Mathematical Preliminaries

Math Models of Human Behavior

Department Psychology

Prof. Schrater

Spring 2006

Linear Algebra

- Need to know or learn
  - How to compute inner products, outer products
  - Multiply, transpose matrices
  - Elements of linear transformations
    - Rotations and scaling

Vectors

- Vectors, Points, constraint lines
- Length, direction, unit circle, rotation
- Addition, subtraction
- Dot product
- Decomposition

(on blackboard)

Linear Algebra Primer

- Why linear algebra?
- What if more than one value is going in and coming out of a system?
  - For example, we have many inputs coming into the retina (photoreceptors) and many outputs (retinal ganglion cells).
  - Any single input affects many different outputs, and any given output is influenced by many different inputs.
- How do we even begin to understand such a system?
  - Linear algebra provides a useful tool for characterizing the behavior of systems where many values must be represented simultaneously. - e.g., . the brain

Vectors

- A vector is simply a list of numbers. A vector is usually denoted in boldface, with an arrow over it, or underlined:

$$\mathbf{x} = \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{OR} \quad \begin{bmatrix} w \\ x \\ \vdots \\ z \end{bmatrix}$$

- An "*n*-dimensional vector" has *n* elements. One can think of an *n*-dimensional vector as either a point in an *n*-dimensional space, or as an arrow drawn from the origin to the point with coordinates **x**.

Vector addition

- The addition of two vectors is their component-wise sum:

$$\mathbf{z} = \mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

- Geometrically, vectors are added by placing them end to end. The vector from the origin to the tip of the last vector is the sum vector.

Dot Product

- The *inner-product* (or) *dot product* of two vectors takes the sum of products of the elements of each vector:

$$\mathbf{x}\cdot\mathbf{y} = \langle \mathbf{x},\mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$$

$$\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

- This provides a measure of the similarity of two vectors (provided you know the length of each vector). If you divide the inner product by the length of each vector, you get the cosine of the angle between them

$$\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \quad \text{where} \quad \|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_i x_i}$$

Dot Prod. Cont'd

- An alternative geometric interpretation of the inner product is that gives you the length one vector after it has been projected onto the other.
  - Thus, orthogonal vectors have an inner product of zero.
  - The inner product is oftentimes also denoted , where the superscript *T* denotes "transpose." $\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$

- The transpose of a vector simply tilts it on its side so it is written as a row of numbers (instead of a column)

$$\mathbf{x}^T = [x_1, x_2, \cdots, x_n]$$

Matrices

- A matrix **M** is just a 2D array of numbers. It is used to map a vector into a new vector via the relation:

$$\mathbf{y} = \mathbf{M}\mathbf{x}$$

It takes the place of the constant, k, in a simple one-dimensional multiplication. An *m* x *n* matrix **M** has elements:

$$\mathbf{M} = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \\ M_{m1} & M_{m2} & \cdots & M_{mn} \end{bmatrix}$$

Matrix-vector multiplication

- Multiplying an *m* x *n* matrix and an *n*-dimensional vector produces an *m-dimensional vector,* and we can write the resulting vector elementwise as:

$$\mathbf{y} = \mathbf{M}\mathbf{x} = \begin{bmatrix} M_{11}x_1 + M_{12}x_2 + \cdots + M_{1n}x_n \\ M_{21}x_1 + M_{22}x_2 + \cdots + M_{2n}x_n \\ \vdots \\ M_{m1}x_1 + M_{m2}x_2 + \cdots + M_{mn}x_n \end{bmatrix}$$

- Analysis/synthesis: the *rows* vectors of a matrix tell you what part of the input space (vectors on the right side of the matrix) the matrix *analyzes (think about the projection interpretation of dot products)*. The *column* vectors tell you what part of the output space (i.e., vectors coming out the left side) the matrix can *synthesize* from any input.

Matrix properties

- Just as a linear scalar system can produce only a limited set of remappings of scalar values, so too is a matrix operation limited in the set of remappings it can produce. What do these remappings look like? - rotations (orthonormal matrices) and scalings (diagonal matrices) and combinations thereof.
- The concatenation of two matrix operations is just another matrix operation. The combined matrix is obtained via matrix multiplication.
- Any matrix can be decomposed in terms of a rotation, a scaling, and another rotation.

Matlab Intro

- "BASIC for people who like linear algebra"
- Full programming language
  - Interpreted language (command)
  - Scriptable
  - Define functions (compilable)

Data

- Basic- Double precision arrays

```
A = [ 1 2 3 4 5]
A = [ 1 2; 3 4]
B = cat(3,A,A) %three dimensional array
```

Advanced- Cell arrays and structures

```
    A(1).name = ‘Paul’
    A(2).name = ‘Harry’

A = {‘Paul’;’Harry’;’Jane’};
>> A{1}                    =>        Paul
```

Almost all commands Vectorized

- A = [1 2 3 4 5] ; B = [2 3 4 5 6]
  - C = A+B
  - C = A.*B
  - C = A*B'
  - C = [A;B]
  - sin(C), exp(C)

Useful commands

- Colon operator
  - Make vectors: a = 1:0.9:10; ind = 1:10
  - Grab parts of a vector: a(1:10) = a(ind)
  - A = [1 2; 3 4]
  - A(:,2)
  - A(:) = [1
    3
    2
    4]

Vectorwise logical expressions

a = [1 2 3 1 5 1]

a = =1 => [1 0 0 1 0 1]

size( ), whos, help, lookfor

ls, cd, pwd,

Indices = find(a = =1) => [1 4 6]

Stats Commands

- Summary statistics, like
  - Mean(), Std(), var(), cov(), corrcoef()
- Distributions:
  - normpdf(),
- Random number generation
  - P = mod(a*x+b,c)
    rand(), randn(), binornd()
- Analysis tools
  - regress(), etc

Probability

For each event $A \subseteq S$, we assume there is a number $P(A)$ called the <u>probability of event A</u>, satisfying the conditions:

i. $0 \leq P(A) \leq 1$

ii. $P(S) = 1$

iii. If $A_1, A_2, A_3, \ldots$ are mutually exclusive ($A_i \cap A_j = \varnothing$, $i \neq j$), then $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

Observe that

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$$

So

$$P(A^c) = 1 - P(A)$$

Law of Total Probability

If $A_1, A_2, A_3, \ldots \subseteq S$ are mutually exclusive such that

$A_i \cap A_j = \varnothing$ for $i \neq j$, and $S = \bigcup_{i=1}^{\infty} A_i$,

then exactly one of the events A_i will occur

(in other words, $\sum_{i=1}^{\infty} P(A_i) = 1$)

and for any event $B \subseteq S$, $P(B) = \sum_{i=1}^{\infty} P(B \cap A_i)$

Conditional Probability

For two events A and B in S ($A,B \subseteq S$), the <u>conditional probability of A given B</u> is the probability that A occurs given that B has already occurred. It is denoted $P(A|B)$ and satisfies

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Note: this makes sense only when $P(B) > 0$.

Independence

Two events A and B in S ($A,B \subseteq S$) are <u>independent</u> if

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

Note that by the definition of conditional probability, if events A and B are independent, then

$$P(A \mid B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

Two events that are not independent are said to be <u>dependent</u>.

Bayes' Formula

Consider two events A and B in S ($A,B \subseteq S$). Since B and B^c are mutually exclusive

$$P(A) = P(A \cap B) + P(A \cap B^c) \quad \text{(law of total probability)}$$

$$= P(A|B)P(B) + P(A|B^c)P(B^c) \quad \text{(def. of conditional probability)}$$

Then, for $B_1, B_2, \ldots, B_n$ mutually exclusive with $\bigcup_{i=1}^{n} B_i = S$

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Suppose that event A has occurred and we want to know whether B_j has occurred…

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)}$$

$$= \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}$$

Bayes Example

Does patient have cancer or not?

- A patient takes a lab test and the result comes back positive. The test returns a correct positive result in 98% of the cases in which the disease is actually present, and a correct negative result in 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer.

P (cancer) = P (not cancer) =

P (+|cancer) = P (–|cancer) =

P (+|not cancer) = P (–|not cancer) =

WHAT YOU NEED TO KNOW

Joint Probability

$$P(x, y) = P(A \cap B)$$

Conditioning

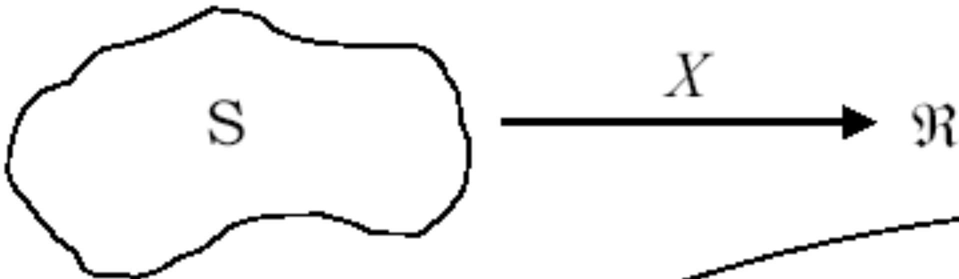
$$P(y \mid x) = P(x, y) / P(x)$$

Marginalization

$$P(x) = \sum_{y} P(x, y)$$

Random Variables

A <u>random variable</u> is a function that associates a (real) number with each outcome in the sample space.



Example: Consider the roll of two fair dice.

roll of die #2

roll of die #1	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Let the random variable X equal their sum.

X	$P(X)$
1	0
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Matlab code for computing sum of two die

```
% Need to enumerate all possibilities
% die1 = 1:6;
% die2 = 1:6;

% Now a basic control structure
%
for die1value=1:6,
    for die2value = 1:6,
        possibilities(die1value,die2value) = die1value +
        die2value;
    end
end

% possibilities =
%
%     2     3     4     5     6     7
%     3     4     5     6     7     8
%     4     5     6     7     8     9
%     5     6     7     8     9    10
%     6     7     8     9    10    11
%     7     8     9    10    11    12

% sort our table into a long list

possibilities = reshape(possibilities,[1,36])
```

```
% possibilities =
%
%   Columns 1 through 18
%
%     2     3     4     5     6     7     3     4     5     6     7     8
%     4     5     6     7     8     9
%
%   Columns 19 through 36
%
%     5     6     7     8     9    10     6     7     8     9    10    11
%     7     8     9    10    11    12

% now the minimum value of the sum is 2
%                                  and the max is 12

for sumvalues = 2:12,
    testifequal = (possibilities ==  sumvalues);
    % testifequal returns a new list of the same size
    % possibilities with a 1 for every element in
    % possibilities that is equal to the current sumvalue
    % (2,3,4, etc) and zero for all other values
    count(sumvalues-1) = sum(testifequal);
end

probsum = count/36
```

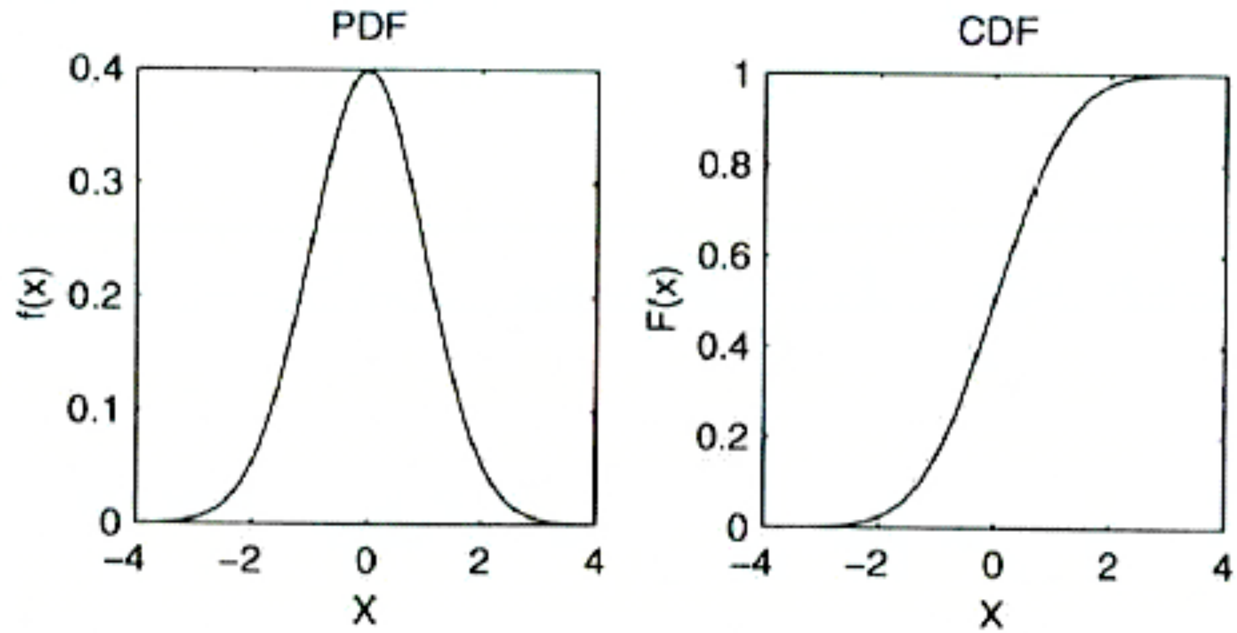


FIGURE 2.2
This shows the probability density function on the left with the associated cumulative distribution function on the right. Notice that the cumulative distribution function takes on values between 0 and 1.

Probability Distribution Function

Given a random variable X, its <u>cumulative distribution function</u> (CDF) is defined as

$$F(b) = P(X \leq b)$$

for any real number b, where $-\infty < b < \infty$.

Properties of the CDF include:

i. $F(b)$ is a non-decreasing function of b

ii. $\lim_{b \to \infty} F(b) = F(\infty) = 1$

iii. $\lim_{b \to -\infty} F(b) = F(-\infty) = 0$

In general, all probability questions about X can be answered in terms of the CDF. For example, for $a < b$

$$P(a < X \leq b) = F(b) - F(a)$$

Discrete Random Variables

A random variable is discrete if it can take on a countable number of values. Example: $X \in \{2, 3, 4, \ldots, 12\}$

For a discrete random variable X, we define the probability mass function as

$$p(a) = P(X = a)$$

So the CDF for a discrete random variable satisfies

$$F(a) = P(X \le a) = \Sigma_{x \le a} P(X = x) = \Sigma_{x \le a} p(x)$$

Consider the case where the possible values of X can be enumerated by $x_1, x_2, \ldots, x_n$. Then,

$$p(x_i) > 0 \qquad \text{for } i = 1, 2, \ldots, n$$

$$p(x) = 0 \qquad \text{for all other values of } x$$

and

$$\sum_{i=1}^{n} p(x_i) = 1$$

Important Discrete Random Variables

Bernoulli Random Variable with parameter (p) (where $0 \le p \le 1$)

$X \in \{0,1\}$ $\quad p(0) = P\{X = 0\} = 1\text{-}p$

$p(1) = P\{X = 1\} = p$

Binomial Random Variable with parameters (n,p) (where $n \ge 0$, $0 \le p \le 1$)

$X \in \{0,1,2, \ldots, \mathrm{n}\}$ $\quad p(i) = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$

Geometric Random Variable with parameter (p) (where $0 \le p \le 1$)

$X \in \{1,2,3,\ldots\}$ $\quad p(n) = P\{X = n\} = (1\text{-}p)^{n\text{-}1}\, p$

Poisson Random Variable with parameter (λ) (where $\lambda \ge 0$)

$X \in \{0,1,2,\ldots\}$ $\quad p(i) = P\{X = i\} = e^{-\lambda} \dfrac{\lambda^i}{i!}$

Binomial Events

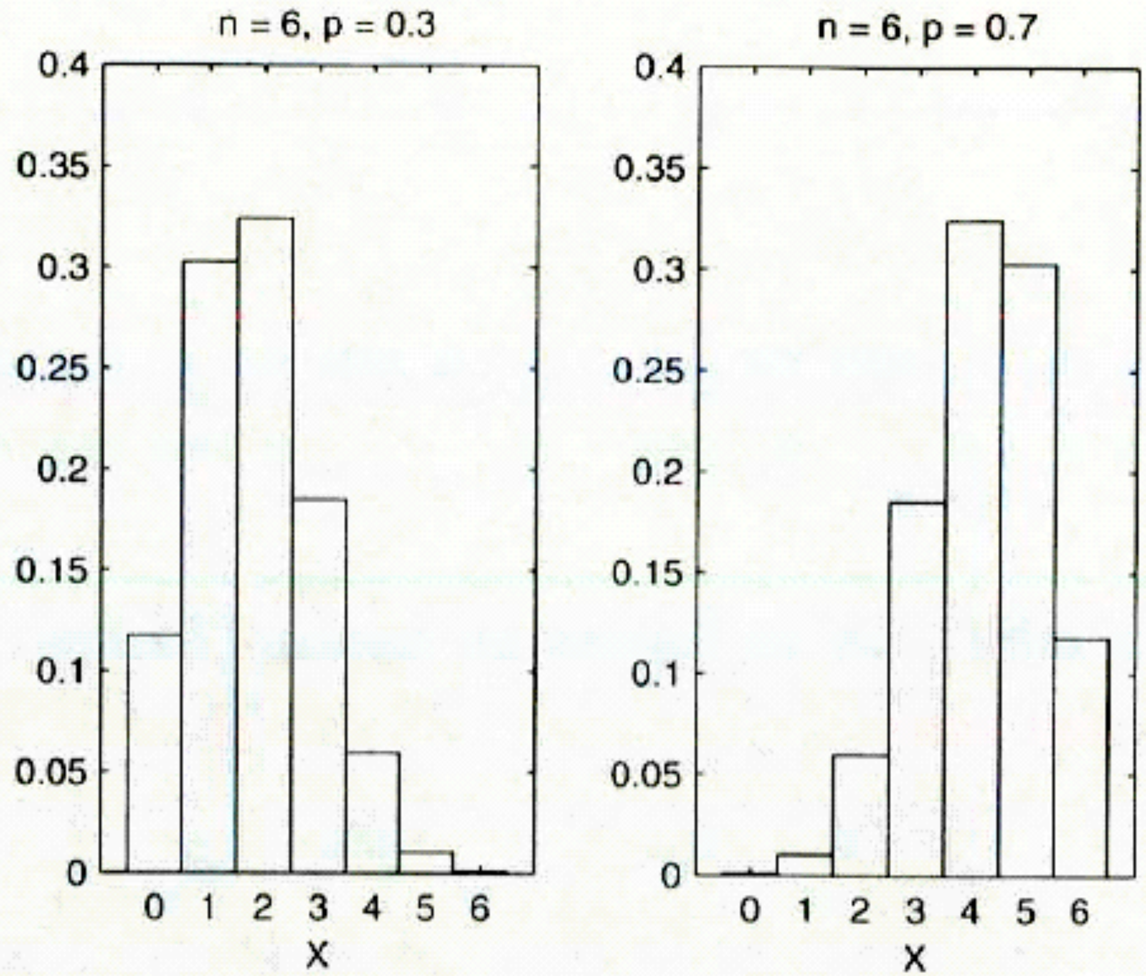


FIGURE 2.3
Examples of the binomial distribution for different success probabilities.

Birthday Problem

What's the chance two people in a room share the same birthday?

Events?

Define probabilities?

Answer in MATLAB code:

```
n=0:60;  p = 1-cumprod((365-n)/365)
```

Birthday Problem

What's the chance two people in a room share the same birthday?

P(B) = 1- P(Not B) = 1-P(NOONE shares the same birthday.)

Events?

Events A_i = Person **i**'s birthday is different from Persons {0,…,**i**-1}

Define probabilities?

$P(A_i)$ = P(**i's** birthday is different from preceding persons)

= (365-i)/365 *(i.e. How many chances out of the total)*

P(Not B) = $P(A_0 \,\&\, A_1 \,\&\ldots\&\, A_i) = P(A_0, A_1, \ldots, A_i)$

$= P(A_0)P(A_1)\ldots P(A_i) = \prod_{j=0}^{i} P(A_j)$

$= (365*364*\ldots.(365-i))/365^i = 365!/((365-i)!\,365^i)$

Continuous Random Variables

A random variable is <u>continuous</u> if it can take on a continuum of possible values. Example: $X \in [0,1]$

For a continuous random variable, we define the <u>probability density function</u> $f(x)$ for all real values $-\infty < x < \infty$

$$F(a) = P(X \le a) = \int_{-\infty}^{a} f(x)dx$$

and more generally

$$P(a \le X \le b) = \int_{a}^{b} f(x)dx$$

This definition implies the following:

$$P(X = a) = \int_{a}^{a} f(x)dx = 0 \qquad P(-\infty \le X \le \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

$$\frac{d}{da}F(a) = f(a)$$

Important Continuous Random Variables

<u>Uniform Random Variable</u> with parameters (α,β)

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \alpha<x<\beta\\ 0 & \text{otherwise}\end{cases}\qquad F(a)=\begin{cases}0 & a\le\alpha\\ \dfrac{a-\alpha}{\beta-\alpha} & \alpha<x<\beta\\ 1 & a\ge\beta\end{cases}$$

<u>Exponential Random Variable</u> with parameter (λ)

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & \text{x}<0\end{cases}\qquad F(a)=1-e^{-\lambda a}\quad a\ge 0$$

<u>Normal Random Variable</u> with parameters (μ,σ^2)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}\qquad F(a)=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}dx$$

Define $Y=(X-\mu)/\sigma$. If $X\sim N(\mu,\sigma^2)$, then $Y\sim N(0,1)$ is known as the <u>standard (unit) random variable</u>. $\Phi(a)=P\{Y\le a\}$

Expected Value

The expected value of a random variable X is

$$E(X) = \sum_{\text{all } x} x\, p(x) \qquad E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

(if X is discrete) (if X is continuous)

and is also known as the expectation, mean, or first moment of X.

Examples:

- Let X be Bernoulli with parameter p.

$$E[X] = 1(p) + 0(1-p)$$
$$= p$$

- Let Y be Uniform with parameters (α, β).

$$E[Y] = \int_{\alpha}^{\beta} \frac{y}{\beta - \alpha}\, dy$$
$$= \left[\frac{y^2}{2(\beta - \alpha)} \right]_{\alpha}^{\beta}$$
$$= \frac{\beta + \alpha}{2}$$

Expected Value for Functions of X

Let $g(X)$ be a function of the random variable X. Then,

$$E[g(X)] = \sum_{\text{all } x} g(x)\, p(x)$$

(if X is discrete)

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

(if X is continuous)

Consider the following important functions:

- When $g(x)=X^m$, then $E[g(X)]$ is known as the m^th^ moment of X

$$E[X^m] = \sum_{\text{all } x} x^m\, p(x) \qquad E[X^m] = \int_{-\infty}^{\infty} x^m\, f(x)\, dx$$

- Let $\mu_x=E[X]$ be the mean of the random variable X. When $g(x)=(x-\mu_x)^2$, then $E[g(X)]$ is known as the variance of X

$$Var(X) = E[(X-\mu_x)^2] = \sum_{\text{all } x} (x-\mu_x)^2\, p(x)$$

$$Var(X) = E[(X-\mu_x)^2] = \int_{-\infty}^{\infty} (x-\mu_x)^2\, f(x)\, dx$$

- In general, $E[(x-\mu_x)^m]$ is known as the m^th^ central moment of X.

Jointly Distributed Random Variables

For any two random variables X and Y we define the joint cumulative probability distribution function of X and Y as

$$F(a,b) = P(X \le a,\ Y \le b) \quad -\infty \le a,b \le \infty$$

In a manner completely analogous to the case of a single random variable, we define:

- Joint probability mass function: $p(x,y)$ (discrete case)
- Joint probability density function: $f(x,y)$ (continuous case)
- Expectation of jointly distributed random variables

Just as we speak of independence of events, we say that two random variables X and Y are independent if

$$P(X \le x,\ Y \le y) = P(X \le x)\, P(Y \le y)$$

By the definition of conditional probability, X and Y are independent if and only if

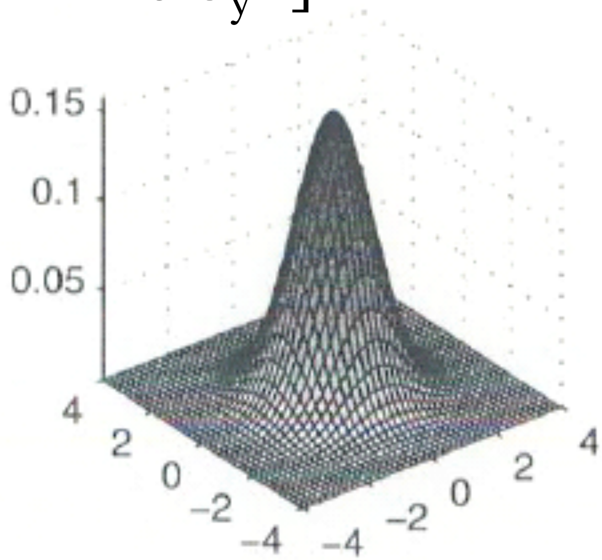
$$P(X \le x \mid Y \le y) = P(X \le x)$$

Normal & Multivariate Normal

$$f(\mathbf{x};\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\},$$

$\Sigma = [\sigma_x^2 \; 0$

$0 \; \sigma_y^2]$



$\Sigma = [\sigma_x^2 \quad \rho\sigma_x\sigma_y$

$\rho\sigma_x\sigma_y \quad \sigma_y^2]$