MIS775 Decision Modelling for Business Analytics

TOPIC 8: Simulation Modelling

DEAKIN BUSINESS SCHOOL

AACSB ACCREDITED

End of Unit Exam

End of Unit exam timetable is released.
Kindly check the MIS775 end of Unit exam, which falls during the exam period

From Monday, 2nd June – Friday 13th June

Special announcements –Assignment 2

- **Assignment 2 for MIS775 has been released** and is now available for you to begin working on.
- This assignment primarily relates to the content covered in **Weeks 7 to 9**. To make the most of the upcoming lectures.
- **Familiarize yourself with the assignment details**

📅 **Due Date: Wednesday, 21st May 2025 (Week 11) by 8:00 PM (Melbourne time)**

Submission Requirements:

Your submission must include the following:

1. **Excel File** – Use the provided template. You are welcome to modify it as needed to suit your analysis and requirements.
2. **Report** – Submit your report in **PowerPoint presentation format**, with a maximum of **30 slides.**

Assignment 2 Discussion Forum:

•The **Assignment 2 Discussion Forum** will be open from **Week 8, 1st of May Thursday**. Before posting your queries, please check existing threads to see if your question has already been addressed.

Online Student Consultation Sessions

Student consultation:

We will be holding a series of student consultation sessions to assist you with any clarifications related to your assignment. These sessions will continue **every Monday and Thursday from 7:00 PM to 7:30 PM until Monday, 19th May as given below.**

Date	Day	Time	Week
1st May	Thursday	7 pm -7:30 pm	8
5th May	Monday	7 pm -7:30 pm	9
8th May	Thursday	7 pm -7:30 pm	9
12th May	Monday	7 pm -7:30 pm	10
15th May	Thursday	7 pm -7:30 pm	10
19th May	Monday	7 pm -7:30 pm	11

📅 Consultation Schedule:
🕖 Time: 7:00 PM – 7:30 PM (Melbourne time)
📍 Days: Every Monday and Thursday from 1st May to 19th May
🔗 Zoom Meeting Link: provided in the Cloud Deakin Unit site

No prior appointment is required. Simply join the Zoom meeting via your unit site, as shown below:

Go to CloudDeakin MIS775 unit site > Table of Contents > Online Classroom and Recordings >Online Zoom Meetings > select the MIS775 Assessment 2 student consultation zoom meeting link as shown below.

AACSB ACCREDITED

DEAKIN BUSINESS SCHOOL

Recap

- **Stochastic decision models** have some random inputs, and therefore its outputs are also random
- We learnt how to structure a problem, and use a framework to build a spreadsheet model
- We also learnt how to use scenario analysis as a first step in developing solutions to these models, but we now know that replacing random variability with a fixed value in a model does not allow us to assess how likely a scenario might be
- **This week we take the next step by showing how to explicitly incorporate randomness in our modelling. This enables decision makers to consider all possible outcomes and their likelihood**

Modelling Uncertainty

- Most business outcomes are affected by random variables that introduce uncertainty and hence risk. Managing risk is an essential element in decision making
- We can use past data to determine appropriate probability distributions for our stochastic inputs, and then use these in our decision model to understand the distribution of the output(s), and hence the risks faced by the business in question
- But often that data is limited, and sometimes non-existent, in which case the modeller needs to have a good understanding of the assumptions underlying the distributions used for the stochastic inputs

Learning Objectives

- Understand the purpose of simulation
- Know the steps required in undertaking a simulation
- Have a thorough understanding of the features of the more commonly used theoretical probability distributions
- Be able to perform simulation in Excel

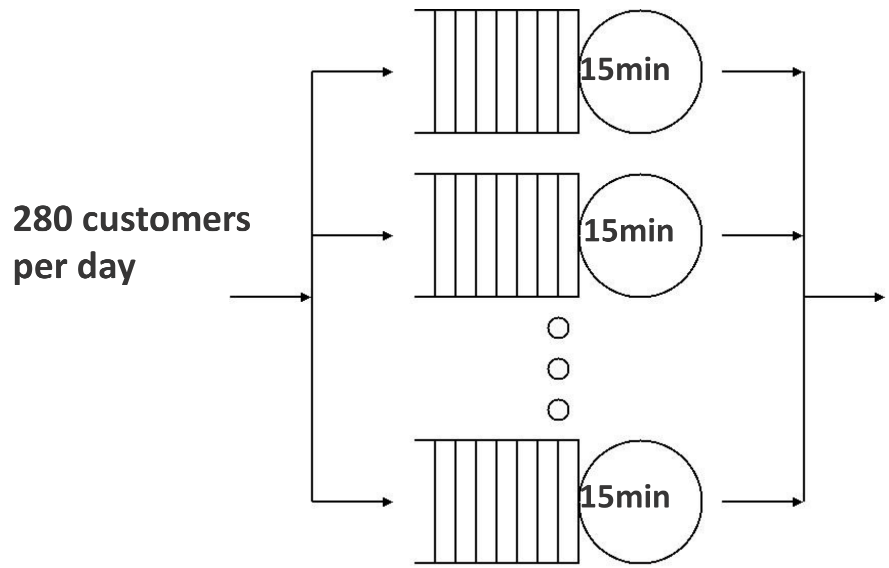
Textbook reading: Chapter 10 (10.2-10.3) ignoring references to @Risk software

A Simple Queueing Problem

- A shop has many customers
- It needs sales assistants to serve them, and they need to be paid
- Each sale takes a certain amount of time
- Each customer spends a certain amount of money, generating a certain amount of profit
- **How many sales assistants should be employed?**

A Simple Queueing Problem

- A shop has, on average, 280 customers per day
- Each sale takes, on average, 15 minutes. 4 customers are served per hour
- Each customer spends a certain amount of money, generating an average profit of $10
- Sales assistants are paid $200 per day for 7 hours work
- **How many sales assistants should be employed?**



wikipedia.org

A Simple Queueing Problem

- Each assistant can serve 28 customers per day (7 hrs x 4 customers/hr)
- To serve all the customers on an average day, 10 assistants are needed (280 Total customers / 28 customers per assistant)
- It costs $2000 each day (=10 assistants x $200 per assistant)
- So if every day is an average day …..
  - Total profit on sales is $10 x 280 = $2800
  - To serve all customers, we need 10 sales assistants and the net profit will be $800 ($2800-$2000)
- **But most days aren't average days!**

A Simple Queueing Problem

- Our simple queueing model, based on averages, gives us a hint of a solution, but it doesn't capture the underlying variability in the inputs
- Number of customers, service time, and profit per customer will all vary around their respective means
- So there will be days when there are not be enough assistants and sales will be lost, and days when there will be idle time for assistants
- And even if the number of customers is 280 and they take exactly 15 minutes to serve, and each generates $10 profit, the inter-arrival time of customers is unlikely to be uniformly spread throughout the day!

Simulation

- In reality, it may be better to have fewer assistants, able to serve fewer customers, than having some assistants idle on the not-so-busy days
- How can we find out?

By using simulation!

- By using probability distributions for stochastic inputs, we can generate data for thousands or tens of thousands of simulation trials in Excel
- We can then analyse the data distribution for each of the key outputs
- We can also experiment with different values for the controllable inputs (i.e. decision variables), and simulate what would happen

Simulation

- In our example, we could change the number of assistants in our simulation model to see what impact it has
- By experimenting in this way, we develop a better understanding of the number of assistants actually required

Simulation - Advantages

- Simulation can be used to model and learn about the behaviour of complex systems that cannot be described analytically
- **Simulation allows more sophisticated "what-if" questions than those illustrated in Topic 7, as it covers the full spectrum of possible outcomes <u>along with their likelihood</u>**
- It can also be used to measure and describe the outputs of a model when the values of multiple input variables are changed <u>at the same time</u>

Simulation - Disadvantages

- Simulation doesn’t generate optimal solutions
- The process of developing, verifying, and validating a simulation model, can be time-consuming

Simulation - Steps

STEP 1. Choose the **most appropriate distribution** for each stochastic input **(Topic 8)**

STEP 2. Simulate values: generate random numbers according to a given distribution **(Topic 8 Seminar)**

STEP 3. Replication: repeat the process for some large number of replications, or trials, using Data Table in Excel **(Topic 9)**

STEP 4. Analyse output data: descriptive statistics **(Topic 9)**

Step 1. Choosing a Distribution

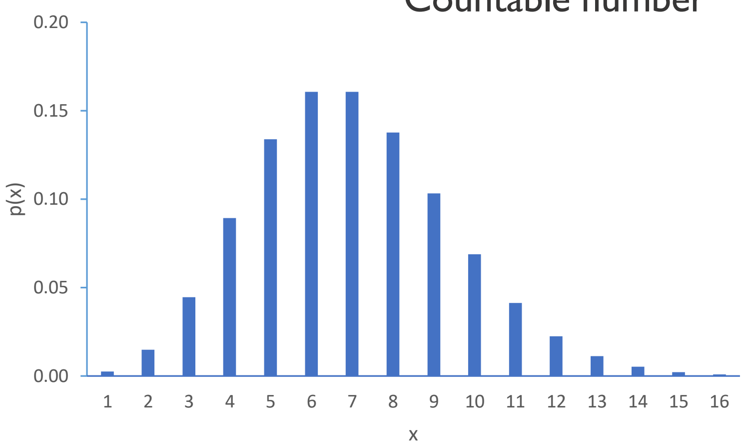
- If you have historical data for a **stochastic input or output variable**, understand your data first, by asking these questions:
  - Can the variable only take **discrete** or **continuous** values?
  - When you produce a bar chart or histogram of the data, is its **shape symmetric or skewed?**
  - What is its range? Does it have a definite **minimum and/or maximum, or is it unbounded?**
  - Do **extreme values** occur frequently or rarely?
  - Do the features suggest* that it could be modelled by one of the commonly used theoretical distributions? If so, what parameters need to be estimated from the data?
    Ex: Mean, Std. deviation….

There are statistical tests called *goodness of fit* tests which can be used here. E.g. Chi-squared "goodness-of-fit" test

Discrete vs Continuous Distributions

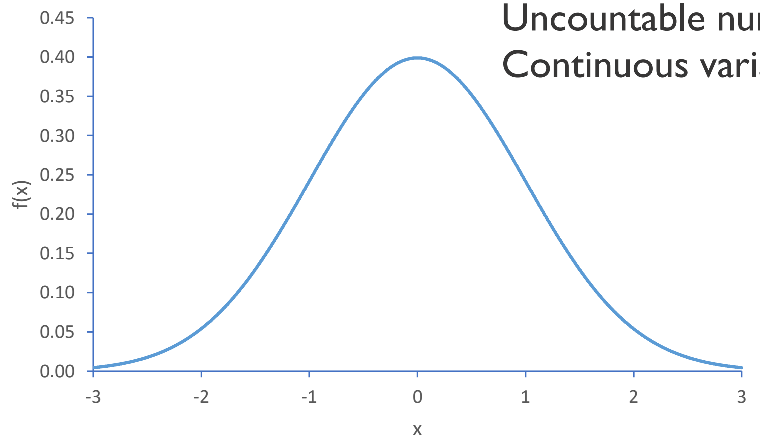
- Discrete random variable can take a countable number of possible values. E.g. 1, rolling a die can only take values from {1,2,3,4,5,6}; e.g. 2, demand for the latest model iPhone can take a wider range of possible values, but it is still countable
- Continuous random variable has an uncountable (infinite) number of possible values. E.g. interest rates, waiting time for an event to occur, etc. countable number

Countable number



Discrete distribution

Uncountable number
Continuous variable

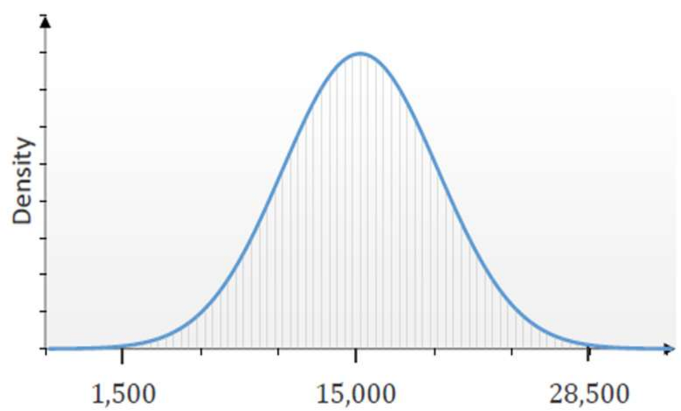


Continuous distribution

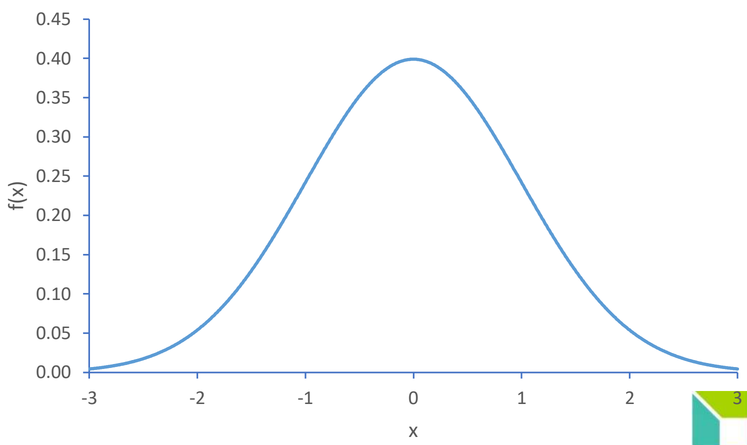
Discrete vs Continuous Distributions

- When a discrete random variable takes on a very large number of possible values, so many, that they become practically uncountable, then it can be approximated by a continuous distribution
- Ex: If demand values range from 1,000 to 40,000, the variable may be treated as continuous, since the large range allows us to assume an uncountable number of possible values.

Discrete variable with Uncountable number

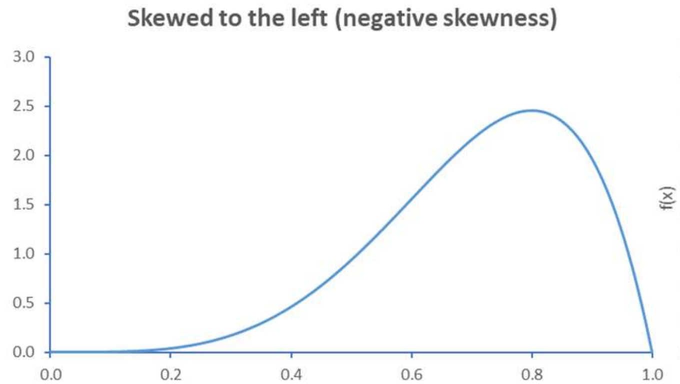


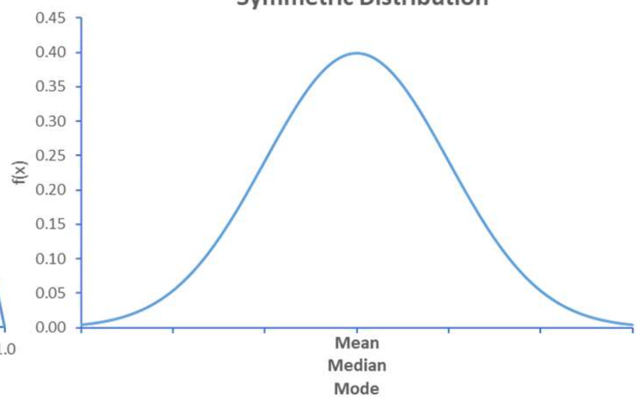
Uncountable number /Continuous variable

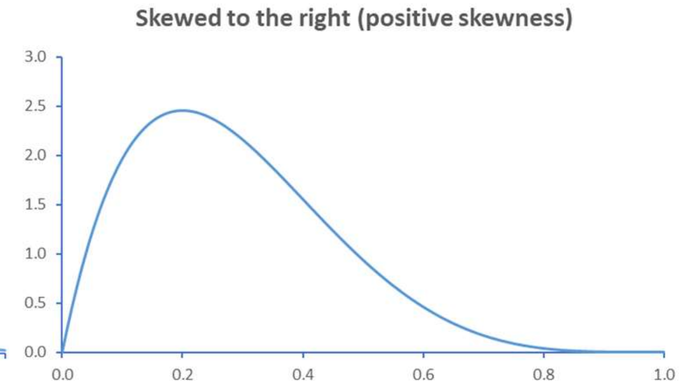


Symmetric vs Skewed

- A probability distribution can either be symmetric, or skewed to the left or right
- Choose between a symmetric and skewed distribution on the basis of which is considered more realistic







Bounded vs Unbounded

- A **probability distribution is bounded** if there are values A and B such that no possible value can be less than A or greater than B
  - A is the **minimum** possible value
  - B is the **maximum** possible value

- Otherwise the **probability distribution is unbounded**

- It is possible for a distribution to be bounded on one side but not the other. E.g. the number of calls received at a call centre is bounded below by zero, but there is no upper limit

Common Distributions

1. Discrete distributions

1.1) Geometric distribution

1.2) Binomial distribution

1.3) Poisson distribution

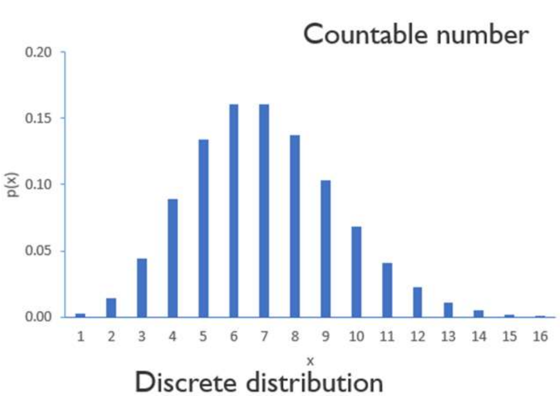
2. Continuous distributions

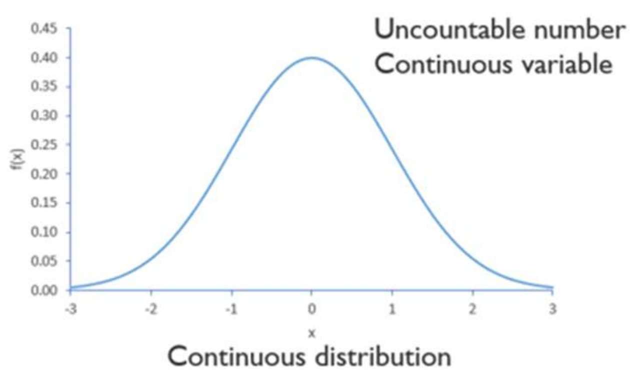
2.1) Uniform distribution

2.2) Normal distribution

2.3) Exponential distribution

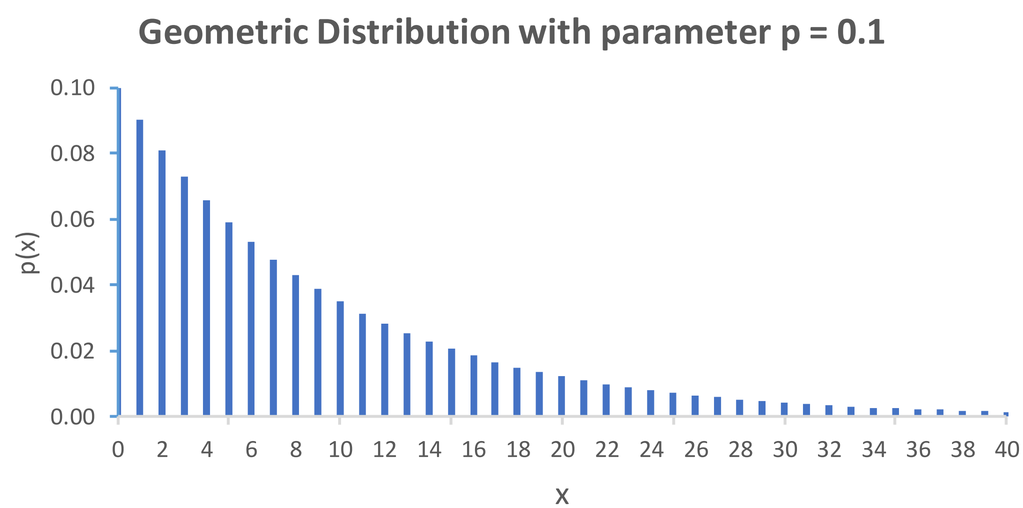
3. Empirical distribution





1.1 Geometric Distribution

- The **Geometric Distribution** models the number of **Bernoulli trials** needed to get the **first success**. Each trial is independent.
- The probability of success is constant in each trial.
- A Geometric random variable is used to model the number of trials *before* an event occurs for the first time
- It takes non-negative integer values
- Parameter, $p = 1 - q$ is the probability of the event occurring on a given trial
- Mean, $\mu = q/p$
- Variance, $\sigma^2 = q/p^2$



Example : Suppose a machine has a fixed probability of 10% of breaking down on any given day. The number of days before the machine breaks down for the first time is a Geometric random variable.

1.1 Geometric Distribution

Ex: Suppose a machine has a fixed probability of 10% chance (p=0.1) breaking down on any given day. The number of days before the machine breaks down for the first time is a Geometric random variable. What is the probability that the first time the machine breakdowns on the 4th day?

Let X be the number of trials until the first success.

$$P(X=k) = (1-p)^{k-1}p$$

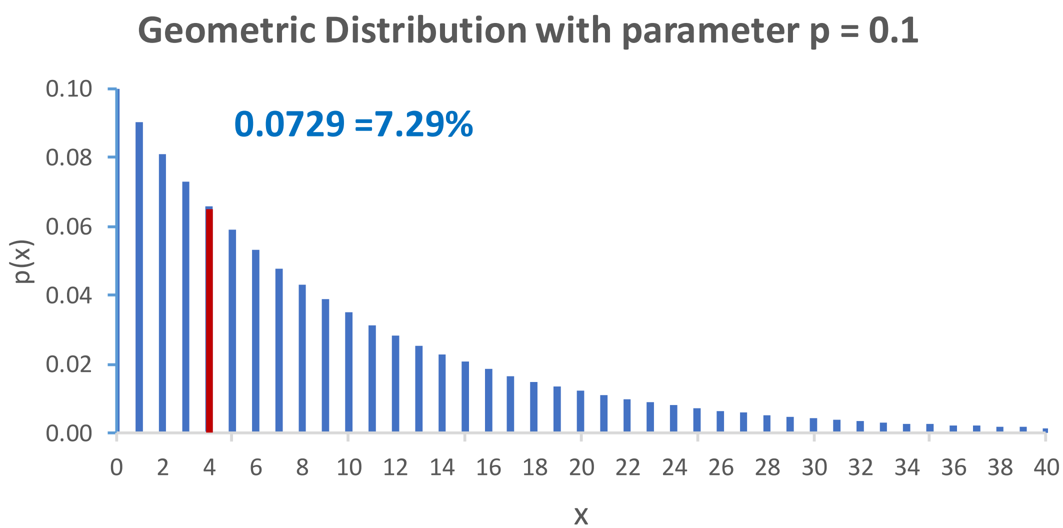
$$P(X=4) = (1-0.1)^{4-1}(0.1)$$

$$= (0.9)^3 \times 0.1$$

$$= 0.729 \times 0.1$$

$$= 0.0729$$

There's about a **7.29% chance** that the machine breakdowns on the 4th day.



1.2 Binomial Distribution

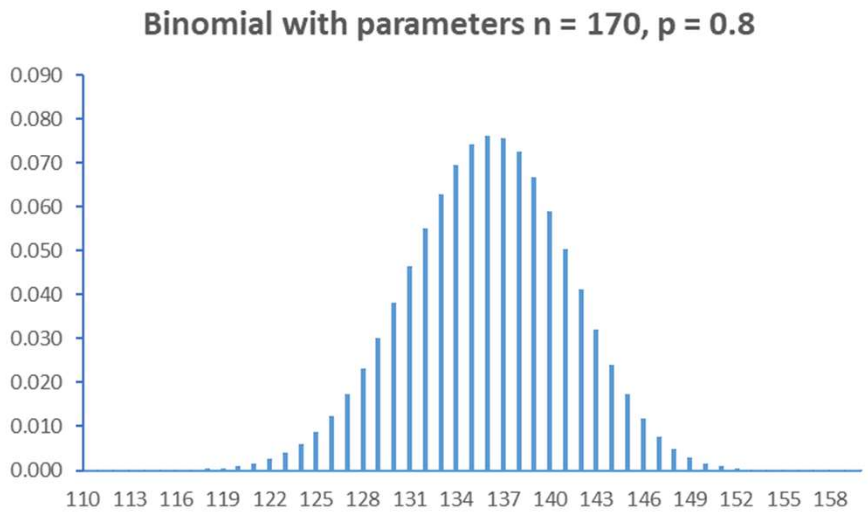
Binomial random variable is the **number of successes** in a sequence of independent trials

It has two parameters:

- Number of trials (n)
- Probability of success on each trial (p)

Mean: $\mu = np$

Variance: $\sigma^2 = np(1-p)$



- An experiment consists of a sequence of n identical trials
- Two outcomes, success and failure, are possible at each trial
- The probability of a success, p, does not change from trial to trial
- The trials are independent

1.2 Binomial Distribution

Suppose a manufacturer produces items with a **success rate of 80% (p = 0.8)** for passing quality control. If **170 items** are produced in a batch, what is the **probability that exactly 140 items** pass the quality check?

$$P(X = 140) = \binom{170}{140} \cdot (0.8)^{140} \cdot (0.2)^{30}$$

n=170 (number of items)

x=140 (number of successes)

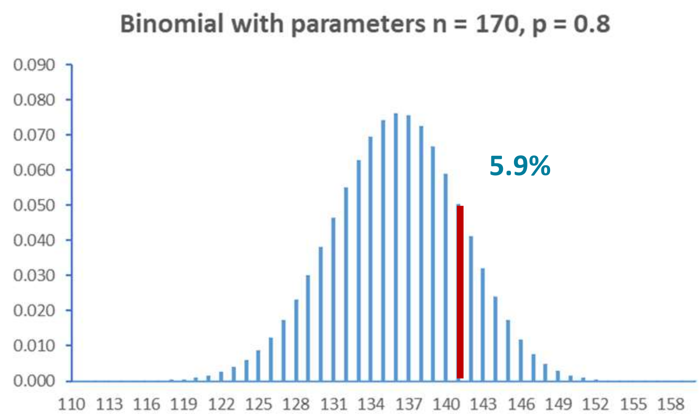
p=0.8 (probability of success)

= 0.059 or 5.9%

=BINOM.DIST(140,170,0.8,FALSE)

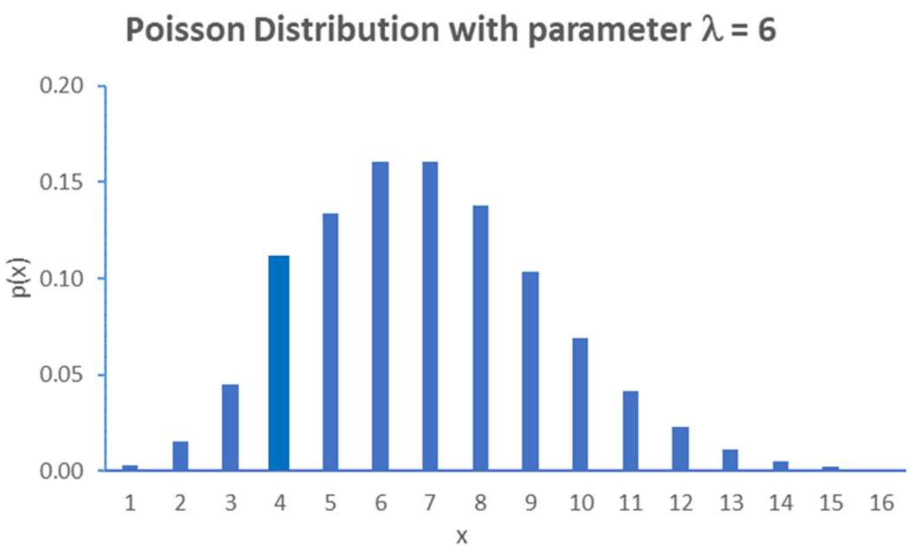
$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

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



1.3 Poisson Distribution

- A Poisson random variable is often used to model the number of occurrences of an event over a given interval of time or space
- It is a discrete random variable that takes non-negative integer values and its mean is the same as its variance (i.e. $\mu = \sigma^2$)
- Poisson distribution has one parameter λ (lambda) which is the mean or expected number of events per interval, i.e. $\mu = \lambda$



- Example : the number of vehicles arriving at a toll booth in one hour

1.3 Poisson Distribution

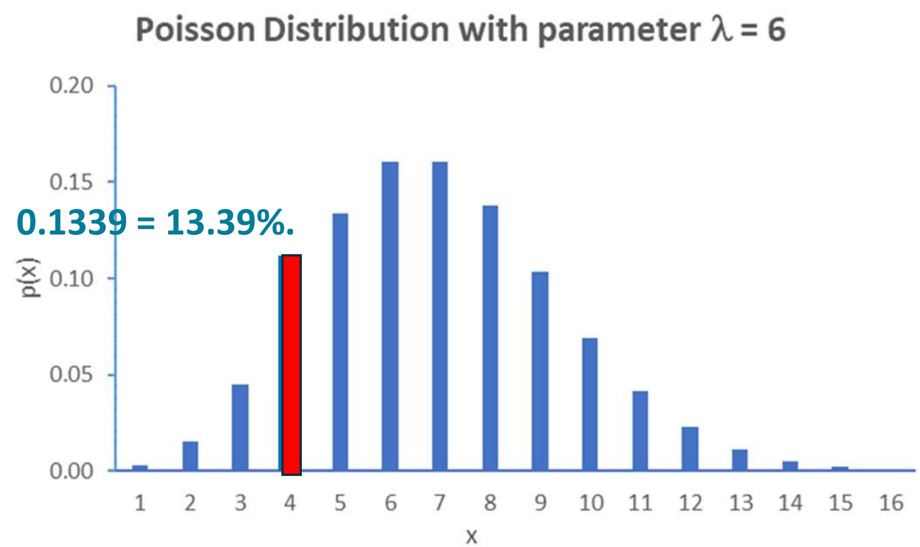
Ex: Average number of vehicles arriving at a toll booth in one hour is 6. We can use the Poisson distribution to find the probability of exactly 4 vehicles arriving in one hour.

where:

- λ is the mean number of arrivals (6 in this case),
- k is the number of arrivals we are interested in (4 here),
- e is the base of the natural logarithm (approximately 2.71828).

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$P(X = 4) = \frac{e^{-6} \cdot 6^4}{4!}$$



=POISSON.DIST(4,6,FALSE)
=0.1339

After calculating, the probability of exactly 4 vehicles arriving at the toll booth in one hour is approximately 0.1339 = 13.39%.

AACSB ACCREDITED

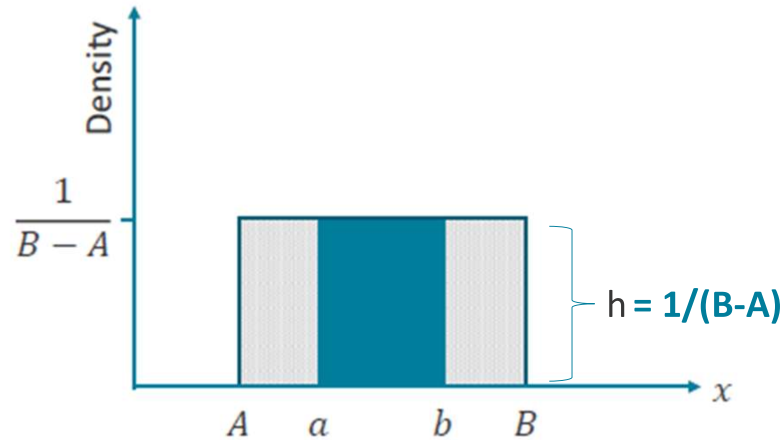
DEAKIN BUSINESS SCHOOL

2.1 Uniform Distribution

1=h*(B-A) or (max-min)
1=h* (B-A)

h = 1/(B-A) or (1/(max-min)

- Uniform distribution on the interval [A, B] is flat and bounded at both ends
- The probability that a Uniform variable will assume a value in a given interval is (a,b) is (b-a)/(B-A)
- Think of it as the "I have no idea" distribution
- Mean and variance: $\mu = \frac{A+B}{2}$
- $\sigma^2 = \frac{(B-A)^2}{12}$



- To generate a random number from a distribution that is Uniform on the 0-1 interval, enter "=RAND()" into a cell. Each time you press the F9 key, Excel will generate another random number from this distribution

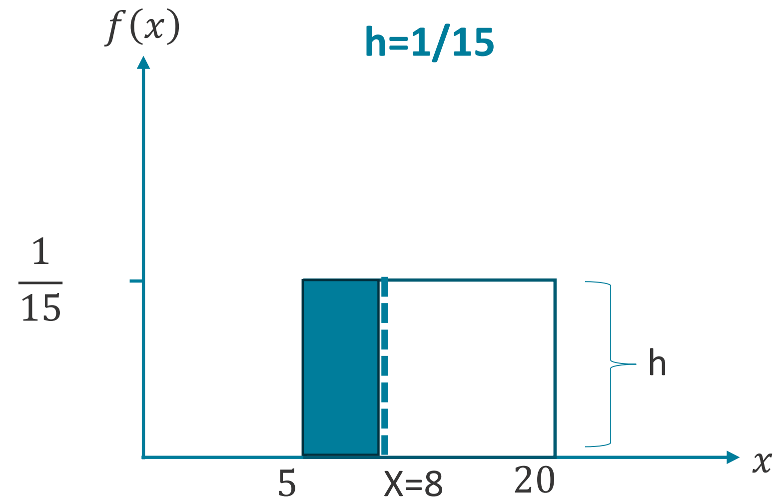
Uniform Distribution

1=h*(20-5)

1=h*15

h=1/15

Ex: A bus **arrives between every 5 and 20 minutes**, and the **arrival time is uniformly distributed between 5 and 20 minutes**. You arrive at the stop at a random time. What is the probability that you wait **less than 8 minutes** for the bus.



Probability = (8-5) * 1/15

= 3/15

= 1/5

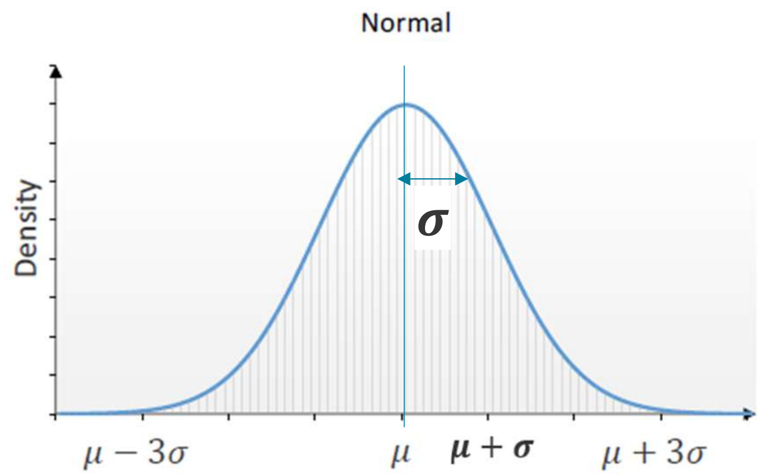
= 0.2 Or 20%

2.2 Normal Distribution

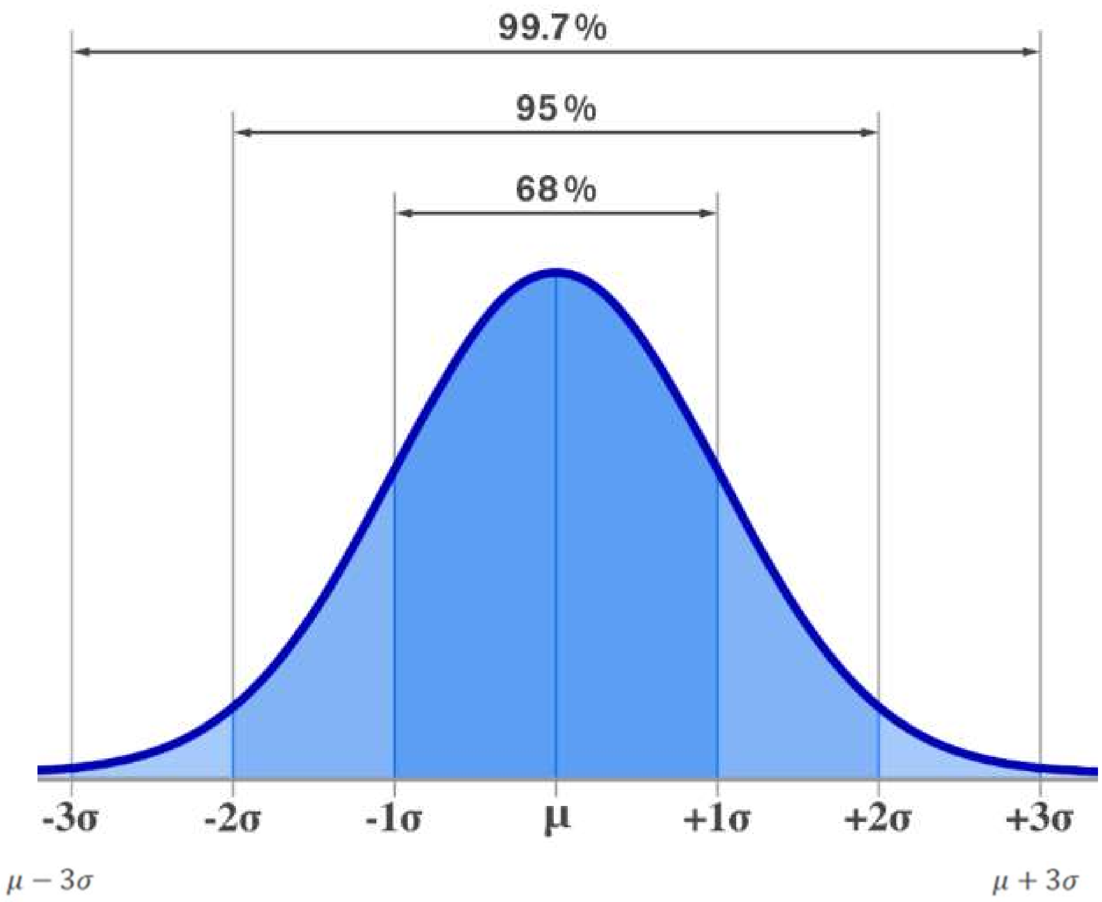
- A Normal random variable with mean, μ and standard deviation, σ has the familiar bell-shaped distribution

- This distribution should only be considered when a random variable is believed to have a symmetric distribution

- A limitation is that it can take negative values, which is usually inappropriate. However if $\mu > 3\sigma$ almost all of the probability lies on the positive part of the x axis, so negative values are unlikely to occur



2.2 Normal Distribution



100% -99.7%
=0.3%
Or 0.003

2.2 Normal Distribution

Ex: Based on laboratory testing, the lifetime of a StarTyre is taken to be normally distributed, with mean 65,107 kilometres (km) and standard deviation of 2,582 km. The tyres carry a customer warranty for 60,000 km.

What proportion of the tyres is expected to fail before the warranty expires?

Let X = lifetime of a tyre (km)

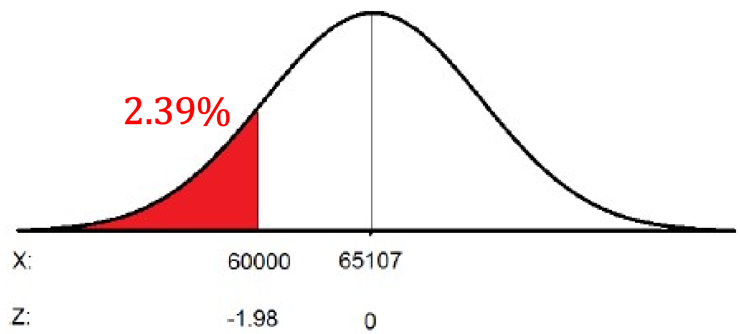
$P(X < 60000)$

$= P(\frac{X-\mu}{\sigma} < \frac{60000-65107}{2582})$

= $NORM.DIST(X, \mu, \alpha, TRUE)$

$= NORM.DIST(60000, 65107, 2582, TRUE)$

$= 0.0239 = 2.39\%$

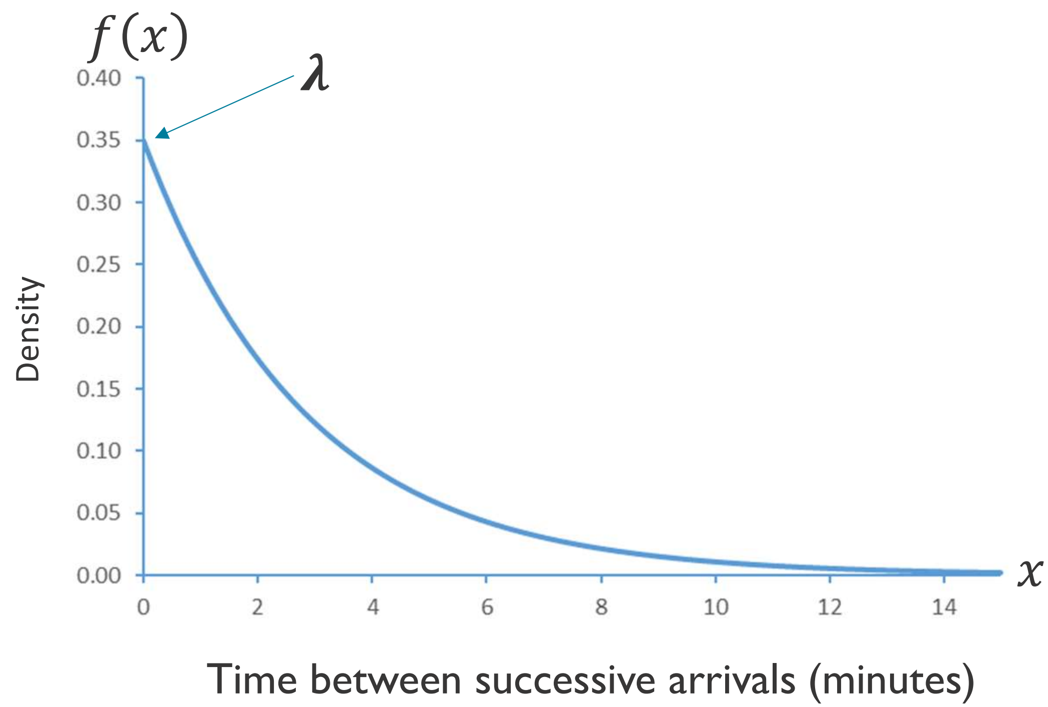


$= NORM.DIST(60000, 65107, 2582, TRUE)$

2.3 Exponential Distribution
(Negative exponential distribution)

- Exponential distribution* can be useful for describing the time it takes for an event to occur
  - **E.g.1 Time between vehicle arrivals at a toll booth**
  - **E.g. 2 Time required to complete a questionnaire**

- It has a single parameter, λ, the expected number of events per interval



2.3 Exponential Distribution
(Negative exponential distribution)

Ex: Vehicles arrive at a toll booth at an average rate of 6 per hour, what is the probability that the time between two successive vehicle arrivals is less than 12 minutes (12/60 = 0.2 hrs)?

The **Exponential distribution** formula for the cumulative probability is:

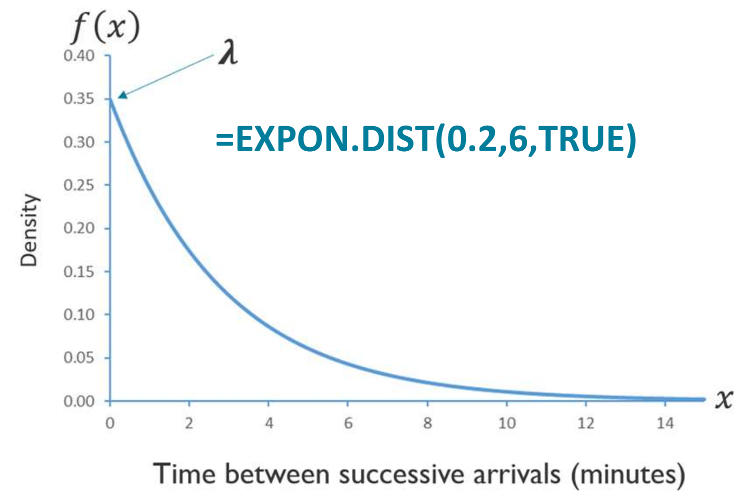
$$P(T < t) = 1 - e^{-\lambda t}$$

where:

- $\lambda = 6,$
- $t = 0.2.$

t =12/60=0.2 hrs

$$P(T < 0.2) = 1 - e^{-6\times 0.2}$$
$$= 1 - e^{-1.2}$$
$$\approx 1 - 0.3012$$
$$= 0.6988$$



Approximately a **69.88% chance** that the time between two successive vehicle arrivals is less than 12 minutes.

Correspondence Between the Exponential & Poisson Distributions

The **Poisson distribution** describes the **number** of occurrences per interval

↓ ↑

The **Exponential distribution** describes the **length of the interval** between occurrences

Average **number of vehicles arriving at a toll booth in one hour is 6**. We can use the Poisson distribution to find the probability of exactly 4 vehicles arriving in one hour

Vehicles arrive at a toll booth at an average rate of 6 per hour, what is the probability that the **time between two successive vehicle arrivals is less than 12 minutes**

3. Creating your own Distribution
(*Empirical Probability Distribution)*

An **empirical distribution** is a **statistical distribution that is based on observed data** rather than a known theoretical model like the normal or binomial distribution or etc...
It shows **how frequently different values occur in your data set**. It's built from actual sample data, not from a known probability distribution.

Key Points:

- **Constructed from observed data**: You take a sample (of 50 days), record the outcomes, and use that to determine probabilities.
- **Non-parametric**: It doesn't assume any specific underlying distribution shape (like normal or exponential).

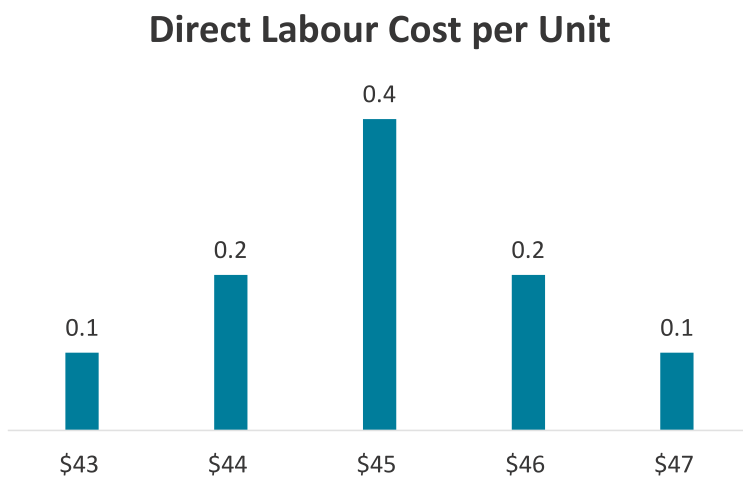
Direct labour cost per unit	frequency (days)	Probability	Sub-intervals*
$43	5	5/50 = 0.1	0.0 - 0.1
$44	10	10/50 = 0.2	0.1 - 0.3
$45	20	20/50 = 0.4	0.3 - 0.7
$46	10	10/50 = 0.2	0.7 - 0.9
$47	5	5/50 = 0.1	0.9 -1.0
	50	1	

AACSB ACCREDITED
DEAKIN BUSINESS SCHOOL

Creating your own Distribution
(Empirical Probability Distribution)

- You can create your own distribution if none of the common ones are an adequate fit. For example, consider *Direct labour cost* in our case study. Sanotronics believes that it can be best modelled by this discrete distribution:

Direct labour cost per unit	frequency (days)	Probability	Sub-intervals*
$43	5	0.1	0.0-0.1
$44	10	0.2	0.1-0.3
$45	20	0.4	0.3-0.7
$46	10	0.2	0.7-0.9
$47	5	0.1	0.9-1.0
	50	**1**	



Creating your own Distribution (*Empirical Probability Distribution)*

- To generate a random value from this distribution, first partition the interval [0, 1] into sub-intervals of widths equal to the corresponding probabilities
- Next use Excel RAND() to generate a random number from [0, 1], e.g. 0.475
- Then find which sub-interval contains this random value. In our example, this value belongs to the sub-interval [0.3, 0.7], which corresponds to a cost of $45. This is then our random value.

Direct labour cost per unit	frequency (days)	Probability	Sub-intervals*
$43	5	0.1	0.0-0.1
$44	10	0.2	0.1-0.3
$45	20	0.4	0.3-0.7
$46	10	0.2	0.7-0.9
$47	5	0.1	0.9-1.0
	50	**1**	

* We can ignore the fact that the sub-intervals overlap at the endpoints because the probability of randomly generating an endpoint is effectively zero

=VLOOKUP(rand(),D11:F15,3)

What if Data is Scarce or Non-Existent?

- Consider *cost of parts* in our case study. Sanotronics believes it will lie somewhere between \$80 and \$100 per unit, but other than that "they have no idea"

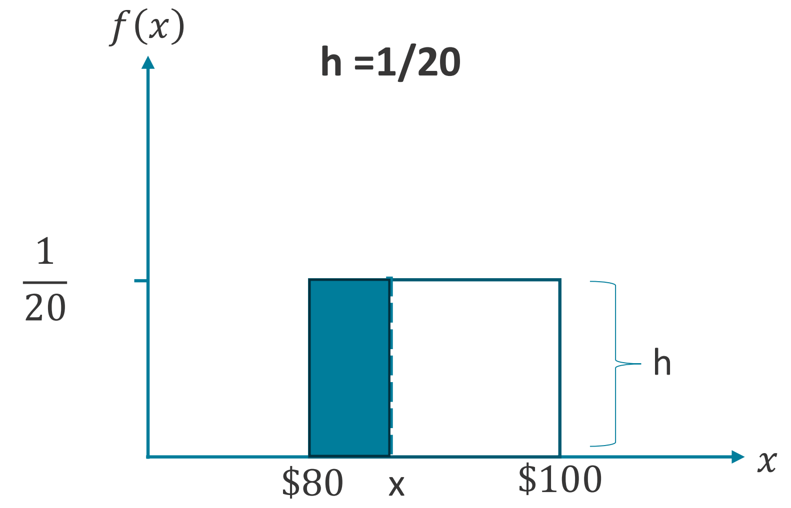
- This is an obvious candidate for a **Uniform distribution** on the interval [\$80, \$100]

- To generate a random value from this distribution, calculate $A + (B - A)$×RAND()

- where $A = \$80$ and $B = \$100$

1 =h*(max - min)
1 =h*20

h =1/20



X = min+ (max-min)*rand()

Probability = (X-80)*1/20
Rand () = (X-min)/(max-min)
X = min + (max - min) * rand()

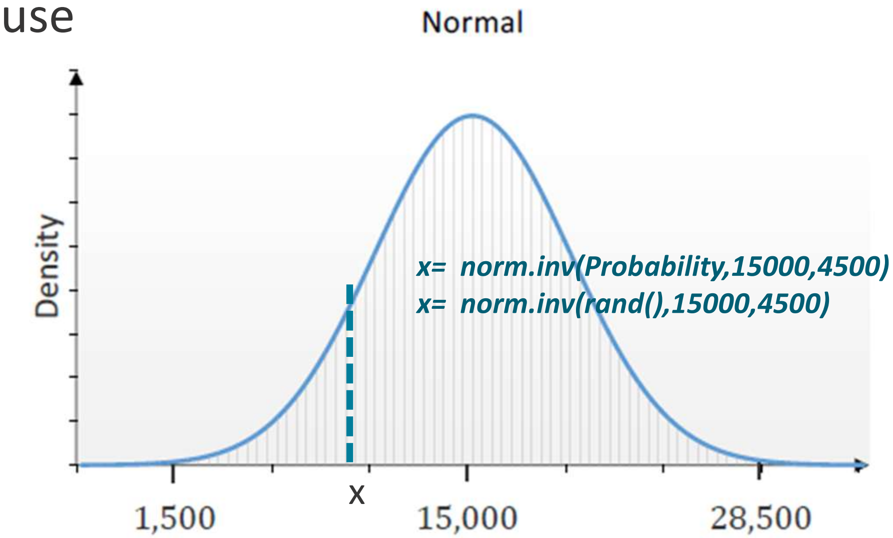
AACSB ACCREDITED
DEAKIN BUSINESS SCHOOL

What if Data is Scarce or Non-Existent?

Consider *demand* in our case study. Sanotronics believes that, at a selling price of \$249, demand will be Normally distributed with $\mu = 15,000$ and $\sigma = 4,500$.

But if price is increased to \$299, then $\mu = 11,000$ and $\sigma = 4,500$

- To generate a random value from this distribution, use the built-in function NORM.INV
- E.g., when price is \$249 demand is given by **NORM.INV(RAND(), 15000, 4500)**
- **To ensure this is non-negative, use MAX(0,NORM.INV(RAND(),15000,4500))**
- **Example when Demand is X then**
- **x= max(0, norm.inv(rand(),15000,4500)**



*15,000 - 3*4500 =1500*

*15,000+3*4500 =28500*

Summary

- Our main focus has been on learning how to determine appropriate distributions for stochastic inputs in a decision model

- We covered the key features of some common theoretical distributions

- We also discussed how to deal with the situation of limited or no data on which to base a distribution

Next Class

Topic 9: Risk Analysis

- **Simulations using Data table in Excel**
- **Case study completed**