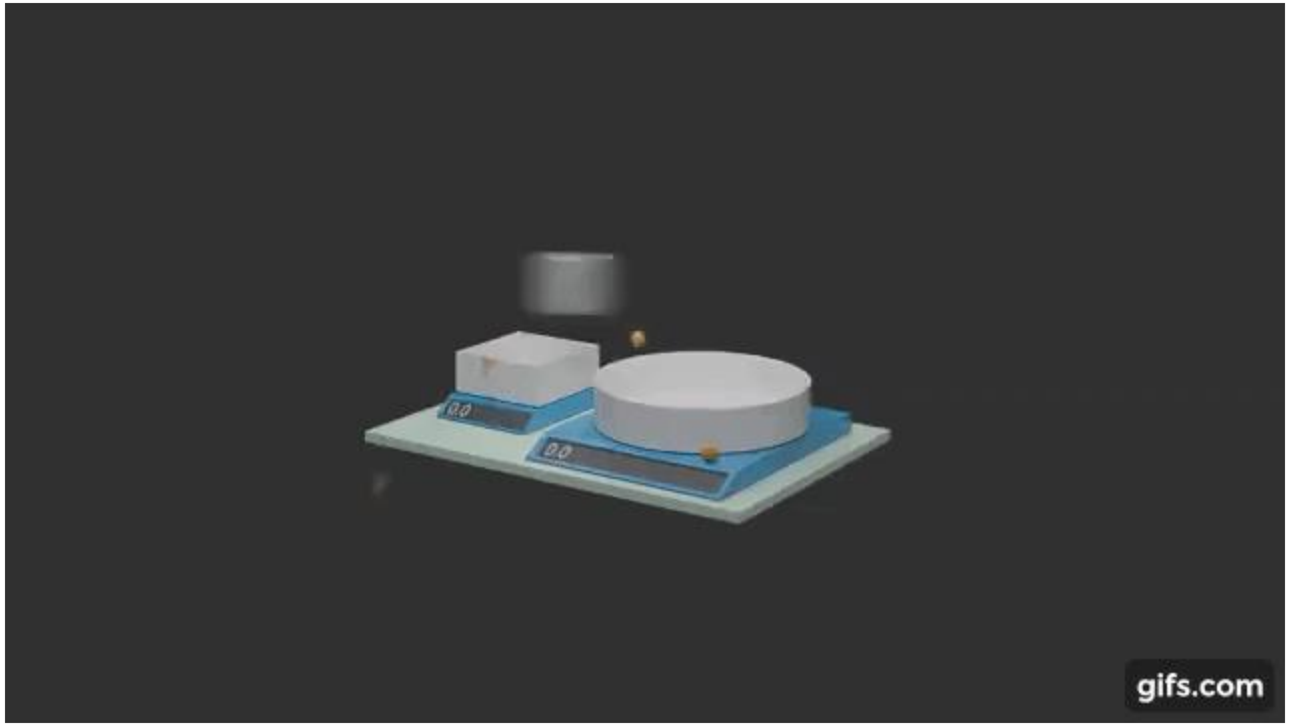
Monte Carlo Simulation

Reference book: Hamdy A. Taha, Operations Research, An Introduction (10th Edition)
Sec. 19.1

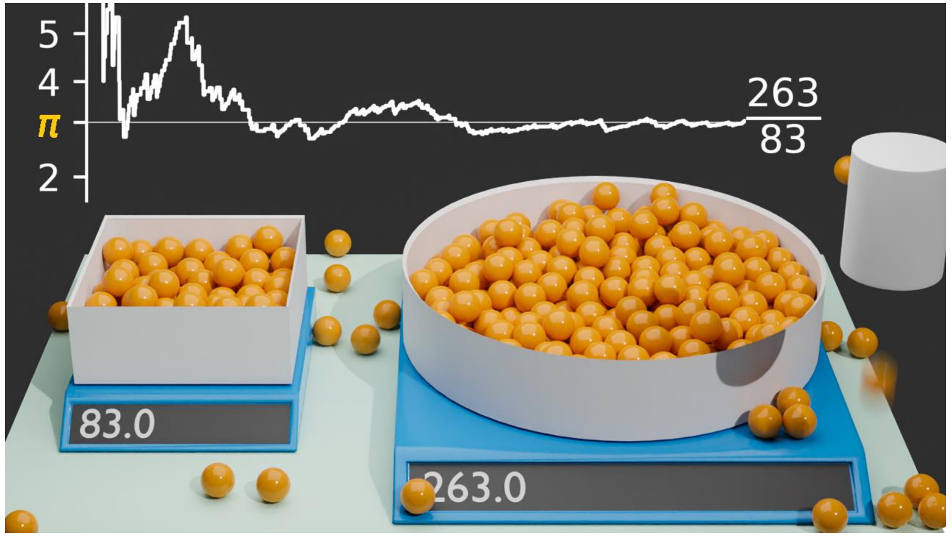
Casino at Monte Carlo

- The name ”Monte Carlo” refers to the city in Monaco, known for its casinos and gambling.
- ”Monte Carlo” is basically used as a synonym for randomness,
- Monte Carlo simulations are simulations evolving in a deliberately random way.

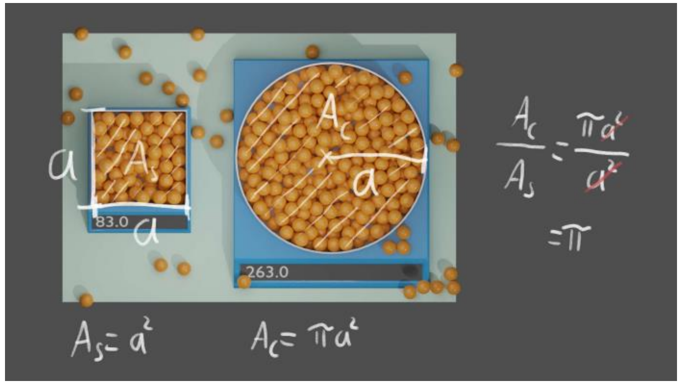


Example

- If we let this simulation evolve for a while and divide the number of marbles in the circular bowl by the number of marbles in the square bowl, the result happens to be roughly pi. Without any advanced math knowledge simply by randomly dropping marbles into two bowls we can estimate pi, a task which is not at all trivial.



Example



- So let's unpack what happened here.
- When we drop a marble in a uniformly random location, the probability for it to end up in one of the bowls is proportional to the bowl's cross-section area.
- When we repeat this random process over and over again, the number of marbles in a bowl will approach a value which is proportional to its cross-section area.
- The area of the square bowl in this example is a^2, and the area of the circular bowl is pi*a^2. That's how we get pi as the fraction of the two areas and consequently as the fraction of marbles in each bowl.

Example

- The idea of obtaining random samples is probably quite familiar to you if you think about how real world studies outside of simulations are done.
- Let's say we wanted to make a study to determine the average body height of all people worldwide.
- In principle, we would need to know the height of each person worldwide and take the average of all these heights.
- Obviously, we can't go out and measure the height of billions of people.
- So what do we do?

Example

- We can measure the heights of a smaller group of people and hope that their average height is a good estimate of the average height of all people worldwide. There are two things that we need to consider to get a good estimate:
- Firstly, the selection of the group needs to be unbiased.
- We can't just measure the height of the next 10 people we meet or the height of all our family members, as we might live in an area with especially tall or short people.
- Also, our family might not be a good representative for the world population.
- Instead, a good way to get an unbiased sample group would be to randomly select people worldwide.

Example

- Secondly, the group of measured people must not be too small.
- If we only measured the height of 5 people, we might have coincidentally picked 5 taller people.
- We can be more and more confident about the resulting average the more people we measure.
- **In probability theory this is often referred to as the law of large numbers.**

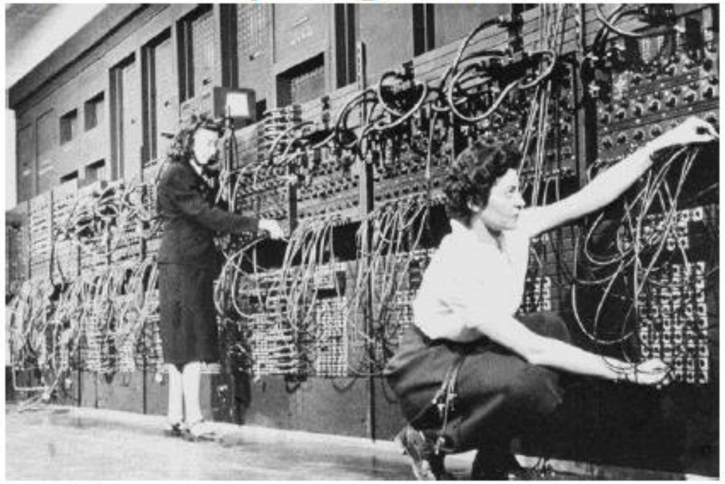
Idea

- The exact same is also true for Monte Carlo simulations. The core idea of a Monte Carlo simulation is that we can obtain a representative group of samples of some large population of possibilities if we allow the simulation to evolve randomly.
- In the marble dropping example, in principle, we would need to test for every possible location whether the marble ends up inside or outside the bowl to determine its cross-section area precisely. Just like we, in principle, would need to measure the height of each person to determine the average height accurately. Instead, we can rely on randomly selected samples, and according to the law of large numbers, we can be more and more confident about the result the more samples we take.

- The concept was invented by the Polish American mathematician, Stanislaw Ulam. Probably more well known for his work on thermonuclear weapons than on mathematics, but he did do a lot of very important mathematics earlier in his life.

A Little History

- Ulam, recovering from an illness, was playing a lot of solitaire
- Tried to figure out probability of winning, and failed
- Thought about playing lots of hands and counting number of wins, but decided it would take years
- Asked Von Neumann if he could build a program to simulate many hands on ENIAC



Monte Carlo Simulation

- A method of estimating the value of an unknown quantity using the principles of inferential statistics
- Inferential statistics
  - *Population*: a set of examples
  - *Sample*: a proper subset of a population
  - Key fact: a *random sample* tends to exhibit the same properties as the population from which it is drawn
- Exactly what we did with random walks

An Example

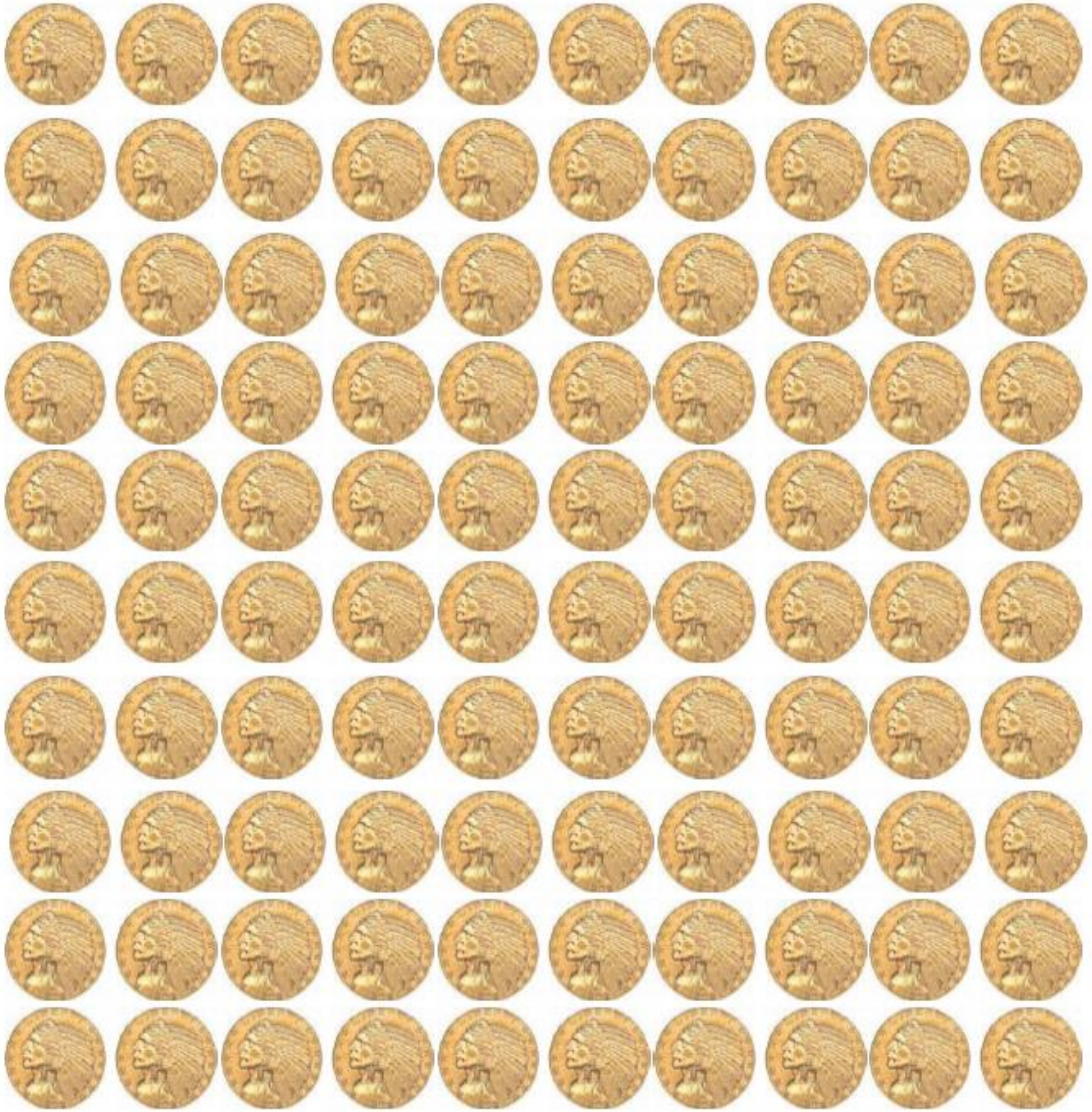
- Given a single coin, estimate fraction of heads you would get if you flipped the coin an infinite number of times
- Consider one flip

How confident would you be about answering 1.0?

Flipping a Coin Twice

Do you think that the next flip will come up heads?

Flipping a Coin 100 Times



Now do you think that the next flip will come up heads?

Flipping a Coin 100 Times

Do you think that the probability of the next flip coming up heads is 52/100?

Given the data, it's your best estimate

But confidence should be low

Why the Difference in Confidence?

- Confidence in our estimate depends upon two things
- Size of sample (e.g., 100 versus 2)
- Variance of sample (e.g., all heads versus 52 heads)
- As the variance grows, we need larger samples to have the same degree of confidence

Methods of Solving Operations Research Problems:

- **Analytical method (Classical Method):** Mathematical techniques such as differential calculus, probability theory etc. to find the solution of a given operations research model.
- **Iterative Method (Numerical Method):** This is trial and error method. First, we set a trial solution and then go on changing the solution under a given set of conditions, until no more modification is possible.
- **The Monte-Carlo Technique (a simulation process):** Based on random sampling of variable's values from a distribution of the variable. A table of random numbers must be available to solve the problems.

Simulation: Some definitions

- a representation of reality through the use of model or other device, which will react in the same manner as reality under a given set of conditions.
- the use of system model that has the designed characteristic of reality in order to produce the essence of actual operation.
- model which depicts the working of a large scale system of men, machines, materials and information operating over a period of time in a simulated environment of the actual real world conditions.
- a numerical technique for conducting experiments on a digital computer, which involves certain types of mathematical and logical relationships necessary to describe the behaviour and structure of a complex real world system over extended period of time.

Overview of Simulation

When do we prefer to develop **simulation model** over an analytic model?

- When not all the underlying assumptions set for analytic model are valid.
- When mathematical complexity makes it hard to provide useful results.
- When “good” solutions (not necessarily optimal) are satisfactory.

❖ A simulation develops a model to numerically evaluate a system over some time period.

❖ By estimating characteristics of the system, *the best alternative* from *a set of alternatives under consideration* can be selected.

Overview of Simulation

- *Continuous simulation systems* monitor the system each time a change in its state takes place.
- *Discrete simulation systems* monitor changes in a state of a system at discrete points in time.
- Simulation of most practical problems requires the use of a computer program.

➢ *Modeling and programming skills, as well as knowledge of statistics are required when implementing the simulation approach.*

Types of simulation

1. **Analog Simulation:** Simulating the reality in physical form (e.g.: Children's park, planetarium, etc.)
2. **Computer Simulation:** For problems of complex managerial decision-making, the analogue simulation may not be applicable. In such situation, the complex system is formulated into a mathematical model for which a computer programme is developed. Using high-speed computers then solves the problem.

- Often, the physical processes we are interested (the **system**) are complex

- To study a certain process, we require may require the use of a **model** (simplified representation of the system).

- When we apply the model to mimic the system's behavior, we say we run **simulations**

  - Models are fundamental tools of science, engineering, business, etc.

  - Abstraction of reality therefore always has limits of credibility

Types of simulation Models

Two distinct types of simulation models exist.

1. Continuous models deal with systems whose behavior changes continuously with time. These models usually use difference-differential equations to describe the interactions among the different elements of the system. A typical example deals with the study of world population dynamics.

Types of simulation Models

2. Discrete models deal primarily with the study of waiting lines, with the objective of determining such measures as the average waiting time and length of the queue. These measures change only when a customer enters or leaves the system. The instants at which changes take place occur at specific discrete points in time (arrivals and departure events), giving rise to the name discrete event simulation.

CLASSIFICATION OF SIMULATION MODELS

- **Simulation of Deterministic models:** the input and output variables are not permitted to be random variables and models are described by exact functional relationship.
- **Simulation of Probabilistic models:** method of random sampling is used. The techniques used for solving these models are termed as **Monte-Carlo technique**.
- **Simulation of Static Models:** do not take variable time into consideration.
- **Simulation of Dynamic Models:** deal with time varying interaction.

Some Advantages of Simulation

- Often **simulation** is the **only type of model possible** for complex systems (e.g assess the impact of a certain

- , fire).

- Process of building simulators can **clarify the understanding** of real systems and sometimes can be more useful than the implementation of the results itself.

- Allows for sensitivity analysis and optimization of real system **without need to operate real system** (e.g no need to burn all stands to infer about fire behavior or its economic impacts).

- Can maintain **better control over experimental conditions** than real system.

Some Disadvantages of Simulation

- May be very **expensive and time consuming** to build simulation
- **Easy to misuse simulation** by “stretching” it beyond the limits of credibility

  - when using commercial simulation packages due to ease of use and lack of familiarity with underlying assumptions and restrictions

  - Slick graphics, animation, tables, etc. may tempt user to assign unwarranted credibility to output

- Monte Carlo simulation usually **requires several (perhaps many) runs** at given input values, whereas analytical solutions provide exact values

MONTE-CARLO SIMULATION

Monte Carlo simulation: method of estimating the value of an unknown quantity using the principles of inferential statistics.

Population: set of examples
Sample: a proper subset of the population
Key fact: a ***random*** sample tends to exhibit the same properties as the population from which it is drawn

- ***Monte Carlo simulation*** *is a computerized mathematical technique that allows to* ***model the probability of different outcomes*** *in a process that* ***cannot be easily predicted*** *due to the intervention of* ***random variables***

- ***Monte Carlo simulation*** *depends on a sequence of* ***random numbers*** *which is generated during the simulation*

MONTE-CARLO SIMULATION

- a simulation technique in which statistical distribution functions are created using a series of random numbers.
- working on the digital computer for a few minutes we can create data for months or years.
- generally used to solve problems which cannot be adequately represented by mathematical models or where solution of the model is not possible by analytical method.
- yields a solution, which should be very close to the optimal, but not necessarily the exact solution.
- yields a solution, which converges to the optimal solution as the number of simulated trials tends to infinity.

Summary of Monte-Carlo simulation :

Step 1: Clearly define the problem:

a) Identify the objectives of the problem.
b) Identify the main factors, which have the greatest effect on the objective of the problem.

Step 2: Construct an approximate model:

a) Specify the variables and parameters of the mode.
b) Formulate the appropriate decision rules, i.e. state the conditions under which the experiment is to be performed.
c) Identify the type of distribution that will be used. Models use either theoretical distributions or empirical distributions to state the patterns of the occurrence associated with the variables.
d) Specify the manner in which time will change.

Problem:

With the help of a single server queuing model having inter-arrival and service times constantly 1.4 minutes and 3 minutes respectively, explain discrete simulation technique taking 10 minutes as the simulation period. Find from this average waiting time and percentage of idle time of the facility of a customer. Assume that initially the system is empty and the first customer arrives at time t = 0.

SOLUTION:

Data: System is initially empty. Service starts as soon as first customer arrives. First customer arrives at t = 0

Average waiting time per customer for those who must wait = Sum of waiting time of all customers/number of waiting times taken= (1.4 + 2.8 + 4.2 + 5.6 + 7.0 + 8.4 + 9.8) / 7 = 18.8 / 7 = 2.7 minutes

Percentage of idle time of server = Sum of idle time of server / total time = 0%

Time	*Event Arr = arrival Dep = departure*	*Customer Number*	*Waiting time.*
0.0	Arr.	1	--
1.4	Arr.	2	-
2.8	Arr.	3	--
3.0	Dep	1	3.00 – 1.40 = 1.6 min. for customer 2.
4.2	Arr.	4	
5.6	Arr.	5	
6.00	Dep	2	6.00 – 2.80 = 3.2 min. for customer 3.
7.00	Arr.	6	
8.4	Arr.	7	
9.00	Dep.	3	9.00 – 4.20 = 4.8 min. for customer 4
9.80	Arr.	8	
10.00	End of given time	-	10.00 – 5.60 = 4.4 min. for customer 5
			10.00 – 7.00 = 3.0 min. for customer 6
			10.00 – 8.4 = 1.6 min. for customer 7
			10.00 – 9.80 = 0.2 min. for customer 8.

Monte Carlo simulation: technique that combines ***distributions*** with ***random number generation***

Random numbers can be generated in different ways

Any variable has a probability distribution for its occurrence

Best way to relate random number to a variable is to use **cumulative probability distribution**

(probability density functions – pdf)

The daily demand for clone packs (80 seedlings) during Spring months was studied and the probabilities are the following:

Relative frequencies (probability)

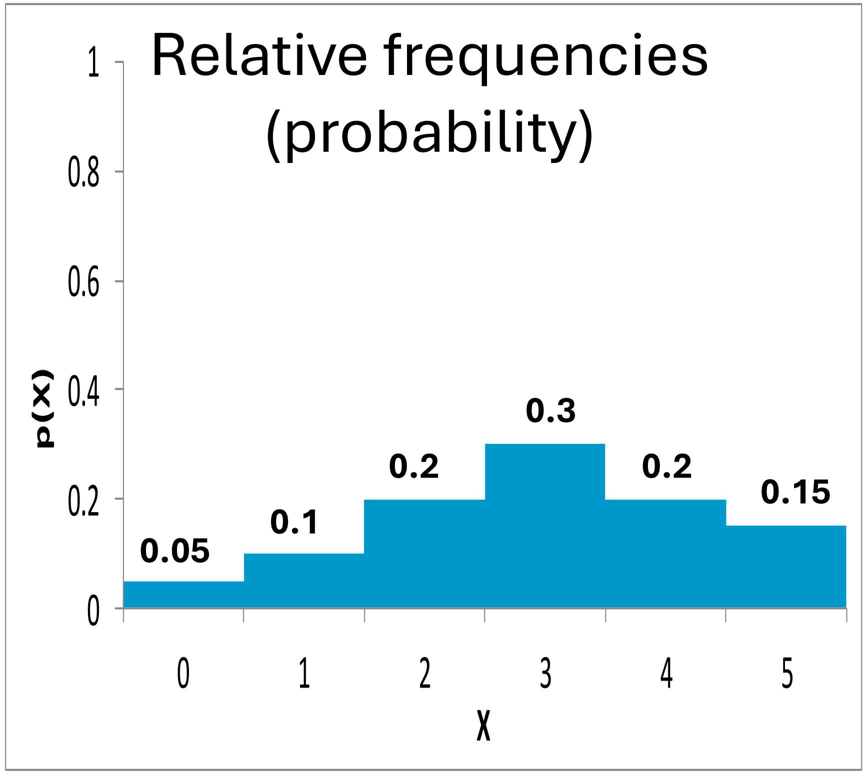
If the distribution is known , **WHY** do we use **random numbers** to simulate it?

Nr packs ordered	probability
0	0.05
1	0.1
2	0.15
3	0.3
4	0.25
5	0.15

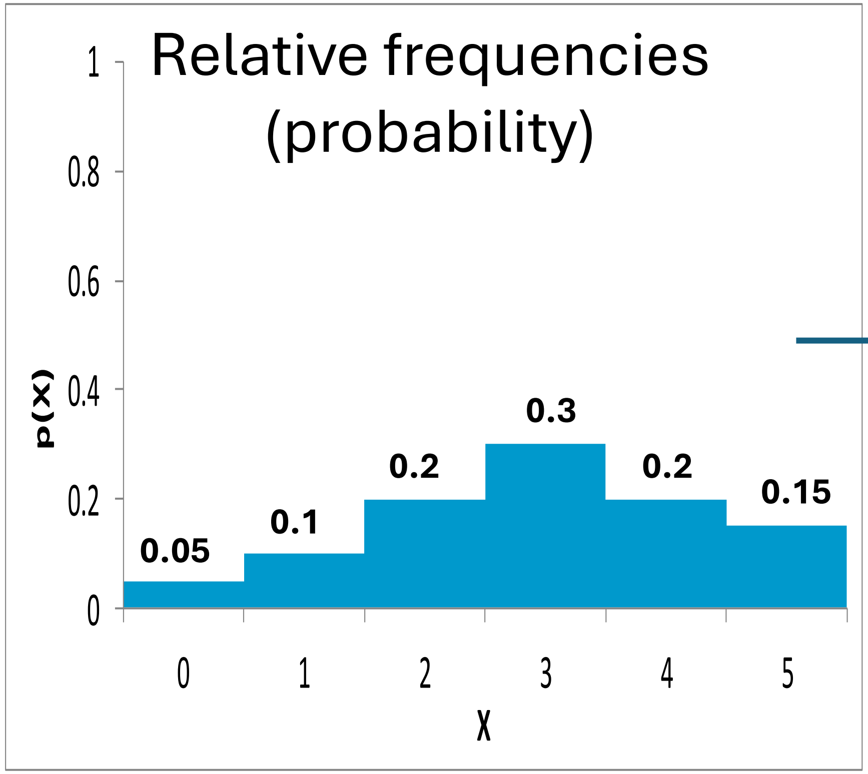
BECAUSE, although the **probability is known** (the relative frequency of each demand level), **the order of occurrence is not**

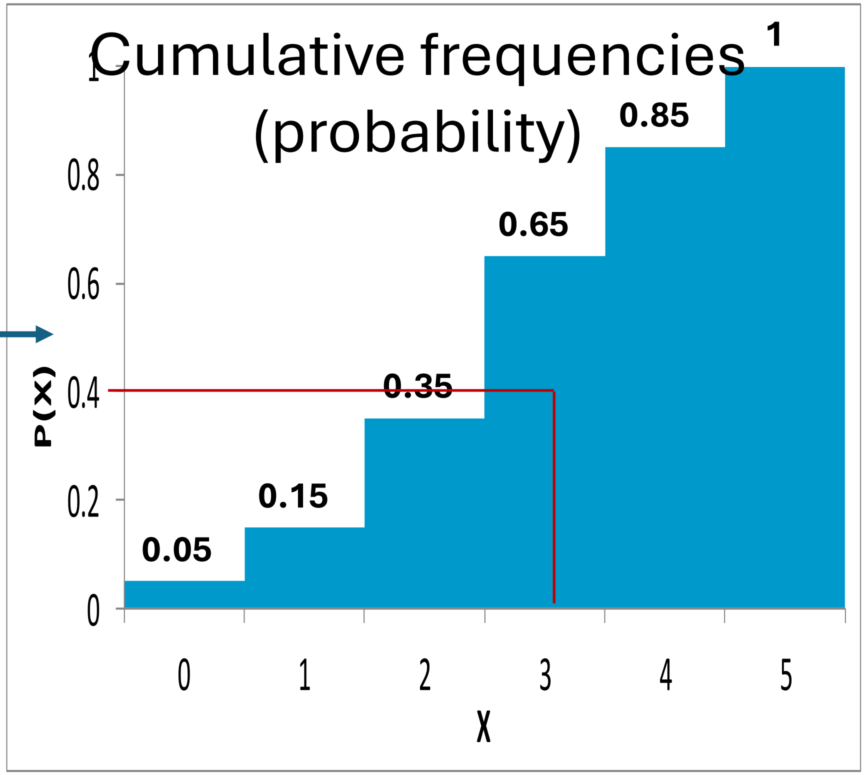
It is **the order of occurrence** (which is assumed random) **which we want to simulate**

Assume that the demand/day is given by:



Assume that the demand/day is given by:

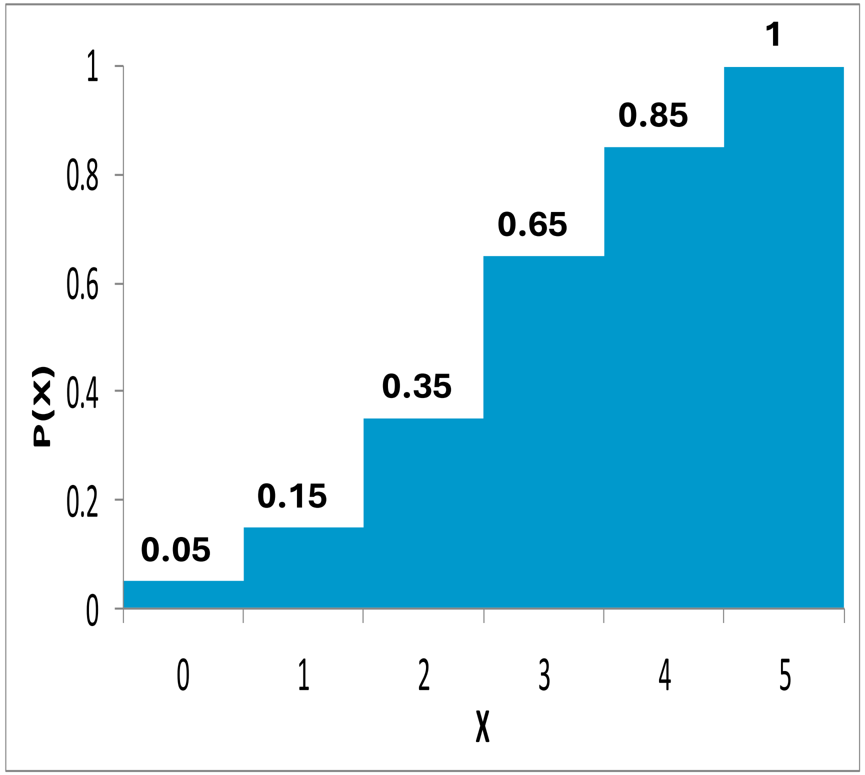




Assume that the demand/day is given by:

Demand (x)	Cumulative frequencies	Interval for random numbers
0	0.05	0 – 4
1	0.15	5 – 14
2	0.35	15 – 34
3	0.65	35 – 64
4	0.85	65 – 84
5	1	85 – 99

Cumulative frequencies (probability)



Assume that the demand/day is given by:

Demand (x)	Cumulative frequencies	Interval for random numbers
0	0.05	0 – 4
1	0.15	5 – 14
2	0.35	15 – 34
3	0.65	35 – 64
4	0.85	65 – 84
5	1	85 – 100

Simulate the demand for 10 days

day	Random number	demand
1	14	1
2		
3		
4		
5		
6		
7		
8		
9		
10		

Assume that the demand/day is given by:

Simulate the demand for 10 days

Demand (x)	Cumulative frequencies	Interval for random numbers
0	0.05	0 – 4
1	0.15	5 – 14
2	0.35	15 – 34
3	0.65	35 – 64
4	0.85	65 – 84
5	1	85 – 100

day	Random number	demand
1	14	1
2	74	4
3		
4		
5		
6		
7		
8		
9		
10		

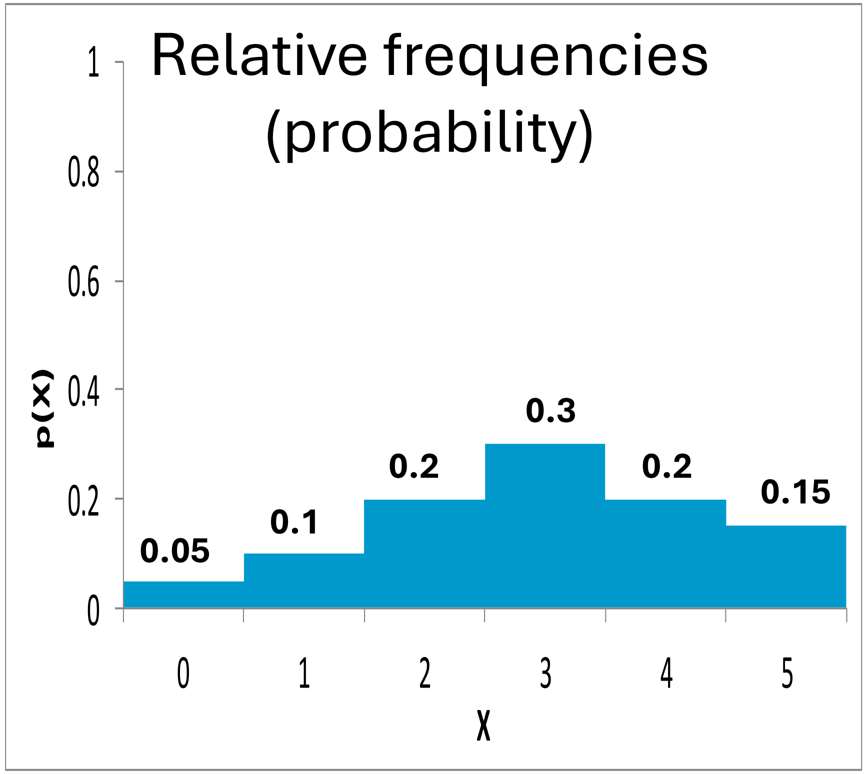
Assume that the demand/day is given by:

Simulate the demand for 10 days

Demand (x)	Cumulative frequencies	Interval for random numbers
0	0.05	0 – 4
1	0.15	5 – 14
2	0.35	15 – 34
3	0.65	35 – 64
4	0.85	65 – 84
5	1	85 – 100

day	Random number	demand
1	14	1
2	74	4
3	24	2
4	87	5
5	7	1
6	45	3
7	26	2
8	66	4
9	26	2
10	94	5

Assume that the demand/day is given by:



If 10000 **random numbers** were drawn it would be expected that the number of observations per class would be:

Demand (x)	frequencies	observations
0	0.05	500
1	0.1	1000
2	0.2	2000
3	0.3	3000
4	0.2	2000
5	0.15	1500

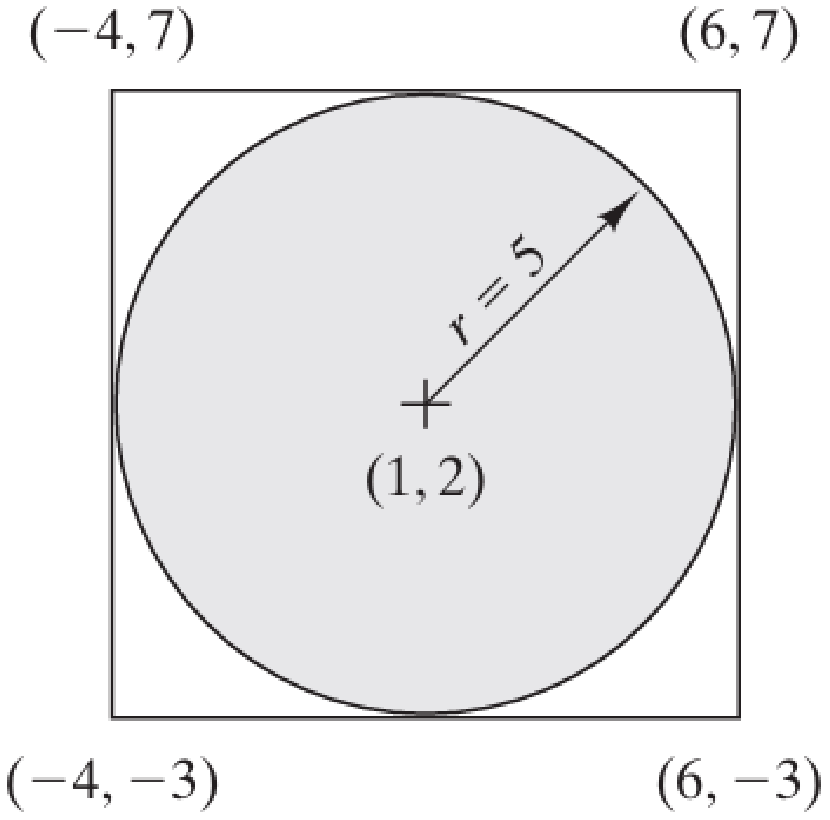
Example 19.1-1

Use Monte Carlo sampling to estimate the area of the following circle: $(x-1)^2 + (y-2)^2 = 25.$

Solution:

The radius of the circle is $r = 5\ cm$, and its centre is $(x, y) = (1,2)$.

The procedure for estimating the area requires enclosing the circle tightly in a square whose side equals the diameter of the circle. The corner points are determined from the geometry of the square.



The estimation of the area of the circle is based on a sampling experiment that gives equal chance to selecting any point in the square. If m out of n sampled points fall within the circle, then

$$\begin{pmatrix}\text{Approximate}\\ \text{area of the circle}\end{pmatrix} = \frac{m}{n}\begin{pmatrix}\text{Area of}\\ \text{the square}\end{pmatrix} = \frac{m}{n}(10 \times 10)$$

To ensure that all the points in the square are equally probable, the coordinates x and y of a point in the square are represented by the following *uniform* distributions:

$$f_1(x) = \frac{1}{10}, -4 \le x \le 6$$

$$f_2(y) = \frac{1}{10}, -3 \le y \le 7$$

TABLE 19.1 A Short List of 0-1 Random Numbers

.0589	.3529	.5869	.3455	.7900	.6307
.6733	.3646	.1281	.4871	.7698	.2346
.4799	.7676	.2867	.8111	.2871	.4220
.9486	.8931	.8216	.8912	.9534	.6991
.6139	.3919	.8261	.4291	.1394	.9745
.5933	.7876	.3866	.2302	.9025	.3428
.9341	.5199	.7125	.5954	.1605	.6037
.1782	.6358	.2108	.5423	.3567	.2569
.3473	.7472	.3575	.4208	.3070	.0546
.5644	.8954	.2926	.6975	.5513	.0305

A pair of 0-1 random numbers, R_1 and R_2, can be used to generate a random point (x, y) in the square by using the following formulas:

$$x = -4 + [6 - (-4)]R_1 = -4 + 10R_1$$

$$y = -3 + [7 - (-3)]R_2 = -3 + 10R_2$$

To demonstrate the application of the procedure, consider $R_1 = .0589$ and $R_2 = .6733$.

$$x = -4 + 10R_1 = -4 + 10 \times .0589 = -3.411$$

$$y = -3 + 10R_2 = -3 + 10 \times .6733 = 3.733$$

This point falls inside the circle because

$$(-3.411 - 1)^2 + (3.733 - 2)^2 = 22.46 < 25$$

Monte Carlo Simulation

- Monte Carlo simulation generates random events.
- Random events in a simulation model are needed when the input data includes random variables.
- To reflect the relative frequencies of the random variables, the *random number mapping* method is used.

JEWEL VENDING COMPANY –
an example for the random mapping technique

- Jewel Vending Company (JVC) installs and stocks vending machines.

- Bill, the owner of JVC, considers the installation of a certain product (“Super Sucker” jaw breaker) in a vending machine located at a new supermarket.

JEWEL VENDING COMPANY –
an example of the random mapping technique

- Data
  - The vending machine holds 80 units of the product.
  - The machine should be filled when it becomes half empty.
  - Daily demand distribution is estimated from similar vending machine placements.
    - P(Daily demand = 0 jaw breakers) = 0.10
    - P(Daily demand = 1 jaw breakers) = 0.15
    - P(Daily demand = 2 jaw breakers) = 0.20
    - P(Daily demand = 3 jaw breakers) = 0.30
    - P(Daily demand = 4 jaw breakers) = 0.20
    - P(Daily demand = 5 jaw breakers) = 0.05

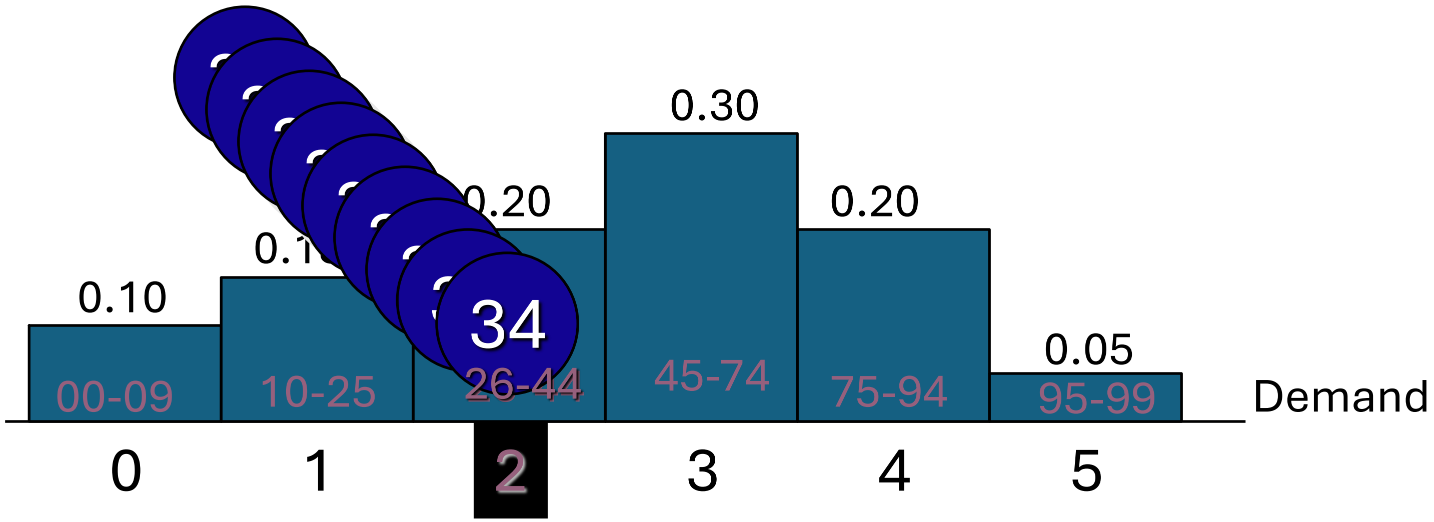
Bill would like to estimate the expected number of days it takes for a filled machine to become half empty.

Random number mapping –
The Probability function Approach

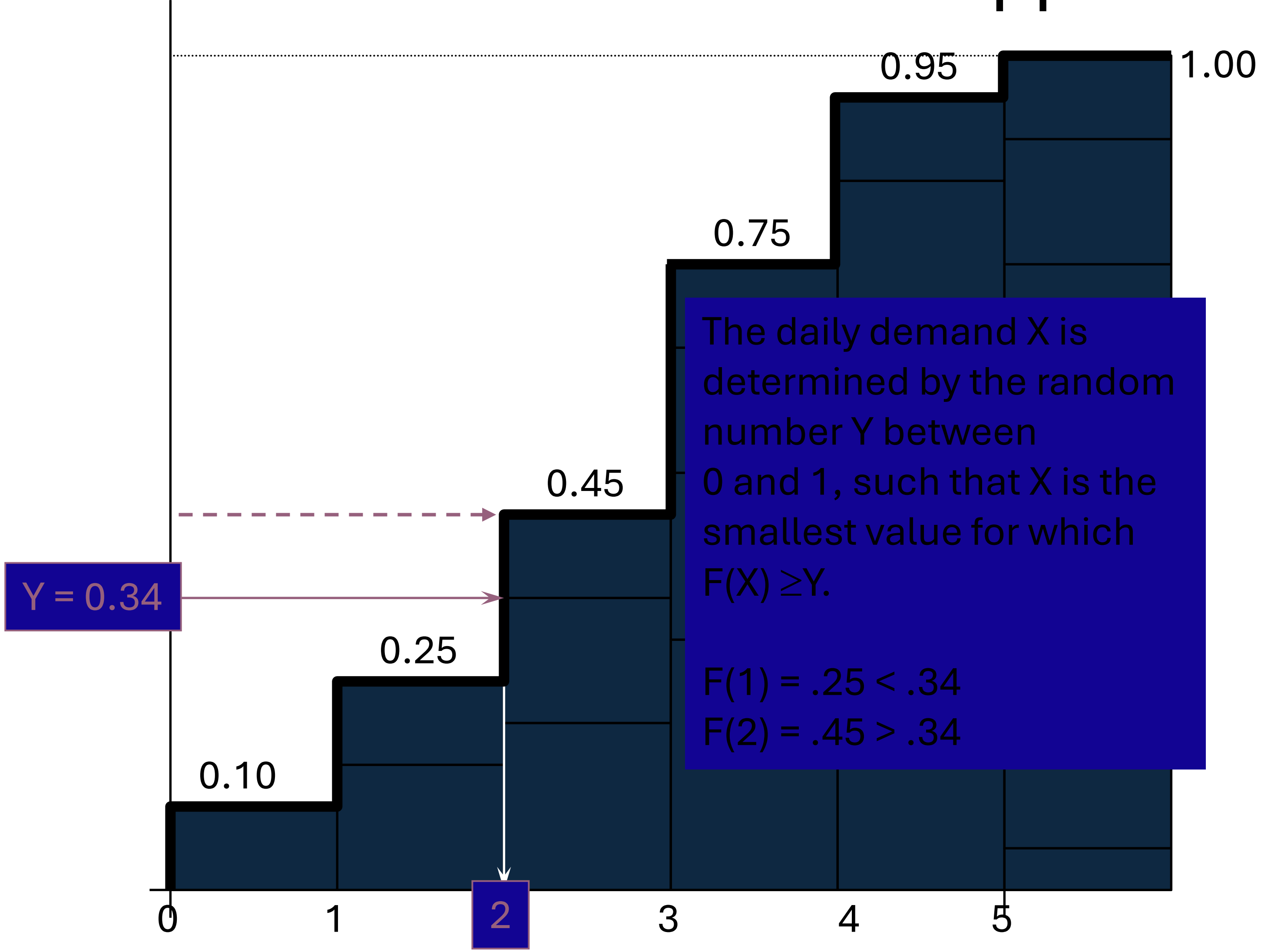
Random number mapping uses the probability function to generate random demand.

A number between 00 and 99 is selected randomly.

The daily demand is determined by the mapping demonstrated below.



Random number mapping –
The Cumulative Distribution Approach



The daily demand X is determined by the random number Y between 0 and 1, such that X is the smallest value for which $F(X) \geq Y$.

$F(1) = .25 < .34$
$F(2) = .45 > .34$

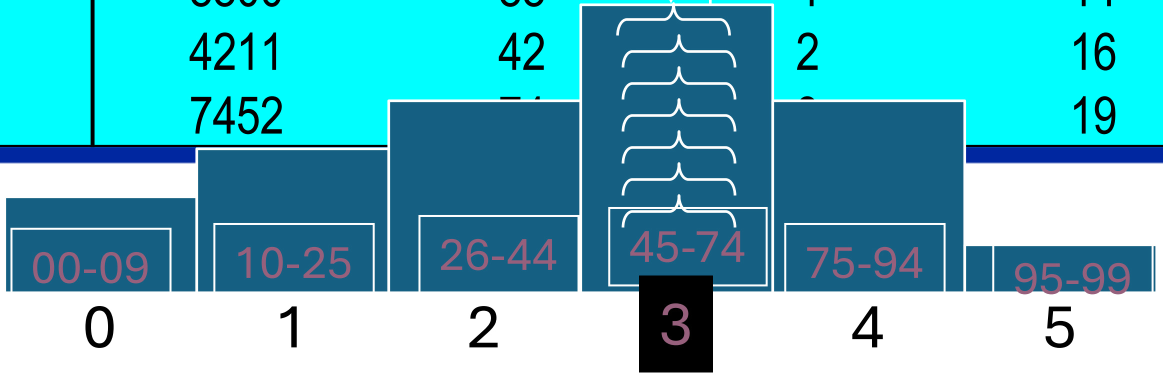
Simulation of the JVC Problem

- A random demand can be generated by hand (for small problems) from a table of pseudo random numbers.
- Using Excel a random number can be generated by
  - The RAND() function
  - The random number generation option (Tools>Data Analysis)

Simulation of the JVC Problem

- An illustration of generating a daily random demand.
- Since we have two digit probabilities, we use the first two digits of each random number.

Day	Random Number	Two First Digits	Demand	Total Demand to Date
1	6506	65	3	3
2	7761	77	4	7
3	6170	61	3	10
4	8800	88	4	14
5	4211	42	2	16
6	7452	[illegible]	[illegible]	19



Simulation of the JVC Problem

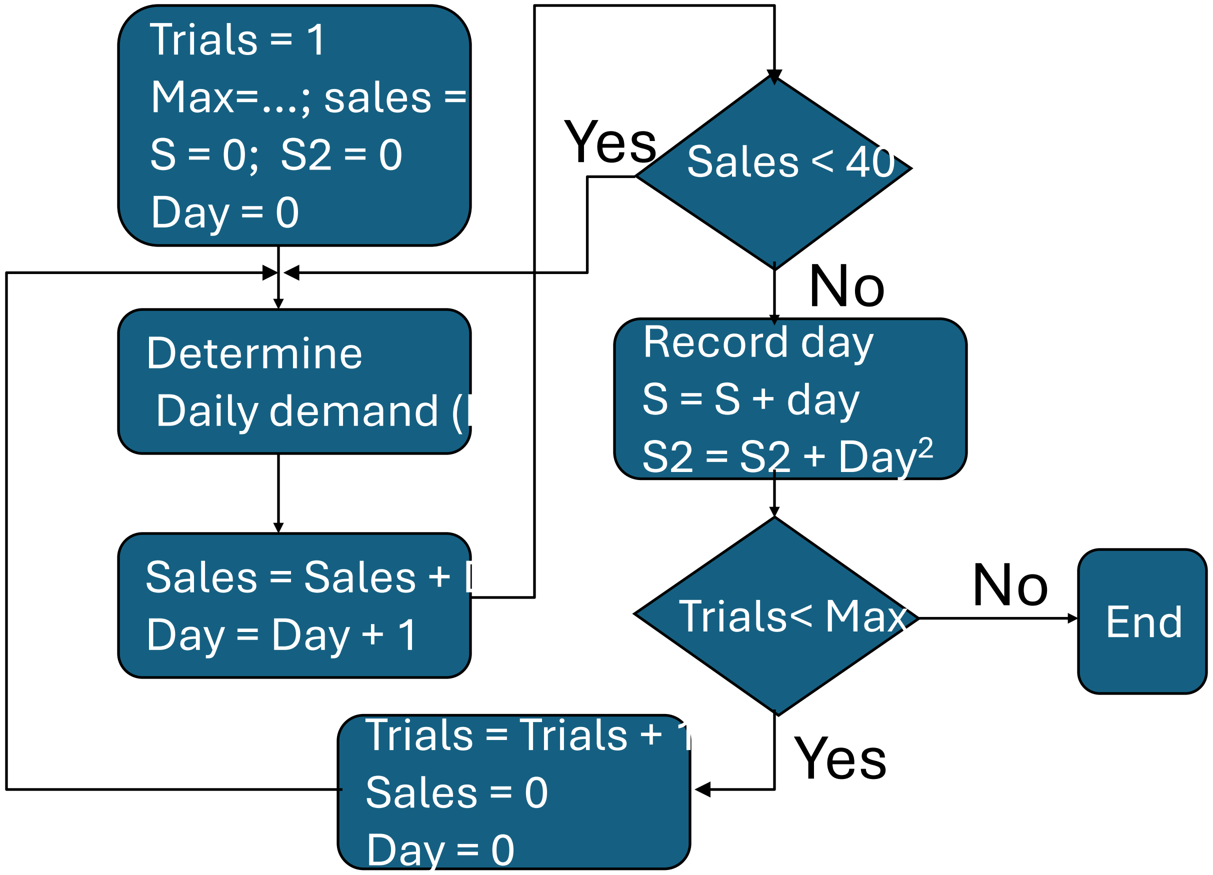
The simulation is repeated and stops once total demand reaches 40 or more.

Day	Random Number	Two First Digits	Demand	Total Demand to Date
1	6506	65	3	3
2				7
3				10
4				14
5				16
6				19

The number of "simulated" days required for the total demand to reach 40 or more is recorded.

JVC – A Flow Chart

Flow charts help guide the simulation program



JVC – Excel Spreadsheet

	A	B	C	E	F	G	H	I	J	K	L
1	Number of Days to Sell 40 Jaw Breakers =				19						
2											
3			Cumulative								
4	Day	Demand	Demand								
5											
6	1	2	2							0	0
7	2	2	4							0.1	1
8	3	3	7							0.25	2
9	4	2	9							0.45	3
10	5	3	12							0.75	4
11	6	0	12							0.95	5
12	7	3	15								
13	8	0	[illegible]								
14	9	2	[illegible]								
15	10	2	19								
16	11	4	23								
17	12	1	24								
18	13	4	28								
19	14	2	30								
20	15	3	33								

=MAX(A5:A1

=IF(C5<40,A5+1,

=IF(C5<40,B6+C5

=IF(C5<40,VLOOKUP(RAND(),$K$6:$L$11,
2),"")

Drag A5:C5 to A105:C105

VLOOKUP TABLE Enter this data

JVC – Excel Spreadsheet

	A	B	C	D	E	F	G	H
1	**Replication**	**Days**		*Days*				
2	1	18						
3	2	15		Mean	16.6			
4	3	14		Standard Error	0.541603			
5	4	18		Median	17		t	1.107823
6	5	17		Mode	18		p-value	0.296665
7	6	17		Standard Deviation	1.712698			
8								
9					527			
10					232			
11								
12					14			
13					19			
14					166			
15					10			

=(E3-16)/E4

=TDIST(ABS(H5),9,2)

- The p-value =.2966… This value is quite high compared to any reasonable significance level.
- Based on this data there is insufficient evidence to infer that the mean number of days differs from 16.