

Data Analytics

ECON 1008, Semester 1, 2019

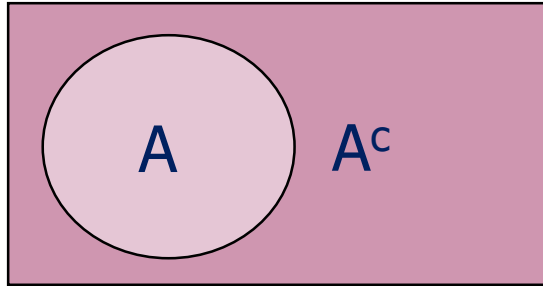
Giulio Zanella
University of Adelaide
School of Economics

Chapter 6, continued...

Probability

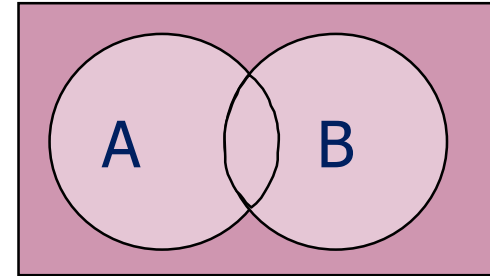
Basic relationships of probability

Complement of event



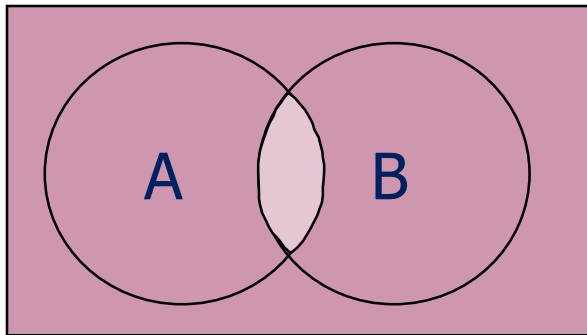
$$P(A^c) = 1 - P(A)$$

Union of events

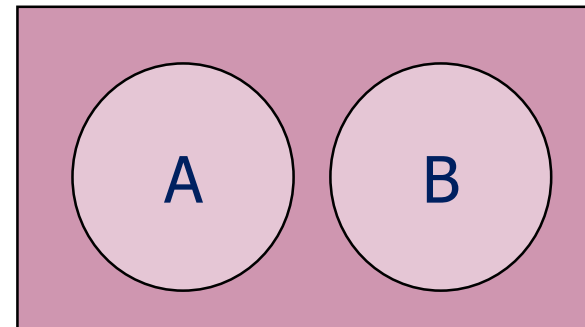


$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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

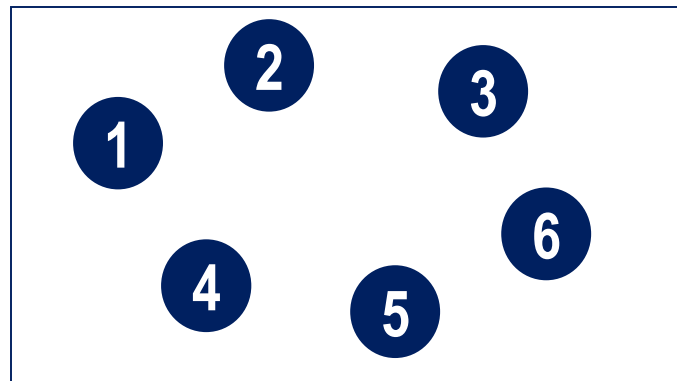
- a) Define the sample space for this random experiment and assign probabilities to the simple events.

Solution

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

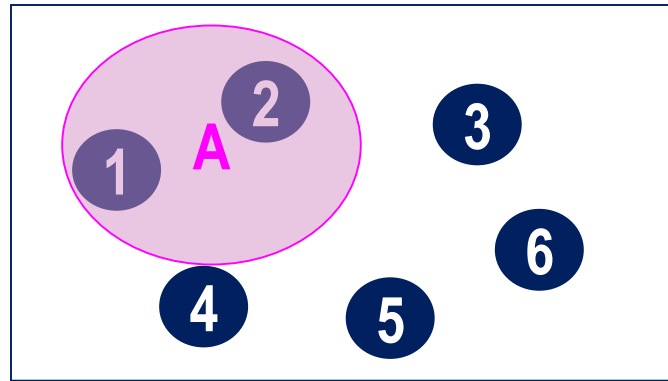
Each simple event is equally likely to occur, thus,
 $P(1) = P(2) = \dots = P(6) = 1/6$.

A Venn diagram



Example 1: Solution...

b) Find $P(A)$.



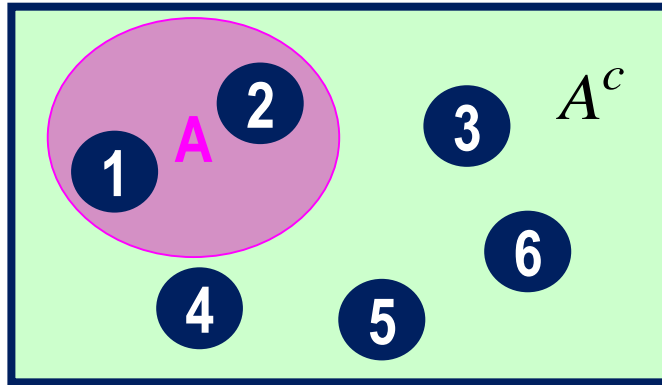
Solution

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

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

Example 1: Solution...

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



Solution

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

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

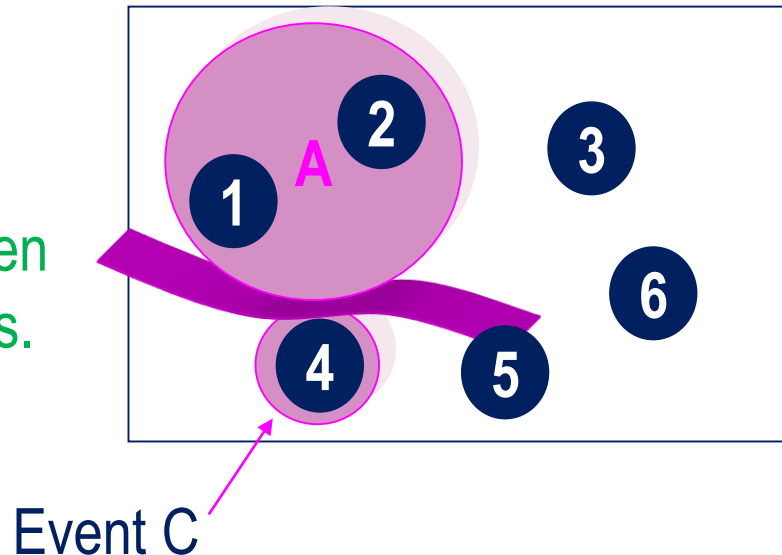
Example 1: Solution...

d) Are events A and C mutually exclusive?

Solution

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

There is no
overlap between
the two regions.



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

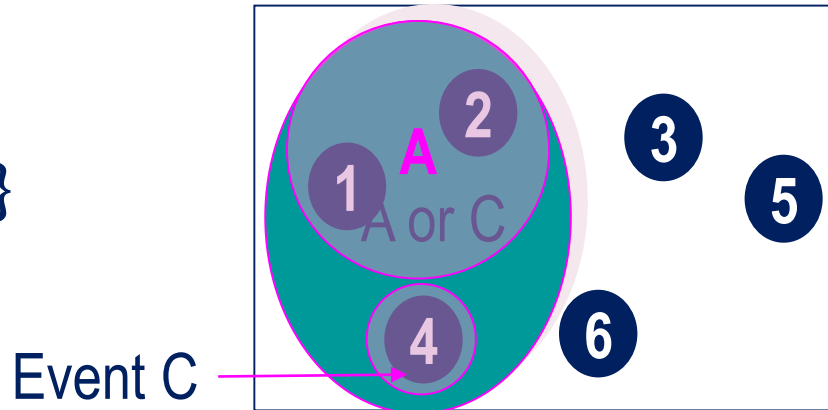
Example 1: Solution...

e) Find $P(A \cup C)$.

Solution

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

$$A \cup C = \{1, 2, 4\}$$



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

or, because A and C are mutually exclusive,

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

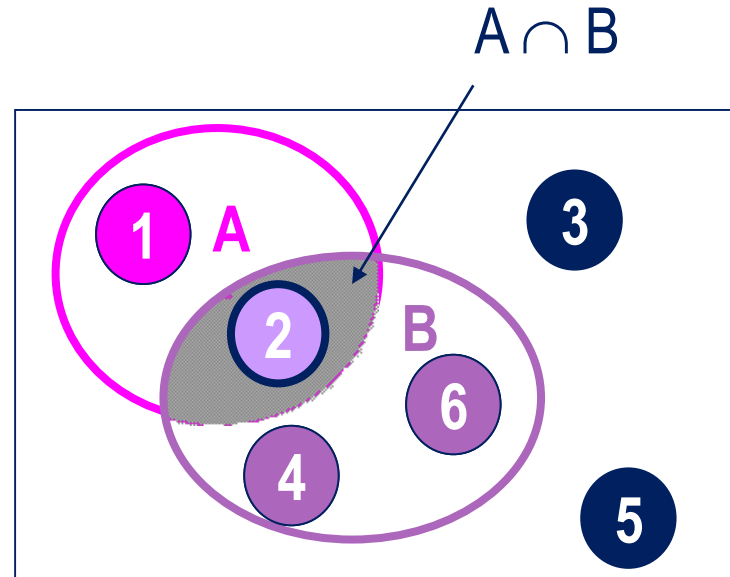
Example 1: Solution...

f) Find $P(A \cap B)$.

Solution

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

$$A \cap B = \{2\}$$



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

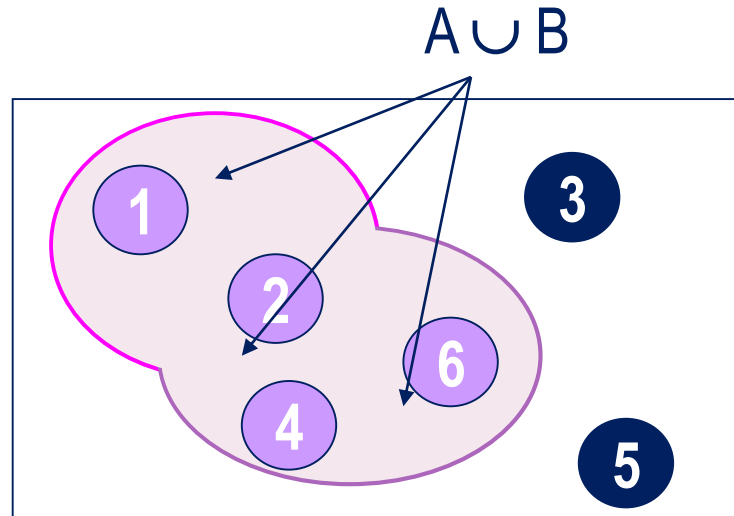
Example 1: Solution...

g) Find $P(A \cup B)$.

Solution

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

$$A \cup B = \{1, 2, 4, 6\}$$



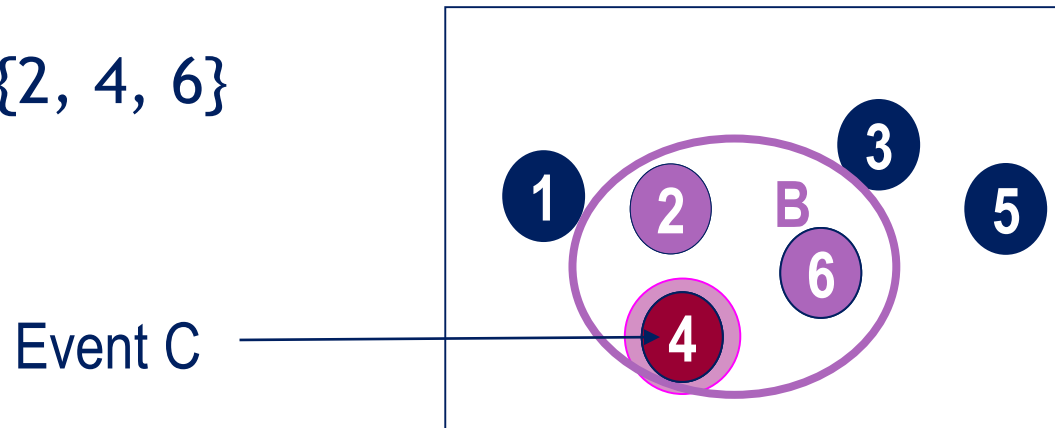
$$P(A \cup B) = P(1, 2, 4, 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = 0.67$$

Example 1: Solution...

h) Find $P(C|B)$.

Solution

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



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

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

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.

$$P(A_1 \cap B_1) = 0.11$$

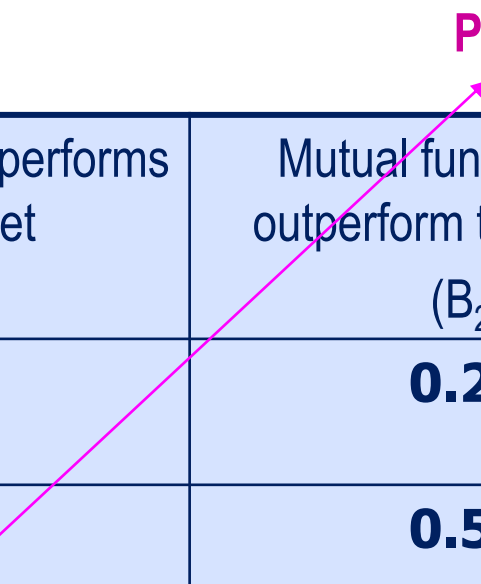
	Mutual fund outperforms the market (B ₁)	Mutual fund doesn't outperform the market (B ₂)
Top 20 MBA program (A ₁)	0.11	0.29
Not top 20 MBA program (A ₂)	0.06	0.54

$$P(A_1 \cap B_1)$$

Example 2: Solution

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

$$P(A_2 \cap B_1) = 0.06$$



	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

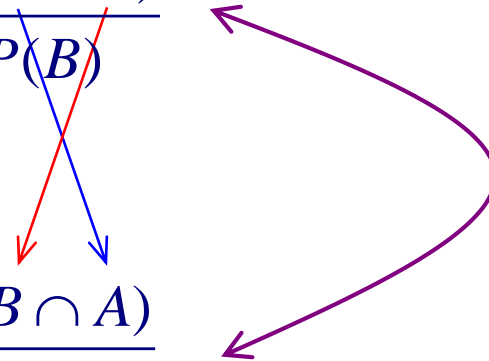
Conditional probability is used to determine how two events are related; that is, we can determine the probability of one event *given* the occurrence of another related event.

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- perform 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 .

$$P(B_2) = 0.83$$

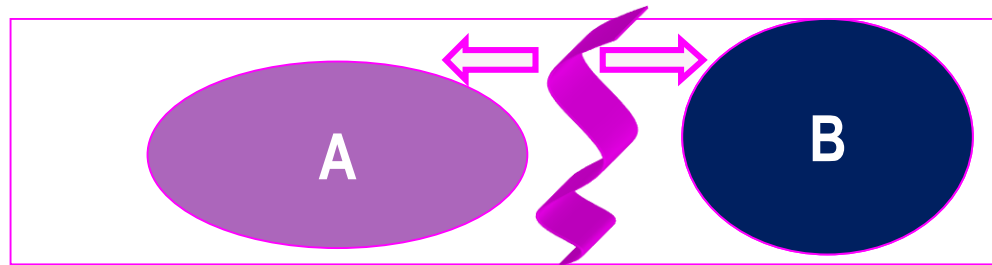
$$P(B_2 | A_2) = P(B_2 \cap A_2) / P(A_2) = 0.54 / 0.60 = 0.90$$

$$\text{Therefore, } P(B_2 | A_2) \neq P(B_2)$$

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...

$$\begin{aligned} P(R) &= \text{Number of promotions} / \text{total number of managers} \\ &= 54 / 270 = 0.20 \end{aligned}$$

$$\begin{aligned} P(R|M) &= P(\text{Number of promotions} \mid \text{Only male managers} \\ &\quad \text{are observed}) \\ &= 46 / 230 = 0.20. \end{aligned}$$

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, **A₁ and B₂** occurs, or **A₂ and B₁** occurs...

	Mutual fund outperforms the market (B ₁)	Mutual fund doesn't outperform the market (B ₂)	A ₁ or B ₁ occurs whenever either: A ₁ and B ₁ occurs,
Top 20 MBA program (A ₁)	0.11	0.29	A ₁ and B ₂ occurs,
Not top 20 MBA program (A ₂)	0.06	0.54	A ₂ and B ₁ occurs.

$$\begin{aligned}
 P(A_1 \cup B_1) &= P(A_1 \cap B_1) + P(A_1 \cap B_2) + P(A_2 \cap B_1) \\
 &= 0.11 + 0.29 + 0.06 = 0.46
 \end{aligned}$$

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

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

- $P(A) = 0.5$; $P(B) = 0.6$; $P(B|A) = 0.9$
- $P(A \cap B) = P(B|A) P(A) = (0.9)(0.5) = 0.45$
- $P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$
 $= 0.5 + 0.6 - 0.45 = 0.65$