Independence of Events

ECE 313

Probability with Engineering Applications

Lecture 4

Professor Ravi K. Iyer

Department of Electrical and Computer Engineering

University of Illinois

Today's Topics and Announcements

- Independence of Events
- Independence vs. Mutual Exclusivity
- Mini Project 1 Description
- Announcements:
  - Mini Project 1, Tasks 0 and 1 are posted today. Remainder will be posted on Feb. 5th
  - Mini Project 1 Task 0 and Task 1 (individual): due on Monday, Feb 6th, 10:00 AM.
  - Project groups will be announced on Monday, Feb 6th .
  - Mini Project 1 Tasks 2-3 (group): due on Friday, Feb 10th, 11:59 PM.
  - Have questions? Post on Piazza
  - Homework 2 is released today, but it will not be graded.
  - There will be a group activity next Wednesday, Feb 6th , in class and it will be graded for each student.

Independence of Events

- We define two events A and B to be **independent** if and only if:

  P(A|B)=P(A)

- From the discussion of conditional probability [provided P(A) $\neq$ 0 and P(B)) $\neq$ 0]:

  $$P(A \cap B) = P(A)P(A \mid B) = P(B)P(A \mid B)$$

- This leads to the following **usual definition of independence**:

  Events A and B are said to be independent if:

  $$P(A \cap B) = P(A)P(B)$$

  Such events are also referred to as "stochastically independent events" or "statistically independent events."

- Note: If A and B are not independent, then $P(A \cap B)$ is computed using the multiplication rule.

A Simple Example: A CPU Board & Memory

A CPU board consists of a microprocessor CPU chip and a random access main memory chip. The CPU is selected from a lot of 100, of which 10 are defective and the memory chip is selected from a lot of 300, of which 15 are defective.

Define A to·be the event "The selected CPU is defective," and
Define B to be the event "The selected memory chip is defective."

Then P(A)=10/100= 0.1, and P(B)=15/300=0.05. Since the two chips are selected from different lots, we may expect the events A and B to be independent.

Check: since there are 10 x15 ways of choosing both defective chips, and there are 100 x 300 ways of choosing any two chips. Thus:

$$P(A \cap B) = \frac{10 \cdot 15}{100 \cdot 300} = 0.005 = 0.10 \cdot 0.05 = P(A)P(B).$$

Some Important Points about the Concept of Independence

- If A and B are two mutually exclusive events, then $A \bigcap B = \emptyset$, (Null Set) which implies $P(A \bigcap B) = 0.$ Now, if they are independent as well, then either P(A)=0 or P(B)=0.
- If an event A is independent of itself, i.e., if A and A are independent, then P(A)=0 or P(A)=1. The assumption of independence yields $P(A \bigcap A) = P(A)P(A)$ or $P(A) = P(A)^2$.

Some Important Points about the Concept of Independence (cont.)

- If the events A and B are independent, then so are events $\overline{A}$ and B, events A and $\overline{B}$, and events $\overline{A}$ and $\overline{B}$. Recall that $A \bigcap B$ and $\overline{A} \bigcap B$ are mutually exclusive events whose union is B, i.e.,

  $$P(B) = P(A \bigcap B) + P(\overline{A} \bigcap B) = P(A)P(B) + P(\overline{A} \bigcap B)$$

  since A and B are independent.

  This implies

  $$P(\overline{A} \bigcap B) = P(B) - P(A)P(B) = P(B)[1 - P(A)] = P(B)(P(\overline{A}).$$

- The independence of A and $\overline{B}$ and $\overline{A}$ and $\overline{B}$ can be shown similarly.
- The concept of independence of two events can be extended to a list of *n* events.

Pairwise Independent Events that are not Independent

- Example (Pairwise Independent Events That Are Not Independent) Let a ball be drawn from an urn containing four balls, numbered 1, 2, 3, 4. Let E= {1,2}, F={1,3}, G={1,4}. If all four outcomes are assumed equally likely, then

$$P(EF) = P(E)P(F) = \frac{1}{4},$$

$$P(EG) = P(E)P(G) = \frac{1}{4},$$

$$P(FG) = P(F)P(G) = \frac{1}{4}$$

- Note $\frac{1}{4} = P(EFG) \neq P(E)P(F)P(G)$

  Hence, even though the events E,F,G are pairwise independent, they are not jointly independent.

Example 1: Demonstrating Dependence

- Suppose that we toss 2 fair dice. Let E_1 denote the event that the sum of the dice is 6 and F denote the event that the first die equals 4. Then

$$P(E_1F) = P(\{(4,2)\}) = \frac{1}{36}$$

$$P(E_1)P(F) = \left(\frac{5}{36}\right)\left(\frac{1}{6}\right) = \frac{5}{216}$$

Hence, E_1 and F are not independent.

- Intuitively, if we are interested in the possibility of throwing a 6 (with 2 dice), if the first die lands on 4 (or, on any of the numbers 1,2,3,4 and 5), we have a possibility of getting a total of 6.
- If the first die landed on 6, we no longer have a chance of getting a total of 6. i.e. the chance of getting a total of 6 depends on the outcome of the first die, thus E_1 and F cannot be independent.

Example 1 continued

- Now suppose that we let E_2 be the event that the sum of the dice equals 7. Is E_2 independent of F ? The answer

, since

$$P(E_2F) = P(\{(4,3)\}) = \frac{1}{36}$$

whereas

$$P(E_2)P(F) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$$

Example 3

- Let E denote the event that the next president is a Republican and F the event that there will be a major earthquake within the next year, then most people would probably be willing to assume that E and F are independent.

- There would probably be some controversy over whether it is reasonable to assume that E is independent of G, the event that there will be a recession within two years after the election

- Next, if E is independent of F, then E is also independent of F^c
- If E is independent of F, then E is also independent of F^c

Proposition

- if E and F are independent, then so are E and F^c
- Assume that E and F are independent. Since $E = EF \cup EF^c and EF^c$ *EF* are obviously mutually exclusive, we have

$$P(E) = P(EF) + P(EF^c)$$
$$= P(E)P(F) + P(EF^c)$$

or equivalently, $P(EF^c) = P(E)[1 - P(F)]$

$$= P(E)P(F^c)$$

- This, if E is independent of F, then the probability of E's occurrence is unchanged by information as to whether or not F has occurred.
- Suppose now that E is independent of F and is also independent of G
- Is E then necessarily independent of FG? The answer is no as the following example demonstrates

Example: Extending to three Events

- Two fair dice are thrown. Let E denote the event that the sum of the dice is 7. Let F denote the event that the first die equals 4 and G denote the event that the second die equals 3.
- From Example 1, we know that E is independent of F, and the same reasoning as applied there shows that E is also independent of G; but E is not independent of FG

  [since $P(E|FG)=1$].

- It follows that an appropriate definition of the independence of three events, E, F, and G would have to go further than merely assuming that all of the pairs of events are independent. We are thus led to the following definition.

Independence of E,F,G

- Three events, E, F, and G are said to be independent if

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G)$$
$$P(FG) = P(F)P(G)$$

- Note that if E, F, and G are independent, then E will be independent of any event formed from F and G. For instance, E is independent of $F \cup G$, since

$$\begin{aligned} P[E(F \cup G)] &= P(EF \cup EG) \\ &= P(EF) + P(EG) - P(EFG) \\ &= P(E)P(F) + P(E)P(G) - P(E)P(FG) \\ &= P(E)[P(F) + P(G) - P(FG)] \\ &= P(E)P(F \cup G) \end{aligned}$$

Independence of n Events

- We can also extend the definition of independence to more than three events. The events $E_1, E_2, \ldots, E_n$ are independent if for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}, r \leq n$ of these events,

$$P(E_{1'}, E_{2'}, \ldots, E_{r'}) = P(E_{1'})P(E_{2'})\ldots P(E_{r'})$$

- Finally, we define an infinite set of events to be independent if every finite subset of those events is independent.

Pairwise independent but not mutually independent

- Consider the above experiment of tossing two dice. Let

  A = “The first die results in a 1, 2, or 3.”

  B = “The second die results in a 4, 5, or 6.”

  C = “The sum of the two faces is 7.”

- Then : $A \cap B = \{(1,4),(1,5),(1,6),(2,4),(2,5),(2,6),(3,4),(3,5),(3,6)$
- And: $A \cap C = B \cap C = A \cap B \cap C = \{(1,6),(2,5)(3,4)\}$
- Therefore:

$$P(A \cap B) = \tfrac{1}{4} = P)A)P(B)$$

$$P(A \cap C) = \tfrac{1}{12} = P(A)P(C)$$

$$P(B \cap C) = \tfrac{1}{12} = P(B)P(C)$$

- But: $P(A \cap B \cap C) = \tfrac{1}{12} \neq P(A)P(B)P(C) = \tfrac{1}{24}$
- In this example, events A, B, and C are pairwise independent but not mutually independent.

Another Dice Example: Pairwise and Mutually independent

- Consider the experiment of tossing two dice. Let the sample space $S = \{(i,j) | 1 \le i,j \le 6\}$. Also assume that each sample point is assigned a probability 1/36. Define the events A, B, and C so that:

A = "First die results in a 1, 2, or 3."

B = "First die results in a 3, 4, or 5."

C = "The sum of two faces is 9."

- Then $A \cap B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$,
- $A \cap C = \{(3, 6)$,
- $B \cap C = \{(3, 6), (4, 5), (5, 4)$, and
- $A \cap B \cap C = \{(3, 6)\}$.

Dice Example (cont.)

- Therefore:

$$P(A \cap B) = 1/6 \neq P(A)\, P(B) = 1/4,$$

$$P(A \cap C) = 1/36 \neq P(A)\, P(C) = 1/18,$$

$$P(B \cap C) = 1/12 \neq P(B)\, P(C) = 1/18,$$

- But:

$$P(A \cap B \cap C) = 1/36 = P(A)\, P(B)\, P(C).$$

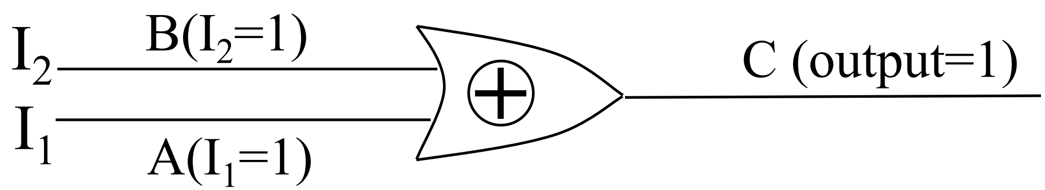
- If the events A1, A2, …, An, are such that every pair is independent, then they are called **pairwise independent**. It does not follow that the list of events is **mutually independent**.

Physical vs. Stochastic Independence

- It may be reasonable to assume that two events are physically independent.
- Example: Coin tossing (One toss does not influence another.)
- Other examples:
  - Arrivals of jobs to a computer system
  - Disk access
  - Phone calls arriving at an exchange
- Physical independence is usually used to assert stochastic independence.
- This assertion can be tested by calculating the relative frequencies (making experimental estimates of probabilities).

Physical vs. Stochastic Independence (cont.)

- Stochastic/statistical independence does not imply physical independence.
- Example:



- Assume that A,B are independent and that P(A)=P(B)=1/2
- C => event, output = 1
- Questions: Are A and C independent? Can you justify/explain the answer?

Indendent vs. mutually exclusive Card Example

- Are 'red card' and 'spade' independent?

$$0 \neq \frac{1}{2}$$

Therefore not independent

- Are they Mutually exclusive?

$$0 = 0$$

Therefore, they are Mutually Exclusive (Disjoint)

Card Example (cont.)

- Are ‘red card’ and ‘ace’ independent?

$$\frac{1}{2} = \frac{1}{2}$$

Therefore they are independent

- Are they Mutually exclusive?

$$\frac{2}{52} \neq 0$$

Therefore, not Mutually Exclusive (not Disjoint)

Mini Project 1

Analysis of Alarm Data from ICU Patient Monitoring Systems

ECE 313

ICU Patient Monitoring Systems

Pictures from: http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1271&context=gs_theses

Sample Logs

- Here is a sample of logs collected from central monitors in an ICU on a given day:

@2530127@ * 00:**01:23**.125 0780:0788 alm_rec.exe **"9NI953: SPO2 LO 86"**
@2530128@ * 00:**01:25**.125 0780:0788 alm_rec.exe **"9NI953: SPO2 LO 87"**
@2530129@ * 00:**01:27**.125 0780:0788 alm_rec.exe **"9N953: SPO2 LO 88"**
@2530130@ * 00:**01:27**.453 0748:0772 audio.exe **MCI_PLAY:ADVISORY**
@2530131@ * 00:**01:29**.125 0780:0788 alm_rec.exe **"9NI953: SPO2 LO 89"**
@2530132@ * 00:01:30.937 0748:1300 audio.exe "9NORTHI953" (display slot 7) now NO_AUDIO_ALARM audio level
@2530133@ * 00:**01:30**.937 0748:0772 audio.exe **MCI_STOP**
@2530134@ * 00:**01:30**.937 0748:1300 audio.exe **Alarms silenced**
@2530135@ * 00:01:41.140 0780:0788 alm_rec.exe **"9NI953: BRADY"**
@2530136@ * 00:01:41.562 3560:3572 ddw.exe GRAPHING BED - 9NORTHI953
@2530137@ * 00:01:41.562 3560:3572 ddw.exe Added status line message 'GRAPHING BED - 9NORTHI953',0 (1)
@2530138@ * 00:01:41.812 3560:3580 ddw.exe (GLL) GRAPHING BED - …
@2530139@ * 00:01:41.843 2920:3044 fdsvr.exe get_admit_pkt::openUNITY() failed (Err = 0) to **"9NORTHI940"**
@2530140@ * 00:01:42.578 3560:3592 ddw.exe >>New DDW status = 80
@2530141@ * 00:**01:42**.937 0748:1300 audio.exe **"9NI953"** (display slot 7) now WARNING_ALARM audio level
@2530142@ * 00:**01:42**.937 0748:1300 audio.exe WARNING_ALARM sounding - volume 70 %
@2530143@ * 00:**01:42**.937 0748:0772 audio.exe **MCI_PLAY:WARNING**
@2530144@ * 00:01:44.937 0748:1300 audio.exe "9NORTHI953" (display slot 7) now NO_AUDIO_ALARM audio level
@2530145@ * 00:**01:44**.937 0748:0772 audio.exe **MCI_STOP**
@2530146@ * 00:**01:44**.937 0748:1300 audio.exe **Alarms silenced**

Analysis of Alarm Data

- Suppose that you are an analyst working at the company ACME
- Your duties consist in analyzing the alarm data collected by the patient monitoring system deployed by ACME
  - i.e., how frequently each alarm happened and what were the common causes for it
- Assume that you wrote an script to process the log files extracted from central monitoring system and structured them into the following format:

Alarm Type	Alarm Cause	Bed No.	Start Time	Stop Time
WARNING	HA_BRADY	953	'00:00:40.953'	'00:00:42.937'
ADVISORY	LOW_OXY_SAT	953	'00:00:48.937'	'00:00:54.937'
ADVISORY	LOW_OXY_SAT	953	'00:01:27.453'	'00:01:30.937'
CRISIS	LOW_OXY_SAT	952	'00:01:27.453'	'00:01:30.937'
WARNING	HA_BRADY	953	'00:01:42.937'	'00:01:44.937'
SYSTEM	LEADS_FAILURE	940	'00:02:45.953'	'00:02:46.953'

Alarm Data Fields

- Each row corresponds to an alarm raised at the central monitor
- **Alarm Type** indicates the severity of alarm and might be a patient or system related alarm:
  - **CRISIS**: Life-threatening
  - **WARNING**: Serious but not life-threatening
  - **ADVISORY**: Events that required monitoring, but not serious or life-threatening
  - **SYSTEM:** System status alarms that are triggered by network or equipment problems.
- **Alarm Cause** indicates the cause of each alarm.
  - The patient alarm causes are related to physiological symptoms, e.g. BRADY corresponds to a very low heart rate event.
  - The system alarm causes could be related to artifacts, sensor/lead failures, network errors, or application errors.

Alarm Data Fields (Cont.)

- **Bed No.** indicates the ID of the patient bed for which the alarm is raised.
- **Start Time** indicates the time that the alarm was asserted at the central monitor.
- **End Time** indicates the time that the alarm was stopped or silenced at the central monitor.

Mini Project 1: Task 0

With the recent change in licensing, most of the students are no longer able to download MATLAB through webstore. Here we suggest two options:

(1) To work on your project in the EWS lab (hours and locations)

(2) Connect remotely through CITRIX
(guide provided in the tutorial section of this description)

- Explore the Matlab software environment and its basic commands by reading provided tutorials on the class website:
https://courses.engr.illinois.edu/ece313/SectionG/projects.html

Mini Project 1: Task 0

This is a preparatory task:

- Import the dataset into Matlab
  - Dataset is in a .mat format. Check the skeleton code provided to you.
- Split the data into different datasets
  - Make two different subsets corresponding to “SYSTEM" and “WARNING” alarm types.
  - See the tutorial on how to split the data sets.
  - Take a screenshot of your datasets.
  - You will use these datasets in Task 2.

Mini Project 1: Task 1

1. Calculate the probability of:
   - alarm types, over all the data: SYSTEM, CRISIS, ADVISORY, WARNING
   - the three most frequent alarm causes and list them.

2. Given the bed number, calculate the conditional probability for each of the four alarm types. Provide your result in a table (bed number vs. alarm type).

3. Calculate the frequency and plot the histograms for all alarms by each hour of the day.

   Use *datenum()* and *datestr()* function to parse "Start Time" fields and count the number of alarms raised in each hour.

   Use *hist()* function to plot histogram of alarms per hour (24 bins). Note that *hist()* will count the frequency of the alarms. Hence you only need to find the hour for each alarm and make an array equal to the size of all alarms you have.

Mini Project 1: Task 2

1. Given the alarm types (e.g., SYSTEM, WARNING), calculate the conditional probability of each of the following alarm causes (use the split data from Task 0): *(e.g., P (APP_ERR | SYSTEM)*
   a. For SYSTEM alarms: APP_ERR, ARTIFACT, LEADS_FAILURE, NW_ ERR
   b. Compare P(APP_ERR | SYSTEM) and P(SYSTEM|APP_ERR). Are they the same? If not, why?
2. Using the results from (1), calculate the probability of P(APP_ERR $\cap$ SYSTEM). Make sure your show your calculation / steps
3. Analyze the duration of each alarm:

   Add a column to the data table showing the duration of each alarm

   Use *array2table()* function to convert the array of results to a table.

   a) Calculate the average duration of each alarm type.
   b) Calculate the average duration of alarms for each hour.
   c) Use the *bar* function to plot the average duration of alarms per hour for the whole day (24 hours).

Mini Project 1: Task 3

We want to find patients in serious conditions.

1. An example of a metric is the count of the total number of alarms per patient. Physicians think that this metric is too simple to be effective. *Propose your own metric for identifying patients in a serious condition using results from Task 1 and Task 2. Justify how such a metric is effective.*

2. Using the metric from (1), find the top two patient beds in a serious condition and analyze the data for the two patient beds: Calculate and compare their alarm types, alarm causes, and average duration of alarms.

***Hint**: you can reuse the codes from Task 1*

3. Compare the data of the two patient beds from (2)

    a. What similarities/difference do you find from the two patient beds?

    - describe your observations in three bullets
    - simple observations, such as the total number of alarms will not be sufficient

    b. For each of the two patients, how does the metric change over time (per hour of the day)

Project Timeline and Grading

- Tasks 0 and Task 1 (individual tasks): **due on Monday,** Feb 6th, 10:00AM.
  - Submit the your answers for Task1 and a working MATLAB script
- Tasks 2 and Task 3 (group tasks): **due on Friday,** Feb 10th, 11:59 PM.
  - Submit the your report and a working MATLAB script
- Grading
  - **Task 0, 1:** 40%
  - **Task 2:** 30%
  - **Task 3:** 30%

Project Timeline and Grading

- **We will announce the groups after Task 1 due date (Feb 6th)**
- **Extra office hours are on Friday, Feb 3rd (Saboo: 11am-noon, Phuong: 4pm-5pm) in CSL 249.**
- **Our strong preference is to use Matlab. If you have sufficient proficiency, you could use Python but with minimal help from the TAs.**
- **Checkout the class website for description of system, tasks, and data set**: https://courses.engr.illinois.edu/ece313/sp2017/sectionG/projects.html

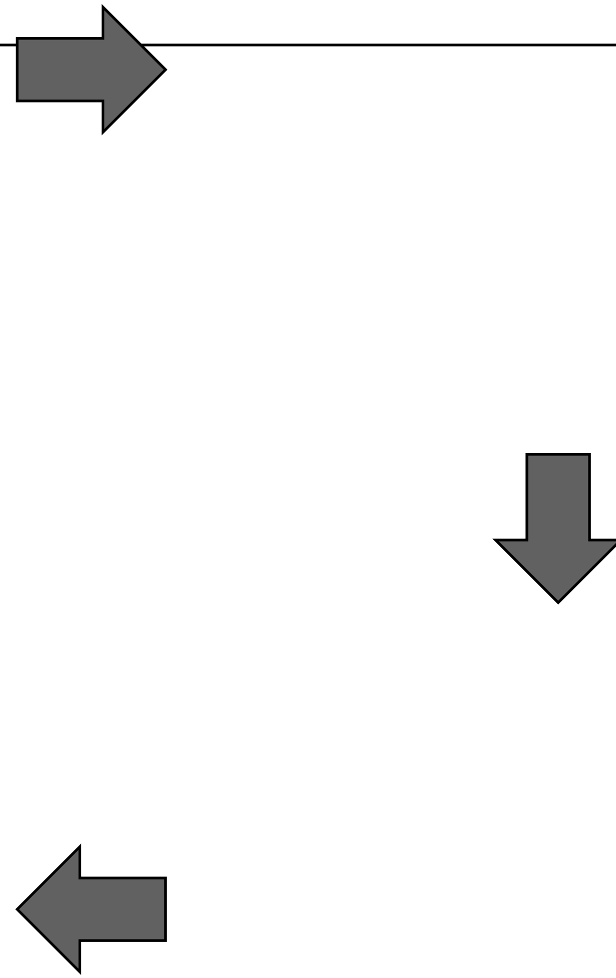
Project Submission

- A **report** describing your work and results **and your Matlab code** must be **submitted through Compass** by the due date
- **Matlab code Instructions:**
  - We provide you with a skeleton file to start from.
  - Fill in the code related to each task in the sections indicated by comments.
  - *fprintf()* functions are provided for structured results. Please replace the variables accordingly to report your results.
  - TAs will execute your code and check the results file. Non-executable code will not be graded.
- **Report Instructions:**
  - Format your report to separate different tasks in separate sections.
  - Copy results generated by *fprintf()* functions into corresponding sections in report.
  - Explain all your work and assumptions made for deriving your results.
  - **Provide insights on the results** generated for each task, in a few bullets.
  - Incorrect final answers without any explanations or comments receive zero credit.
  - Non-readable reports will be returned without a grade.
  - **Include your names, group name.**
  - **Describe the contribution of each member (Who did what task/problem)**

HOWTO: MATLAB through CITRIX

1. **Following "Connecting to CITRIX", install and run CITRIX**
   Note that you have to either
   (1) be on the campus network or
   (2) connected to the campus network through VPN
   (guide to set up the VPN client)
2. **Log in using your netID/password**
3. **Click the + icon on your left**
4. **Browse through 'EWS Lab Software > MATLAB R2015a'**
5. **Clicking on MATLAB will add the application to your CITRIX main page**

Now, you can run MATLAB from your own machine



MATLAB Tutorial

- Files:
  - Function and codes: *.m*
  - Files: *.mat*
- Clear all the workspace environment:
  - *clear all*
- Clear the command window:
  - *clc*
- You can repeat the previous commands by arrow keys or from the command history window

- Import CSV file into MATLAB as a *table*:
  - *T = readtable('myCsvTable.csv')*

MATLAB Tutorial (Cont.)

- Define variables in the command window:
  - *x = 1*
  - *y = X + Z*
  - z = cos(1)
- To define arrays:
  - *T = [1, 2, 3, 4, 5]*
  - *T = 1:5*
  - *W = [1, 1.1, 1.2, 1.3, 1.4, 1.5]*
  - *W = 1:0.1:1.5*
  - *Labels = {'Heart Rate', 'Blood Pressure', 'Electrocardiogram'}*
  - *T = [1, 2, 3; 4, 5, 6]* *(An array of 2 rows and 3 columns)*
- To see the values of variables, just type their name:
  - *x*
  - *W(1)*
  - *T(2:3)*

MATLAB Tutorial (Cont.)

- Define a vector of zeros or ones:
  - *zeros(1,10)* *(A vector of 10 columns, all initialized to 0)*
  - *ones(2,3)* *(An array of 2 rows and 3 columns, all initialized to 1)*
- Find the size of a vector or array:
  - *size(T)*
  - *size(W, 2)*
- Find the indices of interest in a vector or array:
  - *ind = find(X) => locates all nonzero elements of array X*
- Plotting a variable
  - *plot(x)*
  - *plot(t, x)*
  - *bar(x)*
- Plotting a distribution or frequencies
  - *hist()*

MATLAB Tutorial

- **Cell Arrays:**
  - A cell array is a data type with indexed data containers called cells, where each cell can contain any type of data. Cell arrays commonly contain either lists of text strings, combinations of text and numbers, or numeric arrays of different sizes.
  - **Convert cell array to numeric array**
    - *A= cell2mat(C)* converts cell array C with contents of the same data type into a single array, A.
  - **Sub-setting a cell array**
    - *indices = find(ismember(data.Alarm_Type, 'SYSTEM'))*

    Returns the indices to all the rows in the cell array *data* whose *Alarm_Type* field is equal to 'SYSTEM'
    - *subset = data(indices,:)*

    Returns a subset of cell array data containing rows indicated by indices

MATLAB Tutorial (Cont.)

- **Find a list of distinct values in an array or cell array:**
  - *unique(x)*
- **Converting date and time format:**
  - *dateString = '19-May-2001';*
    formatIn = 'dd-mmm-yyyy';
    datenum(DateString,formatIn)

- **Convert date and time to string format:**
  - *datestr('05:32 PM','HH:MM')   => '17:32'*
  - *datestr('05:32:10 PM','HH')   => '17'*

- **Find and replace substring:**
  - *modifiedStr* = `strrep(`*origStr, oldSubstr, newSubstr*)
  - Example: *strrep(data.StartTime,"''","")*

    Will remove single quotes from the Start Time field.

MATLAB Resources

- MATLAB Tutorials and Learning Resources:
  http://www.mathworks.com/academia/student_center/tutorials/launchpad.html
  - Watch short videos and look at MATLAB examples.
  - Read the user's guide.

- MATLAB help:
  - *help hist*
  - *doc hist*