = a \cdot \mu + b$. ■

- $E(X + Y) = E(X) + E(Y)$.
- $E(XY) \neq E(x) \cdot E(Y)$ **unless** X and Y are *independent*.

Definition 4.1.5 — Variance. For a discrete variable X, the *variance* of X is defined to be

$$\sigma^2 = V(X) = E((X - \mu)^2) = \sum_{x \in X} (x - \mu)^2 \cdot P(X = x)$$

Variance captures the spread in *units*^{2}. The standard deviation, $\sigma = \sqrt{\text{variance}}$ is a measure of spread in the same units as the random variable X.

■ **Example 4.7** A mining company needs a type of drill bit for a project. It is known through historical data that these drill bits for similar projects will last 2, 4, or 7 hours with probabilities 0.1, 0.7, and 0.2. Find the variance and standard deviation in the longevity of this type of drill bit. Interpret the values

Let L be the longevity of a random drill bit.

We also found $\mu = 4.4$

l	2	4	7
$P(L = l)$	0.1	0.7	0.2
$(l - \mu)^2$	$(2 - 4.4)^2 = 5.76$	$(4 - 4.4)^2 = 0.16$	$(7 - 4.4)^2 = 6.76$

$$\begin{aligned} V(L) &= {\sigma_L}^2 \\ &= \sum_{l \in L} (l-\mu)^2 \cdot P(L=l) \\ &= 5.76 \cdot 0.1 + 0.16 \cdot 0.7 + 6.76 \cdot 0.2 \\ &= 2.04 \text{ hours}^2 \end{aligned}$$

Then, $SD(L) = \sqrt{V(L)} = \sigma_L$
$= \sqrt{2.04}$
≈ 1.4283 hours

On average, the longevity of the drill bits will vary by about 1.43 hours from the average. Typically, we expect the longevity to be between $(4.4 - 1.4283, 4.4 + 1.4283) = (2.97, 5.83)$ hours.
■

Properties of Variance

For any constants a, b, and discrete variables X, the following are true.

- $V(a) = a$
  Constants do not vary.
- $V(X+a) = V(X) = \sigma^2$
  If the spread of X is $V(X)$, increasing each $x \in X$ will not change how spread out the random variable is.
- $V(aX) = a^2 \cdot V(X) = a^2 \cdot \sigma^2$

Proof.
$$\begin{aligned} V(aX) &= E\left((aX - E(aX))^2\right) \\ &= E\left((ax - a\mu)^2\right) \\ &= E\left((a \cdot (x-\mu))^2\right) \\ &= E\left(a^2 \cdot (x-\mu)^2\right) \\ &= a^2 \cdot E\left((x-\mu)^2\right) \\ &= a^2 \cdot V(X) \qquad = a^2 \cdot \sigma^2 \end{aligned}$$
That is, $V(aX) = a^2 \cdot V(X) = a^2 \cdot \sigma^2$. ■

- $V(aX+b) = a^2 \cdot V(X) = a^2 \cdot \sigma^2$
- $V(X+Y) \neq V(x) + V(Y)$ **unless** X and Y are *independent*.

■ **Example 4.8** A mining company needs a type of drill bit for a project. It is known through historical data that these drill bits for similar projects will last 2, 4, or 7 hours with probabilities 0.1, 0.7, and 0.2.

a) If they ordered 10 drill bits of the same type for replacement once one drill bit fails, how long can they expect these drill bits to last for this project?
   Let L be the longevity of drill bits.

$$\begin{aligned} E(L_1 + L_2 + \cdots + L_{10}) &= \sum_{i=1}^{10} E(L_i) \\ &= 4.4 + 4.4 + \cdots + 4.4 \\ &= 44 \text{ hours} \end{aligned}$$

b) Find the variance and standard deviation in the longevity for the 10 drill bits that were ordered.

All drill bits are independent, so
$$\begin{aligned} V(L_1+L_2+\cdots+L_{10}) &= \sum_{i=1}^{10} V(L_i) \\ &= 2.04+2.04+\cdots+2.04 \\ &= 20.4 \text{ hours}^2 \end{aligned}$$
Then,
$$\begin{aligned} SD(L_1+L_2+\cdots+L_{10}) &= \sqrt{20.4} \\ &\approx 4.52 \text{ hours} \end{aligned}$$
On average, they can expect the drill bits to last 44 ± 4.52 hours. That is, from 39.48 hours to 48.52 hours.

■

An alternative method for calculating the variance of a discrete random variable X with PMF $f(x)$ can be derived as follows:

$$\begin{aligned} E((X-\mu)^2) &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu \cdot E(X) + \mu^2 \\ &= E(X^2) - 2E(X)\cdot E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

This breakdown is "allowed" since $\mu = E(X)$ is an unknown **constant**.

Note that $E(X^2) \neq E(X)^2$. To compute $E(X^2)$, refer back to the definition of expected value. That is,

$$E(X^2) = \sum_{x \in X} x^2 \cdot f(x)$$

4.2 Cumulative Distribution Function

The probability behaviour of a random variable can be represented in many ways, such as with the probability mass function. Another representation is with the *cumulative distribution function*.

Definition 4.2.1 — Cumulative Distribution Function. The *cumulative distribution function* (CDF) $F(x)$ of a discrete random variable with probability mass function $P(x)$ or $f(x)$ is a function that returns the cumulative (total) probability up to and including $X = x$.

$$F(b) = P(X \leq b) = \sum_{x \in \{x \leq b\}} P(x)$$

The domain of the CDF is always over the set of real numbers! As such, CDFs are often represented as a piecewise function.

■ **Example 4.9** Find the cumulative distribution function for PMF below:

x	0	1	2	3
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{6} & \text{if } 0 \leq x < 1 \\ \frac{2}{3} & \text{if } 1 \leq x < 2 \\ \frac{29}{30} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

■

4.2.1 Properties of CDF

CDF of a Discrete Random Variable

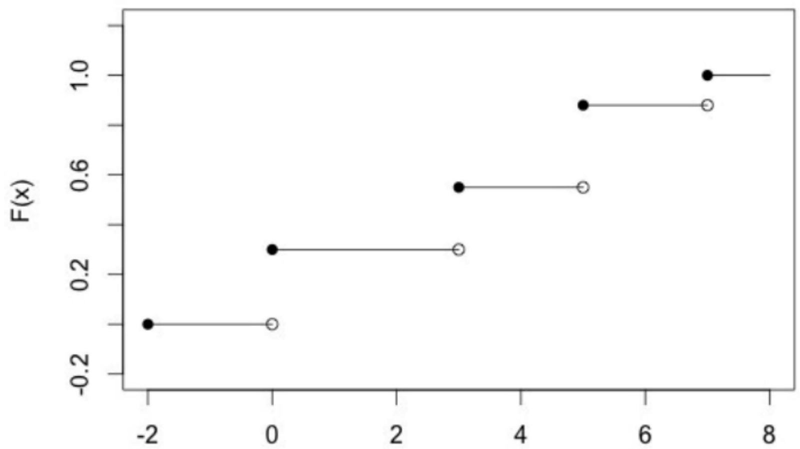
For a discrete random variable X with CDF $F(X)$:

1. The graph of the CDF will be a **non-decreasing step-function**. That is for $a < b$, $F(a) \leq F(b)$.
2. The graph of the CDF is **right continuous**. That is, $\lim_{x \to c^+} F(x) = F(c)$.
3. $\lim_{x \to \infty} F(x) = 1$
4. $\lim_{x \to -\infty} F(x) = -1$

■ **Example 4.10** A discrete random variable X has cumulative distribution function defined by

$$F(X) = \begin{cases} 0 & x < 0 \\ 0.3 & 0 \leq x < 3 \\ 0.55 & 3 \leq x < 5 \\ 0.88 & 5 \leq x < 4 \\ 1 & x \geq 7 \end{cases}$$

a) Plot the CDF.



b) What is the probability that X is less than 5?
$P(x < 5) = 0.55$

c) What is the probability of $X = 2$?

$$\begin{aligned} P(X = 2) &= P(X \leq 2) - P(X < 2) \\ &= 0.3 - 0.3 \\ &= 0 \end{aligned}$$

Note that $P(X = 2) = 0$ has no changes to CDF on interval of $[0, 3)$. The only outcomes with probability mass are $X = 0$ and $X = 3$.

■

4.2.2 Chebyshev's Inequality

For a given random variable X, μ_X and $\sigma^2{}_X$ are measures of two features of the distribution of X: it's 'centre' and it's spread. How can we use these two values to better understand the distribution of X, especially in the absence of the exact distribution such as the probability mass function (PMF) or the cumulative distribution function (CDF)?

Theorem 4.2.1 — Markov's Inequality. Let X be a **non-negative** random variable with mean $E(X)$/ Then, for some constant $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof. Since X is a non-negative random variable, $\forall x \in X, x \geq 0$.

$$\begin{aligned}
\text{Then, } E(X) &= \sum_{x \in X} x \cdot P(X = x) \\
&= \sum_{x<a} x \cdot P(X = x) + \sum_{x \geq a} x \cdot P(X = x) \\
&\geq \sum_{x \geq a} x \cdot P(X = x) \\
&\geq \sum_{x \geq a} a \cdot P(X = x) && \text{since } x \geq a \\
&= a \cdot \sum_{x \geq a} P(X = x) \\
&= a \cdot P(x \geq a)
\end{aligned}$$

That is, $E(X) \geq a \cdot P(X \geq a)$ ■

$$\frac{E(X)}{a} \geq P(X \geq a)$$

Theorem 4.2.2 — Chebyshev's Inequality. Let X be a random variable with mean (expected value) μ and finite variance σ^2. Then for any positive k,

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Chebyshev's Inequality applies to all discrete distributions with **finite** $E(X)$ and $V(X)$ for the random variable X.

Proof. $P(|X - \mu| < k\sigma) = P((x-\mu)^2 < k^2\sigma^2)$ since RV's are non-nagative and $k > 0$

$$= 1 - (P((x-\mu)^2 \geq k^2\sigma^2))$$

By Markov's Inequality, for a non-negative random variable X and a positive constant a, $P(X \geq a) \leq \dfrac{E(x)}{a}$, $a > 0$. Consider $(X - \mu)^2 > 0$ as x and $k^2\sigma^2 > 0$ as a in Markov's Inequality, we have $P((x-\mu)^2 \geq k^2\sigma^2) \geq \dfrac{E((x-\mu)^2)}{k^2\sigma^2} = \dfrac{\sigma^2}{k^2\sigma^2}$

$$= \frac{1}{k^2}$$

That is, $P(|X - \mu| < k\sigma) \geq 1 - \dfrac{1}{k^2}$. ■

■ **Example 4.11** Based on past data, the average daily number of tech support requests at a local call centre is 115 with a standard deviation of 10 calls.

a) What can be said about the fraction of days on which the number of calls received is between 100 and 130?
   - Distribution info: missing
   - We are given: $\mu = 115$, $r = 10$

   Let C be the random number of daily calls.

$$\begin{aligned}
P(100 \leq C \leq 130) &= P(-15 \leq C - 115 \leq 15) \\
&= P(-15 \leq C - \mu \leq 15) \\
&= P(|C - \mu| \leq 15) \\
&= P(|C - \mu| < 16) \\
&= P(|C - \mu| < 1.6\sigma) \\
&\geq 1 - \frac{1}{1.6} = 0.6094
\end{aligned}$$

$\therefore$ At least 60.94% of the time they will have between 100 to 130 calls a day.

b) What number of calls can they expect to receive at least 90% of the time?
We want to find the number of σ that correspond to an interval that has at least 90% chance of occurring.

$P(|X - \mu| < k\sigma) = 1 - \dfrac{1}{k^2} = 90\%$

That is, $1 - \dfrac{1}{k^2} = 0.90$

$$\begin{aligned}
0.1 &= \frac{1}{k} \\
k^2 &= 10 \\
k &= \sqrt{10}
\end{aligned}$$

For all distributions, more than 90% of the outcomes lie within $\sqrt{10} \approx 3.16$ standard derivation of its mean.

$(115 - k\sigma, 115 + k\sigma) = (115 - \sqrt{10} \cdot 10, 115 + \sqrt{10} \cdot 10) \approx (83.38, 146.62)$

$\therefore$ They can expect to receive $[83, 147]$ number of calls at least 90% of the time.

■

4.3 Common Discrete Distributions

4.3.1 Binomial Distribution

Definition 4.3.1 — Bernoulli Trials. A *Bernoulli trial* is a random experiment consisting of exactly one trial involving two possible outcomes, often called a *success* or a *failure*. Let X be the outcome of a Bernoulli trial where

$$\begin{aligned}
X &= 0 \text{ if the outcome is a failure} \\
X &= 1 \text{ if the outcome is a success}
\end{aligned}$$

We define p to be the probability of **success**, and $q = 1 - p$ to be the probability of **failure**. The *probability mass function* is then

$$f(x) = p^x \cdot (1-p)^{1-x}$$

■ **Example 4.12 — Bernoulli Trials.** Below are some examples of Bernoulli trials.

- Whether a randomly selected part is defective
- Whether there is an error in a line of code
- Whether a randomly selected individual is taller than 5’7”
- Whether a switch is in the on or off proposition
- Whether your lotto ticket is the winning number

■

■ **Example 4.13** Consider a multiple choice quiz 4 questions. A student selects an answer at random for each question, and each question is a Bernoulli experiment: the student either guesses correctly (1) or incorrectly (0). The sum of these four Bernoulli experiments will then be the random number of correct answers for a student that completes a similar quiz in this way. Our sample space is:

$$\Omega = \{✓✓✓✓, ✓✓✓✗, ✓✓✗✓, ✓✗✓✓, ✗✓✓✓, ✓✓✗✗, ✓✗✓✗, ✗✓✓✗, ✗✓✗✓, ✗✗✓✓, ✓✗✗✓,$$

$$✗✗✗✓, ✗✗✓✗, ✗✓✗✗, ✓✗✗✗, ✗✗✗✗\}$$

Suppose these MCQs have four options each, and each question only has one correct option. Find the following probabilities.

a) Guessing each question correctly.
$P(✓✓✓✓) = \frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{4^4}$

b) Guessing each question incorrectly.
$P(✗✗✗✗) = \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{3}{4} = \frac{3^4}{4^4}$

c) Guessing exactly 2 questions correctly.
$P(✓✓✗✗) = \frac{1}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{3}{4} = \frac{3^2}{4^4}$
$P(2 \text{ correct}) = {}_4C_2 \cdot \frac{3^2}{4^4}$

■

Often, we are interested in modeling the number of successes among multiple trials instead of the results of a single trial:

Definition 4.3.2 — Binomial Distribution. A *Binomial experiment* consists of n independent and identical Bernoulli trials. The probability of success, p, is fixed for each trial.

Let X be the random variable representing the number of successes among the n trials. Then X can be modeled by the binomial distribution with parameters n and p, denoted as $X \sim \text{Bin}(n, p)$. The binomial distribution has probability mass function:

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

If $X \sim \text{Bin}(n, p)$, we can show that $E(X) = np$ and $V(X) = np(1-p)$

■ **Example 4.14 — Binomial Experiments.** Below are some examples of Binomial experiments.

- The number of people who tried the dalgona candy challenge following ‘Squid Game’
- The number number of randomly selected students who started playing Animal Crossing in 2020
- The number of games won out of 7 independent games with the same opponent

■

■ **Example 4.15** While studying by a window, you find yourself noticing many cars at a nearby intersection that fail to fully come to a stop at the stop sign before passing through the intersection. Based on your months of data, you reliably calculate the probability of drivers failing to do a complete stop to be 70%. Assuming the stopping behaviour of each car is independent of all others, find the probability that among 20 randomly observed cars that...

a) Exactly 5 will come to a complete stop?
Let S be the number of cars that stop.
$S \sim Bin(n = 20, p = 0.3)$.

$$
\begin{aligned}
P(S=5) &= \binom{20}{5} \cdot 0.3^5 \cdot (1-0.3)^{20-5} \\
&= \frac{20!}{(20-5)! \cdot 5!} \cdot 0.3^5 \cdot 0.7^{15} \\
&\approx 0.1789
\end{aligned}
$$

b) At least 3 will come to a complete stop?

$$
\begin{aligned}
P(S \geq 3) &= \sum_{s=3}^{20} \binom{20}{s} \cdot 0.3^s \cdot (1-0.3)^{20-s} && \text{(direct)} \\
&= 1 - P(S < 3) && \text{(indirect)} \\
&= 1 - P(S \leq 2) \\
&= 1 - \left(\binom{20}{0} \cdot 0.7^{20} + \binom{20}{1} \cdot 0.3^1 \cdot 0.7^{19} + \binom{20}{2} \cdot 0.3^2 \cdot 0.7^{18} \right) \\
&\approx 0.9645
\end{aligned}
$$

c) At most 3 will come to a complete stop?

$$
\begin{aligned}
P(S \leq 3) &= P(S=0) + P(S=1) + P(S=2) + P(S=3) \\
&= \binom{20}{0} \cdot 0.7^{20} + \binom{20}{1} \cdot 0.3^1 \cdot 0.7^{19} + \binom{20}{2} \cdot 0.3^2 \cdot 0.7^{18} + \binom{20}{3} \cdot 0.3^3 \cdot 0.7^{17} \\
&\approx 0.1071
\end{aligned}
$$

■

■ **Example 4.16** A local hospital has several backup generators to support critical technologies in the event of a power outage or failure. Each backup generator is identical in make, and operate independently of others. Suppose each backup generator has a 20% chance of failing when used. How many generators should be installed so that the system has at least a 99.5% probability of functioning in the event of a power outage?

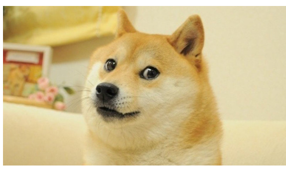
Let n be the number of generators (a fixed quantity).
Let G be the number of generators that functions.
$G \sim Bin(n, p = 0.8)$.
We want to find n such that $P(G \geq 1) \geq 99.5\%$.
That is, by the indirect method, we need

$$
\begin{aligned}
P(G=0) &\leq 0.005 \\
\binom{n}{0} \cdot 0.2^n &\leq 0.005 \\
0.2^n &\leq 0.005 \\
n &\geq 3.29
\end{aligned}
$$

$\therefore$ At least 4 generators should be installed. ■

4.3.2 Poisson Distribution

Consider modeling of the number of Shiba, D, spotted at a nearby park over any 1 day with a probability model. (How is this different from a Binomial model if it still models the number of 'successes'?)



This is **not** a binomial distribution, as trials are *discrete*, while time is *continuous*!

Let's try to formulate this problem so it resembles a Binomial model: first we will arbitrarily divide the 1 day into n equally-sized time interval with the following properties for any one interval:

- $P(D = 1) = p$
- $P(D = 0) = 1 - p$
- $P(D > 1) = 0$ (i.e. the event of Shiba sighting is "rare")

Let us also assume that each time interval behaves independently, and the average (mean) number of Shiba sightings per day is fixed and denoted by λ.

Based on the construction and assumptions, we have n independent trials with equal probabilities of "success" p. This can be modeled as a binomial distribution where $D \sim Bin(n, p)$, which has an expected values of $E(D) = np$.

Since the mean number of daily sightings is constant,

- $E(D) = np = \lambda$, and $p = \frac{\lambda}{n}$
- The number of time intervals is arbitrarily decided, neither n nor p are known.
- In order to ensure daily average $\lambda = np$ remains constant, as n increases, p must decrease so that np remains unchanged

We'll get more accurate probabilities of daily sightings in a day if we allow each time interval to shrink to 0 (not too different from using Riemann sums to approximate area under continuous curves!). Let's see how the binomial PMF behaves as $n \to \infty$ and $p \to 0$.

The resulting function will model the *probability of D number of Shiba sightings* over a *continuous period of 1 day*.

Let D be the number of Shiba sighted in "n" sub-intervals.

$D \sim Bin(n, p = \frac{\lambda}{n})$.

$E(D) = np = \lambda$.

$$
\begin{aligned}
P(D = d) &= \lim_{n\to\infty} \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d} \\
&= \lim_{n\to\infty} \frac{n!}{(n-d)! \cdot d!} \cdot \left(\frac{\lambda}{n}\right)^d \cdot \left(1 - \frac{\lambda}{n}\right)^{n-d} \\
&= \frac{\lambda^d}{d!} \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-d+1)\cancel{(n-d)!}}{\cancel{(n-d)!}} \cdot \frac{1}{n^d} \left(1 - \frac{\lambda}{n}\right)^{n-d} \\
&= \frac{\lambda^d}{d!} \lim_{n\to\infty} \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-d+1}{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-d} \\
&= \frac{\lambda^d}{d!} \lim_{n\to\infty} (1) \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{d-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{n-d} \\
&= \frac{\lambda^d}{d!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-d} \\
&= \frac{\lambda^d}{d!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} \cancel{\left(1 - \frac{\lambda}{n}\right)^{-d}}^{\,1} \\
&= \frac{\lambda^d}{d!} \cdot e^{-\lambda} \qquad \text{as } \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda} \\
&= \frac{\lambda^d e^{-\lambda}}{d!}
\end{aligned}
$$

Definition 4.3.3 — Poisson Distribution. A discrete random variable X denoting the number of (sometimes rare) events of interest in an interval, with the mean number of occurrences per unit interval denoted by λ, is *Poisson Distributed* if it has the probability mass function

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

X has expectation $E(X) = \lambda$ and variance $V(X) = \lambda$.

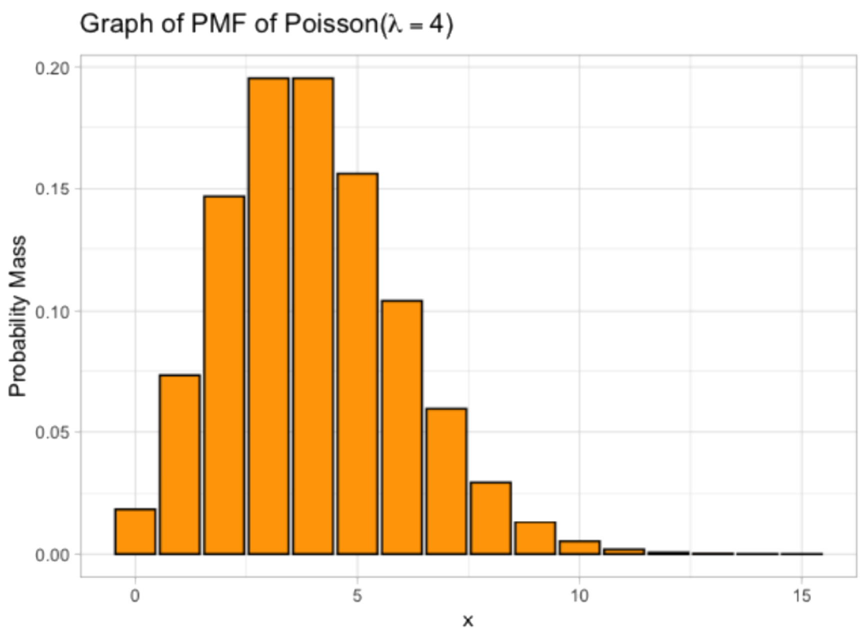
Note that λ is also called the *rare parameter* as it describes the average *rate of occurrance* of the event. λ should always be adjusted for the time interval being considered.

Poisson distribution is appropriate for discrete random variables that count the number of occurrences in a **continuous interval** where

- No more than one occurrence can occur simultaneously
- Non-overlapping intervals have occurrences that behave independently
- Expected number of occurrences in each fixed time interval is constant

As suggested in the construction of the Poisson distribution, this distribution can be used to approximate binomial probabilities where n is large and p is small. The approximation improves as n increases and/or p decreases. The advantage here is that Poisson distribution can be easier to compute since it doesn't involve the $\binom{n}{x}$ factor in its PMF.

What does a Poisson distribution look like? Generally unimodal and right-skewed, there is greater likelihood in observing values near but less than the mean $\lambda = 4$.



■ **Example 4.17** An area of a forest as on average 6 trees per 100 m^2.

a) Do the assumptions for a Poisson model seem appropriate for modeling tree distribution in a forest
   - Two simulations trees (two trees occupying the same spot) is 0
   - Could reasonably assume that the number of trees per area is independent

b) What is the probability of having at least 1 tree in a 100 m^2 area?
   Let T be the number of trees per 100 m^2.
   $T \sim Pois(\lambda = 6)$.

Using the indirect method, $P(T \geq 1) = 1 - P(T = 0)$

$$= 1 - \frac{6^0 e^{-6}}{0!}$$
$$\approx 0.9975$$

$\therefore$ There is 99.75% change of having at least 1 tree in a 100 m^2 area.

■

■ **Example 4.18** The number of car repairs R that arrive at a mechanic can be well modeled by a Poisson random variable with a mean of $3t$, where t denotes the time in hours of operating hours. The average profit per repair is given by $Y = 85R^2 - 60t$. Assuming that vehicle arrival occurs independently,

a) What is the probability that in a 10 hour workday, they will service between 28 to 31 cars?
We can reasonably model $R \sim Pois(\lambda = 3 \cdot 10 = 30)$.
Then, $P(28 \leq R \leq 31) = \sum_{r=28}^{31} P(R = r)$
$$= \frac{30^{28} e^{-30}}{28!} + \frac{30^{29} e^{-30}}{29!} + \frac{30^{30} e^{-30}}{30!} + \frac{30^{31} e^{-30}}{31!}$$
$$\approx 0.2858$$
$\therefore$ 28.58% of the time they will service between 28 to 31 cars.

b) What is the corresponding profit earned for the cars serviced in a)?
$Y = 85R^2 - 60 \cdot 10 = 86R^2 - 600$ with $R \in [28, 31]$.
That is, $y \in [85 \cdot 28^2 - 600, 85 \cdot 31^2 - 600] = [66040, 81085]$
$\therefore$ They are expected to earn from $\$66,040$ to $\$81,085$.

c) What is the expected profit for a typical 8 hour workday?
$R \sim Pois(\lambda = 3 \cdot 8 = 24)$
$E(Y) = E(85R^2 - 60 \cdot 8)$
$= E(85R^2 - 480)$
$= 85E(R^2) - 480$
Since $V(R) = E(R^2) - E(R)^2$, $E(R^2) = V(R) + E(R)^2 = \lambda + \lambda^2 = 24 + 24^2$.
Then, $E(Y) = 85(24 + 24^2) - 480$
$= \$50,520$
$\therefore$ The expected profit for a typcail 8 hour workday is $\$50,520$.

d) Suppose that you know that in an 8 hour workday, $E(R^4) = 416,472$. Determine the variance in profit. Can you determine what interval of profits they can expect to earn with at least 75% probability?
$R \sim Pois(24)$.
$E(Y) = 50,520$
$V(Y) = V(85R^2 - 480) = 85^2 \cdot V(R^2)$
$= 85^2 \cdot (E((R^2)^2) - E(R^2)^2)$
$= 85^2 \cdot (416,472 - (24 + 24^2)^2)$
$= \$^2 408,010,200$
$\sigma_Y = \sqrt{V(Y)} \approx \$20,199.26$
By Chebyshev, $P(|Y - \mu_Y| < k\sigma_Y) \geq 1 - \frac{1}{k^2}$, and we want $1 - \frac{1}{k^2} = \frac{3}{4}$. That is, $k = 2$.
$y \in (50520 - 2(20199.26), 50520 + 2(20199.26)) = (10121.48, 90918.52)$
$\therefore$ They are expected to earn within $(10121.48, 90918.52)$ with at least 75% probability.

■

Index

U

V