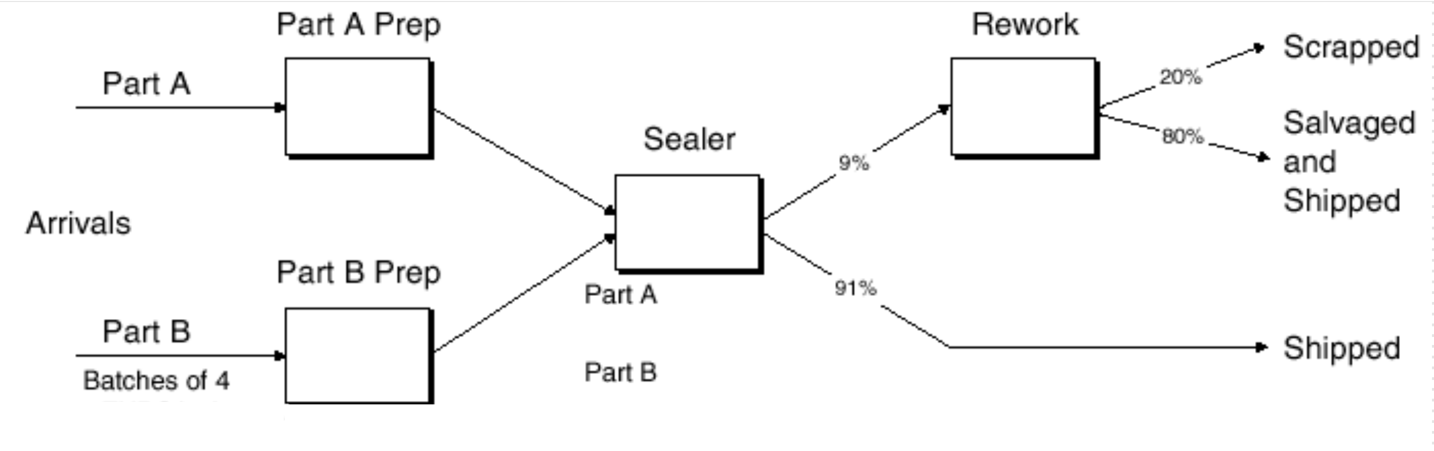
Input Analysis

Input Analysis

- Initial steps of the simulation study have been completed.
- Through a verbal description and/or flow chart of the system operation to be simulated, the random components of the system have been identified.
- The needed data has been identified and collected.
- For random components input analysis will cover how to specify what distributions should be used in the simulation model.

Example


Part A Prep
Part A
Arrivals
Part B Prep
Part B
Batches of 4
Sealer
Part A
Part B
Rework
9%
91%
20%
80%
Scrapped
Salvaged and Shipped
Shipped

Identifying Random System Components

- Examples
  -

- The amount of variability/randomness is relative to the system.
  - E.g., Machine processing time – not fixed but low variability.

Identifying Random System Components

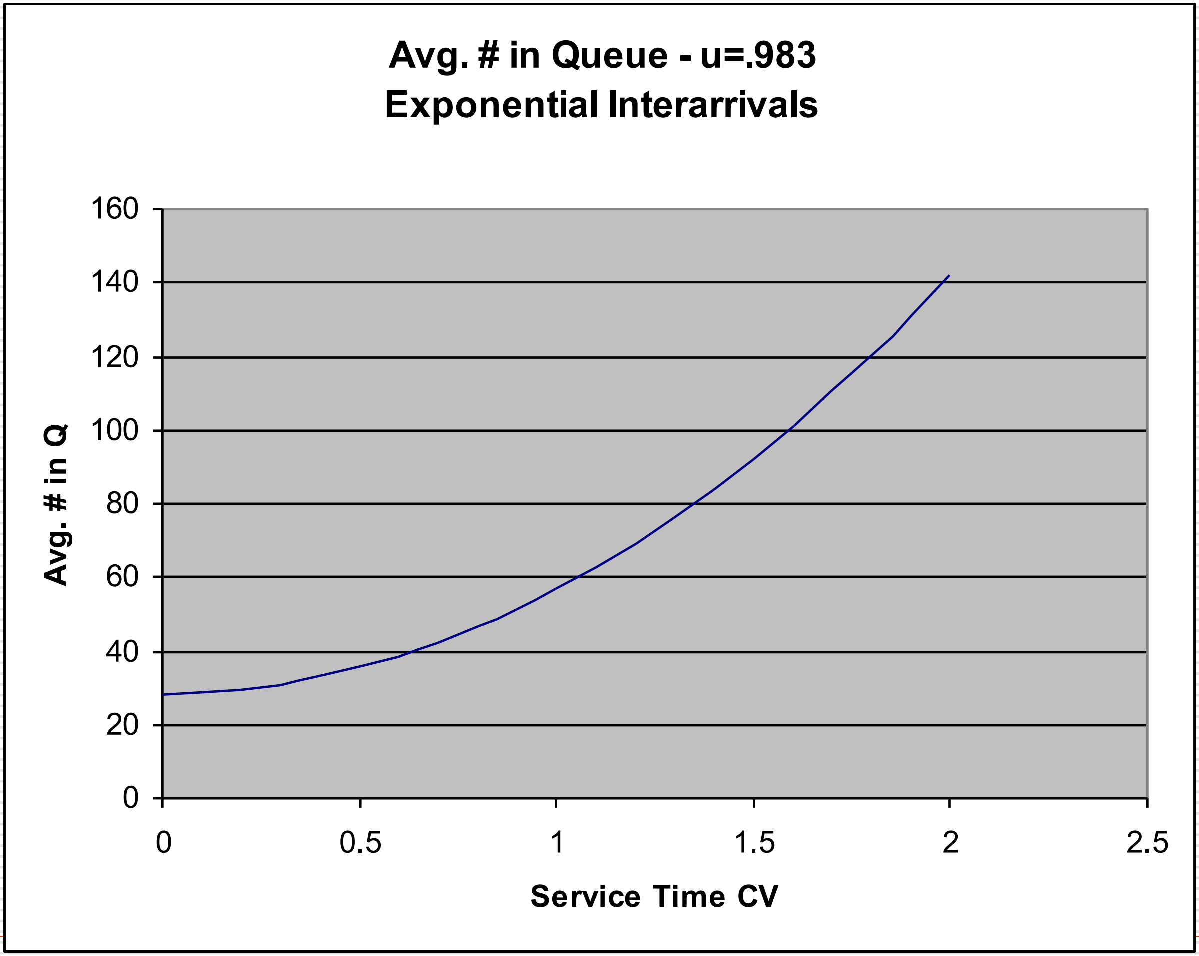
- The concept of relative variability for a random variable X, as measured by the coefficient of variation ($CV(X)$) may help.

$$CV(X) = \frac{\sigma(X)}{\mu(X)}$$

Identifying Random System Components

- Consider the single server model with a utilization =0.983.
  - Both the service times and interarrival times have a CV=1.
  - The average number in queue = 57 jobs.
- If the service time CV=0
  -
- If the service time CV = 0.25
  -

Identifying Random System Components



Identifying Random System Components

- If the system contains many different random components, and data collection resources are limited focus on random components with higher CVs.

Collecting Data

- Collect data that corresponds to what will be simulated.
- More is better.
- Examples
  - Customer/job arrivals.
    - Collect the number of customer arrivals or collect customer interarrival times?
  - Job processing times.
    - What does the time between jobs include?

Collecting Data

- Through observation of the system
- From historical data
  - Must be cautious about using data from databases without a complete understanding of what is included in the data
  - Will often require additional processing

Collecting Data Through Observation - Example

- Simulation of vehicle traffic at an intersection.
  - Arrivals of cars to the intersection – what will you collect?

Collecting Data Through Observation - Planning

- Create your data collection scenario
- Consider
  - What data will be recorded
    -
  - Once data is collected
    -

Input Analysis

- Assume you have "raw" data collected for some random component of the system.
- Example: Service time data (minutes)

Obs #	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
1	3.5	0.4	5.1	1.1	0.8	1.2	1.0
2	5.2	0.6	3.2	1.2	3.6	3.3	4.1
3	0.7	2.0	0.7	3.3	1.2	0.4	0.7
4	8.9	1.7	1.5	0.2	4.7	4.0	1.2
5	0.8	1.3	0.2	3.8	1.1	0.8	0.1
6	2.6	1.0	3.0	1.8	1.0	0.9	1.8
7	0.9	0.1	4.6	4.7	0.7	1.3	0.4
8	1.3	4.4	5.2	5.7	0.9	10.6	0.0
9	1.4	1.2	5.1	0.4	4.8	1.9	1.7
10	1.9	0.9	0.2	4.7	6.4	0.8	0.7

Input Analysis

- The random component of the system will be represented in the simulation model as observations from a probability distribution.
-

Input Analysis

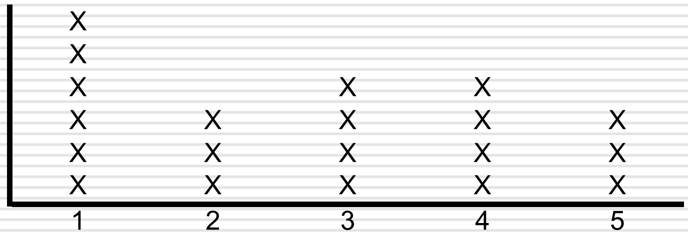
- This process is called "<u>fitting a distribution</u>".
  - Fitting a distribution - Selecting and justifying the proper distribution to use in the simulation to represent the random system component.
- In general there are two choices for the type of distribution fitted.
  -

Input Analysis

- When to use empirical distributions.
  -

Input Analysis

- Empirical distributions
  - Generates observations that match the percentage of different observations in the collected data.
- Example – Discrete case



Input Analysis

- Good features of empirical distributions.
  - Simulated distribution represents actual data collected over a specific period of time, which is good for validation purposes, if performance measures for the same period of time are known.
- Bad features of empirical distributions.
  - The data collected only reflects observations over the time period in which it was collected, and may not accurately represent what could occur.

Input Analysis

- Arena
  - Discrete empirical distribution

DISC(0.25, 1, 0.5, 1.25, 0.8, 1.75, 1.0, 2)

Input Analysis

- Arena
  - Continuous empirical distribution

    CONT(0.25, 1, 0.5, 1.25, 0.8, 1.75, 1.0, 2)

Fitting Theoretical Distributions

- General procedure
  1. Guess/hypothesize an appropriate theoretical distribution with specific parameters.
  2. Apply a statistical test to the hypothesis that the observed data are observations from the theoretical distribution from step 1.
  3. Repeat if necessary.
- Software available to quickly perform fitting.

Fitting Theoretical Distributions

- Step 1 – Suggesting a distribution.
  - Use process knowledge/theory.
  - Descriptive statistics.
  - Histograms.
- Process knowledge/Theory
  - Inspections, large population, inspection sample size =n.
  - Parts processed between failures.

Fitting Theoretical Distributions

- Process knowledge/Theory
  - Customer arrivals to a system from a large population.
  - Normal distribution.

Fitting Theoretical Distributions

- Descriptive statistics
  - Coefficient of variation (for X continuous) – CV(X)
    - Different theoretical distributions have different values or ranges of possible values.
  - X~Exponential : CV= 1.
  - X~U(a,b) :

$$CV(X) = \frac{b-a}{\sqrt{3}(b+a)}$$

  - X~Lognormal : 0 < CV < ∞.

Fitting Theoretical Distributions

- Descriptive statistics
  - Lexis ratio (for X discrete = $Var(X)/E(X)$) denoted $\tau(X)$
    - Different theoretical distributions have different values or ranges of possible values.
  - $X \sim Geometric(p)$: $\tau(X) = 1/p$.
  - $X \sim Binomial(n,p)$: $\tau(X) = 1-p$.
  - $X \sim Poisson$: $\tau(X) = 1$.

Fitting Theoretical Distributions

- When data has been collected, the CV or lexis ratio can be estimated from the data using the sample variance/standard deviation, and sample mean.

Fitting Theoretical Distributions

- Skewness – A measure of symmetry.
  -

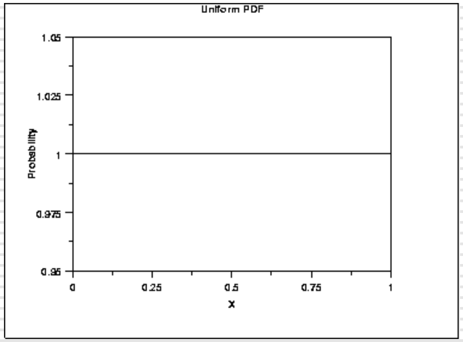
$$\text{Skewness} = \frac{E[(X-\mu)^3]}{\sigma^3}$$

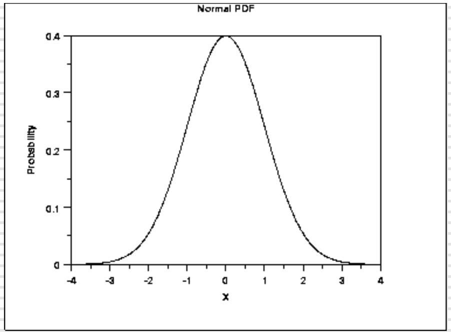
$$\text{Sample skewness} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3 / n}{s(n)^3}, \text{ where } s(n) = \sqrt{\frac{s^2 * (n-1)}{n}}$$

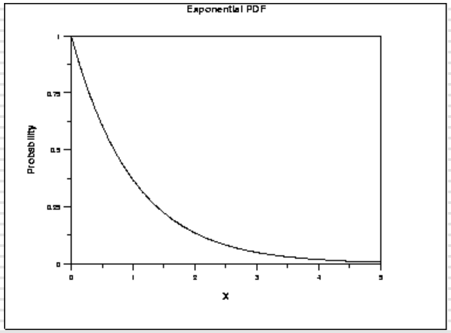
Fitting Theoretical Distributions

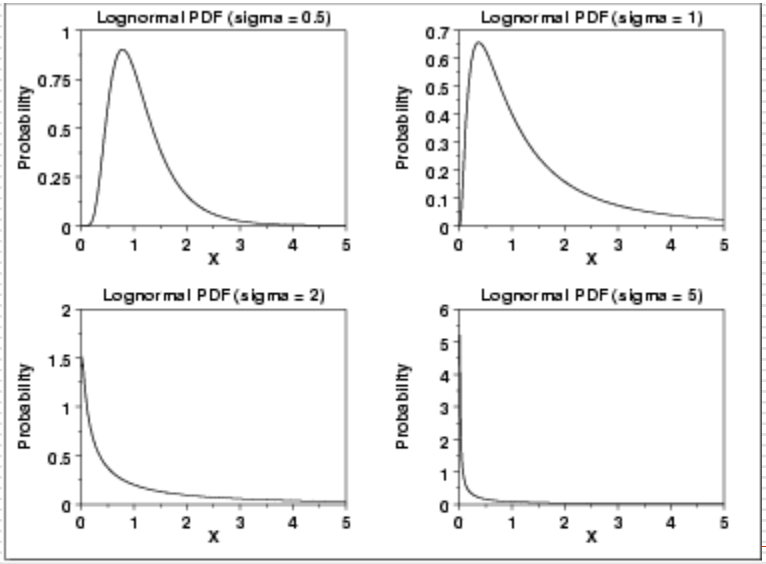
- Histograms – A graph of the frequency of observations within predefined, non-overlapping intervals.
- It is an estimate of the shape of a random variables probability mass function or density function.
- Density functions/probability mass functions tend to have recognizable shapes.

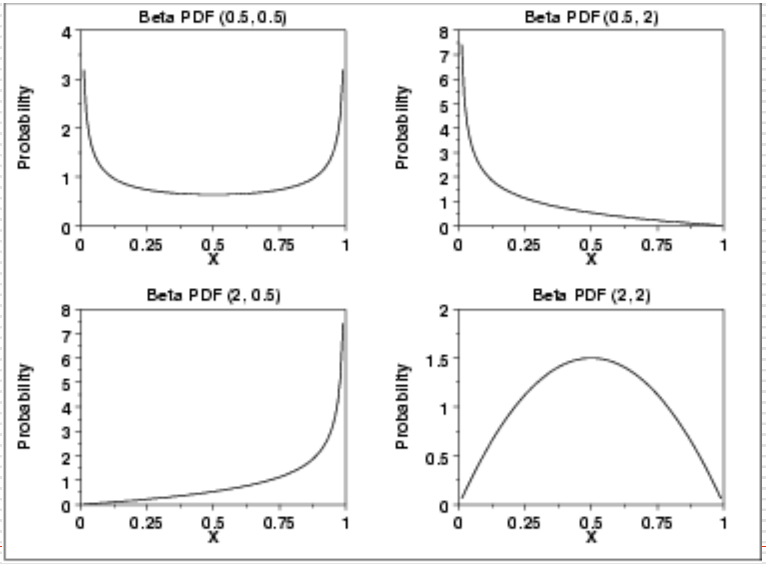
Fitting Theoretical Distributions











Fitting Theoretical Distributions

- Making a histogram
  1. Break up the range of values covered by the data into k disjoint intervals $[b_0,b_1)$, $[b_1,b_2)$,… ,$[b_{k-1},b_k)$. All intervals should be the same width Δb except for the first and last intervals (if necessary).
  2. For $j = 1,2,\ldots,k$, let q_j = the proportion of the observations in the jth interval defined in 1.
  3. Histogram value at x, $h(x) = 0$ if $x < b_0$, q_j if $b_{j-1} \leq x < b_j$, and 0 if $x \geq b_k$.

Fitting Theoretical Distributions

- Making histograms
  - Histograms may be changed subjectively by changing the intervals and Δb (the interval size).
  - Guidelines (Hines, et al. 2003)
    - Keep the number of intervals between 5 and 20.
    - Set the number of intervals to be about the square root of the number of observations.

In-class Exercise

- For data collected from different random components of a system.
  - Estimate the CV.
  - Construct a histogram.
  - Suggest a theoretical distribution to represent the random component.

n	Dist. 1	Dist. 2	Dist. 3
1	6.0	0.9	1.8
2	7.0	1.4	9.9
3	7.5	3.7	0.4
4	5.8	2.0	10.8
5	7.9	0.2	0.6
6	3.3	3.5	2.0
7	5.3	4.6	5.5
8	2.2	1.7	12.4
9	5.5	2.7	2.6
10	5.3	3.6	8.2
11	5.9	3.0	0.8
12	2.7	0.3	3.7
13	4.6	0.9	1.8
14	2.5	3.4	6.3
15	4.5	4.8	22.5
16	6.5	0.6	40.4
17	3.6	0.5	27.2
18	3.4	3.3	29.6
19	2.6	1.9	5.6
20	6.4	3.8	1.9
21	1.7	1.8	5.3
22	4.2	0.1	1.2
23	2.6	4.5	1.4
24	7.3	1.2	0.1
25	6.3	0.6	8.5
26	7.9	4.4	11.3
27	6.6	5.0	5.2
28	2.9	0.6	3.8
29	3.4	3.1	2.5
30	2.9	0.5	49.9
Sample Avg.	4.8	2.3	9.4
StdDev. (est)	1.9	1.6	12.3

In-class Exercise

In-class Exercise

Parameter Estimation

- After suggesting a possible theoretical distribution to represent a random system component, the parameters for that distribution must be estimated from the data ($x_1, x_2, \ldots, x_n$).
- Different distributions may have a different number of parameters.
  -
- Estimating parameters from data for some distributions is intuitive.

Parameter Estimation

- Attention will be restricted to <u>maximum likelihood estimators</u> or MLEs.
  - Good statistical properties (needed as part of a goodness of fit test).
- Examples of MLEs
  - X~Exponential(λ)
    - λ = 1/E[X]
    - The MLE for λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Parameter Estimation

- Examples of MLEs
  - X~Uniform(a,b)
    -
  - X~Normal(μ,σ^2)
    -

Parameter Estimation

- Not all MLEs for distribution parameters are so intuitive.
- Example: X~Lognormal(σ,μ) where σ>0 is called a shape parameter and μ ∈ (-∞, ∞) is called a scale parameter.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln(x_i)}{n}, \qquad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (\ln(x_i) - \hat{\mu})^2}{n}}$$

Statistical Goodness of Fit Tests

- A distribution has been suggested to represent a random component of a system.
- The data has been use to compute estimates (MLE) of the distribution parameters.
- Statistical goodness of fit tests determine whether there is enough evidence to say that the data is not well represented by the suggested distribution.

Statistical Goodness of Fit Tests

- Let $\hat{F}$ be the suggested distribution with parameters θ obtained from the data using MLE estimators.
- The statistical goodness of fit tests will evaluate the null hypothesis
  -

Statistical Goodness of Fit Tests

- Three tests will be reviewed.
  - Chi-square goodness of fit test.
  - Kolmogorov-Smirnov test.
  - Anderson-Darling test.

Chi-square Goodness of Fit Test

- Intuitively think of this test as a formal way to compare a modified histogram to a theoretical distribution.

Chi-square Goodness of Fit Test

- Test procedure
  1. Divide the entire range of the hypothesized distribution into k adjacent intervals – (a_0,a_1), $[a_1,a_2),\ldots,[a_{k-1},a_k)$ where a_0 could be $-\infty$, and a_k could be $+\infty$.
     -

Chi-square Goodness of Fit Test

- Test procedure

2. Tabulate the number of observations falling in each interval – $(a_0,a_1), [a_1,a_2),\ldots,[a_{k-1},a_k)$. Let,
   - N_j = The number of observations falling in the jth interval $[a_{j-1},a_j)$.
   - $\sum N_j = n$:the total number of observations.
3. Calculate the expected number of observations in each interval from the suggested distribution and estimated parameters.
   - Discrete
   - Continuous

Chi-square Goodness of Fit Test

- Discrete

Expected number of observations in interval $j = np_j$
where

$n =$ Total number of observations.

$$p_j = \sum_{i:a_{j-1} \le x_i < a_j} p(x_i) \text{ where } p(x_i) = \text{prob. mass function for the dist.}$$

Chi-square Goodness of Fit Test

- Continuous

Expected number of observations in interval $j = np_j$
where

n = Total number of observations.

$$p_j = \int_{a_{j-1}}^{a_j} f(x)dx \quad \text{where } f(x_i) = \text{prob. density function for the dist.}$$

Chi-square Goodness of Fit Test

- Test procedure
  4. Calculate the test statistic TS

$$TS = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \quad \text{where } N_j = \text{Number of observations in interval } j.$$

Chi-square Goodness of Fit Test

- If the distribution tested is a good fit, should TS be large or small?

- How do we know what large or small is?
  - What do we compare TS to?
  - What is the sampling distribution of the test statistic.

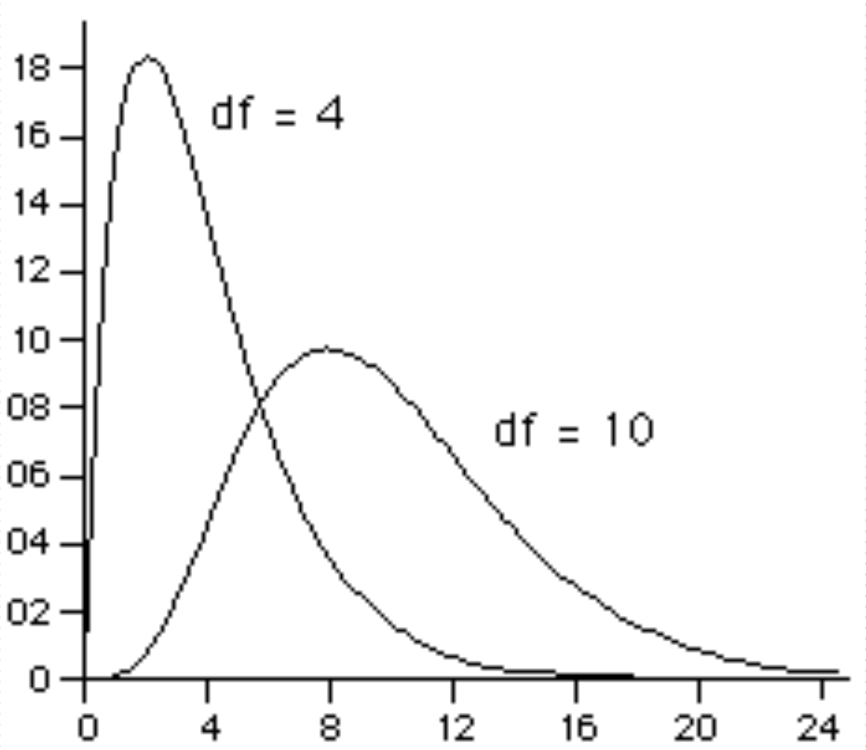
Chi-square Goodness of Fit Test

- Result – If MLE’s are used to estimate the *m* parameters of a suggested distribution, then if H_0 is true, as $n \to \infty$, the distribution of TS (the sampling distribution) converges to a Chi-square distribution that lies between two Chi-square distributions with $k-1$ and $k-m-1$ degrees of freedom.
- If $\chi^2_{df,1-\alpha}$ represents the value of the Chi-square distribution (df = degrees of freedom) such that α of the area under the Chi-square distribution lies to the right of this value, then for the corresponding value for the distribution of TS

$$\chi^2_{k-m-1,1-\alpha} \leq TS_{1-\alpha} \leq \chi^2_{k-1,1-\alpha}$$

Chi-square Goodness of Fit Test

Density Function Plots for the Chi-Square Distribution



Chi-square Goodness of Fit Test

- To be conservative, reject H_0 if

$$TS > \chi^2_{k-1,1-\alpha}$$

- Needed values from the chi-square distribution can be found using the CHIINV function in Excel.

Chi-square Goodness of Fit Test

- As with histograms, there are no precise answers for how to create intervals.
  -

Example

Table 5.6 $n = 87$ **demand sizes sorted into increasing order**

1	3	4	4	5	6	6	6	7	8	8	9	11
1	3	4	5	5	6	6	6	7	8	8	9	12
2	3	4	5	5	6	6	6	7	8	9	9	12
2	3	4	5	5	6	6	7	7	8	9	9	
2	3	4	5	5	6	6	7	7	8	9	9	
2	4	4	5	5	6	6	7	7	8	9	10	
3	4	4	5	6	6	6	7	8	8	9	11	

Example

Table 5.8 A chi-square goodness-of-fit test for the demand-size data

j	Interval	N_j	np_j	$\frac{(N_j - np_j)^2}{np_j}$
1	[−0.5, 3.5)	12	12.267	0.006
2	[3.5, 4.5)	10	11.223	0.133
3	[4.5, 5.5)	12	13.659	0.201
4	[5.5, 6.5)	18	13.920	1.196
5	[6.5, 7.5)	10	12.180	0.390
6	[7.5, 9.5)	20	15.660	1.203
7	[9.5, ∞)	5	8.091	1.181
				$\chi^2 = 4.310$

In-class Exercise

1. Calculate the estimated CV
2. Construct a histogram using 5 intervals.
3. Hypothesize a distribution (exponential or uniform).
4. Conduct a Chi-Square Goddness of fit test (aplha = 0.10).
   using k=4.

Ch-Sq(3,0.9) = 6.25

For an exponential distribution

$$F(x) = 1 - e^{-\lambda x} \text{ where } \lambda = 1/E[X]$$

Chi-Square Goodness of Fit Test Exercise

Observation	Value	Value^2
1	3.36	11.31
2	32.06	1027.54
3	19.59	383.65
4	2.50	6.26
5	21.96	482.26
6	0.41	0.17
7	0.31	0.10
8	1.22	1.50
9	2.89	8.37
10	41.84	1750.24
11	0.09	0.01
12	16.82	282.84
13	3.97	15.75
14	1.59	2.53
15	8.03	64.43
16	6.13	37.54
17	22.86	522.44
18	7.63	58.20
19	5.21	27.11
20	0.76	0.57
21	9.08	82.41
22	1.50	2.24
23	18.55	344.21
24	11.31	127.81
25	7.28	53.01
26	7.75	60.07
27	26.17	684.86
28	3.15	9.94
29	3.14	9.86
30	2.88	8.27
Sum	**290.02**	**6065.51**

In-class Exercise

In-class Exercise

Kolmogorov-Smirnov (K-S) Test

- Pros
  - Does not require partitioning of the data into intervals.
  - When the parameters of a distribution are known, it is a more powerful test than the chi-square test (less probability of a type II error).

Kolmogorov-Smirnov (K-S) Test

- Cons
  - Usually the parameters of a distribution are estimated.
    - Critical value (those values to compare to a test statistic) have been estimated for some hypothesized distributions using Monte Carlo simulation.
      - Normal
      - Exponential
      - Others (see Law and Kelton reference)
    - Difficult to use in the discrete case (ARENA Input Analyzer will generate no K-S test results).

Kolmogorov-Smirnov (K-S) Test

- The parameters known case is used when the distribution parameters are estimated, and estimated critical values are not available.
  - This is conservative – The type I error is smaller than specified.

Kolmogorov-Smirnov (K-S) Test

- The K-S test uses the empirical probability distribution function generated from the data as a basis for computing a test statistic.
  - Suppose there are n data observations collected, $x_1, \ldots, x_n$.
  - The empirical distribution function $F_n(x)$ is

$$F_n(x) = \frac{\text{Number of } x_i's \leq x}{n}$$

  - This is a right continuous step function.

Example

Kolmogorov-Smirnov (K-S) Test

- Intuitive description of the K-S test – Let $\hat{F}$ be the distribution function of the suggested distribution. Overlay $\hat{F}$ on $F_n(x)$ and find the maximum absolute difference.
- K-S test statistic D_n:

$$D_n = \max(D_n^+, D_n^-)$$

where

$$D_i^+ = \frac{i}{n} - \hat{F}(x_{(i)})$$

$$D_i^- = \hat{F}(x_{(i)}) - \frac{i-1}{n}$$

$$x_{(i)} = ith \text{ sorted value of the } x_i's$$

$$D_n^+ = \max_{1 \le i \le n} \{D_i^+\}$$

$$D_n^- = \max_{1 \le i \le n} \{D_i^-\}$$

Kolmogorov-Smirnov (K-S) Test

Kolmogorov-Smirnov (K-S) Test

- Test – Reject $\hat{F}$ if

$$\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) D_n > c_{1-\alpha}$$

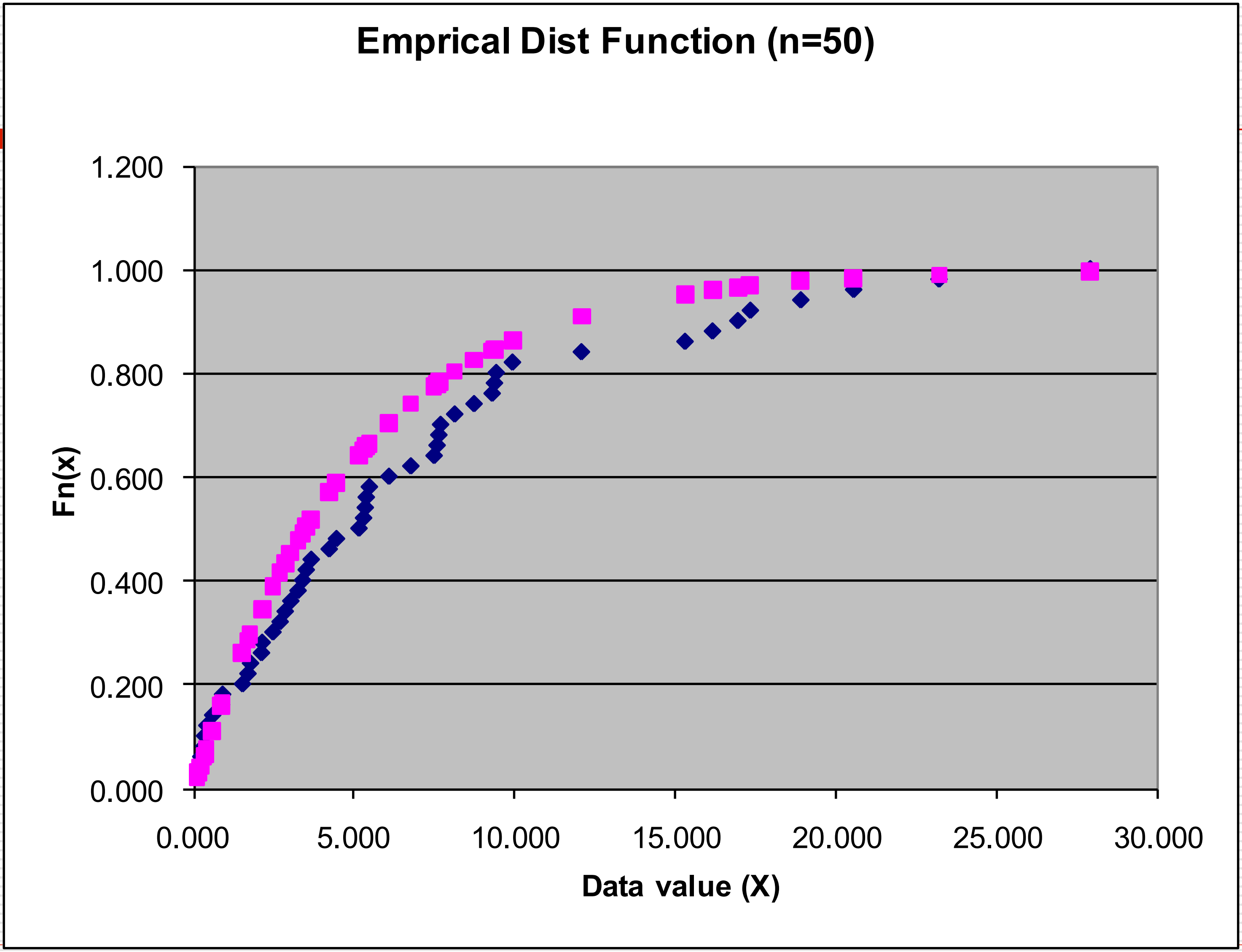
where

	$1-\alpha$			
	.85	.90	.95	.975
$c_{1-\alpha}$	1.138	1.224	1.358	1.480

Kolomogorov-Smirnov Test - Example with 50 data points (not all shown)

	(Expo(5))		Expo(mean=5)		
n	**Sample data**	**Step Value**	**F(x)**	**Dn+**	**Dn-**
0	**0.000**	0.000	0.000	**0.042**	**0.163**
1	0.119	0.020	0.024	-0.004	0.024
2	0.157	0.040	0.031	0.009	0.011
3	0.217	0.060	0.042	0.018	0.002
4	0.311	0.080	0.060	0.020	0.000
5	0.334	0.100	0.065	0.035	-0.015
6	0.405	0.120	0.078	0.042	-0.022
7	0.587	0.140	0.111	0.029	-0.009
8	0.863	0.160	0.158	0.002	0.018
9	0.895	0.180	0.164	0.016	0.004
10	1.517	0.200	0.262	-0.062	0.082
11	1.686	0.220	0.286	-0.066	0.086
12	1.766	0.240	0.298	-0.058	0.078
13	2.107	0.260	0.344	-0.084	0.104
14	2.129	0.280	0.347	-0.067	0.087
15	2.465	0.300	0.389	-0.089	0.109
16	2.688	0.320	0.416	-0.096	0.116
17	2.847	0.340	0.434	-0.094	0.114
18	3.022	0.360	0.454	-0.094	0.114
19	3.247	0.380	0.478	-0.098	0.118
20	3.390	0.400	0.492	-0.092	0.112
21	3.502	0.420	0.504	-0.084	0.104
22	3.657	0.440	0.519	-0.079	0.099
23	4.218	0.460	0.570	-0.110	0.130
24	4.444	0.480	0.589	-0.109	0.129
25	5.146	0.500	0.643	-0.143	0.163
26	5.289	0.520	0.653	-0.133	0.153
27	5.342	0.540	0.656	-0.116	0.136
28	5.373	0.560	0.659	-0.099	0.119
29	5.470	0.580	0.665	-0.085	0.105
30	6.081	0.600	0.704	-0.104	0.124

Test Statistic	**Critical value (alpha = 0.05)**
1.172	1.358



Emprical Dist Function (n=50)
Fn(x)
1.200
1.000
0.800
0.600
0.400
0.200
0.000
0.000
5.000
10.000
15.000
20.000
25.000
30.000
Data value (X)

In-class Exercise

Conduct a K-S test to test the hypothesis that the data are observations from a U(0,10) distribution.

n	Data					
1	1.8					
2	5.2					
3	9.8					
4	6.6					
5	2.6					
6	3.3					
7	2.9					
8	1.1					
9	9.2					
10	3.6					

Anderson-Darling Test

- Think of as a weighted K-S test.
- More weight given to differences when the hypothesized probabilities are low – similar to percentage differences vs. absolute differences

Anderson-Darling Test

- Critical values have been estimated for some hypothesized distributions
  -

Goodness of Fit Tests

- If a p-value is computed
  - Low p-values (< 0.05) indicate a poor fit.

Assessing Independence

- It has been assumed that all observations of data for some random system component are independent.
- The chi-square, K-S, and A-D tests are testing for independent identically distributed random variables.
- Most discrete event simulation models and software assume random components are independent.

Assessing Independence

- When presenting Monte Carlo simulation we talked about correlation.
  - Two or more random components of a simulation (e.g., engineering economic calculation) may be correlated.
    - E.g., Repair costs in year one and year two.
    - There was no simulated time component in these simulations.
- In discrete event simulation correlation may occur,
  - Between different random components.
  - Within the same random component.

Assessing Independence

- Examples
  - Different random components.
    -
  - Within the same random component.
    -

Recognizing Non-independence

- Method 1 – Correlation estimates and hypothesis testing.
  - E.g., Rank correlation estimates.
- Method 2 – Correlation plot.
- Method 3 – Scatter diagram.

Recognizing Non-independence

- Let $x_1, x_2, \ldots, x_n$ be collected data listed in the time order of collection.
- Sample correlation $\hat{\rho}_j = \dfrac{\hat{C}_j}{\hat{\sigma}^2}$

$$\hat{C}_j = \frac{\sum_{i=1}^{n-j}(x_i - \bar{x})(x_{i+j} - \bar{x})}{n-j}$$

- As j approaches n this estimate is very poor since the number of values used for the estimate is small.

Recognizing Non-independence

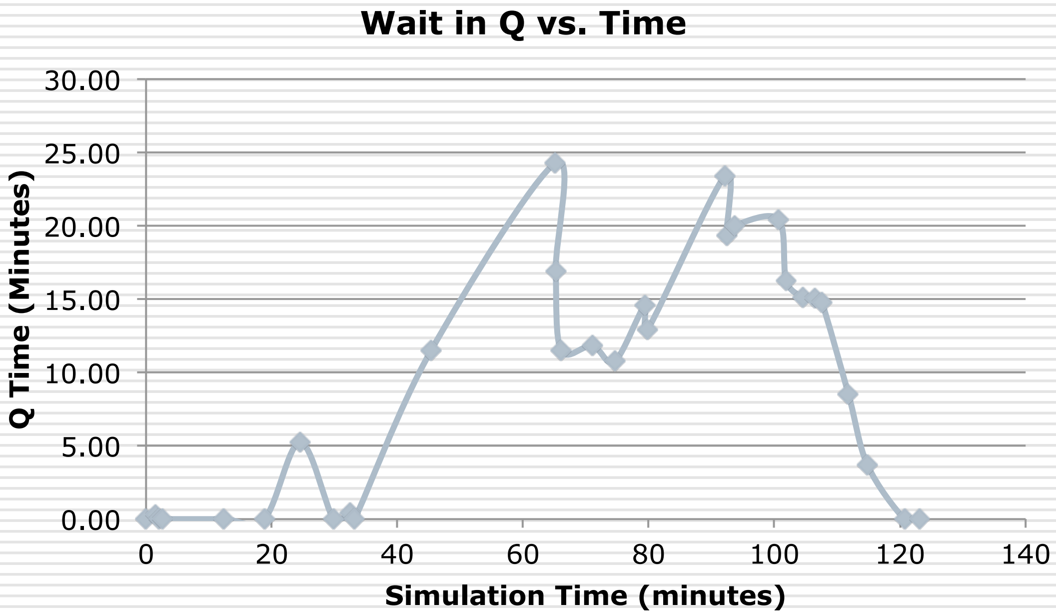
- Scatter plot.
  - Plot of (x_i, x_{i+1}) or (x_i, y_i).
  - Visual subjective interpretation.

Recognizing Non-independence

- Correlation plot – Applicable when examining correlation within a single random component.
  - Plot C_j vs. j.

Example

- Wait in queue. Correlated?
  - Single server system – Utilization = 0.98



Example – Correlation Estimates

Queue Times from an M/M/1 system - Util = .98

Average	Variance
9.23	69.47

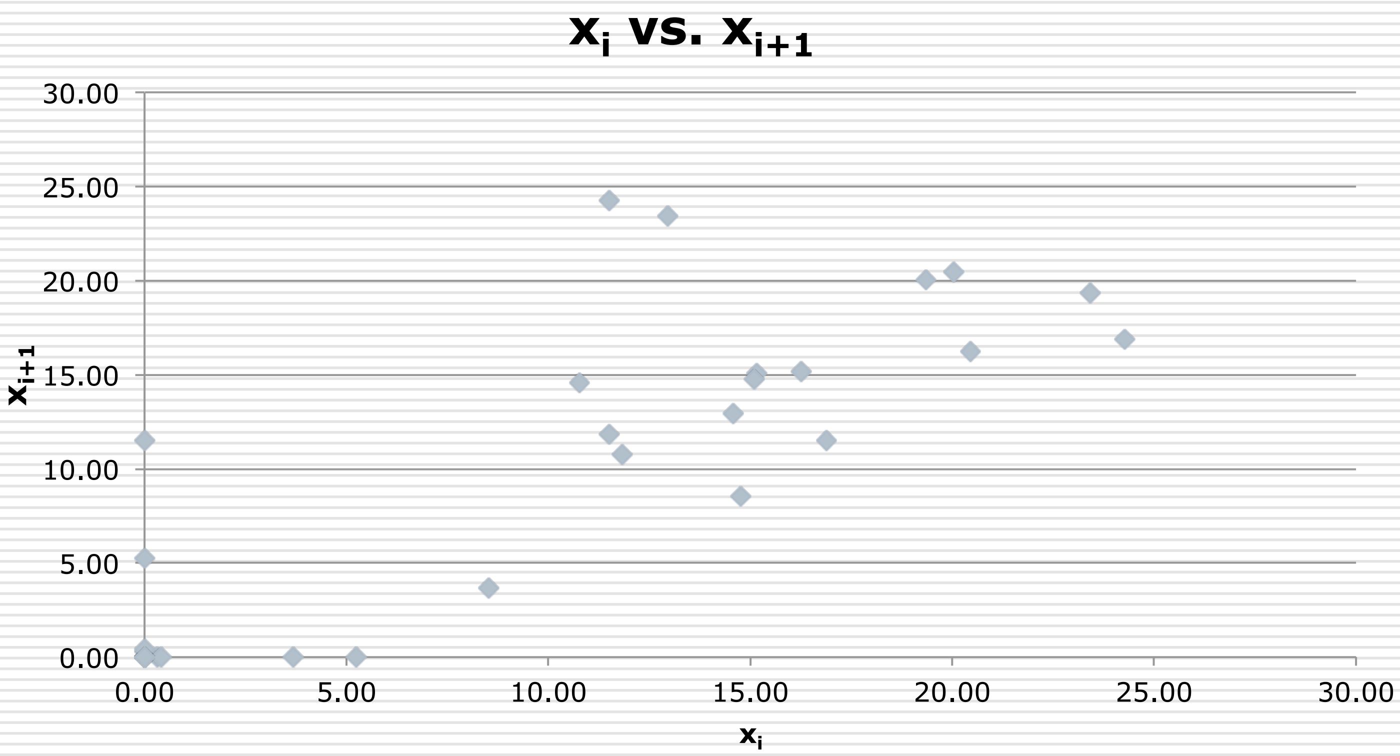
Customer	Time (minutes)	Queue Time (min)
1	0	0.00
2	1.48	0.31
3	2.10	0.00
4	2.57	0.00
5	12.28	0.00
6	18.81	0.00
7	24.53	5.24
8	29.84	0.00
9	32.40	0.42
10	33.18	0.00
11	45.36	11.50
12	65.15	24.29
13	65.21	16.90
14	66.07	11.51
15	71.03	11.83
16	74.61	10.77
17	79.38	14.58
18	79.85	12.95
19	92.13	23.42
20	92.42	19.35
21	93.71	20.04
22	100.67	20.45
23	101.88	16.26
24	104.54	15.16
25	106.49	15.09
26	107.59	14.77
27	111.80	8.52
28	114.83	3.69
29	120.73	0.00
30	123.18	0.00

Correlation Estimates

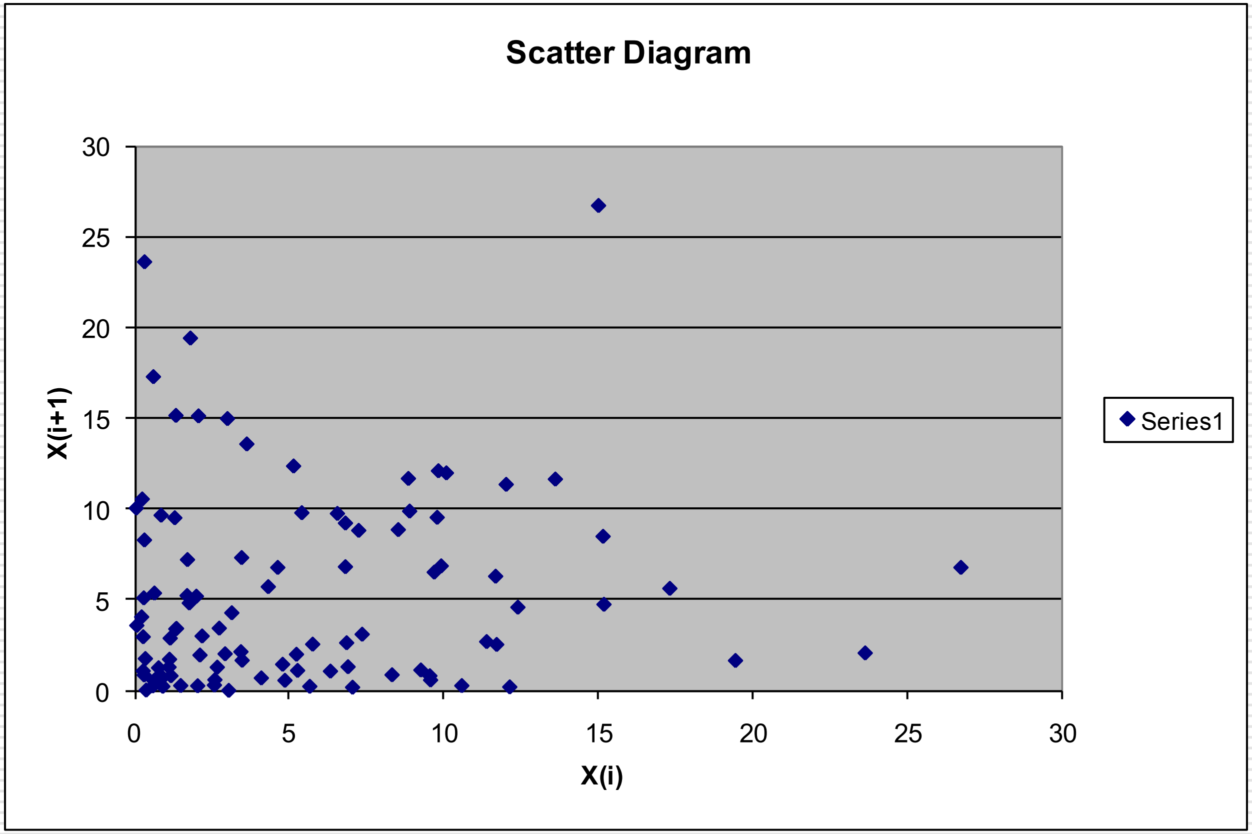
Lag 1	Lag 2
0.79	0.58

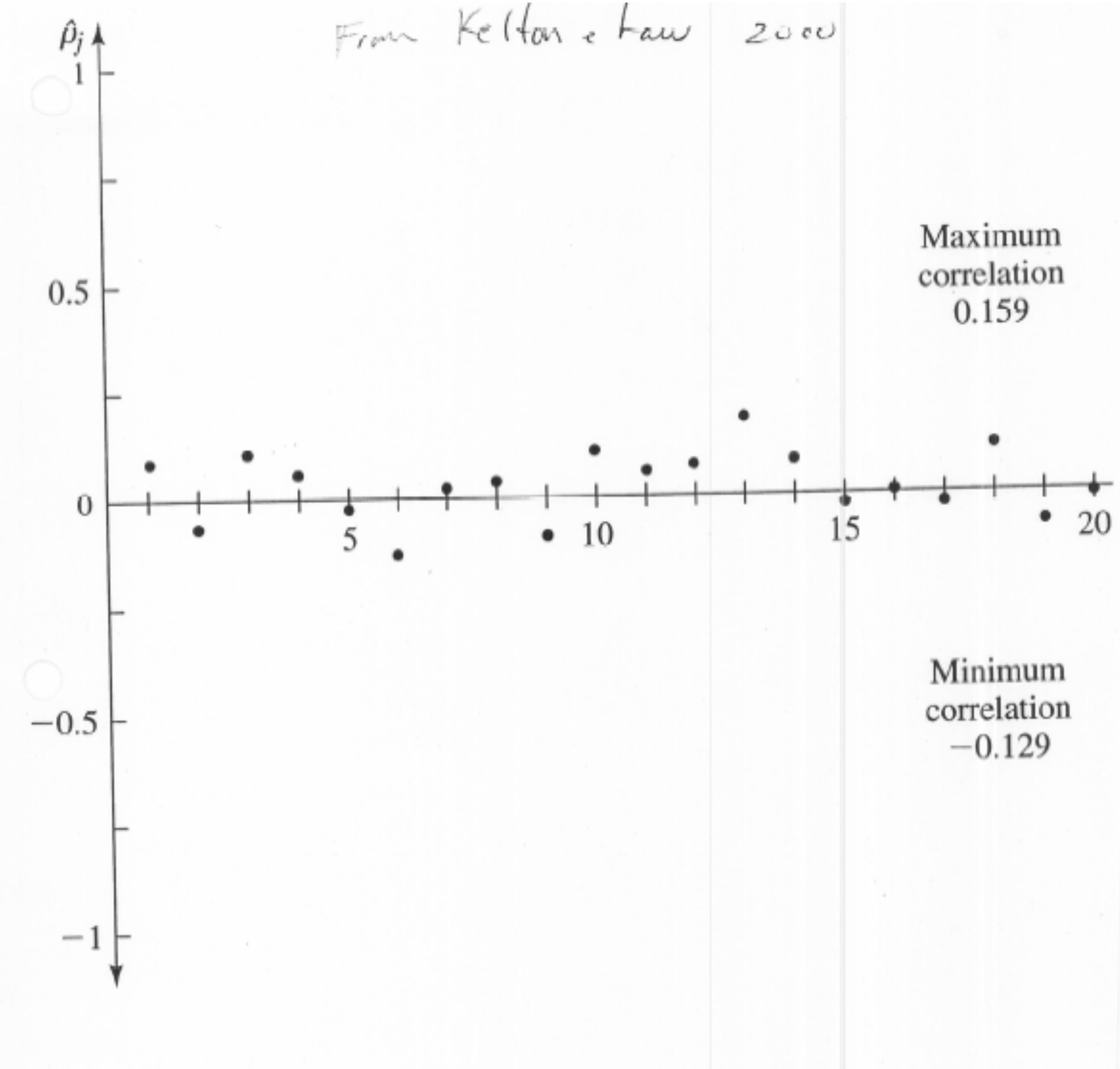
$(x_i\text{-xbar})(x_{i+1}\text{-xbar})$	$(x_i\text{-xbar})(x_{i+2}\text{-xbar})$
82.43	85.27
82.43	82.43
85.27	85.27
85.27	85.27
85.27	36.86
36.86	85.27
36.86	35.18
81.40	85.27
81.40	-19.95
-20.90	-139.00
34.07	17.34
115.35	34.18
17.40	19.90
5.90	3.49
4.00	13.88
8.23	5.72
19.86	75.84
52.72	37.59
143.55	153.25
109.28	113.48
121.16	75.87
78.79	66.45
41.61	41.10
34.66	32.77
32.37	-4.18
-3.95	-30.69
3.96	6.59
51.22	51.22
85.27	

Example – Scatter Diagram



Example Scatter Diagram – Zero Correlation




From Kelton & Law 2000
$\hat{\rho}_j$
1
0.5
0
−0.5
−1
5
10
15
20
Maximum correlation 0.159
Minimum correlation −0.129

Recognizing Non-independence

- If you determine that correlations exist, what do you do?
  -

What to do with Minimal Information?

- Suppose you cannot collect data or that none is available.
- Suggested no data distributions.

Possible No-Data Distributions		
Distribution	**Characteristics**	**Examples**
Exponential	CV = 1 Non-negative (minimum = 0) No maximum value Non symmetric	Interarrival Times Time between failures
Triangular	Symmetric or non-symmetric Finite range Easily interpreted parameters (Min, Max, Mode)	Activity times
Uniform	Finite range All values equally likely	Little known about process

What to do with Minimal Information?

- Getting information about distribution parameters.
  -

Non-stationary Arrival Processes

- A very common feature of many systems (in particular service systems) is that the arrival rate (e.g., customers) varies as a function of time.
  - The arrival process is called <u>non-stationary</u>.

Non-stationary Arrival Processes

- A practical method to approximate a non-stationary arrival process is to use a “piecewise constant” rate function.
- Over predefined time intervals assume a constant (and different) arrival rate.

Non-stationary Arrival Processes

- Example – A very common arrival process used in simulation is the Poisson process.
  -

Non-stationary Arrival Processes

- Implementation of a non-stationary Poisson process.
  - Common but incorrect approach.
    - Have the mean interarrival time (or interarrival rates) be represented with a variable. Change the value of this variable for each interval.

Non-stationary Arrival Processes

- What's wrong with this approach?

Non-stationary Arrival Processes

- One correct approach for implementing a non-stationary Poisson Process is called thinning.

Suppose there are n different arrival rates $\lambda_1, \ldots, \lambda_n$ in the piecewise constant rate function.

Let $\lambda_{\max} = \max(\lambda_1, \ldots, \lambda_n)$

1. Generate Poisson arrivals at the rate of $\lambda_{\max}$.

2. Reject arrivals in interval i with probability $= 1 - \dfrac{\lambda_i}{\lambda_{\max}}$.

Non-stationary Arrival Processes

Arena Basic Process Panel Online Help

Hide Back Forward Print Options

Contents Index Search

Schedule Module—Examples

Examples | Prompts | Remarks | Main Topic

Example 1—Resource Schedule

Prompt	Entry
Name	Daily Schedule
Type	Capacity
Time Units	hours
Value (Capacity)	4, 2, 4
Duration	4, 2, 3

The Daily Schedule is a resource schedule, as it is specified as type Capacity. The number of resources available will be four for the first 4 hours of the simulation (morning hours). The duration then drops to two for 2 hours. This may represent a lunch time period while two accountants go to lunch for a 1 hour period, then the other two go to lunch for the second hour. For the last 3 hours of the day, all four of the accountants are available (afternoon hours).

Example 2—Arrival Schedule

Prompt	Entry
Name	Spring Weekly Schedule
Type	Arrival
Time Units	days
Scale Factor	1
Value	25, 40, 55, 70, 60, 20, 5
Duration	1, 1, 1, 1, 1, 1, 1

The *Spring Weekly Schedule* demonstrates the use of an arrival type schedule. Each data set in the arrival schedule specifies *Value* entity arrivals **per hour** over *Time Duration*. An exponential distribution is used to evenly distribute the *Value* arrivals over each hour. Assuming the hours per day is 24, *Spring Weekly Schedule* states that approximately 600 entities will arrive within the first day, 960 the second day, 1320 the third day, 1680 the fourth, 1440 the fifth, 480 during the sixth, and 120 during the last day.

The scale factor in this example is set to 1. To increase the arrival rate by 10%, the scale factor can be set to 1.1, while specifying a scale factor of 0.9 will decrease the arrival rate by 10%.

Non-stationary Arrival Processes