Note # 3: Probability

Problem 1. You decide to flip a fair coin many times and calculate the proportion of heads.

a. On the first ten-coin flips, 9 of them are heads and 1 is tails. What is the proportion of heads on the first 10 flips?

b. You flip the coin another 10 times-half are heads and half are tails. What is the proportion of heads on the first 20 flips?

c. You flip the coin another 100 times-half are heads and half are tails. What is the proportion of heads on the first 120 flips?

d. You flip the coin another 1000 times-half are heads and half are tails. What is the proportion of heads on the first 1120 flips?

e. You flip the coin another 10,000 times-half are heads and half are tails. What is the proportion of heads on the first 11,120 flips?

f. What does this tell us about long term probabilities?

Solution:

a. $P(Heads) = \frac{9}{10} = 0.90$

b. $P(Heads) = \frac{9+5}{20} = \frac{14}{20} = 0.70$

c. $P(Heads) = \frac{14+50}{120} = \frac{64}{120} = 0.53$

d. $P(Heads) = \frac{64+500}{1120} = \frac{564}{1120} = 0.504$

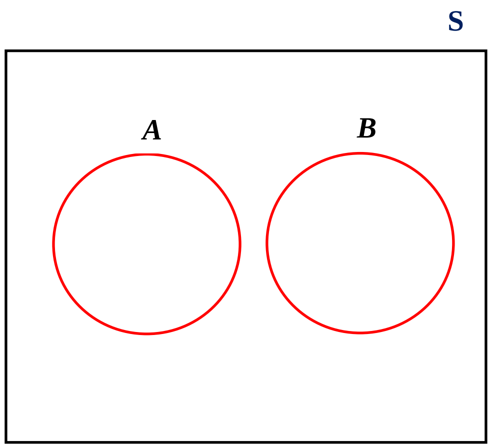
e. $P(Heads) = \frac{564+5000}{11{,}120} = \frac{5564}{11{,}120} = 0.5004$

f. Even though we had "extra heads" in the 1st ten flips, the proportion of heads still approaches what we would expect to happen [$P(Heads) = 0.50$] as our number of coin flips increases.

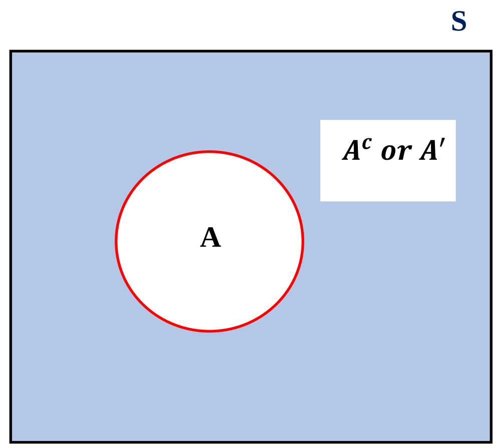
Problem 2. Use the space below to illustrate the following four concepts with Venn Diagrams: Disjoint Events, Complementary Events, Union of Two Events, Intersection of Two Events.

Solution:

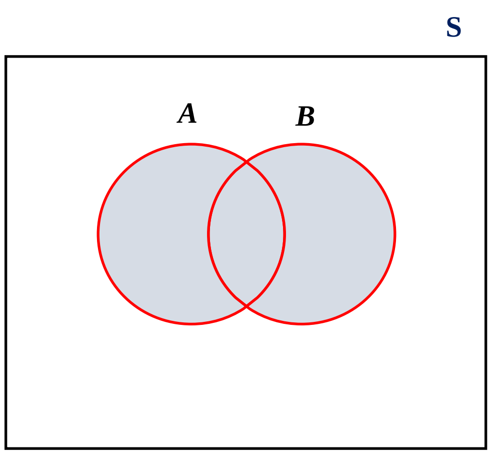
Disjoint/Mutually Exclusive



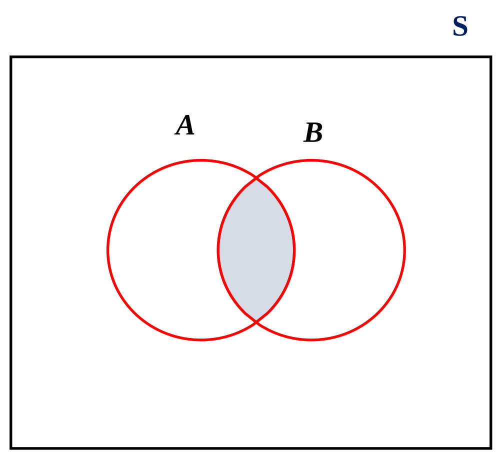
Complementary Events



Union $[A \cup B]$



Intersection $[\boldsymbol{A} \cap \boldsymbol{B}]$



Problem 3. Many times, when we are looking at probability problems, we are not asking about just one event, but rather a combination of events. Let's look at an example where we are looking at two different events: A student is taking a history class that has one midterm and one final exam. They believe that the chance of them getting an A on the midterm is 50%. They also believe the chance of them getting an A on the final exam is 50%. The student states that their chance of getting an A on at least one of the two tests is 100%. Is their reasoning correct? Why or why not? If it is incorrect, how would you solve for the probability of them getting an A on at least one of the two tests?

Solution:

- Getting A on MT and getting A on final are not disjoint.

$\because P(A \cup B) = P(A) + P(B) - P(A \cap B)$, if A and B are disjoint: $P(A \cap B) = 0$

$P(get\ an\ A\ on\ MT) = 0.50$ and $P(get\ an\ A\ on\ final) = 0.50$

$P(get\ an\ A\ on\ MT \cup get\ an\ A\ on\ final) = P(A\ on\ MT) + P(A\ on\ final) - P(A\ on\ MT \cap A\ on\ final)$

- We would need to subtract the probability of the intersection.

Problem 4. Let's look at another example of a probability question involving two events. A software company provides an email filtering service to protect email users from spam. The company has advertised that their software is 95% accurate. This could mean one of four things. Listed below are the four things this could mean. Write out each as a conditional probability statement.

a. 95% of the blocked emails are spam.
b. 95% of spam emails are blocked.
c. 95% of the valid emails are allowed through.
d. 95% of the emails allowed through are valid.

Solution:

$S = Spam \quad S^c = Valid$

$B = Blocked \quad B^c = Allowed$

a. $P(S|B) = 0.95$
b. $P(B|S) = 0.95$
c. $P(B^c \mid S^c) = 0.95$
d. $P(S^c \mid B^c) = 0.95$

Problem 5. While the most common die has six sides, there are a number of other varieties of die that are used in various games. One such die is the four-sided die. Assume we have two fair four-sided die, each labeled with the numbers from one to four.

a. What do we mean by a fair die?
b. What is the sample space for rolling one die?
c. What is the sample space for rolling two die?
d. Assume you roll both die and add them together. What is the sample space?
e. Assume you roll both die and add them together. What is the probability the sum is greater than 6? It may be a good idea to use the sample space from part c to solve this.

Solution:

a. Probability of each outcome is equal $P(1) = P(2) = P(3) = P(4) = \frac{1}{4} = 0.25$
b. $S = \{1,2,3,4\}$
c. $S = \{(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)\}$
d. $S = \{2,3,4,5.6,7,8\}$
e. $P(sum > 6) = P(sum = 7 \cup sum = 8) = P(sum = 7) + P(sum = 8) = \frac{2}{16} + \frac{1}{16} = \frac{3}{16} = 0.1875$

Problem 6. Many times, we organize our information about a random event by creating a table or diagram that describes the probability distribution. The table below shows the probability distribution of Educational Attainment for Americans 25 years and older in 2019, as reported by the Census Bureau.

x	**None**	**No HS Diploma**	**HS Diploma**	**Some College**	**Assoc.**	**Bach**	**Bach+**
P(x)	0.003	0.096	0.281	0.157	0.103	0.225	0.135

a. What type of variable is educational attainment?

b. What two properties must be true for a probability distribution to be valid? Is this a valid probability distribution?

c. What is the probability that a randomly selected person does not have a high school diploma?

d. What is the probability that a randomly selected person has some college degree?

Solution:

a. Categorical - ordinal

b. Sum =1, Each probability is between 0 and 1. Yes, it is a valid probability distribution.

c. $P(didn't\ graduate\ HS) = P(none \cup no\ HS\ diploma) = P(none) + P(no\ HS\ diploma) =$

$0.003 + 0.096 = 0.099$

d. $P(some\ college\ degree) = P(assoc.\ \cup bach \cup bach+) = P(assoc.) + P(bach) +$

$P(bach+) = 0.103 + 0.225 + 0.135 = 0.463$

Problem 7. Let's continue looking at the 2019 Educational Attainment data, however, now let's look at it split up by gender.

	None	No HS Diploma	HS Diploma	Some College	Assoc.	Bach	Bach+	Total
Males	0.002	0.048	0.141	0.075	0.045	0.107	0.063	0.482
Females	0.002	0.047	0.140	0.082	0.058	0.118	0.072	0.518
Total	0.004	0.095	0.281	0.157	0.103	0.225	0.135	1

a. Are having an associate's degree and being female mutually exclusive?

b. What is the probability that a randomly selected person has an associate's degree or is a female?

c. What is the probability that a randomly selected person has an associate's degree and is a female?

d. What is the probability that a randomly selected female has an associate's degree?

e. What is the probability that a randomly selected person has an associate's degree if they are a male?

f. Does it appear as though gender and educational attainment are dependent or independent? Why?

Solution:

a. Mutually exclusive/disjoint $\rightarrow P(intersection) = 0$

$P(female \cap accoc.) = 0.058 \neq 0$ a person can be female and have an associate's degree, so the events are not disjoint

b. $P(assoc. \cup female) = P(assoc.) + P(female) - P(assoc. \cap female) = 0.103 + 0.518 - 0.058 = 0.563$

c. $P(assoc. \cap female) = 0.058$

d. $P(assoc. | female) = \frac{P(assoc. \cap female)}{P(female)} = \frac{0.058}{0.518} = 0.112$

e. $P(assoc. | male) = \frac{P(assoc. \cap male)}{P(male)} = \frac{0.045}{0.482} = 0.093$

f. $\because P(assoc.) = 0.103$, $P(assoc. | female) = 0.112$, $P(assoc. | male) = 0.093$

we are not sure, there is a difference in $P(assoc.)$ for different genders, it is a small difference based on a sample.

Problem 8. A professor surveyed his students and asked them, on a typical weekend, how many days do you spend studying. The probability distribution is shown below.

x	**0**	**1**	**2**	**3**
P(x)	0.10	0.35	0.25	0.30

- **a.** Verify that this is a valid probability distribution.
- **b.** What is the probability that a randomly selected individual studied at least 2 days?
- **c.** What is the probability that a randomly selected individual studied no more than 2 days?
- **d.** What is the expected value?

Solution:

- **a.** All probabilities are between 0 and 1, sum of all probabilities = 1 [0.10+0.35+0.25+0.30 =1].
- **b.** $P(x \geq 2) = P(x = 2) + P(x = 3) = 0.25 + 0.30 = 0.55.$
- **c.** $P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2) = 0.10 + 0.35 + 0.25 = 0.70.$
- **d.** $\mu = E(x) = \sum x_i \times P(X = x_i) = (0)(0.10) + (1)(0.35) + (2)(0.25) + (3)(0.30) = 1.75$

Problem 9. An airline charges the following baggage fees: $25 for the first bag and $35 for the second bag. They do not allow individuals to bring more than two bags. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage, and 12% have two pieces.

- **a.** Draw the probability model for the amount of money spent by passengers on luggage.
- **b.** What is the expected value?

Solution:

a.

Event	$\boldsymbol{x}$	$\boldsymbol{P(X = x_i)}$	$\boldsymbol{x \times P(X = x_i)}$
No Bags	$\mathbf{\$0}$	$\mathbf{0.54}$	$\mathbf{(0) \times (0.54) = \$0}$
1 Bag	$\mathbf{\$25}$	$\mathbf{0.34}$	$\mathbf{(25) \times (0.34) = \$8.50}$
2 Bags	$\mathbf{\$60}$	$\mathbf{0.12}$	$\mathbf{(60) \times (0.12) = \$7.20}$

b. $\mu = E(x) = \sum x_i \times P(X = x_i) = \$0 + \$8.50 + \$7.20 = \$15.7$