Data Analytics

ECON 1008, Semester 1, 2019

Giulio Zanella
University of Adelaide
School of Economics

CHAPTER 5, CONTINUED...

Numerical descriptive measures

Coefficient of Variation

The ***coefficient of variation*** is the standard deviation divided by the mean.

$$\text{Sample coefficient of variation: } \mathrm{cv} = \frac{\mathrm{s}}{\bar{\mathrm{x}}}$$

$$\text{Population coefficient of variation: } \mathrm{CV} = \frac{\sigma}{\mu}$$

Coefficient of Variation

$$\text{Sample coefficient of variation: } cv = \frac{s}{\bar{x}}$$

$$\text{Population coefficient of variation: } CV = \frac{\sigma}{\mu}$$

It's **a proportionate measure of variation**.

Why? A standard deviation of 10 may be perceived as large when the mean is 100, but small when the mean is 500.

Measures of relative standing

Measures of relative standing are designed to provide information about the ***position*** of particular values ***relative*** to the entire data set.

Percentile: the p^{th} percentile is the value for which p percent are *less than* that value and (100-p)% are *greater* than that value (**Q.: what's the median?**)

If you scored in the 60^{th} percentile on your final exam, then 60% of the other students' scores were *below* yours, while 40% of scores were *above* yours.

Percentiles

The p^{th} percentile of a set of measurements is the value for which

- at most p% of the measurements are less than that value
- at most (100-p)% of all the measurements are greater than that value.

For example, suppose 77 is the 68^{th} percentile of a statistics exam score. Then

Quartiles

We have special names for the 25th, 50th and the 75th percentiles, namely **quartiles** (because they divide the distribution in four quarters)

- First (lower) quartile, $\mathbf{Q_1}$ = 25th percentile (p_{25})
- Second (middle) quartile, $\mathbf{Q_2}$ = 50th percentile (p_{50}) (which is also the median)
- Third (upper) quartile, $\mathbf{Q_3}$ = 75th percentile (p_{75})

We can also convert percentiles into quintiles (fifths) and deciles (tenths).

Commonly Used Percentiles…

First (lower) decile	= 10^{th} percentile
First (lower) quartile, Q_1	= 25^{th} percentile
Second (middle)quartile,Q_2	= 50^{th} percentile
Third quartile, Q_3,	= 75^{th} percentile
Ninth (upper) decile	= 90^{th} percentile

NOTICE: if your exam mark places you in the 80th percentile, that doesn't mean you scored 80% on the exam – it means that 80% of your peers scored **lower** than you and 20% scored **higher** than you in the exam. It is about your position relative to others, not the actual mark.

Example

Find the quartiles of the following set of measurements

7, 18, 12, 17, 29, 18, 4, 27, 30, 2, 4, 10, 21, 5, 8

Example: Solution

First sort the measurements

$\boxed{2, 4, 4,}$ 5, $\boxed{7, 8, 10, 12, 17, 18, 18, 21, 27, 29, 30}$

15 measurements

The first quartile

At most (0.25)(15) = 3.75 measurements should appear below the first quartile. Check the first 3 measurements on the left hand side.

At most (0.75)(15)=11.25 measurements should appear above the first quartile. Check 11 measurements on the right hand side.

If the number of measurements is even, two measurements will remain unchecked. In this case choose the midpoint between these two measurements (exactly like we did for the median)

Location of Percentiles

Find the location of any percentile using the formula

$$L_P = (n+1)\frac{P}{100}$$

where L_P is the location of the P^{th} percentile

Think of easy cases, e.g., the median, to make sense of this formula. You will realise it's very intuitive!

Example

Calculate the 25th, 50th, and 75th percentile of the data:

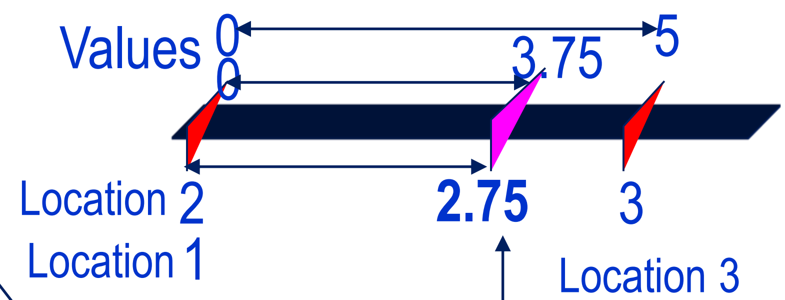
0, 7, 12, 5, 33, 14, 8, 0, 9, 22

Example: Solution

After sorting the data we have

	0,	0,	5,	7,	8,	9,	12,	14,	22,	33.
Location	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

$$L_{25} = (10+1)\frac{25}{100} = 2.75$$



The 2.75th location translates to the value

$$p_{25} = 0 + (.75)(5 - 0) = 3.75$$

2nd observation   3rd observation   2nd observation

Example: Solution…

$$L_{50} = (10+1)\frac{50}{100} = 5.5$$

The 50th percentile is halfway between the fifth and sixth observations (in the middle between 8 and 9), that is 8.5. That is,

$$p_{50} = 8 + (0.5)(9 - 8) = 8.5$$

5th observation

6th observation

Example: Solution…

$$L_{75} = (10+1)\frac{75}{100} = 8.25$$

The 75th percentile is one quarter of the distance between the eighth and ninth observation. That is

$$p_{75} = 14 + .25(22 - 14) = 16.$$

8th observation

9th observation

Location of Percentiles…

Please remember…

position 2.75

value 16

0 0 | 5 7 8 9 12 14 | 22 33

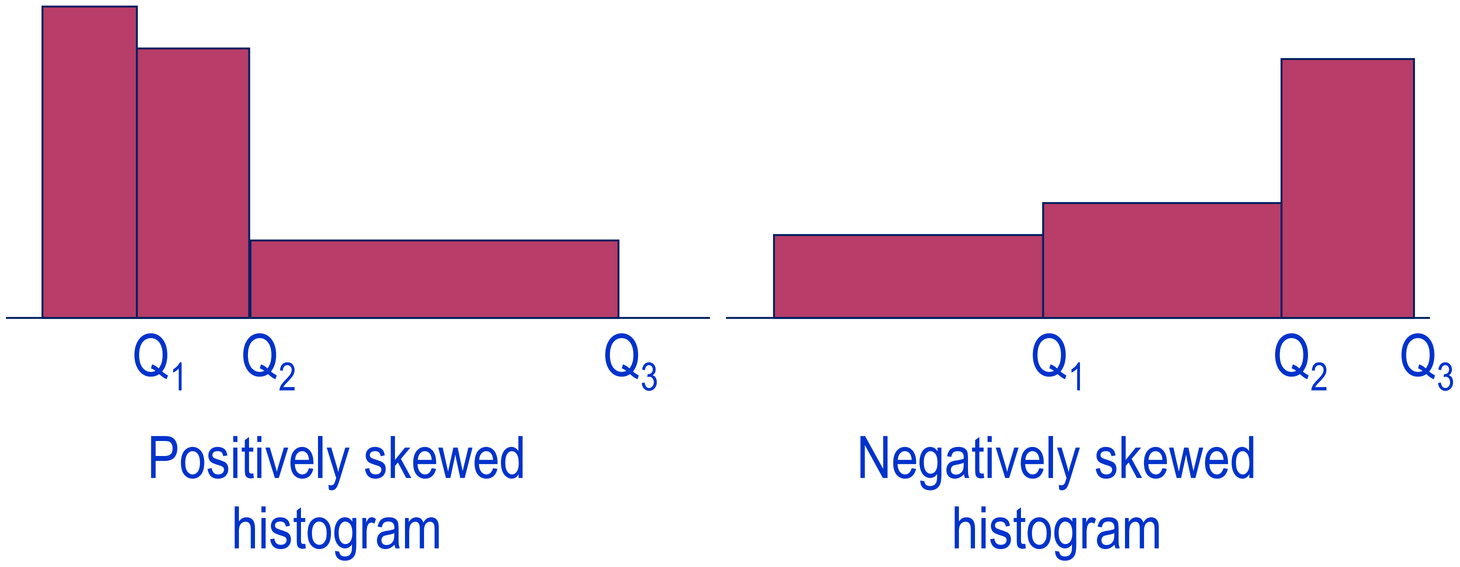
value 3.75

position 8.25

L_p determines the **position** in the data set where the percentile value lies, not the value of the percentile itself.

Quartiles and Variability

Quartiles can provide an idea about the shape of a histogram.



Interquartile Range…

The quartiles can be used to create another measure of variability, the ***interquartile range***, which is defined as follows:

$$\text{Interquartile Range (IQR)} = Q_3 - Q_1$$

The interquartile range measures the spread of the middle 50% of the observations.

Large values of this statistic mean that the 1^{st} and 3^{rd} quartiles are far apart, indicating a high level of variability.

Chapter 6

Probability

Assigning probabilities to events

Random experiment

A ***random experiment*** is a process or course of action whose outcome is uncertain.

Examples

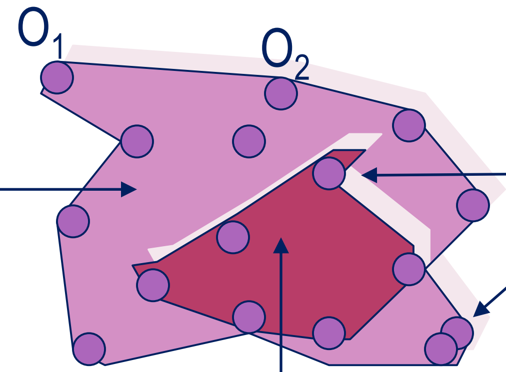
Experiment	*Outcomes*
Flip a coin	Heads and tails
Record statistics test marks	Numbers between 0 and 100
Measure the time to assemble a computer	Numbers from 0 and above

Sample Space

A list of all possible outcomes of a random experiment is called a sample space.

Sample Space: $S = \{O_1, O_2, \ldots, O_k\}$

Sample Space: $S = \{O_1, O_2, \ldots, O_k\}$



Sample space
A sample space of a random experiment is a list of all possible outcomes of the experiment. The outcomes must be mutually exclusive and exhaustive.

Event, A
An event is any collection of one or more possible outcomes.

Simple events
The individual outcomes are called *simple events*. Simple events cannot be further decomposed into constituent outcomes.

Our objective is to determine **P(*A*)**, the probability that event *A* will occur.

Approaches to Assigning Probabilities…

There are three ways to assign a probability, $P(O_i)$, to an outcome, O_i, namely:

1. ***Classical approach***: based on counting "possible" and "favourable" events.

2. ***Relative frequency***: assigning probabilities based on experimentation or historical data.

3. ***Subjective approach***: Assigning probabilities based on the assignor's (subjective) judgment.

Classical Approach

If an experiment has *n* possible outcomes, this method would assign a probability of 1/*n* to each outcome. It is necessary to determine the number of possible outcomes.

Experiment 1: Rolling a ***die***

Outcomes {1, 2, 3, 4, 5, 6}

Probabilities: Each sample point has a 1/6 chance of occurring.

Classical Approach…

Experiment 2: Rolling two dice and observing the total (X)

Outcomes: {X} = {2, 3, …, 12}

Examples:

P(X=2) = 1/36

P(X=6) = 5/36

P(X=10) = 3/36

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Relative Frequency Approach

Bits & Bytes Computer Shop tracks the number of desktop computer systems it sells over a month (30 days):

Desktops Sold	Number of Days
0	1
1	2
2	10
3	12
4	5

For example,

10 days out of 30 days

2 desktops were sold.

From this we can construct

the probabilities of an event

(i.e. the number of desktops sold on a given day)…

Relative Frequency Approach…

Desktops Sold	Number of Days	P(Desktops Sold)
0	1	1/30 = 0.03
1	2	2/30 = 0.07
2	10	10/30 = 0.33
3	12	12/30 = 0.40
4	5	5/30 = 0.17
	$\sum$ =30	$\sum$ = 1.00

'There is a 40% chance that 3 desktops will be sold on any given day.'

Subjective Approach

‘In the subjective approach we define probability as the degree of belief that we hold in the occurrence of an event.’

Example: Weather forecasting’s ‘P.O.P.’

‘Probability of Precipitation’ (or P.O.P.) is defined in different ways by different forecasters, but basically it’s a subjective probability based on past observations combined with current weather conditions, computed through a model.

POP 60% – based on current conditions, there is a 60% chance of rain (say).

Interpreting Probability

No matter which method is used to assign probabilities all will be interpreted in the relative frequency approach.

For example, a government lottery game where 6 numbers (of 49) are picked. The classical approach would predict the probability for any one number being picked as 1/49 = 2.04%.

We interpret this to mean that in the long run each number will be picked 2.04% of the time.

Assigning Probabilities

Given a sample space S = {O_1, O_2, …, O_n}, the following characteristics for the probability P(O_i) of the simple event O_i must hold:

1. $0 \le P(O_i) \le 1 \quad for \quad each \quad i$

2. $\sum_{i=1}^{n} P(O_i) = 1$

Probability of an event: the probability P(*A*) of event *A* is the sum of the probabilities assigned to the simple events contained in *A*.

Complement, union and intersection of events

We study methods to determine probabilities of events that result from ***combining*** other events in various ways.

There are several types of combinations and relationships between events:

- Complement of an event
- Intersection of events
- Union of events
- Mutually exclusive events
- Dependent and independent events
- Conditional event

Probability of Combinations of Events

If A and B are two events, then

$P(A \cup B)$ = P(A occur or B occur or both)

$P(A \cap B)$ = P(A and B both occur)

$P(\bar{A})$ = P(A does not occur)

$P(A|B)$ = P(A occurs given that B has occurred)

The union of A and B

The intersection of A and B

The complement event of A

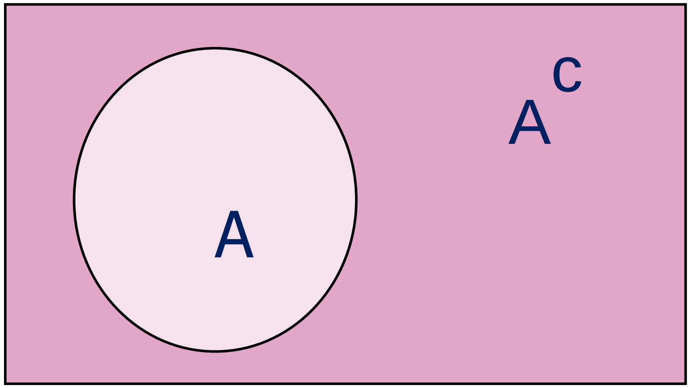
The conditional probability of A given B

Complement of an Event

The ***complement of event A*** is defined to be the event consisting of all sample points that are ‘not in A’.

Complement of A is denoted by $\bar{A}$ or A^c

The Venn diagram below illustrates the concept of a complement.



$P(A) + P(A^c) = 1$

Complement of an Event…

For example, the rectangle stores all the possible tosses of 2 coins S = {(H,H), (H,T), (T,H), (T,T)}

Let A = observing at least one head

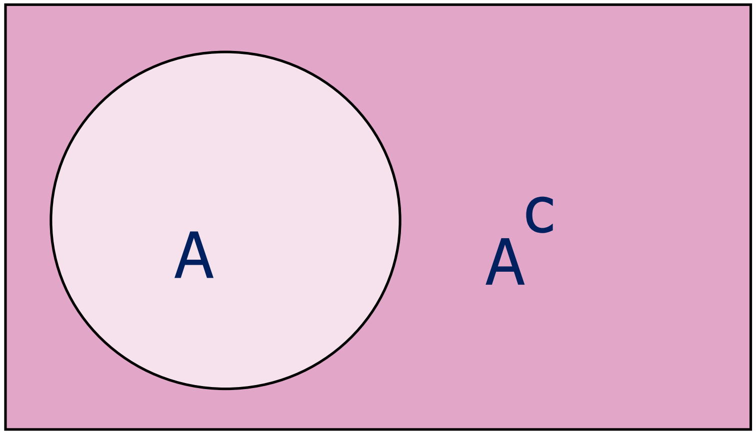
= {(H,H), (H,T), (T,H)}

A^c = {(T,T)}

$P(A) = ¾$, $P(A^c) = ¼$

$P(A) + P(A^c) = P(S) = 1$

$P(A^c) = 1 – P(A)$

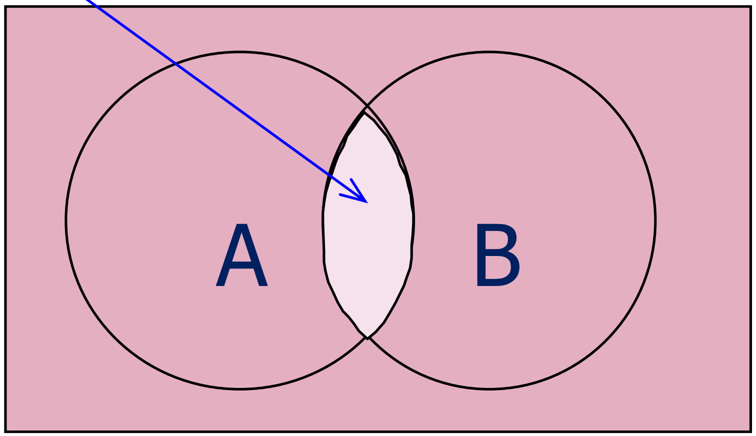


Intersection of two events

The ***intersection of events*** A and B is the set of all sample points that are in both $A \cap B$.

The intersection is denoted: **$A \cap B$**.

The ***joint probability*** of A and B is the probability of the intersection of A and B, i.e. $P(A \cap B)$.



Intersection of two events…

For example, consider all possible tosses of two dice.

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

Let A = tosses where first toss is 1

$= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}$

and B = tosses where the second toss is 5

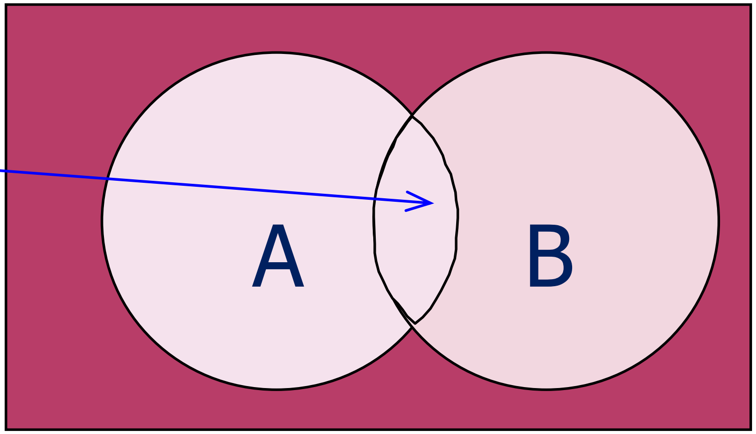
$= \{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\}$

The intersection, $A \cap B = \{(1,5)\}$.

The *joint probability* of A and B is

The probability of the intersection

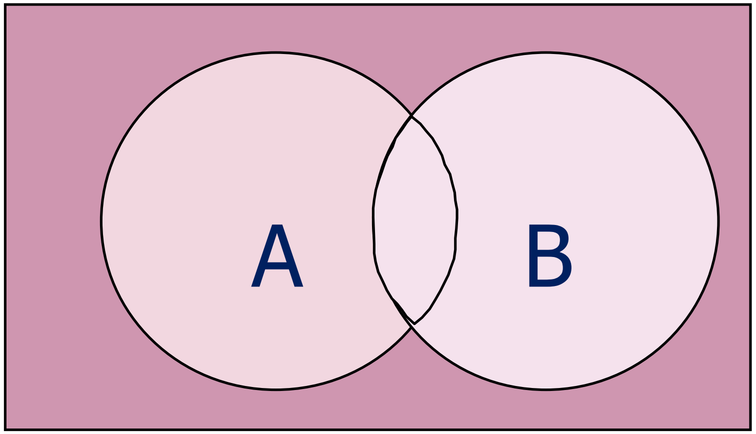
of A and B, i.e. $P(A \cap B) = 1/36$



Union of two events

The ***union of two events*** A and B, is the event containing all outcomes that are in A or B or both:

Union of A and B is denoted: $\mathbf{A \cup B}$



Union of two events…

For example, consider all possible tosses of two dice.

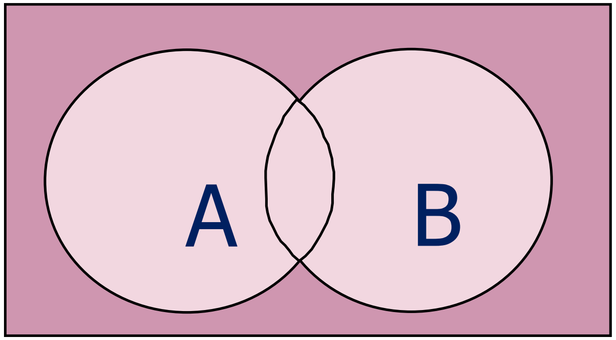
$S = \{(1,1),(1,2), \ldots,(6,6)\}$.

Let A = tosses where first toss is 1

$= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}$

and B = tosses where the second toss is 5

$= \{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\}$



Union of A and B, $A \cup B$ is

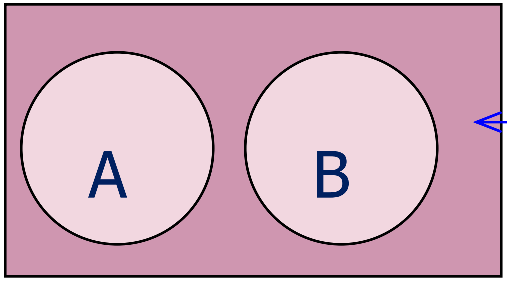
$$A \cup B = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,5),(3,5),(4,5),(5,5),(6,5)\}$$

The ***probability*** of $A \cup B$,

$P(A \cup B) = 11/36$

Mutually exclusive events

When two events are ***mutually exclusive*** (that is the two events have no outcomes in common), their joint probability is 0, hence:



A and B are mutually exclusive; there are no outcomes in common…

For example, consider all possible tosses of two dice.

S = {(1,1),(1,2), …,(6,6)}.

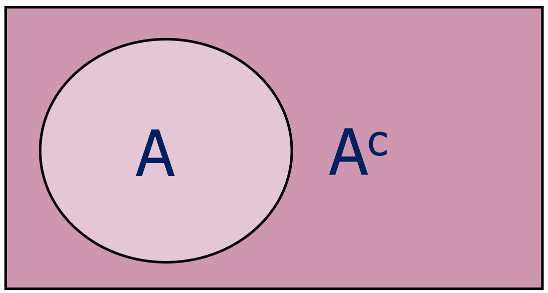
A = tosses totalling 9 = {(3,6),(6,3),(4,5),(5,4)} and

B = tosses totalling 11 = {(5,6),(6,5)}

Therefore, $A \cap B = \{\}$ = Empty set. $P(A \cap B) = 0$

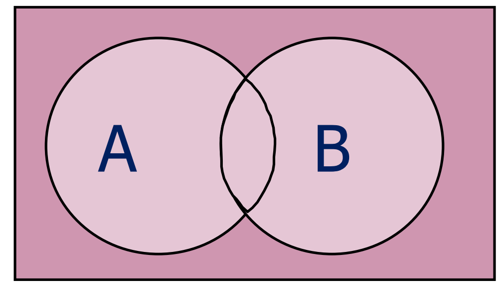
Basic relationships of probability

Complement of event



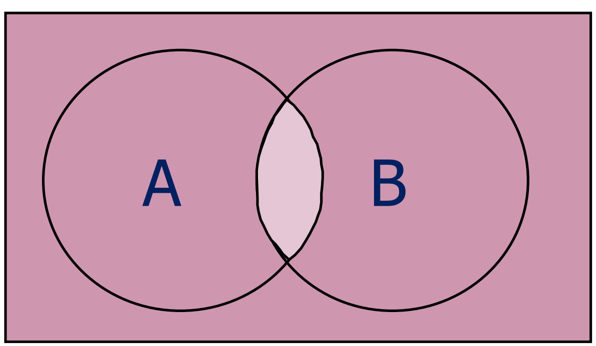
$$P(A^c) = 1 - P(A)$$

Union of events

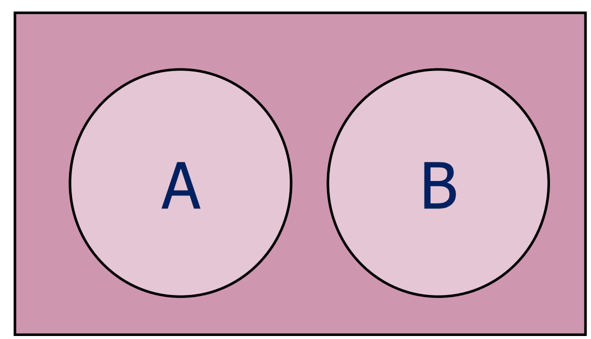


$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Intersection of events



Mutually exclusive events



Example 1

The number of spots turning up when a 6-sided die is tossed is observed.

Consider the following events:

A: The number observed is at most 2.

B: The number observed is an even number.

C: The number 4 turns up.

Answer the following questions.

Example 1

a. Define the sample space for this random experiment and assign probabilities to the simple events.
b. Find $P(A)$.
c. Find $P(A^C)$.
d. Are events A and C mutually exclusive?
e. Find $P(A \cup C)$.
f. Find $P(A \cap B)$.
g. Find $P(A \cup B)$.
h. Find $P(C|B)$.

Example 1: Solution

a) Define the sample space for this random experiment and assign probabilities to the simple events.

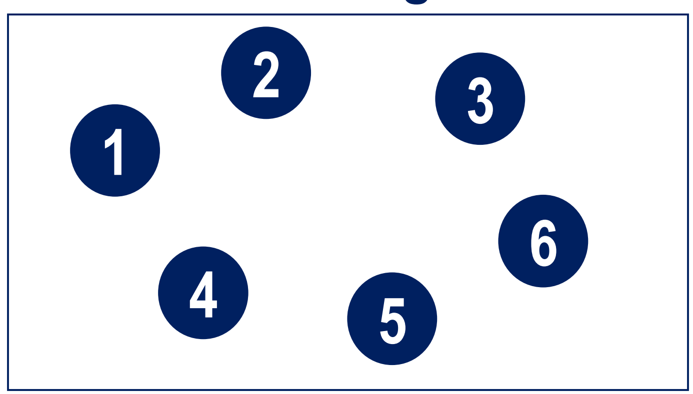
Solution

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

Each simple event is equally likely to occur, thus,

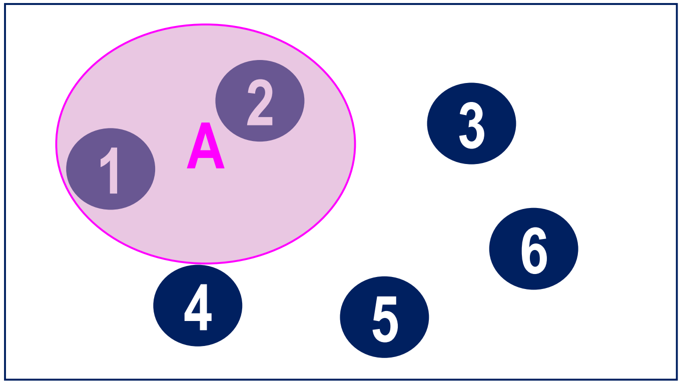
$P(1) = P(2) = \ldots = P(6) = 1/6.$

A Venn diagram



Example 1: Solution…

b) Find P(A).



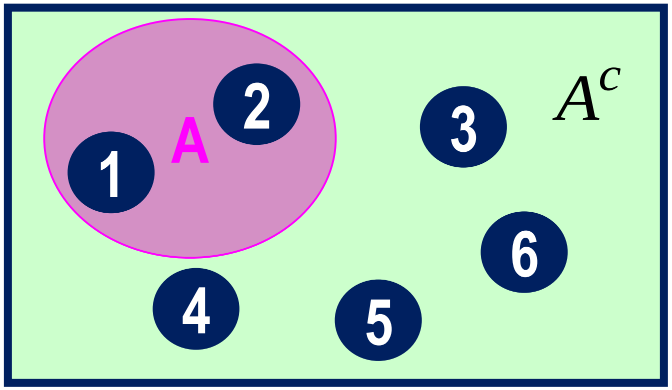
Solution

$$P(A) = \mathrm{P}\{1, 2\}$$

$$P(A) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 0.33$$

Example 1: Solution…

c) Find $P(A^c)$.



Solution

The complement of event A is $A^C = \{3,4,5,6\}$

$$P(A^c) = P(3) + P(4) + P(5) + P(6) = \frac{4}{6} = 0.67$$

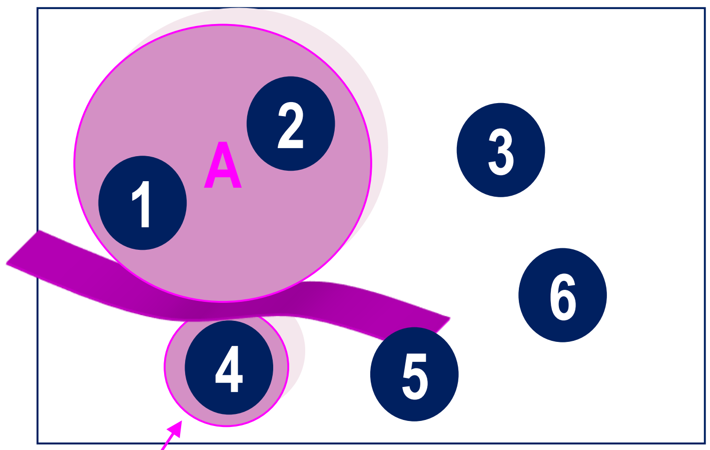
Example 1: Solution…

d) Are events A and C mutually exclusive?

Solution

A = {1,2}, C = {4}

There is no overlap between the two regions.



Event C

Events A and C are mutually exclusive because they cannot occur simultaneously.

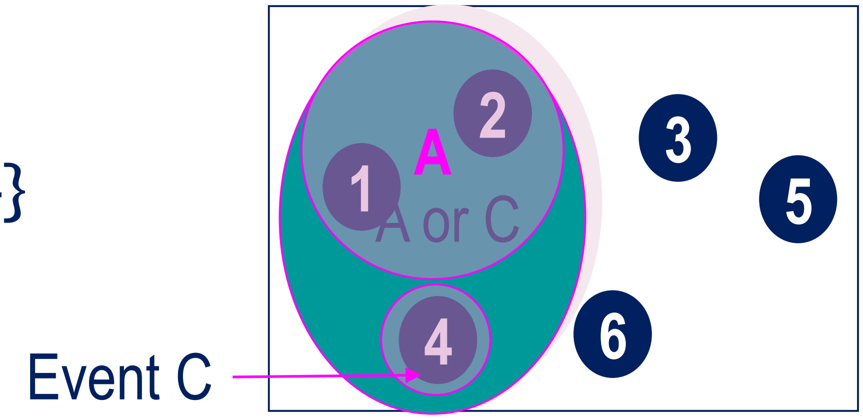
Example 1: Solution…

e) Find P(A ∪ C).

Solution

A = {1, 2}, C = {4}

A ∪ C = {1, 2, 4}



$$P(A \cup C) = P(1,2,4) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 0.5$$

or, because A and C are mutually exclusive,

$$P(A \cup C) = P(A) + P(C) = \frac{2}{6} + \frac{1}{6} = \frac{3}{6} = 0.5$$

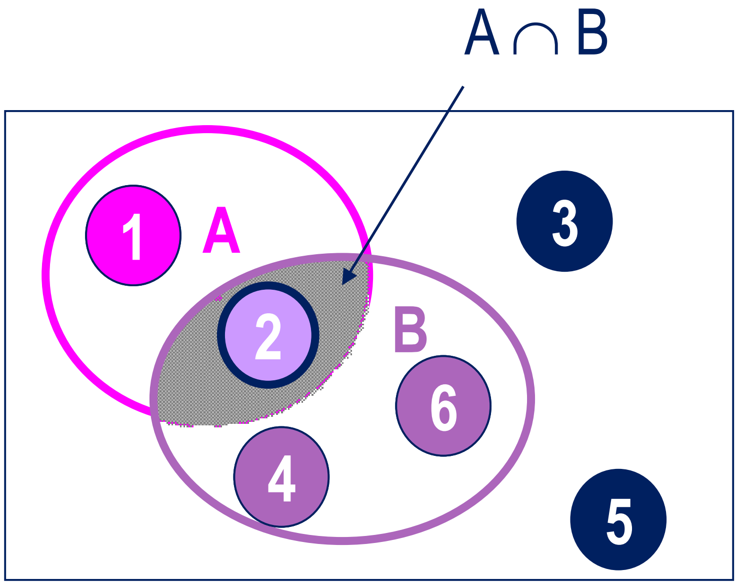
Example 1: Solution…

f) Find P(A ∩ B).

Solution

A = {1, 2}, B = {2, 4, 6}

A ∩ B = {2}



$$P(A \cap B) = P(2) = \frac{1}{6} = 0.167$$

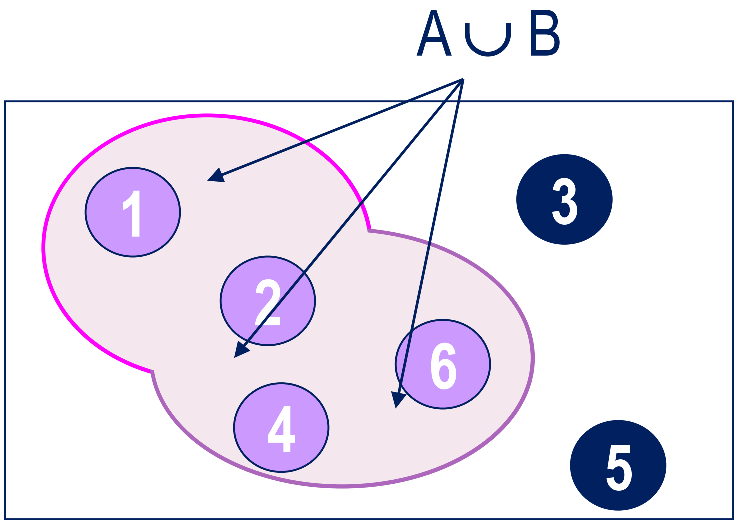
Example 1: Solution…

g) Find P(A ∪ B).

Solution

A = {1, 2}, B = {2, 4, 6}

A ∪ B = {1, 2, 4, 6}



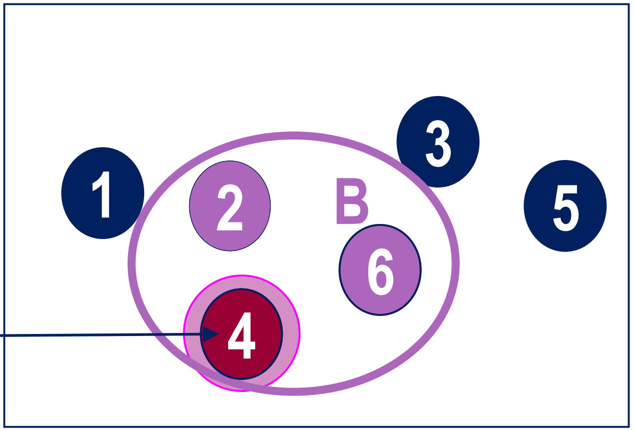
$$P(A \cup B) = P(1,2,4,6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = 0.67$$

Example 1: Solution…

h) Find P(C|B).

Solution

C = {4}, B = {2, 4, 6}



Event C

$$P(C|B) = P(\text{The number is 4 }|\text{The number is even})$$

$$= P(4|\{2,4,6\}) = \frac{1}{3} = 0.33$$

Joint, marginal and conditional probability

We have studied methods to determine probabilities of events that result from combining other events in various ways.

We have also learnt that there are several types of combinations and relationships between events:

- intersection of events
- union of events
- dependent and independent events
- complement events
- conditional events

Intersection

The ***intersection*** of event A and B is the event that occurs when both A and B occur.

The intersection of events A and B is denoted by $(A \cap B)$ or (A and B) .

The ***joint probability*** of A and B is the probability of the intersection of A and B, and is denoted by $P(A \cap B)$.

Example 2

A potential investor examined the relationship between the performance of mutual funds and the university where the fund manager earned his/her MBA.

The following table describes the *joint probabilities*.

	Mutual fund outperforms the market	Mutual fund doesn't outperform the market
Top 20 MBA program	**0.11**	**0.29**
Not top 20 MBA program	**0.06**	**0.54**

E.g. This is the probability that a mutual fund outperforms **AND** the manager was in a top-20 MBA program; it's a **joint probability**.

Example 2: Solution

The joint probability of
[mutual fund outperforms …] **and** [… from a top 20 …] = 0.11.

$P(A_1 \cap B_1) = 0.11$

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)
Top 20 MBA program (A_1)	**0.11**	**0.29**
Not top 20 MBA program (A_2)	**0.06**	**0.54**

$P(A_1 \cap B_1)$

Example 2: Solution

The joint probability of
[mutual fund outperforms …] **and** [… not from a top 20 …] = 0.06

$P(A_2 \cap B_1) = 0.06$

$P(A_2 \cap B_1)$

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)
Top 20 MBA program (A_1)	**0.11**	**0.29**
Not top 20 MBA program (A_2)	**0.06**	**0.54**

Marginal probabilities

Marginal probabilities are computed by adding across rows and down columns; that is they are calculated in the **margins** of the table.

	Mutual fund outperforms the market (B_1)		Mutual fund doesn't outperform the market (B_2)		Marginal prob. $P(A_i)$
Top 20 MBA program (A_1)	0.11	+	0.29	=	0.40
Not top 20 MBA program (A_2)	0.06	+	0.54	=	0.60
Marginal probability $P(B_j)$					1.00

Marginal probabilities

These probabilities are computed by adding across rows and down columns.

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)	Marginal prob. $P(A_i)$
Top 20 MBA program (A_1)	.11 +	.29 +	.40
Not top 20 MBA program (A_2)	.06	.54	.60
Marginal probability $P(B_j)$	.17	.83	1.0

Marginal probabilities

"What's the probability a fund manager isn't from a top school?"

$P(A_2) = 0.06 + 0.54$

	B_1	B_2	$P(A_i)$
A_1	**0.11**	**0.29**	**0.40**
A_2	**0.06**	**0.54**	**0.60**
$P(B_j)$	**0.17**	**0.83**	**1.00**

$P(B_1) = 0.11 + 0.06$

"What's the probability a fund outperforms the market?"

BOTH margins must add to 1 (useful error check)

Conditional probability

Conditional probability is used to determine how two events are related; that is, we can determine the probability of one event ***given*** the occurrence of another related event.

Conditional probabilities are written as **P(A|B)** and read as ‘the probability of A *given* B’ and is calculated as:

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

Conditional probability…

Again, the probability of an event ***given*** that another event has occurred is called a conditional probability…

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

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

Note how 'A given B' and 'B given A' are related…

Example 2a (continued from Example 2)

Find the conditional probability that a randomly selected fund is managed by a top 20 MBA program graduate, given that it did not outperform the market.

	Mutual fund outperforms the market	Mutual fund doesn't outperform the market
Top 20 MBA program	**0.11**	**0.29**
Not top 20 MBA program	**0.06**	**0.54**

Example 2a…

Find the conditional probability that a randomly selected fund is managed by a top 20 MBA program graduate, given that it did not outperform the market.

Solution:

A_1 = Fund manager graduated from a top-20 MBA program

A_2 = Fund manager did not graduate from a top-20 MBA program

B_1 = Fund outperforms the market

B_2 = Fund does not outperform the market

We need to find $P(A_1|B_2)$.

$$P(A_1|B_2) = \frac{P(A_1 \cap B_2)}{P(B_2)}$$

Example 2a: Solution

We need to find $P(A_1|B_2)$.

For event $A_1|B_2$, the new information that B_2 is given reduces the relevant sample space for $A_1|B_2$ to 83% of event B_2..

$$P(A_1|B_2) = \frac{P(A_1 \cap B_2)}{P(B_2)}$$

$$= \frac{0.29}{0.83} = 0.3494$$

	Mutual fund does out-perfom the market (B_1)	Mutual fund doesn't out-perform the market (B_2)	Marginal prob. $P(A_i)$
Top 20 MBA program (A_1)	0.11	0.29	0.40
Not top 20 MBA program (A_2)	0.06	0.54	0.60
Marginal probability $P(B_j)$	0.17	0.83	

Conditional probability…

Before the new information becomes available, we have

$$P(A_1) = 0.40$$

After the new information on B_2 becomes available, $P(A_1)$ changes to

$$P(A_1 | B_2) = 0.3494$$

Since the occurrence of B_2 has changed the probability of A_1, the two events are related (ie, A_1 and B_2 are not independent) and are called ‘dependent events’.

Independence

One of the objectives of calculating conditional probability is to determine whether two events are related.

In particular, we would like to know whether they are ***independent***, that is, if the probability of one event is ***not affected*** by the occurrence of the other event.

Two events A and B are said to be ***independent*** if

$$P(A|B) = P(A)$$

or

$$P(B|A) = P(B)$$

Example 2b (continued from Example 2)

Is ‘Fund outperforms the market’ dependent on the event that the ‘Manager graduated from a top-20 MBA program’?

	Mutual fund outperforms the market	Mutual fund doesn’t outperform the market
Top 20 MBA program	**0.11**	**0.29**
Not top 20 MBA program	**0.06**	**0.54**

Example 2b…

Is ‘Fund outperforms the market’ dependent on the event that the ‘Manager graduated from a top-20 MBA program’?

Solution:

A_1 = Fund manager graduated from a top-20 MBA program

A_2 = Fund manager did not graduate from a top-20 MBA program

B_1 = Fund outperforms the market

B_2 = Fund does not outperform the market

In order to answer this question, it suffices to show that, for at least one i and j, $P(B_i | A_j) \neq P(B_j)$

If B_2 and A_2 are independent, then $P(B_2 | A_2) = P(B_2)$

Example 2b: Solution…

We have already seen the dependency between A_1 and B_2, as $P(A_1|B_2) = 0.3494 \neq P(A_1) = 0.40$

Now, let us check the independence of A_2 and B_2.

$P(B_2) = 0.83$

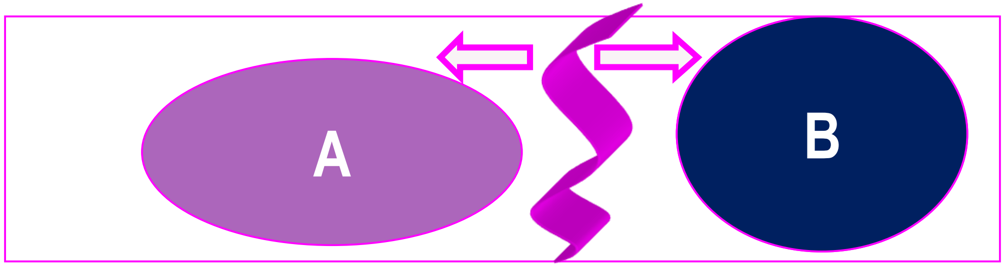
$P(B_2|A_2)=P(B_2 \cap A_2)/P(A_2) = 0.54/0.60 = 0.90$

Therefore, $P(B_2|A_2) \neq P(B_2)$

Conclusion: A_2 and B_2 are dependent. That is, the probability of event B_2 is affected by the occurrence of the event A_2. This means that Fund outperforms the market does not depend on whether the manager graduated from a top-20-MBA program.

Note that independent events and mutually exclusive events are not the same!

A and B are two mutually exclusive events and A *can* take place, that is, $P(A) > 0$. Can A and B be independent?



Let's assume event B has occurred.

The conditional probability that A occurs given that B has occurred is zero, that is, $P(A|B) = 0$, because $P(A \cap B) = 0$.

However, $P(A)>0$, thus $P(A|B) \neq P(A)$ and therefore events **A and B *cannot be independent***.

Example 3

The personnel department of an insurance company has compiled data on promotion, classified by gender. Are promotion and gender dependent on each another?

Manager	Promoted (R)	Not promoted (R^c)	Total
Male (M)	46	184	230
Female (M^c)	8	32	40
Total	54	216	270

Events of interest:

M: A manager is a male — **R:** A manager is promoted

M^c: A manager is a female — **R^c:** A manager is not promoted

Example 3: Solution

If gender has no effect on promotion, then R and M are independent and $P(R|M) = P(R)$. If this equality holds, there is no difference in the probability of promotion between male and female managers.

Manager	Promoted (R)	Not promoted (R^c)	Total
Male (M)	46	184	230
Female (M^c)	8	32	40
Total	54	216	270

Example 3: Solution…

$P(R)$ = Number of promotions / total number of managers
= 54 /270 = 0.20

$P(R|M)$ = P(Number of promotions **|** Only male managers are observed)
= 46 / 230 = 0.20.

Therefore, $P(R|M) = P(R)$. That is, R is not influenced by M, and R *and M are independent*. Similarly, we can show that this results holds for all possibilities of conditional events in the table.

Conclusion: There is no gender discrimination in awarding promotions.

Union

As discussed earlier, the **union event of A and B** is the event that occurs when either A or B or both occur.

It is denoted $(A \cup B)$ or (A or B).

Example 3

Determine the probability that a randomly selected fund outperforms the market or the manager graduated from a top 20 MBA program.

That is, we want to calculate $P(A_1 \cup B_1)$.

Example 3: Solution

A1 or B1 occurs whenever: **A1 and B1** occurs, **A_1 and B_2** occurs, or **A_2 and B_1** occurs…

	Mutual fund outperforms the market (B_1)	Mutual fund doesn't outperform the market (B_2)
Top 20 MBA program (A_1)	0.11	0.29
Not top 20 MBA program (A_2)	0.06	0.54

A_1 or B_1 occurs whenever either: A_1 and B_1 occurs,

A_1 and B_2 occurs,

A_2 and B_1 occurs.

$$P(A_1 \cup B_1) = P(A_1 \cap B_1) + P(A_1 \cap B_2) + P(A_2 \cap B_1)$$
$$= 0.11 + 0.29 + 0.06 = 0.46$$

<u>Alternatively,</u> $P(A_1 \cup B_1) = 1 - P(A_2 \cap B_2) = 1 - 0.54 = 0.46$

Rules of probability

Complement rule

Each simple event must belong to either A or its complement, A^c.

Since the sum of the probabilities assigned to all simple events is one, we have for any event A:

$$P(A^c) = 1 - P(A)$$

A^c is also denoted by $\bar{A}$.

Rules of Probability…

Addition rule

For any two events A and B

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

Rules of Probability…

Multiplication rule

For any two events A and B

$$
\begin{aligned}
P(A \cap B) &= P(B|A)\,P(A) \\
&= P(A|B)\,P(B)
\end{aligned}
$$

When A and B are independent

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

Example 4

A stock market analyst feels that

- The probability that a certain mutual fund will receive increased contributions from investors is 0.6.
- The probability of receiving increased contributions from investors becomes 0.9 if the stock market goes up.
- There is a probability of 0.5 that the stock market rises.

The events of interest are:

A: The stock market rises

B: The company receives increased contribution

Example 4…

Calculate the following probabilities

- The probability that both A and B will occur, **$P(A \cap B)$** [sharp increase in earnings].
- The probability that either A or B will occur, **$P(A \cup B)$** [at least moderate increase in earnings].

Solution

- $P(A) = 0.5$; $P(B) = 0.6$; $P(B|A) = 0.9$
- $P(A \cap B) = P(B|A)\,P(A) = (0.9)(0.5) = 0.45$
- $P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$

$$= 0.5 + 0.6 - 0.45 = 0.65$$