

## Lecture 3: Review of Probability, MATLAB, Histograms

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN

Scribed by: Chad Conley and Ken Varghese

---

---

This is preliminary work and has not been reviewed by instructor. If you have comments about typos, errors, notation inconsistencies, etc., please email Chad Conley and Ken Varghese at [cconley8@student.gsu.edu](mailto:cconley8@student.gsu.edu), [kvarghese2@student.gsu.edu](mailto:kvarghese2@student.gsu.edu).

### 3.1 Introduction

This lecture covers a broad review of probability including Bernoulli, binomial, exponential, and Gaussian distribution. It also covers the use of some MATLAB commands helpful for completing Mini-project 1 as well as the use of histograms.

### 3.2 Basics of Probability

by: Ken Varghese

**Probability** The Science of analyzing events that may or may not happen. There are multiple models that are used to help provide theoretical outcomes that mirror real world events. Each has their strengths and weaknesses. We cover four models in class. In order to understand those models, we need to understand some terms.

### 3.3 Terms

by: Ken Varghese

- **Random Variables** are numeric variables that fall randomly. They are represented by a curly  $x$ .
- Functions determine the potential outcomes of random variables, and these can be represented as  $f(x)$ .
- The phrase **i.i.d.** stands for independently, identically distributed, and it states that all inputs to a function, such as coin flips, are independent of each other (they don't have any impact on previous or future trials), they all have the same probability  $P$ , and they can fall anywhere defined by the function.

### 3.4 Bernoulli Distribution

by: Ken Varghese

A **Bernoulli distribution** describes a simple 50/50 chance in probability, defined by

$$f(k; p) = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \end{cases} \quad (3.1)$$

in which each event has an equal chance of happening. There are a limited number of outcomes; the function is **discrete**.

### 3.5 Binomial Distribution

by: Ken Varghese

A **Binomial distribution** represents a function with two parameters  $n$  and  $p$ , where  $n$  is the number of individual Bernoulli-distributed experiments, and  $p$  is the probability of each one. In other words, a Binomial distribution will represent the outcome of a sequence of Bernoulli trials. The popular usage of this distribution is to check for statistical significance. It is defined as such:

$$f(x = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (3.2)$$

### 3.6 Exponential Distribution

by: Ken Varghese

The Exponential distribution represents an exponential function using lambda:

$$f(X) = \lambda e^{-\lambda x} \quad (3.3)$$

Where lambda is the slope of the curve (the function is in slope-intercept form).  $x$  would be the variable. As lambda increases, the slope gets more steep.

### 3.7 Gaussian/Normal distribution

by: Chad Conley

**Gaussian distribution** also known as **normal distribution** is a function that represents the distribution of variable as a symmetrical bell-curve. It is used to model things that we expect to see in real life. A random value has a higher probability of being closer to the mean of the data set. The probability density function of the normal distribution is represented by the following equation:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (3.4)$$

**Example 3.1** (Manipulating the probability density function.). Suppose:

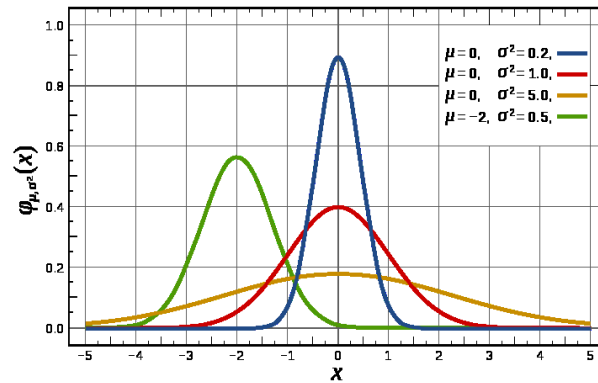
$$X \sim \mathcal{N}(\mu, \sigma^2). \quad (3.5)$$

Where:

$\mathcal{N}$  - Normal/Gaussian

$\mu$  - mu - mean

$\sigma^2$  - sigma squared - variance



The standard normal distribution is represented by the red curve. A change in  $\mu$  causes a lateral shift of the function. A change in  $\sigma^2$  changes the maximum value of the function where a larger variance equates a wider less focused distribution of data points.

## 3.8 MATLAB

by: Chad Conley

MATLAB commands useful to complete Mini-project 1:

<b>command</b>	<b>description of command</b>
<b>help</b> name	Displays the help text for the functionality specified by name, such as a function, method, class, toolbox or variable.
$X = \mathbf{rand}(m,n)$ $X = \mathbf{randn}(m,n)$	Returns an m-by-n matrix of random numbers. Returns an m-by-n matrix of normally distributed random numbers.
<b>figure</b> (n)	Finds a figure in which the Number property is equal to n, and makes it the current figure. If no figure exists with that property value, MATLAB creates a new figure and sets its Number property to n.
<b>hist</b> (x)	Creates a histogram bar chart of the elements in vector x.
<b>plot</b> (X,Y)	Creates a 2-D line plot of the data in Y versus the corresponding values in X.
<b>hold on</b>	Retains plots in the current axes so that new plots added to the axes do not delete existing plots.
$B = \mathbf{reshape}(A,sz)$	Reshapes A using the size vector, sz, to define size(B). For example, reshape(A,[2,3]) reshapes A into a 2-by-3 matrix. sz must contain at least 2 elements, and prod(sz) must be the same as numel(A).
$B = \mathbf{repmat}(A,n)$	Returns an array containing n copies of A in the row and column dimensions. The size of B is size(A)*n when A is a matrix.
$S = \mathbf{sum}(A,dim)$	Returns the sum along dimension dim. For example, if A is a matrix, then sum(A,2) is a column vector containing the sum of each row.
$k = \mathbf{find}(X)$	Returns a vector containing the linear indices of each nonzero element in array X.
$M = \mathbf{max}(A)$ $M = \mathbf{max}(A,[],dim)$	Returns the largest elements of A. Returns the largest elements along dimension dim. For example, if A is a matrix, then max(A,[],2) is a column vector containing the maximum value of each row.
$Y = \mathbf{abs}(X)$	Returns the absolute value of each element in array X.

## 3.9 Histograms

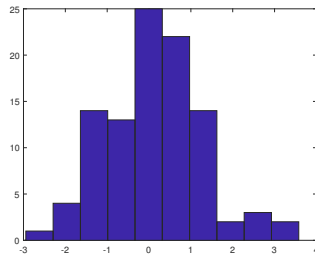
by: Chad Conley

**histogram** - the distribution of your sample represented as a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval.

**Example 3.2** (Creating a histogram in MATLAB). Suppose:

```
 $X = randn(1, 100);$   
 $hist(X);$ 
```

A histogram is generated in MATLAB of a 1x100 vector.



This can be used to help determine that a Gaussian distribution has been achieved by the `randn` function of MATLAB and can be compared to the following random generated distribution:

```
 $X = rand(1, 100);$   
 $hist(X);$ 
```

