Lecture Date – 26/3/04

Statistical Signal Processing

We started by learning about Random Variables.

A random variable X is real valued and is something, which is unknown to us. It is the value the variable x would have in the future. It is possible that the event that gives x a value has taken place, but till we don't know the result, x remains to be the random variable X.

Random variables are like a secret (a real value) in a closed box, and we guess what this could be, without opening the box.

This guesswork is based on what we call a "belief system". E.g.: Suppose we wanted to guess the outcome of a fair coin toss. Some of us would say that there is an equal probability of getting a head or tail. We give the occurrence of head and tail both a probability of ½ each because we believe that for some reason.

Since we don't know the value of the random variable, it can take any one value from a given set of finite values. Such variables are known as Discrete Random Variables. With each such value a probability is associated (depending upon our belief system). This set of probabilities is the Probability Mass function (pmf).
E.g. If we throw a dice, 1,2,3,4,5 or 6 can occur each with lets say a probability of 1/6. The pmf for this is {1/6, 1/6, 1/6, 1/6, 1/6, 1/6}.

It is denoted by $p_{\mathbf{X}}(x)$, the pmf of the random variable X. It is the probability that the value of X obtained on performance of the experiment is equal to x. 'x' is the argument of the function and is a variable and very different from X.
Some important properties of pmf

1. Since $p_{\mathbf{X}}(x)$ is probability, $0 \leq p_{\mathbf{X}}(x) \leq 1$, $\forall\, x$.
2. $\sum_{x} p_{\mathbf{X}}(x) = 1$. This true as the random variable will be assigned some value definitely.

Similarly if we have 2 events, E.g. throwing 2 dice. Then the pmf associated with this event (which, is made up of 2 events) is given by what is called the 'Joint pmf'. It is important to see that a Joint pmf contains more information then the individual pmfs.
Consider the example of throwing 2 dice. Suppose the outcome of this event was that the 1 dice always tries to match the other. If we looked at the die individually, we would think that all the values are equally probable and may not realize that they actually trying to be equal.

The joint pmf for this would be 2-D table, with values of dice X on 1 dimension and that of Y on the other.

$X\backslash Y$	1	2	3	4	5	6
1	1/6	0	0	0	0	0
2	0	1/6	0	0	0	0
3	0	0	1/6	0	0	0
4	0	0	0	1/6	0	0
5	0	0	0	0	1/6	0
6	0	0	0	0	0	1/6

Fig 1: Joint pmf of 2 die when they are thrown such that 1 always tries to match the other.

Like this we can have the joint pmf of 'n' random variables, which would be an N dimensional table. The joint pmf for 2 random variables can be written as

$p_{\mathbf{X},\mathbf{Y}}(x, y)$, the probability that X=x and Y=y.

In terms of closed boxes joint pmf means that we have 1 big box containing 2 boxes with X and Y inside them. Opening the big box means opening the 2 smaller boxes automatically and the joint pmf is the label on the big box.

Suppose, if we had to open only one of these boxes say X, then we can find the pmf for Y from the joint pmf as below. This is called the 'conditional pmf' and is read as the pmf of Y given X equal to 'x'

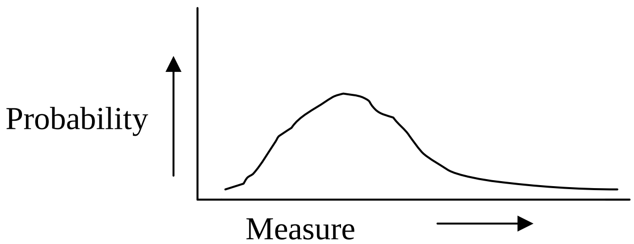
$$p_{\mathbf{Y}|\mathbf{X}=\mathrm{x}}(y) = \frac{p_{\mathbf{X},\mathbf{Y}}(\mathrm{x}, y)}{\sum_{y} p_{\mathbf{X},\mathbf{Y}}(\mathrm{x}, y)}$$

Suppose given the joint pmf of X,Y, we want to find the pmf of X. This is like saying we don't care about Y at all i.e. given the joint pmf we want to find the individual pmfs. This is called the 'Marginal pmf' and is found as

$$p_{\mathbf{X}}(x) = \sum_{i} p_{\mathbf{X},\mathbf{Y}}(\mathrm{x}, y_i)$$

Pmf is possible in the discrete domain, but for continuous domain, the random variable can take a value from infinite values. Thus for continuous valued random variables we define the Probability density function (pdf). It is the ratio of the probability of the domain to the measure of that domain and at a point it is the limit of this ratio as the measure tends to zero.

E.g. We have a dart board, and want the probability of hitting a point on the board. In the 1-dimensional case, the measure would be length and the pdf could be shown as below.



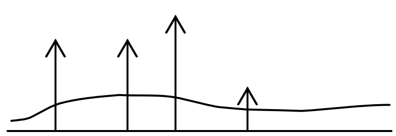
It is denoted as $f_X(x)$.

The pdf satisfies the following properties

1. $f_X(x) > 0 \,, \forall\, x$.
2. The area under the curve must be 1 i.e

$$\int_x f_X(x)\,dx = 1$$

Suppose we have lumped and continuous probabilities together like in the regulator of a fan. We have lumped probability at the standard fan speeds, but continuous between them. This can be shown by dirac delta functions in the pdf as shown below.



<u>Expected value of a random variable X</u>

This is the expected value of - the average value of X over all the possible futures. In other words, if we could perform an experiment 'n' times to find the value of X, then the mean of all those values would be the expected value of the random variable X.

Mathematically, $E[X]$ the expected value of X is

$$E[X] = \sum_{x} p_X(x)\, x \qquad \textit{or}$$

$$E[X] = \int_{x} f_X(x)\, x\, dx \qquad (f_X(x)\, dx \text{ is the probability of } x)$$

Linearity of expectations

E[X+Y] = E[X] + E[Y]

Proof:

$$E[X+Y] = \int_{x}\int_{y} (x+y)\, f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{x}\int_{y} x\, f_{X,Y}(x,y)\, dy\, dx \;+\; \int_{x}\int_{y} y\, f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{x} x \int_{y} (f_{X,Y}(x,y)\, dy)\, dx \;+\; \int_{y} y \int_{x} (f_{X,Y}(x,y)\, dx)\, dy$$

$$= \int_{x} x\; f_X(x)\; dx \;+\; \int_{y} y\, f_Y(y)\, dy$$

$$= E[X] + E[Y]$$

Also, if we have a function g(X) (i.e. a function which depends only on X), then

$$E[g(X)] = \sum_{x} p_X(x)\, g(x)$$

This can be seen from the fact that,

$$p_{g(X)}(g(x)) = \sum_{x:\, g(x)} p_X(x)$$

Estimators of X

We want to predict the value of X, such that the expected value of the mean square error is minimum i.e. $E[(X-a)^2]$ is minimum, where a is the estimated value of X.

$$E[(X-a)^2] = E[X^2] - 2a\, E[X] + a^2$$

Differentiating w.r.t to a

– $2\,E[X] + 2a = 0$

$a = E[X]$

The value of 'a' comes out to be $E[X]$. This is the best estimator for X, which is the expected value of X itself.

Suppose we had the Joint pdf of X & Y. Now, if we open the box for Y, we can predict the value of X given Y=y.

From above we observe 'a' will be,

$a = E[X \mid Y = y]$ (the centroid of the row with Y=y)

Thus 'a' is a function of y. This again is the best estimator of X given Y, and 'a' can be any function of y such that the cost is minimum.

If we had to restrict 'a' to be only a linear function of Y, then we would be basically trying to minimize $E[(X - aY)^2]$ and 'a' comes out to be

$$a = \frac{E[XY]}{E[Y^2]}$$

We notice that this expression is the exactly the same as that for – the scalar 'a', when two vectors y and x are given and we want to fit them together (we would scale one of them by a the scalar 'a').

We realize that random variables are vectors. They can be scaled and added just like any two vectors and the above expression enforces this fact.

What remains to be seen is why would we want a linear estimation for X as opposed to the best estimation.

1. One reason is that it is mathematically simpler.
2. A lot less data is required for the linear estimate than for the best estimate. For the linear estimate we just need E[XY] i.e. is covariance of X & Y (if X , Y are zero mean) whereas for the best estimate we need the entire joint pdf.
3. In a lot of cases, the linear estimate is the best estimate, so basically it is good enough.