unit 6: INPUT MODELING

6. INPUT MODELING

- Input data provide the driving force for a simulation model. In the simulation of a queuing system, typical input data are the distributions of time between arrivals and service times.
- For the simulation of a reliability system, the distribution of time-to-failure of a component is an example of input data.

There are four steps in the development of a useful model of input data:

- Collect data from the real system of interest. This often requires a substantial time and resource commitment. Unfortunately, in some situations it is not possible to collect data
- Identify a probability distribution to represent the input process. When data are available, this step typically begins by developing a frequency distribution, or histogram, of the data.
- Choose parameters that determine a specific instance of the distribution family. When data are available, these parameters may be estimated from the data.
- Evaluate the chosen distribution and the associated parameters for good-of-fit. Goodness-of-fit may be evaluated informally via graphical methods, or formally via statistical tests. The chisquare and the Kolmo-gorov-Smirnov tests are standard goodness-of-fit tests. If not satisfied that the chosen distribution is a good approximation of the data, then the analyst returns to the second step, chooses a different family of distributions, and repeats the procedure. If several iterations of this procedure fail to yield a fit between an assumed distributional form and the collected data

6.1 Data Collection

- Data collection is one of the biggest tasks in solving real problem. It is one of the most important and difficult problems in simulation. And even if when data are available, they have rarely been recorded in a form that is directly useful for simulation input modeling.
The following suggestions may enhance and facilitate data collection, although they are not all – inclusive.

1. A useful expenditure of time is in planning. This could begin by a practice or pre-observing session. Try to collect data while pre-observing.

2. Try to analyze the data as they are being collected. Determine if any data being collected are useless to the simulation. There is no need to collect superfluous data.

3. Try to combine homogeneous data sets. Check data for homogeneity in successive time periods and during the same time period on successive days.

4. Be aware of the possibility of data censoring, in which a quantity of interest is not observed in its entirety. This problem most often occurs when the analyst is interested in the time required to complete some process (for example, produce a part, treat a patient, or have a component fail), but the process begins prior to, or finishes after the completion of, the observation period.

5. To determine whether there is a relationship between two variables, build a scatter diagram.

6. Consider the possibility that a sequence of observations which appear to be independent may possess autocorrelation. Autocorrelation may exist in successive time periods or for successive customers.

7. Keep in mind the difference between input data and output or performance data, and be sure to collect input data. Input data typically represent the uncertain quantities that are largely beyond the control of the system and will not be altered by changes made to improve the system.

6.2 Identifying the Distribution with Data.

- In this section we discuss methods for selecting families of input distributions when data are available.

6.2.1 Histogram

- A frequency distribution or histogram is useful in identifying the shape of a distribution. A histogram is constructed as follows:
  1. Divide the range of the data into intervals (intervals are usually of equal width;
however, unequal widths however, unequal width may be used if the heights of the frequencies are adjusted).

2. Label the horizontal axis to conform to the intervals selected.
3. Determine 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.

- If the intervals are too wide, the histogram will be coarse, or blocky, and its shape and other details will not show well. If the intervals are too narrow, the histogram will be ragged and will not smooth the data.
- The histogram for continuous data corresponds to the probability density function of a theoretical distribution.

Example 6.2: The number of vehicles arriving at the northwest corner of an intersection in a 5 min period between 7 A.M. and 7:05 A.M. was monitored for five workdays over a 20-week period. Table shows the resulting data. The first entry in the table indicates that there were 12:5 min periods during which zero vehicles arrived, 10 periods during which one vehicles arrived, and so on.

Table 6:1 Number of Arrivals in a 5 Minute period

<table>
<thead>
<tr>
<th>Arrivals Per period</th>
<th>Frequency</th>
<th>Arrivals Per Period</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>
6.2.2 Selecting the Family of Distributions

- Additionally, the shapes of these distributions were displayed. The purpose of preparing histogram is to infer a known pdf or pmf. A family of distributions is selected on the basis of what might arise in the context being investigated along with the shape of the histogram.

- Thus, if interarrival-time data have been collected, and the histogram has a shape similar to the pdf in Figure 5.9, the assumption of an exponential distribution would be warranted.

- Similarly, if measurements of weights of pallets of freight are being made, and the histogram appears symmetric about the mean with a shape like that shown in Fig 5.12, the assumption of a normal distribution would be warranted.

- The exponential, normal, and Poisson distributions are frequently encountered and are not difficult to analyze from a computational standpoint. Although more difficult to analyze, the gamma and Weibull distributions provide array of shapes, and should not be overlooked when modeling an underlying probabilistic process. Perhaps an exponential
distribution was assumed, but it was found not to fit the data. The next step would be to examine where the lack of fit occurred.

- If the lack of fit was in one of the tails of the distribution, perhaps a gamma or Weibull distribution would more adequately fit the data.

- Literally hundreds of probability distributions have been created, many with some specific physical process in mind. One aid to selecting distributions is to use the physical basis of the distributions as a guide. Here are some examples:

6.2.3 Quantile-Quantile Plots

- Further, our perception of the fit depends on widths of the histogram intervals. But even if the intervals are well chosen, grouping of data into cells makes it difficult to compare a histogram to a continues probability density function

- If X is a random variable with cdf F, then the q-quintile of X is that y such that F(y) = P(X < y) = q, for 0 < q < 1. When F has an inverser, we write y = F\(^{-1}\)(q).

- Now let \{Xi, i = 1, 2, ..., n\} be a sample of data from X. Order the observations from the smallest to the largest, and denote these as \{yj, j = 1, 2, ..., n\}, where y1 < y2 < ... < yn. Let j denote the ranking or order number. Therefore, j = 1 for the smallest and j = n for the largest. The q-q plot is based on the fact that y1 is an estimate of the \((j - 1/2)/n\) quantile of X other words,

\[
Y_j \text{ is approximately } F^{-1} \left( \frac{j - 1/2}{n} \right)
\]

- Now suppose that we have chosen a distribution with cdf F as a possible representation of the distribution of X. If F is a member of an appropriate family of distributions, then a plot of yj versus F\(^{-1}\)((j - 1/2)/n) will be approximately a straight line.
6.3 Parameter Estimation

- After a family of distributions has been selected, the next step is to estimate the parameters of the distribution. Estimators for many useful distributions are described in this section. In addition, many software packages—some of them integrated into simulation languages—are now available to compute these estimates.

6.3.1 Preliminary Statistics: Sample Mean and Sample Variance

- In a number of instances the sample mean, or the sample mean and sample variance, are used to estimate of the parameters of hypothesized distribution;

- If the observations in a sample of size n are X1, X2, ..., Xn, the sample mean (X) is defined by

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

9.1

and the sample variance, \( s^2 \) is defined by

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

9.2

If the data are discrete and grouped in frequency distribution, Equation (9.1) and (9.2) can be modified to provide for much greater computational efficiency. The sample mean can be computed by

\[
\bar{X} = \frac{\sum_{j=1}^{n} f_j X_j}{n}
\]

9.3
And the sample variance by

\[ X^2 = \frac{\sum_{j=1}^{k} f_j X_j^2 - n \overline{X}^2}{n - 1} \]

where \( k \) is the number of distinct values of \( X \) and \( f_j \) is the observed frequency of the value \( X_j \) of \( X \).

### 6.3.2 Suggested Estimators

- Numerical estimates of the distribution parameters are needed to reduce the family of distributions to a specific distribution and to test the resulting hypothesis.
- These estimators are the maximum-likelihood estimators based on the raw data. (If the data are in class intervals, these estimators must be modified.)
- The triangular distribution is usually employed when no data are available, with the parameters obtained from educated guesses for the minimum, most likely, and maximum possible value's; the uniform distribution may also be used in this way if only minimum and maximum values are available.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameter</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>( \alpha )</td>
<td>( \hat{\alpha} = \overline{X} )</td>
</tr>
<tr>
<td>Exponential</td>
<td>( \lambda )</td>
<td>( \hat{\lambda} = \frac{1}{\overline{X}} )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( \beta, \theta )</td>
<td>( \hat{\theta} = \frac{1}{\overline{X}} )</td>
</tr>
<tr>
<td>Normal</td>
<td>( \mu, \sigma^2 )</td>
<td>( \hat{\mu} = \overline{X}, \hat{\sigma^2} = S^2 )</td>
</tr>
<tr>
<td>Lognormal</td>
<td>( \mu, \sigma^2 )</td>
<td>( \hat{\mu} = \overline{X}, \hat{\sigma^2} = S^2 )</td>
</tr>
</tbody>
</table>
6.4 Goodness-of-Fit Tests

- These two tests are applied in this section to hypotheses about distributional forms of input data. Goodness-of-fit tests provide help full guidance for evaluating the suitability of a potential input model.
- However, since there is no single correct distribution in a real application, you should not be a slave to the verdict of such tests.
- It is especially important to understand the effect of sample size. If very little data are available, then a goodness-of-fit test is unlikely to reject any candidate distribution; but if a lot of data are available, then a goodness-of-fit test will likely reject all candidate distribution.

6.4.1 Chi-Square Test

- One procedure for testing the hypothesis that a random sample of size n of the random variable X follows a specific distributional form is the chi-square goodness-of-fit test.
- This test formalizes the intuitive idea of comparing the histogram of the data to the shape of the candidate density or mass function. The test is valid for large sample sizes, for both discrete and continuous distribution assumptions. When parameters are estimated by maximum likelihood.

\[
X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}
\]

- where 0, is the observed frequency in the ith class interval and Ei, is the expected frequency in that class interval. The expected frequency for each class interval is computed as Ei=npi, where pf is the theoretical, hypothesized probability associated with the ith class interval.
- It can be shown thatX02 approximately follows the chi-square distribution with k-s-1 degrees of freedom, where s represents the number of parameters of the hypothesized distribution estimated by sample statistics. The hypotheses are:
H0: the random variable, X, conforms to the distributional assumption with the parameter(s) given by the parameter estimate(s)

H1: the random variable X does not conform

- If the distribution being tested is discrete, each value of the random variable should be a class interval, unless it is necessary to combine adjacent class intervals to meet the minimum expected cell-frequency requirement. For the discrete case, if combining adjacent cells is not required,

\[ P_i = P(X_i) = P(X \geq X_i) \]

Otherwise, \( P_i \) is determined by summing the probabilities of appropriate adjacent cells.

- If the distribution being tested is continuous, the class intervals are given by \([a_{i-1}, a_i)\), where \(a_{i-1}\) and \(a_i\) are the endpoints of the \(i\)th class interval. For the continuous case with assumed pdf \(f(x)\), or assumed cdf \(F(x)\), \( P_i \) can be computed by

\[ P_i = \int_{a_{i-1}}^{a_i} f(x) \, dx = F(a_i) - F(a_{i-1}) \]

### 6.4.2 Chi-Square Test with Equal Probabilities

- If a continuous distributional assumption is being tested, class intervals that are equal in probability rather than equal in width of interval should be used.
- Unfortunately, there is as yet no method for determining the probability associated with each interval that maximize the power of a test of a given size.

\[ E_i = n \cdot p_i \]

- Substituting for \(p_i\) yields \(n/k\)
- and solving for \(k\) yields \(k \approx n/5\)
6.4.3 Kolmogorov - Smirnov Goodness-of-Fit Test

- The chi-square goodness-of-fit test can accommodate the estimation of parameters from the data with a resultant decrease in the degrees of freedom (one for J each parameter estimated). The chi-square test requires that the data be placed in class intervals, and in the case of continues distributional assumption, this grouping is arbitrary.

- Also, the distribution of the chi-square test statistic is known only approximately, and the power of the test is sometimes rather low. As a result of these considerations, goodness-of-fit tests, other than the chi-square, are desired.

- The Kolmogorov-Smirnov test is particularly useful when sample sizes are small and when no parameters have been estimated from the data.

- (Kolmogoro-Smirnov Test for Exponential Distribution)

Ho : the interarrival times are exponentially distributed
H1: the interarrival times are not exponentially distributed

- The data were collected over the interval 0 to T = 100 min. It can be shown that if the underlying distribution of interarrival times { T1, T2, … } is exponential, the arrival times are uniformly distributed on the interval (0,T).
• The arrival times $T_1$, $T_1+T_2$, $T_1+T_2+T_3$,……,$T_1+…..+T_{50}$ are obtained by adding interarrival times.
• On a $(0,1)$ interval, the points will be $[T_1/T, (T_1+T_2)/T,……,(T_1+…..+T_{50})/T]$.

6.5 Selecting Input Models without Data

Unfortunately, it is often necessary in practice to develop a simulation model for demonstration purposes or a preliminary study—before any data are available.) In this case the modeler must be resourceful in choosing input models and must carefully check the sensitivity of results to the models.

**Engineering data** : Often a product or process has performance ratings provided by the manufacturer.

**Expert option** : Talk to people who are experienced with the processes or similar processes. Often they can provide optimistic, pessimistic and most likely times.

**Physical or conventional limitations** : Most real processes have physical limits on performance. Because of company policies, there may be upper limits on how long a process may take. Do not ignore obvious limits or bound: that narrow the range of the input process.

**The nature of the process** It can be used to justify a particular choice even when no data are available.

6.6 Multivariate and Time-Series Input Models
The random variables presented were considered to be independent of any other variables within the context of the problem. However, variables may be related, and if the variables appear in a simulation model as inputs, the relationship should be determined and taken into consideration.

**Step 1.** Generate $Z_1$ and $Z_2$, independent standard normal random variables.

**Step 2.** Set $X_1 = \mu_1 + \sigma_1 Z_1$

**Step 3.** Set $X_2 = \mu_2 + \sigma_2 \left( \rho Z_1 + \sqrt{1-\rho^2} Z_2 \right)$
6.7 Time series input model:
If $X_1, X_2, \ldots, X_n$ is a sequence of identically distributed, but dependent and covariance stationary random variables, then there are a number of times series models that can be used to represent the process. The two models that have the characteristics that the autocorrelation take the form:

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

for $h=1,2,\ldots,n$ that the log-h autocorrelation decreases geometrically as the lag increases.

**AR(1) Model:**
consider the time series model

$$X_t = \mu + \phi(X_{t-1} - \mu) + \epsilon_t$$

for $t=2,3,\ldots,n$ where $\epsilon_2, \epsilon_3$ are the independent and identically distributed with mean 0 and variance $\sigma^2_\epsilon$ and $-1 < \phi < 1$. If the initial value $x_1$ is chosen appropriately, then $x_1, x_2, \ldots$ are all normal distributed with mean $\mu$ and variance $\sigma^2_\epsilon/(1-\phi^2)$.

1. Generate $X_1$ from the normal distribution with mean $\mu$ and variance $\sigma^2_\epsilon/(1-\phi^2)$. Set $t = 2$.
2. Generate $\epsilon_t$ from the normal distribution with mean 0 and variance $\sigma^2_\epsilon$.
3. Set $X_t = \mu + \phi(X_{t-1} - \mu) + \epsilon_t$.
4. Set $t = t + 1$ and go to Step 2.

**EAR(1) Model:**
Consider the time series model

$$X_t = \begin{cases} \phi X_{t-1}, & \text{with probability } \phi \\ \phi X_{t-1} + \epsilon_t, & \text{with probability } 1-\phi \end{cases}$$

for $t=2,3,\ldots,n$ where $\epsilon_2, \epsilon_3$ are the independent and identically distributed with mean $1/\lambda$ and $0 < \phi < 1$. If the initial value $x_1$ is chosen appropriately, then $x_1, x_2, \ldots$ are all exponentially distributed with mean $1/\lambda$ and variance $\sigma^2_\epsilon/(1-\phi^2)$. 
Step 1. Generate $X_1$ from the exponential distribution with mean $1/\lambda$. Set $t = 2$.

Step 2. Generate $U$ from the uniform distribution on $[0, 1]$. If $U \leq \phi$, then set

$$X_t = \phi X_{t-1}$$

Otherwise, generate $e_t$ from the exponential distribution with mean $1/\lambda$ and set

$$X_t = \phi X_{t-1} + e_t$$

Step 3. Set $t = t + 1$ and go to Step 2.
Goodness-of-Fit Tests

0 Chi-Square Test with Poisson Assumption

Step 1: Compute Poisson distribution using

\[ P(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \]  \( x = 0, 1, 2, \ldots \)

Step 2: Compute expected frequency

\[ E_i = n \cdot P_i \]  where \( n \) is sum of sample data

Reduce interval i.e.

\[ E_i > 5 \]

Step 3: Compute Chi-Square test i.e.

\[ \chi^2 = \sum_{i=0}^{n} \left( \frac{(O_i - E_i)^2}{E_i} \right) \]

Step 4: Obtain Chi-Square test value from Table A.6

\[ \chi^2_{0.05, K-S-1} \]

Step 5: Check hypothesis or Null hypothesis

\[ \chi^2 < \chi^2_{0.05, K-S-1} \text{ Accepted / Rejected} \]

Null hypothesis
Chi-Square test for Exponential distribution (Equal probability)

Step 1: Determine the probability

\[ P = \frac{1}{k} \]  
where \( k \) is interval

Step 2: Determine the mean

\[ \lambda = \frac{1}{k} \]

\[ \bar{x} = \frac{\sum_{i=0}^{n} x_i}{n} \]

Step 3: Compute class interval

\[ a_i = -\frac{1}{\lambda} \ln(1 - ip) \]

\( i = 0, 1, 2, \ldots, k \)

Step 4: Compute expected frequency

\[ E_i = \frac{N}{k} \]

\( N \) - Sum of Sample data

\( k \) - interval

Step 5: Compute Chi-Square test

\[ \chi^2 = \frac{\sum_{i=0}^{k} (O_i - E_i)^2}{E_i} \]

Step 6: Obtain Chi-Square test value from table A.6

\[ \chi^2_{\alpha, k-\epsilon-1} \]

Step 7: Check hypothesis or null hypothesis

\[ \chi^2 < \chi^2_{\alpha, k-\epsilon-1} \]  
accepted
Kolmogorov-Smirnov test for exponential distribution

Step 1: Calculate interarrival points

\[ R_i = \sum_{j=1}^{i} \frac{T_j}{T}, \quad \frac{(T_1+T_2)/T_1}{T}, \quad \frac{(T_1+T_2+T_3)/T_1}{T} \]

- \( T \) is total number of sample data
- \( T_i \) is the sample data

Step 2: Compute

\[
D^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - R_1 \right\}
\]

\[
D^- = \max_{1 \leq i \leq n} \left\{ R_i - \frac{i-1}{n} \right\}
\]

Step 3: Compute

\[
D = \max(D^+, D^-)
\]

Step 4: Obtain KS-test value from table A.8

\[ D_{\alpha, n} \]

Step 5: Check hypothesis or null hypothesis

\[ D \leq D_{\alpha, n}^{\alpha} \text{ accepted} \]
UNIT-6: INPUT MODELLING

I. Chi Square Test using Poisson Assumption

1. Using goodness of fit test, test whether random Nos. are uniformly distributed based on poisson assumption with level of significance \( \alpha = 0.05 \).

\( \chi^2 = 3.64 \). Sample data are:

Interval : 0 1 2 3 4 5 6 7 8 9 10 11
Observed : 12 10 19 17 10 8 7 5 5 3 3 1

Frequency

Given:
\( \alpha = 0.05 \)
\( \chi^2 = 3.64 \)

\( N = 12 + 10 + 19 + 17 + 10 + 8 + 7 + 5 + 5 + 3 + 3 + 1 = 100 \)

Step 1: Compute Poisson Distribution

\( P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \) where \( x = 0, 1, \ldots, 11 \)

\( P(0) = \frac{e^{-3.64} \times (3.64)^0}{0!} = 0.026 \)

\( P(1) = 0.096 \)
\( P(2) = 0.174 \)
\( P(3) = 0.211 \)
\( P(4) = 0.192 \)
\( P(5) = 0.140 \)
\( P(6) = 0.085 \)
\( P(7) = 0.044 \)
\( P(8) = 0.020 \)
Step 2: Apply Chi Square with poisson assumption.

<table>
<thead>
<tr>
<th>$X_i$</th>
<th>$O_i$</th>
<th>$E_i = n \cdot P_i$</th>
<th>$O_i - E_i$</th>
<th>$(O_i - E_i)^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12</td>
<td>2.6</td>
<td>9.8</td>
<td>96.04</td>
<td>7.87</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>9.6</td>
<td>1.6</td>
<td>2.56</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>17.4</td>
<td>-4.1</td>
<td>16.81</td>
<td>0.79</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>21.1</td>
<td>-9.2</td>
<td>84.64</td>
<td>4.41</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>19.2</td>
<td>-9.2</td>
<td>84.64</td>
<td>4.41</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>14.8</td>
<td>-6</td>
<td>36</td>
<td>2.57</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8.5</td>
<td>-1.5</td>
<td>2.25</td>
<td>0.27</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>4.4</td>
<td>9.4</td>
<td>88.36</td>
<td>11.63</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>2.0</td>
<td>3.6</td>
<td>12.96</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0.8</td>
<td>2.2</td>
<td>4.84</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0.3</td>
<td>2.7</td>
<td>7.29</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0.1</td>
<td>0.9</td>
<td>0.81</td>
<td></td>
</tr>
</tbody>
</table>

We have $k = 7$, $s = 1$, $\chi^2 = 27.69$

Step 3: Compute level of Significance from Tabu A6

$\chi^2, k \geq k - s - 1 = \chi^2, 7 - 1 - 1$

$= \chi^2, 0.05, 5 = 11.1$

Step 4: Check whether Random Nos are uniformly distributed.

Compare $\chi^2$ & $\chi^2, 0.05, 5$

$\therefore 27.69 > 11.1 = \) Random Nos are not uniformly distributed.
2. Using goodness of fit test, check whether Random Nos are uniformly distributed over interval [0, 1] using poisson assumption with level of significance = 0.05. Simulation table for critical values is given:

Interval (Xi) : 0 1 2 3 4 5 6 7
Frequency (fi) : 5 10 5 8 12 10 8 12

Given: $\alpha = 0.05$

$n = 5 + 10 + 5 + 8 + 12 + 10 + 8 + 12 = 70$

$\bar{X} = ?$

$\bar{X} = \frac{\sum_{i=1}^{n} fi \cdot Xi}{n} = \frac{0 + 10 + 10 + 24 + 48 + 50 + 48 + 84}{70} = \frac{274}{70} = 3.91$

Step 1: Compute Poisson Distribution

$p(x) = \frac{e^{-x} \cdot x^x}{x!}, x = 0, 1, \ldots, 7, \& \alpha = 3.91$

$p(0) = 0.020$
$p(1) = 0.048$
$p(2) = 0.077$
$p(3) = 0.099$
$p(4) = 0.111$
$p(5) = 0.111$
$p(6) = 0.099$
$p(7) = 0.084$
Step 2: Apply Chi Square test with Poisson assumption

| \( x_i \) | \( o_i \) | \( E_i = np_i \) | \( O_i - E_i \) | \( (O_i - E_i)^2 \) | \( \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5 ( { 15 )</td>
<td>1.4 ( { 6.86 )</td>
<td>8.14</td>
<td>66.26</td>
</tr>
<tr>
<td>1</td>
<td>10 ( { 5.46 )</td>
<td>5.71</td>
<td>-5.71</td>
<td>32.60</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>10.71</td>
<td>-5.93</td>
<td>35.16</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>13.93</td>
<td>-5.93</td>
<td>35.16</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>13.65</td>
<td>-1.65</td>
<td>2.72</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10.71</td>
<td>-0.71</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>8 ( { 20 )</td>
<td>6.93 ( { 10.85 )</td>
<td>9.15</td>
<td>83.72</td>
</tr>
<tr>
<td>7</td>
<td>12 ( { 3.92 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Here \( k = 6 \), \( S = 1 \)

\[ \chi^2 = \frac{23.18}{9.66} \]

Step 3: Compute level of Significance from Table A6

\[ \chi^2, k-S-1 = \chi^2 0.05, 6-1-1 = 9.49 \]

Step 4: Check whether Random No.s are uniformly distributed.

Compare \( \chi^2 \) & \( \chi^2 0.05, 4 \)

\[ \chi^2 > 9.49 \Rightarrow \text{Random No.s are not uniformly distributed} \]
Apply goodness of fit test, check whether Random Nos are uniformly distributed over Interval [0, 1] with given size of data 100. Assume \( \alpha = 0.01 \). Simulation table to check critical value using Poisson assumption is given below:

<table>
<thead>
<tr>
<th>Interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>8</td>
<td>6</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>

Given: \( \alpha = 0.01 \) \( \alpha' = ? \)
\[ n = 100 \]
\[ \bar{X} = \frac{\sum f_i x_i}{n} = \frac{586}{100} = 5.86 \]

**Step 1:** Compute Poisson Distribution

\[ P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{where} \quad x = 0, 1, 2, \ldots, 10 \quad \text{and} \quad \lambda = 5.86 \]

\[ P(1) = 0.017 \]
\[ P(2) = 0.049 \]
\[ P(3) = 0.096 \]
\[ P(4) = 0.140 \]
\[ P(5) = 0.164 \]
\[ P(6) = 0.160 \]
\[ P(7) = 0.134 \]
\[ P(8) = 0.098 \]
\[ P(9) = 0.064 \]
\[ P(10) = 0.038 \]
**Step 2:** Apply Chi Square with Poisson Assumption

<table>
<thead>
<tr>
<th>(X_i)</th>
<th>(O_i)</th>
<th>(E_i = n \cdot P_i)</th>
<th>(O_i - E_i)</th>
<th>((O_i - E_i)^2)</th>
<th>(\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>7.7</td>
<td>7.4</td>
<td>54.96</td>
<td>8.29</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4.9</td>
<td>0.4</td>
<td>0.16</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>9.6</td>
<td>-0.4</td>
<td>0.16</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>14.0</td>
<td>-3</td>
<td>9</td>
<td>0.64</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>16.4</td>
<td>-4.4</td>
<td>19.36</td>
<td>1.18</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>16.0</td>
<td>-8</td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>13.4</td>
<td>-3.4</td>
<td>11.56</td>
<td>0.86</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>9.8</td>
<td>2.2</td>
<td>4.84</td>
<td>0.49</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>6.4</td>
<td>6.4</td>
<td>16.384</td>
<td>16.06</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(k = 8, \ \delta = 1\)

\(\chi^2 = 31.54\)

**Step 3:** Compute level of Significance from Table A6

\(\chi^2_{\alpha}, k - \delta - 1 = \chi^2_{0.01}, 8 - 1 - 1 = 20.1\)

**Step 4:** Check whether Random Nos are uniformly distributed

Compare \(\chi^2\) & \(\chi^2_{0.05}\).

\(31.54 > 20.1 \Rightarrow\) Random Nos are not uniformly distributed
Chi-Square Test with Equal Probability (Exponential Dist.)

1. Apply goodness of fit test to check whether random numbers are uniformly distributed over [0, 1] using equal probability. Use $\alpha = 0.05$, interval $k = 8$ to check whether given sample data are accepted or rejected.

4.562

Given: $k = 8$, $\alpha = 0.05$

**Step 1:** Compute mean

$$\bar{X} = \frac{1}{n} \sum X_i$$

$$\bar{X} = \frac{594.674}{50} = 11.894$$

$$\lambda = \frac{1}{11.894} = 0.084$$

**Step 2:** Compute class intervals

$$P = \frac{1}{k} = \frac{1}{8} = 0.125$$

$$a_i = -\frac{1}{\lambda} \ln \left(1 - i \times P\right) \text{ where } i = 0, \ldots, 8$$

$\lambda = 0.084$

$P = 0.125$
\[
\begin{align*}
\alpha &= 0.1589 \\
\alpha_1 &= 0.1589 \\
\alpha_2 &= 3.425 \\
\alpha_3 &= 5.595 \\
\alpha_4 &= 8.252 \\
\alpha_5 &= 11.677 \\
\alpha_6 &= 16.504 \\
\alpha_7 &= 24.755 \\
\alpha_8 &= \infty
\end{align*}
\]

**Step 3**: Compute Chi Square with equal probability

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Oi</th>
<th>Ei = \frac{N}{k} (50%)</th>
<th>Oi - Ei</th>
<th>(Oi - Ei)^2</th>
<th>\chi^2 = \frac{\sum (Oi - Ei)^2}{Ei}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 1.589</td>
<td>19</td>
<td>6.25</td>
<td>12.75</td>
<td>162.563</td>
<td>26.01</td>
</tr>
<tr>
<td>1.589 - 3.425</td>
<td>10</td>
<td>6.25</td>
<td>3.75</td>
<td>14.063</td>
<td>2.25</td>
</tr>
<tr>
<td>3.425 - 5.595</td>
<td>3</td>
<td>6.25</td>
<td>-3.75</td>
<td>10.563</td>
<td>1.69</td>
</tr>
<tr>
<td>5.595 - 8.252</td>
<td>6</td>
<td>6.25</td>
<td>-0.25</td>
<td>0.0625</td>
<td>0.01</td>
</tr>
<tr>
<td>8.252 - 11.677</td>
<td>1</td>
<td>6.25</td>
<td>-5.25</td>
<td>27.563</td>
<td>4.41</td>
</tr>
<tr>
<td>11.677 - 16.504</td>
<td>1</td>
<td>6.25</td>
<td>-5.25</td>
<td>27.563</td>
<td>4.41</td>
</tr>
<tr>
<td>16.504 - 24.755</td>
<td>4</td>
<td>6.25</td>
<td>-2.25</td>
<td>5.0623</td>
<td>0.81</td>
</tr>
<tr>
<td>24.755 - \infty</td>
<td>6</td>
<td>6.25</td>
<td>-0.25</td>
<td>0.0625</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 39.60 \]

**Step 3**: Compute level of significance from Table A.6.

\[ \chi^2 \alpha, k-8-1 = \chi^2_{0.05}, 8-1-1 = 12.6 \]

**Step 4**: Check whether random Nos. are uniformly distributed.

\[ 39.60 > 12.6 \Rightarrow \text{Random Nos. are rejected.} \]
Consider goodness of fit test using Chi Square test with equal probability. Given k = 6, \( \alpha = 0.05 \). Sample data:

0.34 0.90 1.88 1.90 0.74 2.62 2.67 3.53 4.91
5.50 1.10 1.03 1.73 1.00 2.69 1.49 2.16 0.80
0.48 5.60 0.45 0.26 0.24 0.63 0.36 1.28 0.82
2.18 0.05 0.04 0.89 0.21 0.99 0.53 3.53 2.62
0.53 1.50 2.81

Given: k = 6 \( \alpha = 0.05 \) N = 39

**Step 1:** Compute Mean

\[ \bar{x} = \frac{\sum x_i}{N} = \frac{61.61}{39} = 1.579 \]

\[ \bar{x} = \frac{1}{1.579} = 0.63 \]

**Step 2:** Compute class intervals

\[ p = \frac{1}{k} = \frac{1}{6} = 0.17 \]

\[ a_i = \frac{-1}{\bar{x}} \ln \left( 1 - i \times p \right) \]

\[ a_0 = 0 \]
\[ a_1 = 0.29 \]
\[ a_2 = 0.66 \]
\[ a_3 = 1.13 \]
\[ a_4 = 1.81 \]
\[ a_5 = 3.01 \]
\[ a_6 = \infty \]
<table>
<thead>
<tr>
<th>Class Interval</th>
<th>O_i</th>
<th>E_i = \frac{N}{K}</th>
<th>O_i - E_i</th>
<th>(O_i - E_i)^2</th>
<th>\frac{\sum (O_i - E_i)^2}{E_i}</th>
<th>X_0^2 = \frac{\sum (O_i - E_i)^2}{E_i}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 0.29</td>
<td>5</td>
<td>6.5</td>
<td>-1.5</td>
<td>2.25</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>0.29 - 0.66</td>
<td>8</td>
<td>6.5</td>
<td>1.5</td>
<td>2.25</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>0.66 - 1.13</td>
<td>8</td>
<td>6.5</td>
<td>1.5</td>
<td>2.25</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>1.13 - 1.81</td>
<td>4</td>
<td>6.5</td>
<td>-2.5</td>
<td>6.25</td>
<td></td>
<td>0.96</td>
</tr>
<tr>
<td>1.81 - 8.01</td>
<td>9</td>
<td>6.5</td>
<td>2.5</td>
<td>6.25</td>
<td></td>
<td>0.96</td>
</tr>
<tr>
<td>3.01 - \infty</td>
<td>5</td>
<td>6.5</td>
<td>-1.5</td>
<td>2.25</td>
<td></td>
<td>0.35</td>
</tr>
</tbody>
</table>

\[
X_0^2 = 3.32
\]

**Step 3:** Compute level of significance from table A6

\[
X_0^2, k - 8 - 1 = X_0^2, 6 - 1 - 1 = 9.49
\]

**Step 4:** Check whether Random Noses are uniformly distributed.

\[
: 3.32 < 9.49 \Rightarrow \text{Random Noses are accepted}
\]

### III

**K-S Test**

1. Apply goodness of fit test to check whether Random Noses are uniformly distributed over (0, T) for an interval 100. Take \( \alpha = 0.05 \)

Simulation table for critical values:

0.44 0.53 0.74 2.00 0.80 2.54 0.52 1.02 1.89

Given: \( \alpha = 0.05 \), \( n = 10 \), \( T = 100 \)

\[
R(t) = \left\{ \frac{T_1}{T}, \frac{T_1 + T_2}{T}, \ldots, \frac{T_1 + T_2 + \ldots + T_n}{T} \right\}
\]
Step 1:
\[ R(c) = \{ 0.0044, 0.0097, 0.0301, 0.0575, 0.0795, 0.0805, 0.1059, 0.1111, 0.1313, 0.1502 \} \]

Step 2:

<table>
<thead>
<tr>
<th>i</th>
<th>( R(c) )</th>
<th>( i/n )</th>
<th>( i-1/n )</th>
<th>( D^+ = \max \left{ \frac{i}{n}, R(c) \right} )</th>
<th>( D = \max \left{ \frac{i}{n} - \frac{i-1}{n} \right} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0044</td>
<td>0.1</td>
<td>0</td>
<td>0.0956</td>
<td>( \underline{0.0044} )</td>
</tr>
<tr>
<td>2</td>
<td>0.0097</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1903</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.0301</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3699</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0.0575</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3425</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0.0795</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4225</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0.0805</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5195</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0.1059</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5941</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0.1111</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6889</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0.1313</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7687</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0.1502</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8498</td>
<td>2</td>
</tr>
</tbody>
</table>

Step 3:

\[ D = \max \left\{ D^+, D^- \right\} = \max \left\{ 0.8498, 0.0044 \right\} \]

\[ D = 0.8498 \]

Step 4:

\[ D_{0.05, 10} = 0.410 \]

Step 5:

\[ 0.8498 > 0.410 = \text{Random Nos are rejected} \]
Consider sample data. Perform KS Test.

\[ \begin{align*}
0.10 & \quad 0.42 & \quad 0.46 & \quad 0.07 & \quad 1.09 & \quad 0.96 & \quad 5.53 & \quad 3.93 & \quad 1.07 \\
2.26 & \quad 2.88 & \quad 0.67 & \quad 1.12 & \quad 0.26 \\
\end{align*} \]

Interval \( (0,7) \): 100 min \( \alpha = 0.05 \ \ n = 14 \)

**Step 1:**

\[ R(i) = \{ 0.0010, 0.0152, 0.0198, 0.0305, 0.0314, 0.039, 0.0943, 0.1336, 0.1443, 0.1669, 0.1957, 0.2024, 0.2136, 0.2163 \} \]

**Step 2:**

<table>
<thead>
<tr>
<th>i</th>
<th>( R(i) )</th>
<th>( i/n )</th>
<th>( i/n - R(i) )</th>
<th>( D^+ = \max { i/n - R(i) } )</th>
<th>( D = \max { R(i) - i/n } )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0010</td>
<td>0.0714</td>
<td>0</td>
<td>0.0714</td>
<td>0.0010</td>
</tr>
<tr>
<td>2</td>
<td>0.0152</td>
<td>0.1429</td>
<td>0.0714</td>
<td>0.1214</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0198</td>
<td>0.2143</td>
<td>0.1429</td>
<td>0.1329</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>0.0305</td>
<td>0.2857</td>
<td>0.2143</td>
<td>0.0714</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.0314</td>
<td>0.3571</td>
<td>0.2857</td>
<td>0.0714</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>0.039</td>
<td>0.4286</td>
<td>0.3571</td>
<td>0.0714</td>
<td>0.4</td>
</tr>
<tr>
<td>7</td>
<td>0.0943</td>
<td>0.5</td>
<td>0.4286</td>
<td>0.0714</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>0.1336</td>
<td>0.5714</td>
<td>0.5</td>
<td>0.4714</td>
<td>0.4</td>
</tr>
<tr>
<td>9</td>
<td>0.1443</td>
<td>0.6429</td>
<td>0.5714</td>
<td>0.0714</td>
<td>0.5</td>
</tr>
<tr>
<td>10</td>
<td>0.1669</td>
<td>0.7143</td>
<td>0.6429</td>
<td>0.0714</td>
<td>0.6</td>
</tr>
<tr>
<td>11</td>
<td>0.1957</td>
<td>0.7857</td>
<td>0.7143</td>
<td>0.0714</td>
<td>0.7</td>
</tr>
<tr>
<td>12</td>
<td>0.2024</td>
<td>0.8571</td>
<td>0.7857</td>
<td>0.0714</td>
<td>0.8</td>
</tr>
<tr>
<td>13</td>
<td>0.2136</td>
<td>0.9286</td>
<td>0.8571</td>
<td>0.0714</td>
<td>0.9</td>
</tr>
<tr>
<td>14</td>
<td>0.2162</td>
<td>1</td>
<td>0.9286</td>
<td>0.0714</td>
<td>1</td>
</tr>
</tbody>
</table>

**Step 3:**

\[ D = \max \{ 0.0714, 0.0010 \} = 0.0714 \]

**Step 4:**

\[ D = 0.05, \ 14 = 0.349 \ \ \text{(table)} \]

**Step 5:**

\[ 0.7838 > 0.349 \Rightarrow \text{Random No.'s are rejected} \]