

Lecture 1 - 4th Sept 2024

Statistics: the science of understanding data and making decisions in the face of variability and uncertainty.

Probability: A branch of mathematics concerned with describing and modeling uncertain events.

Preliminaries:

- Experiment: the process of obtaining an observed result of some phenomenon.
- Trial: the performance of an experiment.
- Outcome: the result of a single trial (attempt) of an experiment.
- Event: one or more outcomes of an experiment.
- Probability: the measure of how likely an event is.

Sample Space: the set of ALL possible distinct outcomes in a random experiment, denoted by S .

↳ note: one and only one of the outcomes occurs in any single trial of the experiment.

↳ example:

- Roll a six-sided die: $S = \{1, 2, 3, 4, 5, 6\}$
- Flip a coin: $S = \{\text{heads, tails}\}$
- Waiting time for a bus: $S = \{t \in \mathbb{R}, 0 \leq t \leq 10\}$
- A sample space is finite if it consists of a

finite number of outcomes, say $\{a_1, a_2, \dots, a_n\}$

- A sample space can be a set of countable infinite outcomes, say $\{a_1, a_2, \dots\}$, where the outcomes can be put into a one-on-one correspondence with positive integers.

↳ Example: suppose our experiment consists of tossing a coin until it lands on heads.

↳ the possible outcomes in the sample space are:

H, TH, TTH, TTTH, TTTTH, ...

∴ the outcomes are countable but infinite.

Discrete Sample Space: when a sample space is either finite or countably infinite. Else, it's non-discrete.

↳ Example:

- Roll a six-sided die: discrete
- Flip a coin: discrete
- Waiting time for a bus: non-discrete

Event: a subset of sample space S.

↳ we say A is an event if $A \subset S$.

↳ note: A and S are both sets!

Example: consider tossing two coins.

a) what is the event of obtaining at least one head?

$$S = \{ HH, HT, TH, TT \}$$

$$\therefore A = \{ HH, HT, TH \}$$

b) What is the event of obtaining at least one head and at least one tail?

$$A = \{ HT, TH \}.$$

Elementary / Simple Event: an event that contains only one outcome of the experiment. Eg, $A = \{ \alpha, \beta \}$.

Compound Event: an event made up of two or more simple events. Eg: $A = \{ \alpha_1, \alpha_2 \}$.

Set Notations and Terminology:

- A set is a collection of elements. Eg, $A = \{ \text{apple, orange} \}$.
- The notation $\alpha \in A$ or $\alpha \notin A$ will mean that α is or is not an element of A .
- The empty / null set is \emptyset .
- Union: $A \cup B = \{ \alpha \mid \alpha \in A \text{ or } \alpha \in B \}$
- Intersection: $A \cap B = \{ \alpha \mid \alpha \in A \text{ and } \alpha \in B \}$
- The notation $B \subset A$ means that B is a subset of A .
- Complement: $A^c = \{ \alpha \mid \alpha \in S, \alpha \notin A \} = \bar{A}$
- Two sets A and B are disjoint if $A \cap B = \emptyset$.

- The set $A \cap B^c$ is "A but not B"
- The set $A^c \cap B^c$ is "neither A nor B"

- If A_1, \dots, A_n is a finite collection of sets,

$\hookrightarrow A_1 \cap \dots \cap A_n = \bigcap_{i=1}^n A_i$ is "in all A_i ;"

$\hookrightarrow A_1 \cup \dots \cup A_n = \bigcup_{i=1}^n A_i$ is "in at least one A_i ;"

✓ very similar to disjoint!

Mutually Exclusiveness: when two events A and B if $A \cap B = \emptyset$. If events are exclusive, they have no outcomes in common.

\hookrightarrow Example: tossing two coins. A is the event "at least one head", and B is the event "both tails".

$$\hookrightarrow S = \{HH, HT, TH, TT\}.$$

$$A = \{HH, HT, TH\}, B = \{TT\}.$$

$$A \cap B = \{HH, HT, TH\} \cap \{TT\} = \emptyset !$$

\therefore , A and B are mutually exclusive.

\hookrightarrow Events A_1, A_2, A_3, \dots are said to be mutually exclusive if they are pairwise mutually exclusive. That is,

$$A_i \cap A_j = \emptyset, \text{ wherever } i \neq j.$$

Probability Modeling : assigning a probability $P(A)$ to each event A . This probability measures how likely it is that A will happen when the experiment is conducted.

- ↳ think of $P(A)$ as a (set) function, whose domain is a collection of sets (events), and the range of which is a subset of real numbers.
- ↳ not all set functions are appropriate for assigning probabilities to events.

• Let $S = \{\alpha_1, \alpha_2, \alpha_3, \dots\}$ be a discrete sample space. We assign to each elementary event $A_i = \{\alpha_i\}$ for $i = 1, 2, \dots$, a number $P(A_i) = P(\{\alpha_i\})$, such that

$$\bullet 0 \leq P(A_i) \text{ and } \sum_{\text{all } i} P(A_i) = 1$$

• We call $P(A)$ the probability of A , and call the set of probabilities $\{P(A_i), i=1, 2, \dots\}$ the probability distribution on S .

↳ General Properties :

- The null event (empty set) has a probability of 0. ∴ $P(\emptyset) = 0$.
- If A and B are two mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$
- If A_1, A_2, \dots, A_k is a finite collection of pairwise mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \dots \cup A_R) = P(A_1) + P(A_2) + \dots + P(A_R)$$

$$\cdot P(\bar{A}) = 1 - P(A)$$

↳ Example: Suppose a six-sided die is rolled once.
If A is the event that an even number is obtained.
What is $P(A)$?

- $S: \{1, 2, 3, 4, 5, 6\}$. Since each number is equally likely, $P(\{i\}) = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$.

$$\begin{aligned} A &= \{2, 4, 6\}. \rightarrow P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}. \end{aligned}$$

↳ Example: Toss a coin twice. What's the probability of getting exactly one head?

- $S: \{HH, HT, TH, TT\}$. Each option is uniformly likely.
- $A = \{HT, TH\} \rightarrow P(A) = P(\{HT\}) + P(\{TH\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Random Selection: when choosing an object from a finite set where each object has an equal chance of being selected.

Uniform Probability Model: consider a discrete finite sample space $S = \{a_1, a_2, \dots, a_n\}$, where each

simple event has a probability of $\frac{1}{N}$. This is a uniform distribution over the set $\{\alpha_1, \alpha_2, \dots, \alpha_N\}$.

For a compound event A, we can calculate $P(A)$ by:

$$P(A) = \frac{n(A)}{N},$$

where $n(A)$ is the number of outcomes in A.

The formula $P(A) = \frac{n(A)}{N}$ is classical probability.

(trivial) example: draw one card at random from a well-shuffled deck without jokers. What is the probability that the card is a spade?

$$S = \{A\heartsuit, 2\heartsuit, \dots\}.$$

$$A_s = \{A\spadesuit, 2\spadesuit, \dots, K\spadesuit\}.$$

$$P(A_s) = \frac{n(A_s)}{N} = \frac{13}{52} = \frac{1}{4}.$$

Example: suppose that P is a probability, and A and B are events such that A and B are mutually exclusive, and $P(\bar{A}) = 0.8$ and $P(\bar{B}) = 0.7$. Is each statement true or false?

A) A and B are each non-empty

Since $P(\bar{A}) = 0.8$ and $P(\bar{B}) = 0.7$, we know that

$$P(A) = 0.2 \text{ and } P(B) = 0.3.$$

we also know that $P(\emptyset) = 0$. Since $0.2 \neq 0 \neq 0.3$, we know A and B cannot be empty.
 \therefore the statement is true.

B) $P(A \cap B) = 0$

Since we're told that A and B are mutually exclusive, we know that $A \cap B = \emptyset$.

we also know that $P(\emptyset) = 0$, so the statement is true!

C) $P(A \cup B) = 0.4$

we know that $\boxed{P(A \cup B)} = P(A) + P(B)$

"the probability of
A or B"

$$P(A) + P(B) = 0.2 + 0.3 = 0.5 \neq 0.4.$$

\therefore the statement is false!

Example: suppose two six-sided die are rolled where the outcomes are of the form (die 1, die 2). If A is the event that the first die is even, and B is the event that the sum of both rolls is 6, compute the event $C = \bar{A} \cap B$

$$\hookrightarrow A = \{(2,1), (2,2), \dots, (4,1), (4,2), \dots, (6,1), (6,2), \dots, (6,6)\}.$$

$$B = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$$

\bar{A} means the event that A does not happen, so in other words, \bar{A} is the event that the first die is odd.

$$\therefore \bar{A} = \{(1,1), \dots, (1,6), (3,1), \dots, (3,6), (5,1), \dots, (5,6)\}.$$

$\bar{A} \cap B$ is A AND B.

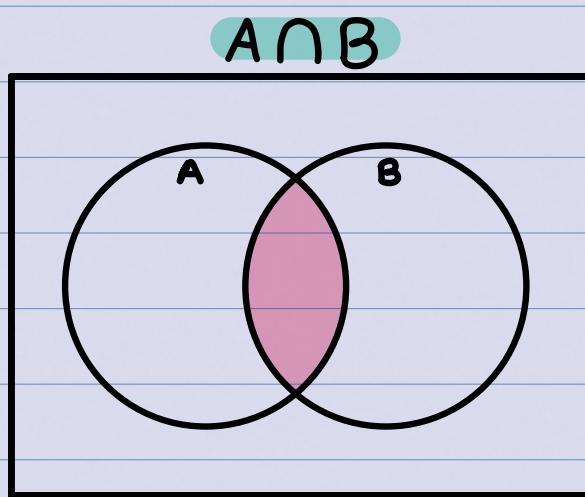
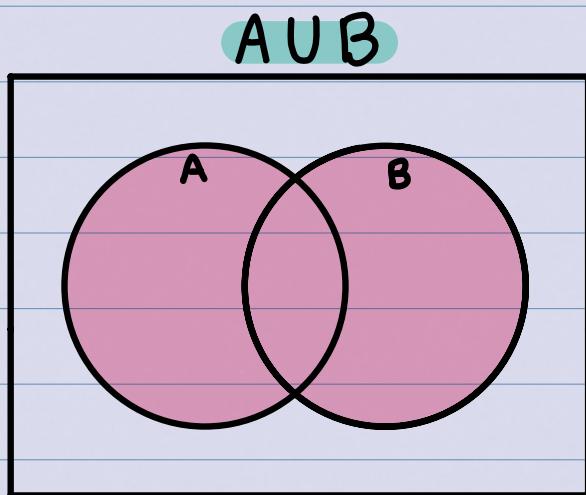
$$\hookrightarrow \therefore C = \{(1,5), (3,3), (5,1)\}.$$

Venn Diagram: a tool to illustrate the relationship among sets (or events).

↳ made up of:

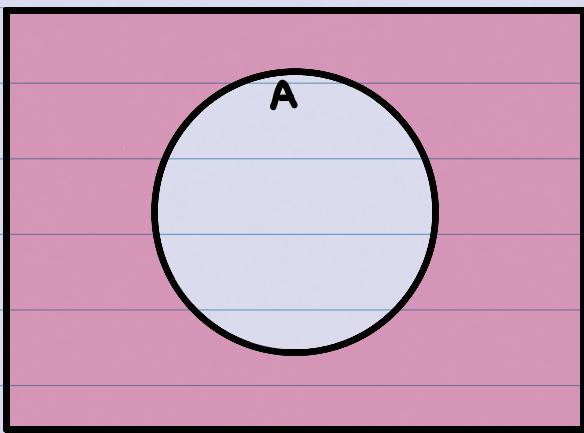
- a rectangle to represent the sample space S.
- Circles within the rectangle to illustrate the events.

The Union and Intersection of Two Events



The Compliment of an Event:

\bar{A}



"S but not A"

De Morgan's Laws:

$$\cdot \overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\cdot \overline{A \cap B} = \bar{A} \cup \bar{B}$$

Probability Rules

- Probability of the Complement of an Event
 - $P(A) = 1 - P(\bar{A})$
- Proof: $A \cup \bar{A} = S$, so $P(A \cup \bar{A}) = P(A) + P(\bar{A})$
since $P(A \cup \bar{A}) = P(S) = 1$, $1 = P(A) + P(\bar{A})$
bc mutually exclusive!

Example: an experiment involves tossing a coin four times. The event A is "at least one head". Find the probability of A, ie, find $P(A)$.

- the only case where there is NOT at least one head is $\bar{A} = \{\text{TTTT}\}$.

flipping a coin 4 times would result in 2^4 possible

outcomes, and \bar{A} is only 1 of these.

$$\therefore P(\bar{A}) = \frac{1}{16}, \text{ so } P(A) = 1 - P(\bar{A}) = 1 - \frac{1}{16} = \frac{15}{16}.$$

probability of complement!

Example: two ordinary dice are rolled. Find the probability that at least one lands on a 6.

$$S = \{(1,1), \dots, (1,6), (2,1), \dots, (2,6), \dots, (6,1), \dots, (6,6)\}.$$

A = the event of at least one six.

$\therefore \bar{A}$ is the event that neither die is a six.

classical probability of each die to NOT be a six: $\frac{n(x)}{N}$.

where $n(x)$ is the number of outcomes that satisfy the roll not being a 6 - ie; $\{1, 2, 3, 4, 5\} \Rightarrow 5$, and N is the total number of outcomes in the set - ie; $\{1, 2, 3, 4, 5, 6\} \Rightarrow 6$.

$$\therefore \frac{n(x)}{N} = \frac{5}{6} !$$

since both dice are independent, the probability of both NOT being a six (aka \bar{A}) is $\frac{5}{6} \cdot \frac{5}{6}$.

$$\therefore P(\bar{A}) = \frac{25}{36}, \text{ so } P(A) = 1 - \frac{25}{36} = \frac{11}{36} !$$

Rules for Union:

• Union of two events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example: At a university, 70% of the students own a car, 60% of students live on campus, and 50% both live on campus and own a car. If a student is chosen at random, what is the probability that the student does not own a car and does not live on campus?

- Let A be the event that a student has a car.
- Let B be the event that a student lives on campus.

we want: $P(\bar{A} \cap \bar{B})$.

↳ we know that $\bar{A} \cap \bar{B} = \overline{A \cup B}$.

$$\hookrightarrow P(\bar{A} \cap \bar{B}) = 1 - P(\overline{A \cup B}) = 1 - P(A \cup B)$$

from the union of two events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$\begin{aligned}\therefore P(\bar{A} \cap \bar{B}) &= 1 - (P(A) + P(B) - P(A \cap B)) \\ &= 1 + P(A \cap B) - P(A) - P(B) \\ &= 1 + 0.5 - 0.7 - 0.6\end{aligned}$$

$$\hookrightarrow \therefore P(\bar{A} \cap \bar{B}) = 0.2 = 20\%.$$

Union of Three Events: for any 3 events A, B, and C:



$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Conditional Probability: determining the probability of some event A, while knowing that some related event B has occurred.

Example: consider that a fair die has been rolled and you are asked to give the probability that it was a 5.

$$\hookrightarrow A = \{5\}, S = \{1, 2, 3, 4, 5, 6\}.$$

$$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{1}{6}.$$

However, if you're told that the dice has landed on an odd number, the odds change:

$$\hookrightarrow A = \{5\}, S = \{1, 3, 5\}.$$

$$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{1}{3}.$$

Example: a box contains 100 microchips, some produced by Factory 1 and the rest by Factory 2. Among these microchips, some are defective, while others are good. Let A be the event of "obtaining a defective microchip" and B be the event of "the microchip was produced in Factory 2."

the number of microchips in each category:

	B	\bar{B}	totals
A	15	5	20
\bar{A}	45	35	80
totals	60	40	100

a) Find the probability of obtaining a defective microchip.

the total of A is 20, and there are 100 total microchips.

$$\hookrightarrow P(A) = \frac{n(A)}{N} = \frac{20}{100} = \frac{1}{5} \quad (20\%)$$

b) Given that a microchip is from factory 1, find the probability that it is defective.

the sample space S shrinks from all 100 microchips to only the 60 made by factory 1 (event B).

$$\hookrightarrow \therefore P(A) = \frac{n(A)}{N} = \frac{15}{60} = \frac{1}{4} \quad (25\%)$$

Conditional Probability mathematical definition:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (P(B) > 0)$$

↪ "the probability of A given B"

$$\hookrightarrow 0 \leq P(A|B) \leq 1$$

- $P(A|B) = 1 - P(\bar{A}|B)$

Pretty much the same as
the regular union of two events

- $P(A_1 \cup A_2 | B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2 | B)$

Multiplication Theorem of Probability: a way to compute the joint occurrence of A and B:

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

Law of Total Probability: if B_1, B_2, \dots, B_k is a collection of mutually exclusive events and $B_1 \cup B_2 \cup \dots \cup B_k = S$, then for any event A,

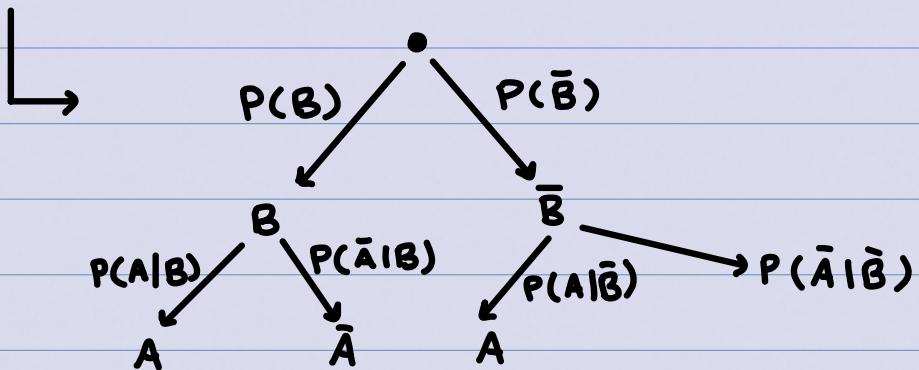
$$P(A) = \sum_{i=1}^k P(B_i) \cdot P(A|B_i)$$

Note: for a special case $k=2$:

only for
 $k=2$!!

$$P(A) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B})$$

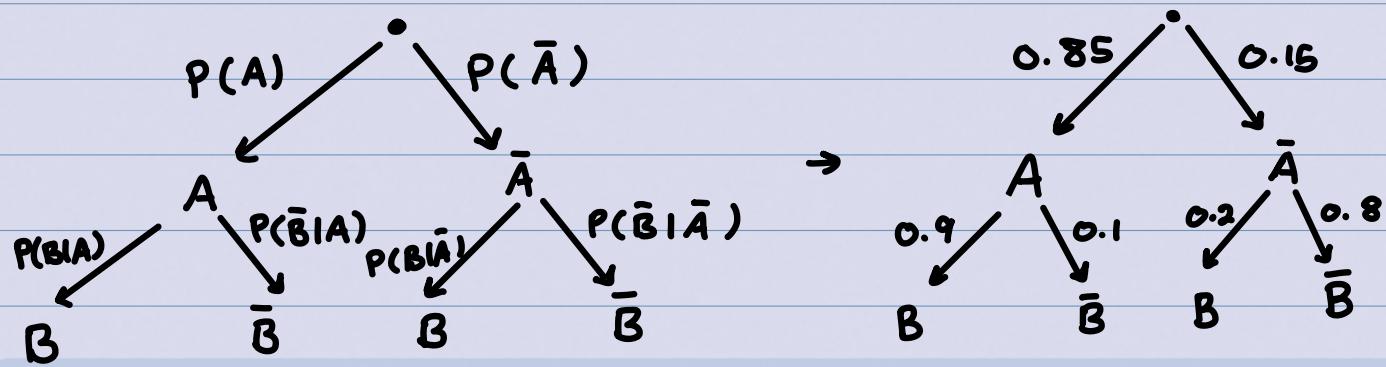
for event A to occur, it must happen with either B or \bar{B} . \therefore only $A \cap B$ or $A \cap \bar{B}$ can occur.



Example: without water, a plant will die with a

probability of 0.8, and with water, it will die with a probability of 0.1. I'll remember to water the plant with a probability of 0.85. Represent this info using a tree diagram.

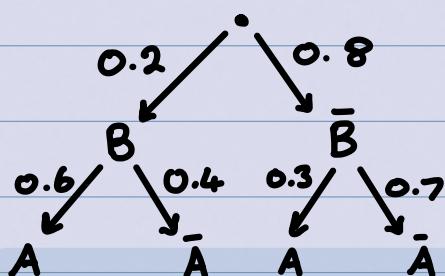
Let A be the event that I remember to water the plant, and B be the event that the plant survives.



Example: in a typical year, 20% of the days have a high temperature greater than 22°C . 40% of these days, there is no rain. During the rest of the year, when the temperature is $\leq 22^{\circ}\text{C}$, 70% of the days have no rain.

a) represent this information in a tree diagram.

let A be the event that it rains, and let B be the event that the temperature peaks above 22°C .



b) solve for the proportion of days in the year which have rain and a temperature $\leq 22^{\circ}\text{C}$.

↳ aka, find $P(A \cap \bar{B})$

$$\text{we know } P(A|\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{P(A \cap \bar{B})}{1 - P(B)}$$

$$\begin{aligned} \therefore P(A \cap \bar{B}) &= P(A|\bar{B}) \cdot (1 - P(B)) \\ &= 0.3 \cdot (1 - 0.2) = 0.3 \cdot 0.8 = 0.24 \\ &= 24\% . \end{aligned}$$

Baye's Rule: $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$,

and more generally, $P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^k P(B_i)P(A|B_i)}$ ↗ for each $j=1, 2, \dots, k$.

Example: revisiting the plant problem from before, if the plant is alive, what's the probability that I remembered to water it?

↗ same as previous time

↳ let A be the event that I remember to water the plant, and B be the event that the plant survives.

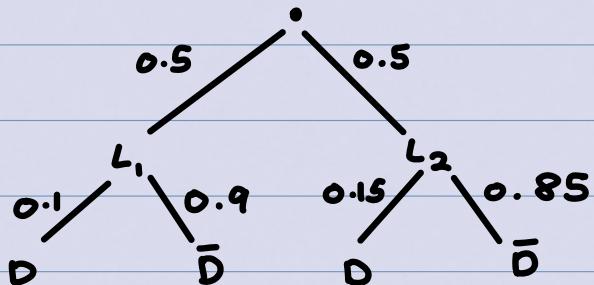
$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} = \frac{(0.9)(0.85)}{(0.9)(0.85) + (0.2)(0.15)} \\ &= 0.9623 = 96.23\% ! \end{aligned}$$

Example: electric motors from two assembly lines are

pooled in a common stockroom, with equal numbers of motors from each line. Motors are periodically sampled for testing. It's known that 10% of the motors from line 1 are defective, while 15% of the motors from line two are defective.

a) if a motor is randomly selected, what is the probability that it is defective?

Let D be the event that a motor is defective.



$$\begin{aligned} \rightarrow P(D) &= P(D \cap L_1) + P(D \cap L_2) \\ &= 0.5 \cdot 0.1 + 0.5 \cdot 0.15 \\ &= 0.125 \end{aligned}$$

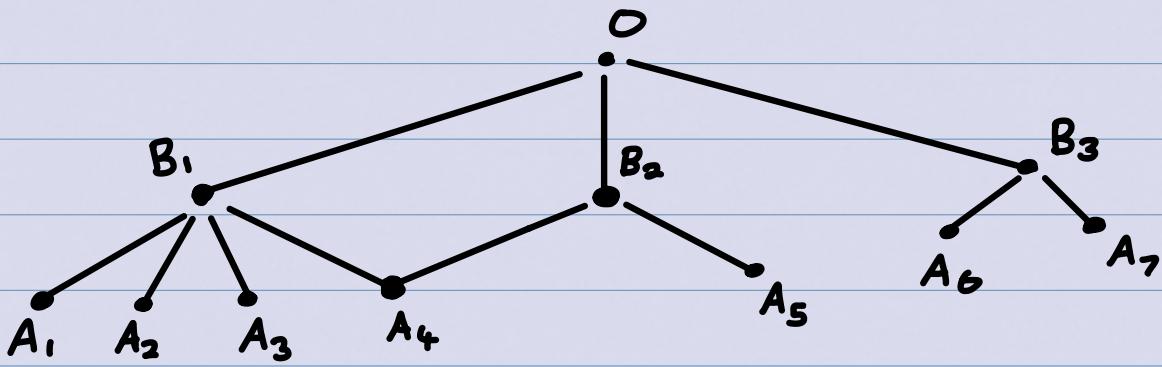
↳ 12.5% chance.

b) if the randomly selected motor is defective, what is the probability that it came from line 1?

$$P(L_1 | D) = \frac{P(D | L_1) P(L_1)}{P(D)} = \frac{0.1 \cdot 0.5}{0.125} = 0.4$$

= 40% chance.

Example: a man starts at point O on this map. He chooses a path at random and follows it to point B₁, B₂, or B₃. From that point, he chooses a new path at random and follows it to one of the points A, ... A₇.



a) what is the probability that the man arrives at point A_4 ?

$$P(A_4) = P(A_4 | B_1) + P(A_4 | B_2) = \frac{1}{4} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12} + \frac{1}{6}$$

$$= \frac{1}{4} = 0.25$$

$= 25\%$!

b) suppose that the man arrives at point A_4 . What is the probability that he passed through B_1 ?

$$P(B_1 | A_4) = \frac{P(A_4 | B_1)}{P(A_4)} = \frac{\frac{1}{4} \cdot \frac{1}{3}}{\frac{1}{4}} = \frac{\frac{1}{12}}{\frac{1}{4}} = \frac{4}{12} = \frac{1}{3}$$

\therefore , there was an 33.33% chance he came from B_1 .

Independent Event: events A and B are called independent if and only if $P(A \cap B) = P(A)P(B)$. Otherwise, A and B are called dependent events.

↳ an equivalent formulation can be given in terms of conditional probability :

→ If A and B are events such that $P(A) > 0$ and $P(B) > 0$, then A and B are called independent if and only if either of the following hold:

$$P(A|B) = P(A), \text{ or } P(B|A) = P(B)$$

Note: independence \neq mutually exclusive!

unless $P(A) = 0$!

General rule: $P(A \cap B) = P(B|A) \cdot P(A)$

Independence of Compliments: Suppose events A and B are independent. Then the following events are also independent:

- \bar{A} and \bar{B}
- \bar{A} and B
- A and \bar{B}

Independence of n events: the k events A_1, A_2, \dots, A_k are said to be independent, or mutually independent, if and only if:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

forall possible distinct i_1, i_2, \dots, i_k chosen from 1, 2, ..., n.

→ Note: for a sequence of events to be independent, all of their subsets must also be independent.

Example: suppose a fair die is rolled twice. Let A be the event that the first roll is a 6, and let B be the event

that the second roll is a 6. Show that A and B are independent events.

↳ we want to show that $P(A \cap B) = P(A) \cdot P(B)$

$$P(\{6,6\}) = P(A \cap B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

$$P(A) = \frac{n(A)}{N} = \frac{1}{6} = \frac{n(B)}{N} = P(B). \therefore P(A) = P(B) = \frac{1}{6}.$$

$$P(A) \cdot P(B) = \frac{1}{36} = P(A \cap B).$$

↳ ∴, the events are independent.

Counting Techniques

often, manually counting many outcomes manually can be very time-consuming and tedious.

1) Addition Rule : "OR" interpreted as addition

↳ Eg: suppose job 1 can be done in p ways and job 2 can be done in q ways. ∴, we can do either job 1 or job 2 in $p+q$ ways.

2) Multiplication Rule: "AND" interpreted as multiplication.

↳ Eg: suppose job 1 can be done in p ways and for each of these ways, we can do job 2 in q ways. ∴, we can do both Job 1 and Job 2 in $p \times q$ ways.

Example: how many are there to answer a 20

question true/false test?

$$Q_1 \cdot Q_2 \cdot \dots \cdot Q_{20} = 2^{20} \text{ ways.}$$

Sampling with Replacement: what we get on the first selection does not affect what we get on second selection.

↳ Eg: every time an object is selected, it's put back into the pool of possible objects.

Sampling without Replacement: what we get on the first selection will affect what we get on the second selection.

↳ Eg: once an object is selected, it stays out of the pool of possible objects.

Example: A bag contains 3 blue marbles and 5 red marbles. If you pick two marbles from the bag,

a) what is the probability of picking 2 blue marbles if the selection is with replacement?

let A be the event of picking two marbles.

↳ $P(A) = P(\text{blue on first pick}) + P(\text{blue on second pick})$ ↗ using independence property!

$$\hookrightarrow P(A) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64} = 14.06\%$$

b) what about without replacement?

$$P(A) = \frac{3}{8} \cdot \frac{2}{7} = \frac{6}{56} = \frac{3}{28} = 10.71\%$$

Example: a personal identification number (PIN) of length 4 is formed by randomly selecting 4 digits from the set of digits $\{0, 1, 2, \dots, 9\}$. If selection is done with replacement, find the probability that:

a) The PIN is even:

for the PIN to be even, the last digit must be a multiple of 2. $\therefore A = \{0, 2, 4, 6, 8\}$.

$$\therefore P(A) = \frac{n(A)}{N} = \frac{5}{10} = \frac{1}{2} \rightarrow 50\%$$

b) the PIN contains at least one 0?

$$\therefore P(\text{at least one } 0) = 1 - P(\text{no } 0s \text{ in PIN})$$

\therefore there are 9 ways to fill each of the digits, as we want the probability of no zeros.

$$\therefore P(\text{no } 0s) = 9 \cdot 9 \cdot 9 \cdot 9 = 9^4 \quad \begin{matrix} \xrightarrow{\text{multiplication counting rule}} \\ \hookrightarrow \text{not } 0 \text{ AND not } 0 \text{ AND...} \end{matrix}$$

total sample space is 10^4 (including 0s).

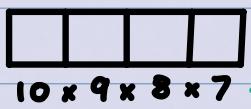
$$\therefore P(\text{at least one zero}) = 1 - \frac{n(\text{no } 0s)}{n(S)} = 1 - \frac{9^4}{10^4}$$

c) Redo a and b but without replacement:

i) even PIN:

A: event of getting an even number. $\{0, 2, 4, 6, 8\}$.

S: the sample space



we have less options
for each digit, bc
no replacement!

A: $n(A) = 9 \times 8 \times 7 \times 5$
 $9 \times 8 \times 7 \times 5 \rightarrow \text{bc } \text{len}(\{0, 2, 4, 6, 8\}) = 5$.

$$\hookrightarrow P(A) = \frac{9 \times 8 \times 7 \times 5}{10 \times 9 \times 8 \times 7} = \frac{5}{10} = \frac{1}{2} = 50\%$$

ii) at least one zero:

\curvearrowleft no 0 in PIN

$$P(B) = 1 - P(\bar{B}) = 1 - \frac{9 \times 8 \times 7 \times 6}{10 \times 9 \times 8 \times 7} = 1 - \frac{6}{10} = \frac{2}{5} = 40\%$$

Example: a fair die is tossed 3 times. What is the probability that exactly one of the tosses produces a number greater than 4?

let A be the event that the first toss is greater than 4, and the second and third tosses are ≤ 4 .

$$\hookrightarrow \{5, 6\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4\}.$$

"AND"
 $\therefore n(A) = 2 \times 4 \times 4 = 32$.

let B be the event where the 2nd toss is > 4 , and let C be the event where it's the 3rd toss. $\rightarrow n(A) = n(B) = n(C)$

\therefore total number of ways to get exactly one number greater than 4 is $3(2 \times 4 \times 4) = 96$.

$$\therefore P(\text{exactly one toss} > 4) = \frac{96}{6^3} = 44.44\%.$$

Counting Permutations:

- Permutation: an ordered arrangement of a set of objects.

Example: how many different ordered arrangements of the letters a, b, and c are possible?

↳ abc, acb, bac, bca, cab, cba.

∴ 6 arrangements!

This will very quickly become very tedious and slow.

- The number of permutations of n distinguishable objects is:

$$n \times (n-1) \times \dots \times 1 = n!$$

- Given n distinct objects, a permutation of length r is an ordered subset of r objects.

- The number of permutations of length r taken from n objects is denoted as $n^{(r)}$, where:

$$n^{(r)} = n \times (n-1) \times \dots \times (n-r+1) = \frac{n!}{(n-r)!}$$

↑ "n to r factors"

Example: five separate awards are to be presented to some students from a class of size 30. How many different outcomes are possible if

a) a student can receive any number of awards?

$$\frac{30}{1^{\text{st}}} \times \frac{30}{2^{\text{nd}}} \times \frac{30}{3^{\text{rd}}} \times \frac{30}{4^{\text{th}}} \times \frac{30}{5^{\text{th}}} = 30^5.$$

$$\therefore n(A) = 30^5$$

b) each student can receive, at most, one award?

$$\frac{30}{1^{\text{st}}} \times \frac{29}{2^{\text{nd}}} \times \frac{28}{3^{\text{rd}}} \times \frac{27}{4^{\text{th}}} \times \frac{26}{5^{\text{th}}} = 30 \cdot 29 \cdot 28 \cdot 27 \cdot 26$$

$$\hookrightarrow n(A) = 30^{(5)} = \frac{30!}{25!}$$

Sometimes the outcomes in the sample space are subsets of a fixed size:

Example: suppose we have three books: b_1 , b_2 , and b_3 . We choose two of the books to read. In how many ways can the two books be read if:

a) the order that the books are read in matters:

(b_1, b_2) , (b_2, b_1)

(b_1, b_3) , (b_3, b_1) → 6 ways!

(b_2, b_3) , (b_3, b_2)

b) the order that the books are read in does not matter?

$$\{b_1, b_2\}$$

$$\{b_1, b_3\}$$

$$\{b_2, b_3\}$$

→ 3 ways!

Given n distinct objects, a combination of size r is an unordered subset of r objects chosen from n distinct objects.

↳ the number of combinations (or subsets) of size r chosen from n distinct objects is given by:

$$\binom{n}{r} = \frac{n^{(r)}}{r!} = \frac{n!}{(n-r)r!}$$

↳ "n choose r"

Example: we randomly select a subset of 3 digits from $\{0, 1, \dots, 9\}$. All the digits in each outcome are unique, but the order of the elements in a subset is not relevant.

Find the probability that:

a) All the digits in the selected subset are even:

$$S = \{\{0, 1, 2\}, \{0, 1, 3\}, \dots, \{7, 8, 9\}\}$$

$$n(S) = \binom{10}{3}.$$

↗ length of S

the set of even digits is $\{0, 2, 4, 6, 8\}$.

let A be the event that all 3 digits are even.

$$\hookrightarrow n(A) = \binom{5}{3}$$

$$\hookrightarrow P(A) = \frac{n(A)}{n(S)} = \frac{\binom{5}{3}}{\binom{10}{3}} = \frac{1}{12} = 8.3\dot{3}\%$$

b) At least one of the digits in the selected subset is less than or equal to 5.

$$P(\text{at least one digit } \leq 5)$$

$$= 1 - P(\text{no digit } \leq 5)$$

$$= 1 - P(\text{all digits } > 5)$$

the set of digits > 5 : $\{6, 7, 8, 9\}$.

$$= 1 - \frac{\binom{4}{3}}{\binom{10}{3}} = \frac{29}{30} = 96.6\dot{6}\%$$

Example: there are 30 geese of which 6 were tagged. Later, 5 of the geese are randomly captured.

a) How many samples of 5 are possible? $\binom{30}{5}$

b) How many samples of 5, which include 2 of the tagged geese, are possible?

$$\hookrightarrow \binom{6}{2} \cdot \binom{24}{3}$$

c) if the five captured geese represent a simple random sample drawn from the 30

geese, find the probability that:

i) two of the 5 captured geese are tagged.

$$P(A) = \frac{\binom{6}{2} \binom{24}{3}}{\binom{30}{5}} = 0.2130 = 21.3\%$$

ii) none of the 5 captured geese are tagged.

$$P(B) = \frac{\binom{24}{5}}{\binom{30}{5}} = 0.2983 = 29.83\%$$

→ Key Properties of Permutations and Combinations:

1) $n^{(r)} = n \times (n-1)^{(r-1)}$ for $r \geq 1$.

2) $\binom{n}{r} = \frac{n^{(r)}}{r!}$

3) Symmetry Property: $\binom{n}{r} = \binom{n}{n-r}$ for $r \geq 0$.

4) since $0! = 1$, $\binom{n}{0} = \binom{n}{n} = 1$.

5) $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$

6) Binomial Theorem:

$$(1+x)^n = \binom{0}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n \text{ for } x \in \mathbb{R}.$$

Summary of Counting Techniques:

• Addition rule : OR (+)

• Multiplication rule : AND (×)

- Factorial: $n!$ is the number of arrangements of n distinct objects where order matters.
- Permutation: $n^{(k)}$ is the number of ways to choose k objects from n distinct objects where order matters.
- Combination: $\binom{n}{k}$ is the number of ways to choose k objects from n objects where order does NOT matter.

Example: a person has 10 friends, of whom 6 will be invited to a party. How many choices are there if:

a) 2 of the friends will only attend together?

$$\hookrightarrow \binom{8}{4} \cdot \binom{2}{2}$$

b) 2 of the friends will not attend if the other is also attending?

$$\hookrightarrow \binom{8}{6} \cdot \binom{2}{0} + \binom{8}{5} \cdot \binom{2}{1}$$

OR (+)! \rightsquigarrow only one or the other can occur.

Example: 13 cards are picked at random from a standard deck of 52 cards without replacement. Find the probability that we pick:

a) at least one ace:

$\hookrightarrow A =$ the event we pick at least one Ace, so

$\bar{A} =$ the event we pick no aces

$$\hookrightarrow P(A) = 1 - P(\bar{A}) = 1 - \frac{\binom{48}{13}}{\binom{52}{13}}$$

b) 6 spades, 4 hearts, 2 diamonds, and one club:

$$\hookrightarrow \frac{\binom{13}{6} \cdot \binom{13}{4} \cdot \binom{13}{2} \cdot \binom{13}{1}}{\binom{52}{13}}$$

Example: consider drawing 3 numbers at random with replacement from the digits $\{0, 1, \dots, 9\}$. What is the probability that there is a repeated number among the 3?

$$S = 10 \cdot 10 \cdot 10 = 10^3$$

A = event of repeated number

\bar{A} = event of no repeated number

$\hookrightarrow 10 \cdot 9 \cdot 8$ options

$$\therefore P(A) = 1 - P(\bar{A}) = 1 - \frac{n(\bar{A})}{n(S)} = 1 - \frac{\frac{10^3}{10^3}}{10^3} = \frac{7}{25}.$$

Example: Melissa selects 7 numbers between 1 and 50, and then a computer does the same. What is the probability that at least 5 of Melissa's numbers match the computer's?

$\hookrightarrow S = \binom{50}{7}$, and Melissa must match 5, 6, or 7 of the numbers: $\binom{7}{5} \binom{43}{2} + \binom{7}{6} \binom{43}{1} + \binom{7}{7} \binom{43}{0}$

$$\therefore P = \frac{\binom{7}{5} \binom{43}{2} + \binom{7}{6} \binom{43}{1} + \binom{7}{7} \binom{43}{0}}{\binom{50}{7}}$$

Example: a box contains 4 coins - 3 fair coins and 1 biased coin which lands on heads 80% of the time. A coin is randomly picked and tossed 6 times. It lands on heads 5 times. What is the probability the coin is fair?

let H_5 be the event of getting 5 heads in 6 tosses.
 let F be the event the coin is fair, and let B
 be the event the coin is biased.

$$P(F | H_5) = \frac{P(F \cap H_5)}{P(H_5)} = \frac{P(H_5 | F) P(F)}{P(H_5 | F) P(F) + P(H_5 | B) P(B)}$$

$$= \frac{6 \times 0.5^5 \times 0.5 \times \frac{3}{4}}{6 \times 0.5^5 \times 0.5 \times \frac{3}{4} + 6 \times 0.8^5 \times 0.2 \times \frac{1}{4}} = 0.417$$

Example: there are 4 passengers on a 5-floor elevator. What is the probability that:

a) the passengers all get off on different floors?

↳ 5 people getting off on 4 different floors
 is $5^{(4)}$, out of total possibilities $S = 5^4$.

$$\therefore \frac{5^{(4)}}{5^4} = \frac{24}{125} = 19.2\%$$

b) 2 passengers get off on floor two and 2
 get off on floor 3?

$$\frac{\binom{4}{2} \binom{2}{2}}{5^4}$$

c) 2 passengers get off one a floor and 2
 passengers get off on a different floor?

$$\frac{\binom{4}{2} \binom{2}{2} \times \binom{5}{2}}{5^4}$$

Random Variable: a function, denoted by X , that assigns a real number $x = X(\alpha)$ to each possible outcome α in a sample space S .

aka, we say X is a random variable if $X: S \rightarrow \mathbb{R}$.

- Random variables are denoted by capital letters such as X , Y , and Z .
- Their possible values/realizations are denoted by lowercase letters such as x , y , and z .

Example: Consider an experiment where a fair coin is tossed 3 times.

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

For each outcome α in S , the value of the function $X(\alpha)$ is the number of heads corresponding to α .

↳ Eg: if $\alpha = \text{THH}$, then $X(\alpha) = 2$.

∴ Value of X	Definition of the event
$X = 0$	$\{\text{TTT}\}$
$X = 1$	$\{\text{TTH}, \text{THT}, \text{HTT}\}$
$X = 2$	$\{\text{HHT}, \text{HTH}, \text{THH}\}$
$X = 3$	$\{\text{HHH}\}$

↳ basically, if in the naturals

Discrete Variable: if a random variable's range is a discrete subset of \mathbb{R} (a finite set $\{x_1, \dots, x_n\}$ or a countably infinite set $\{x_1, x_2, \dots\}$).

eg: number of coin flips, number of people, ...

eg: monthly rainfall, time, ...

Continuous Variable: if a random variable's range is an interval that is a subset of \mathbb{R} .

also referred to as discrete probability density function (discrete pdf)

Probability Mass Function (pmf / pf): the probability function of a discrete random variable X is:

$$\rightarrow P\{\omega \in S : X(\omega) = x\}$$

$$f_x(x) = P(X=x)$$

Probability distribution of X : the set of pairs $\{(x, f(x)) : x \in A\}$.

Example: a fair coin is tossed 3 times. $X(a)$ is the number of heads corresponding to a .
Find the probability distribution of a .

$$\hookrightarrow f(0) = P(X=0) = P(\{\text{TTT}\}) = 1/8$$

$$f(1) = P(X=1) = P(\{\text{TTH}, \text{HTH}, \text{HTT}\}) = 3/8$$

$$f(2) = P(X=2) = P(\{\text{HHT}, \text{HTH}, \text{THH}\}) = 3/8$$

$$f(3) = P(X=3) = P(\{\text{HHH}\}) = 1/8$$

$$\hookrightarrow f(0) + f(1) + f(2) + f(3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{8}{8} = 1 !$$

A function $f(x)$ is a discrete pf if and only if it satisfies both of the following properties:

$$1) 0 \leq f(x) \leq 1 \text{ for all } x \in \mathbb{R}$$

$$2) \sum_{\text{all } x} f(x) = 1$$

Probability Histogram: the graph of a probability mass function of a discrete random variable.

Example: the random variable x has a pmf:

x	0	1	2	3	4
$f(x)$	$0.1c$	$0.2c$	$0.5c$	c	$0.2c$

a) Find the value of c :

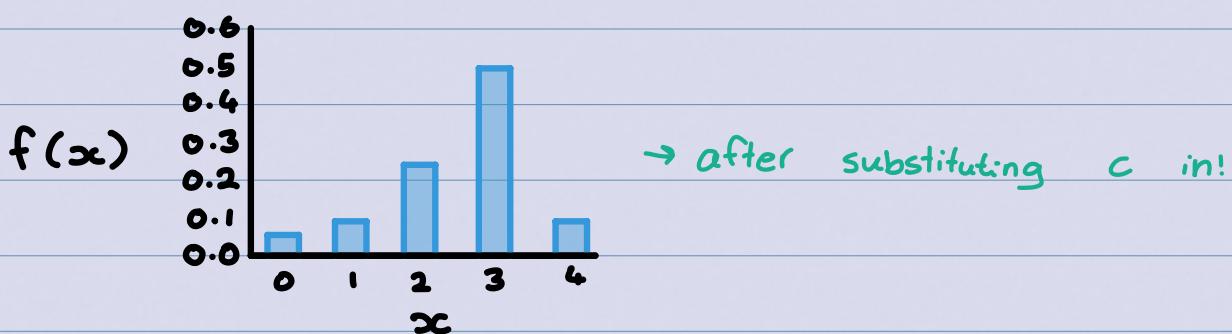
$$\hookrightarrow \sum f(x) = 1 \rightarrow 0.1c + 0.2c + 0.5c + c + 0.2c = 2c.$$

$$\therefore 2c = 1, \text{ so } c = \frac{1}{2} = 0.5.$$

b) Find $P(x > 2)$

$$\begin{aligned} P(x > 2) &= P(x = 3) + P(x = 4) \\ &= c + 0.2c = 1.2c \\ &\Rightarrow 1.2c = 0.6 \end{aligned}$$

c) Plot the probability histogram of $f(x)$



Cumulative Distribution Function (cdf): the cdf

of a random variable x is defined for any real x by:

$$F(x) = P(X \leq x)$$

If X is a discrete random variable with probability function $f(x)$, then:

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$$

Example: find the cdf of random variable x from the previous example:

$$F_x(x) = \begin{cases} 0 & x < 0 \\ 0.05 & 0 \leq x < 1 \\ 0.15 & 1 \leq x < 2 \\ 0.4 & 2 \leq x < 3 \\ 0.9 & 3 \leq x < 4 \\ 1 & 4 \leq x \end{cases}$$

- For discrete random variables, the CDF $F(x)$ is represented as a step function.

(General) Example: Consider rolling two fair six-sided dice, and let the random variable X be the minimum of the rolls. What is $f_x(2)$?

$$\hookrightarrow f_x(2) = P(X=2)$$

$$= P(\{(2,2), (2,3), (2,4), (2,5), (2,6), (3,2), (4,2), (5,2), (6,2)\})$$

↳ $P(X=2) = 9/36.$

Properties of Cumulative Distribution Functions:

↳ let $F(\cdot)$ be a cdf. Then:

- 1) $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$
- 2) $F(x) \leq F(y)$ for $x < y \rightarrow F(x)$ is a non-decreasing function of $x \quad \forall x \in \mathbb{R}$.
- 3) $\lim_{h \rightarrow 0^+} F(x+h) = F(x) \rightarrow F(x)$ is continuous on the right
- 4) $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.

Example: Students A, B, and C each independently answer a question. The probability of getting the correct answer is 0.9 for A, 0.7 for B, and 0.4 for C. Let X be the number of people who get the answer correct.

a) Compute the probability function of X :

Since the events are independent, $P(ABC) = P(A)P(B)P(C)$

↓

$$f(0) = P(X=0) = P(\bar{A}\bar{B}\bar{C}) = 0.1 \times 0.3 \times 0.6 = 0.018$$

$$\begin{aligned} f(1) &= P(X=1) = P(A\bar{B}\bar{C}) + P(\bar{A}B\bar{C}) + P(\bar{A}\bar{B}C) \\ &= 0.9 \times 0.3 \times 0.6 + 0.1 \times 0.7 \times 0.6 + 0.1 \times 0.3 \times 0.4 \\ &= 0.216 \end{aligned}$$

$$f(2) = P(X=2) = P(AB\bar{C}) + P(A\bar{B}C) + P(\bar{A}BC)$$

$$= 0.9 \times 0.7 \times 0.6 + 0.9 \times 0.3 \times 0.4 + 0.1 \times 0.7 \times 0.4$$

$$= 0.514$$

$$f(3) = P(x=3) = P(ABC) = 0.9 \times 0.7 \times 0.4 = 0.252$$

$$\hookrightarrow f_x(x) = \begin{cases} 0.018 & x=0 \\ 0.216 & x=1 \\ 0.514 & x=2 \\ 0.252 & x=3 \end{cases}$$

b) compute the cdf of X

$$\hookrightarrow F_x(x) = \begin{cases} 0.018 & x < 0 \\ 0.234 & 0 \leq x < 1 \\ 0.748 & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

c) compute the probability that at least one person gets the answer correct.

$$\hookrightarrow P(x \geq 1) = P(x=1) + P(x=2) + P(x=3)$$

$$= 0.216 + 0.514 + 0.252 = 0.982.$$

$$\text{OR, } P(x \geq 1) = 1 - P(x=0) = 1 - 0.018 = 0.982.$$

$\cdot F(x)$ can be obtained from $f(x)$, and vice versa!



Example: the random variable x has a cdf given by:

x	1	2	3	4
$F(x)$	0.2	0.5	0.8	1

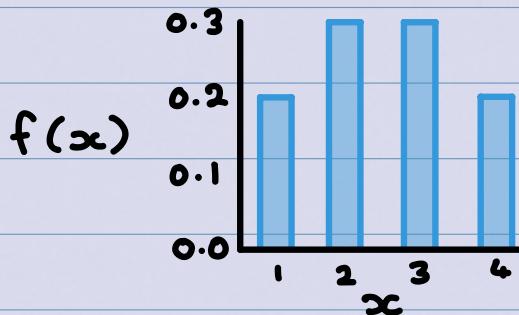
a) Find $f(3)$.

$$\hookrightarrow f_x(3) = F_x(3) - F_x(2) = 0.8 - 0.5 = 0.3$$

b) Calculate $P(1 < x \leq 3)$.

$$\begin{aligned} \hookrightarrow P(1 < x \leq 3) &= P(x \leq 3) - P(x \leq 1) \\ &= F(3) - F(1) = 0.8 - 0.2 = 0.6. \end{aligned}$$

c) Plot $f(x)$.



paka, "mean of x " or "first moment of x "

Expected Value: suppose X is a discrete random variable with probability function $f_x(x)$. Then, $E(X)$ is called the expected value of X and is defined by :

$$E(X) = \sum_{x \in A} x \cdot f_x(x)$$

Example: X denotes the outcome of one fair six-sided die roll. Compute $E(X)$.

paka, the odds of $x=1, \dots, 6 = 1/6$.

we know $f(1) = f(2) = \dots = f(6) = 1/6$

$$\therefore E(X) = \sum_{x \in \{1, 2, \dots, 6\}} x \cdot f(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Note(s):

- 1) The expected value of X may be a value that X can never actually take.
- 2) both μ and $E(x)$ refer to the same quantity which is the expected value of random variable X .

Example: a random variable X only takes two values, 0 or 1. If $P(X=0) = 0.4$, find $E(x)$.

$$P(X=1) = \overline{P(X=0)} = 1 - P(X=0) = 1 - 0.4 = 0.6$$

$$E(x) = \sum_{x \in \{0,1\}} x \cdot f(x) = 0 \times 0.4 + 1 \times 0.6 = 0.6.$$

Example: consider rolling 2 fair dice. Let x be the sum of the rolls, and let $Y = X \pmod{4}$. Find Y 's pmf.

→ we know $S = \{(1,1), (1,2), \dots, (2,1), (2,2), \dots, (6,1), (6,2), \dots, (6,6)\}$
 $\therefore A = \{2, 3, 4, \dots, 12\}$

x	$f_x(x)$	$y = x \pmod{4}$
2	$1/36$	2
3	$2/36$	3
4	$3/36$	0
5	$4/36$	1
6	$5/36$	2
7	$6/36$	3
8	$5/36$	0
9	$4/36$	1
10	$3/36$	2
11	$2/36$	3
12	$1/36$	0

$$f_y(0) = P(Y=0) = P(X=4) + P(X=8) + P(X=12) \\ = \frac{3}{36} + \frac{5}{36} + \frac{1}{36} = \frac{9}{36} = \frac{1}{4}.$$

$$f_y(1) = P(Y=1) = P(X=5) + P(X=9) \\ = \frac{4}{36} + \frac{4}{36} = \frac{8}{36} = \frac{2}{9}$$

$$f_y(2) = P(Y=2) = P(X=2) + P(X=6) + P(X=10) \\ = \frac{1}{36} + \frac{5}{36} + \frac{3}{36} = \frac{9}{36} = \frac{1}{4}$$

$$f_y(3) = P(Y=3) = P(X=3) + P(X=7) + P(X=11) \\ = \frac{2}{36} + \frac{6}{36} + \frac{2}{36} = \frac{10}{36} = \frac{5}{18}$$

$\forall y \notin \{0, 1, 2, 3\}, f_y(y) = 0.$

• General Formula when $Y = aX + b$

↳ in the case of a linear function of the form

$Y = g(x) = ax + b$, where $a \neq 0$:

$$f_y(y) = f_x\left(\frac{y-b}{a}\right)$$

• General Formula when $Y = g(x)$

↳ Let $Y = g(x)$ where $g(\cdot)$ is an injective function.
Since $g(\cdot)$ is injective, its inverse $g^{-1}(\cdot)$ exists.

aka one-to-one

Then, we see that:

$$\begin{aligned}
 f_Y(y) &= P(Y = y) \\
 &= P(g(X) = y) \\
 &= P(X = g^{-1}(y)) \\
 &= f_X(g^{-1}(y))
 \end{aligned}$$

$$\therefore f_Y(y) = f_X(g^{-1}(y))$$

Let A be a discrete random variable with the range A and pmf $f(x)$. The expected value of some function $g(\cdot)$ of X is given by:

$$E[g(x)] = \sum_{x \in A} g(x) \cdot f(x)$$

Example: a TV station sells 15, 30, and 60 second advertising spots. X is the length of a random commercial and suppose that it's probability distribution is given by:

x	15	30	60
$f(x)$	0.1	0.3	0.6

a) Find $E(x)$.

$$\hookrightarrow E(x) = \sum_{x \in \{15, 30, 60\}} x \cdot f(x) = 15 \times 0.1 + 30 \times 0.3 + 60 \times 0.6 = 46.5 \text{ seconds.}$$

b) if a 15s spot is \$500, a 30s spot is \$800, and a 60s spot is \$1000, find the average amount paid for a commercial on this station.

$$\hookrightarrow E(\text{amount paid}) = 500 \times 0.1 + 800 \times 0.3 + 1000 \times 0.6 = \$890.$$

Example: Consider the discrete random variable X with pmf of the form $f_X(x) = \frac{1}{3}$, for $x = -1, 0, 1$. If we define the random variable $Y = X^2$, calculate $E(Y)$.

$$E(Y) = \sum_{x \in A} g(x) f(x) = \frac{1}{3}(-1)^2 + \frac{1}{3}(0)^2 + \frac{1}{3}(1)^2 = \frac{2}{3} !$$

Linearity Property of Expectation

The expectation $E(\cdot)$ is a linear operator, satisfying the properties:

i) For constants a and b ,

$$E[a g(x) + b] = a E[g(x)] + b$$

2) For constants a and b and for functions g_1 and g_2 ,

$$E[a g_1(x) + b g_2(x)] = a E[g_1(x)] + b E[g_2(x)]$$

• If $g(x)$ is a linear function, then $E[g(x)] = g[E(x)]$.

This is not necessarily true if $g(x)$ is non-linear!

Example: let random variable X have the following probability function:

x	1	2	3	4
$f(x)$	0.4	0.3	0.1	0.2

a) Find $E(x)$.

$$\hookrightarrow \sum_{x \in \{1, 2, 3, 4\}} x \cdot f(x) = 1 \times 0.4 + 2 \times 0.3 + 3 \times 0.1 + 4 \times 0.2 = 2.1$$

b) Find $E(\frac{1}{x})$.

$$\hookrightarrow \sum_{x \in \{1, 2, 3, 4\}} \frac{1}{x} \cdot f(x) = \frac{1}{1} \times 0.4 + \frac{1}{2} \times 0.3 + \frac{1}{3} \times 0.1 + \frac{1}{4} \times 0.2 = 0.633.$$

• Variance: the variance of a random variable X is:

$$\sigma^2 = \text{Var}(X) = E[(X - \mu)^2]$$

↳ can also be calculated as: $\sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2$

aka $\mu!$

Standard Deviation: the standard deviation of a random variable X is:

$$\sigma = \text{sd}(X) = \sqrt{\text{Var}(X)} = \sqrt{E[(X - \mu)^2]}$$

• Important Properties:

- for real constants a and b , $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- if $Y = aX + b$, then
 - 1) $E(Y) = aE(X) + b$
 - 2) $\text{Var}(Y) = a^2 \text{Var}(X)$
 - 3) $\text{sd}(Y) = |a| \text{sd}(X)$

Example: X has $\text{Var}(X) = 2$. Compute the variance of random variable Y , where $Y = -2X + 3$.

↳ we know that $\text{Var}(Y) = a^2 \text{Var}(X) = (-2)^2 (2) = 8$.

• Properties of Variance:

- 1) If X is a r.v., $\text{Var}(X) \geq 0$.
- 2) big variances indicate that the probability distribution of X is more "spread out" around the mean.
- 3) $\text{Var}(X) = 0$ if and only if $P(X = E(X)) = 1$.
↳ no variance if the expected value is guaranteed!

Example: X is a random variable with probability function given by:

x	1	2	3	4	5
$f(x)$	0.1	0.1	0.3	0.35	0.15

If $Y = 4 - 2X$, find $E(Y)$, $\text{Var}(Y)$, and $\text{sd}(Y)$:

$$\cdot E(X) = \sum_{x \in \{1, \dots, 5\}} x \cdot f_x(x) = 1 \cdot 0.1 + 2 \cdot 0.1 + 3 \cdot 0.3 + 4 \cdot 0.35 + 5 \cdot 0.15 = 3.35$$

$$E(Y) = 4 - 2E(X) = 4 - 2(3.35) = -2.7$$

$$\begin{aligned} E(X^2) &= \sum_{x \in \{1, \dots, 5\}} x^2 \cdot f_x(x) = 1^2 \cdot 0.1 + 2^2 \cdot 0.1 + 3^2 \cdot 0.3 + 4^2 \cdot 0.35 + 5^2 \cdot 0.15 \\ &= 12.55 \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - \mu^2 = E(X^2) - E(X)^2 = 12.55 - 3.35^2 \\ &= 1.3275 \end{aligned}$$

$$\text{Var}(Y) = (-2)^2 \text{Var}(X) = 4(1.3275) = 5.31$$

$$\text{sd}(Y) = \sqrt{\text{Var}(Y)} = \sqrt{5.31} = 2.3043.$$

$$\therefore E(Y) = -2.7, \text{Var}(Y) = 5.31, \text{and } \text{sd}(Y) = 2.3043!$$

Discrete Uniform Distribution: suppose the range of random variable X is $\{a, a+1, \dots, b\}$, where a and b are integers and all values are equally likely. Then, X is said to be a discrete uniform distribution on the set $\{a, a+1, \dots, b\}$. We may write $X \sim U[a, b]$, where a and b are the parameters of the distribution.

Example: if X is the number obtained from rolling a fair die, then X has a discrete uniform distribution over the set $\{1, 2, \dots, 6\}$. Here, $a=1$, $b=6$, and $X \sim U[1, 6]$.

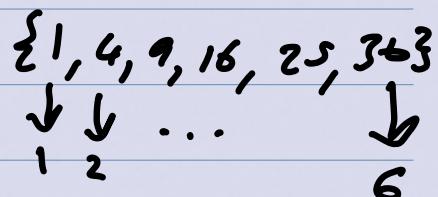
Suppose $X \sim U[a, b]$. This means that we have $b-a+1$ possible values for X , which are all equally likely. Therefore, the probability function of X is:

$$f(x) = P(X=x) = \frac{1}{b-a+1} \text{ for } x=a, a+1, \dots, b! \quad \text{otherwise, } f(x)=0$$

$$E(X) = \frac{a+b}{2}$$

$$\text{Var}(X) = \frac{(b-a+1)^2 - 1}{12}$$

$$F(X) = \frac{\lfloor x \rfloor - a + 1}{b-a+1}$$



$$\frac{1+6}{2} = \underline{3.5}$$

Example: X is the outcome of a 6-sided die roll.

a) Compute $E(X)$

$\hookrightarrow X \sim U[1, 6] \rightarrow E(X) = \frac{a+b}{2}$, where $a=1$, $b=6$

$$\hookrightarrow \frac{1+6}{2} = \frac{7}{2} = 3.5$$

b) Compute $\text{Var}(X)$

$$\hookrightarrow X \sim U[1, 6] \rightarrow \text{Var}(X) = \frac{(b-a+1)^2 - 1}{12}$$

$$\hookrightarrow \text{Var}(X) = \frac{(6-1+1)^2 - 1}{12} = \frac{35}{12}$$

c) If $g(x) = x^2$, then compute $E(g(X))$.

$$\hookrightarrow E(g(x)) = E(x^2).$$

$$\hookrightarrow E(x^2) = 1 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6}$$
$$= \frac{1}{6} (1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} \approx 15.17.$$

Bernoulli Trial : a random experiment that the result is either a success or failure and the probability of success in every trial is p for $0 < p < 1$. Then the probability of failure is $1-p$ in every trial.

↓

closely related to a bernoulli trial is a binomial experiment

↳ same as bernoulli, but repeated independently for a number of n times, and record X as the number of successes.

↳ then the RV X has binomial distribution as denoted by $X \sim \text{Binomial}(n, p)$

↳ p is probability of success!

• Key assumptions:

- 1) There are multiple trials, and on each trial, there are only two outcomes (S or F)
- 2) the probability of success (p) is constant over all n trials
- 3) the outcome (S/F) is independent of all other trials.

• Probability Function

for a RV $X \sim \text{Binomial}(n, p)$, we want to determine $f(x) = P(X = x)$.

if we have x successes, then there must be $n-x$ failures. The total number of different arrangements of S's and F's is:

$$\frac{n!}{x!(n-x)!} = \binom{n}{x}$$

since the trials are independent, each arrangement (of x successes and $n-x$ failures) is achieved with probability:

$$p^x (1-p)^{n-x}$$

$$\therefore f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

Expected Value / Variance of Binomial Distribution:



if $X \sim \text{Binomial}(n, p)$, then:

$$\hookrightarrow E(X) = n \cdot p$$

$$\text{Var}(X) = np(1-p)$$

Example: 75% of students use Instagram. A sample of 5 students are chosen. What is the probability that at least 3 students use Instagram?

$\hookrightarrow X \sim \text{Binomial}(n, p)$, where $n=5$ and $p=0.75$

$$\therefore X \sim \text{Binomial}(5, 0.75)$$

$$P(X \geq 3) = P(X=3) + P(X=4) + P(X=5)$$

$$= \binom{5}{3} 0.75^3 \cdot 0.25^2 + \binom{5}{4} 0.75^4 \cdot 0.25^1 + \binom{5}{5} 0.75^5 \cdot 0.25^0$$

$$= 0.8965.$$

Example: in a weekly lottery, you have a 0.02 chance of winning a prize with a single ticket. If I buy one ticket a week for 52 weeks, what is the probability that I win no prizes?

$\hookrightarrow X \sim \text{Binomial}(n, p)$ where $n=52$ and $p=0.02$

$\hookrightarrow X \sim \text{Binomial}(52, 0.02)$

$$P(X=0) = \binom{52}{0} 0.02^0 (1-0.02)^{52} = 0.3498$$

What's the probability I win 2 or more prizes?

$$1 - P(X < 2) = 1 - (P(X=0) + P(X=1))$$

$$\hookrightarrow 1 - (0.3498 + \binom{52}{1} 0.02^1 (1-0.02)^{51})$$

$$= 1 - 0.3498 - 0.3712 = 0.2790 !$$

• Geometric Series Formula: $\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$ if $|r| < 1$

• Partial Geometric Series: $\sum_{i=0}^R r^i = \frac{1-r^{R+1}}{1-r}$

• Binomial Theorem: $(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$

• Taylor Series for e^x : $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$

• Geometric Distribution:

↳ same as bernoulli's, but repeat the experiment independently until the first success occurs.

record X as the number of failures before the first success.

Then the RV X has a geometric distribution denoted as $X \sim \text{Geo}(p)$.

Geometric Distribution Probability Function
for a RV $X \sim \text{Geo}(p)$, we want to compute $f(x) = P(X=x)$.

• there is only one way to arrange x failures!

the geometric distribution has the following pmf:

$$f(x) = (1-p)^x p, \quad x=0, 1, 2, \dots$$

• If $X \sim \text{Geo}(p)$, then

$$\hookrightarrow \begin{aligned} \cdot E(X) &= \frac{1-p}{p} \\ \cdot \text{Var}(X) &= \frac{1-p}{p^2} \end{aligned}$$

Example: at campus pizza, 60% of orders are for takeout.

a) What is the probability that the 5th order is the first takeout order?

$$\hookrightarrow X \sim \text{Geo}(p) \text{ where } p = 0.6$$
$$\hookrightarrow X \sim \text{Geo}(0.6)$$

$P(X=4) = (1-0.6)^4 \cdot 0.6 = 0.01536$

p^4 non-takeouts first

b) What is the probability that more than 3 orders are ordered before the first takeout?

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - (P(X=3) + P(X=2) + P(X=1) + P(X=0)) \\ &= 1 - ((1-0.6)^3 \cdot 0.6 + (1-0.6)^2 \cdot 0.6 + (1-0.6)^1 \cdot 0.6 \\ &\quad + (1-0.6)^0 \cdot 0.6) = 0.0256 \end{aligned}$$

Binomial / Geometric Models Summary

- 1) Two outcomes in each Bernoulli trial,
- 2) Each trial has same probability of success,
- 3) Trials are independent of each other.

Poisson Distribution

- The RV X represents the number of events of some type.
 - The event occurs with a known constant mean rate, denoted by μ , where $\mu > 0$.
 - One occurrence is independent of the next
- ↳ then the RV X has a Poisson Distribution, denoted by $X \sim \text{Poisson}(\mu)$

↳ Poisson Distribution pmf: $f(x) = \frac{e^{-\mu} \mu^x}{x!}$

r for $x=0, 1, 2, \dots$

- If $X \sim \text{Poisson}(\mu)$, then:

- ↳
- $E(X) = \mu$
 - $\text{Var}(X) = \mu$

Poisson Distribution from Binomial:

↳ If $X \sim \text{Binomial}(n, p)$, then for each value of $x = 0, 1, 2, \dots$, and $p \rightarrow 0$ with $np = u$ constant,

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-u} u^x}{x!}$$

- This allows us to express the Poisson Distribution with $u = np$ as a close approximation to the binomial distribution when n is very large and p is small.
- As a general rule, the approximation gives reasonable results provided $n \geq 100$ and $p \leq 0.01$, and when x is close to np .
- When $n \geq 100$ and $np \leq 10$, the approximation will generally be excellent.
- When $n \geq 20$ and $p \leq 0.05$, the Poisson Distribution will be a good approximation.

Example: 1 out of 20 cups are winners.

Suppose you buy 100 cups.

- a) What is the probability that there are exactly two winning cups?

$$X \sim \text{Binomial}(100, 0.05)$$

$$P(X=2) = \binom{100}{2} 0.05^2 (1-0.05)^{98} = 0.08118$$

b) Approximate the probability using a Poisson approximation.

$$\mu = np = 100(0.05) = 5$$

$$P(X=2) = \frac{e^{-\mu} \mu^2}{2!} = \frac{e^{-5} 5^2}{2} = 0.0842$$

Example: you roll a fair dice until you get a 6. Let RV X denote the number of rolls to get the first 6 (including the trial where the 6 occurs). What is the expected value of X ?

$Y = \# \text{ trials before first 6: } Y \sim \text{Geo}(1/6)$

$$E(Y) = \frac{1-p}{p} = \frac{1-1/6}{1/6} = \frac{5/6}{1/6} = 5$$

including the trial where the 6 occurs:

$$E(X) = E(Y) + 1 = 6.$$

• Hypergeometric Distribution:

• Suppose a collection consists of a finite number of N items which can be classified into two distinct types:

- There are r items of type 1 (or success),
- The remaining $N-r$ items are of type 2 (or failure).

Suppose $n \leq N$ items are picked at random without replacement.

↓

If X represents the number of type 1 items that are drawn, then X has a hypergeometric distribution with parameters N, r , and n .

$$\hookrightarrow X \sim HG(N, r, n)$$

• Hypergeometric Probability Mass Function:

For a RV $X \sim HG(N, r, n)$, we want to compute $f(x) = P(X=x)$

- There are $\binom{N}{n}$ points in the sample space
- There are $\binom{r}{x}$ ways to choose x type-1 objects from the r type-1 objects available
- There are $\binom{N-r}{n-x}$ ways to choose the remaining $n-x$ type-2 objects from the $N-r$ type-2 objects available.

$$\therefore f(x) = P(X=x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

where $x \geq \max\{0, n-(N-r)\}$ and $x \leq \min\{r, n\}$.

- If $X \sim HG(N, r, n)$,

$$\hookrightarrow E(X) = \frac{nr}{N}$$

$$\cdot \text{Var}(X) = n \frac{r}{N} \left(1 - \frac{r}{N}\right) \frac{N-n}{N-1}$$

Example: a box has 100 microchips, 80 are good and 20 are defective. I select 10 microchips at random without replacement.

a) What's the probability that the 10 chips selected include no more than 2 defectives?

$X = \#$ defective chips in sample of 10.

↳ ∴, $X \sim HG(N, r, n)$, where $N=100$, $r=20$, and $n=10$.

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= \sum_{x=0}^2 \frac{\binom{20}{x} \binom{80}{10-x}}{\binom{100}{10}} = 0.681$$

b) now we look at 30 chips instead of 10. Find the probability of getting exactly 6 defectives.

↳ $X \sim HG(N, r, n)$, where $N=100$, $r=20$, $n=30$

$$P(X=6) = \frac{\binom{20}{6} \binom{80}{24}}{\binom{100}{30}} = 0.214$$

• Hypergeometric vs Binomial Distributions

- Both have 2 outcomes
- Experiment repeated n times in both
- The RV X records the number of successes

- But, the binomial distribution requires n independent trials where p is constant

- In hypergeometric, n draws are made w/o replacement, so trials are NOT independent!

• Negative Binomial Distribution

- ↳ same as regular binomial, but repeated independently until a specified number of r successes have been obtained.
- record X as the number of trials to obtain r successes.

Then the RV X has a negative binomial distribution, denoted as $X \sim NB(r, p)$

Negative Binomial Probability Mass Function

For a RV $X \sim NB(r, p)$, we want to compute $f(x) = P(X = x)$.

- One must obtain the r^{th} success on the x^{th} trial by obtaining $(r-1)$ successes in the first $(x-1)$ trials, in any order.
- ↳ There are $\binom{x-1}{r-1}$ different orders.
- Then obtain a success on the x^{th} trial

$$\hookrightarrow f(x) = \binom{x-1}{r-1} p^r (p-1)^{x-r}, \quad x=r, r+1, \dots$$

If $X \sim NB(r, p)$,

$$\hookrightarrow E(X) = r/p$$

$$\cdot \text{Var}(X) = \frac{r(1-p)}{p^2}$$

Example: a start-up wants 5 investors. Each investor will independently agree to invest with a 20% probability. The founder asks one investor at a time until he gets 5 "yes's. Let X represent the total number of investors asked.

a) what is the pmf of X ?

$$\hookrightarrow P(X=x) = \binom{x-1}{4} (0.2)^5 (0.8)^{x-5}, x=5, 6, \dots$$

b) Find $P(X=6)$

$$P(X=6) = \binom{5}{4} (0.2)^5 (0.8)^1 = 0.00128$$

Binomial vs Negative Binomial Distribution

- Can be distinguished by looking at what is specified / known in advance
- Binomial: n independent trials but we do not know the number of successes we will obtain
- NB: we know the number of successes to be obtained, but don't know the number of trials needed to obtain the r successes.

→ Continuous Probability Distributions (Chapter 3)

- Continuous random variables

↳ take on values from \mathbb{R} (no longer countable)

↳ they are treated differently than discrete random variables because now $P(X=a)=0$, since the chance of obtaining the specific value of a among an infinite number of possibilities is extremely remote!

∴ we specify the probability over intervals rather than individual points.

Cumulative Distribution Function (CDF)

↳ we can use the cdf to describe the distribution of a continuous RV: $F(x) = P(X \leq x)$

- Must satisfy the following properties:

- 1) $F(x)$ is defined $\forall x$

- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$

- 3) $F(x)$ is a non-decreasing function of x .

- 4) $P(a < X \leq b) = F(b) - F(a)$

For a continuous RV X , the probability density function (pdf) $f(x)$ is the derivative of the cdf and is given by:

$$f(x) = \frac{d F(x)}{dx}$$

Properties of a PDF:

- 1) $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$

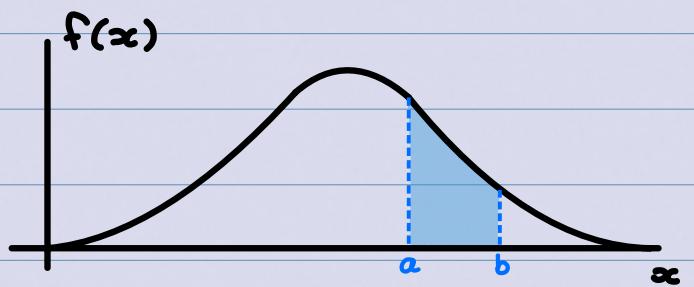
- 2) $f(x) \geq 0$ (since $F(x)$ non-decreasing)

- 3) $\int_{-\infty}^{\infty} f(x) = \int_{all \infty} f(x) = 1$

- 4) $F(x) = \int_{-\infty}^x f(u) du$

Interval Probability

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



Percentile: suppose X is a continuous RV, with the cdf $F(x)$. The $100p$ th percentile of X (or the distribution) is the value q such that:

$$P(X \leq q) = p$$

Example: if $F(1) = 0.8$, then the 80% percentile of the distribution is 1.

• If $p = 0.5$, then q is the median of X !

Example: Let X be a RV with cdf given by:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x/4, & 0 < x \leq 4 \\ 1, & x > 4 \end{cases}$$

a) find the pdf $f(x)$.

$$\hookrightarrow f(x) = \frac{dF(x)}{dx} \rightarrow f(x) = \begin{cases} \frac{1}{4}, & 0 < x < 4 \\ 0, & \text{otherwise} \end{cases}$$

b) find the 90th percentile, which is the value of x such that the area under the curve to the left of x is 90%

$$\hookrightarrow F(x) = \frac{x}{4} = 0.9 \rightarrow x = 3.6.$$

Example: let X be a RV with the following pdf:

$$f(x) = \begin{cases} C(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

a) find C .

\hookrightarrow since $\int_{-\infty}^{\infty} f(x) dx = \int_0^2 f(x) dx = 1$, we know

$$\int_0^2 C(4x - 2x^2) dx = 1.$$

$$\hookrightarrow C \int_0^2 4x - 2x^2 dx = C \left[2x^2 - \frac{2x^3}{3} \right]_0^2 = 1$$

$$\hookrightarrow C \left(8 - \frac{8}{3} \right) = 1 \rightarrow C = \frac{3}{8}$$

b) find $F(x)$

$\hookrightarrow F(x)$ must be 0 for $x \leq 0$.

If $0 < x < 2$:

$$\hookrightarrow F(x) = \int_{-\infty}^x f(u) du = \int_0^x \frac{3}{8}(4u - u^2) du$$

$$= \frac{3}{8} \left[2u^2 - \frac{2}{3}u^3 \right]_0^x = \frac{3}{4}x^2 - \frac{x^3}{4}$$

If $x > 2$, $F(x) = 1$.

$$\therefore F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{3}{4}x^2 - \frac{x^3}{4}, & 0 < x < 2 \\ 1, & x \geq 2 \end{cases}$$

c) Find $P(X > 1)$

$$\hookrightarrow P(X > 1) = 1 - P(X \leq 1) = 1 - F(1)$$

$$= 1 - \left(\frac{3}{4} - \frac{1}{4}\right) = 1 - \frac{1}{2} = \frac{1}{2}.$$

• When X is a continuous RV with pdf $f(x)$, the expected value of X is given by:

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

In general, we define:

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

The variance of X is defined as:

$$\sigma^2 = \text{Var}(X) = E[(x - \mu)^2] = E(x^2) - E(x)^2$$

where:

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

The standard deviation of X is:

$$\sigma = \text{sd}(X) = \sqrt{\text{Var}(X)}$$

Example: suppose a continuous RV X has pdf $f(x) = cx^2$ for $0 < x < 2$, and 0 otherwise.

Find $P(X > 1)$:

$$\hookrightarrow \text{we know } \int_{-\infty}^{\infty} f(x) dx = \int_0^2 f(x) dx = 1$$

$$\hookrightarrow c \int_0^2 x^2 dx = c \left[\frac{x^3}{3} \right]_0^2 = \frac{8}{3} c = 1$$

$$\therefore c = \frac{3}{8}.$$

$$P(X > 1) = \int_1^{\infty} f(x) dx = \int_1^2 \frac{3}{8} x^2 dx = \frac{3}{8} \left[\frac{x^3}{3} \right]_1^2 \\ = \frac{1}{8} (8 - 1) = \frac{7}{8}.$$

• All of the earlier properties still hold in the continuous case!



- $E(\alpha g(x) + b) = \alpha E(g(x)) + b$
- $E(\alpha g_1(x) + b g_2(x)) = \alpha E(g_1(x)) + b E(g_2(x))$
- $\text{Var}(\alpha g(x) + b) = \alpha^2 \text{Var}(g(x))$

Example: let X be a RV with the following pdf:

$$f(x) = \begin{cases} kx^2, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

a) Find k : $\int_{-\infty}^{\infty} f(x) dx = 1 \rightarrow k \int_{-1}^1 x^2 dx = 1$

$$\therefore k \left[\frac{x^3}{3} \right]_{-1}^1 = k \left[\frac{1}{3} + \frac{1}{3} \right] = \frac{2k}{3} = 1 \rightarrow k = \frac{3}{2}.$$

b) Find the cdf of x for all values of x :

$$\hookrightarrow F(x) = 0 \text{ if } x \leq -1.$$

If $-1 < x < 1$:

$$\begin{aligned}\hookrightarrow F(x) &= \int_{-\infty}^x f(u) du = \frac{3}{2} \int_{-1}^x u^2 du = \frac{3}{2} \left[\frac{u^3}{3} \right]_{-1}^x \\ &= \frac{1}{2} (x^3 + 1).\end{aligned}$$

$$F(x) = 1 \text{ if } x \geq 1.$$

$$\therefore F(x) = \begin{cases} 0, & x \leq -1 \\ \frac{1}{2}(x^3 + 1), & -1 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

c) Find $P(-\frac{1}{3} < x < 0.5)$:

$$\hookrightarrow P(-\frac{1}{3} < x < \frac{1}{2}) = F(\frac{1}{2}) - F(-\frac{1}{3})$$

$$= \frac{1}{2} \left((\frac{1}{2})^3 - (-\frac{1}{3})^3 \right) = \frac{35}{432}$$

d) Find $E(x)$

$$\begin{aligned}\hookrightarrow E(x) &= \int_{-\infty}^{\infty} x f(x) dx = \frac{3}{2} \int_{-1}^1 x \cdot x^2 dx \\ &= \frac{3}{2} \left[\frac{x^4}{4} \right]_{-1}^1 = \frac{3}{2} \left[\frac{1}{4} - \frac{1}{4} \right] = 0.\end{aligned}$$

e) Find $\text{Var}(x)$

$$\hookrightarrow \text{Var}(x) = E(x^2) - E(x)^2.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \frac{3}{2} \int_{-1}^1 x^2 \cdot x^2 dx$$

$$= \frac{3}{2} \left[\frac{x^5}{5} \right]_{-1}^1 = \frac{3}{2} \left(\frac{1}{5} + \frac{1}{5} \right) = \frac{3}{5}.$$

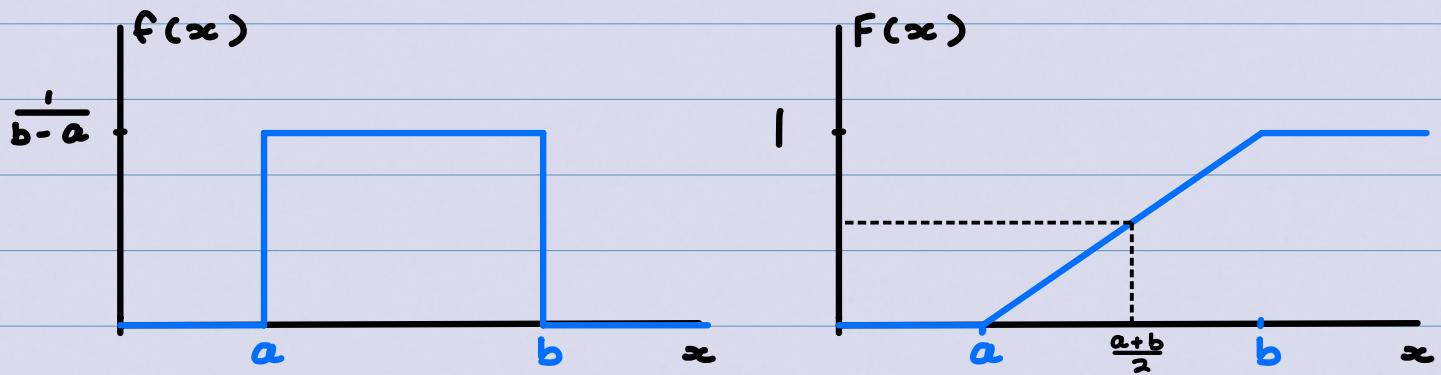
$$\hookrightarrow \text{Var}(X) = \frac{3}{5} - 0^2 = \frac{3}{5} = 0.6.$$

Continuous Uniform Distribution: a RV X has a continuous uniform distribution on $[a, b]$ if X has a pdf that is constant on $[a, b]$ and 0 elsewhere.

↪ The pdf of X is defined as:

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

• Note: it does not matter whether the interval is open or closed.



→ For a continuous RV X where $X \sim U[a, b]$,

$$\bullet E(X) = \frac{a+b}{2}, \quad \bullet \text{Var}(X) = \frac{(b-a)^2}{12}$$

Exponential Distribution: a continuous RV X is said to have an exponential distribution with the rate $\lambda > 0$ if X has the probability density:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

the cdf of X is given by:

$$F(x) = 1 - e^{-\lambda x} \quad \text{for } x > 0$$

It's also common to use parameter $\Theta = \frac{1}{\lambda}$ in the exponential distribution.

$$f(x) = \begin{cases} \frac{1}{\Theta} e^{-x/\Theta}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\Theta > 0$, and $F(x) = 1 - e^{-x/\Theta}$ for $x > 0$

We write $X \sim \text{Exp}(\Theta)$ or $X \sim \text{Exponential}(\Theta)$

For a continuous RV X where $X \sim \text{Exp}(\Theta)$,

- $E(X) = \Theta$, $\text{Var}(X) = \Theta^2$

Example: the amount of time in hours that a computer survives before breaking down is exponentially

distributed with a mean of 100 hours.

- a) Find the probability that a computer will function between 50 and 150 hours before breaking down.

↳ let X be the time a computer functions before breaking down.

$X \sim \text{Exp}(\theta)$, where $\theta = 100$.

$$P(50 < X < 150) = F(150) - F(50)$$

$$= (1 - e^{-150/100}) - (1 - e^{-50/100})$$

$$= e^{-1/2} - e^{-3/2} = 0.383$$

- b) Find the probability that will function for fewer than 100 hours.

$$\hookrightarrow P(X < 100) = F(100) = 1 - e^{-100/100} = 1 - e^{-1}$$

- c) If a computer survives more than 100 hours, what is the probability that it'll survive 50 more?

$$P(X > 100 + 50 | X > 100) = \frac{1 - F(150)}{1 - F(100)} = \frac{e^{-3/2}}{e^{-1}} = e^{-1/2}$$

• Memoryless Property of the Exponential Distribution

↳ shown in part c of previous example

$$\hookrightarrow P(X > b + c \mid X > b) = P(X > c)$$

↳ basically, given that you've waited b units of time for the next event, the probability that you wait an additional c units of time does NOT depend on b and only depends on c .

Example: the length of a phone call is an exponential RV with a mean of 10 minutes. If someone arrives immediately ahead of you at a telephone booth, find the probability you'll wait:

a) more than 10 minutes

↳ let X be the minutes you wait $\rightarrow X \sim \text{Exp}(\theta)$, where $\theta = 10$.

$$\hookrightarrow P(X > 10) = 1 - P(X \leq 10) = 1 - F(10)$$

$$= 1 - (1 - e^{-x/\theta}) = e^{-10/10} = e^{-1} = 0.3679$$

b) between 10 and 20 minutes:

$$\hookrightarrow P(10 < X < 20) = F(20) - F(10)$$

$$= 1 - e^{-20/10} - 1 - e^{-10/10} = e^{-2} - e^{-1} = 0.2325$$

c) more than an additional 7 minutes, given that you've already waited for over 10.

$$\hookrightarrow P(X > 10 + 7 \mid X > 10) = P(X > 7)$$

$$= 1 - P(X \leq 7) = 1 - F(7) = 1 - (1 - e^{-7/10})$$

$$= e^{-7/10} = 0.4966$$

Normal Distribution: a continuous RV X is said to have a normal distribution if X has the pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are parameters.

We write $X \sim N(\mu, \sigma^2)$

For a continuous RV X where $X \sim N(\mu, \sigma^2)$, we write:

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2$$

- the mean, μ , shifts the distribution along the x -axis.

$\hookrightarrow \mu$ is a location parameter.

- the variance, σ^2 , stretches out or pulls in the distribution

$\hookrightarrow \sigma^2$ is a scale parameter.

- The standard normal distribution is $Z \sim N(0, 1)$.

↳ every normal distribution is a version of the standard normal distribution that has been transformed / scaled!

• The cdf of the normal distribution $N(\mu, \sigma^2)$ is:

$$F(x) = \int_{-\infty}^x f(u) du = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{u-\mu}{\sigma})^2} du, \quad x \in \mathbb{R}$$

↳ horrible to solve, so let's use a different method!

Finding Normal Probabilities via a $N(0, 1)$ table

↳ let $x \sim N(\mu, \sigma^2)$, and define:

$$z = \frac{x - \mu}{\sigma}.$$

Then, $z \sim N(0, 1)$, and

$$P(x \leq x) = P(z \leq \frac{x - \mu}{\sigma})$$

Example: let $z \sim N(0, 1)$. Find the following probabilities:

a) $P(z < 2.11)$

↳ from the table of $N(0, 1)$ cdf values, we get $P(z < 2.11) = 0.98257$.

b) $P(z > 1.06)$

↳ from the table, we get:

$$P(z > 1.06) = 1 - P(z < 1.06) = 1 - 0.85543 \\ = 0.14457$$

c) $P(z < -1.06)$

Since a $N(0, 1)$ is symmetrical about the y-axis,
 $P(z < -1.06) = P(z > 1.06) = 0.14457$!

d) $P(-1.06 < z < 2.11)$

$$\hookrightarrow P(-1.06 < z < 2.11) = P(z < 2.11) - P(z < -1.06) \\ = 0.982 - 0.144 = 0.838$$

Example: let $z \sim N(0, 1)$.

a) Find 85% percentile of z

\hookrightarrow the 85% percentile of z is the level u such that $F(u) = 0.85$. $\therefore u = F^{-1}(0.85)$.

From the table, we see $u = 1.0364$.

b) Find a number b such that $P(z > b) = 0.9$

\hookrightarrow since 90% of the distribution lies to the right of b , and since $N(0, 1)$ is symmetrical, we know b must be negative!

$$P(z < b) = 1 - P(z > b) = 1 - 0.9 = 0.1$$

$$\therefore P(Z > |b|) = 0.1 \rightarrow P(Z < |b|) = 0.9$$

$$\rightarrow |b| = F^{-1}(0.9) = 1.282 \rightarrow b = -1.282!$$

c) Find a number c such that $P(|Z| < c) = 0.95$

$$\hookrightarrow P(|Z| < c) = P(-c < Z < c) = 0.95$$

$$\rightarrow P(Z < c) = 0.95 + \frac{0.05}{2} = 0.975$$

$$\rightarrow c = F^{-1}(0.975) = 1.96$$

Example: $X \sim N(10, 2)$. Calculate $P(|X-10| \leq 3)$

For a RV $X \sim N(10, 2)$, it follows that $\frac{X-10}{\sqrt{2}} \sim N(0, 1)$

$$P(|X-10| \leq 3) = P(-3 \leq X-10 \leq 3)$$

$$= P\left(-\frac{3}{\sqrt{2}} \leq \frac{X-10}{\sqrt{2}} \leq \frac{3}{\sqrt{2}}\right)$$

$$= P(-2.12 \leq Z \leq 2.12)$$

$$= P(Z \leq 2.12) - (1 - P(Z \leq 2.12))$$

$$= 2 \cdot P(Z \leq 2.12) - 1$$

$$= 2 \cdot 0.983 - 1$$

$$= 0.966!$$

Example: $X \sim N(-7, 14)$. Calculate $P(|X+7| \geq 8)$

For an RV $X \sim N(-7, 14)$, it follows that $\frac{X+7}{\sqrt{14}} \sim N(0, 1)$

$$\begin{aligned} P(|X+7| \geq 8) &= P\left(\frac{|X+7|}{\sqrt{14}} \geq \frac{8}{\sqrt{14}}\right) \\ &= P(|Z| \geq 2.138) \quad \text{bc symmetrical!} \\ &= 2 \cdot P(Z \geq 2.138) \\ &= 2 \cdot (1 - P(Z < 2.138)) \\ &= 2 \cdot (1 - 0.9837) = 0.032. \end{aligned}$$

Example: IQ is normally distributed. The mean is 100 and the standard deviation is 15.

a) What percentage of the population has IQ scores of 140 or more?

↪ let the RV X represent the IQ score $\rightarrow X \sim N(100, 15^2)$

$$P(X \geq 140) = P(X - 100 \geq 140 - 100)$$

$$= P\left(\frac{x-100}{15} \geq \frac{140-100}{15}\right)$$

$$\begin{aligned} &= P(Z \geq \frac{8}{3}) \\ &= 1 - P(Z < \frac{8}{3}) \\ &= 1 - P(Z < 2.66) \\ &= 1 - 0.996 = 0.004. \end{aligned}$$

b) What percentage of the population have IQ scores between 80 and 120?

$$\hookrightarrow P(80 < x < 120)$$

$$\begin{aligned}
 P(80 < x < 120) &= P(80 - 100 < x - 100 < 120 - 100) \\
 &= P\left(\frac{-20}{15} < \frac{x - 100}{15} < \frac{20}{15}\right) \\
 &= P(-\frac{4}{3} < z < \frac{4}{3}) \\
 &= P(z < \frac{4}{3}) - P(z < -\frac{4}{3}) \\
 &= P(z < \frac{4}{3}) - P(z > \frac{4}{3}) \\
 &= P(z < \frac{4}{3}) - (1 - P(z < \frac{4}{3})) \\
 &= 2P(z < \frac{4}{3}) - 1 \\
 &= 2 \cdot 0.908 - 1 = 0.816
 \end{aligned}$$

c) solve for the 95th percentile, which is the IQ level such that 95% of the population has an IQ below it.

$$P(x \leq \infty) = P\left(\frac{x - 100}{15} \leq \frac{\infty - 100}{15}\right) = 0.95$$

$$\rightarrow \frac{\infty - 100}{15} = F^{-1}(0.95) = 1.6449$$

$$\hookrightarrow x = (15)(1.6449) + 100 = 124.67$$

Multivariate Probability Distributions (chapter 4)

multiple RVs can be denoted by $x, y, z / x_1, x_2, x_3$

The joint pmf of 2 discrete RVs: suppose x and y are two discrete random variables. The

Joint probability mass function of (X, Y) is:

$$f(x, y) = P(\{X=x\} \cap \{Y=y\})$$

as shorthand, $f(x, y) = P(X=x, Y=y)$

Properties:

- 1) $f(x, y) \geq 0 \quad \forall (x, y)$
- 2) $\sum_{\text{all } x} f(x, y) = \sum_{\text{all } y} f(x, y) = 1$

more generally, $f(x_1, x_2, \dots, x_n) = P(X_1=x_1, X_2=x_2, \dots, X_n=x_n)$

Example: I either get a small or a medium coffee along with two or three timbits. Let X be the size of the coffee and let Y be the number of timbits. The joint pmf of what I buy is shown below:

		coffee	
		S	M
timbits	2	0.1	0.2
	3	0.3	C

Show that the probability I buy a medium coffee and 3 timbits is 0.4.

$$\sum P(X=x, Y=y) = 1 \rightarrow \sum_{\text{all } x} \sum_{\text{all } y} P(X=x, Y=y) = 1$$

$$\begin{aligned} & \rightarrow P(X=S, Y=2) + P(X=S, Y=3) + P(X=M, Y=2) + P(X=M, Y=3) \\ & = 0.1 + 0.3 + 0.2 + C = 1 \end{aligned}$$

$$\hookrightarrow C = 0.4.$$

Marginal Distribution: given the joint pmf of X and Y , the marginal probability function of X is:

$$f_X(x) = \sum_{\text{all } y} f(x, y)$$

Similarly, the marginal probability function of Y is:

$$f_Y(y) = \sum_{\text{all } x} f(x, y)$$

Note: always check that these add up to 1!

Example: X and Y have a joint probability function given by:

		x		
		0	1	2
y	0	0.2	0.3	0.1
	2	0.25	0.13	0.02

Compute the marginal probability function of X .

$$P(X=0) = P(X=0, Y=1) + P(X=0, Y=2) = 0.2 + 0.25 = 0.45$$

$$P(X=1) = P(X=1, Y=1) + P(X=1, Y=2) = 0.3 + 0.13 = 0.43$$

$$P(X=2) = P(X=2, Y=1) + P(X=2, Y=2) = 0.1 + 0.02 = 0.12$$

also we have that $\sum_{x=0,1,2} f_X(x) = 1$!

Independent Random Variables: suppose that X and Y are discrete random variables with

joint probability mass function $f(x, y)$ and marginal probability functions $f_x(x)$ and $f_y(y)$. RVs X and Y are independent if and only if:

$$f(x, y) = f_x(x) f_y(y) \text{ for pairs of values } (x, y)$$

or equivalently,

$$P(X=x, Y=y) = P(X=x) P(Y=y) \text{ for } (x, y).$$

Example: X and Y have the following joint pmf:

		x			$f(y)$	
		0	1	2		
y	0	0.2	0.3	0.1	0.6	
	2	0.25	0.13	0.02	0.4	
		$f(x)$	0.45	0.43	0.12	1

Are X and Y independent?

↳ for X and Y to be independent, we must have $f(x, y) = f_x(x) f_y(y)$ for (x, y) pairs.

let's just check $f(0, 0) = f_x(0) f_y(0)$

$$\hookrightarrow 0.2 = 0.45 \times 0.6 = 0.27 \rightarrow 0.2 \neq 0.27!$$

$\therefore f(x, y) \neq f_x(x) f_y(y)$, so they're dependent!

Conditional Probability Functions

↳ the conditional probability of X given $Y=y$ is:

$$f(x|y) = P(X=x | Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{f(x,y)}{f_y(y)}.$$

provided that $P(Y=y) > 0$.

- Note that this conditional pf is defined over all x in the range of the RV X with y fixed.
- Also, $f(x|y) \geq 0$ and $\sum_{\text{all } x} f(x|y) = 1$!

Example: continuing the coffee problem:

		coffee		
		S	M	
timbits	2	0.1	0.2	
	3	0.3	0.4 ($= c$)	
$f(x)$		0.4	0.6	

Find the conditional probability of Y , given $X=M$:

$$P(Y=2 | X=M) = \frac{P(Y=2, X=M)}{P(X=M)} = \frac{0.2}{0.6} = 1/3$$

$$P(Y=3 | X=M) = \frac{P(Y=3, X=M)}{P(X=M)} = \frac{0.3}{0.6} = 1/2.$$

Conditional Probability Function for 2 Independent RVs

↓

since $f(x,y) = f_x(x) f_y(y)$,

$$f(x|y) = P(X=x, Y=y) = \frac{f(x,y)}{f_y(y)} = \frac{f_x(x) f_y(y)}{f_y(y)} = f_x(x)$$

↳ the conditional pmf becomes the marginal pmf in the case of independence!

Functions of Random Variables: Suppose that

$g: \mathbb{R}^2 \rightarrow \mathbb{R}$, and X and Y are RVs with a joint pmf $f(x, y)$. The probability function of RV $U = g(X, Y)$ is given by:

$$f_U(u) = P(U = u) = \sum_{\substack{\text{all } (x, y) \\ g(x, y) = u}} f(x, y)$$

- 1) If $X \sim \text{Poisson}(\mu_1)$ and $Y \sim \text{Poisson}(\mu_2)$, and X and Y are independent, then $X + Y \sim \text{Poisson}(\mu_1 + \mu_2)$
- 2) If $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$, and X and Y are independent, then $X + Y \sim \text{Binomial}(n+m, p)$

Example: suppose X and Y have joint pmf:

		x			$f(y)$
		0	1	2	
y	0	0.2	0.3	0.1	0.6
	2	0.25	0.13	0.02	0.4
		$f(x)$	0.45	0.43	0.12
					1

Let $U = X + Y$.

a) Calculate the probability function of U .

the possible values of RV U are 0, 1, 2, 3, and 4. Therefore, we have:

$$P(U=0) = P(X=0, Y=0) = f(0,0) = 0.2$$

$$P(U=1) = P(X=1, Y=0) = f(0,1) = 0.3$$

$$\begin{aligned}P(U=2) &= P(X=2, Y=0) + P(X=0, Y=2) = f(0,2) + f(2,0) \\&= 0.35\end{aligned}$$

$$P(U=3) = P(X=1, Y=2) = f(1,2) = 0.13$$

$$P(U=4) = P(X=2, Y=2) = f(2,2) = 0.02$$

we have $\sum_{u=1,2,3,4,0} f(u) = 1$!

b) Find $P(X+Y > 3)$

$$P(U > 3) = P(U = 4) = 0.02.$$

c) Find $P(XY = 0)$

We are looking for the event $X=0$ or $Y=0$, which would be the probability of the union of these two events.

$$\begin{aligned}\hookrightarrow P(XY=0) &= P(X=0) + P(Y=0) - P(X=0, Y=0) \\&= 0.6 + 0.45 - 0.2 = 0.85.\end{aligned}$$

Expectation for Multivariate Distributions:

Suppose X and Y are discrete RVs with joint pmf $f(x, y)$. Then for a function $g: \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$E[g(X, Y)] = \sum_{\text{all } (x,y)} g(x, y) f(x, y)$$

More generally, for n discrete random variables

X_1, X_2, \dots, X_n with joint pmf $f(x_1, x_2, \dots, x_n)$, if
 $g: \mathbb{R}^n \rightarrow \mathbb{R}$, then

$$E[g(x_1, x_2, \dots, x_n)] = \sum_{\text{all } (x_1, x_2, \dots, x_n)} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n)$$

Properties of Expected Value:

$$E(X+Y) = \sum_x x f_X(x) + \sum_y y f_Y(y) = E(X) + E(Y)$$

and in general,

$$E[\alpha g_1(x, y) + b g_2(x, y)] = \alpha E[g_1(x, y)] + b E[g_2(x, y)]$$

Example: X and Y have the joint pmf:

		x		
$f(x, y)$		0	1	2
y	0	0.2	0.3	0.1
	2	0.25	0.13	0.02

Compute $E(XY)$.

		x		
$g(x, y) = xy$		0	1	2
y	0	0	0	0
	2	0	2	4

$$\begin{aligned} E(XY) &= 0 + 0 + 0 + 0 + 2(0.13) + 4(0.02) \\ &= 0.34 \end{aligned}$$

Multinomial Distribution Definition: Physical Setup

Consider an experiment which is repeated independently n times.

↳ Each trial has k possible outcomes

↳ let p_1, p_2, \dots, p_k denote the probability of each outcome ($p_1 + \dots + p_k = 1$)

↳ let RV X_i denote the number of outcomes of type i for $i = 1, \dots, k$.

We say X_1, X_2, \dots, X_k have a multinomial distribution with parameters n and p_1, p_2, \dots, p_k .

Shorthand : $(X_1, X_2, \dots, X_n) \sim \text{Multinomial}(n; p_1, p_2, \dots, p_n)$

Multinomial distribution is a generalization of binomial distribution.

Multinomial Distribution Definition: PMF

If X_1, X_2, \dots, X_k have a joint multinomial distribution with parameters n and p_1, \dots, p_n , then their joint probability function is:

$$f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

where $x_i = 0, 1, \dots, n$ and $x_1 + x_2 + \dots + x_k = n$.

Note: the term $\frac{n!}{x_1! x_2! \dots x_k!}$ is called a multinomial coefficient. It counts the number of ways we can arrange x_1 items of type-1, x_2 items of type-2, ..., x_k items of type-k.

x_k items of type- k among the k type of total items.