Department of Systems Engineering

George Mason University

SYST 302: Systems Methodology and Design II #4

Kuo-Chu Chang
Fairfax, Virginia

Statistical Methods

- **Descriptive Statistics**
  - **Analysis techniques to describe sample data**
  - **Frequency distribution (histogram)**
  - **Measures of central tendency (mean)**
  - **Measures of dispersion (variance)**
- **Inferential Statistics**
  - **Drawing inference about a population from the sample**
  - **Estimate population parameters**
  - **Hypothesis testing**

Parameter Estimation

- **Estimation:** the process of inferring the value of a quantity of interest from indirect, inaccurate, and uncertain observations
  - Determine model parameters for physical system - system identification
  - Determine message characteristics from noisy channel - communication
- **Optimal Estimator:** the algorithm that yields an estimate of interest which minimizes a certain error criterion
  - Unbiased estimate
  - Maximum likelihood estimate
  - Maximum A Posterior estimate
  - Minimum Mean Square Error estimate

Maximum Likelihood and Maximum A Posterior Estimates

Non – Bayesian :

Maximum Likelihood Estimate (MLE):

$$\hat{x}^{ML}(Z) = \max_x \Lambda_Z(x) = \max_x p(Z|x)$$

Bayesian :

Maximum A Posterior Estimate (MAP):

$$\hat{x}^{MAP}(Z) = \max_x p(x|Z) = \max_x [p(Z|x)p(x)]$$

ML Estimator Example

Maximum Likelihood Estimate (MLE) :

$$\hat{x}^{ML}(Z) = \max_x \Lambda_Z(x) = \max_x p(Z \mid x)$$

For example: (parameter estimation)

Given a static parameter x with Gaussian prior

$p(x) \sim N(x; \bar{x}, \sigma_0^2)$

with observation model : $z = x + \omega,$

where $\omega \sim N(0, \sigma^2),$ then with ML estimator

$$\Lambda(x) = p(z|x) = N(z; x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-x)^2}{2\sigma^2}}$$

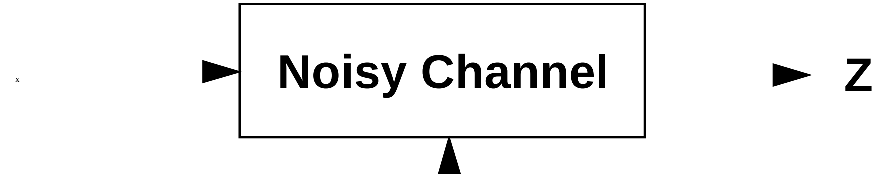
$$\Rightarrow \hat{x}^{ML} = \max_x \Lambda(x) = z$$

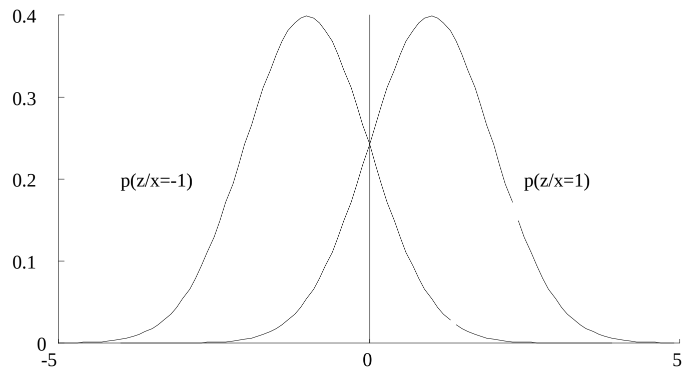
Communication System Example

$$z = x + w$$

$$x = \begin{cases} 1 \\ -1 \end{cases}, \quad w = N(0,1)$$

$$\Rightarrow p(z|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}}$$





Maximum likelihood decision policy: if z > 0, x=1 if z < 0, x=-1

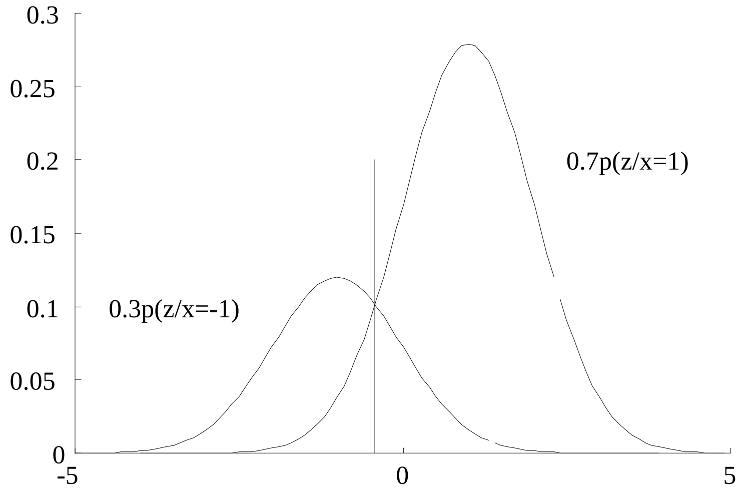
$$\text{Prob(error)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{\frac{-(t-1)^2}{2}} dt = erf(-1) = 0.1587$$

MAP Estimate Example

$$z = x + w$$

$$P(x=1) = 0.7,\ P(x=-1) = 0.3$$

$$w = N(0,1)$$



$$\text{MAP Estimate}: \max_{x} p(x \mid z) = \max_{x} \frac{1}{C} p(z \mid x) P(x)$$

$$\text{Decision boundary}: p(z \mid x=1)P(x=1) = p(z \mid x=-1)P(x=-1)$$

$$\Rightarrow 0.7e^{-\frac{(z-1)^2}{2}} = 0.3e^{-\frac{(z+1)^2}{2}} \Rightarrow z = \ln(3/7)/2 = -0.4236$$

$$\text{Prob(error)} = 0.7\,erf(-1.4236) + 0.3[1 - erf(1-0.4236)]$$

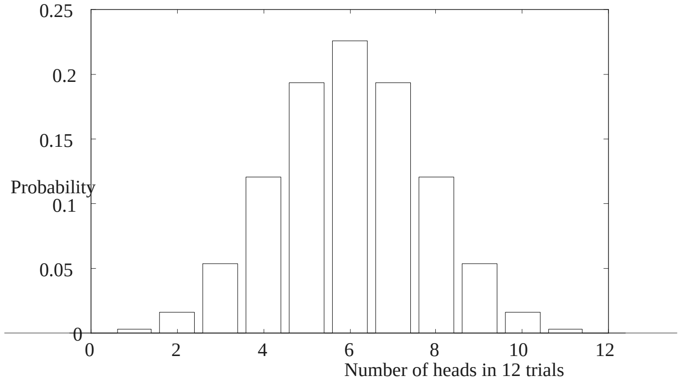
$$= 0.7 * 0.0773 + 0.3 * [1 - 0.7178] = 0.0541 + 0.0847 = 0.1388$$

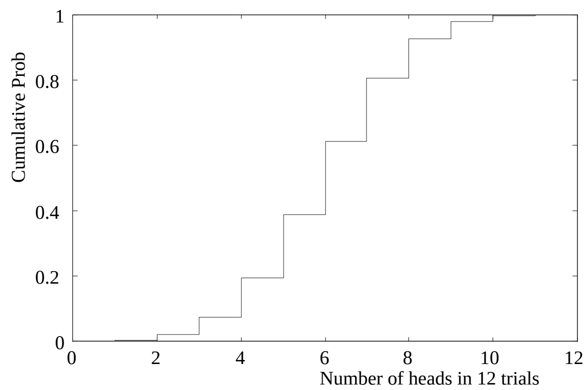
Hypothesis Testing

- **Hypothesis Testing**
  - Formulate a hypothesis concerning the population parameters (null hypothesis)
  - Specify acceptance or rejection criteria (level of significance)
  - Conduct the statistical test (making the decision)
- **Possible Outcomes**
  - Correct decision (accept true hypothesis or reject false hypothesis)
  - Type I error (reject true hypothesis)
  - Type II error (accept false hypothesis)
- **Designing a Test**
  - Trade-off between accuracy and complexity
  - Dependency between type I and type II errors

Example - Testing a Coin

- **Testing Steps**
  - **A coin is suspected to be biased, set up a** *null hypothesis* **stating that the coin is unbiased**
  - **Choose level of significance (α) and number of trials (12)**





$$P_k = \frac{n!}{(n-k)!k!} P^k (1-P)^{n-k}, n = 12, p = 0.5$$

Example - Testing a Coin

Set $\alpha = 0.01$

$$P_0 = \frac{12!}{12!0!} P^0 (1-P)^{12} = P_{12} = 0.0002$$

$$P_1 = \frac{12!}{11!1!} P^1 (1-P)^{11} = P_{11} = 0.0029$$

$$P_2 = \frac{12!}{10!2!} P^2 (1-P)^{10} = P_{10} = 0.0161$$

With two-tails test

$$P\{2 \le k \le 10\} = 1 - P_0 - P_1 - P_{11} - P_{12} = 0.9936 > 0.99 = 1 - \alpha$$

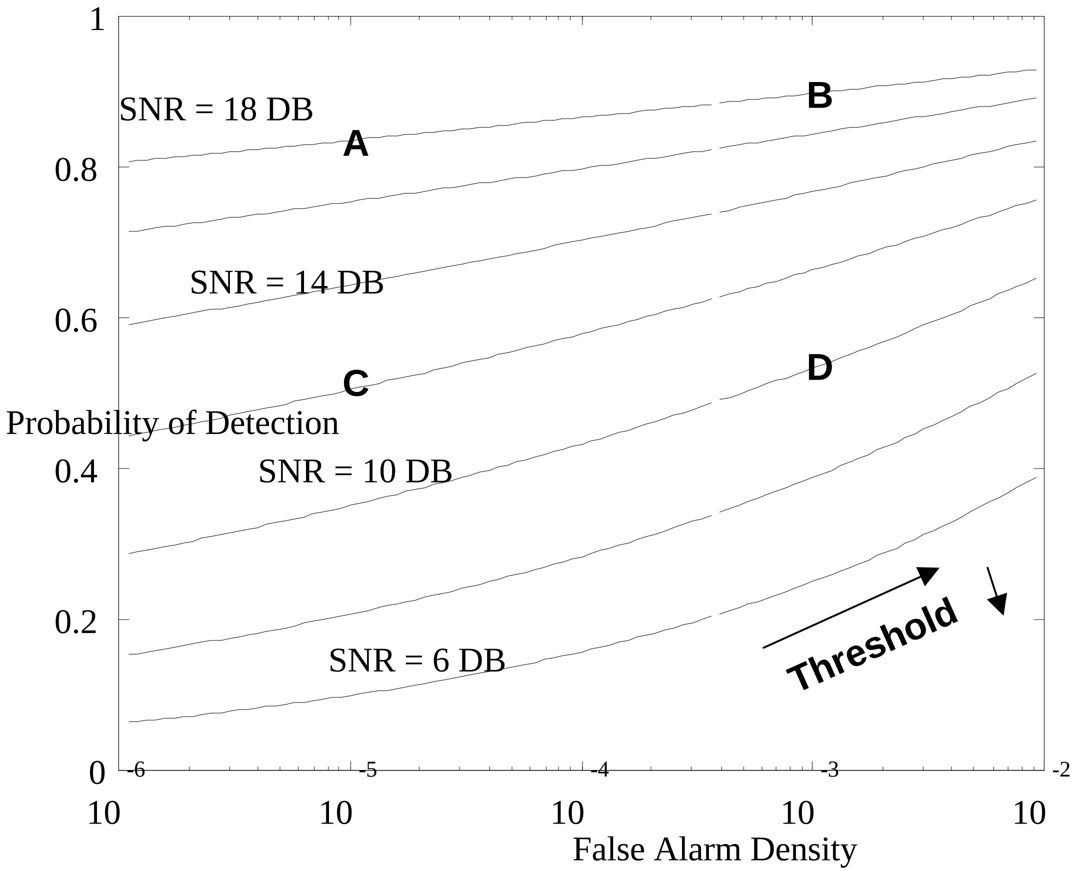
$$P\{3 \le k \le 9\} = 1 - P_0 - P_1 - P_2 - P_{10} - P_{11} - P_{12} = 0.9614 < 0.99 = 1 - \alpha$$

Decision: Reject the null hypothesis when either 11 or 12 heads or tails are obtained. This decision is subject to a Type I error

Type II error (actually biased but detect no bias) is more difficult to determine

Example - Sensor Systems

System Operating Characteristic



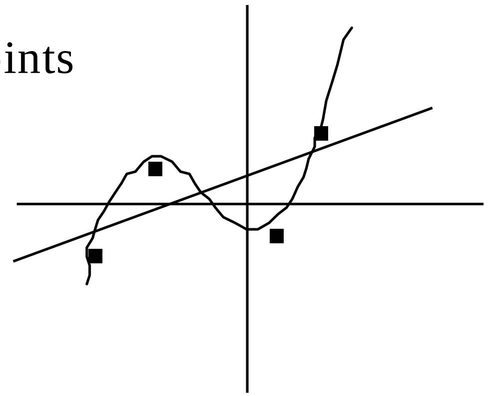
Probability of Detection vs. False Alarm Density **(SNR = 10log (S / N))**

Performance Analysis

- **Operating Points**
  - **Lower detection threshold, increase PD, increase Pfa, A -> B**
  - **Smaller SNR, same PD, increase Pfa, C -> D**
  - **Smaller SNR, lower PD, same Pfa, B -> D**
- **Performance**
  - **Operating point A (PD = 0.85, Pfa = 10^-5)**
    - **Type I error = 1-PD = 0.15, Type II error = Pfa**
  - **Operating point B (PD = 0.9, Pfa = 10^-3)**
    - **Type I error = 1-PD = 0.1, Type II error = Pfa**
- **Selecting an Operating Point**
  - **Trade-off between detection probability and and false alarm probability**
  - **Dependency between type I and type II errors**
  - **Type I: a target exists, but miss the detection**
  - **Type II: no target exists, but falsely detected**

Linear Regression

Linear regression: to identify linear relationship between two variables



Fit the straight line $y = a + bx$ through the given points $(x_1, y_1), \ldots, (x_n, y_n)$ such that

$$q = \sum_{j=1}^{n} (y_j - a - bx_j)^2 \text{ is minimized}$$

necessary conditions:

$$\left.\begin{aligned} \frac{\partial q}{\partial a} &= -2\sum_{j=1}^{n} (y_j - a - bx_j) = 0 \\ \frac{\partial q}{\partial b} &= -2\sum_{j=1}^{n} x_j (y_j - a - bx_j) = 0 \end{aligned}\right\} \quad \begin{aligned} an + b\sum x_j &= \sum y_j \\ a\sum x_j + b\sum x_j^2 &= \sum x_j y_j \end{aligned}$$

$$\Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_j \\ \sum x_j y_j \end{bmatrix}$$

Curve Fitting

Fit a polynomial of degree m: $p(\mathrm{x}) = b_0 + b_1 x + \cdots + b_m x^m$

through the given points $(x_1, y_1),\ldots,(x_n, y_n)$ where $m \le n - 1$

then $q = \sum_{j=1}^{n} (y_j - p(x_j))^2$

necessary conditions: $\dfrac{\partial q}{\partial b_0} = 0,\ldots,\dfrac{\partial q}{\partial b_m} = 0$

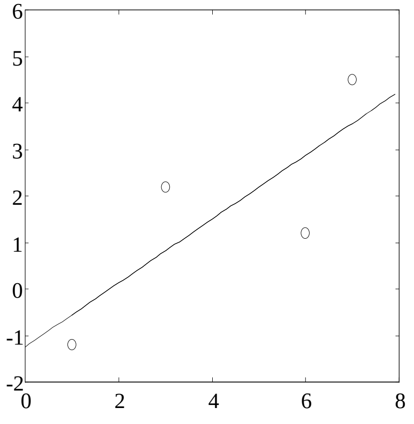
Example: 2nd order polynomial $p(\mathrm{x}) = b_0 + b_1 x + b_2 x^2$

ICBES that $\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_j & \sum x_j^2 \\ \sum x_j & \sum x_j^2 & \sum x_j^3 \\ \sum x_j^2 & \sum x_j^3 & \sum x_j^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_j \\ \sum x_j y_j \\ \sum x_j^2 y_j \end{bmatrix}$

MATLAB Tools for Curve Fitting

- **Least Squares**
  - given a set of data
  - find a curve that is closest to the points
  - minimizing sum of squared distance from each point to the curve
  - linear or polynomial (nonlinear) regression
  - coef = polyfit (x , y , n)

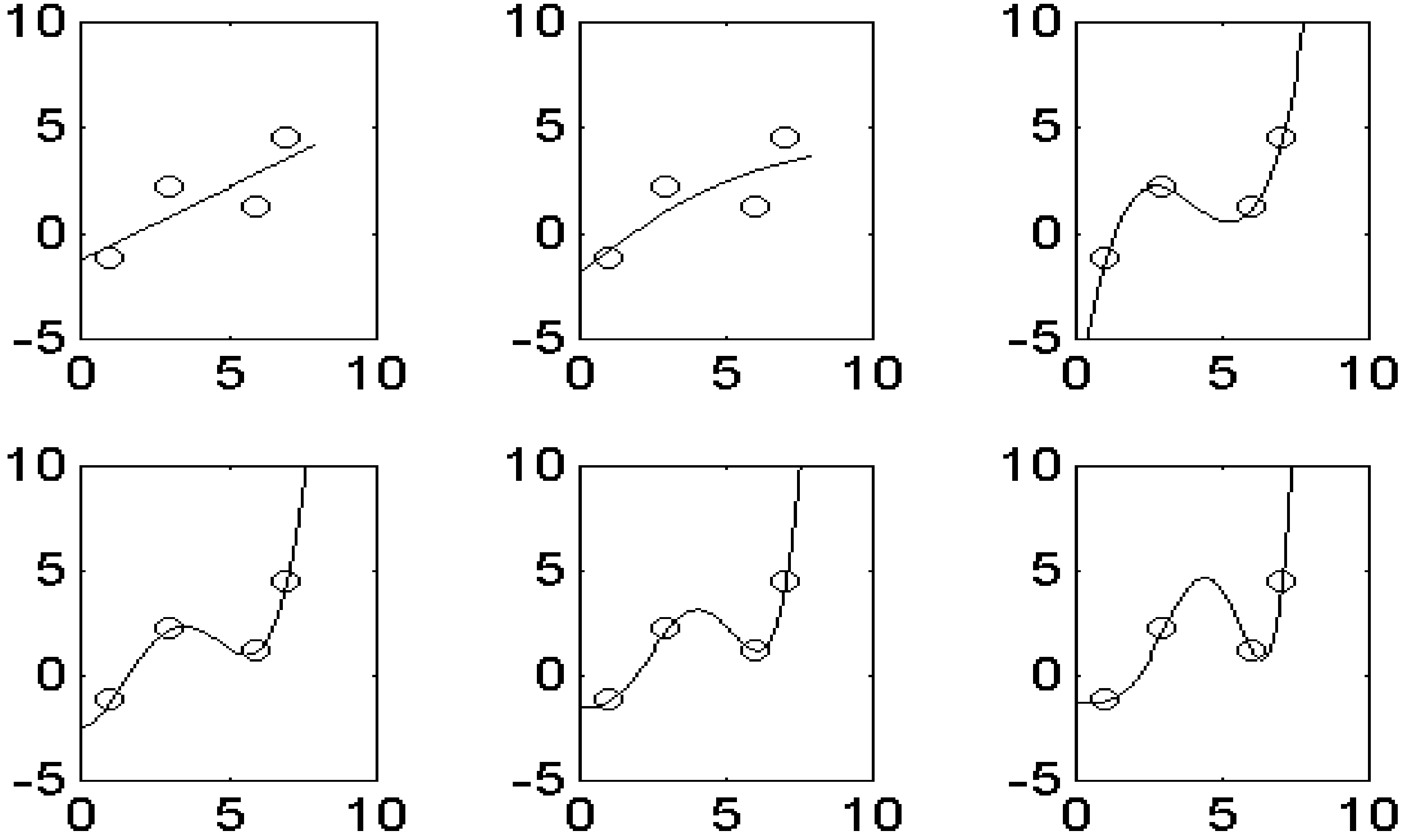
$$y = f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$$




```
x = [ 1, 3, 6, 7 ];
y = [ -1.2, 2.2, 1.2, 4.5 ];
coef = polyfit ( x, y, 1 );
newx = 1 : 0.1 : 7;
newy = polyval ( coef, newx );

plot ( x, y, ‘o’, newx, newy, x, y )
```

Polynomial Fits Examples



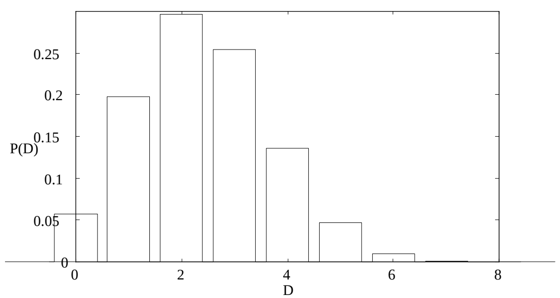
Monte Carlo Simulation

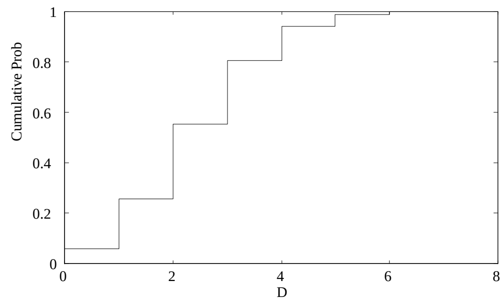
- **Motivation**
  - Uncertain decision environment requires probabilistic models
  - Formal mathematic solutions may be difficult or impossible
  - Monte Carlo analysis provides a powerful tool through simulation
- **Applications**
  - System performance evaluation
  - System operation characteristics analysis

Lead Time Demand Example

$$\boldsymbol{L : Lead\ Time,}\ f(L) = 0.25\,e^{-0.25L} \Rightarrow F(L) = \int_0^L f(L)\,dL = 1 - e^{-0.25L}$$

$$\boldsymbol{D : Demand,}\ P(D)$$

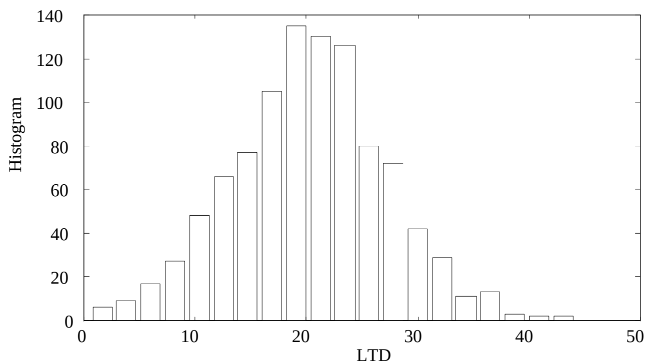




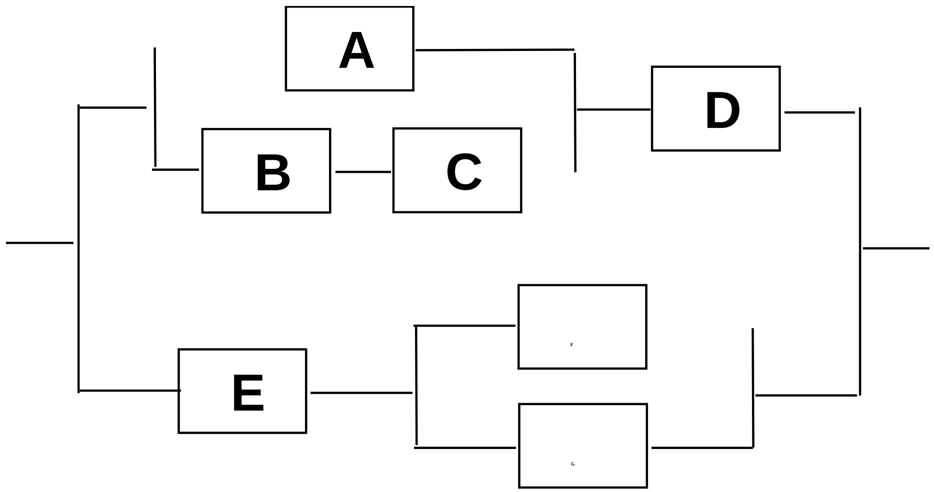
Simulation Results: $(LTD = \sum_L D)$

$Trial$	$L(rounded)$	D	LTD
1	5	3,1,2,4,6	16
2	4	6,2,3,4	15
3	7	3,1,2,4,3,2,5	20
4	2	2,4	6
$\vdots$	$\vdots$	$\cdots$	$\vdots$

$$\text{Average LTD} = \sum_i LTD_i$$



Reliability Example



Given component reliabilities:

P(A)=P(B)= P(E)=0.7, P(C)=P(D)=0.9, P(F)=P(G)=0.8

What is the system overall reliability?

One approach is to use Monte Carlo simulation

serial connection: AND logic
parallel connection: OR logic