# Mod-1

Model characterization refers to the process of describing and understanding the properties, features, and behavior of a model. It involves analyzing various aspects of the model to gain insights into its structure, assumptions, limitations, and performance characteristics.

Here are some key elements of model characterization:

1. **Structure**: Describing the components and relationships within the model. This includes identifying input variables, output variables, parameters, equations, algorithms, and any underlying assumptions.
2. **Assumptions**: Explicitly stating the assumptions made in developing the model. Understanding the assumptions helps users interpret the model's outputs and assess its applicability to different scenarios.
3. **Limitations**: Identifying the constraints or limitations of the model. No model perfectly represents reality, so it's essential to understand its limitations and potential sources of error.
4. **Validation**: Assessing the accuracy and reliability of the model by comparing its outputs to real-world data or other validated models. Validation helps establish confidence in the model's predictive capabilities.
5. **Sensitivity Analysis**: Analyzing how changes in input parameters or assumptions affect the model's outputs. Sensitivity analysis helps identify which factors have the most significant impact on the results and how robust the model is to variations in input.
6. **Uncertainty**: Characterizing the uncertainty inherent in the model outputs. This involves quantifying uncertainties in input data, parameters, and assumptions and understanding their implications for decision-making.
7. **Performance Metrics**: Defining metrics to evaluate the performance of the model, such as accuracy, precision, bias, and computational efficiency. These metrics help assess the model's suitability for its intended purpose.

**Table 1.1** Examples of Systems and Components

| *System* | *Entities* | *Attributes* | *Activities* | *Events* | *State Variables* |
|---|---|---|---|---|---|
| Banking | Customers | Checking-account balance | Making deposits | Arrival; departure | Number of busy tellers; number of customers waiting |
| Rapid rail | Riders | Origination; destination | Traveling | Arrival at station; arrival at destination | Number of riders waiting at each station; number of riders in transit |
| Production | Machines | Speed; capacity; breakdown rate | Welding; stamping | Breakdown | Status of machines (busy, idle, or down) |
| Communications | Messages | Length; destination | Transmitting | Arrival at destination | Number waiting to be transmitted |
| Inventory | Warehouse | Capacity | Withdrawing | Demand | Levels of inventory; backlogged demands |

# ORDER-UPTO LEVEL INVENTORY SYSTEM

# REFRIGERATORY INVENTORY SYSTEM

Consider the refrigerator inventory system with maximum inventory level(M) is 11 units and the review period (N) is 5 days. Estimate by simulation. The average ending units in inventory and number of days when shortage condition occurs. Initially the simulation is started with inventory level of 3 units and an order of 8 units.Schedule to arrive in two days time

Simulate for 3 cycles

Random digit for daily demand : 24,35,65,81,54,3,87,27,73,70,47,45,,48,17,9

Random digit for lead time : 5,0,3

Solution:

Random distribution for daily demand distribution

| Demand | Probability | Cumulative probability | Random digit assignment |
|---|---|---|---|
| 0 | 0.1 | | |
| 1 | 0.25 | | |
| 2 | 0.35 | | |
| 3 | 0.21 | | |
| 4 | 0.09 | | |

Consider the refrigerator inventory system with maximum inventory level(M) is 11 units and the review period (N) is 5 days. Estimate by simulation. The average ending units in inventory and number of days when shortage condition occurs. Initially the simulation is started with inventory level of 3 units and an order of 8 units,Schedule to arrive in two days time

Simulate for 3 cycles

Random digit for daily demand : 24,35,65,81,54,3,87,27,73,70,47,45,,48,17,9

Random digit for lead time : 5,0,3

Solution:

Random distribution for daily demand distribution

| Demand | Probability | Cumulative probability | Random digit assignment |
|---|---|---|---|
| 0 | 0.1 | 0.10 | 01-10 |
| 1 | 0.25 | 0.35 | 11-35 |
| 2 | 0.35 | 0.70 | 36-70 |
| 3 | 0.21 | 0.91 | 71-91 |
| 4 | 0.09 | 1.00 | 92-00 |

3:50 / 19:54 • Cumulative Probability >

Random distribution for lead time distribution

| Lead time | Probability | Cumulative probability | Random digit assignment |
|---|---|---|---|
| 1 | 0.6 | 0.6 | 1-6 |
| 2 | 0.3 | 0.9 | 7-9 |
| 3 | 0.1 | 1.0 | 0 |

Each cycle review period = 5 days

3 cycles= 3*5= 15days

There are 3 inventory level on each day, so there is 3 lead time

5:38 / 19:54 • Lead Time Distribution >

Simulation table for M-N inventory system

| Day | cycle | Day with in cycle | Beginning inventory | RD for demand | Demand | Ending inventory | Shortage quantity | Order quantity | RD for lead time | Lead time | Day until order arrives |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 3 | 24 | 1 | 2 | 0 | | | | |
| 2 | | 2 | 2 | 35 | 1 | 1 | 0 | | | | |
| 3 | | 3 | 1+8=9 | 65 | 2 | 7 | 0 | | | | |
| 4 | | 4 | 7 | 81 | 3 | 4 | 0 | | | | |
| 5 | | 5 | 4 | 54 | 2 | 2 | 0 | 11-2=9 | 5 | 1 | 1 |
| 6 | 2 | 1 | 2 | 03 | 0 | 2 | 0 | | | | |
| 7 | | 2 | 9+2=11 | 87 | 3 | 11-3=8 | 0 | | | | |
| 8 | | 3 | 8 | 27 | 1 | 7 | 0 | | | | |
| 9 | | 4 | 7 | 73 | 3 | 4 | 0 | | | | |
| 10 | | 5 | 4 | 70 | 2 | 2 | 0 | 11-2=9 | 0 | 3 | 3 |
| 11 | 3 | 1 | 2 | 47 | 2 | 0 | 0 | | | | |
| 12 | | 2 | 0 | 45 | 2 | 0 | 2 | | | | |
| 13 | | 3 | 0 | 48 | 2 | 0 | 2+2=4 | | | | |
| 14 | | 4 | 9 | 17 | 1 | 9-5=4 | 0 | | | | |
| 15 | | 5 | 4 | 09 | 0 | 4 | 0 | 11-4=7 | 3 | | |


| | | | | | | |
|---|---|---|---|---|---|---|
| 14 | 3 | 4 | 9 | 17 | 1 | 9-5=4 |
| 15 | | 5 | 4 | 09 | 0 | 4 |


Average ending inventory= $\frac{\textit{Total number of ending inventory}}{\textit{number of days}} = \frac{47}{15} =$

Number of day shortage units =2

# Mod-2

1. **Random Numbers**: True random numbers are generated from processes or phenomena that are inherently unpredictable, such as radioactive decay, atmospheric noise, or chaotic systems. These numbers have no discernible pattern and cannot be predicted in advance. True randomness is difficult to achieve in computational systems because computers operate based on deterministic algorithms.
2. **Pseudo-Random Numbers**: Pseudo-random numbers are generated using deterministic algorithms that produce sequences of numbers that appear to be random. These algorithms start with an initial value called a seed and use mathematical formulas to generate subsequent numbers in the sequence. While the numbers generated by pseudo-random number generators (PRNGs) may exhibit statistical properties similar to those of true random numbers, they are ultimately deterministic and repeatable given the same initial seed. As a result, pseudo-random numbers are not truly random but are instead deterministic sequences that mimic randomness for practical purposes.

**Properties of random Numbers:**
A sequence of random number R1, R2 … must have two important statistical properties, uniformity and independence.

**Uniformity** :
If the interval (0, 1) is divided into „n" classes or subintervals of equal length , the expected number of observations in each interval is N/n, where N is total number of observations.

**Independence**:
The probability of observing a value in a particular interval is independent of the previous drawn value.

**Problems faced in generating random numbers:**
1. The generated number may not be uniformly distributed.
2. The number may be discrete valued instead of continuous values.
3. The mean of the numbers may be too high or low
4. The variance of the number may be too high or low.
5. The numbers may not be independent
   e.g. a. Autocorrelation between numbers
   b. Numbers successively higher or lower than adjacent numbers.

**Criteria for random no. generator**:
1. The routine should be fast.
2. The routine should be portable.
3. The routine should have a sufficient long cycle. The cycle length or period represents the length of random number sequence before previous numbers begin to repeat themselves in an earlier order. A special case of cycling is degenerating. A routine degenerates when some number appears repeatedly which is unacceptable.
4. The random number should be replicable.
5. Most important, the generated random numbers should closely approximate to the ideal statistical properties of uniformity and independence.

K-S Test :-

1) Arrange the given no in ascending order.
2) Compute :-

$$D^{+} = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_i \right\}$$

$$D^{-} = \max_{1 \le i \le N} \left\{ R_i - \left( \frac{i-1}{N} \right) \right\}$$

N - total no. of random no
degree of freedom
Sample Size

3) Compute :-

$$D = \max (D^{+}, D^{-})$$

4) Compute theoretical value $D_\alpha$ for a given level of significance $\alpha$ from standard KS Test table.
5) If $D < D_\alpha$ then accept hypothesis, else reject

o/p determined by Null Hypothesis

**For chi square test sample size n ke lye kitne interval lene hai**

**Table 9.5** Recommendations for Number of Class Intervals for Continuous Data

| *Sample Size, n* | *Number of Class Intervals, k* |
|---|---|
| 20 | Do not use the chi-square test |
| 50 | 5 to 10 |
| 100 | 10 to 20 |
| >100 | $\sqrt{n}$ to $n/5$ |

- Runs Up & Runs Down Algo :-

S1 :- Define Hypothesis as -

$H_0$ :- $R_i \sim$ independently

$H_1$ :- $R_i \nsim$ independently

S2 :- Write sequence of runs up & runs down

S3 :- Count total no. of runs (a) present in sequence

S4 :- Count mean & variance a :-

$$\mu_a = \frac{2N-1}{3} \qquad \sigma_a^2 = \frac{16N-29}{90}$$

S5 :- Std normal statistics :-

$$Z_0 = \frac{a-\mu_a}{\sigma_a} \; ; \; Z_0 \sim N[0,1]$$

S6 :- Determine critical value, $-Z_{\alpha/2} \le Z_0 \le Z_{\alpha/2}$

If above cond$^n$ checked

$\Rightarrow$ $H_0$ not rejected.

**Example 2**: The theory predicts the proportion of beans, in the four groups A, B, C and D should be 9:3:3:1. In an experiment among 1600 beams, the number of four groups was 882, 313, 287 and 118. Does the experiment result support the theory?

**Solution:**

**Step 1:** Null Hypothesis: There is no significant between the experiment value and the theory.

$$9:3:3:1 \qquad \frac{9}{16} \quad \frac{3}{16} \quad \frac{3}{16} \quad \frac{1}{16}$$

Chi-Square Test: Goodness of Fit

**Step 2:** Calculate $E$ and $\chi^2$:

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

$= 4.726$

| Group | $O$ | $E$ | $(O-E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|
| A | 882 | $\frac{9}{16} \times 1600 = 900$ | 324 | 0.360 |
| B | 313 | $\frac{3}{16} \times 1600 = 300$ | 169 | 0.563 |
| C | 287 | $\frac{3}{16} \times 1600 = 300$ | 169 | 0.563 |
| D | 118 | $\frac{1}{16} \times 1600 = 100$ | 324 | 3.240 |
| Total | 1600 | 1600 | | 4.726 |

**Step 3:** d.f. = 4 - 1 = 3

Tabulated $\chi^2(0.05)$ for 3 d.f. is 7.81.

12:40 / 22:58 • Chapter 11 >

- Runs Above & Below Mean Test :-

S1 :- Define Hypothesis for independence

S2 :- seq of runs above & below mean

S3 :- Count no. of obs above mean $(n_1)$ & below mean $(n_2)$ & b = Total no. of runs.

S4 :- Compare mean & variance of b,

$$\mu_b = \frac{2n_1 . n_2}{N} + \frac{1}{2}$$

$$\sigma_b^2 = \frac{2n_1 n_2 (2n_1 n_2 - N)}{N^2 (N-1)}$$

S5 :- Compute std normal

$$Z_0 = \frac{b - \mu_b}{\sigma_b}, \quad Z_0 \sim N[0,1]$$

S6 :- same as RURD

S7 :- Same as RURD

S1 :- Define Hypothesis.

S2 :- for decimal no', take mean 0.495 & for whole no's take mean 49.5.

Now, we take mean as 0.495.

If no's are < 0.495 put '−' else '+'

$\therefore \underbrace{- \quad - \quad - \quad - \quad -}_{1} \quad \underbrace{+}_{2} \quad \underbrace{-}_{3} \quad \underbrace{+ \quad +}_{4}$

$b = 4$.

$n_1 = 3 \quad n_2 = 6$

S3 :- $\mu_b = \frac{2n_1 . n_2}{N} + \frac{1}{2} = \frac{2.3.6}{9} + \frac{1}{2} = 4.5$.



$\sigma_b^2 = \frac{2n_1 . n_2 (2n_1 n_2 - N)}{N^2(N-1)}$

$= \frac{2 \times 3 \times 6 (2 \times 3 \times 6 - 9)}{9^2 (9-1)} = 1.5$.

S4 : $z_0 = \frac{b - \mu_b}{\sigma_b} = \frac{4 - 4.5}{\sqrt{1.5}} = -0.4$.

S5 − $-z_{\alpha/2} \le z_0 \le z_{\alpha/2}$

$-1.96 \le -0.4 \le 1.96$

# Mod-3

# Mod-4

# Binomial

A production process manufactures computer chips on the average at 2% nonconforming. Every day, a random sample of size 50 is taken from the process. If the sample contains more than two nonconforming chips, the process will be stopped. Compute the probability that the process is stopped by the sampling scheme.

Consider the sampling process as $n = 50$ Bernoulli trials, each with $p = 0.02$; then the total number of nonconforming chips in the sample, $X$, would have a binomial distribution given by

$$p(x) = \begin{cases} \binom{50}{x}(0.02)^x(0.98)^{50-x}, & x = 0, 1, 2, \dots, 50 \\ 0, & \text{otherwise} \end{cases}$$

It is much easier to compute the right-hand side of the following identity to compute the probability that more than two nonconforming chips are found in a sample:

$$P(X > 2) = 1 - P(X \le 2)$$

The probability $P(X \le 2)$ is calculated from

$$\begin{aligned} P(X \le 2) &= \sum_{x=0}^{2} \binom{50}{x}(0.02)^x(0.98)^{50-x} \\ &= (0.98)^{50} + 50(0.02)(0.98)^{49} + 1225(0.02)^2(0.98)^{48} \\ &= 0.92 \end{aligned}$$

Thus, the probability that the production process is stopped on any day, based on the sampling process, is approximately 0.08. The mean number of nonconforming chips in a random sample of size 50 is given by

$$E(X) = np = 50(0.02) = 1$$

and the variance is given by

$$V(X) = npq = 50(0.02)(0.98) = 0.98$$

The cdf for the binomial distribution has been tabulated by Banks and Heikes [1984] and others. The tables decrease the effort considerably for computing probabilities such as $P(a < X \le b)$. Under certain conditions on $n$ and $p$, both the Poisson distribution and the normal distribution may be used to approximate the binomial distribution [Hines and Montgomery, 1990].

Certainly! Let's explore random sampling and shuffling in detail:

### 1. Random Sampling:

**Definition**: Random sampling is a method used to select a subset of individuals or items from a larger population in such a way that each member of the population has an equal chance of being selected. It is widely used in statistics, research, and data analysis to obtain representative samples for analysis.

**Techniques**:

- **Simple Random Sampling**: Each member of the population has an equal probability of being selected, and every possible sample of a given size has the same probability of being chosen.
- **Stratified Sampling**: The population is divided into homogeneous subgroups (strata), and samples are then randomly selected from each subgroup.
- **Cluster Sampling**: The population is divided into clusters, and a random sample of clusters is selected. All individuals within the selected clusters are then included in the sample.
- **Systematic Sampling**: Members of the population are selected at regular intervals, using a fixed sampling interval.

**Applications**:

- Opinion polls and surveys
- Market research
- Quality control in manufacturing
- Medical studies and clinical trials
- Environmental studies

**Benefits**:

- Representative samples: Random sampling helps ensure that the selected sample is representative of the population.
- Unbiased estimates: Random sampling reduces the risk of bias in estimates and results.

- Statistical validity: Random sampling provides a basis for making statistical inferences about the population.

### 2. Shuffling:

**Definition**: Shuffling is a process of randomizing the order of items in a collection. It is commonly used in various applications such as gaming, data analysis, and cryptography.

**Techniques**:

- **Fisher-Yates Shuffle (Knuth Shuffle)**: A popular algorithm for shuffling a sequence of items. It iteratively selects a random element from the unshuffled portion of the list and swaps it with the last unselected element.

- **Random Permutation**: Generating a random permutation of the items in the collection using a random number generator.

- **Cryptographic Shuffling**: Shuffling methods designed to ensure randomness and unpredictability, often used in cryptographic applications.

**Applications**:

- Randomizing decks of cards in card games

- Shuffling playlists or recommendations in music and video streaming services

- Data shuffling in machine learning and data analysis to ensure randomness and prevent bias

- Shuffling cryptographic keys and data to enhance security and privacy

**Benefits**:

- Randomness: Shuffling ensures that the order of items is unpredictable and random.

- Fairness: In gaming and competitive scenarios, shuffling ensures fairness by preventing players from predicting the order of items.

- Security: In cryptography, shuffling enhances security by introducing randomness and preventing attacks based on predictable patterns.

In summary, random sampling and shuffling are important techniques used in various fields for obtaining representative samples, randomizing data, and ensuring fairness and security. Understanding these techniques and their applications is crucial for conducting unbiased research, analysis, and decision-making.

**Techniques**:

- **Simple Random Sampling**: Each member of the population has an equal probability of being selected, and every possible sample of a given size has the same probability of being chosen.
- **Stratified Sampling**: The population is divided into homogeneous subgroups (strata), and samples are then randomly selected from each subgroup.
- **Cluster Sampling**: The population is divided into clusters, and a random sample of clusters is selected. All individuals within the selected clusters are then included in the sample.
- **Systematic Sampling**: Members of the population are selected at regular intervals, using a fixed sampling interval.

**Applications**:

- Opinion polls and surveys
- Market research
- Quality control in manufacturing
- Medical studies and clinical trials
- Environmental studies

## 2. Finite-State Markov Chains:

**Definition**: A Finite-State Markov Chain is a stochastic process with a finite number of possible states, where transitions between states occur based on certain probabilities or rates. The system evolves over discrete time steps according to the Markov property, meaning that future states depend only on the current state and are independent of past states.

**Key Characteristics**:

- **State Space**: The state space consists of a finite set of distinct states, denoted by $S = \{S_1, S_2, ..., S_n\}$.
- **Transition Probabilities**: Transitions between states are governed by transition probabilities $P_{ij}$, representing the probability of transitioning from state $S_i$ to state $S_j$ in one time step.
- **Markov Property**: The future state of the system depends only on the current state and is independent of the past history of states.
- **Stationary Distribution**: Under certain conditions, the process may converge to a stationary distribution where the probabilities of being in different states remain constant over time.

**Applications**:

- **Modeling Random Processes**: Finite-State Markov Chains are used to model various random

## Algorithm for Random Variate Generation using Inverse Transform Technique:

1. **Define the Probability Distribution**: Start by defining the probability distribution function (PDF) of the desired random variable.
2. **Calculate the Cumulative Distribution Function (CDF)**: Compute the cumulative distribution function (CDF) of the probability distribution. This function represents the cumulative probabilities up to each value of the random variable.
3. **Find the Inverse CDF**: Calculate the inverse of the cumulative distribution function (CDF). This function maps probabilities to corresponding values of the random variable.
4. **Generate Uniform Random Numbers**: Generate random numbers $U$ from a uniform distribution in the range [0, 1]. These random numbers will serve as probabilities.
5. **Transform Uniform Random Numbers**: Apply the inverse CDF to the generated uniform random numbers $U$ to obtain random variates from the desired distribution. This is done by evaluating the inverse CDF function at each $U$ value.

Let's illustrate the algorithm with an example of generating random variates from an exponential distribution:

1. **Define the Exponential Distribution**: Assume we want to generate random variates from an exponential distribution with parameter $\lambda$.
2. **Calculate the CDF**: The cumulative distribution function (CDF) of the exponential distribution is given by:

$$F(x) = 1 - e^{-\lambda x}$$

3. **Find the Inverse CDF**: Solve the equation $U = 1 - e^{-\lambda x}$ for $x$ to obtain the inverse CDF:

$$x = -\frac{1}{\lambda}\ln(1-U)$$

4. **Generate Uniform Random Numbers**: Generate $n$ uniform random numbers $U_1, U_2, \ldots, U_n$ from a uniform distribution in the range [0, 1].
5. **Transform Uniform Random Numbers**: Apply the inverse CDF to each uniform random number $U_i$ to obtain the corresponding random variates:

$$X_i = -\frac{1}{\lambda}\ln(1-U_i)$$

**Acceptance/Rejection Technique**

The following procedure describe **Acceptance/Rejection Technique for generating a Poisson random variates.**

**Step 1.** Set $n = 0$, $P = 1$.

**Step 2.** Generate a random number $R_{n+1}$, and replace $P$ by $P \cdot R_{n+1}$.

**Step 3.** If $P < e^{-\alpha}$, then accept $N = n$. Otherwise, reject the current $n$, increase $n$ by one, and return to step 2.

**N = n, says there were exactly n arrivals during one unit of time and P is Probability**

# Mod-5

- Input models are the distributions of time b/w arrivals and of service times.
- four steps in the development of a useful model of input data
  1. collect the data from the real system of interest
  2. identify the probability distribution to represent the input process.
  3. choose the parameters that determine a specific instance of the distribution family
  4. evaluate the chosen distribution and associated parameters for goodness of fit test.

## 1. DATA COLLECTION

- Collect data from the real system of interest
- Requires substantial amount of time and resource commitment
- When data is not available, expert opinion and knowledge of the process must be used to make educated guesses
- Even though if the model structure is valid, if the input data is inaccurately collected, inappropriately analysed then simulation output will be misleading.

# SUGGESTIONS TO ENHANCE DATA COLLECTION

1. A useful expenditure of time is in planning which could begin by practice or pre observing session
2. Try to analyse the data as they are being collected
3. Try to combine homogeneous data sets
4. Be aware of the possibility of data censoring
5. Discover the relationship between the two variables by using scatter diagram
6. Check for autocorrelation of data collected from the customers
7. Difference between input data and output or performance data must be given importance

# 2. IDENTIFYING THE DISTRIBUTION WITH DATA

- Shape of distribution is identified by
  - Frequency distribution
  - Histograms

# STEPS INVOLVED IN CONSTRUCTION OF HISTOGRAM

1. Divide the range of the data into interval
2. Label the horizontal axis to conform to the intervals selected
3. Find the frequency of occurrences within each interval
4. Label the vertical axis so that the total occurrences can be plotted for each interval
5. Plot the frequencies on the vertical axis

- The no .of intervals depends on the no. of observations & on the amount of scatter or dispersion in the data
- Histogram should not be too ragged or to coarse as shown

# 2.2. SELECTING THE FAMILY OF DISTRIBUTION

- Purpose of preparing a histogram is to infer a known pmf or pdf
- Family of distribution is chosen based on what might arise in the context being investigated along with the shape of the histogram
- There are many probability distributions created , few are

- Binomial :
  - Models the no. of successes in n trails , where the trails are independent with common success probability p
  - Eg: number of defective chips out of n chips
- Poisson:
  - Models the number of independent events that occur in a fixed amount of time or space
  - Eg : the number of customer that arrive to a store in one hour
- Normal:
  - Models the distribution of a process that can be thought of as the sum of a number of component process
  - Eg: time assemble a product that is the sum of the times required for each assembly operation

- Exponential:
  - Models the time between independent event
  - E.g.: time between the arrivals
- Gamma:
  - Models non-negative random variables and it is flexible distribution
- Beta
  - Extremely flexible distribution used to model bounded random variable
- Weibull
  - Models the time to failure for compounds
- Discrete or continuous uniform
  - Models complete uncertainty
- Triangular
  - Models a process which only min most likely and max values of the distribution known

# 3 PARAMETER ESTIMATION

- After the selection of family of distributions , the next step is to estimate the parameters of the distribution.

**Preliminary statistics: sample mean and sample variance**

- Sample mean and variance will be calculated depending on the type of data whether discrete or continuous .

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

sample variance, $S^2$, is defined by

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

# SUGGESTED ESTIMATORS:

- Numerical estimates of the distribution parameters are needed to reduce the family of distributions to a specific distribution

- These estimators are the likelihood estimators based on the raw data.

# 4 GOODNESS-OF-FIT TESTS

- Apply the different types of tests for goodness based on the family of distribution selected.
- Use the corresponding estimators based on the family of distribution and verify for the goodness of fit.
- General tests can be applied are
  1. Chi-square test
  2. Kolmogorov-Smirnov test

## GOODNESS-OF-FIT TESTS

- **Chi-square test**: the test procedure starts by arranging 'n' observations into k class intervals or cells.
- The test statistic is given by $X_0^2 = \Sigma (O_i - E_i)^2 / E_i$, where $O_i$ is the observed frequency at $i^{th}$ interval & $E_i$ is the expected frequency in that class interval.
- The expected frequency for each class interval can be computed as $E_i = nP_i$, where, Pi is the theoretical hypothesized probability associated with the $i^{th}$ class

# FITTING A NON STATIONARY POISON PROCESS (NSPP)

- Approaches used are
  - Choose a very flexible model with lots of parameters and fit it with a method such as maximum likelihood
  - Approximate the arrival rate as being constant over some basic interval of time , such as an hour, or a day or a month best by carrying from time interval to time interval

# MULTIVARIATE AND TIME SERVICE INPUT MODELS

- If we have finite numbers of random variables then we have multivariate model
- If we have infinite number of random variable then we have time series model

# COVARIANCE AND CORRELATION

Covariance $\quad Cov(x, y) = \frac{1}{n-1}\left[\Sigma xy - n\bar{x}\bar{y}\right]$

Correlation $\quad \rho = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y}$

where $\sigma_x = \sqrt{\frac{\Sigma x^2 - n\bar{x}^2}{n-1}}$ and $\sigma_y = \sqrt{\frac{\Sigma y^2 - n\bar{y}^2}{n-1}}$

# TIME SERIES INPUT MODEL

- If $X_1, X_2, X_3, \ldots$ is a sequence of identically distributed but dependent and covariance stationary random variables there are number of time series models that can be used to represent the process
- Two models that have the characteristics that auto correlation take the form

$$\rho_h = \mathrm{corr}(X_t, X_{t+h}) = \rho^h$$

- AR(1) model
- EAR (1) model

# Difference between verification and validation

| Verification | Validation |
| --- | --- |
| • Verification is concerned with building the model right. | Validation is concerned with building the right model. |
| • It is utilized in comparison of the conceptual model to the computer representation that implements that conception. | • It is utilized to determine that a model is an accurate representation of the real system. It is usually achieved through the calibration of the model. |
| • The purpose of model verification is to assure that the conceptual model is reflected accurately in the operational model. | • Validation is the overall process of comparing the model and its behavior to the real system and its modeler. |