LINEAR PREDICTION TECHNIQUES

27 March 2004

Summary of the lecture

We continued the theory of joint probability function, and understood what is meant by covariance of two random variables.

Then, we discussed how to express a random variable X as linear function of another random variable Y, minimizing the error in the L2 sense.

Then we extended this to a more unrestricted case where we expressed X as an affine function of Y. This yields an estimation technique called Wiener's method. Lastly, we extended this to a case where we express a random variable X in terms of N other random variables

Recapitulation of the previous lecture

What we saw in the last lecture is that associated with every random variable, is a belief structure, called its probability mask function or the probability density function.

When there is more than one variable their combined random variable is described in terms of the joint probability distribution.

Then we saw the concept of marginalization and conditional probability distribution.

Finally, we saw what is expected value of a random variable. We also found that if we were to predict a value for outcome of a random variable X, such that the error (X - guessVal) is minimum in the mean square or L2 sense, and then the guessVal should be the expected value of X.

If we restrict ourselves to X being estimated as a linear function "aY" of another random variable Y, then best value of 'a' = $E[XY]/E[Y^2]$

Covariance of Random Variables

What is the meaning of the symbol "E [XY]" that we saw last time ?

XY is a random variable such that it takes the outcome of variable X and multiplies it to the outcome of variable Y, and that is the value of random variable XY.

The expected value of the covariance of X and Y, namely E [XY] cannot be expressed in terms of E [X] and E [Y]. Let us see why.

Consider the examples given below

1. A coin is tossed. Suppose X is the random variable such that the value of X is +1 when the coin lands heads up, and has value -1 when it lands tails up.

Suppose Y is the same coin toss. Assuming a fair coin, the joint probability function is as shown below in figure 1.

$E[XY] = 0.5 * (1)(1) + 0.5 * (-1)(-1) = 1$ But, E[X] = 0 E[Y] = 0

2. Consider X to be outcome of a fair coin toss, same as X above. But Y is another coin toss, this time. So both the random variables are independent of each other.

The joint probability function would be,

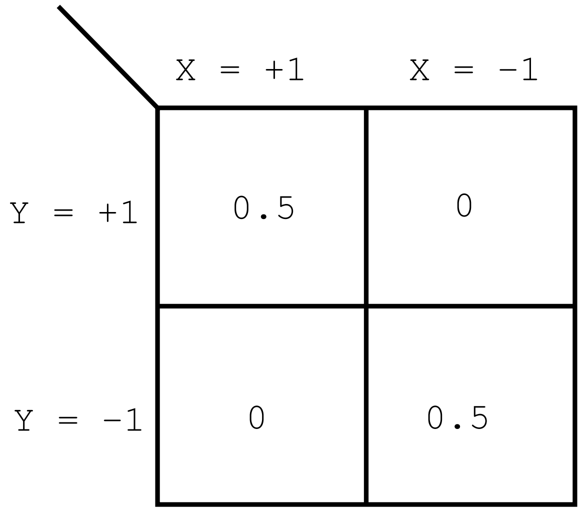

Figure 1: Joint Probablility distribution of two zero mean dependent random variables

$E[XY] = 0.5 * (1) + 0.5 * (-1) = 0$ But, E[X] = 0 E[Y] = 0

So, these two examples show that the E [XY] cannot be expressed in terms of E [X] and E[Y], as for the same values of E [X] and E [Y] different values of E [XY] are obtained.

In fact, random variables can be uncorrelated even if they are not independent. So if X and Y are independent, then surely they are uncorrelated. But, converse is not true, i.e. if X and Y are uncorrelated it does not necessarily imply their independence.

Mathematically,

$E[XY] = \int\limits_x \int\limits_y F_{X,Y}(x, y) x.y.dy.dx$

Linear Estimation Technique

If we were to estimate X using a vector of real numbers 'a', then 'a' is the vector that contains the centroid of each row.For each Y=y , the row would be some values. The centroid of , say row Y=+2,is the best estimate a_1 is the best estimate,i.e. $a_1 = E[X|Y = +2], a_2 = E[X|Y = +1]$....

Refer to Figure 3.

The curved line 'a' is the vector that minimizes the function (X-a).

Suppose we have to estimate the random variable X using a variable Y. In this technique we try to find out real numbers 'a' and 'b' such that the error function "X - aY - b " is minimum in the L2 sense.

We recall that if X is independent of Y, then the scale 'a' = 0, and b is simply the expected value of X. This is the first case discussed in the previous lecture, when the error function is (X - b).

But in general X and Y may not be independent.

Suppose X and Y are zero mean random variables. Zero mean random variables are ones whose expected values are equal to zero. Then, we can prove that 'b' is 0. Why this is so is simple to see. We know that X = a Y + b. So E [X] = a E[Y] + b, If E [X] = E[Y] =0, then b =0.

This implies that line aY + b passes through (0,0).

	X = +1	X = −1
Y = +1	0.5	0
Y = −1	0	0.5

Figure 2: Joint Probablility distribution of two zero mean independent random variables

Mathematically,

$E[(X - aY - b)^2]$

$= E[X^2 + a^2Y^2 + b^2 - 2aXY - 2bX - 2baY]$

$= E[X^2] + E[a^2Y^2] + E[b^2] - 2E[aXY] - 2bE[X] - 2bE[aY]$

Taking partial derivatives w.r.t a and b, we get

$a = E[XY]/E[Y^2]$

as, $E[X] = E[Y] = 0$

and

$b = E[X] - bE[Y] = 0$

Now, suppose that X and Y are not zero mean random variables, or E[X] and E[Y] are not zero. The line 'X = aY +b' is say somewhere , (in Fig 4, as line 1).

We can prove that line 'X= aY +b ' passes through (E[X] , E[Y]).

Let us make the random variables X and Y zero mean. This means we arithmetically shift the outcomes of X and Y by -E[X] and -E[Y] respectively.Now, that line stays where it is , but the labels of each row and column have have changed. So, the co-ordinate system has shifted. Now, we know that this line 'X = aY + b' has to pass through (0,0) in the shifted co-ordinate system.This corresponds to line 2 in Fig 4. The point now called (0,0) was called (E[X] , E[Y]) earlier.So that line passes through (E[X],E[Y]) in the "unshifted" system.

So, we know that '$X = aY + b$' always passes through ($E[X]$, $E[Y]$).

$E[X] = aE[Y] + b$ This implies that $b = E[X] - aE[Y]$

To find 'a', We try to minimize the mean square of error function E $[(X - aY - b)^2]$. If we do so we would find that we obtain a similar result as in the case of zeromean random variables.

$a = E[(X - E[X])(Y - E[Y])]/E[(Y - E[Y])^2]$

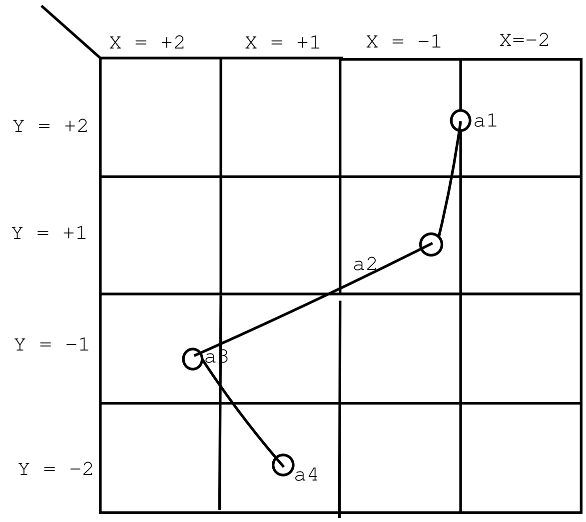

Figure 3: Expected Value curve, where expected values are a vector of real numbers

and the constant shift is,

$b = E[X] - aE[Y]$

This is just the same as for zero mean random variable case but for that both random variables are shifted by their expected values.

This method is called as Wiener's method of estimation.

Why is the linear estimator used?

1. Its computationally faster to compute and is easier to apply. This is so because we just have to calculate scalars 'a' and 'b' rather than functions. Once we have a, b it is just a multiplication and an addition to get a vector element of estimated X, rather than a table lookup.

2. The estimation can be done with much lesser data, i.e. only $E[XY]$ and $E[Y^2]$.

3. Many times it so happens that even the unrestrained vector is an affine function of Y. So in such cases, we might as well save computation time by estimating using linear or affine estimation.

Y estimated as a linear combination of N other random variables

Suppose we want to estimate the random variable Y using a linear combination of random variables $X_1, X_2..X_N$. This can be viewed as estimating the value that will come out of the unopened box , given the outcomes of N other random variables boxes.

So,

$Y = a_1.X_1 + a_2.X_2 + ...a_N.X_N$

Again, in this case we will try to see what values of a's will minimize the mean square estimation error. The error function is

$Y - A^T X$

where A is a vector of real nos. ,Y is a random variable and X is a vector of random variables.

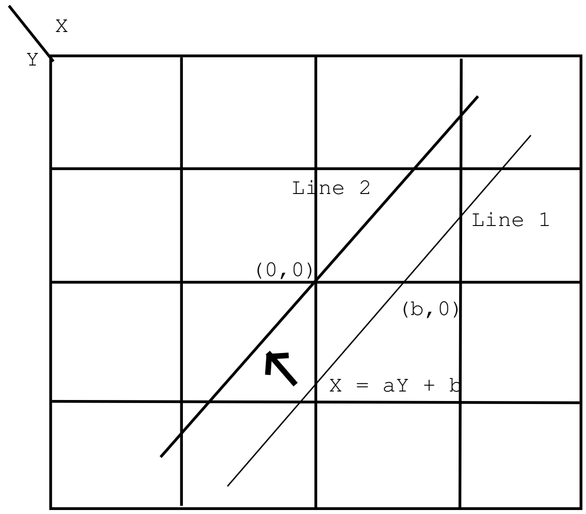

Figure 4: Expected Value curve, where expected values are "aY + b"

$$A = \begin{pmatrix} a_1 & a_2 & . & . & . \end{pmatrix}$$

The expected Value of

E $[(Y - A^T X)^2]$

is to be minimized

$E[(Y - A^T X)^2]$

$= E[Y^2 - 2YA^T X + A^T XA^T X]$

By Linearity of expectation,

$= E[Y^2] - 2A^T E[XY] + E[A^T XA^T X]$

Since $A^T X$

$$= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} \begin{pmatrix} X_1 & X_2 & . & . & . \end{pmatrix}$$

$$= \begin{pmatrix} X_1 \\ X_2 \\ \vdots \end{pmatrix} \begin{pmatrix} a_1 & a_2 & . & . & . \end{pmatrix}$$

$= X^T A$

So, using this to modify the last term of the above equation,

$= E[Y^2] - 2A^T E[XY] + E[A^T XX^T A]$

$= E[Y^2] - 2A^T E[XY] + A^T E[XX^T]A$

This is a familiar quadratic in A. The solution of this is,

$A = E[XX^T]^{-1}E[XY]$

This is again very similar to the result we obtained in the 2nd lecture, except that now we have shown the same concept in terms of random variables, i.e. in the vector space defined by random variables X ,Y.

The application of this result will be seen in the coming lectures.