Review of Underlying Mathematical Techniques

By four PhD graduate students

Section 1 Vectors & Matrices

By Yu Bi

Why is it necessary to review Vectors & Matrices (I)

- Vectors & Matrices are basic mathematical objects in linear system theory.

  Example:

  Suppose for given linear system, We recorded inputs and outputs over a time interval 1<= t<=N. Its linear difference equation is:

  $y(t) = -a_1 y(t-1) - \ldots - a_n y(t-n) + b_1 u(t-1) + \ldots + b_m u(t-m)$

Why is it necessary to review Vectors & Matrices (II)

We introduce the vectors:

$\theta = [a_1 \ldots a_n \; b_1 \ldots b_m]$

$\varphi(t) = [- \; y(t-1) \ldots -y(t-n) \; u(t-1) \ldots u(t-m)]^T$

$y(t) = \varphi^T(t)\theta$

Then select an approach so as to fit the calculated values. One general approach is least squares method, which will be introduced by Yongshen Pan

Therefore vectors & matrices are basic mathematics for us to develop a system.

Vector Operation(I)

- Vector: an array of elements, arranged in a column;
  $x = [\ x_1\ x_2 \dots x_n]^T$
- Vector Addition :
  $x+y = [x_1+y_1\ x_2+y_2 \dots x_n+y_n]^T$
- Scalar Multiplication:
  $kx = [\ kx_1\ kx_2 \dots kx_n]^T$

Vector Operations(II)

- Inner Product: $x^Ty = x_1y_1+x_2y_2+\dots+x_ny_n$
- Outer Product: it yields the matrix

$$xy^T = \begin{matrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ . & . & \dots & . \\ x_ny_1 & x_ny_2 & \dots & x_ny_n \end{matrix}$$

Vector Operaton(III)

- Vector Derivative

$$\dot{x}(t) = \lim_{\Delta t -> 0}\left(\begin{bmatrix} x_1(t+\Delta t) \\ x_2(t+\Delta t) \\ \cdots\cdots\cdots \quad \cdots\cdots \\ x_n(t+\Delta t) \end{bmatrix} \times \frac{1}{\Delta t} - \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdots\cdots \\ x_n(t) \end{bmatrix} \times \frac{1}{\Delta t} \right)$$

$$= \lim_{\Delta t -> 0}\left(\begin{bmatrix} x_1(t+\Delta t) - x_1(t) \\ x_2(t+\Delta t) - x_2(t) \\ \cdots\cdots\cdots \quad \cdots\cdots\cdots \quad \cdots\cdots \\ x_n(t+\Delta t) - x_n(t) \end{bmatrix} \times \frac{1}{\Delta t} \right)$$

$$= \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \cdots\cdots \\ \dot{x}_n(t) \end{bmatrix}$$

Matrix Operation (I)

- Matrix is an m×n rectangular array of elements in m rows and n columns.

  $A = [a_{ij}]$

- Maxtrix Addition

  $A + B = [a_{ij} + b_{ij}]$

  Prerequisite: the same number of rows and columns

- Scalar Multiplication

  $kA = [ka_{ij}]$

Maxtrix Operation(II)

- Matrix Multiplcation

$AB = C = [c_{ij}]$

$c_{ij} = \sum a_{ik} b_{kj}$

Prerequisite : A and B are conformable, that is the number of columns of A is equal to the number of rows of B.

- Vecotor-Matrix Product

$y = Ax$

$y_i = \sum a_{ij} x_j \quad (j = 1 \ldots n)$

Matrix Operation (III)

- Matrix Derivative and Integral

$\dot{A}(t) = [\dot{a}_{ij}]$ (Analogous to Vector Derivative)

- Matrix Identity Matrix

$I = [\zeta_{ij}] \quad \zeta_{ij} = 1$ if $i = j$ else 0;

- Matrix Determinant

$$|A| = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \cdots \sum_{\substack{l=1 \\ l \neq i, j \ldots k}}^{n} a_{1i} a_{2j} \cdots a_{nl}$$

Matrix Operation (IV)

- The inverse of a Matrix

  $A^{-1}A = AA^{-1} = I$

  $(AB)^{-1} = B^{-1}A^{-1}$

  all square matrices do not have inverses only those that are nonsingular have an inverse.

$$A^{-1} = \frac{1}{|A|} adjA$$

  adjA is the adjoint of A

Matrix Operation (V)

- Trace

$$Trace[A] = \sum_{i=1}^{n} a_{ii}$$

  Note: only square matrix has trace

  Trace[AB] = trace[BA]

- Rank

  The rank of matrix A is the dimension of the largest square matrix contained in A which has a nonzero determinant.

Matrix Operation (VI)

- Matrix Pseudo inverse

for nonsquare matrices used to describe systems of equations where the number of equations does not equal the number of unknows

if a matrix A has more rows than column

$A^{\#} = (A^T A)^{-1} A^T$

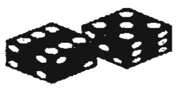
In this situation of linear equations this is the overdetermined case. The resulting solution, $x = A^{\#} y$ is best in a least-squares sense.

If a matrix A has ~~more~~ rows than column

$A^{\#} = A^T (A A^T)^{-1}$ less

This situation is underdetermined case. This solution resulting from Pseudo inverse is best in a least-squares sense, the vector $x = A^{\#} y$ is the solution of minimum length.

Probability and Random Variable



Junqing Sun
Sep. 2, 2004

1

- As the experiment is run repeatedly, the probability of an event is the theoretical limit of the *relative frequency* of the event (the ratio of the number of times the event occurred to the number of runs), as the number of runs increases to infinity.

- $0 \leqslant \Pr(E_i) \leqslant 1$ (E_i, i=1,2,…,n : All outcomes of the experiment.)
- $\sum_{i=1}^{n} \Pr(Ei) = 1$

2

- **Unions**
- **Mutually Exclusive Events**

  P(A + B+C+...) = P(A) + P(B)+ P(C) ...
- **Non-Mutually Exclusive Events**

  P(A or B) = P(A) + P(B) - P(A and B)

- **Intersections**
- **Independent events**

  P(A B) = P(A) * P(B)
- **Dependent Events**

  The occurrence of one event does affect the probability of the other occurring.

3

- **Conditional Probability**
- The probability of event B occurring that event A has already occurred is read "the probability of B given A" and is written: P(B|A)
- P(A and B) = P(A) * P(B|A)

- **Example:**
- P(A) = 0.20, P(B) = 0.70, P(B|A) = 0.40
- A good way to think of P(B|A) is that 40% of A is B. 40% of the 20% which was in event A is 8%, thus the intersection is 0.08.

4

Example

	B	B'	Marginal
A	0.08	0.12	0.20
A'	0.62	0.18	0.80
Marginal	0.70	0.30	1.00

5

- **Independence Revisited**
- The following four statements are equivalent

1. A and B are independent events
2. P(A and B) = P(A) * P(B)
3. P(A|B) = P(A)
4. P(B|A) = P(B)

6

Bayes' Theorem

- The Given (D) is made of three parts, the part of D in A, the part of D in B, and the part of D in C.

$$P(B \mid D) = \frac{P(B \text{ and } D)}{P(A \text{ and } D) + P(B \text{ and } D) + P(C \text{ and } D)}$$

- Inserting the multiplication rule for each of these joint probabilities gives

$$P(B \mid D) = \frac{P(D \mid B) * P(B)}{P(D \mid A) * P(A) + P(D \mid B) * P(B) + P(D \mid C) * P(C)}$$

7

Random Variables

Technically, a random variable is a function defined on the sample space. A *statement* about a random variable (such as an equation or inequality) defines an *event* (namely the set of outcomes for which the statement is true).

- ***Examples***

1. A coin is tossed ten times. The random variable X is the number of tails that are noted. X can only take the values 0, 1, ..., 10, so X is a discrete random variable.
2. A light bulb is burned until it burns out. The random variable Y is its lifetime in hours. Y can take any positive real value, so Y is a continuous random variable.

8

Distribution Function F(x)

- Specifying the probability with which different values are taken by the random variable.
- $F(x)=\Pr(X \leqslant x)$
- Probability Density Funtion
- $f(x)=dF(x)/dx$

9

- f(x) means the density of probability of the event that X takes a value in the vicinity of x.

$$f(x)=\lim_{dx\to 0}\frac{F(x+dx)-F(x)}{dx}=\lim_{dx\to 0}\frac{\Pr(x<X\leq x+dx)}{dx}$$

- If X takes a set of discrete values xi -- with nonzero probabilities Pi – f(x) is infinite at these values of x, such as, the outcome of the roll of a die

$$f(x)=\sum_i Pi\delta(x-xi)$$

10

Joint Probability Distribution

- In case of two, the probability of the occurrence of pairs of values of a given range:

  F(x,y)=Pr(Xx and Yy)

- Joint Probability Density Function:

$$f_2(x, y) = \frac{\partial^2 F_2(x, y)}{\partial x \partial y}$$

- The individual probability distribution and density functions for X and Y:

$$F(x) = F_2(x, \infty) = \int_{-\infty}^{x} dx \int_{-\infty}^{\infty} f_2(x, y)\, dy$$

$$f(x) = \int_{-\infty}^{\infty} f_2(x, y) dy$$

- If X and Y are independent:

  F(x,y)=Pr(Xx and Yy)= Pr(Xx) Pr(Yy)=F(x)F(y)

  The joint probability function is, then,

  f(x,y)= f(x)f(y)

- **Expected Value**
- Discrete random variable:

$$\mu = E(X) = \sum xi\, p(xi)$$

- Continuous random variable:

$$\mu = E(X) = \int x f(x)\, dx$$

- ***Example***
- Discrete case : When a die is thrown, each of the possible faces 1, 2, 3, 4, 5, 6 (the xi's) has a probability of 1/6 (the p(xi)'s) of showing. The expected value of the face showing is therefore:
- μ = E(X) = (1 x 1/6) + (2 x 1/6) + (3 x 1/6) + (4 x 1/6) + (5 x 1/6) + (6 x 1/6) = 3.5
- Notice that, in this case, E(X) is 3.5, which is not a possible value of X.

13

- **Mean squared value**

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

The quantity E[X]is also called the second moment of X.

- ***Variance***
- A measure of the 'spread' of a distribution about its average value.

$$V(X) = \sigma^2 = E(X - E(X))^2 = E(X^2) - E(X)^2$$

- *Notes*

Taking the square root of the variance gives the standard deviation, i.e.:

$$\sqrt{V(X)} = \sqrt{\sigma^2} = \sigma = s$$

14

$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$

- For independent Xi,

$$X = \sum_{i=1}^{n} X_i$$

- Then $\sigma_X^2 = \sum_{i=1}^{n} \sigma_{Xi}^2$
- **Covariance**

E[(X-E[x])(Y-E[Y])]==E[XY]-E[x]E[Y]

Degree to which one variable is related to another.

Correlation coefficient

$$\rho = \frac{E[XY] - E[x]E[Y]}{\sigma_X \sigma_Y}$$

It is a measure of the degree of linear dependence between X and Y.

15

Characteristic Function

$$g(t) = E[\exp(jtX)] = \int_{-\infty}^{\infty} \exp(jtx) f(x) dx$$

probability density function can be determined from:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-jtx) g(t) dt$$

The moments of x can be generated directly from :

$$E[X^n] = j^{-n} \frac{d^n g(t)}{dt^n} \Big|_{t=0}$$

16

Normal Distribution

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

We write $X \sim N(\mu, \sigma^2)$

The expected value μ.

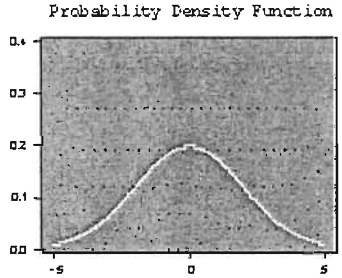
The variance is σ.

Standard Normal Distribution has expected value zero and variance one and is written as N(0,1).

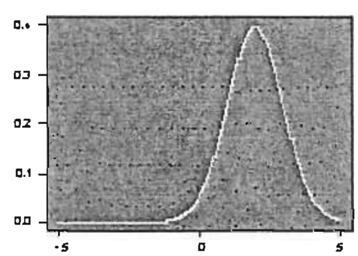
17

Examples:

Normal (0,2)
Probability Density Function



Normal (2,1)
Probability Density Function



18

Normal Distribution

- The distribution of the sum of normally distributed variables is also normal, whether they are independent or not.
- Under certain circumstances the distribution of the sum of independent random variables, each having an arbitrary distribution, tends toward the normal distribution as the number of variables in the sum tends toward infinity.

19

Normal Distribution

The bivariate normal distribution.

$$f_2(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp - \left[\frac{\frac{x_1^2}{\sigma_1^2} - 2\rho\frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \frac{x_2^2}{\sigma_2^2}}{2(1-\rho^2)} \right]$$

For n random variables,

$$f_n(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}|p|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{m})^T P^{-1}(\underline{x}-\underline{m}) \right]$$

With $\underline{x}^T = (x_1 x_2 \ldots x_n)$

The quantities $\underline{m} = E[\underline{x}]$

and $P = E[(\underline{x}-\underline{m})(\underline{x}-\underline{m})^T]$

are the mean and covariance of the vector, respectively.

20

- **Poisson Distribution**
- Typically, a Poisson random variable is a count of the number of events that occur in a certain time interval or spatial area.
- For example, the number of cars passing a fixed point in a 5 minute interval, or the number of calls received by a switchboard during a given period of time.

21

Poisson probability distribution

$$P(X=x) = \frac{m^x}{x!} e^{-m}$$

where $x = 0, 1, 2, \ldots, n$ and $m > 0$.

The following requirements must be met:

a. the length of the observation period is fixed in advance;
b. the events occur at a constant average rate;
c. the number of events occurring in disjoint intervals are statistically independent.

22

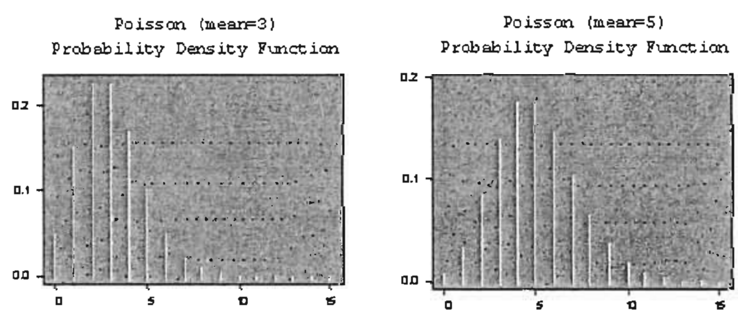
Poisson probability distribution

- To Poisson distribution
  E(X) = m; V(X) = m

- The Poisson distribution can sometimes be used to approximate the Binominal distribution with parameters n and p. When the number of observations n is large, and the success probability p is small, the Bi(n,p) distribution approaches the Poisson distribution with the parameter given by m = np.

23

Examples



24

Binomial Distribution

Typically, a binomial random variable is the number of successes in a series of trials, for example, the number of 'heads' occurring when a coin is tossed 50 times.

probability distribution :

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$x = 0, 1, 2, \ldots\ldots., n$

25

Binomial Distribution

•The trials must meet the following requirements:

a.the total number of trials is fixed in advance;

b.there are just two outcomes of each trial; success and failure;

c.the outcomes of all the trials are statistically independent;

d.all the trials have the same probability of success.

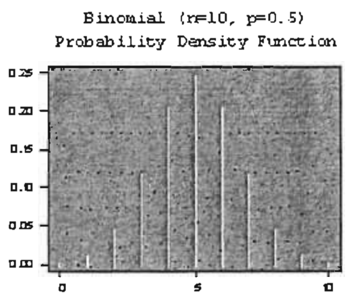
Expected Value: $E(X) = np$

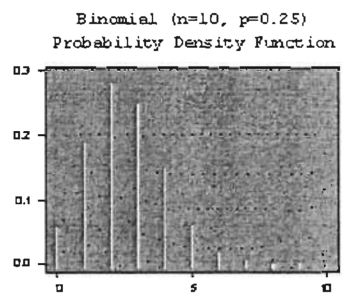
and

Variance: $V(X) = np(1-p)$.

26

Examples





27

Geometric Distribution

Typically, a Geometric random variable is the number of trials required to obtain the first failure, for example, the number of tosses of a coin until the first 'tail' is obtained, or a process where components from a production line are tested, in turn, until the first defective item is found.

28

Geometric distribution :

$P(X=x) = p^{x-1}(1-p)$

where p = success probability; $0 < p < 1$

$E(X)= 1/(1-p)$

$V(X)= p/(1-p)^2$

The trials must meet the following requirements:

a. the total number of trials is potentially infinite;

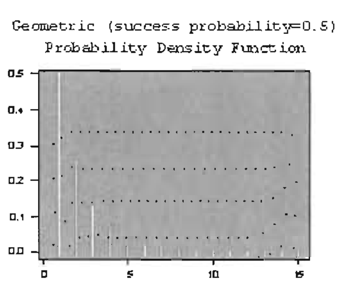
b. there are just two outcomes of each trial; success and failure;

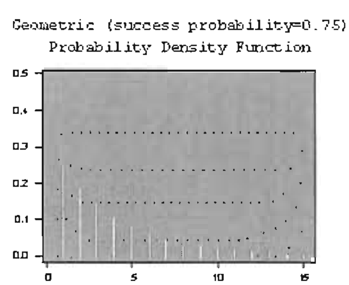
c. the outcomes of all the trials are statistically independent;

d. all the trials have the same probability of success.

29

Examples





30

?

Questions

31

Thank You !

32