Optimal Estimation Methods

(Lecture 6 – Basic Probability Concepts)

Dr. John L. Crassidis

University at Buffalo – State University of New York
Department of Mechanical & Aerospace Engineering
Amherst, NY 14260-4400
johnc@buffalo.edu
http://www.buffalo.edu/~johnc

University at Buffalo The State University of New York

Discrete Variables (i)

- Consider a single throw of a "true" die
  - The probability of the occurrence of each of the events 1, 2, 3, 4, 5, or 6 is exactly the same on a given throw
  - If a given discrete-values experiment is conducted N times and N_j is the number of times that the j^{th} event $x(j)$ occurred, then it is intuitively reasonable to define the probability of the occurrence of $x(j)$ as

$$p(x(j)) \equiv \lim_{N \to \infty} \frac{N_j}{N}$$

  - For example, for a throw of a single die the probability of obtaining a value of 3 is given by $p(3) = 1/6$
  - A discrete-valued random variable, x, is defined as a function having finite number of possible values $x(j)$; with the associated probability of $x(j)$ occurring being denoted by $p(x(j))$
  - We most often use the notation x and $p(x)$ to simplify notation

Discrete Variables (ii)

- Throw two dice and compute probabilities

Sum	Count	$p(x)$
2	1	$1/36$
3	2	$2/36$
4	3	$3/36$
5	4	$4/36$
6	5	$5/36$
7	6	$6/36$
8	5	$5/36$
9	4	$4/36$
10	3	$3/36$
11	2	$2/36$
12	1	$1/36$

We now have 36 possible outcomes over the entire set

When multiple dice, $n > 2$, are used this table is much more difficult to produce. Fortunately, can sometimes use a *generating function* to obtain the result

$$f(x) = \left(x + x^2 + x^3 + x^4 + x^5 + x^6\right)^n$$

The coefficients of the powers of x can be used to form the "count" column. The probability of each event is given by the count divided by 6^n. Check $n = 2$ case:

$$f(x) = x^2 + 2x^3 + 3x^4 + 4x^5 + 5x^6 + 6x^7 + 5x^8 + 4x^9 + 3x^{10} + 2x^{11} + x^{12}$$

Discrete Variables (iii)

- Consider another experiment involving four flips of a coin
  - Look at number of ways a heads appears for the 16 total outcomes

No. of Heads	Results	No. of Ways
0	TTTT	1
1	HTTT THTT TTHT TTTH	4
2	TTHH THTH HTTH THHT HTHT HHTT	6
3	THHH HTHH HHTH HHHT	4
4	HHHH	1

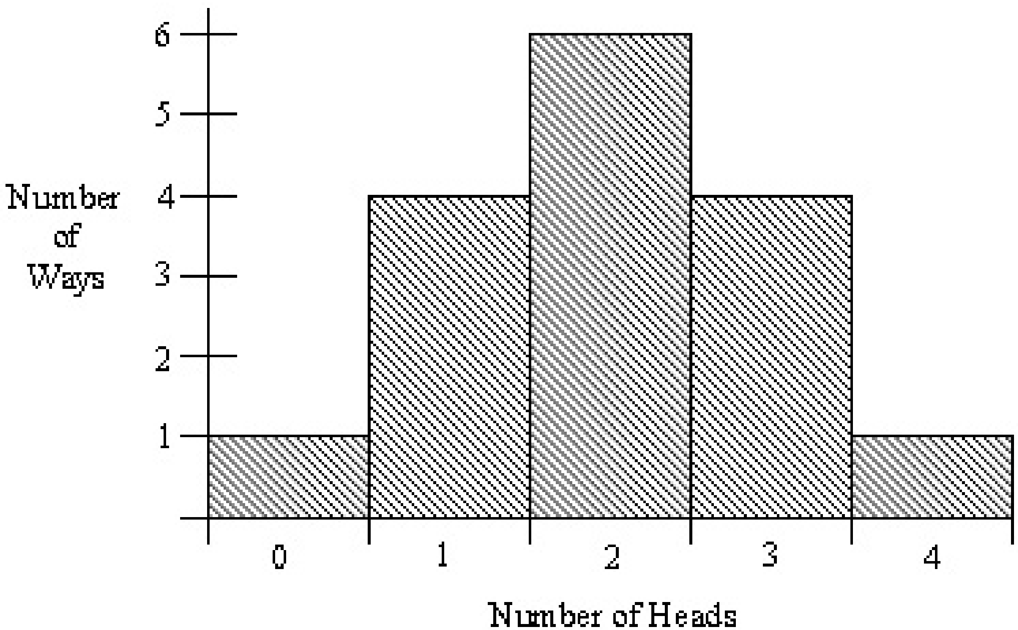
Probability is number of ways divided by number of total possible outcomes

No. of Heads	No. of Ways	Probability
0	1	$1/16 = 0.0625$
1	4	$4/16 = 0.25$
2	6	$6/16 = 0.375$
3	4	$4/16 = 0.25$
4	1	$1/16 = 0.0625$

Note that the sum of all the probabilities is 1!

Discrete Variables (iv)

- Look at histogram of results
  - Note, starting to look like a standard “bell curve” already
  - Looks more and more like a bell curve with more flips



Histogram of the Number of Ways a Heads Appears for Four Flips

- General case for n flips can also be obtained

Mathematically, the number of ways to obtain x heads in n flips is spoken as the “number of combinations of n things taken x at a time”

$$\text{Number of Ways} \equiv \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

For example if $n = 4$ and $x = 2$, then the number of ways is computed to be 6, which matches table result. Probability is given by

$$p(x) = \frac{\binom{n}{x}}{2^n} = \frac{n!}{x!\,(n-x)!\,2^n}$$

For example if $n = 4$ and $x = 2$, then $p(2) = 0.375$, which matches table result.

Discrete Variables (v)

$\cup$ - or
$\cap$ - and

- Compound Event
  - Defined as the occurrence of "either $x(j)$ or $x(k)$"; the probability of a compound event is defined as

$$p(x(j) \cup x(k)) = p(x(j)) + p(x(k)) - p(x(j) \cap x(k))$$

  where $x(j) \cup x(k)$ denotes "$x(j)$ or $x(k)$" and $x(j) \cap x(k)$ denotes "$x(j)$ and $x(k)$"
  - The probability of obtaining one event and another event is known as the joint probability of $x(j)$ and $x(k)$
  - If $p(x(j) \cap x(k)) = 0$, then the individual probabilities are summed to determine the overall probability
    - For example, the probability of obtaining less than 3 heads in 4 flips is given by $1/16 + 4/16 + 6/16 = 0.6875$
    - Note that calculating the probability of obtaining 4 or less heads gives a value of $1/16 + 4/16 + 6/16 + 4/16 + 1/16 = 1$

Discrete Variables (vi)

- It is clear that a probability mass function $p(x(j))$ has the following properties

$$0 \leq p(x(j)) \leq 1$$

$$\sum_j p(x(j)) = 1$$

  - If events $x(j)$ and $x(k)$ are independent, then we have

$$p(x(j) \cap x(k)) = p(x(j))\, p(x(k))$$

  - For example, the probability of obtaining one head in two successive trials is given by $1/4 \times 1/4 = 1/16$

Discrete Variables (vii)

- We now define the conditional probability of $x(j)$ given $x(k)$, which is denoted by $p(x(j)|x(k))$
  - Suppose we know that an event $x(k)$ has occurred
  - Then $x(j)$ occurs if and only if $x(j)$ and $x(k)$ occur
  - Therefore, the probability of $x(j)$, given that we know $x(k)$ has occurred, should intuitively be proportional to $p(x(j) \cap x(k))$
  - However, the conditional probability must satisfy the properties of probability shown previously
  - This forces a proportionality constant of $1 \,/\, p(x(k))$, so that

x(j) given x(k) occurs

$$p(x(j)|x(k)) = \frac{p(x(j) \cap x(k))}{p(x(k))}$$

  - In a similar manner the conditional probability of $x(k)$ given $x(j)$ is

$$p(x(k)|x(j)) = \frac{p(x(k),\, x(j))}{p(x(j))}$$

where $p(x(k), x(j)) \equiv p(x(k) \,\cap\, x(j))$

Bayes' Rule

- Combining the previous equations leads to *Bayes' Rule*

$$\boxed{p(x(j)|x(k)) = \frac{p(x(k)|x(j))\,p(x(j))}{p(x(k))}}$$

$P(x_j)$ given x_k occurs = probability of x_k given x_j occurs, times prob of x_j divided by prob of x_k

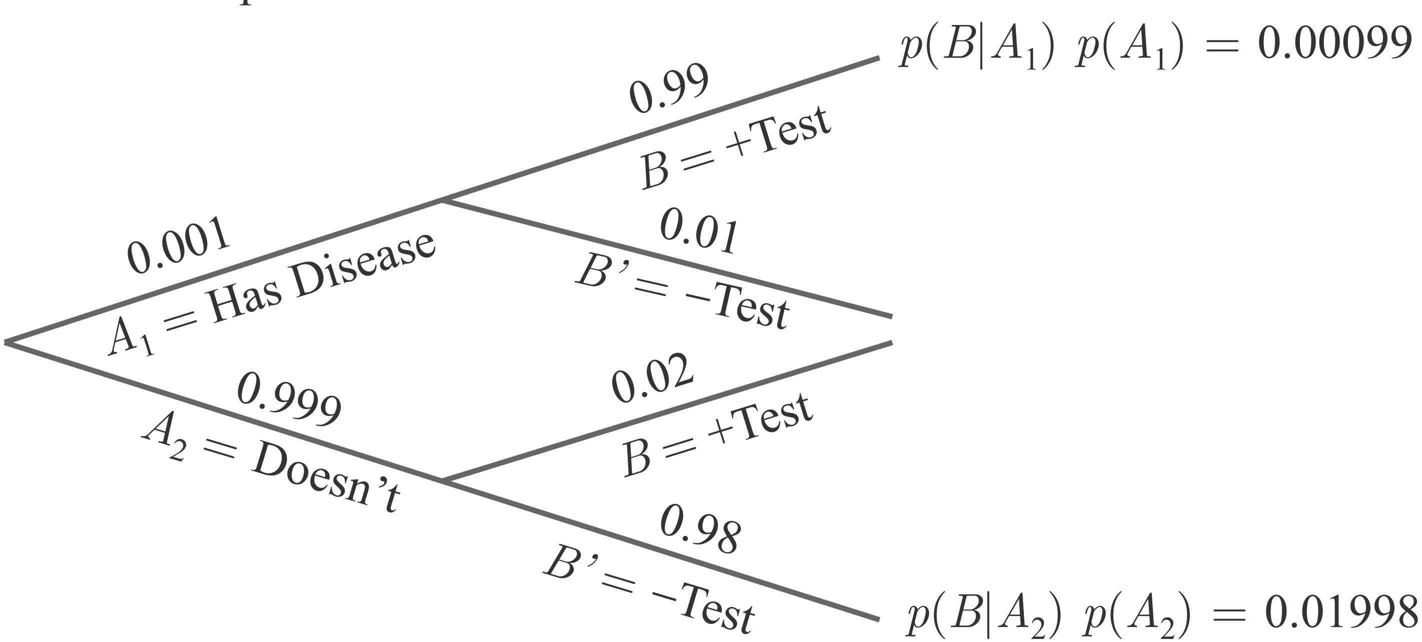
- Widely used in estimation theory
  - We'll apply it to least squares-type applications later

Bayes' Rule Example (i)

- Say 1 out $1{,}000$ people have a rare disease
- Tests show that 99% are positive when they have a disease and 2% are positive when they don't
- What is the probability when the test is positive that they actually have a disease?

$$\begin{aligned} A_1 &= \text{person with disease} \\ A_2 &= \text{person without disease} \\ B &= \text{positive test result} \end{aligned}$$

$$\begin{aligned} p(A_1) &= 0.001, & p(A_2) &= 0.999 \\ p(B|A_1) &= 0.99, & p(B|A_2) &= 0.02 \end{aligned}$$



Bayes' Rule Example (ii)

- Compute probability of a positive test result

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) = 0.00099 + 0.01998 = 0.02097$$

- Use Bayes' Rule; want $p(A_1|B)$

$$p(A_1|B) = \frac{p(B|A_1)p(A_1)}{p(B)} = \frac{0.00099}{0.02097} = 0.047$$

- Seem counterintuitive, why?
  - Disease is rare and test is only moderately reliable
  - Most positives arise from errors rather than from diseased people (2% equates to 20 people when only 1 has the disease)
  - Note: if a 25% incidence rate is given, then the probability is 0.94, which is line with our intuition; $p(A_1) = 0.25$, $p(A_2) = 0.75$

$$p(B) = p(B|A_1)P(A_1) + p(B|A_2)p(A_2) = 0.2475 + 0.015 = 0.2625$$

$$p(A_1|B) = \frac{p(B|A_1)p(A_1)}{p(B)} = \frac{0.2475}{0.2625} = 0.94$$

Monte Hall Problem (i)

- Famous problem
  - The Monty Hall Problem’s origin is from the TV show, “Let’s Make A Deal” hosted by Monty Hall
    - “You are a contestant in a game show in which a prize is hidden behind one of three curtains. You will win the prize if you select the correct curtain. After you have picked one curtain but before the curtain is lifted, Monty Hall lifts one of the other curtains, revealing an empty stage, and asks if you would like to switch from your current selection to the remaining curtain. How will your chances change if you switch?”
  - The question was originally proposed by a reader of “Ask Marilyn,” a column by Marilyn vos Savant in Parade Magazine in 1990
    - Her solution caused an uproar among mathematicians, as the answer to the problem is unintuitive: while most people would respond that switching should not matter, the contestant’s chances for winning in fact double if she switches curtains
    - Part of the controversy, however, was caused by the lack of agreement on the statement of the problem itself

Monte Hall Problem (ii)

- Without loss of generality, let us call the curtain picked by the **contestant curtain** a, the curtain opened by **Monty Hall curtain** b, and the third curtain c. Define the following events
  - A, B, and C are the events that the prize is behind curtains a, b, and c, respectively
  - O is the event that Monty Hall opens curtain b
- The Monty Hall Problem can be restated as follows: Is $p(A|O) = p(C|O)$?
- Use Bayes' Rule

$$p(A|O) = \frac{p(O|A)\,p(A)}{p(O)}$$

$$p(C|O) = \frac{p(O|C)\,p(C)}{p(O)}$$

- Assume that the prize is randomly placed behind the curtains

$$p(A) = p(B) = p(C) = 1/3$$

Monte Hall Problem (iii)

- Compute all of the conditional probabilities for event O

$$p(O|A) = 1/2, \quad \text{if prize is behind } a\text{, Monty Hall can open either } b \text{ or } c$$
$$p(O|B) = 0, \quad \text{if prize is behind } b\text{, Monty Hall cannot open } b$$
$$p(O|C) = 1, \quad \text{if prize is behind } c\text{, Monty Hall can only open } b$$

(remember, the contestant picked curtain a)

- Compute $p(O)$ by

$$\begin{aligned} p(O) &= p(O|A)\,p(A) + p(O|B)\,p(B) + p(O|C)\,p(C) \\ &= (1/2)\cdot(1/3) + 0\cdot(1/3) + (1)\cdot(1/3) = 1/2 \end{aligned}$$

- Then

$$\begin{aligned} p(A|O) &= \frac{p(O|A)\,p(A)}{p(O)} \\ &= \frac{(1/2)\cdot(1/3)}{1/2} = 1/3 \end{aligned} \qquad \begin{aligned} p(C|O) &= \frac{p(O|C)\,p(C)}{p(O)} \\ &= \frac{1\cdot(1/3)}{1/2} = 2/3 \end{aligned}$$

If we switch we double our chances!

- Note that since $p(B|O) = 0$ then $p(C|O) = 1 - p(A|O)$ $= 1 - 1/3 = 2/3$

Monte Hall Problem (iv)

```
function [pwinSwitch,pwinNoSwitch]=montyhall(nTrials)
pwinSwitch = 0;
pwinNoSwitch = 0;
for trials = 1:nTrials
 win = randi(3,1);
 choice = randi(3,1);
 if choice == win
  pwinNoSwitch = pwinNoSwitch + 1;
 end
% Now, reveal an empty door
 doors = ones(3,1);
% Don't reveal a door with the prize
 doors(win) = 0;
% Don't reveal your choice
 doors(choice) = 0;
% If there is more than one door to reveal, pick one to not reveal
 if nnz(doors) > 1
  reveal = find(doors);
  doors(reveal(randi(2,1))) = 0;
 end
```

Monte Hall Problem (v)

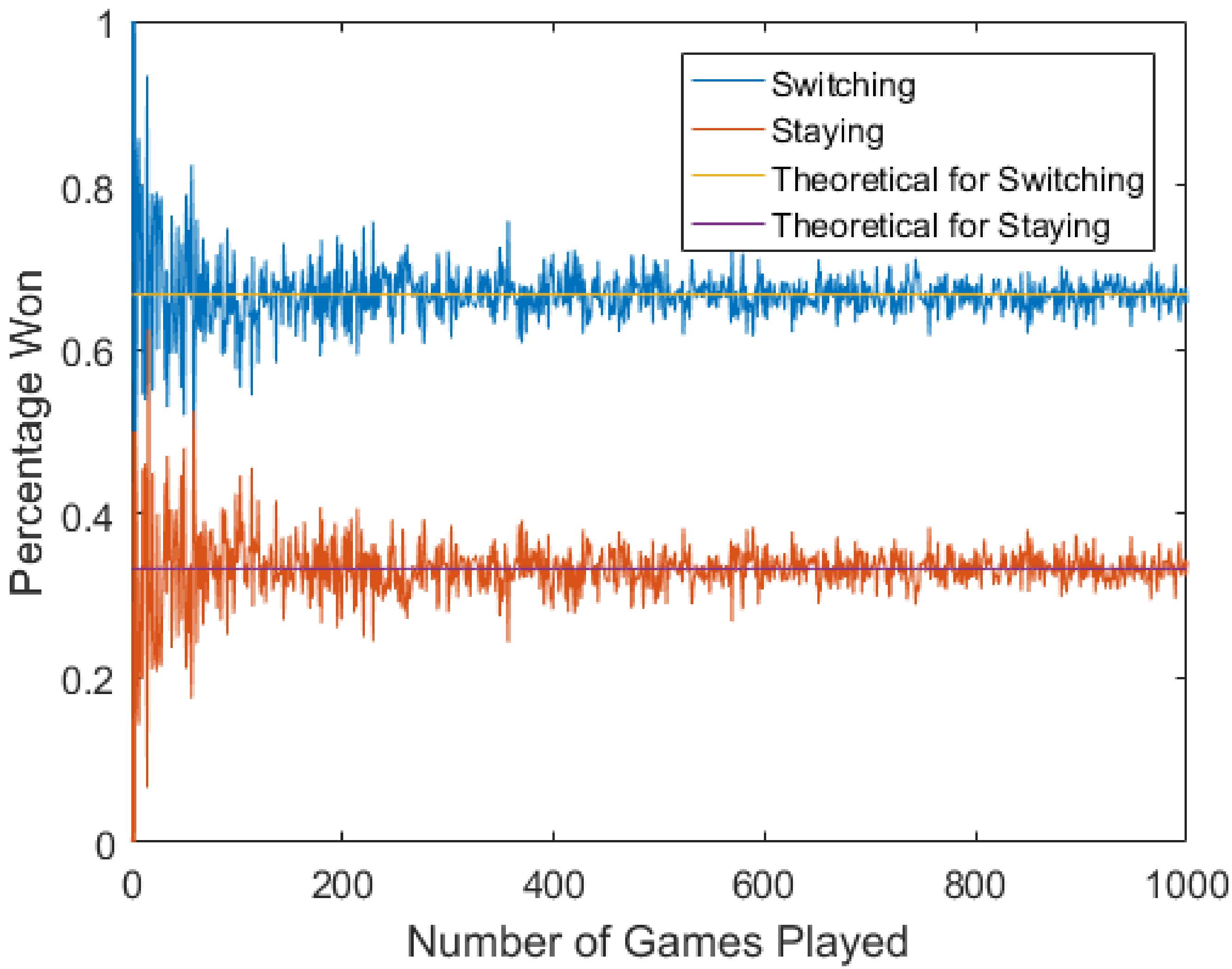
```
% Make the swap by putting choice back, and finding the index that
% has not been revealed, and was not your choice
 doors(choice)=1;
 newChoice = find(~doors);
 if newChoice == win
  pwinSwitch = pwinSwitch + 1;
 end
end
pwinSwitch = pwinSwitch/nTrials;
pwinNoSwitch = pwinNoSwitch/nTrials;
%fprintf(1, 'Total number of trials: %d\n', nTrials);
%fprintf(1, 'Probability of win without switch: %g\n', pwinNoSwitch);
%fprintf(1, 'Probability of win with switch: %g\n', pwinSwitch);
```

Monte Hall Problem (vi)

```
% Script to call code for 1:m number of runs
m=1000;winSwitch=zeros(m,1);winNoSwitch=zeros(m,1);
 for i=1:m,
  [pwinSwitch, pwinNoSwitch]=montyhall(i);
  winSwitch(i)=pwinSwitch;
  winNoSwitch(i)=pwinNoSwitch;
 end

plot([1:m]',winSwitch,[1:m]',winNoSwitch,[1:m]',2/3*ones(m,1),[1:m]',1/3*ones(m,1))
set(gca,'fontsize',12)
ylabel('Percentage Won')
xlabel('Number of Games Played')
legend('Switching', 'Staying', 'Theoretical for Switching', 'Theoretical for Staying')
```

Monte Hall Problem (vii)



Moments

- The random variable x is usually described in terms of its moments
  - The first two moments of x are given by the mean μ of x

$$\mu \equiv \sum_j x(j)\, p(x(j))$$

and the variance σ^2 of x

$$\sigma^2 \equiv \sum_j (x(j) - \mu)^2 p(x(j))$$

This is actually the "central second moment"

  - The quantity σ is often called the standard deviation
  - If $p(x)$ is considered to be a function defining the mass of several discrete masses located along a straight line, then μ locates the center of mass and σ^2 is the moment of inertia of the system of masses about their centroid

Expected Value

- The expected value or "average value" of a function $f(x)$ of a discrete random variable x is defined as

$E(x) = \mu$

$E(x-\mu)^2 = \sigma^2$

$$E\{f(x)\} = \sum_j f(x(j))\, p(x(j))$$

  - Note, the mean and variance are the expected values of x and $(x-\mu)^2$, respectively

- Notice that the expected value operator is linear so that

$$E\{a\, f(x) + b\, g(x)\} = a\, E\{f(x)\} + b\, E\{g(x)\}$$

for a and b arbitrary deterministic scalars, and $f(x)$ and $g(x)$ arbitrary functions of the random variable x

Functions of Discrete Variables

- If each scalar element x_i of $\mathbf{x}$ can take on a finite number, m_i, of discrete values $x_i(j_i)$, for $j_i = 1, 2, \ldots, m_i$, then there are $m_1 m_2 \cdots m_n$ possible vectors
  - For a complete probabilistic characterization of $\mathbf{x}$, its *joint probability function* $p(j_1, j_2, \ldots, j_n)$ is the probability that x_1 has its j_1th value, x_2 has its j_2th value, …, x_n has its j_nth value
  - The function $p(j_1, j_2, \ldots, j_n)$ is often written $p(x_1, x_2, \ldots, x_n)$
- The marginal probability mass function is

$$p(j_1) = \sum_{j_2=1}^{m_2} \sum_{j_3=1}^{m_3} \cdots \sum_{j_n=1}^{m_n} p(j_1, j_2, \ldots, j_n)$$

  - Note that $p(j_1)$ is the probability of a compound event; that x_1 takes on its j_1th value while $x_2, x_3, \ldots, x_n$ take on arbitrary possible values
    - Thus, a scalar random variable may represent an elementary or compound event, depending upon the dimension of the underlying space of events

Mean and Covariance

- The marginal probability functions are sufficient to fully probabilistically characterize the components of x, but to fully characterize x, it is necessary to specify $p(x_1, x_2, \ldots, x_n)$
  - As in the scalar case, it is customary to describe $p(x_1, x_2, \ldots, x_n)$ and x in terms of the moments of x
  - The mean of x is given by

$$\boldsymbol{\mu} \equiv E\left\{\mathbf{x}\right\} = \sum_{j_1=1}^{m_1} \cdots \sum_{j_n=1}^{m_n} \begin{bmatrix} x_1(j_1) \\ \vdots \\ x_n(j_n) \end{bmatrix} p(j_1, j_2, \ldots, j_n)$$

  - The covariance of x is given by

Note this is a symmetric matrix

$$R \equiv E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right\}$$

$$= E\left\{\begin{bmatrix} (x_1-\mu_1)^2 & (x_1-\mu_1)(x_2-\mu_2) & \cdots & (x_1-\mu_1)(x_n-\mu_n) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)^2 & \cdots & (x_2-\mu_2)(x_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (x_n-\mu_n)(x_1-\mu_1) & (x_n-\mu_n)(x_2-\mu_2) & \cdots & (x_n-\mu_n)^2 \end{bmatrix}\right\}$$

Covariance

- We adopt the following notations

$$\sigma_i^2 \equiv E\left\{(x_i - \mu_i)^2\right\} = \text{ variance of } x_i$$
$$\sigma_{ij} \equiv E\left\{(x_i - \mu_i)(x_j - \mu_j)\right\} = \text{ covariance of } x_i \text{ and } x_j$$

- The covariance matrix is commonly written as

$$R \equiv \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix}$$

where ρ_{ij} is the correlation of x_i and x_j, defined by

$$\rho_{ij} \equiv \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

- The correlation coefficient gives a measure of the degree of linear dependence between x_i and x_j
  - If x_i is linear in x_j, then $\rho_{ij} = \pm 1$; however, if x_i and x_j are independent of each other, then $\rho_{ij} = 0$
  - If $p(x_1, x_2, \ldots, x_n) = p(x_1)\, p(x_2) \cdots p(x_n)$ for all possible values of $\{x_1, x_2, \ldots, x_n\}$, then the random variables are independent, as discussed previously
  - Note that while pairwise independence is sufficient to ensure zero correlation of $\{x_1, x_2, \ldots, x_n\}$, it is not sufficient to ensure independence of $\{x_1, x_2, \ldots, x_n\}$

Covariance

- To investigate the definiteness of R in general, let

$$\boldsymbol{\mu} = E\left\{\mathbf{x}\right\} \quad \text{and} \quad z = \mathbf{c}^T\left(\mathbf{x} - \boldsymbol{\mu}\right)$$

where $\mathbf{c}$ is $n \times 1$ vector of arbitrary constraints

- Investigating the moments of z, we find

$$\mu_z \equiv E\left\{z\right\} = E\left\{\mathbf{c}^T\left(\mathbf{x} - \boldsymbol{\mu}\right)\right\} = \mathbf{c}^T\left(\boldsymbol{\mu} - \boldsymbol{\mu}\right) = 0$$

and

$$\begin{aligned}\sigma_z^2 &\equiv E\left\{\left(z - \mu_z\right)^2\right\} = E\left\{\mathbf{c}^T\left(\mathbf{x} - \boldsymbol{\mu}\right)\left(\mathbf{x} - \boldsymbol{\mu}\right)^T\mathbf{c}\right\} \\ &= \mathbf{c}^T E\left\{\left(\mathbf{x} - \boldsymbol{\mu}\right)\left(\mathbf{x} - \boldsymbol{\mu}\right)^T\right\}\mathbf{c} = \mathbf{c}^T R\, \mathbf{c}\end{aligned}$$

Since $\sigma_z^2 \geq 0$ and since $\mathbf{c}$ is an arbitrary vector, then R is always *at least* positive semi-definite. For diagonal R, the positive semi-definiteness of R agrees with our intuitive interpretation of σ_i^2; since $\sigma_i^2 < 0$ implies "better than perfect knowledge" or "less than zero uncertainty" in x_i, which is impossible!

- By letting $N \to \infty$ with the probability mass function $p(x_1(j_1), \ldots, x_n(j_n))$ being replaced by a *probability density function* $p(x_1, \ldots, x_n)$; then

$$p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n$$

is the probability that the components of $\mathbf{x}$ lie within the differential volume given by $dx_1, dx_2, \ldots, dx_n$ centered at $x_1, x_2, \ldots, x_n$

- Since all possible $\mathbf{x}$-vectors are located in the infinite sphere, it follows that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n = 1$$

  - Common shorthand notation

$$\int_{-\infty}^{\infty} p(\mathbf{x})\, d\mathbf{x} = 1$$

Probability (i)

- To obtain the probability use

$$p(a \le x \le b) = \int_a^b p(x)\ dx, \quad a \le b$$

  - Example: find k and $p(0.5 \le x \le 1)$

$$p(x) = \begin{cases} k\,e^{-3\,x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

  - Compute the following to find k

$$\int_{-\infty}^{\infty} p(x)\ dx = \int_0^{\infty} k\,e^{-3\,x}\ dx = -\frac{k}{3}e^{-3\,x}\Big|_0^{\infty} = \frac{k}{3}$$

    - So k must be 3 to be a valid pdf

  - Compute the probability

$$p(0.5 \le x \le 1) = \int_{0.5}^{1} 3\,e^{-3\,x}\ dx$$

$$= -e^{-3\,x}\Big|_{0.5}^{1} = -e^{-3} + e^{-1.5} = 0.173$$

Probability (ii)

- The distribution function is given

$$P(x) \equiv p(X \leq x) = \int_{-\infty}^{x} p(t)\ dt$$

- Note that

$$p(x) = \frac{d\,P(x)}{dx}$$

- Useful because $p(a \leq x \leq b) = P(b) - P(a)$
- Last example

$$P(x) = \int_{-\infty}^{x} p(t)\ dt = \int_{0}^{x} 3\,e^{-3\,t}\ dt$$

$$= -e^{-3\,t}\Big|_0^x = 1 - e^{-3\,x}$$

- Compute the probability

$p(0.5 \leq x \leq 1) = P(1) - P(0.5)$

No other integrals required if distribution function can be found

$$= \left(1 - e^{-3}\right) - \left(1 - e^{-0.5}\right) = -e^{-3} + e^{-1.5} = 0.173$$

Expectation

- The expected value of an arbitrary function $g(x_1, \ldots, x_n)$ is defined in terms of the density function as

$$E\left\{g(x_1, \ldots, x_n)\right\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n)\; dx_1 \cdots dx_n$$

  - Thus the summation signs of the discrete results are replaced by integral signs to obtain the corresponding continuous result
- Important to always think of the expectation as an integral
  - Standard integration rules apply for expectation
  - Helps to determine what to "pull out" of the expectation!
  - Uniform distribution example

$\int p(x)\, \partial x = P(x)$

Probability Density Function

$$p(x) = \frac{1}{\beta - \alpha}, \quad \text{for } \alpha \leq x \leq \beta$$

$$E\{x\} = \int_{-\infty}^{\infty} x\, p(x)\; dx = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}\; dx = \left.\frac{x^2}{2(\beta - \alpha)}\right|_{\alpha}^{\beta} = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{\beta + \alpha}{2}$$

Moments

- The r^{th} moment is defined by

$$E\{x^r\} = \int_{-\infty}^{\infty} x^r p(x)\ dx$$

If $x^r = y^2$

- Variance (*second central moment*) is given by

$E\{x\} = \mu$

$y^2 = (x - \mu)^2$

$$\begin{aligned} \text{var}\{x\} &\equiv \int_{-\infty}^{\infty} (x - E\{x\})^2 p(x)\ dx \\ &= E\{x^2\} - E^2\{x\}, \end{aligned}$$

comes from the *parallel axis theorem*

$(x^2)(p(x))$

- Uniform distribution example

$$\begin{aligned} \text{var}\{x\} &= \left(\int_{\alpha}^{\beta} \frac{x^2}{\beta - \alpha}\ dx\right) - \left(\frac{\beta + \alpha}{2}\right)^2 \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} - \frac{(\beta + \alpha)^2}{4} \\ &= \frac{1}{12}(\beta - \alpha)^2 \end{aligned}$$

see above

$E\{x\}^2$

Gaussian Random Variables

- Most widely used distribution for state estimation
  - Taking the limit as the number of coin flips, used to produce the histogram previously, approaches infinity leads to the *Gaussian* or *normal* density function for x

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

  with mean given by μ and variance given by σ^2
  - This function can also be expanded to the multidimensional case for a vector $\mathbf{x}$

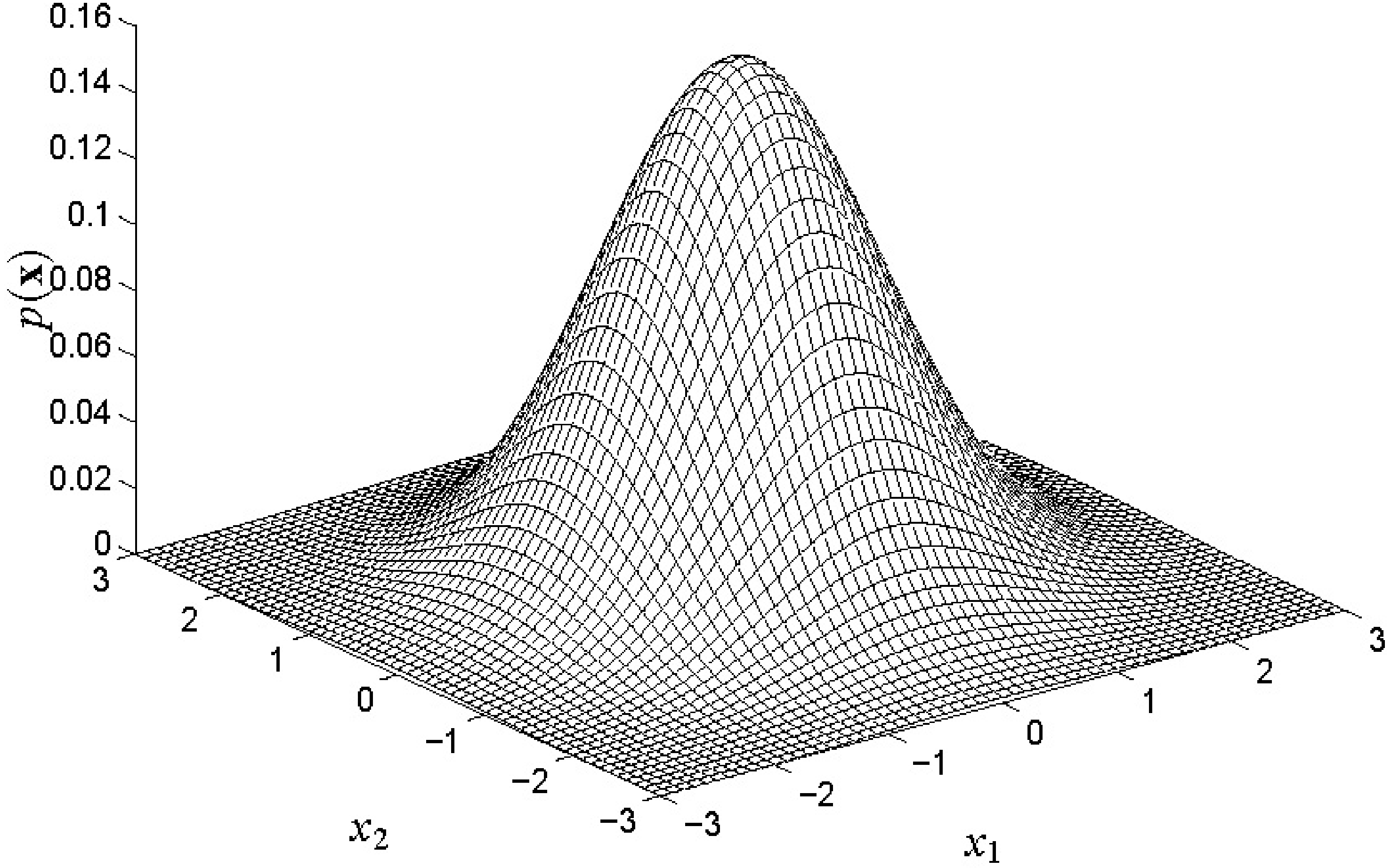
$$p(\mathbf{x}) = \frac{1}{[\det(2\pi R)]^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T R^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

  - The mean and covariance are sufficient enough to define this distribution
    - Simple notation for this distribution is given by

$$\mathbf{x} \sim N(\boldsymbol{\mu}, R)$$

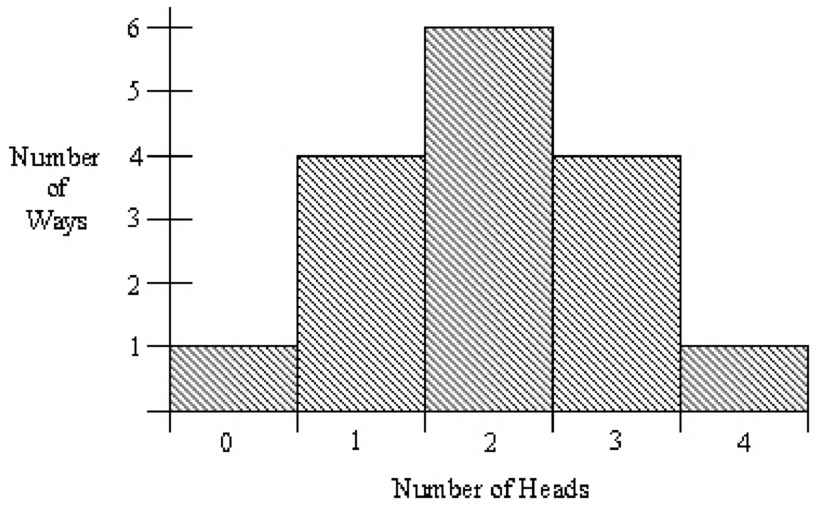
covariance matrix

- Plot of Gaussian function for two variables, with $\boldsymbol{\mu} = \mathbf{0}$ and $R = I_{2\times 2}$



Gaussian Random Variables

- The Gaussian distribution is important because of a very useful property that involves any distribution
  - The central limit theorem states that given a distribution with mean μ and variance given by σ^2, the sampling distribution (no matter what the shape of the original distribution) approaches a Gaussian distribution with mean μ and variance σ^2/N as N, the sample size, increases



As stated before even for a relatively small sample size the histogram looks like the classic “bell shape” form of the Gaussian distribution

- Simulation for general covariance matrix $\mathbf{x} \sim N(\mathbf{0}, R)$
  - One way is to take eigenvalue/eigenvector decomposition

$$R = V \Lambda V^T$$

$R = E\{(x-\mu)(x-\mu)^T\}$

$R = E\{(x)(x)^T\}$

  - Define new variable $\mathbf{z} \sim N(\mathbf{0}, \Lambda)$
    - Can easily sample $\mathbf{z}$ now since its covariance is a diagonal matrix
  - Then $\mathbf{x} = V\mathbf{z}$; let's check to make sure

zero mean

$$\rightarrow E\{\mathbf{x}\,\mathbf{x}^T\} = E\{V\,\mathbf{z}\,\mathbf{z}^T V^T\} = V \Lambda V^T = R \;\checkmark$$

    - Note, simply add the mean to $\mathbf{x}$ if it is not zero mean
  - Another way involves the Cholesky decomposition

$$R = L\,L^T$$

  - Now $\mathbf{z} \sim N(\mathbf{0}, I)$, which is a normal distribution
    - Then $\mathbf{x} = L\mathbf{z}$
  - Cholesky approach works only if $R > 0$, while the eigenvalue approach works if $R \geq 0$

```
function v=correlated_noise(r,m)
%function v=correlated_noise(r,m)
%
% The inputs are:
%      r = covariance matrix (nxn)
%      m = number of points
%
% The output is:
%      v = noise matrix (mxn)

% Decompose the Covariance Matrix
n=length(r);
[u,r_diag]=eig(r);

% Get Uncorrelated Noise
v_uncorr=randn(m,n)*r_diag.^(0.5);

% Get Correlated Noise
v=(u*v_uncorr')';
```

```
>> r=randn(3);r=r*r',m=10000;

r =

    2.2588    4.2062   -0.3237
    4.2062    8.8793    0.2851
   -0.3237    0.2851    2.4609

>> v=correlated_noise(r,m);cov(v)

ans =

    2.2469    4.1842   -0.3268
    4.1842    8.8298    0.2792
   -0.3268    0.2792    2.4259
```