Q. What is Montecarlo Simulation

A **Monte Carlo simulation** is a **statistical simulation technique** that provides approximate solutions to problems expressed mathematically. It utilizes a sequence of random numbers to perform the simulation.

Q. What are the three basic types of parameters to describe the distribution

The three basic types of parameters to describe the distribution of a data set are:

1. Measures of central tendency: These parameters describe where the center of the distribution lies. The three commonly used measures of central tendency are **mean, median, and mode.**
2. Measures of variability: These parameters describe how spread out the data is. The commonly used measures of variability are **standard deviation, variance**.
3. Measures of shape: These parameters describe the shape of the distribution, which can be **symmetrical, skewed, or peaked**. Some commonly used measures of shape include skewness, kurtosis, and measures based on quantiles such as the quartile skewness and the quartile coefficient of skewness.

Q. What is Simulation?

Simulation is the process of creating a computer model or a mathematical representation of a system, process, or phenomenon to study and analyze its behavior or performance under different conditions. The goal of simulation is to provide insights into how a system or process works, and to explore different scenarios to understand how changes in variables or parameters affect the outcome.

Q. Why do we need Simulation?

- To understand existing (natural or human) systems
- To build new systems
- Show the eventual real effects of alternative conditions and courses of action.
- The real system cannot be engaged, because it may not be accessible, or it may be dangerous or unacceptable to engage, or it is being designed but not yet built, or it may simply not exist.

It is useful when experimentation with the real system is expensive, dangerous or likely to cause significant disruption (e.g., transport systems, nuclear reactor and airline systems).

It might also be an option when mathematical modelling of a system is impossible.

Logistics operations management requires more sophisticated technologies, such as simulation that can handle the inherent uncertainty of real-world logistics systems.

Q. Key Steps in Simulation

- Collecting valid and relevant source information about the system (including key characteristics and behaviors)
- Setting, approximations, and assumptions within the simulation
- Justifying the fidelity and validity of the simulation outcomes (data)

Q. Briefly state the procedure of simulation

- Define the problem: Identify the system or process to be modeled and specify the objectives and scope of the simulation.
- Develop a conceptual model: Construct a high-level, simplified representation of the system, based on knowledge of the relevant physical, mathematical, and logical principles.
- Create a mathematical model: Translate the conceptual model into a mathematical representation, using equations, algorithms, and other analytical tools.
- Implement the model: Use a programming language or simulation software to code the mathematical model and set up the simulation environment.
- Define inputs and parameters: Specify the values and ranges of input variables, as well as any parameters or assumptions used in the model.
- Run the simulation: Execute the simulation, using the inputs and parameters defined in step 5, and collect the output data.
- Analyze the results: Use statistical methods and visualization tools to analyze the output data and draw conclusions about the behavior and performance of the system.
- Validate and refine the model: Compare the simulation results with real-world data and refine the model as necessary to improve accuracy and predictive power.
- Communicate the findings: Present the simulation results in a clear, concise, and meaningful way to stakeholders, decision-makers, and other interested parties.

Q. What is Chi-Square Godness fit test

The chi-square goodness of fit test is used to test whether the frequency distribution of a categorical variable is different from your expectations.

Q. How to perform a chi-square test

The exact procedure for performing a Pearson's chi-square test depends on which test you're using, but it generally follows these steps:

1. Create a table of the observed and expected frequencies. This can sometimes be the most difficult step because you will need to carefully consider which expected values are most appropriate for your null hypothesis.
2. Calculate the chi-square value from your observed and expected frequencies using the chi-square formula.
3. Find the critical chi-square value in a chi-square critical value table or using statistical software.
4. Compare the chi-square value to the critical value to determine which is larger.
5. Decide whether to reject the null hypothesis. You should reject the null hypothesis if the chi-square value is greater than the critical value. If you reject the null hypothesis, you can conclude that your data are significantly different from what you expected.

Bernoulli Distribution: Bernoulli (p)

- Used to model the distribution with two possible outcomes
  - E.g. coin flipping
  - A customer will click a login button or not
  - A server is up or down
- Parameter: p

$$p(x) = \begin{cases} 1-p & if \quad x = 0 \\ p & if \quad x = 1 \\ 0 & otherwise \end{cases}$$

Mean: *p*, Variance: *p*(1−*p*)

Chi-Square:

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Where:

- X^2 is the chi-square test statistic
- Σ is the summation operator (it means “take the sum of”)
- *O* is the observed frequency
- *E* is the expected frequency

Binomial Distribution: bin(n,p)

- Used to model the number of *x* successes in *n* Bernoulli trials with probability *p* of success on each trial
  - e.g. number of defective items in a batch of size *n*
  - no. of packets that reach the destination without loss
- Constant probability for each observation

✓✓××✓×××✓✓✓××××.... ××

n trials, where x is the number of ✓

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \; \forall x \in \{0,1,\cdots\}$$

$$where \binom{n}{x} = \text{combinations of selecting x items out of n objects}$$

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

Mean: *np*, Variance: *np*(1−*p*)

Geometric Distribution: geom(p)

- To model the number of failures before the first success in a sequence of independent Bernoulli trials with probability p of success on each trial
  - E.g. Number of items inspected before encountering the first defective item
  - E.g. life time (discrete, e.g. days) of disk access before crash

××××××...××✓

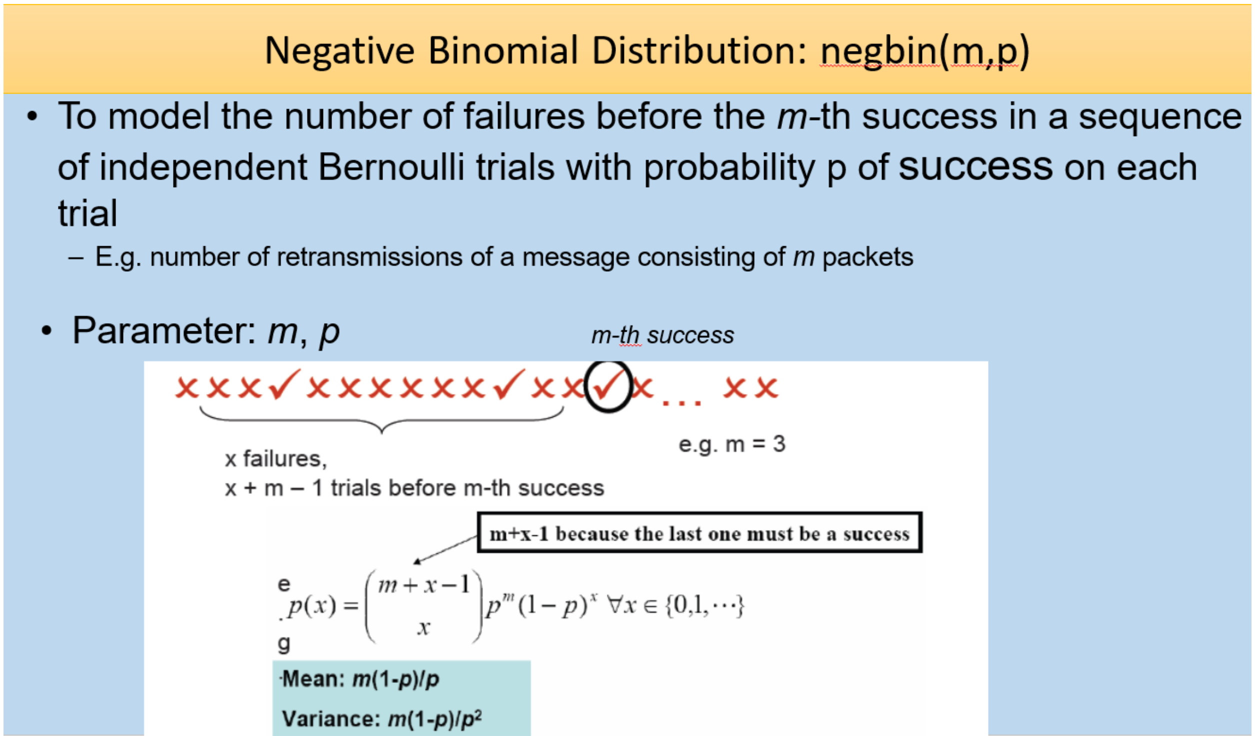
Parameter: *p*

$$p(x) = p(1-p)^x \; \forall x \in \{0,1,\cdots\}$$

Mean: (1-*p*)/*p*

Variance: (1−*p*)/*p*²

A Poisson distribution is a discrete probability distribution, meaning that it gives the probability of a discrete (i.e., countable) outcome. For Poisson distributions, the discrete outcome is the number of times an event occurs, represented by k.



Mean: the "center" of the entire population

Average: the "center" of a sample of the population

Q. What is the difference between a chi-square test and a correlation?

Both correlations and chi-square tests can test for relationships between two variables. However, **a correlation** is used when you have **two quantitative variables**, and a chi-square test of independence is used when you have two categorical variables.

Correlation describes an association between types of variables: when one variable changes, so does the other. A correlation is a statistical indicator of the relationship between variables. These variables change together.

Weibull Distribution: Weibull(α , β)

- Used to model time to complete a task
  - Instead of having a constant mean service time (such as exponential), such value is a variable in Weibull (depending on α)
    - If the service time decreases over time, then $\alpha < 1$
    - If the service time is constant over time, then $\alpha = 1$ (= exponential)
    - If the service time increases over time, then $\alpha > 1$

$$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} \quad \forall x > 0$$

$$\text{Mean}: \quad \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$$

$$\text{Variance}: \frac{\beta^2}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}$$

Hypothesis testing is a formal procedure for investigating our ideas about the world using statistics. It is most often used by scientists to test specific predictions, called hypotheses, that arise from theories.

There are 5 main steps in hypothesis testing:

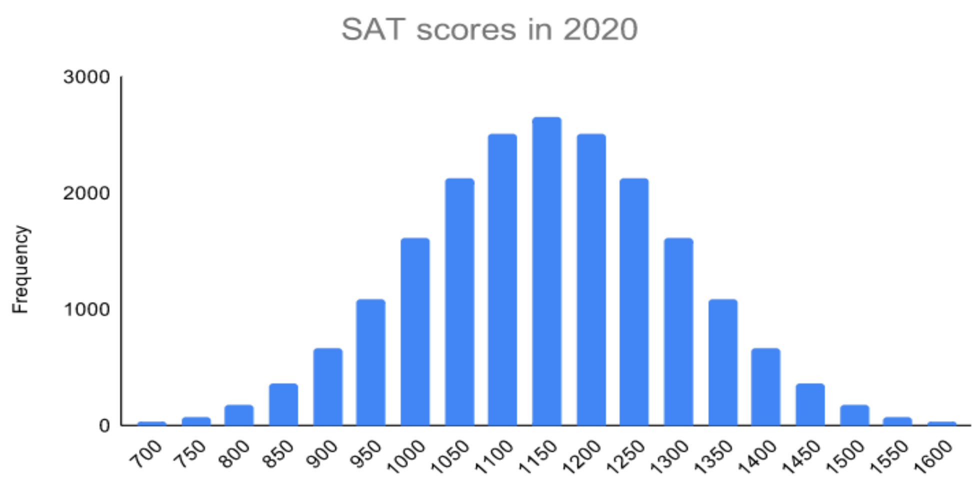
1. State your research hypothesis as a null hypothesis and alternate hypothesis (H_o) and (Ha or H1).
2. Collect data in a way designed to test the hypothesis.
3. Perform an appropriate statistical test.
4. Decide whether to reject or fail to reject your null hypothesis.
5. Present the findings in your results and discussion section.

Normal Distribution | Examples, Formulas, & Uses

Published on October 23, 2020 by Pritha Bhandari. Revised on November 25, 2022.

In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center.

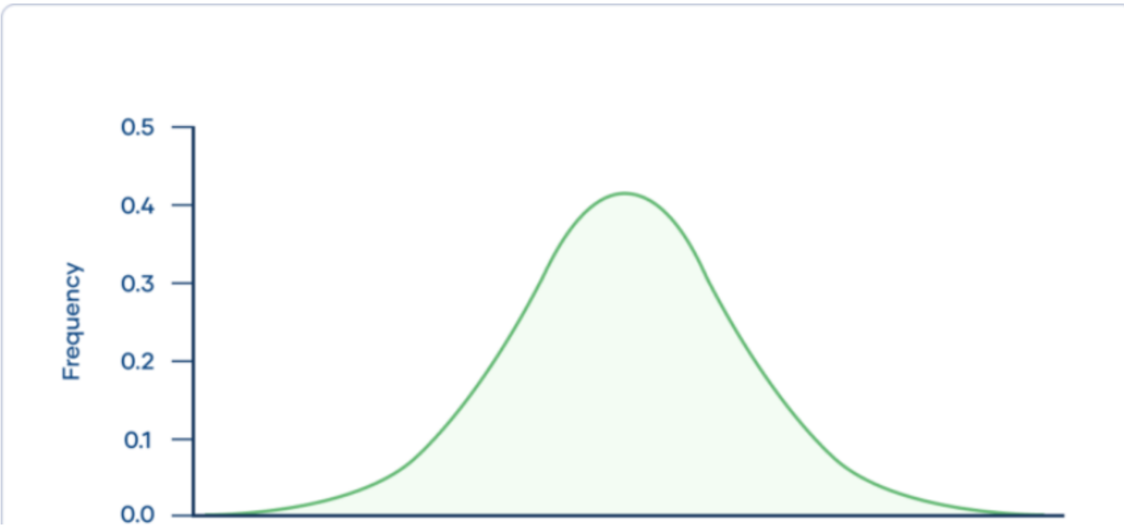
Normal distributions are also called Gaussian distributions or bell curves because of their shape.



What are the properties of normal distributions?

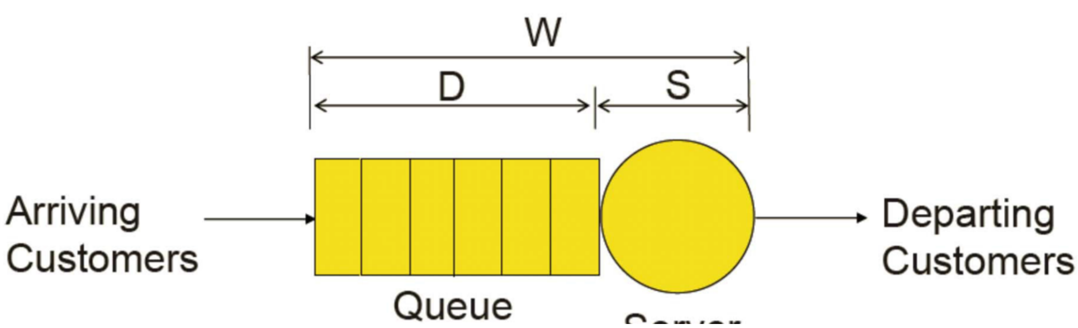
Normal distributions have key characteristics that are easy to spot in graphs:

- The mean, median and mode are exactly the same.
- The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.
- The distribution can be described by two values: the mean and the standard deviation.



Single-Queue Single-Server System

- Single-server queueing system
  - A Queueing System consists of one or more servers that provide service to arriving customers
  - A customer who arrive to find all servers busy join one in front of the server
  - Single-Queue Single-Server system: a queueing system with only one server which has only one queue
- For a customer: W = total waiting time in the system, D = delay in queue, S = service time



- A single-queue single-server system
- When a customer arrives
  - If the server is idle: enters service immediately
  - If the server is busy: wait at the end of the queue
- Service discipline: FIFO (first-in-first-out)
- Random variables:
  - Interarrival times A_1, A_2, … are IID
  - Service times S1, S2, … are IID
  - The two sets variables are independent of each other

- Initial state: “empty-and-idle”
  - time t = 0
  - no customers
  - idle server
- Wait for the arrival of the first customer
  - occurs at the first interarrival time A_1
- Stopping-rule: the n-th customer enters the service
  - A fixed number of customers have completed their delays in queue the system
  - Note: the time the simulation ends is also a random variable!

- *d(n): Expected average delay in queue* of n-th customer
  - excluding service time
  - "Expected": the average delay is a random variable
  - "the average of averages"

- *q(n): Expected average number of customers in queue* (excluding any in service)
-
- *u(n): Expected utilization of the server*
  - proportion of time busy
  - How busy the server is

- Q(t): the queue length at time t

- B(t): the busy function against t

$$B(t) = \begin{cases} 1 & \text{if the server is busy at time } t \\ 0 & \text{if the server is idle at time } t \end{cases}$$

- T(n): the time required to observe n delays in queue
  - time elapsed when the n-th customers completes its delay in queue and start being served
  - In our example, simulation will ends at T(6)

Advantages of modelling and simulation

- Can be safer and cheaper than the real world.
- Able to test a product or system works before building it.
- Able to explore 'what if…' questions.
- Can speed things up or slow them down to see changes over long or short periods of time.
- Results are accurate in general, compared to analytical model.
- Help to find un-expected phenomenon, behavior of the system.

Disadvantages of modelling and simulation

- Mistakes may be made in the programming or rules of the simulation or model.
- The cost of a simulation model can be high.
- The cost of running several different simulations may be high.
- Time may be needed to make sense of the results.
- People's reactions to the model or simulation might not be realistic or reliable.
- Sometimes it is difficult to interpret the simulation results.

Q. Use of Simulation in research

Simulation is a powerful tool that can be used in research to model complex systems and explore the effects of different variables and scenarios. Simulation involves creating a mathematical or computer model of a system or process, and then using the model to generate data and explore hypothetical scenarios.

1. Explore hypothetical scenarios: Simulations allow researchers to test different scenarios and conditions without having to conduct expensive or time-consuming experiments.
2. Generate data: Simulations can generate data that can be used to test hypotheses and inform decision-making.
3. Optimize systems: Simulations can be used to optimize complex systems by identifying bottlenecks, improving efficiency, and reducing costs.
4. Predict outcomes: Simulations can be used to predict the outcomes of different scenarios and inform policy decisions.

Q.

Explain the mechanisms of Next-event time advance and Fixed-increment time advance in simulation.

In a discrete-event simulation, the state of the system is updated at specific points in time, known as "events." There are two main approaches for advancing the simulation from one event to the next: next-event time advance and fixed-increment time advance. Next-event time advance: In next-event time advance, the simulation advances to the next event that is scheduled to occur. This means that the simulation time is set to the exact time at which the next event is scheduled to occur. For example, if the next event is a customer arriving at the queue at time t=5, the simulation time would be advanced to t=5 and the event would be processed. Fixed-increment time advance: In fixed-increment time advance, the simulation advances by a fixed amount of time (also known as a "time step") at each iteration. For example, if the time step is 1 minute, the simulation time would be advanced by 1 minute at each iteration, regardless of whether an event is scheduled to occur at that time. Both next-event time advance and fixed-increment time advance have their own benefits and limitations, and the appropriate approach will depend on the specific characteristics of the system being modeled.

Q. Why does Single-server queue is called Discrete-Event Simulation?

A single-server queue is called a discrete-event simulation because it models the behavior of a system as a sequence of discrete events. In a single-server queue, the events might include the arrival of a new customer, the departure of a customer who has completed service, and other events such as the starting and ending of breaks for the server. Each of these events occurs at a specific point in time, and the simulation tracks the evolution of the system over time by advancing from one event to the next. This approach is called a discrete-event simulation because it models the system as a sequence of discrete events rather than as a continuous process

Q. Write the application of these distribution in simulation.

The negative binomial distribution and the geometric distribution are often used in simulation to model the number of failures that occur before a certain number of successes are achieved. Here are some examples of the use of these distributions in simulation:

1. Reliability testing: The negative binomial distribution can be used to model the number of times a machine must be used before it fails, given a certain number of successes (such as the number of hours of operation). This information can be used to predict the reliability of the machine and to identify potential failure points. **2. Marketing campaigns:** The geometric distribution can be used to model the number of times a customer must be contacted before making a purchase, given the probability of making a purchase on each contact. This information can be used to optimize marketing campaigns and to predict the likelihood of different outcomes occurring. **3. Quality control**: The negative binomial distribution can be used to model the number of defective items that are produced before a certain number of good items are produced, given the probability of producing a good item. This information can be used to identify and address quality control issues and to optimize production processes. **4. Inventory management**: The geometric distribution can be used to model the number of days that an item must be stored before it is sold, given the probability of selling the item on each day. This information can be used to optimize inventory management and to predict the likelihood of different outcomes occurring.