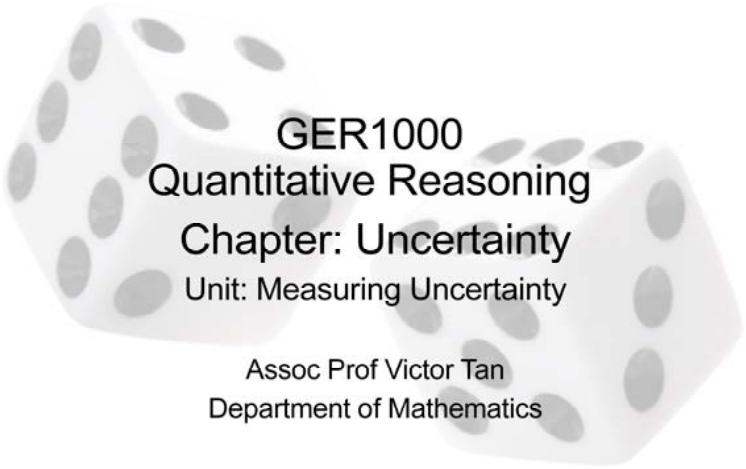
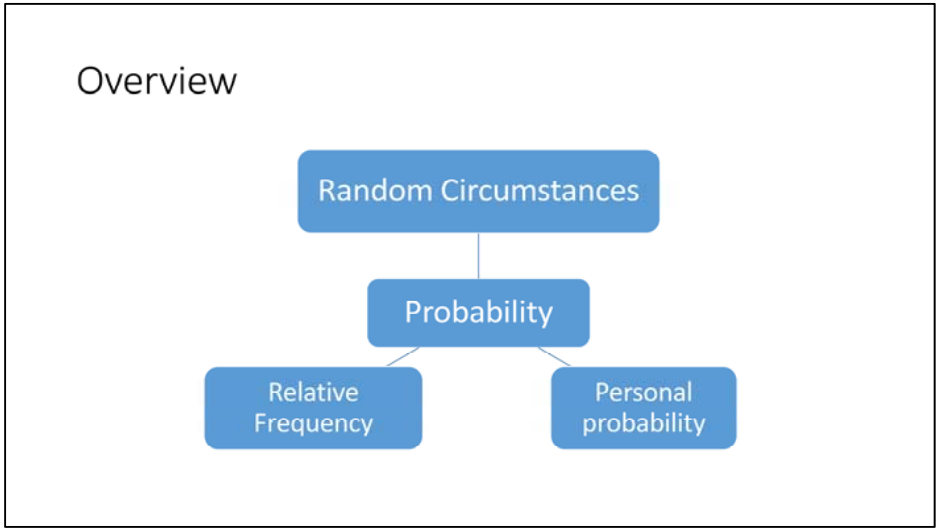



Hi there, my name is Victor Tan.
In this new chapter, we are going to talk about "Uncertainty".
We face with uncertainty everyday.
For example: Will it rain today? Will the bus arrive on time? Will the MRT train break down?
Sometimes we want to know how likely is it that such a scenario will happen?
Many of our decisions are based on an intuitive sense of likelihood
Is there a quantitative way for us to measure uncertainty? The answer is yes.






The underlying quantitative concept in this chapter is Probability.
In this introductory unit, we will start by understanding what is random circumstances.
This will lead to the discussion of probabilities of the possible outcomes under the random circumstances.
Depending on the circumstances, there are two interpretations of probability: the relative frequency and the personal probability interpretation, which will occupy the main part of this unit.

## Random Circumstances

- A random circumstance is one in which the outcome is uncertain.
- The outcome is not determined until we observe it.

❖ Whether it will rain tomorrow

How to measure?

❖ Whether you will get an A grade for this module


A random circumstance is a situation, a scenario, or an activity, in which the outcome is uncertain. In other words, we cannot be 100% sure what will happen.
Very often, we are not able to tell the outcome of a random circumstance until it actually happen.
For example,
When we ask: whether it will rain tomorrow? We are referring to the weather, and the outcome is that either it will rain or it will not rain. This is a random circumstance, as no one can say for sure what the weather will be like until we observe it tomorrow.
When you ask: whether you will get an A grade for this module? You are referring to the activity of taking the module, and the outcome is the grade that you are going to get. Again, you will not know your grade until the result is released.

We shall now discuss how to construct a way to measure the uncertainty of outcomes of these circumstances.



## Probability – the measure of likelihood

- Probability is a measure of how likely something will happen.
- It is a value between 0 and 1 assigned to an outcome of a random circumstance.

❖The probability that it will rain tomorrow is 0.7

| Probability | Outcome happening |
|---|---|
| 1 | Definitely happening |
| 0 | Definitely not happening |
| Close to 1 | More likely to happen |
| Close to 0 | Less likely to happen |

What is probability?
Simply speaking, it is a quantity that measure how likely it is that something will happen.
More precisely, given a random circumstance, the probability that a certain outcome happens, is a value between 0 and 1.
For example, we say: the probability that it will rain tomorrow is 0.7. This means, we assign the value 0.7 to the outcome that it will rain tomorrow.
Look at this table.
If the outcome is definitely going to happen, we assign the value 1 to the outcome, and say the probability of the outcome happening is 1.
On the other hand, if the outcome is definitely not going to happen, we assign the value 0 to the outcome, and say the probability of the outcome happening is 0.
Most outcomes will have a probability value strictly between 0 and 1.
For outcomes that are more likely to happen, we will assign a larger value, closer to 1; for outcomes that are less likely to happen, we will assign a smaller value, closer to 0.
When we assign values to represent probabilities of outcome, we should do it in a consistent and meaningful way.

## Probability VS Chance

A more commonly used term: Chance

- The probability that it will rain tomorrow is 0.7
- There is a 70% chance of raining tomorrow

Perhaps a term that is more commonly used in everyday life and has a similar meaning as probability is “chance”.
For example, the probability that it will rain tomorrow is 0.7.
Another way of saying it is: there is a 70% chance of raining tomorrow.
Note that there is no difference between 0.7 and 70% mathematically.
So it is equally correct to say there is a 0.7 chance of raining tomorrow, thought it is more common to describe “chance” in terms of percentage.



## Interpretation of Probability

| Relative frequency | Personal probability |
|---|---|
| ❖ Will you win the lottery? | ❖ Will you be working overseas after you have graduated? |
| Can be quantified exactly | Cannot be quantified exactly |
| Based on repeated observation of outcomes | Based on our own personal belief |

There are essentially two ways to interpret probabilities.
One is the relative frequency interpretation, and the other one is the personal probability interpretation.

Let us illustrate these two interpretations using examples.
First, the probability that you will win the lottery
This probability can be quantified exactly and its value can be assigned using the relative frequency interpretation.
It is based on repeated observation of the outcomes of the activity.
In our example, assuming you buy lottery tickets regularly, the probability can be determined by observing how often you win the lottery.

The second example is the probability that you will be working overseas after you have graduated.
This probability cannot be quantified exactly and has to be estimated using the personal probability interpretation.
It is based on our own personal belief.

## Relative Frequency Interpretation

For circumstances that we can observe repeatedly

Probability ⟷ proportion of times occur (or relative frequency) over the long run

- Bought lottery tickets 100 times, and won 3 times.
  - Proportion: 3/100 = 0.03
  - Probability that you win lottery ≈ 0.03
- Among 500 randomly chosen people, 20 have certain disease.
  - Proportion: 20/500 = 0.04
  - Probability that an individual has disease ≈ 0.04

Relative frequency interpretation of probability applies in circumstances that we can repeat many times, or observe repeatedly.
For example, winning lottery, raining, breakdown of MRT trains and so on.
For this interpretation, probability of a specific outcome is given by the proportion of times it would occur over the long run.
This is also called the relative frequency of that particular outcome in the long run.
For example, if you have bought lottery tickets 100 times in the past, and won a total of 3 times, then the proportion of times occur is 3/100 which is 0.03.
Therefore, the probability that you will win lottery has an estimated value of 0.03.
Note that under the relative frequency interpretation, we cannot assess the probability of a particular outcome by observing it only a few times. We need to do it repeatedly for sufficient number of times, say 100 times or even more to have a good estimation.
This interpretation also includes the description of the proportion of individuals who have a certain characteristic.
For example, among a group of 500 randomly chosen people from a population, if we observe that 20 of them have a certain disease, then the proportion of the sick people is 20/500, or 0.04.
So the probability that an individual from the population has the disease has an estimated value of 0.04.



## From earlier chapter

### Diabetes Risk

| | Diabetic | Healthy | Row total |
|---|---|---|---|
| Female | 72,000 | 144,000 | 216,000 |
| Male | 52,000 | 156,000 | 208,000 |
| Column total | 124,000 | 300,000 | 424,000 |

Risk _female = Diabetes risk for female = $\frac{72000}{216000}$ = 0.333

Probability that a female has diabetes

Risk _male = Diabetes risk for male = $\frac{52000}{208000}$ = 0.25

Probability that a male has diabetes

Recall the example on diabetes risk in the earlier chapter on Association.
The given two-by-two table gives the number of diabetic cases in a certain population.
The diabetes risk for female is computed using the first row of the table.
This is in fact the same as the probability that a female from the population has diabetes.
Similarly, the diabetes risk for male is the same as the probability that a male has diabetes.

## Equally likely outcomes


relative frequency in the long run are the same

| Circumstance | Tossing fair coin | Rolling fair die |
|---|---|---|
| Outcomes | Head & Tail | 1, 2, 3, 4, 5, 6 |
| No. of outcomes | 2 | 6 |
| Probability of each outcome | $\frac{1}{2}$ | $\frac{1}{6}$ |

Circumstances with exactly N outcomes

- Suppose each outcome is equally likely.
- Then the probability of each outcome is $\frac{1}{N}$.

In a typical introductory course on probability, the standard examples used for illustrations are tossing coins and rolling die, as these examples involve the notion of equally likely outcome.

Let's consider a random circumstance with exactly 2 possible outcomes.
If we say each of the two outcomes is equally likely, it means that we expect the relative frequency of their occurrences in the long run to be the same. In other words, they each occur half the time.
Therefore, the probability of each outcome is $1/2$ or 0.5.
For example, when we toss a fair coin, the 2 possible outcomes are head and tail. Being a fair coin, we expect that, when we toss the coin repeatedly, half the time the outcome will be head, and half the time tail.
So the probability of tossing a head is $1/2$. and the probability of tossing a tail is also $1/2$.
More generally, for a random circumstance with exactly N possible outcomes, if each outcome is equally likely, then we expect the number of time each outcome will occur to be the same if it is observed repeatedly many times. Hence the relative frequency and probability of each of the N outcomes is $1/N$.
For example, when we roll a fair die, there are 6 equally likely outcomes: 1, 2, 3, 4, 5, 6.
So, the probability of rolling a 4 is $1/6$. Same for rolling any of the other five numbers.



## Relative frequency as Probability

Assume the coin is fair

| Tossing a coin | Weather forecast |
|---|---|
| probabilities can be computed precisely | probabilities cannot be computed precisely |
| assumption about the physical realities | circumstances repeated and outcome observed |
| Use probability to predict relative frequency | Use relative frequency to estimate probability |

In some random circumstances like coin tossing and die rolling, we can compute the probabilities precisely based on what we think about the physical realities. In other words, we have to make some assumption.
For example, to assign the probability of $1/2$ for tossing a head, we have to assume that the coin is fair. so that both tossing a head and tossing a tail are equally likely.

There are circumstances whereby the probability of an outcome cannot be computed exactly, for example, in weather forecast, how do we determine the probability that it will rain tomorrow?

For such cases, we may use the relative frequency to assign probabilities to the outcomes, provided the circumstances can be repeated numerous times and the outcome can be observed each time.

## Weather Forecast


Weather is an example of a random phenomenon with an uncertain outcome, but it does have a regular pattern over time.

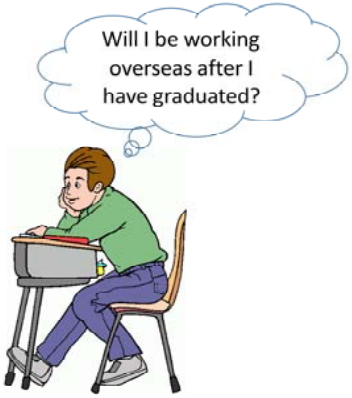
When the weather forecast announced that there is a 70% chance of raining tomorrow, what does it really mean?
Simply speaking, "70% chance of raining tomorrow" qualitatively means it is quite likely that it will rain tomorrow.
But quantitatively, it means, based on similar scenarios observed in the past, 7 out of 10 times ended up raining.

Meteorologists use mathematical models to combine weather inputs to predict what will happen in the next few days based on what has happened in the past when similar scenarios have been observed.



## Personal Probability Interpretation

- Circumstances that:
  - are not repeatable
  - only apply to particular individuals
  - only happen once and will never happen again
- Personal probability
  - is the degree to which one believes outcome will happen.
  - may take into consideration similar events in the past

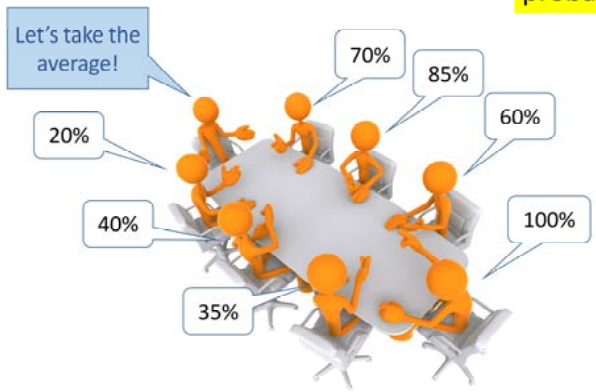



Uncertainty is a characteristic of most circumstances, whether or not they are repeatable under similar conditions.
Sometime, people need to make decision based on how likely they think the future will evolve.
For example, how likely you will be working overseas after you have graduated.
Very often, such uncertainty arises from circumstances that cannot be repeated or duplicate.
These are situations that apply to particular individuals. They may only happen once, and will never happen again.

So how do we assign probabilities in such situation?
This is where the personal probability interpretation comes in.
The personal probability of an outcome is the degree to which a given individual believes that the outcome will happen.
Sometimes data from similar events in the past and other knowledge are taken into consideration when an individual assigns values to the personal probability.

## Using personal probabilities

Subjective probability




People routinely base their decisions on personal probabilities.
For example, suppose an organizing committee is deciding whether to hold a certain event.
Each member of the committee may have different assessment of the event, and disagree on the probability that the event will be successful.
The committee can make use of the personal probabilities of its members to reach a consensus quantitatively.

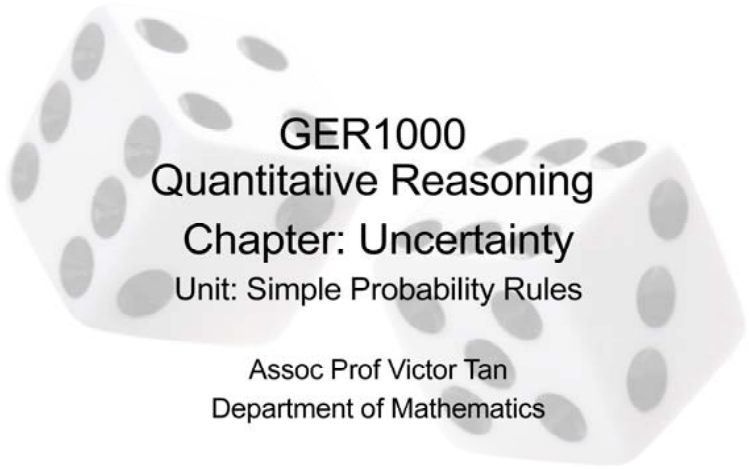
When applying personal probability on a common situation, each individual may assign a different value based on his or her own knowledge and experiences. Nobody could be considered wrong.
That's why this interpretation is also known as subjective probability.



## Summary

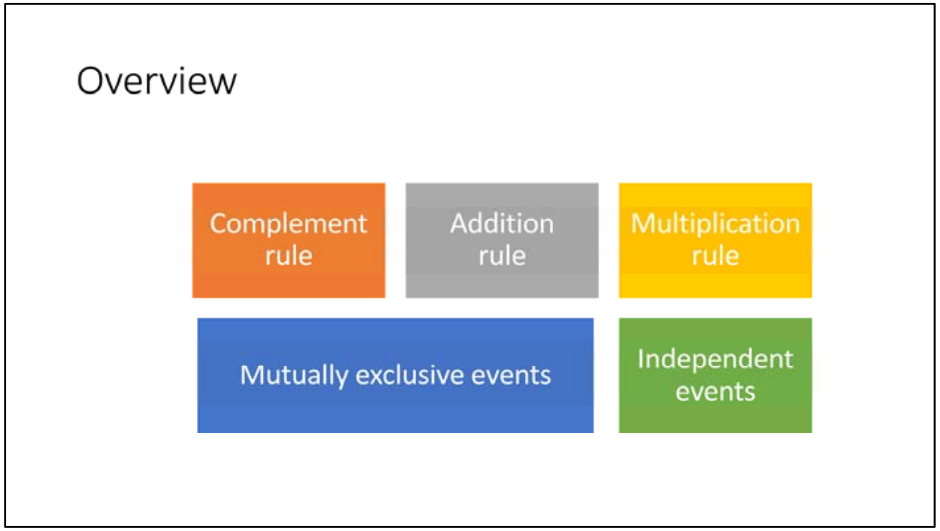
- Two interpretations of Probability
  - Relative frequency probability
  - Personal probability
- Relative frequency can be determined by
  - Making an assumption about the physical world to define relative frequency
  - Observe relative frequencies over many repetitions
- Personal probability is determined based on
  - Individual's belief and past experience of similar situations

To sum up this unit, we have introduced the concept of probability, which is a quantitative way to measure the likelihood of something happening in a random circumstance.
There are essentially two interpretations of probability, namely relative frequency and personal probability, depending on the nature and context of the circumstance.
Under the relative frequency interpretation, probability can either be computed exactly by making assumption about the physical world related to the circumstance,
or it can be approximated by observing the relative frequencies over many repetitions of the circumstance.
On the other hand, personal probability interpretation is more subjective, and usually apply to situations that cannot be repeated. It depends on individual's belief and past experience of similar situations.




In the first unit of this chapter, we talk about the probability of outcomes of random circumstances, which are values between 0 and 1.
In this unit, we are going to see how simple outcomes can be put together to form some complex events,
and how we can determine the probability of the combination of outcomes from the probabilities of the individual outcomes.
In order to do that, we need to introduce some simple probability rules.






There are three basic mathematical rules about probability that will be useful for our discussion.
Namely, complement rule, addition rule and multiplication rule. They will only involve simple arithmetic.
To understand how to apply these rules, we need to introduce the concepts of mutually exclusive events and independent events.

## Events

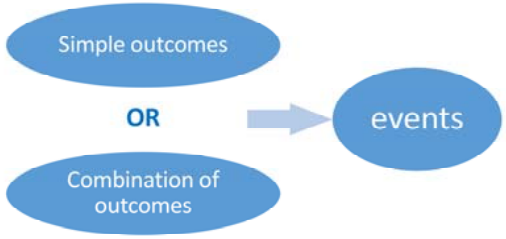
❖Rolling a fair die

Simple outcome:

- Rolling a 1
- Rolling a 2
- Rolling a 3
- Rolling a 4
- Rolling a 5
- Rolling a 6

"Combinations" of outcomes:

- Rolling an even number
- Rolling a number greater than 4




In the previous unit, we have seen the example of rolling a fair die. There are 6 simple outcomes.
Rolling a 1 is a simple outcome; so is rolling a 2, rolling a 3 etc.
Sometimes, we may also consider other combinations of outcomes.
For example, instead of a specific number, we may be interested in rolling an even number, which is a combination of rolling a 2, rolling a 4 and rolling a 6.
We may also want to consider rolling a number that is greater than 4. This is a combination of rolling a 5 and rolling a 6.
All these are examples of events.
From now on, we will collectively refer both simple outcomes, or combinations of outcomes as events.



## Probability of Events

P(an event) : probability of the event

Probability of complex event from probabilities of simpler events

❖Rolling a fair die

i. P(rolling even) = P(rolling 2, 4 or 6)

$$= P(\text{rolling 2}) + P(\text{rolling 4}) + P(\text{rolling 6})$$

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

ii. P(rolling > 4) = P(rolling 5 or 6)

$$= P(\text{rolling 5}) + P(\text{rolling 6})$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

iii. P(rolling even or > 4) = P(rolling even) + P(rolling > 4) ?

In our subsequent discussions, to represent the probability of an event, we shall use the letter P, followed with a pair of parenthesis enclosing a brief description of the event.
Sometime it is possible to obtain the probability of a more complex event from the individual probabilities of simpler events
Here are some examples.
For the event of rolling an even number, we need to roll either a 2, a 4 or a 6. Then its probability is the sum of the probability of the 3 simpler events.
In other words, we add up the probabilities of rolling a 2, rolling a 4 and rolling a 6.
In the previous unit, we have seen that the probability of each of these equally likely outcomes is 1/6, therefore the probability of the event of rolling an even number is given by ½.
Similarly, for the probability of rolling a number greater than 4, we just sum up the probability of rolling a 5 and the probability of rolling a 6, which gives 1/3.
In this third example, we combine the event of rolling an even number, and the event of rolling a number greater than 4. In other words, we are hoping the outcome is either one of the even numbers, or a number greater than 4.
The question is, can we compute the probability of this complex event by summing up the probability of the two individual events?

## Mutually Exclusive Events

❖Rolling a fair die

P(rolling even or > 4) = P(rolling even) + P(rolling > 4) ?

$= \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ ← Incorrect

Rolling even: rolling 2, 4, 6

Rolling > 4: rolling 5, 6

an overlapping of outcome

P(rolling even or > 4) = P(rolling 2, 4, 5, 6) $= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$

The events "rolling even" and "rolling > 4" are not mutually exclusive.

Two events are mutually exclusive
if they cannot occur at the same time.

Let's examine this event closely.
Suppose we add up the probabilities of the two individual events: rolling an even number, and rolling a number greater than 4, the value that we get will be ½ + 1/3 which is 5/6.

Now let us look at this event in terms of the individual outcomes.
Rolling an even number means getting the outcome 2, 4 or 6.
Rolling a number greater than 4 means getting the outcome 5 or 6.
Therefore, if we want the outcome to be either an even number or a number that is greater than 4, we must roll either a 2, 4, 5 or 6.
So the probability of this event is the probability of the combination of rolling a 2, rolling a 4, rolling a 5 and rolling a 6.
And hence it is the sum of the probabilities of the 4 individual outcomes, which gives a value of 2/3.
This means that the value calculated above by summing the probability of rolling an even number, and rolling a number greater than 4 is incorrect.
The reason that the first approach fails is because there is an overlapping outcome 6 in the two events.
We say that the two events are not mutually exclusive.
On the other hand, we say two events are mutually exclusive if they cannot occur at the same time. In other words, if one event happens, then the other event cannot happen.



## Addition Rule

When two (or more) events are mutually exclusive, the probability of either one of these events occurring is the sum of their individual probabilities.

❖Rolling a fair die

i. "Rolling a 5" and "Rolling a 6" are mutually exclusive

P(rolling 5 or 6) = P(rolling 5) + P(rolling 6) $= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

ii. "Rolling > 4" and "Rolling < 4" are mutually exclusive

P(rolling > 4 or < 4) = P(rolling > 4) + P(rolling < 4) $= \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$

The addition rule of probability says that: given two mutually exclusive events A and B, the probability that either A or B occurring, is the sum of the individual probability of A and B occurring.

In other words, this rule tells us when we are allowed to add up two or more events to get the probability of the combined event.

Let's look at some examples.
When we roll a die, and we get a 5, then we can't get a 6 at the same time.
So rolling a 5 and rolling a 6 are mutually exclusive. According to the addition rule, the probability of rolling a 5 or 6 can be obtained by adding the probability of rolling a 5, and that of rolling a 6.

The event of rolling a number greater then 4, and the event of rolling a number less than 4 are again mutually exclusive.
So the probability of the combined event can be given by the sum of the probabilities of these two individual events.

## Complement Rule

The probability of an event not occurring
is 1 minus the probability of the event occurring.

❖Rolling a fair die

i. Event: rolling even
   - Event not occurring same as not rolling even same as rolling odd
   - P(rolling odd) = 1 – P(rolling even) = $1 - \frac{1}{2} = \frac{1}{2}$

ii. Event: rolling > 4
   - Event not occurring same as not rolling > 4 same as rolling ≤ 4
   - P(rolling ≤ 4) = 1 – P(rolling > 4) = $1 - \frac{1}{3} = \frac{2}{3}$

The next rule is called complement rule, which involves subtraction.
It says: the probability of an event A not happening is 1 minus the probability of the event A happening.

Let's consider the event of rolling an even number.
What does it mean by the event not occurring? It means not rolling an even number, which means rolling an odd number.
So according to the complement rule, the probability of rolling an odd number is 1 minus the probability of rolling an even number, which is ½.

For the event of rolling a number greater than 4.
To say that the event not occurring is the same as the event of rolling a number less than or equal to 4.
Again, we can apply the complement rule to see that the probability of this event is 1 minus the probability of rolling a number greater than 4.



## Odds revisited

In earlier chapter

$$\text{Odds for an event} = \frac{\#\text{ of events}}{\#\text{ of non-events}}$$

$$= \frac{\text{Probability of event occurring}}{\text{Probablility of event not occurring}}$$

For example, if the probability of having a disease is 0.25

$$\text{Odds of having the disease} = \frac{P(\text{disease})}{P(\text{no disease})} = \frac{0.25}{1-0.25} = \frac{1}{3}$$

The concept of odds was introduced in the chapter of Association.
Recall that the odds for an event is defined as the number of events divided by the number of non-events.
This can also be calculated in terms of probability, namely, the probability of "the event occurring" divided by the probability of "the event not occurring".
Let's look at a quick example. Suppose we know that the probability of an individual in a population having a certain disease is 0.25.
Then the odds of an individual having the disease is given by the probability that "an individual has the disease" divided by the probability that "an individual does not have the disease".
By complement rule, the probability that an individual does not have the disease is 1 – 0.25. Hence the odds of having the disease can be calculated as 1/3.

## More Complex Scenario

i. Rolling a fair die twice
(repeat a simple activity consecutively)

 X 2

Example of an outcome: (2, 4).
There are 6 x 6 = 36 outcomes
Probability of outcome (2,4): $\frac{1}{36}$

ii. Rolling a fair die and tossing a fair coin
(carry out 2 "unrelated" activities concurrently)

Example of an outcome: (5, T).
There are 6 x 2= 12 outcomes
Probability of outcome (5, T): $\frac{1}{12}$

So far we have seen simple activities like tossing a coin or rolling a die.
Very often we need to combine simpler activities into more complex ones.
For example, rolling a die twice , is repeating a simple activity consecutively.
Rolling and die and tossing a coin, is carrying out 2 unrelated activities concurrently.
For the first example, the outcome will be a pair of number, one for each roll.
For instant, if the first roll is a 2, and second roll is a 4, then the outcome is the ordered pair (2, 4).
Since there are 6 possible outcomes for the first roll, and 6 possible outcomes for the second roll, in total there are 36 outcomes for rolling a die twice.
If we assume the die is fair, then all the 36 outcomes are equally likely.
In particular, the probability of getting outcome (2, 4) is 1/36.

For the second example, the outcome will be a number, pairing with a head or a tail.
For instant, if the roll is a 5, and the toss is a tail, then the outcome is the pair (5, T), where T represents tail.
Since there are 6 possible outcomes for rolling the die, and 2 possible outcomes for tossing the coin, in total there are 12 outcomes for this activity.
With a fair die and a fair coin, the 12 outcomes are all equally likely.
So the probability of getting outcome (5, T) is 1/12.



## Independent Events

i. Rolling a die twice

The outcome of 1st roll does not affect the chance of the outcome of 2nd roll

ii. Rolling a die and tossing a coin

The outcome of the die does not affect the chance of the outcome of the coin

If two events do not affect each other's chance of occurrence, the events are said to be independent of each other.

Observe that, when we roll a die twice, whatever outcome of the 1st roll we get, it will not affect the probability of the outcome of the 2nd roll, and vice versa.
Likewise, when we roll a die and toss a coin, what number appears on the die will have no impact on whether it is more or less likely to get a head or tail on tossing the coin.

We say that two events are independent of each other, if the occurrence of one event does not affect the probability of the occurrence of the other.

## Multiplication Rule ✖

Given two independent events.
The probability of these events both occurring (at the same time) is the product of their individual probabilities.

i. Rolling a die twice
Probability of outcome (2, 4)

$= P(\text{rolling a 2}) \times P(\text{rolling a 4}) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

ii. Rolling a die and tossing a coin
Probability of outcome (5, T)

$= P(\text{rolling a 5}) \times P(\text{tossing a T}) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

The multiplication rule of probability says that: given two events A and B which are independent of each other, the probability that A and B occurring at the same time, is the product of the individual probabilities of A and B occurring.

In other words, this rule tells us when we are allowed to multiply two or more events to get the probability of them happening concurrently.

Go back to our two examples.
Since getting a 2 in the 1st roll and getting a 4 in the 2nd roll are independent of each other, the probability of the outcome (2,4) when we roll a die twice can be simply obtained by multiplying the two individual probabilities of rolling a 2 and rolling a 4, which is 1/6 x 1/6.
We see that the product 1/36 is indeed the probability of the outcome (2, 4)

Similarly for the example of rolling a die and tossing a coin.



## Putting All Together


A couple will **continue** having children **until they** have a **boy**.
Suppose:

- Probability of a birth resulting in a girl is 0.49. — Assumption I
- Outcomes of births are **independent** of each other. — Assumption II

Find the probability for the couple to make at most 3 tries to have the first boy.

$P(\text{boy}) = 1 - P(\text{girl})$ — Complement rule

$= 1 - 0.49 = 0.51$ — Assumption I

We are going to illustrate how all the three rules are applied in the following scenario.
Suppose a married couple likes to have a boy, and is determined to continue trying until they give birth to a boy.
Given that the probability of giving birth to a girl is 0.49, and the sex of one birth does not affect the sex of the subsequent births. In other words, the outcome of births are independent of each other.
What is the probability for the couple to make at most 3 tries to have the first boy?

First of all, note that the two given conditions are assumptions made based on data gathered in the past. Let's call them assumption I and assumption II.

With these assumptions, we can first use the complement rule to determine the probability of giving birth to a boy.
Since we are given the probability of giving birth to a girl is 0.49, that of giving birth to a boy is 0.51.

## Putting All Together


Probability to make at most 3 tries to have the first boy:

P(at most 3 tries) = P(1 try, 2 tries or 3 tries) mutually exclusive
= P(1 try) + P(2 tries) + P(3 tries) Addition rule
= 0.51 + 0.2499 + 0.1201 = 0.88

High chance

P(1 try) = P(boy) = 0.51

P(2 tries) = P(1st girl, 2nd boy) Assumption II: independent
= P(girl) x P(boy) Multiplication rule
= 0.49 x 0.51 = 0.2499

P(3 tries) = P(1st girl, 2nd girl, 3rd boy) Assumption II: independent
= P(girl) x P(girl) x P(boy) Multiplication rule
= 0.49 x 0.49 x 0.51 = 0.1201

When we say the couple will make at most 3 tries to have the first boy, it means they will have their first boy with either 1 try, 2 tries, or 3 tries.
It is not hard to see that having first boy in 1 try, 2 tries and 3 tries are three mutually exclusive events. Therefore, we can apply the addition rule here to compute the probability of making at 3 tries by summing up the individual probabilities of 1 try, 2 tries and 3 tries separately.
So it boils down to finding the probability of these three events.
Having a boy with just 1 try simply means giving birth to a boy.
So the probability for the couple to make only 1 try is 0.51.
Having a boy with 2 tries would mean the first birth is a girl, while the second birth is a boy.
Since we are given that outcome of births are independent, the probability for the couple making two tries is the product of the individual probabilities of having a girl and having a boy using multiplication rule.
This gives a value of 0.2499.
Finally, having a boy with 3 tries would mean the first 2 births are girls, while the third is a boy.
Again, since each of the 3 births are independent, we can compute the probability for the couple making three tries using multiplication rule. And this gives a value of 0.1201.
Now we can go back to find the probability for the couple to make at most 3 tries by adding up 0.51, 0.2499 and 0.1201, which gives a value of 0.88.
What this means qualitatively is that, there is quite a high chance for the couple to have a boy within three tries.



## Summary

Mutually exclusive VS Independent

- Two events are mutually exclusive if the occurrence of one prevents the other from happening
- Two events are independent if the occurrence of one does not change the chances for the other

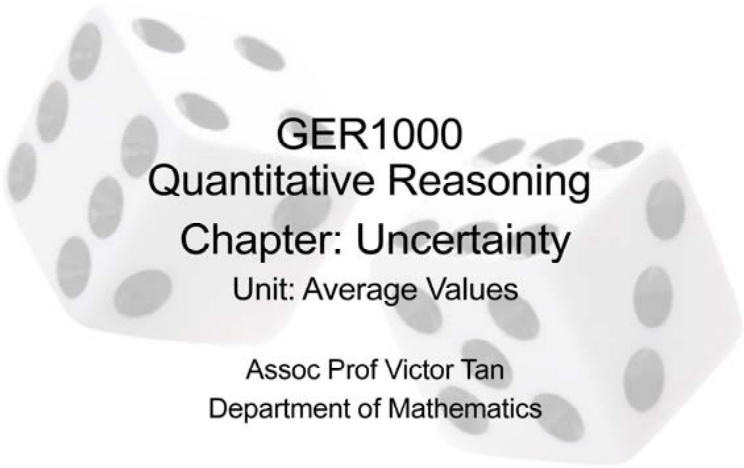
Addition VS Multiplication

- The addition rule calculates the probability that at least one of two events happen
- The multiplication rule calculates the probability that two events both happen

To conclude this unit, let me make a comparison between mutually exclusive and independent events, as well as between addition and multiplication rules.

The concepts of mutually exclusive and independent both involve at least two events.
When two events are mutually exclusive, it means the occurrence of one event prevents the other event from happening.
On the other hand, when two events are independent, it means the occurrence of one event does not change the chances for the other event to occur.

Both addition and multiplication rules are ways of combining probability of two or more events.
On one hand, the addition rule calculates the probability that at least one of two events happen.
On the other hand, the multiplication rule calculates the probability that two events both happen.



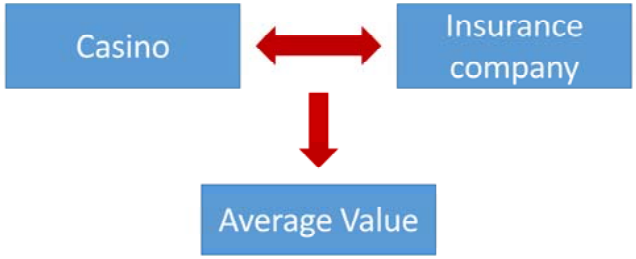
# GER1000
# Quantitative Reasoning
## Chapter: Uncertainty
### Unit: Average Values

Assoc Prof Victor Tan
Department of Mathematics

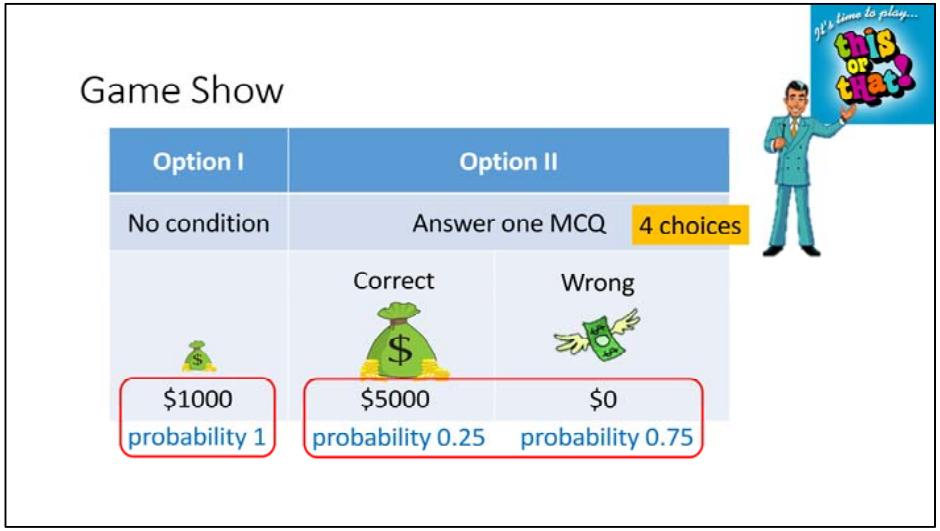
In this unit, we will discuss the concept of average values, which is also known as expected value or mean value in statistics.
For certain random circumstances, every outcome may be associated with a numerical value. Consequently, when we carry out the activities repeatedly, we can take the average value of the outcomes. Some of these average values may give us interesting information.
We are going to see some real life applications of average values. We will also see how average values can help individuals to make decision quantitatively.



## Overview



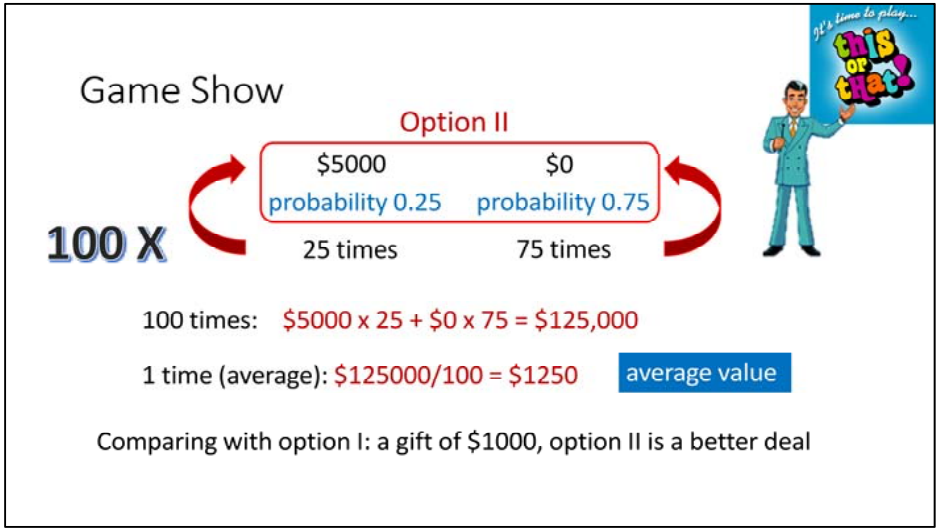

Casino and insurance company: what do they have in common?
In this unit, we shall see that these industries ensure that their businesses will make profits by applying probability to work out their average gains or losses.
The underlying quantitative concept in these examples is the average value.



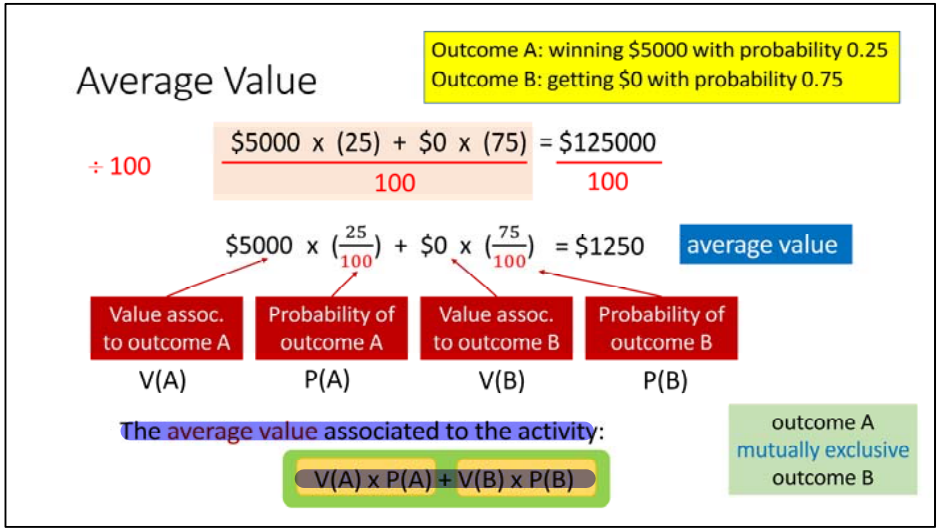

Let's start with a simple example.
You have entered a game show and are given two options:
Option 1 is to take away the cash prize of $1000 with no condition attached.
Option 2 is to answer an MCQ. You win $5000 if you answer correctly; and get nothing if you answer wrongly
Which option would you choose?

If you choose option 1, you will be guaranteed a definite value of $1000
However, if you choose option 2, you will be facing with uncertainty involves two outcomes
What is the chance of winning $5000?
We need to have more information here:
Suppose there are 4 choices given for the MCQ. Furthermore, you do not know what question will be asked until you have decided to go with option 2.
Then the outcome of winning $5000 will have a probability of 0.25, assuming you will pick 1 out of the 4 options purely by chance.
By complement rule, the outcome of getting nothing has a probability of 0.75.
But how do we compare option II with option I using all these quantities?



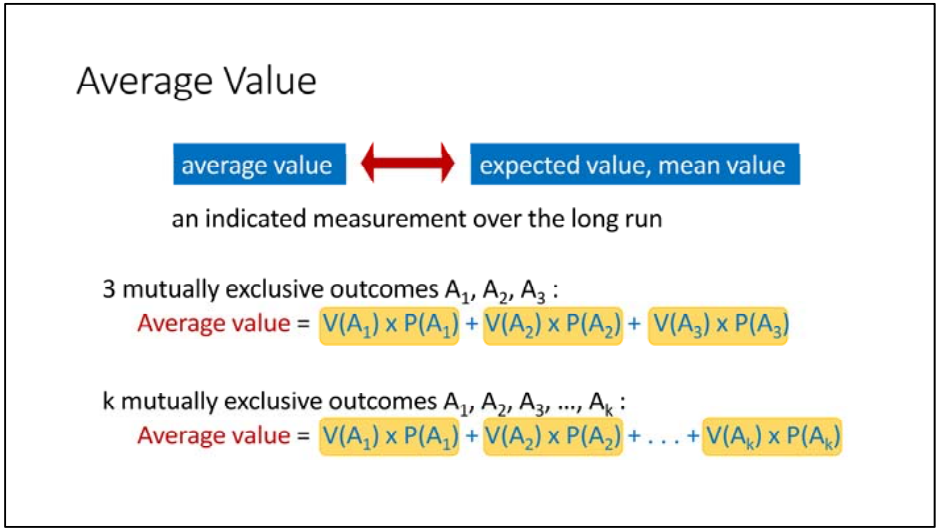



We need to find the average value of option 2.
Suppose we repeat this option 100 times.
We expect to win $5000 for 25 times, and get $0 for 75 times.
In total, we expect to win: $125,000 by adding up all the winning amount for the 100 times.
On average, each time we expect to win: $1250.
Comparing with option 1 which gives a fixed amount of $1000, we can see that option 2 is a better deal.
We call the value $1250 the average value associated to the activity in option 2.
In reality, the average value will be even larger, as you may actually know how to answer the question, and hence raise the probability of winning the $5000.




In the previous example, we calculate the average value by first totaling up the amount and then dividing by 100.
If we look at the expression on the left hand side, it can be rewritten as
5000 x 25/100 + 0 x 75/100
We identify the 4 terms on the left hand side as the quantity that arises directly from option 2 of the game show:
Namely 5000 is the value associated to outcome A, while 25/100 or 0.25 is the probability of outcome A.
Likewise, 0 is the value associated to outcome B, while 75/100 or 0.75 is the probability of outcome B.
Let's denote these 4 quantities by V(A), P(A), V(B) and P(B) respectively.
In general, given an activity with two possible mutually exclusive outcomes A and B with probabilities P(A) and P(B) respectively.
Suppose the outcomes: are associated with values V(A) and V(B).
The average value associated to the activity is multiply the value and probability for each outcome, and sum up the products.






Average value is also called expected value, or mean value in statistics.
It gives an indicated measurement over the long run or over a large sample size.
In general, it does not represent any typical value for any one occurrence of the activity.
It can be extended to activities with more than 2 mutually exclusive outcomes:
For example, suppose an activities has precisely 3 possible outcomes which are mutually exclusive, denoted by A1, A2 and A3.
Then the average value of the activity is given by summing over all the outcomes the product of each pair of value and probability as shown.
More generally, if there are k mutually exclusive outcomes in the activity, the average value is computed in a similar way.

## Casino Gambling


Long run

average value

In a typical casino, gamblers place bets on all kinds of games like slot machines, roulette wheel, pokers and blackjack.
Some players win, others lose; some people gets rich, others go broke.
There is one thing for sure: in the long run, the casinos always make money. How do they do it?
Well, the casino's performance is the combined result of huge number of individual bets.
Sometimes the House wins, some time they lose.
When these random events are repeated over and over again, the gain or loss of the casino will get closer and closer to its average value.
Moreover, every casino game is designed in such a way that is slightly in the casino's favor.
So it is a certainty that the casino will make money over the long run.



## Roulette

- 38 spots
- 18 red
- 18 black
- 2 green

- A player bet $10 on red
- 18 out of 38, player win
- 20 out of 38, player lose

To see how this works, consider the game of roulette.
A standard roulette wheel consists of 38 spots; the numbers 1 to 36 occupy alternately the red and black spots. So there are 18 red spots and 18 black spots.
There are also two special numbers 0 and 00 occupying the green spots.
When the wheel is spun, the marker ball is equally likely to land in any one of the 38 spots.
Suppose a player bet $10 on red. In other words, he will win $10 if the ball lands on one of the 18 red spots, but will lose $10 if the ball lands on the other 20 spots.
So the player will win $10 with a 18 out of 38 chance,
And the player will lose $10 with a 20 out of 38 chance.

## Roulette

average value = V(A) x P(A) + V(B) x P(B)

Outcome A: player win \$10 V(A) = \$10, P(A) = $\frac{18}{38}$

Outcome B: player lose \$10 V(B) = -\$10, P(B) = $\frac{20}{38}$

The average value: \$10 x $\frac{18}{38}$ + (-\$10) x $\frac{20}{38}$ = -\$0.526

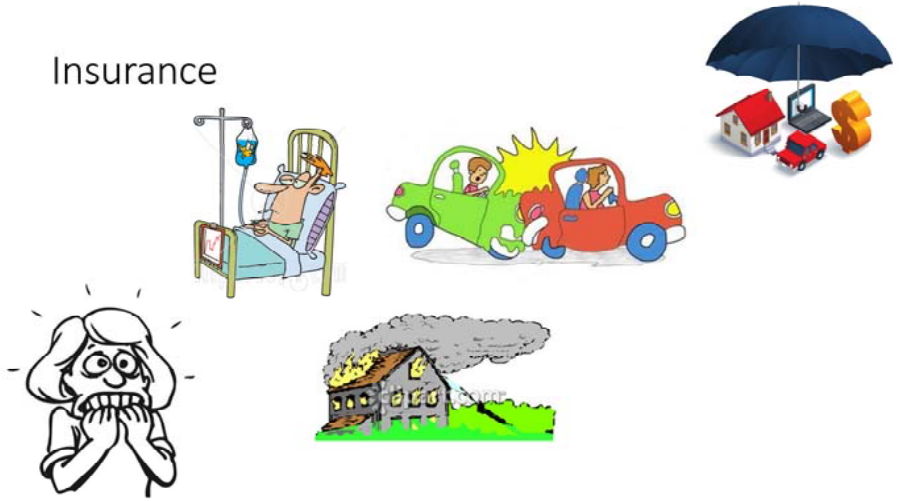
player "lose about 52 cents" in the long run

Let's work out the average gain of the player.
In this activity, there are two possible outcomes:
Outcome A is when the player win \$10. Hence we have V(A) = 10 with probability P(A) = 18/38.
Outcome B is when the player lose \$10. In this case V(B) = -10, where the negative sign indicate making a loss, and the probability P(B) = 20/38.

Hence, the average value can be computed as -\$0.526
In other words, on average, the player will lose about 52 cents by making the bet.
This does not mean the player will actually lose 52 cents on any one bet. He will either win \$10 or lose \$10.
However, if he keeps betting on this over and over again, in the long run, he will lose about 52 cents for each bet.



## Insurance



Let's move on to our next example: Insurance.
When you buy insurance you're gambling.
In this case the gamble is one you hope to lose: you don't want to get sick, or have your house burn down, or lose your car.
In each of those situations, if you have bought insurance, you've made a small advance payment in order to cover your losses in case a catastrophic event with small probability happens.

## Insurance


Use probabilities to determine premium for policies

When considering insurance,
a potential customer will ask:
"In the long run, am I likely to pay more than what I can claim? Or receive more than what I pay?"
On the other hand, insurance companies use probabilities to determine the premium for their policies, taking into consideration the amount they need to cover their expenses and make a profit.
In the long run, on average, their customers never get all their money back.



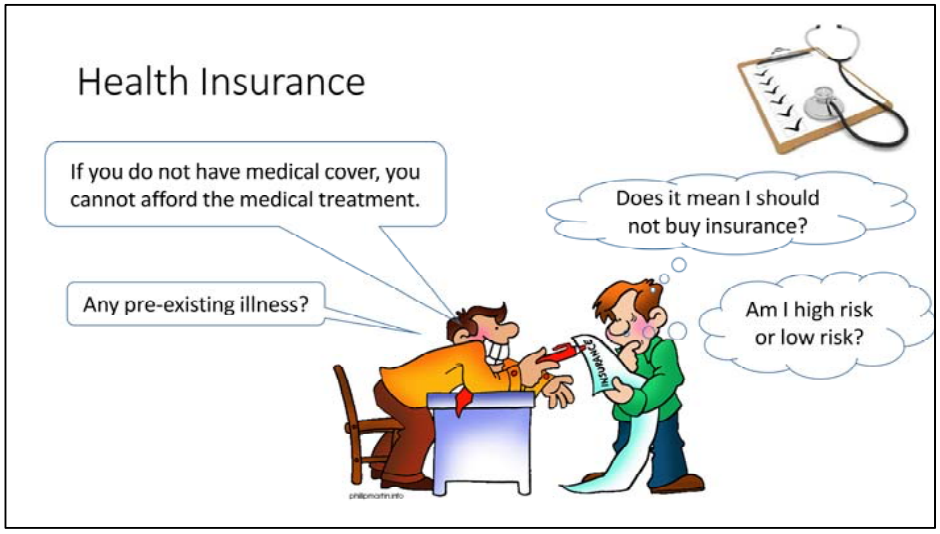
## Health Insurance

- Premium: $500 per year
- Payout: $10,000 per year
  - 2% customers claim $10,000 a year
  - 5% customers claim $5,000 a year
  - The rest make no claim

| Outcomes | Amount gain | | Probability | |
|---|---|---|---|---|
| A: Claim $10,000 | -$9500 | V(A) | 0.02 | P(A) |
| B: Claim $5,000 | -$4500 | V(B) | 0.05 | P(B) |
| C: No claim | $500 | V(C) | 0.93 | P(C) |

Average Gain (insurance company) = Average loss (each customer)
= (-$9,500) x 0.02 + (-$4,500) x 0.05 + $500 x 0.93 = $50

Take a simple case: An insurance company charged $500 per year for a certain health insurance policy with a maximum payout of $10000 when customer makes a claim.
For simplicity, suppose each year, 2% of the customers receive a full claim of $10,000, 5% receive half claim of $5000, while the rest of the customers do not submit a claim. Will the insurance company make money by selling this policy, and how much can the insurance company expect to make?
Let's set up a table showing the three possible outcomes of the policy.
There is either a full claim, a half claim or no claim for each policy. Lets call them outcome A, B and C respectively.
The second column indicates the amount gained by the company for each of the three outcomes. For outcomes A and B, the negative amount indicates a loss, as the company receives $500 from a customer, but pays out $10,000 and $5000 respectively to the customer who makes a claim. We denote the quantities by V(A), V(B) and V(C).
The third column gives the probabilities for the three outcomes. We are given that each year, there are 2% full claim, so we may take the probability that a customer will make a full claim to be 0.02. Similarly, there are 5% half claim, so the probability that a customer will make a half claim will be 0.05. By complement rule, the probability that a customer will not make a claim is 0.93. We denote the probabilities by P(A), P(B) and P(C).
We are now ready to compute the average gain of insurance company, which work out to be $50. This same amount also represents the average loss of each customer. Given the profitability of insurance company, it is certain that, on average, customers will pay more than they will receive.




Does this mean you should not buy insurance?
Though mathematically, you will make a loss in the long run, the real decision is not as straight forward.
For example, when you fall ill and do not having the medical cover, you may not be able to afford the medical treatment, which may lead to other disastrous consequences. So you might have no choice but to buy the insurance.
The probabilities used for computing the average gain is based on the insurance company's data.
Individual customers should take into account other risk factors to assess whether he or she has a higher or lower chance of falling ill. That will affect the average gain or loss of the customer.
On the other hand, insurance company usually will ask the customer to make declaration on relevant factors that might increase the customer's chance of falling ill, and adjust the premium accordingly.

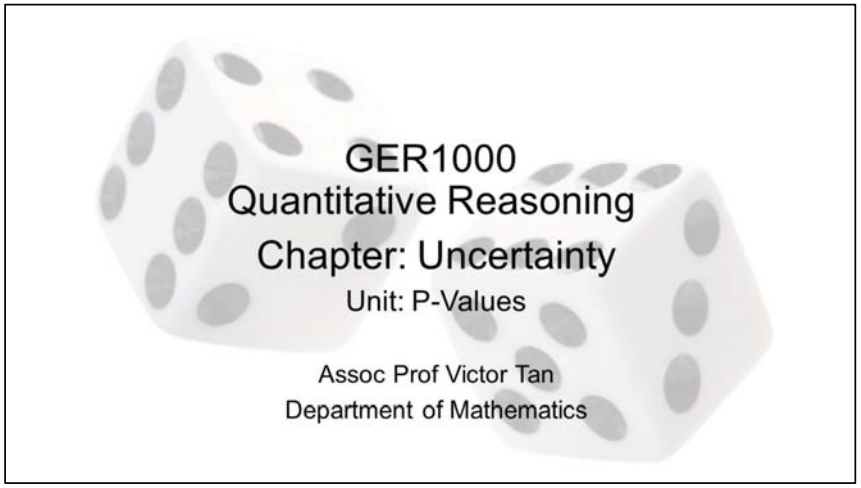


## Summary

- Identify the possible (mutually exclusive) outcomes of the activity:
  A, B, C, …
- Determine the value of each outcome: V(A), V(B), V(C), …
- Determine the probability of each outcome: P(A), P(B), P(C), …
- Compute the average value:

$$V(A) \times P(A) + V(B) \times P(B) + V(C) \times P(C) + \dots$$

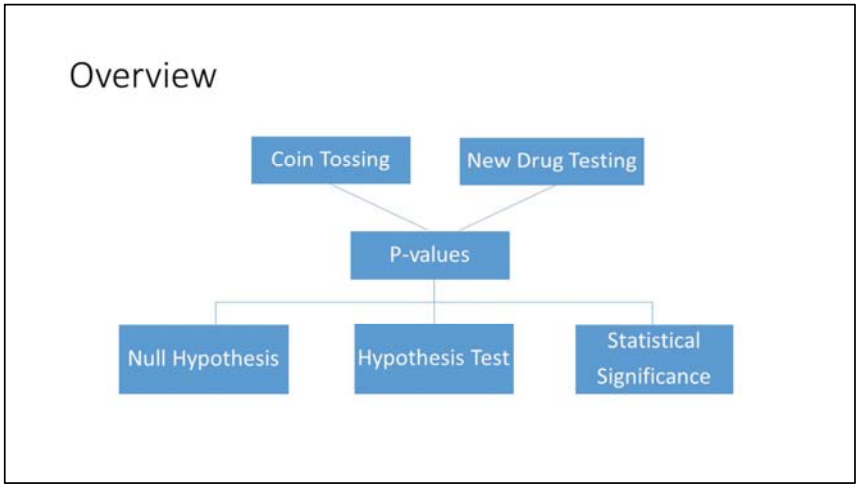
Let's do a quick recap of how to find the average value of a random activity.
First of all, identify what are all the possible outcomes of the activity. Make sure these outcomes are mutually exclusive.
Determine the value of each outcome;
as well as the probability of each outcome.
Then we compute the average value by taking the sum of the product of value and probability of every outcome.



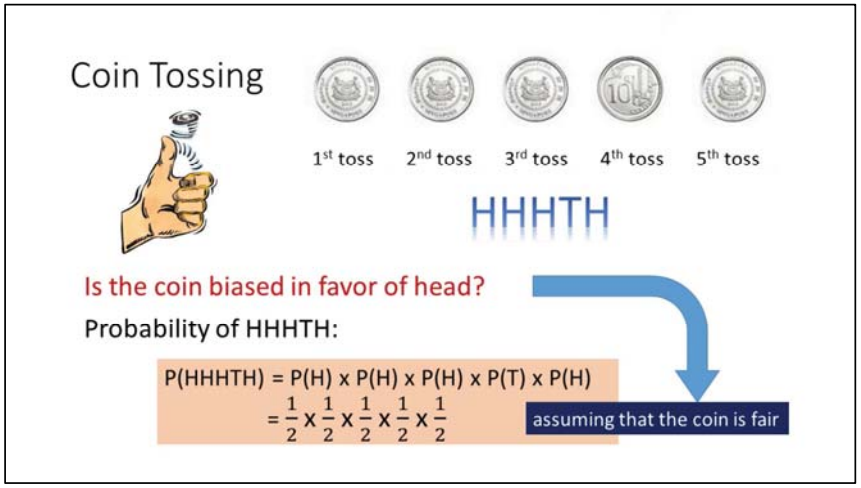

Suppose you toss a coin five times in a row and observe a head four out of 5 times. Will you suspect that the coin might not be a fair one?
To answer this question, we shall introduce the concept of p-value, which is a probability based on our observation.

P-values are commonly used in many fields of science and social sciences, such as economics, psychology, biology, criminal justice and sociology.



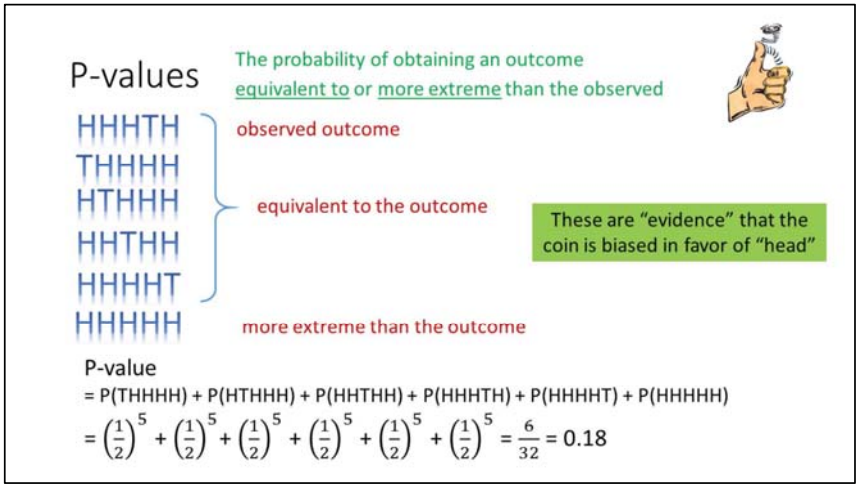



In this unit, we shall be using two examples to illustrate the concepts of P-values.
In the coin tossing example, we want to know whether a certain coin is fair, and in the new drug testing example, our objective is to determine whether a certain new drug is effective on patients.
We shall see how the P-values will be used in a process called hypothesis testing to give us the desired answers.
Along the way, we will also introduce the notion of null hypothesis and statistical significance.




Let's look at our coin tossing example.
You toss a coin five times and suppose the observed outcome is head, head, head, tail, head in that order.
Let's simply refer to this outcome as HHHTH
Can you conclude that the coin is biased in favor of head? In other words, can we say that the coin is not fair?
To answer this question, let us compute the probability of what we have observed.
Since the 5 tosses of the coin are independent of each other, we can apply multiplication rule.
In other words, the probability of HHHTH is just the product of the probabilities of the individual outcomes.
Now let us assume that the coin is fair. Then, base on this assumption,
the probability of tossing a head, and the probability of tossing a tail *are both equal to* $\frac{1}{2}$
.
So the probability of HHHTH is $\frac{1}{2}$ multiplying itself 5 times.
Note that at this point, it is not certain that the coin is fair. Nevertheless, we need to make this assumption in order for us to compute the desired probability, which will in turn help us decide whether it is reasonable to make this assumption or not.






The probability that we have computed on the previous slide is not sufficient for us to make conclusion.
We need to also consider a few more outcomes that are related to the one observed.
Recall that our observed outcome is HHHTH.
If we take the order into consideration, then THHHH is another equivalent outcome.
In fact, there are altogether 5 outcomes that are equivalent to the one observed.
These are evidence which support that the coin is biased in favor of head, and hence is not fair.
Of course, any outcome that is more extreme than the one observed will also serve as such an evidence.
In our case, the outcome HHHHH is more extreme as it consists of more than 4 heads.
Now we are ready to define P-value. It is the probability of obtaining an outcome that is equivalent to or more extreme than the observed.
For our example, the P-value is the sum of the probabilities of all the 6 outcomes listed above.
Note that we apply the addition rule here.
To compute the probability of each outcome, we can apply multiplication rule as before.
So the probability of each term is just $\frac{1}{2}$ multiplying itself 5 times.
Therefore the P-value can be computed as such, which gives a value of 0.18.
How do we use this P-value to make conclusion whether the coin is fair? Before answering this question, we need to introduce the notion of null hypothesis.

## Null Hypothesis

P-value

$= P(THHHH) + P(HTHHH) + P(HHTHH) + P(HHHTH) + P(HHHHT) + P(HHHHH)$

$= \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{6}{32} = 0.18$

Null Hypothesis: the coin is fair

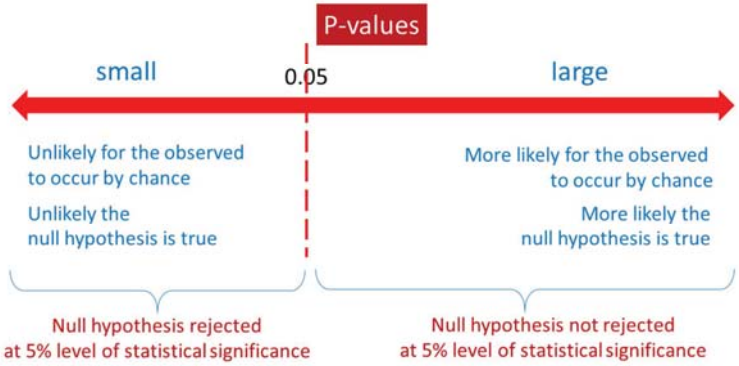
If the null hypothesis is true,
when we toss the coin "many" times, there will be roughly the same number of heads and tails;

purely due to chance →


Recall that, when we compute the probability here, we need to assume that we have a fair coin.
In other words, the P-value that we have computed is based on the assumption that the coin is fair.
This assumption is called the null hypothesis. We want to determine whether this hypothesis is true or false, based on our observation.
If our null hypothesis is true, then when we toss the coin again and again many times, we should expect roughly equal number of heads and tails. In this case, the disproportionate occurrence of heads and tails that we have observed is purely by chance, and not because the coin is not fair.



## What do P-values tell us ?




We are now ready to interpret the P-value.
If the p-value is small, that means it is unlikely for what we have observed to occur by chance.
This in turn means that it is unlikely for the null hypothesis to be true.
In other words, we have strong evidence against the null hypothesis.

On the other hand, if the p-value is large, this will mean it is more likely that what we have observed occur by chance, which mean we do not have strong evidence against the null hypothesis.

But how do we quantify large and small p-values?
For many studies, the traditional standard is 0.05, or 5%.
In other words, for a P-value less than 0.05, the null hypothesis will be rejected. We say that it is rejected at 5% level of statistical significance.
For a p-value larger than 0.05, we say that the null hypothesis is not rejected at 5% level of statistical significance.
However, this "standard" value is rather arbitrary.
Some statisticians feel that the P-value should be less than 0.01 in order to be statistically significant.

## Statistical Significance

Using 5% level of statistical significance

| Observed: HHHTH | Observed: HHHHH |
|---|---|
| P-value = 0.18 > 0.05 | P-value = $\left(\frac{1}{2}\right)^5$ = 0.03 < 0.05 |
| Do not reject null hypothesis at the 5% significance level | Reject null hypothesis at the 5% significance level |
| Cannot conclude that the coin is not fair. | Can conclude that the coin is not fair. |

In our example, suppose we adopt the 5% level of statistical significance.
With the outcome HHHTH that we have observed, the P-value = 0.18, which is greater than 0.05. So we do not reject the null hypothesis at the 5% significance level. In other words, we do not have enough evidence to conclude that the coin is not fair.
On the other hand, suppose we repeat the experiment and this time the outcome is 5 heads (HHHHH), then the P-value would be just the product of 5 individual probabilities of getting a head. This gives a value of 0.03, which is smaller than 0.05.
In this case, we reject the null hypothesis at the 5% significance level. In other words, this time we are more confident to conclude that the coin is not fair.



## Hypothesis Testing

1. Identify the question: is the coin biased? **Frame**
2. State the null hypothesis: the coin is fair **Specify**
3. Conduct the experiment: toss the coin five times and observe the number of heads **Collect**
4. Compute P-value: probability of outcomes that are equivalent or more extreme than observed **Analyse**
5. Make conclusion about the null hypothesis: whether to reject that the coin is fair at certain level of statistical significance **Communicate**

The whole process above is known as hypothesis testing, which is aligned with the QR framework.
The first step is to identify what we are testing, In our example, we want to test whether the coin is biased in favour of head. This is framing in the QR framework.
The second step is to state the null hypothesis. For our example, the hypothesis is that the coin is fair. This correspond to specifying the assumption we made in our study.
Next step is to carry out the experiment. In our case, that is simply the activity of tossing the coin 5 times and count the total number of heads. This is collecting the data.
The 4th step is to compute the P-value, which is a probability based on the data we have collected as well as the hypothesis. This is analysing.
In the final step, we use the P-value to conclusion whether there are enough evidence to reject the null hypothesis. In our example, we make claim whether the coin is fair or not, based on the P-value calculated and the level of statistical significance. This is communicating.

## New Drug Testing

- Disease D, 40% fatal.
- New drug claimed to reduce fatality level.
- Three patients with D given the drug. 
- Suppose all three patients survive.
- Can we be sure the drug is effective? Or is it just coincident?

In an earlier chapter on Design of Studies, we talked about testing a new drug against an old one, which involves two samples.
Here we look at another example of drug testing using p-values.
A certain disease, let's call it D, is fatal in 40% of cases from a population.
A pharmaceutical company develops a new drug that they claim reduce the fatality level of D.
To test the claim, three patients from the population suffering from D are given the drug.
Suppose all three patients survive the disease.
You might think: without the drug, 40% of patients die, but with the drug, it seems that all survive. So the drug must be effective.
Can we be sure about this conclusion? Or is it just coincident?
i.e. Did all the three patients get better purely by chance, and the drug had no effect at all?



## Hypothesis Testing

1) Identify the question : Is the drug effective in some population?
2) State the null hypothesis (assumption): the drug has no effect in that population
3) Conduct the experiment: give drug to three patients and observe number of patients survived 

4) Compute P-value: next slide
5) Make conclusion about the null hypothesis: whether the drug is effective in that population at certain level of statistical significance

We can again use the hypothesis testing.
We want to know whether the new drug is effective in a certain population.
So we make the null hypothesis that the new drug has no effect in that population.
Then the experiment to be carried out is to give the new drug to 3 patients randomly chosen from the population, and observe how many patients survived the disease.
Note that this is a sampling experiment as we have seen in the previous chapter. Here the 3 patients given the drug is a sample taken from the population concerned.
Next we analyze the data by computing the P-value.
Finally we make conclusion whether the drug is effective in that population at a certain level of statistical significance.

## P-values

Disease D, 40% fatal

Suppose drug has no effect (null hypothesis).
40% fatal → 60% survived
P(a patient survive) = 0.6
P(all three patients survive) = 0.6 x 0.6 x 0.6 = 0.216
P-value = 0.216

Most extreme outcome

Using 5% level of statistical significance
P-value > 0.05
Do not reject the null hypothesis at 5% significance level.
Does not establish that the drug is effective.

In order to compute the p-value, we need to know the probability of a patient from the population that survive the disease.

We are given that disease D has a 40% fatality rate.
So, suppose the drug has no effect on the patient from the population, then the patient would have a 60% chance of surviving.
In other words, the probability is 0.6.
It follows that the probability of all three patients surviving the disease would be 0.216.
Since what we have observed is the most extreme outcome, we conclude that in the study, the drug seems to reduce the disease's fatalities with a P-value of 0.216.
Suppose we use the 5% level of statistical significant as bench mark.
Since the P-value is greater than 0.05, we do not reject the null hypothesis at 5% significance level.
Thus this study does not establish that the drug is effective in reducing the fatality of disease D in the population.



## Sample size

In the above study, sample size = 3.

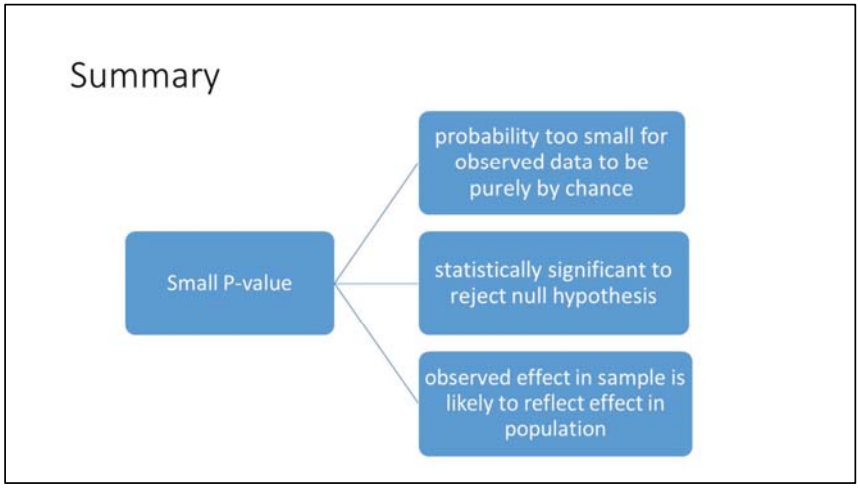
No conclusion about effectiveness of the drug.
Increase the sample size:

| Sample size | P-value (all survive) | Conclusion |
|---|---|---|
| 3 | 0.216 | No |
| 4 | 0.1296 | No |
| 5 | 0.0778 | No |
| 6 | 0.0467 | Yes |

With sample size 6 and all patients survive, we may conclude that the drug is effective.

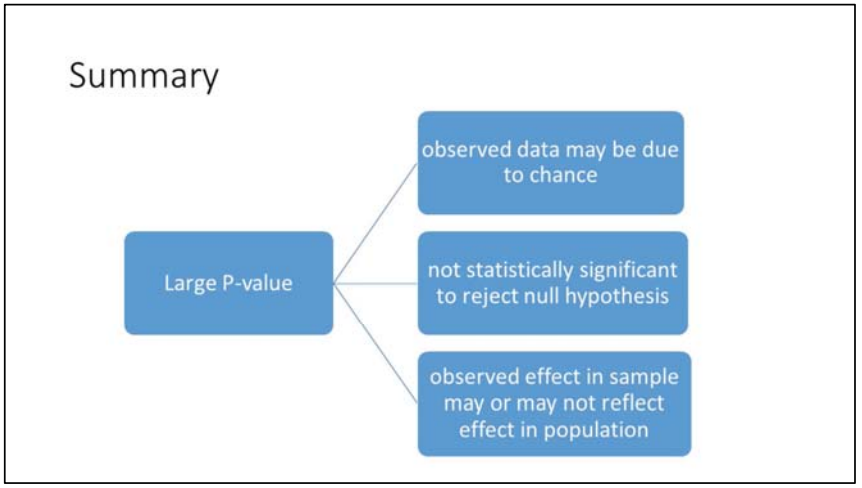
In the above study, we only tested the drug with three patients. In other words, the sample size is 3.
The result does not lead to a clear conclusion about the effectiveness of the drug.
If we increase the sample size, there is a higher chance for us to make a strong conclusion.
In this table, we show the respective P-values of sample sizes from 3 to 6, assuming all the patients in each sample survive the disease. We see that, as the sample size increases, the p-values decreases.
When the sample size is 6, with all 6 patients survived. The p-value is less than 0.05 and hence the test is statistically significant at 5% level.
Therefore, we have a stronger evidence to conclude that the drug reduced the fatality of disease D in the whole population.

What if not all patients survive? For example, if 2 out of the 6 patients who took the drug died of the disease, then to calculate the P-value, we need to include outcomes that are equivalent and more extreme to the observed, namely, the probabilities of 4 patients surviving, 5 patients surviving and 6 patients surviving. This will again raise the p-value, and will change the conclusion of the test.
In practice, for drug test or other scientific studies, the sample size is usually much bigger.

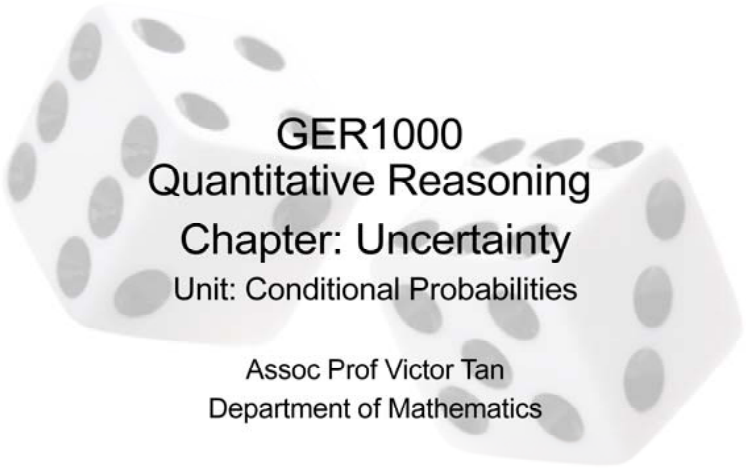



Here's a summary of this unit.
We introduce the concept of p-value and its role in hypothesis testing.
You have seen the definition and computation of p-values. But most importantly you should know how to use it to make correct interpretation.
When P-value is small enough, it is unlikely that the observed data or outcome is purely by chance.
This means it is unlikely that the null hypothesis is true. Therefore, it is statistically significant to reject the hypothesis.
Consequently, the observed effect in the sample is likely to reflect the effect in the population.






On the other hand, when P-value is big, it is probable that the observed data or outcome is purely by chance.
This means we do not have enough evidence to say the null hypothesis is false.
So it is not statistically significant to reject the hypothesis. In this case, what we observe in the sample may or may not reflect the actual effect in the population.



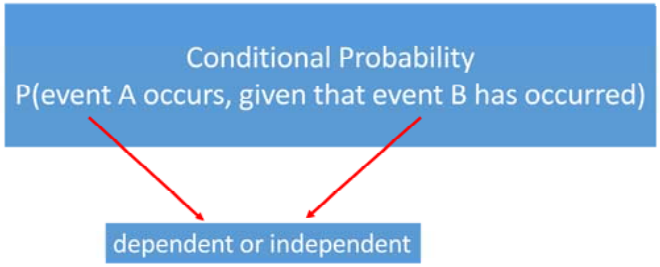
# GER1000
# Quantitative Reasoning
## Chapter: Uncertainty
### Unit: Conditional Probabilities

Assoc Prof Victor Tan
Department of Mathematics

Suppose you have taken the mid-term test of a module and received an A grade.
Based on this information, you want to know the chance for you to get an A grade for the final.
In this scenario, getting an A grade for final is uncertain, while getting an A grade for mid-term is confirmed.
Calculating probability of an uncertain event based on some known relevant information is very common in real situation.
Such probabilities are called conditional probabilities.



## Overview

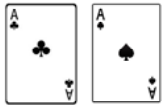



Conditional probability involves two events: the uncertain event that we are interested to find the probability, and the certain event that we use as additional information.
We shall see that this probability depends on whether the two events are dependent or independent of each other.

## Drawing a Card

52 cards

- Randomly draw a card from the deck.
- What is the probability of getting a black ace?
- P(black ace) = $\frac{2}{52} = \frac{1}{26}$.



2 black aces

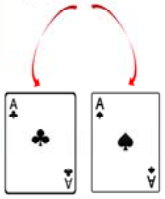
Let's begin with the example of drawing a card from a deck of 52 playing cards.
Suppose the deck of card is shuffled and you randomly draw a card from the deck.
What is the probability of getting a black ace?
Since there are two black aces among the 52 cards,
the probability of drawing a black ace is 2/52 or 1/26.



## Drawing Two Cards (without replacement)

| First card drawn |  |  |
|---|---|---|
| No. of black ace left | 2 | 1 |
| P(second card is black ace) | $\frac{2}{51}$ ≠ | $\frac{1}{51}$ |

51 cards



Conditional probability

Now suppose we draw two cards consecutively, without replacement.
If the first card is the queen of hearts, without replacing the card, what is the probability of drawing a black ace for your second card?
Since there are still two black aces among the remaining 51 cards,
so the probability of second draw being a black ace is $2/51$
Repeat the activity again and suppose this time the first card is the ace of clubs.
This time there is only one black ace left among the remaining 51 cards,
so the probability of second draw being a black ace is $1/51$.
We see that the value of the probability of this event depends on the outcome of the first draw. This is an example of conditional probability.

## Leaving the room

Event: 2nd is boy
(Second person leaving the room is a boy)

- If the first person leaves the room is a boy, then $P(2^{nd} \text{ is boy}^*) = \frac{1}{2}$
- If the first person leaves the room is a girl, then $P(2^{nd} \text{ is boy}^{**}) = 1$

additional information

Conditional probabilities

Here is another simple example.
Suppose there are three person in a room, two boys and one girl.
They are to leave the room one at a time in random order.
Let's say we are interested in the event that the second person leaving the room is a boy.
What is the probability of this event?
Suppose we know that the first person leaving the room is a boy, then there will be one boy and one girl left.
So this probability is equal to $1/2$ .
On the other hand, suppose the first person leaving the room is a girl, then there will be two boys left. So we are certain that the second person must be a boy, and hence the probability is equal to $1$ .
Note that the information about the first person have an effect on the probability of second person leaving the room being a boy.
So the probabilities that we get are actually conditional probabilities given the gender of the first person leaving.



## Dependent events

Events A and B are dependent

Event B occurs

P(event A) $\neq$ P(event A, given event B)

Conditional probability

When we carry out an activity, the probability of an event usually depend on what is known about the situation.
If we subsequently get some additional information, it may change the probability of the event.
In particular, if two events A and B are dependent, the occurrence of event B will affect the probability of the occurrence of event A.
Now suppose the additional information is that event B occurs.
Then the probability of event A, given that B occurs, will be different from the probability of event A without this additional information.
We call this the conditional probability.
So how do we calculate the conditional probability of event A given that event B has occurred?

## Conditional Probability

Two events: A and B

The probability of A occurring given that B has occurred:

- called the conditional probability of A given B
- denoted by P(A | B)

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} \quad \text{assume } P(B) > 0$$

$$P(A \text{ and } B) = P(B) \times P(A \mid B)$$

not equal

Multiplication rule for dependent events

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} \quad \text{assume } P(A) > 0$$

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

For any two events A and B,
the probability of A occurring given that B has occurred is called the conditional probability of A given B
and is denoted by P(A | B).
This conditional probability can be computed as the probability of A and B both occurring divided by the probability of B.
Essentially the conditional probability is the ratio of these two probabilities.
When considering conditional probability of A given B, we may assume that there is a chance for B to occur, and hence the probability of B will not be 0.
With a bit of arithmetic, we can rewrite the equation as P(A and B) = P(B) x P(A | B).
We may regards this as the multiplication rule for dependent events.
Please note the difference between P(A and B) and P(A | B).
By symmetry, we have a similar formula for the conditional probability of B given A, and the corresponding multiplication rule.
Note that, in general, the probability of A given B is not the same as the probability of B given A.



## Leaving the Room

All possible orders by gender:

i. 1st boy, 2nd boy, 3rd girl
ii. 1st boy, 2nd girl, 3rd boy
iii. 1st girl, 2nd boy, 3rd boy

$P(1^{st} \text{ is boy and } 2^{nd} \text{ is boy}) = \frac{1}{3}$ (case i)

$P(1^{st} \text{ is boy}) = \frac{2}{3}$ (cases i and ii)

$$P(2^{nd} \text{ is boy} \mid 1^{st} \text{ is boy}) = \frac{P(1^{st} \text{ is boy and } 2^{nd} \text{ is boy})}{P(1^{st} \text{ is boy})} = \frac{1}{2}$$

dependent events

Let's come back to this example, and look at all possible order by gender that the three person leave the room:
Case 1 is boy, boy, girl; case 2 is boy, girl, boy; case 3 is girl, boy, boy.

The event that the first two person leaving are boys occurs in only one out of the three cases. Hence the probability of this event is 1/3.
The event that the 2nd person leaving is a boy occurs in two out of the three cases. Hence the probability of this event is 2/3.
According to our formula, the conditional probability that 2nd is a boy given that 1st is a boy can be computed by taking the ratio of the two probabilities above, which gives the value ½.

I leave it for you to find the conditional probability of 2nd is a boy given 1st is a girl using this approach.
Note that the event that the 2nd is a boy and the event that 1st is a boy are dependent on each other.

## Independent Events

Events A and B are independent $P(A \text{ and } B) = P(A) \times P(B)$

$P(A \mid B)$ given that event B occurred $=$ $P(A)$ without given event B occurred

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

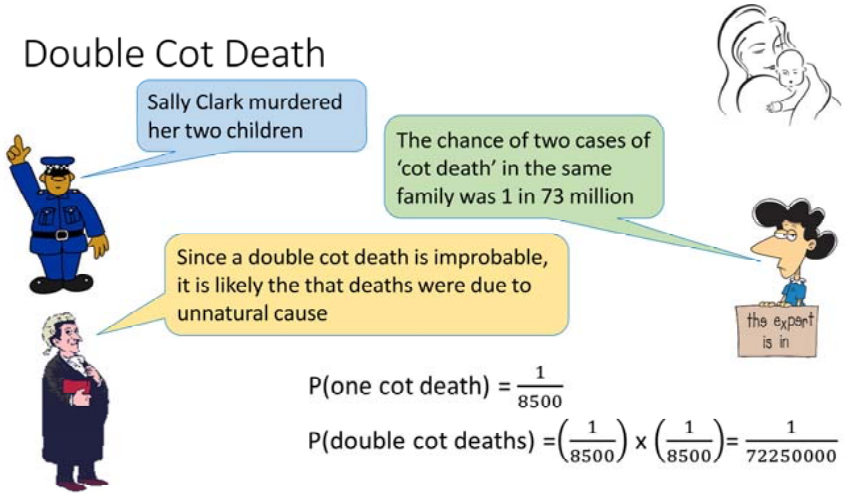
$P(B \mid A)$ given that event A occurred $=$ $P(B)$ without given event A occurred

If events A and B are independent,
then the conditional probability of A given B is the same as the probability of A.
Note that in the left hand side, the information that event B occurred is given, while the right hand side is without the information.
In other words, with or without knowing event B has occurred, it does not affect the probability of A occurring, if A and B are independent events.
This can be verified mathematically using multiplication rule.
Recall that the probability of both A and B occurring is the product of the individual probabilities of A and B occurring.
Then, by our formula for conditional probability, P of A given B can be reduced to P of A as shown.

Similarly, the probability of B given A is the same as the probability of B, when A and B are independent events.



## Double Cot Death




$$P(\text{one cot death}) = \frac{1}{8500}$$

$$P(\text{double cot deaths}) = \left(\frac{1}{8500}\right) \times \left(\frac{1}{8500}\right) = \frac{1}{72250000}$$

This is a classical example involving conditional probability of a real incident took place in 1999.
Sally Clark was charged for murdering her first two children. Her first child died at the age of three months. and second child died at the age of two months
A pediatrician gave expert testimony claiming that the chance of two cases of 'cot death' in the same family was 1 in 73 million.
In other words, a double cot death is highly unlikely, which implies that it is likely the deaths were due to unnatural cause.
How did the expert came up with the figures?
Base on available data: a baby born into family similar to the Clarks has a 1 in 8500 chance of dying in a cot death, i.e. P(one cot death) = $1/8500$.
The expert then calculated the probability of two deaths using the multiplication rule: which gives the probability of double cot deaths to be $1/72250000$

## Independent VS Dependent

A: "1st child dies"
B: "2nd child dies"

Are these two events independent?

Expert's calculation: assume A & B independent
P(double cot death) = P(A and B) = P(A) x P(B)

assumption made by expert

If A and B are dependent:
P(double cot death) = P(A and B) = P(A) x P(B | A)

Consider the two events:
A is the event that the 1st child dies; and B the event that the 2nd child dies.
According to the expert's calculation, the probability of double cot death, which is the probability that A and B happening at the same time, is the product P(A) x P(B).
This is the multiplication rule based on the assumption that A and B are independent events

But could the death of the 2nd child be actually dependent on the death of the 1st child?
If that is the case, then P(double cot death) should be P(A) x P(B | A)

We can see that the expert had made the assumption that:
P(B | A) = P(B), which is valid only when A and B are independent.
So before making any conclusion, we should question whether the death of two children from the same mother are really independent?



## P(D | E) VS P(E | D)

D: "double cot deaths"
E: "Sally is innocent"

not equal

| P(D \| E) | P(E \| D) |
|---|---|
| Observing the evidence of double cot deaths given that Sally is innocent | Sally is innocent given the evidence of double cot deaths |
| Probability quoted by the expert | Probability needed by the prosecutor |
| Small value ⇒ chances for double cot death is small | Small value ⇒ chances for Sally to be innocent is small |


There is another flaw in the use of the probabilities in this case:
Consider these two events.
D is the event of double cot deaths; and E the event that Sally is innocent.
How do we interpret the conditional probability of D given E, and the probability of E given D?

P(D | E) is the probability of observing the evidence of double cot deaths, given that Sally is innocent.
This is the probability quoted by the expert, which is 1 in 73 million.
P(E | D) is the probability that Sally is innocent given the evidence of double cot deaths.
This is the probability needed by the prosecutor.
To prosecute Sally, the prosecutor needs to prove that she is guilty beyond reasonable doubt based on the evidence.
What the prosecutor needs is a small value of P(E | D), which means the chance that Sally is innocent based on the given evidence is reasonably small.
The probability P(D|E) provided by the expert is insufficient, even if it is a small value. In general, the two probabilities are not equal.

## From earlier chapter

### Smoking and heart disease (HD)

| | Heart disease | No heart disease | Row total |
|---|---|---|---|
| Smokers | 38 | 14,962 | 15,000 |
| Non-smokers | 44 | 84,956 | 85,000 |
| Column total | 82 | 99,918 | 100,000 |

rate(smoke | HD) = rate of smoking among HD = $\frac{38}{82}$ × 100% = 46.3%

Conditional probability of a person being a smoker given that he has HD

rate(HD | smoke) = rate of HD among smokers = $\frac{38}{15,000}$ × 100% = 0.25%

Conditional probability of a person having HD given that he is a smoker

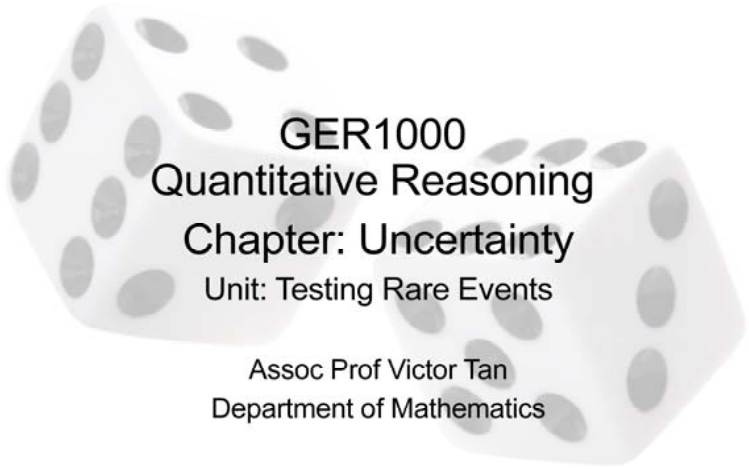
Before I end this unit, let me refer you back to an example in the earlier chapter on Design of Studies.
Recall the example of association between smoking and heart disease.
From the given two-by-two table, we computed the rate of smoking among people with heard disease.
You should realize that this is the same as the conditional probability of a person being a smoker given that he has heart disease.
Similarly, the rate of heart disease among smokers is the same concept as the conditional probability of a person having heart disease given that he is a smoker.



## Summary

- P( A | B) is the conditional probability of event A given that another event B has occurred.
- $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$
- If A and B are dependent, then $P(A \mid B) \neq P(A)$.
- If A and B are independent, then $P(A \mid B) = P(A)$.
- In general, $P(A \mid B) \neq P(B \mid A)$.

Here is a quick summary of this unit.
A conditional probability P( A | B) measures the probability of event A given that another event B has occurred
and it is given by the quotient of two probabilities.
If A and B are dependent events, then P(A|B) is not equal to P(A),
If A and B are independent events, then P(A|B) is equal to P(A).
In general P(A|B) is not equal to P(B|A).



# GER1000
# Quantitative Reasoning
## Chapter: Uncertainty
### Unit: Testing Rare Events

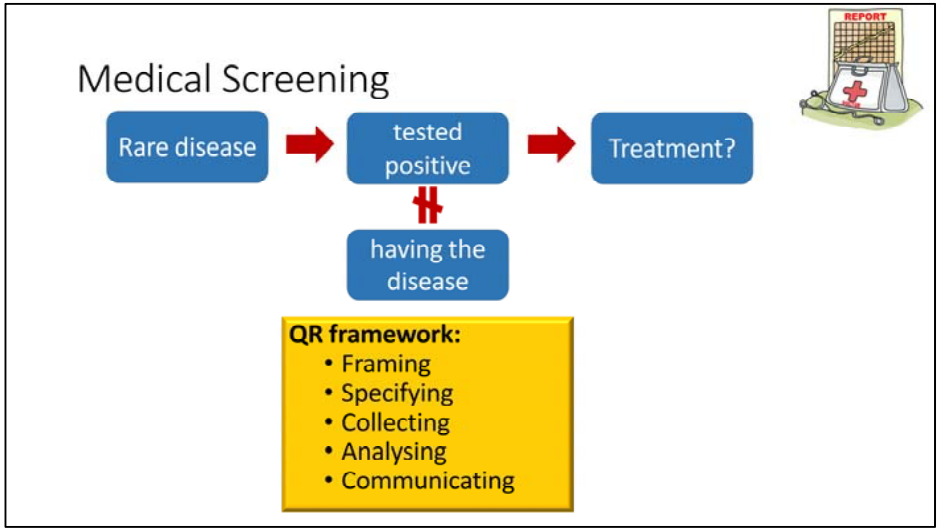
Assoc Prof Victor Tan
Department of Mathematics

In this last unit of the chapter om Uncertainty, we will continue with the discussion of the concept of conditional probability.
In particular, we will apply the concepts in the circumstances of testing or detecting rare events which has a very low chance to occur.



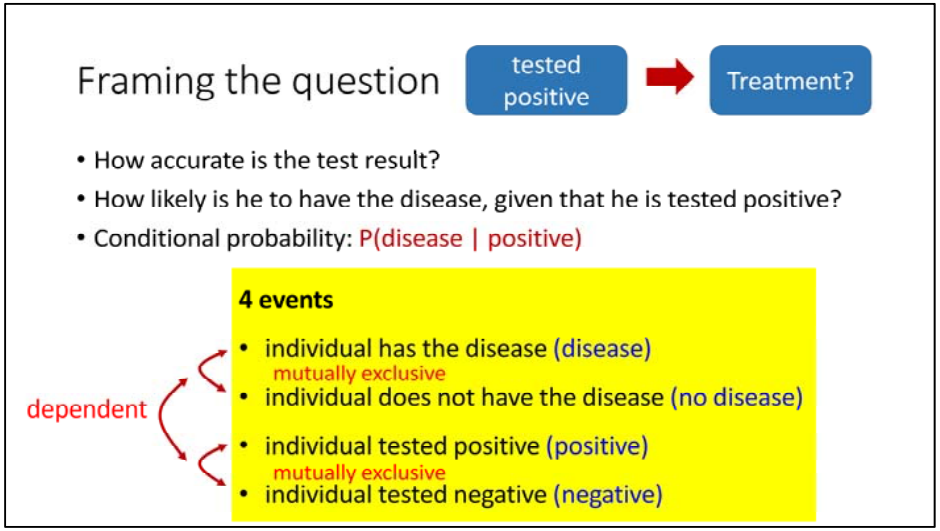
## Overview


This unit will be focused on an example of medical screening of rare diseases.
We will begin with a question about going through treatment of the disease based on a medical screening result.
To answer the question, we will go through the 5 step process in the QR framework.
Along the way, we will need to apply the concepts of conditional probability.




Here is a hypothetical example on medical screening.
Suppose one of your family members has gone through a medical screening for a certain rare disease and was tested positive.
Would you advise this family member to go for a treatment?
Before you say yes, do note that being tested positive for the disease does not mean the person actually has the disease.
In other words, the test result for the medical screening may not be 100% accurate.
To help us make a decision quantitatively, we shall go through the QR framework.
Let's recall the 5 steps: Framing the question, Specifying what and how to measure, Collecting the data, Analysing the data, and Communicating the findings
.






Let's try to frame the question quantitatively.
In order to decide whether to go for treatment, we need to know whether the test result is accurate.
But since there is uncertainty in the screening test, we can't give a definite answer of yes or no.
Instead, we should rephrase our question as, how accurate is the test result.
To be more precise, we should ask: How likely is he to have the disease, given that he is tested positive?
So we are looking at probability.
Let's consider 4 events, namely: individual has the disease; individual does not have the disease; individual tested positive; and individual tested negative.
In short, we denote the 4 events by: disease, no disease, positive. and negative respectively.
Note that the events "disease" and "no disease" are mutually exclusive; and the events "positive" and "negative" are also mutually exclusive.
On the other hand, the events of "disease" and "no disease", are dependent of the events "positive" and "negative".
So, our question can now be phrased mathematically as finding the conditional probability of "having the disease" given "the test is positive".

## Specifying what to measure

Want to find: P(disease | positive)

| | |
|---|---|
| Base Rate | P(disease) |
| Sensitivity of the test | P(positive \| disease) |
| Specificity of the test | P(negative \| no disease) |

To determine this probability, there are three things that we need to measure: the base rate, the sensitivity of the test, and the specificity of the test.
Base rate is the probability that an individual is likely to have the disease, without any knowledge of the test results. This is an unconditional probability.
Sensitivity of the test is the probability that an individual is correctly tested positive when he or she actually has the disease. So this is a conditional probability.
Specificity of the test is the conditional probability that an individual is correctly tested negative when he or she doesn't have the disease.
Recall that, our question is to find the conditional probability of having the disease given the test is positive.
Note that this is not the same as the conditional probability of testing positive given that the person has the disease.



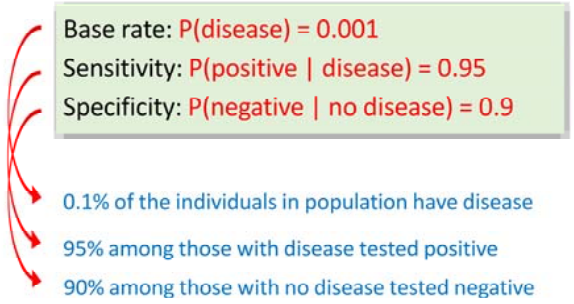
## Collecting the data

1. Carry out a study on a random sample (N) from a population
2. Record the number of people with disease (A) and those without disease (B)
3. Among those with disease, record the number of people that are tested positive (C).
4. Among those without the disease, record the number of people that are tested negative (D).

Base rate: P(disease) = A/N
Sensitivity: P(positive | disease) = C/A
Specificity: P(negative | no disease) = D/B

The collection of such data is usually carried out by experts through a clinical study. Essentially this involves studying a random sample from a population. Say the sample size is the number N.
From the sample, determine and record the number of people having the disease and those who do not have the disease. Let these two numbers be A and B respectively.
Then among those with the disease, record the number of people that have been tested positive. Let this number be C.
Last but not least, among those without the disease, record the number of people that have been tested negative. Let this number be D.
Then we can use the above numbers to get the information we want, namely:
the base rate is given by A/N; the sensitivity is C/A; and the specificity is D/B.

## Collecting the data (example)




Let's assume the following are the data that we have collected:
Base rate is 0.001.
In other words, about 0.1% of the individuals in a certain population have the disease (or every 1 in 1000).
Sensitivity is 0.95. This means, of those who have the disease, 95% is tested positive.
Specificity is 0.9. which means among those who do not have the disease, 90% is tested negative.



## Analyzing the data

Base rate: P(disease) = 0.001
Sensitivity: P(positive | disease) = 0.95
Specificity: P(negative | no disease) = 0.9

### Contingency Table

Assume a population of 100,000

| | Test positive | Test negative | Row sum |
|---|---|---|---|
| Have disease | 95 | 5 | 100 |
| Do not have disease | 9990 | 89910 | 99900 |
| Column Sum | 10085 | 89915 | 100000 |

We are now ready to analyse the data.
From the given data, we can set up a contingency table as shown.
For convenient, let us assume a population of 100,000.
(This number is chosen so that when we breakdown the numbers in the table, we will get whole numbers, and hence easier to work with.)
The rows in the table record the number of people in the population with and without the disease.
The columns in the table record the number of people in the population tested positive and negative.
Let's fill in the table.
We start with the last column: We put in the size of the population which is 100,000.
From the base rate, we get the total number of people having the disease to be 0.001 x 100,000 which gives 100.
From there, we get the rests of the population 99,900 without the disease by subtraction.
Now let's fill in the first row: among the 100 people with the disease, 95% were tested positive, as given by the sensitivity of 0.95.
In other words, there are 95 people. We deduce that the balance of 5 people were tested negative.
Let's move on to the second row: among the 99900 people without the disease, 90% were tested negative, as given by the specificity of 0.9.
So there are 89,910 people. And the balance of 9990 people were tested positive.
Finally, we can add up the numbers column-wise to get the total number of people tested positive and tested negative.

## Analyzing the data

Base rate: P(disease) = 0.001
Sensitivity: P(positive | disease) = 0.95
Specificity: P(negative | no disease) = 0.9

### Contingency Table

Assume a population of 100,000

True positive

False negative

| | Test positive | Test negative | Row sum |
|---|---|---|---|
| Have disease | 95 | 5 | 100 |
| Do not have disease | 9990 | 89910 | 99900 |
| Column Sum | 10085 | 89915 | 100000 |

False positive

True negative

Before we move on, let me also introduce 4 terminologies for future reference:
For people who are tested positive and have the disease, we call them the "true positive".
For people who are tested positive but do not have the disease, we call them the "false positive".
Similarly, For people who are tested negative and have the disease, we refer to them as the "false negative",
while for people who are tested negative and does not have the disease, we refer to them as the "true negative".



## Analyzing the data

Base rate: P(disease) = 0.001
Sensitivity: P(positive | disease) = 0.95
Specificity: P(negative | no disease) = 0.9

| | Test positive | Test negative | Row sum |
|---|---|---|---|
| Have disease | 95 | 5 | 100 |
| Do not have disease | 9990 | 89910 | 99900 |
| Column Sum | 10085 | 89915 | 100000 |

$$P(\text{disease} \mid \text{positive}) = \frac{\text{No. of true positive}}{\text{No. people tested positive}}$$

$$= \frac{95}{10085} = 0.0094$$

To compute the conditional probability that someone has the disease given that he is tested positive, we take the number of true positive divided by the total of number of people tested positive.
The two numbers are represented by these two cells in the table.
Hence we compute the value to be 0.0094, which is a very small probability.

## Analyzing the data

Base rate: P(disease) = 0.001
Sensitivity: P(positive | disease) = 0.95
Specificity: P(negative | no disease) = 0.9

| | Test positive | Test negative | Row sum |
|---|---|---|---|
| Have disease | 95 | 5 | 100 |
| Do not have disease | 9990 | 89910 | 99900 |
| Column Sum | 10085 | 89915 | 100000 |

$$P(\text{no disease} \mid \text{positive}) = \frac{\text{No. of false positive}}{\text{No. people tested positive}}$$

$$= \frac{9990}{10085} = 0.9906$$

Let's also compute the conditional probability that someone does not have the disease given that he is tested positive.
Similarly, we can take the number of false positive divided by the total of number of people tested positive, which gives us the value 0.9906.
This is a large probability, which means that, most of the time when people are tested positive, they do not have the disease.



## Communicating the findings

- The test has high sensitivity and specificity P(positive | disease) = 0.95 P(negative | no disease) = 0.9
- Less than 1% tested positive have the disease P(disease | positive) = 0.0094
- More than 99% tested positive have no disease P(no disease | positive) = 0.9906
- If tested positive, not confident that test is correct
- Happens whenever disease is rare P(disease) = 0.001

So how should we interpret and communicate the outcome of our analysis?
First of all, we have found that the test is very accurate in the sense that it has high sensitivity and specificity.
However, we have also found that there is a very low chance, in fact less than 1%, for those tested positive to actually have the disease.
Equivalently, it means that, there is a very high chance for those tested positive to not have the disease.
So if someone is tested positive for the disease, we have no confident that the test is correct.
This phenomenon always happens whenever a disease is rare, like what we have in our example.
In this case, we may need to repeat the test a few times. If the results are mostly positive, then we can make a diagnosis about the disease.

## To test or not to test


| To test | Not to test |
| --- | --- |
| No alternative test | Alternative more reliable test |
| Test is inexpensive and more expensive 2nd test | Test is expensive |
| Good chance of successfully treatment | Unreliable treatment |

So should you do screening tests for rare diseases in the first place?
Here are some points for consideration.
If there are basis, such as family history, that suggest for a screening of the disease, and there are no alternative tests available, then perhaps you should do it.
If the test is inexpensive and there's a second, but more expensive test that can weed out the false positives, it makes sense to start with the cheaper test first.
Of course, if the disease can be treated successfully if detected, perhaps the screening is a good idea.
On the other hand, if all the people who are tested positive must undergo expensive painful unreliable treatment, which would be unnecessary for 90% of them, then the screening is probably a bad investment of health care resources.



## Summary

- P(event happening | event suspected)
- Base rate, sensitivity and specificity
- Contingency table
- True positive and false positive

The framework that we used in the example of medical screening can be applied in a similar way to making decision in many other real life situations, such as detection of natural disasters, matching DNA for criminal evidence, blocking of spam mails and so on. They all involve asking about the likelihood of certain event happening given that the event being detected.
To answer the question, we need to know the base rate, sensitivity and specificity. These values are themselves conditional, or unconditional probabilities.
Using these information, we can set up the contingency table.
This table consists of figures including the true positive and false positive, which can then be used to compute the conditional probability that help us to answer the questions we are interested.