Week 4
Concept of Probability

Mudrik Alaydrus
Faculty of Computer Sciences
University of Mercu Buana, Jakarta

mudrikalaydrus@yahoo.com

Introduction :

A Speech Recognition System
computer communication network

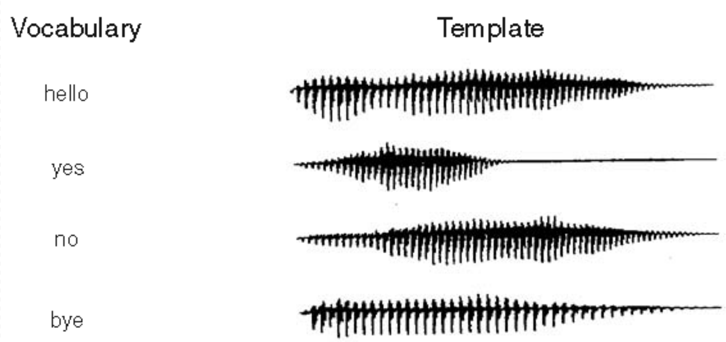
What we will learn today :

Concept of Probability

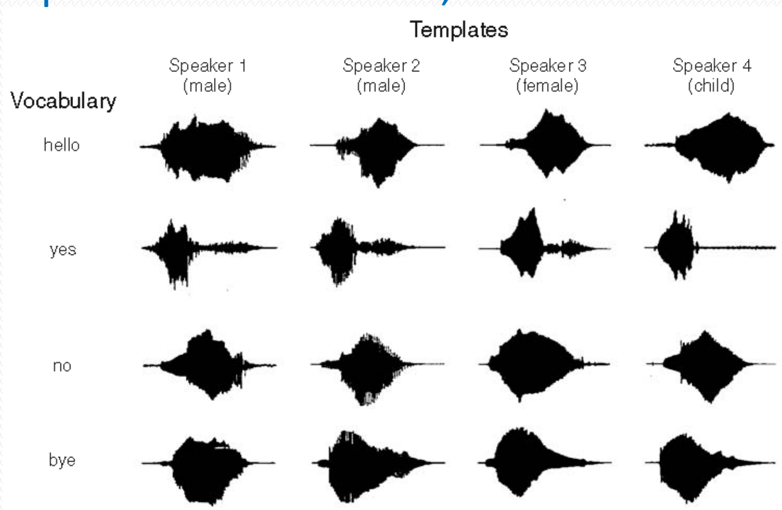
Distributions

A Speech Recognition System

Example: Recognizing commands spoken to a computer.
Method used Template matching :
Data base of vocabularies, or a set of possible words for a computerized dictionary.

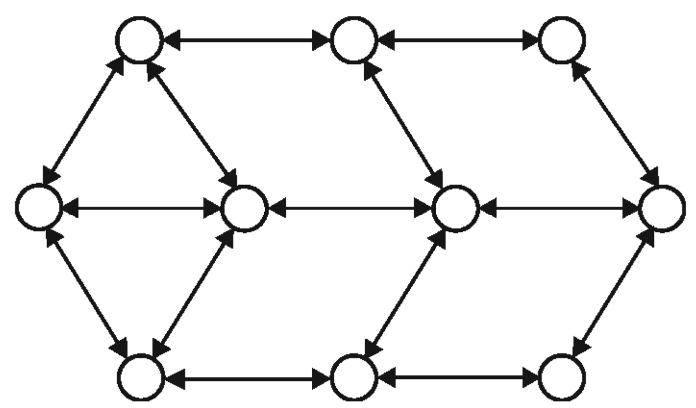


Speech recognition is a complicated task, because so factors that make this task so difficult : interference from the surroundings
variability in the amplitude and duration, etc.



Describing by the theory of probability and random processes.

Computer communication network



Restriction:

A node has a maximal rate of data transfer and a max. buffering capability

Job an network engineer:

Think how big a buffer must be, if certain traffic given

Or if the buffer given, what is the max. traffic

Sending information from a node to other node:

Router must determine the route from the source to the destination.

Some nodes in the network may be more congested than others.

The routing decision must be made using probability. Which route should the packet follow so that it is least likely to be dropped along the way?

Protocols for routing, flow control, and the like are all based in the foundations of probability theory.

Motivation

Probability is the mechanism by which we can manage the uncertainty that underlies all real world data and phenomena.

For example:

- To understand and use statistical hypothesis testing, one needs knowledge of the sampling distribution of the test statistic.

- To evaluate the performance (e.g., standard error, bias, etc.) of an estimate, we must know its sampling distribution.

- To adequately simulate a real system, one needs to understand the probability distributions that correctly model the underlying processes.

- To build classifiers to predict what group an object belongs to based on a set of features, one can estimate the probability density function that describes the individual classes.

Probability

A random experiment is defined as a process or action whose outcome cannot be predicted with certainty and would likely change when the experiment is repeated.

The ability to model and analyze the outcomes from experiments is at the heart of statistics.

Some examples of random experiments:

Engineering: Data are collected on the number of failures of piston rings in the legs of steam-driven compressors.

Medicine: The oral glucose tolerance test is a diagnostic tool for early diabetes mellitus.

Software Engineering: Engineers **measure the failure times in CPU seconds** of a command and control software system. These data are used to obtain models to **predict the reliability of the software system**

The sample space is the **set of all outcomes** from an experiment.
Examples of these sample spaces are:

• When observing piston ring failures, the sample space is {1,0}, where 1 represents a failure and 0 represents a non-failure.

• If we roll a six-sided die and count the number of dots on the face, then the sample space is {1,2,3,4,5,6}

The outcomes from random experiments are often represented by an uppercase variable such as *X.*
This is called a ***random variable,*** and its value is **subject to the uncertainty intrinsic to the experiment.**

Random variables can be **discrete or continuous**.
A ***discrete random variable*** can take on values from a finite or countably infinite set of numbers.
A ***continuous random variable*** is one that can take on values from an interval of real numbers.

Probability is a measure of the likelihood that some event will occur.

a way to quantify the likelihood that an observed measurement will take on values within some set or range of values.

Probabilities always range between 0 and 1.

Two classical methods for assigning probabilities:

- the equal likelihood model and
- the relative frequency method.

Equal likelihood model: Experiment with each of n outcomes is equally likely, probability mass **1/n** to each outcome.

Coin experiment ½, one die experiment 1/6 .

Relative frequency method: Conducting the experiment n times and record the outcome.

The probability of event E is assigned by **P(f)= f/n**,

where f denotes the number of experimental outcomes that satisfy event E.

Another way is to use **Probability Density Function (pdf)** denoted by **f(x)**

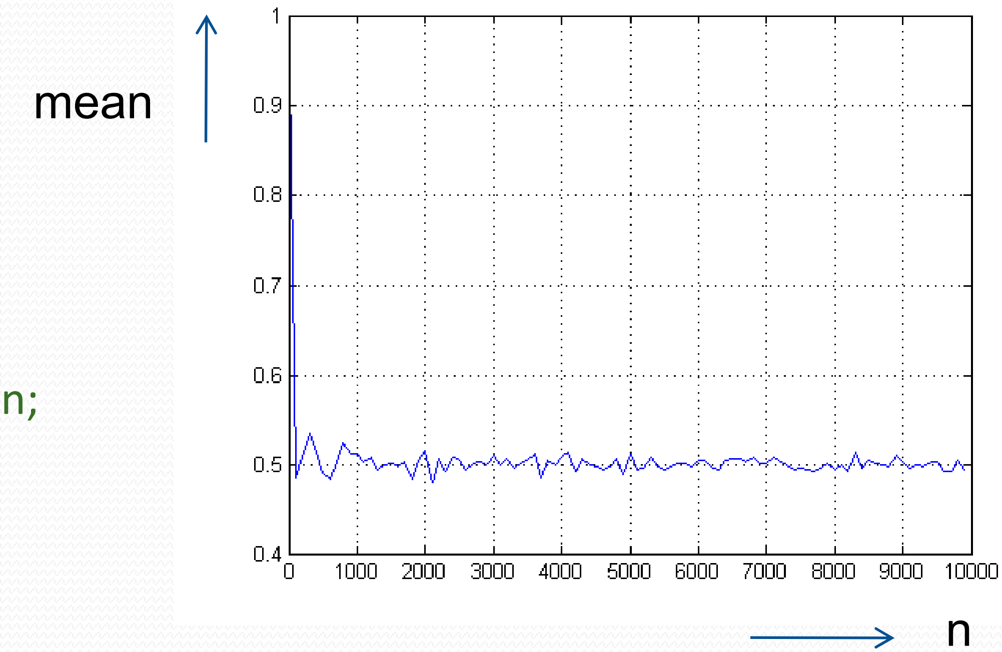
In Matlab we use the function rand(m,n)

This function generates a m rows and n column matrix, his elements are random numbers lying between 0 and 1.

By flipping a coin (coin experiment) we use 0 and 1, so that we can use round(rand(1))

If you repeat it n times, you will find about the half are zeros and the other half ones.

```
clear;
NN=1000
ii=0;
for n=1:10:NN
  ii=ii+1;
  N(ii)=n;
  y(ii)=sum(round(rand(n,1)))/n;
end
plot(N,y);grid
```



```
% Simulation of coin flipping
coin_flip=round(rand(1))

% Simulation of die tossing  (one die)
die_toss=ceil(6*rand(1))

% Simulation of die tossing  (one dice)
dice_toss=ceil(6*rand(1,2))
```

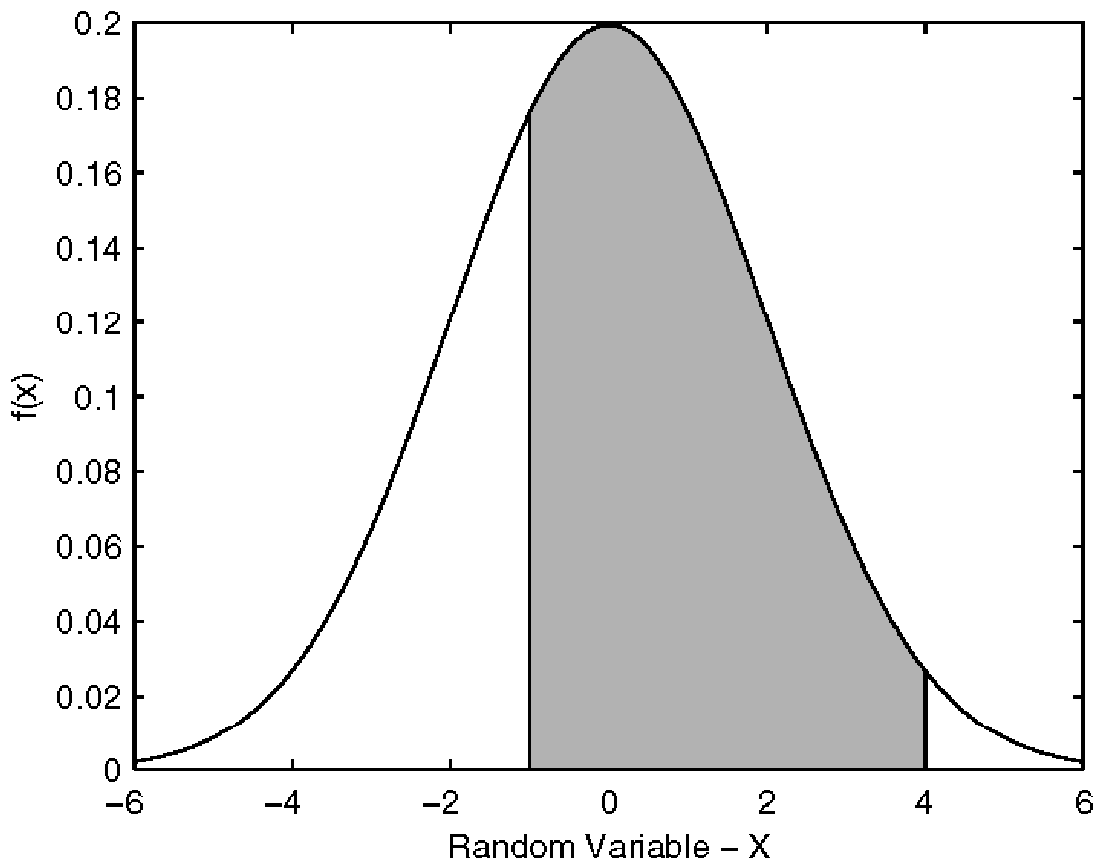
Check in the class both others !!

Probability Density Function (pdf)

Calculation of probability that a continuous random variable falls in a particular interval of real numbers:

we have to calculate the appropriate area under the curve of f(x) (integration)

$$P(a \le X \le b) = \int_a^b f(x)dx.$$



Axioms of Probability

S : Sample space of an experiment
E : Event that is a subset of S.

AXIOM 1 : The probability of event E must be between 0 and 1

$$0 \leq P(E) \leq 1$$

AXIOM 2

$$P(S) = 1$$

AXIOM 3
For mutually exclusive events E1, E1 .. Ek etc

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = \sum_{i=1}^{k} P(E_i).$$

Expectation

used to describe distributions

The mean value of a random variable: a measure of central tendency of the distribution (approximately as average value)

discrete case

$$\mu = E[X] = \sum_{i=1}^{\infty} x_i f(x_i)$$

continuous case

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

The variance: a measure of dispersion in the distribution.

If the variance is large: observed value is more likely to be far from the mean

discrete case

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i)$$

continuous case

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

Common Distributions

Bernoulli

Experiment with outcome 'success' or a 'failure'.

Define a success with X=1 with probability p, or P(X=1) = p, so
an unsuccessful with X=0, P(X=0) = 1 – p

Binomial

Repetition of Bernoulli n times, assume k times we get 1 and (n-k) times get 0, so

$$\xi_k = (\overbrace{1, 1, \ldots, 1,}^{k \text{ times}} \overbrace{0, 0, \ldots, 0}^{n-k \text{ times}}) .$$

$$\Pr(\xi_k) = \Pr(1, 1, \ldots, 1, 0, 0, \ldots, 0) = \Pr(1)\Pr(1) \ldots \Pr(1)\Pr(0)\Pr(0) \ldots \Pr(0)$$

$$= (\Pr(1))^k (\Pr(0))^{n-k} = p^k (1-p)^{n-k}. \quad (2.$$

Because, the sequence is irrelevant, we have other outcomes with k times 1 and (n-k) times 0, it leads

$$P_X(k) = \Pr(A_k) = (\text{\# of outcomes in } A_k) * (\text{probability of each outcome in } A_k)$$

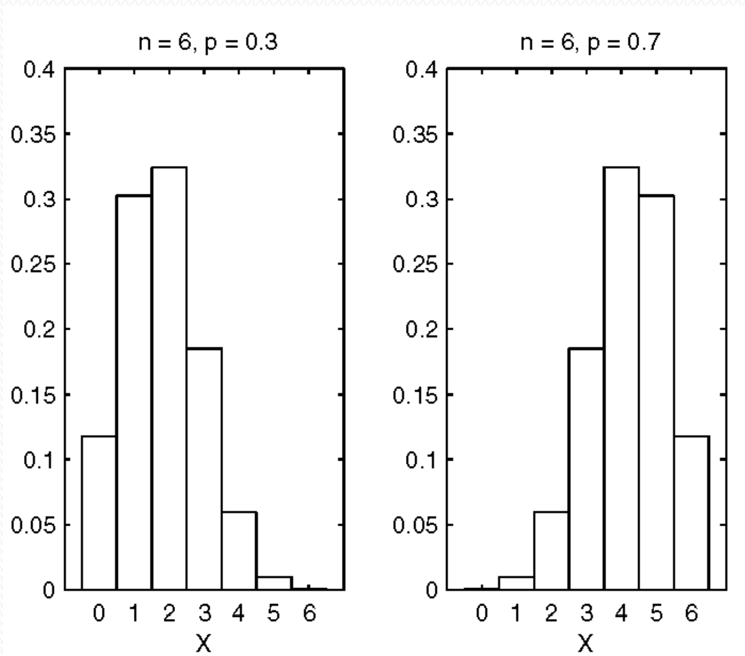
$$= \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Application of Binomial random variables:

In digital communication system, a packet of n bits may be transmitted Some arrive in error.

Or, perhaps a bank manager might be interested in the number of tellers who are serving customers at a given point in time.



Examples (do in class !)

• A drug has probability 0.90 of curing a disease. It is administered to 100 patients, where the outcome for each patient is either cured or not cured. If *X is the number of patients cured, then X is a* binomial random variable with parameters *(100, 0.90).*

• The National Institute of Mental Health estimates that there is a 20% chance that an adult American suffers from a psychiatric disorder. Fifty adult Americans are randomly selected. If we let *X* represent the number who have a psychiatric disorder, then *X takes* on values according to the binomial distribution with parameters (50, 0.20).

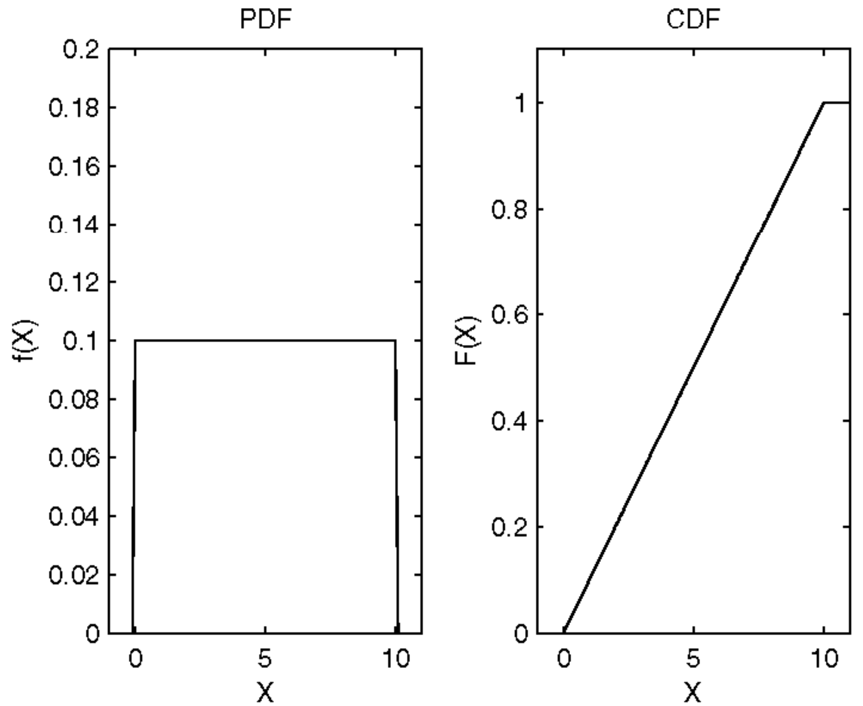
• A manufacturer of computer chips finds that on the average 5% are defective. To monitor the manufacturing process, they take a random sample of size 75. If the sample contains more than five defective chips, then the process is stopped. The binomial distribution with parameters (75, 0.05) can be used to model the random variable *X, where X represents the number of defective chips.*

Uniform Distribution

Perhaps one of the most important distributions is the uniform distribution for continuous random variables.

$$f(x;a,b) = \frac{1}{b-a}; \quad a < x < b\,.$$

```
x = -1:.1:11;
pdf = unifpdf(x,0,10);
cdf = unifcdf(x,0,10);
subplot(1,2,1);plot(x,pdf);
title('PDF');xlabel('X');ylabel('f(X)');
axis([-1 11 0 0.2]);axis square;
subplot(1,2,2);plot(x,cdf);
title('CDF');xlabel('X');ylabel('F(X)');
axis([-1 11 0 1.1]);axis square
```



EXAMPLE 3.6:

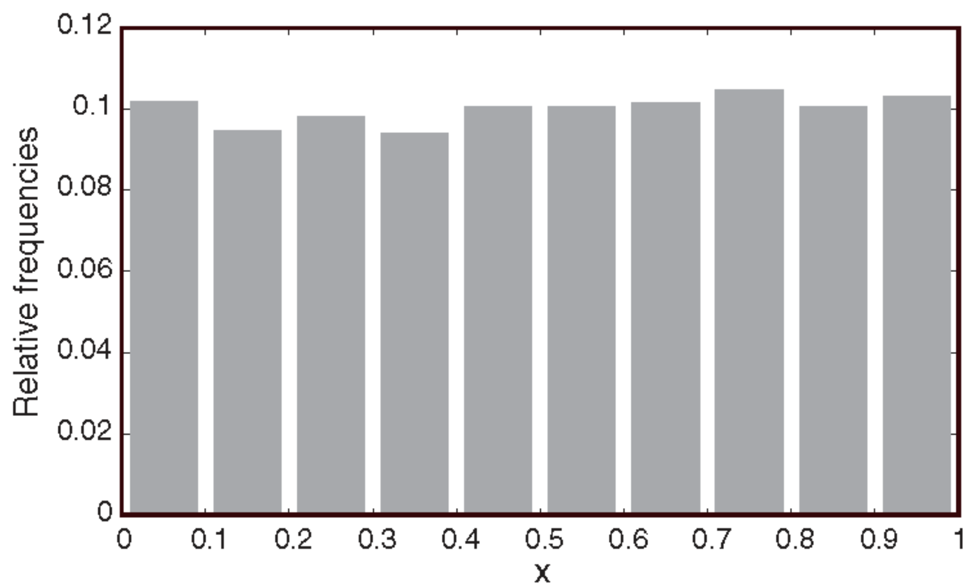
The MATLAB function *rand* generates random numbers that are uniformly distributed in the interval (0,1)

To construct a histogram for the random numbers generated by rand, we write a script that calls rand repeatedly.
Since we can do this only a finite number of times, we quantize the range of the random numbers into increments of 0.1.

We then calculate the number of times a random number falls in each quantized interval and divide by the total number of numbers generated for the example.

Try changing the value of *N or the number and width of the bins in this example to see* how results vary.

```
N=10,000;                       % Do N times
x=rand(1,N);                    % Produce N random numbers.
bins=[0.05:0.1:0.95];           % Create 10 bins,
                                % with centers at 0.05,0.15, ....
[yvalues,xvalues]=hist(x,bins); % Define xvalues and yvalues.
yvalues=yvalues/N;              % Normalize to produce
                                % relative frequencies.
bar(xvalues,yvalues);           % Plot bar graph.
xlabel('x')
ylabel('Relative Frequencies')
```



Normal Distribution

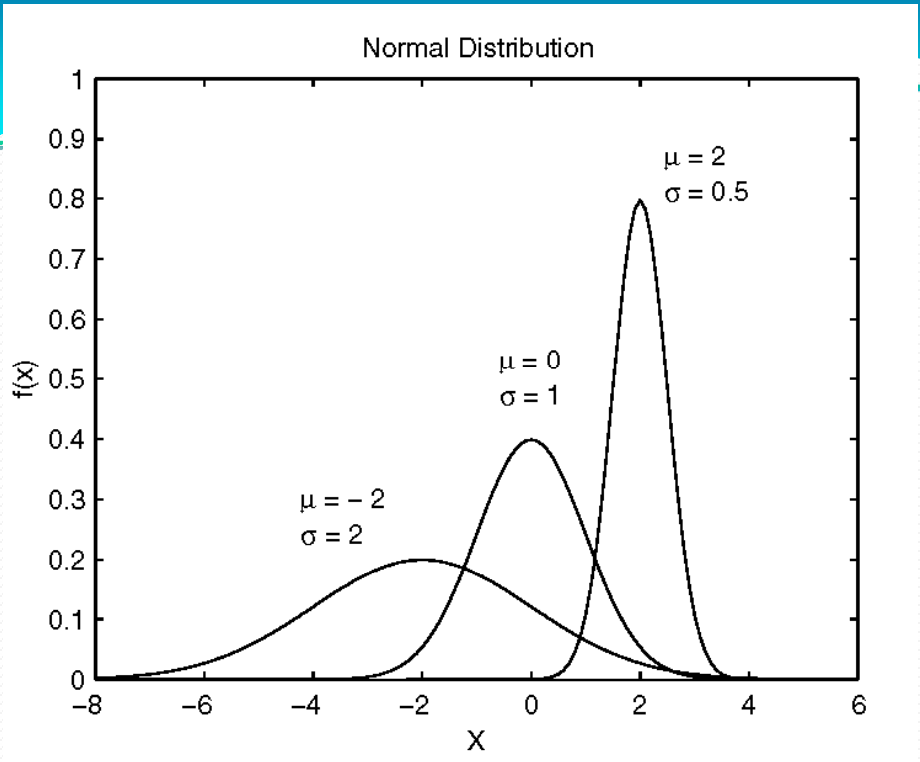
A well known distribution in statistics and engineering is the normal distribution. Also called the Gaussian distribution

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

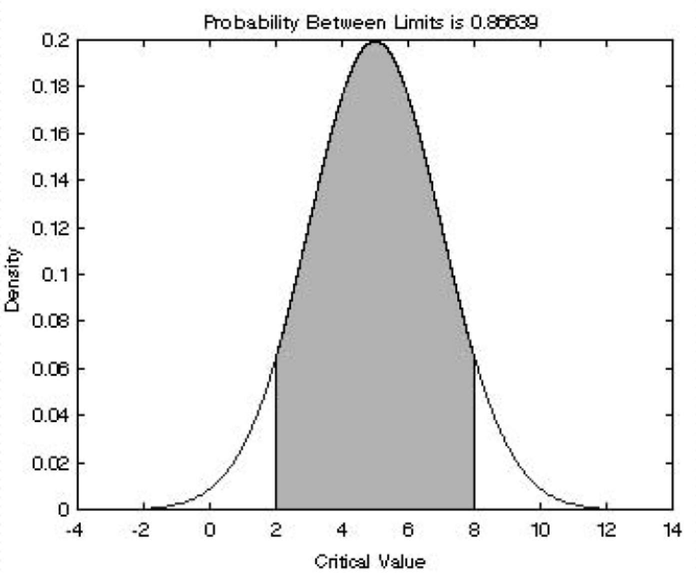
$-\infty < x,\mu < +\infty$ and $\sigma > 0$, which μ is mean and σ is standard deviation.

Some special properties of the normal distribution are given here.

- The value of the probability density function approaches zero as *x* approaches positive and negative infinity.
- The probability density function is centered at the mean μ, and the maximum value of the function occurs at $x = \mu$.
- The probability density function for the normal distribution is symmetric about the mean μ.



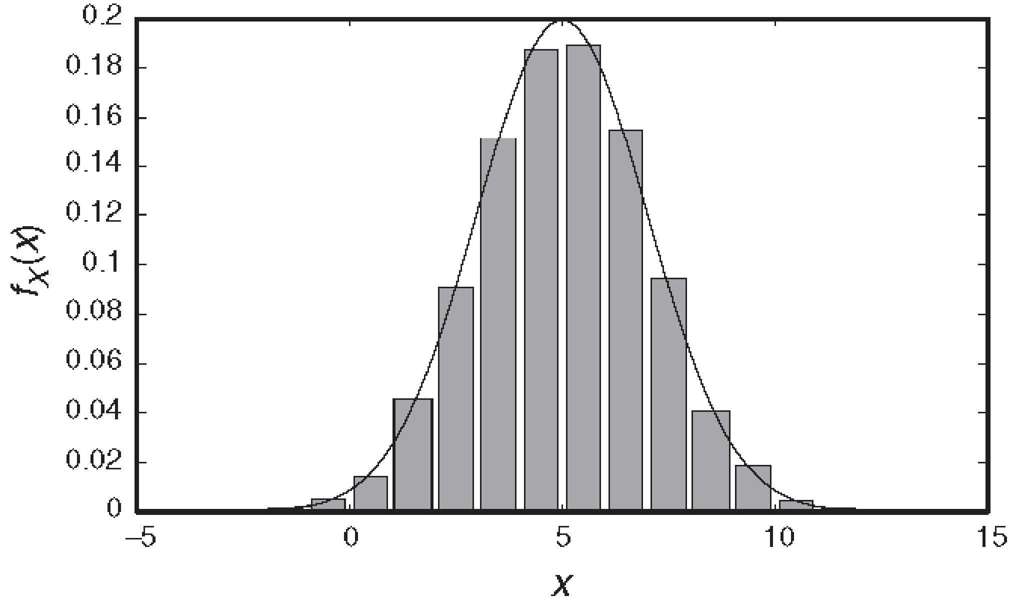
normpdf(x,mu,sigma)




```
mu = 5;
sigma = 2;
% the two element vector 'specs'.
specs = [2, 8];
prob = normspec(specs, mu, sigma);
```

EXAMPLE 3.9: MATLAB also has a built-in function, randn, which generates random variables according to a Gaussian or normal distribution.

```
N=10,000;
m=5; sigma=2;
x=m+sigma*randn(1,N);                 % Produce N random numbers.
left=-4.5; width=1; right=14.5;       % Set bin parameters.
bins=[left:width:right];              % Create bins with centers at
                                      % left, left+width,..., right.
[yvalues,xvalues]=hist(x,bins);       % Define xvalues and yvalues.
yvalues=yvalues/(N*width);            % Normalize to produce
                                      % probability densities.
bar(xvalues,yvalues);                 % Plot bar graph.
z=[left-width/2:width/10:right+width/2];
pdf=exp(-(z-m).^2/(2*sigma^2));       % Compute true PDF
pdf=pdf/sqrt(2*pi*sigma^2);
hold on                               % Place plot of true PDF on
plot(z,pdf)                           % top of histogram.
xlabel('x')
ylabel('f_X(x)')
```



Histogram of random numbers produced by `randn` along with a Gaussian PDF, where $m = 5$, $\sigma = 2$.

Other distributions

- Poisson
- Weibull
- Chi-Square
- Exponential
- Gamma
- Beta

Distribution	MATLAB Function
Beta	csbetap, csbetac
Binomial	csbinop, csbinoc
Chi-square	cschip, cschic
Exponential	csexpop, csexpoc
Gamma	csgammp, csgammc
Normal - univariate	csnormp, csnormc
Normal - multivariate	csevalnorm
Poisson	cspoisp, cspoisc
Continuous Uniform	csunifp, csunifc
Weibull	csweibp, csweibc

3.16 A digital communication system sends two messages, $M = 0$ or $M = 1$, with equal probability. A receiver observes a voltage which can be modeled as a Gaussian random variable, X, whose PDFs conditioned on the transmitted message are given by

$$f_{X|M=0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad \text{and}$$

$$f_{X|M=1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-1)^2}{2\sigma^2}\right).$$

(a) Find and plot $\Pr(M = 0|X = x)$ as a function of x for $\sigma^2 = 1$. Repeat for $\sigma^2 = 5$.

(b) Repeat part (a) assuming that the a priori probabilities are $\Pr(M = 0) = 1/4$ and $\Pr(M = 1) = 3/4$.

3.17 In Problem 3.16, suppose our receiver must observe the random variable X and then make a decision as to what message was sent. Furthermore, suppose the receiver makes a three-level decision as follows:

(1) Decide 0 was sent if $\Pr(M = 0|X = x) \geq 0.9$.

(2) Decide 1 was sent if $\Pr(M = 1|X = x) \geq 0.9$.

(3) Erase the symbol (decide not to decide) if both $\Pr(M = 0|X = x) < 0.9$ and $\Pr(M = 1|X = x) < 0.9$.

Assuming the two messages are equally probable, $\Pr(M = 0) = \Pr(M = 1) = 1/2$, and that $\sigma^2 = 1$, find

(a) the range of x over which each of the three decisions should be made,

(b) the probability that the receiver erases a symbol,

(c) the probability that the receiver makes an error (i.e., decides a 0 was sent when a 1 was actually sent, or vice versa).