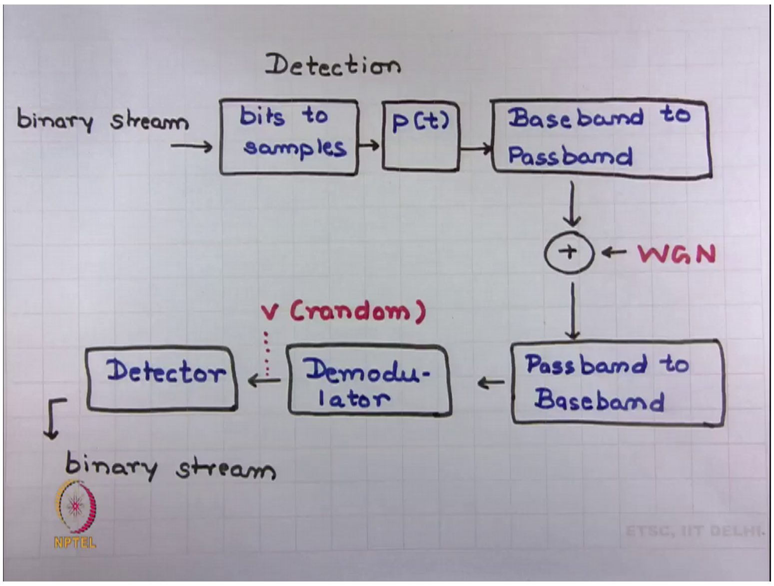
Principles of Digital Communication
Prof. Abhishek Dixit
Department of Electrical Engineering
Indian Institute of Technology, Delhi

Lecture - 30
Detection
Maximum Aposteriori Probability (MAP) Detector & Maximum Likelihood (ML) Detector

Good morning. Welcome to new unit and today, we will be starting with Detection. So, let us first see how detection fits into the scheme of things that we have carried out in this course.

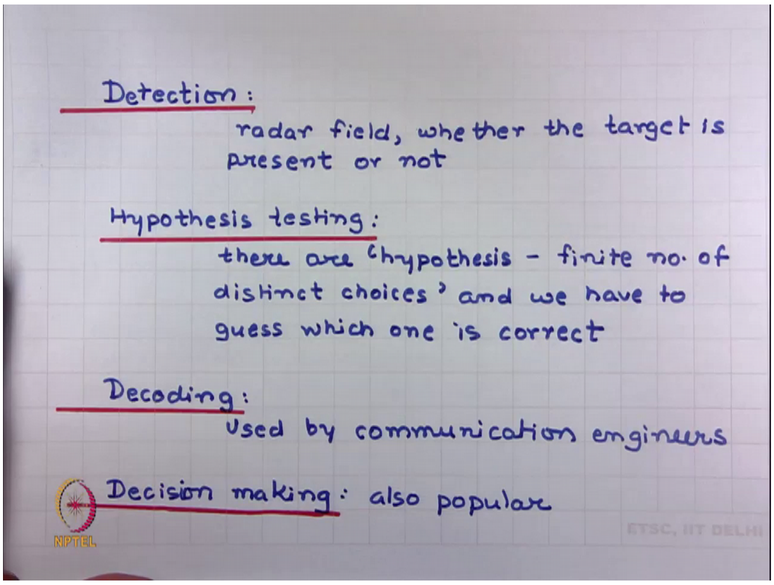
(Refer Slide Time: 00:32)



So, we said that in modulator we start with input as a binary stream, then we have bits to samples converter which takes in this binary stream, converts it into a weighted impulse strain. And, then we have a filter which converts this weighted impulse train to weighted pulse train, then we do baseband to pass band conversion, then you have a channel and, after the channel you expect that some noise has added in and then you go from this pass band to baseband domain. And, then we have a demodulator which carries out baseband demodulation.

Then we have a detector which based on what it is receiving, it is trying to get this binary stream out and today we would be starting with detection, where we will be focusing on this block right. So, we till now have seen modulator which is this part, we have seen and understood what is channel, we have understood what is wide Gaussian noise. We have clearly understood this pass band to baseband conversion, we have looked into some examples of demodulators like matched filters and correlators and today we will be seeing about this detector. So, what is the job of a detector? Detector sees this random variable and based on this random variable it guesses about this binary stream. So, let us look at some synonyms of detection.

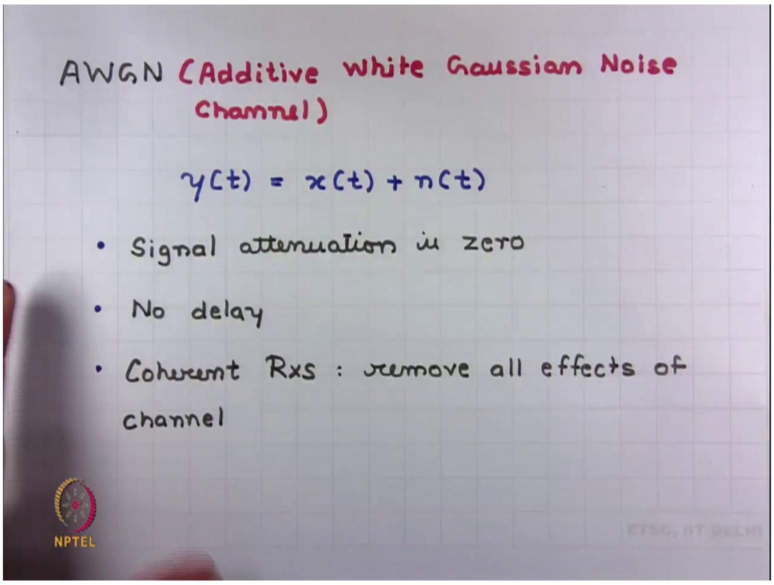
(Refer Slide Time: 02:15)



So, we use some other words like hypothesis testing, decoding and decision making for detection and we will also use them interchangeably because, they mean one and the same thing. So, let us look at some history and the context of these words for example, this word detection is influenced from radar theory. So, in radar the main object is to identify whether the target is present or not. In hypothesis testing there are hypothesis, what is hypothesis? Hypothesis means that you have finite number of distinct choices, this finite is important and then based on some observation you have to guess which choice is correct ok. In decoding we also do similar thing, but this is the word which is loved by communication engineers and thus, we will also use it. Decision making also looks like an appropriate choice which also means the same thing.

So, we would use these words also interchangeably namely detection, hypothesis testing, decoding and decision making alright. So, let us look at this detector again and what we have in here is we would study this field of detection for a white Gaussian noise channel or more precisely for an additive white Gaussian noise channel. Because, this white Gaussian noise we assume acts to the signal, we have seen this additive white Gaussian noise channel before as well. So, we will study all this detection theory for this additive white Gaussian noise channel. So, the object is same as before the detector observes a random variable and based on this observation it find out which hypothesis is correct. Let us look at this additive white Gaussian noise channel again.

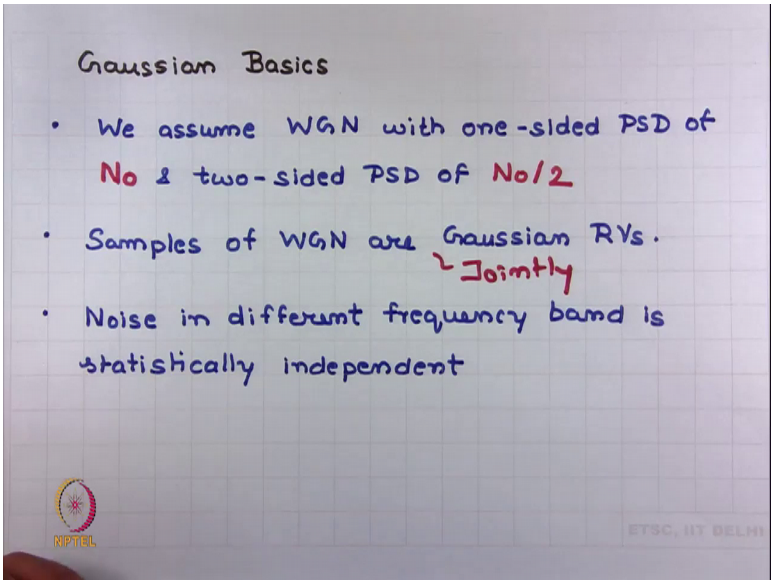
(Refer Slide Time: 40:13)



So, in Additive White Gaussian Noise channel we assume that the output is input plus noise. In practice this input would also be attenuated by the channel, but here you see that we do not assume any attenuation for the input. We assume that the output is input plus noise, there is no attenuation and nothing like that.

We also assume that output arrives at the same time as input is transmitted, we assume that there is no propagation delay because, these issues are dealt independently. The detection issues are completely isolated from the issues of attenuation and from the issues of propagation delays. To start with we assume that we have a coherent receiver and so, all these issues of frequency offsets and phase offsets are not there. So, this is the assumption with which we will like to start with detection. Let us also revise some of the basic stuff from the Gaussian noise.

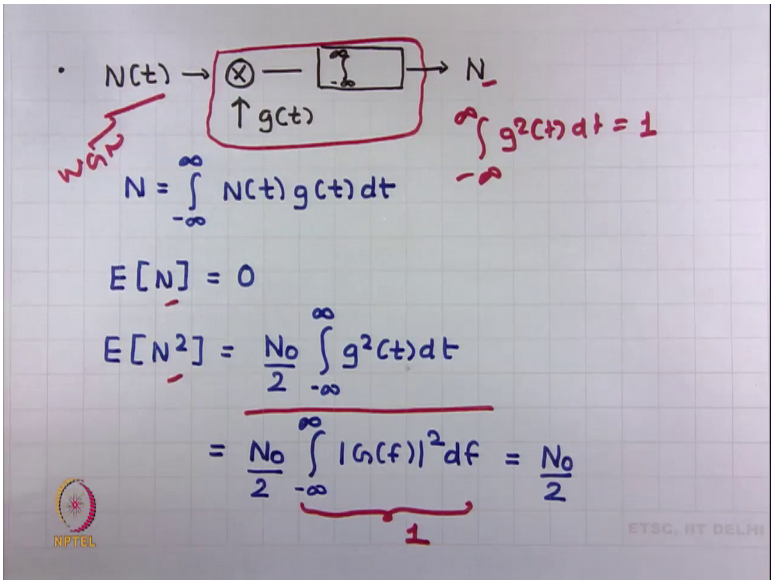
(Refer Slide Time: 05:19)



So, in the context of the noise we assume the noise to be white Gaussian noise, we have described in detail what is this white Gaussian noise. Refer to lecture 17 again, if you have forgotten about this white Gaussian noise, we will assume that we have a white Gaussian noise and the power spectral density or namely one sided power spectral density of that noise is N naught and thus the two sided power spectral density of this noise would be N naught by 2 and you must also know that samples of this white Gaussian noise.

So, if you take the sample of a random process what we get? We get random variables and the samples of this white Gaussian noise are jointly Gaussian random variables ok. So, it is a jointly Gaussian random variables because, it is a Gaussian noise, it is a Gaussian process and because it is a Gaussian process its samples will be jointly Gaussian random variables. We also have seen in lecture 17 that noise and different frequency band is statistically independent. We are revising this because, these ideas we will also use in detection.

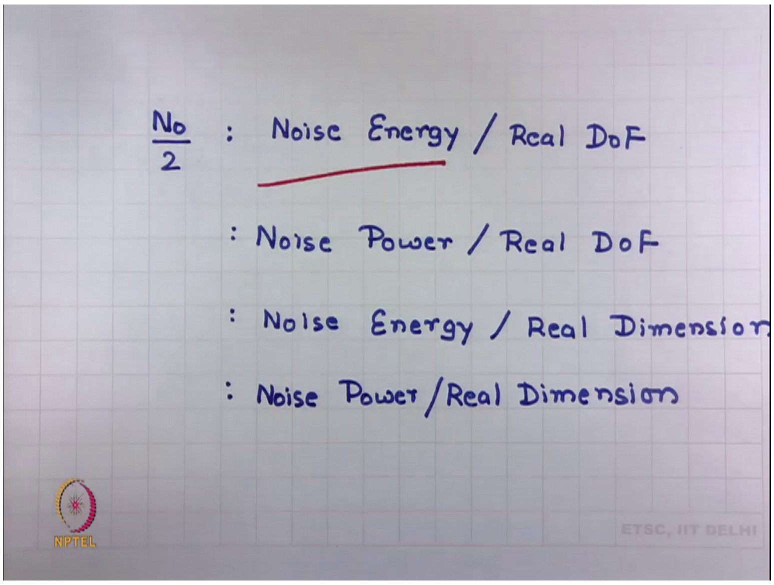
(Refer Slide Time: 06:35)



Also what we have seen is if you have a white Gaussian noise at the input of a correlator. So, this is a correlator structure and this correlator let us assume is fed with an orthonormal function g t; it is an orthonormal; that means, its energy is 1. So, if a white Gaussian noise is fed to a correlator, at the output of a correlator what we have? We have a random variable and let us assume that this random variable is N. So, N is nothing, but it is a linear functional of this random process N t with this deterministic function g t ok. We have seen that the expected value of this random variable is 0 and we have also seen that the variance of this random variable can be easily computed like this ok.

And this is same as this thing from Parseval's theorem and because it is an orthonormal function this is 1. And so, we get that the variance of the output noise is N naught by 2 ok, these ideas must be remembered. So, what I am emphasizing is if you have a white Gaussian noise, if you pass it through a linear filter a correlator is also an example of a linear filter; what you end up with is a Gaussian random variable, particularly if that linear system is a correlator then at the output of a correlator which is fed with an orthonormal function the variance of that random variable that you obtain is N naught by 2.

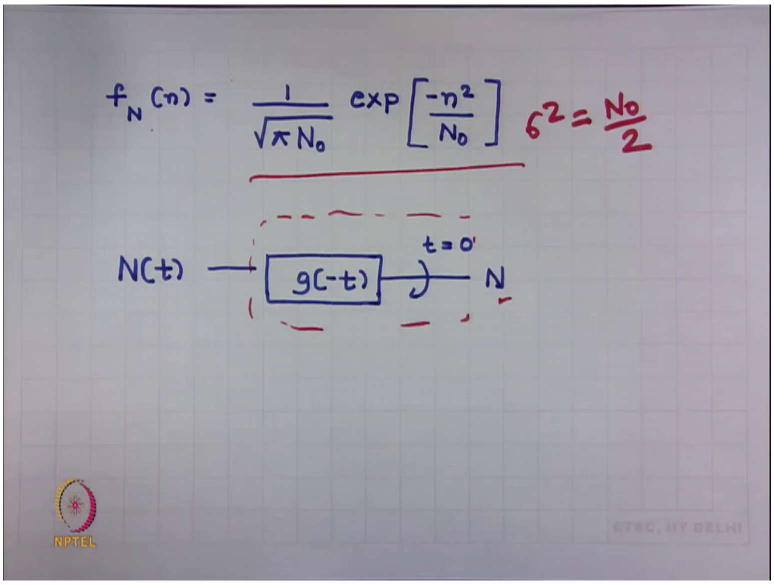
(Refer Slide Time: 08:11)



And what is this N naught by 2? Of course is power spectral density, but this is also the noise energy that is available for real degree of freedom. Why is this real degree of freedom? Because, correlator extracts the signal component along one-dimension right. So, the noise energy that we have would be per real degree of freedom.

We have also seen in one of the lecture that whether you talk about noise energy or noise power it is one and the same thing, if we are talking about per degree of freedom. We can also think about this as noise energy per real dimension or noise power for real dimension, energy and power for a noise per dimension is one and the same thing ok. So, remember these ideas we have covered them before, but I am just revising them for you so, that there are no confusions anymore.

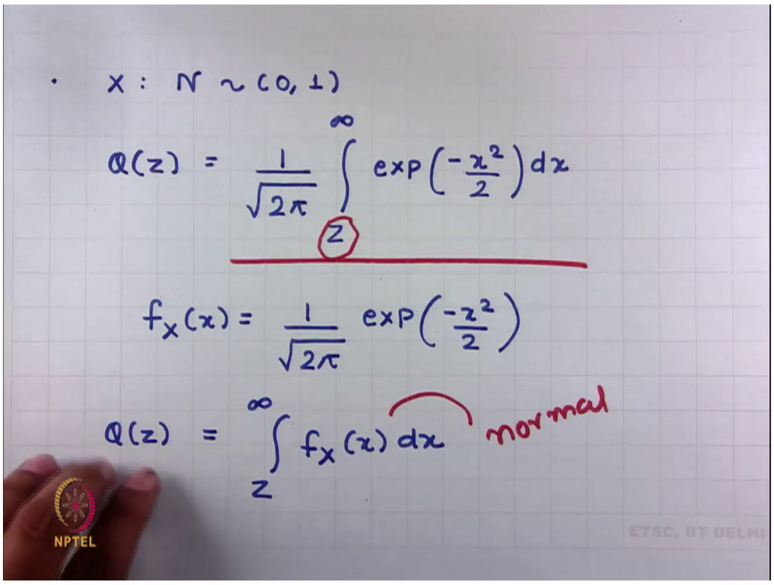
(Refer Slide Time: 09:08)



So, if N is a Gaussian random variable we know that is probability density function will be given by this. So, the variance is N naught by 2 and if you know that you have a Gaussian random variable with mean 0 and variance of N naught by 2; this is the probability density function of that random variable. So, what we show in here is that if you pass this N t white Gaussian noise through a matched filter, we have seen in one of the previous lectures that matched filter is same as a correlator.

So, you would expect the same output. So, again you have at the output of this matched filter random variable N. So, whatever we have said in the context of correlator also holds here. Let us now introduce few more concepts related to these Gaussian random variables ok.

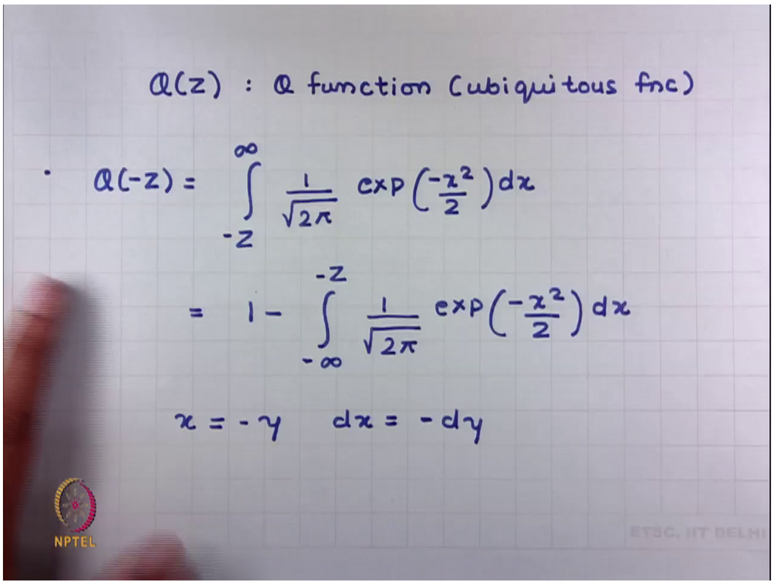
(Refer Slide Time: 10:08)



One very important function that you encounter in detection is this Q function. And what is this Q function? For example, Q of z Q of z is nothing but it is this integration. So, you should notice carefully that this z corresponds to this limit of integration. So, what does this Q of z? Q of z is nothing, but you take the probability density function of a normal random variable. So, if I take probability density function of a normal random variable, normal random variable is a Gaussian random variable with mean 0 and variance of 1.

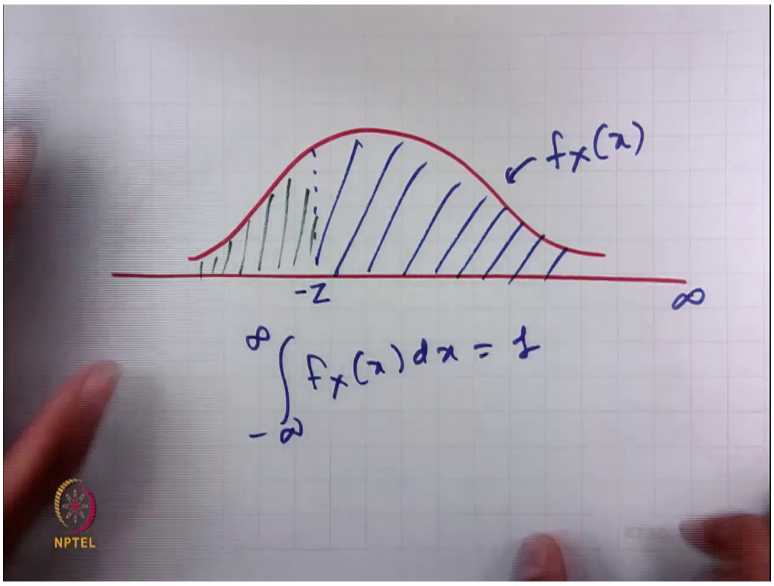
So, if you take a probability density function of a normal random variable which is this and if you integrate this probability density function from z to infinity. This integration corresponds to Q function, particularly Q of z and this is a very important integration that we will find lot of use. And thus, this Q function is really important and it is in ubiquitous function. So, whenever you see digital communication you would see this Q function.

(Refer Slide Time: 11:22)



So, you should make yourself comfortable with this Q function. Let us now see some properties of these Q function. For example, what is Q of minus z? What changes? Nothing changes now this integration starts from minus z right, earlier for Q of z it was starting from z; now if you have minus in here the limits would change from z to minus z the change is trivial ok. Now, let us look at this carefully.

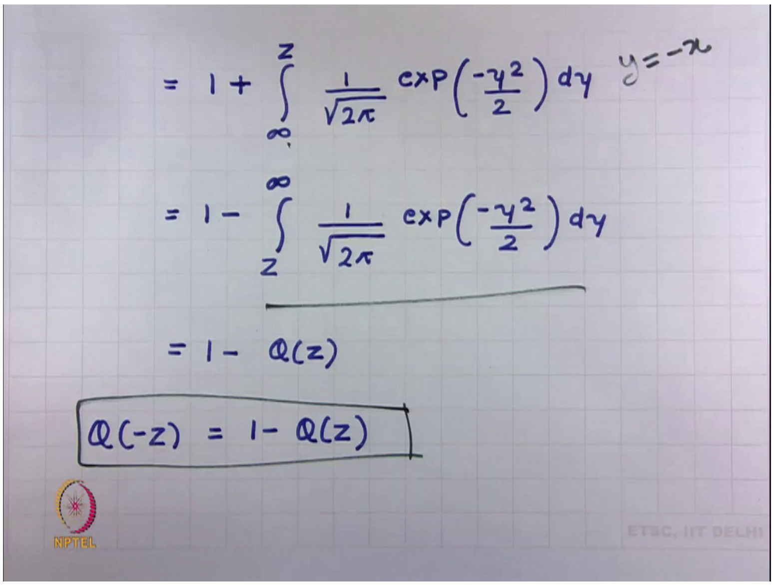
(Refer Slide Time: 11:59)



So, what we are doing in this integration is suppose you have a pdf, let us say we have a pdf; the probability density function. What is the area of a pdf? The area of a pdf we all know is 1 and so, this Q of minus z is calculating the area from minus z to infinity, this area and this area would be nothing, but 1 minus this area; it is a green area. So, the total area is 1, the blue area is 1 minus green area.

So, this is the idea that we are using. So, this gives me a blue area and this blue area is 1 minus green area. What is the limit? To compute this area you have to compute the integration from minus infinity to minus z and the probability density function is same. Now, let us make some changes in the variable. So, let us assume that x is minus y so, d x will be minus d y. In that case what would happen?

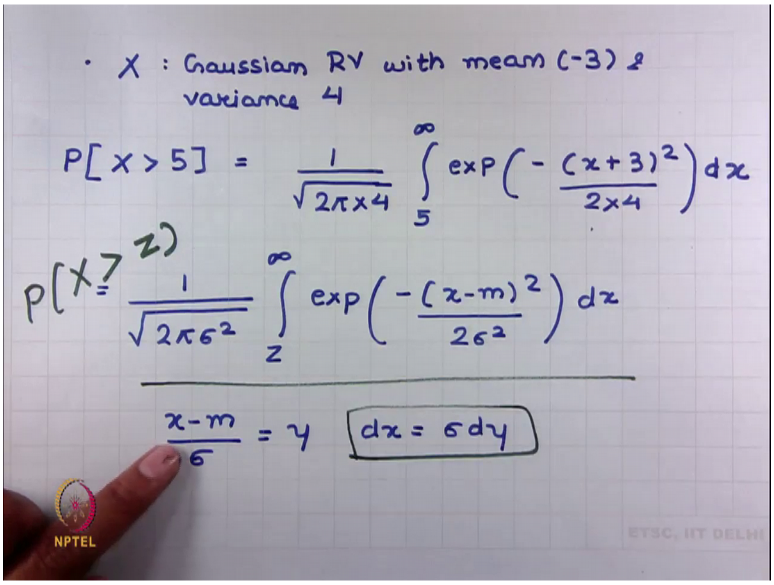
(Refer Slide Time: 13:24)



So, first we would see that if d x is minus d y there would be a sign change here. So, we will have 1 plus because x is minus infinity and y is minus x, this will become infinity minus that will become z, minus x square will become minus y square, d x becomes d y with minus, but we have already absorbed this in here. So, this integration boils down to this integration alright.

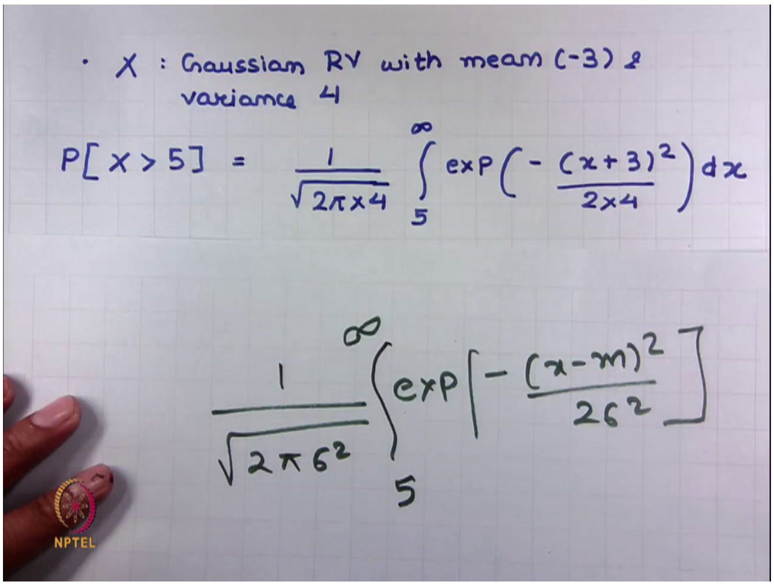
And what is this? So, you can again bring in a minus sign here by changing the order of this limits. So, I can have a negative sign here and I can flip the limits. So, instead of this integration going from infinity to z you have to calculate this integration from z to infinity. And what is this thing? This is nothing, but Q of z. So, we have got a neat property that Q of minus z is 1 minus Q of z and this property must be known and must be remembered as well.

(Refer Slide Time: 14:44)



Now, let us try to make our self more familiar with this Q function and to do so, let us pose a simple question to you. Let us now assume that X is a Gaussian random variable, it is not a normal random variable, it is a Gaussian random variable. So, we have got some mean, mean of minus 3 and a variance of 4 and the question that we are asking is what is the probability for which this X is greater than 5. So, we have to write the probability density function of this random variable X and you know that the probability density function for a Gaussian random variable is given by this right.

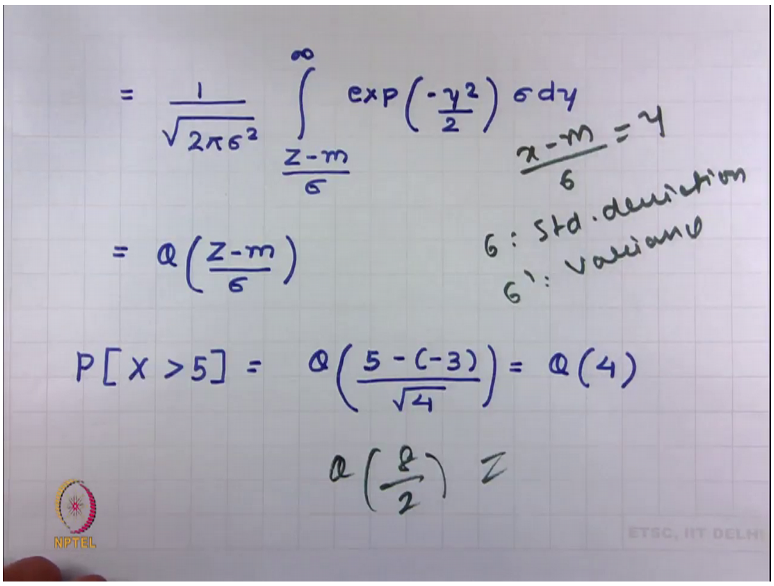
(Refer Slide Time: 15:19)



So, here the variance is 4. So, instead of sigma square we have to substitute 4 and mean is minus 3. So, we have plus 3 here because, here we have x minus m. So, it becomes x plus 3, sigma square is 4 and we have to find the probability for which X is greater than 5; that means, you have to integrate this probability density function from 5 to infinity, X should be greater than 5. So, this integration goes from 5 to infinity alright. So, we can find this probability X greater than 5 by simply computing this integration and to think about this integration let us think about a general case because, then you can easily substitute for a special case. So, for a general case if I have to find the probability of let us say a random variable X greater than Z.

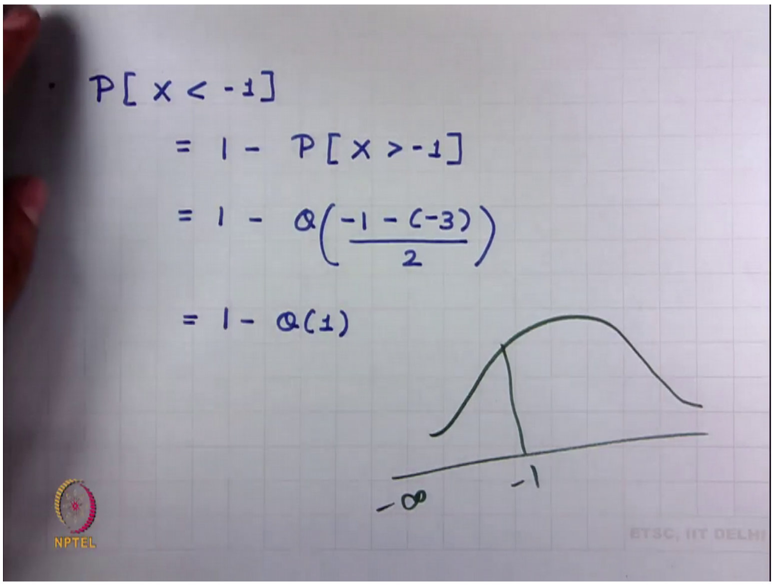
So, the limit should go from z to infinity and in general this random variable can have a mean of m and a variance of sigma square ok. This is our general case, this is a specific case; now let us substitute x minus m by sigma as y. So, d x becomes sigma times d y so, this is the substitution that we did. So, x minus m square by sigma square is simply y square d x is sigma d y and x goes from z to infinity. So, y will go from z minus m by sigma to infinity. So, by just doing the change in variables I can easily go from this integration to this integration. And what is this? This is nothing, but Q of z minus m by sigma. So, in general case if you have to think about a random variable X being greater than Z.

(Refer Slide Time: 17:46)



And, if this random variable has gotten some mean variance you can easily compute this in terms of Q function and this would be Q of Z minus m by sigma. So, if you have to calculate for example, the special case that we were talking about and wondering about before, if you have to think about what is the probability of X greater than 5. So, Z is 5, mean of X is minus 3, standard deviation of X is square root of variance, variance is 4. So, from this we get this is nothing, but Q of 4. So, this is Q of 8 by 2 which is Q of 4 ok, sigma is a standard deviation and sigma square is the variance this we know alright. So, we now have learnt how to compute these integrations in terms of Q function. Let us do one more example. Suppose we have to find what is the probability that X is less than minus 1.

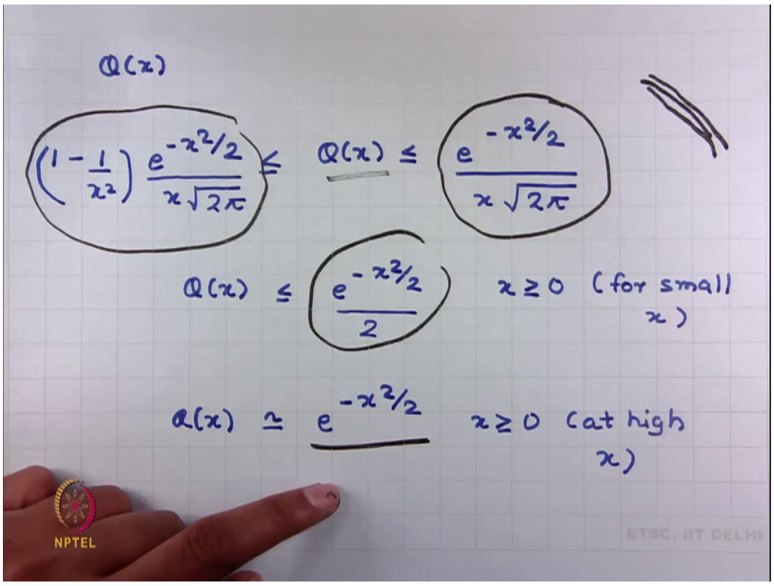
(Refer Slide Time: 18:49)



So, we are wondering about X lying between minus infinity to minus 1 and this thing will be 1 minus probability of X greater than minus 1. Because, the total area is 1, if you have to think about this part you can think this in terms of this part. So, this part is 1 minus this part and that is what we are doing in here; probability of X less than minus 1 is 1 minus probability of X greater than minus 1 and we know how to solve this.

So, minus 1 minus the mean, mean is minus 3 divided by the standard deviation which is 2. So, we get this probability as 1 minus Q of 1. So, these examples hopefully will make you comfortable with this Q function, most of the error formulas that we will obtain will be in terms of this Q function. So, it would be a good idea to do some more examples on this Q function by yourself.

(Refer Slide Time: 20:01)

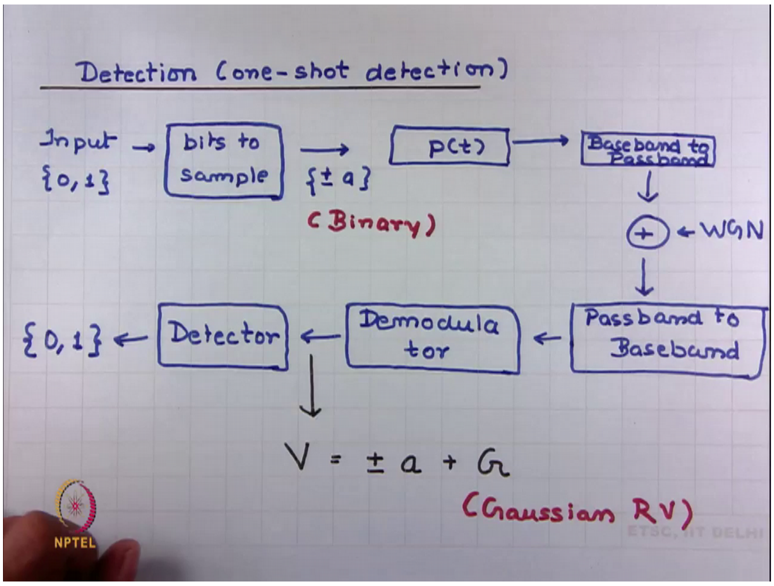


Also it would be a good idea to think about some approximations of this Q function. For example, Q of x always lies between this function and this function, you do not have to remember this by heart. What you should know is that some approximations of Q of x exist. So, if you ever want to approximate Q of x in terms of these exponential functions which are easier to deal with, you can look at them and can find out. So, Q of x always lies between these two functions and this is really very tight bound.

For example, Q of x is very closely bound between these two functions. So, very tight bound, for large x this actually no difference between Q of x and these functions. You can also use this bound if x is small, you can think about the Q of x as being less than this thing or if your x is pretty large you can also approximate Q of x as this function right. So, there are several approximations of this Q of x that exist in literature and you can use one of them appropriately as per your convenience. It depends upon which regions you are interested in whether you are trying to find Q of x, if x is small or whether you are trying to find Q of x if x is large ok.

So, we have revised the Gaussian basics. So, this is what we would need, you need to clearly understand what is this white Gaussian noise and how this noise interacts with this filters and correlators and matched filters. How can you find out the power spectral density of the noise and the output of this filters, how can you find out the noise variance, energy, power at the output of these filters. So, this is what you must know. Let us come back to this problem of detection.

(Refer Slide Time: 22:17)



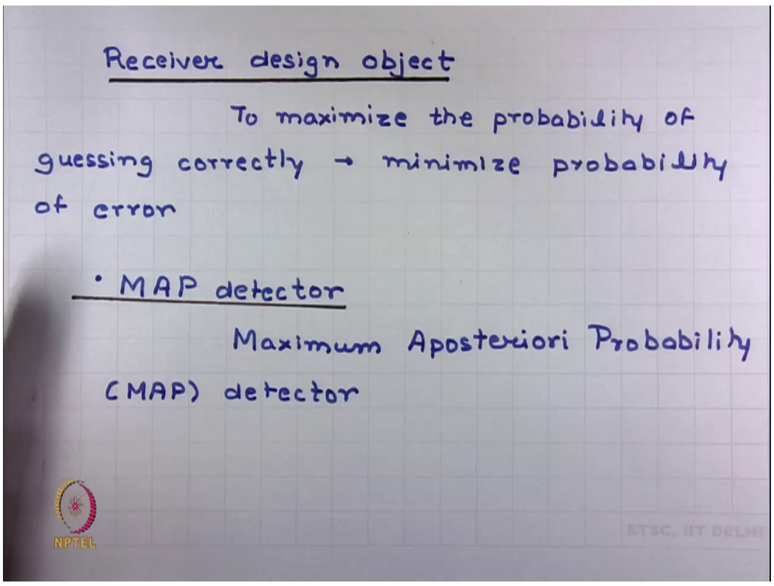
And this detection theory we would start by looking into very simple case and this case we call as one shot detection and then we will generalize this further on. One shot detection means that you have transmitted a symbol, the detector should detect that transmitted symbol. So, what we are saying that instead of transmitting sequence of symbols which we usually do, we will look at this problem for a specific case when we just transmit one symbol. We will see how to do detection in that case, that is the case with which we will start on and then we will generalize this case for the case when we are transmitting a sequence of symbols instead of just one symbol.

So, the detection theory that we would deal or start dealing with is a one shot detection theory. Just consider the detection and reception of one symbol faithfully rather than thinking about detecting a sequence of symbols ok. And, detection of sequence of symbols we will see later on is no different from detection of one symbol. But let us not complicate this issue from starting, in a starting we like to make things as easy as possible. Also when we will start this detection theory, we will forget about all these blocks for a starting. You would not worry about this demodulator, this pass band to baseband conversion and so on so forth. We will just assume to start with that output of this demodulator we have some random variable G which is a Gaussian random variable because, of this white Gaussian noise plus the symbol that we have transmitted.

Here in this case we are showing that we are transmitting binary symbol; that means, we are transmitting either plus or minus a. So, this we have already seen in polar signaling mechanism. So, if you are transmitting plus a or minus a; it is a binary symbol because, there are only two symbols involved. And, for this picture what we would say is we would get G plus or minus a, where is the symbol that we are transmitting and G is a Gaussian random variable coming out because, of this Gaussian noise. So, this is what we will start with, later on we will look into the issues like what happens when this pass band noise is converted to baseband noise. And what happens when this interacts with demodulator and things like that, but to start with the problem statement that we have is pretty simple problem statement.

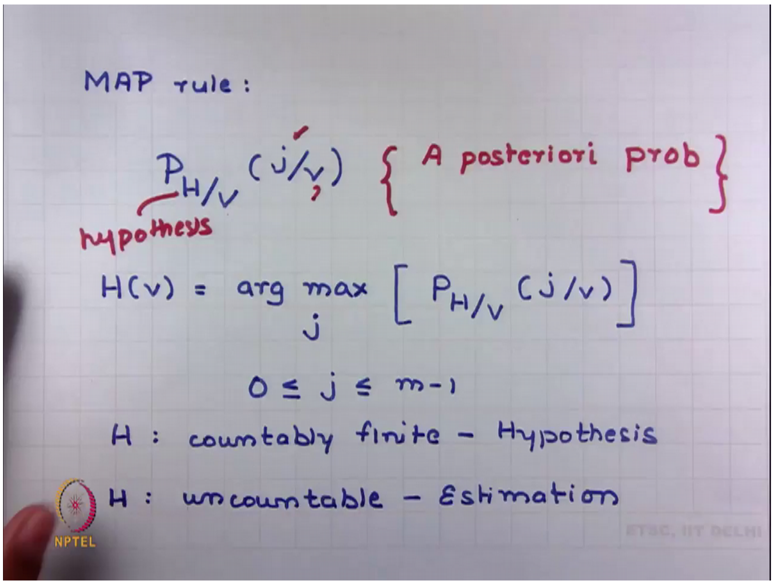
So, the problem statement that we have is we are having a Gaussian random variable V and this Gaussian random variable is shifted because, of the transmitted symbol ok. So, if you have a symbol plus a transmitted, this is a Gaussian random variable with the mean at plus a. And, if you are transmitting minus a, this is a Gaussian random variable with a mean at minus a, but V is also a Gaussian random variable; if G is a Gaussian random variable. So, I hope I have summarized the assumptions that we will take to start with this theory of detection. What is the receiver design objective?

(Refer Slide Time: 26:08)



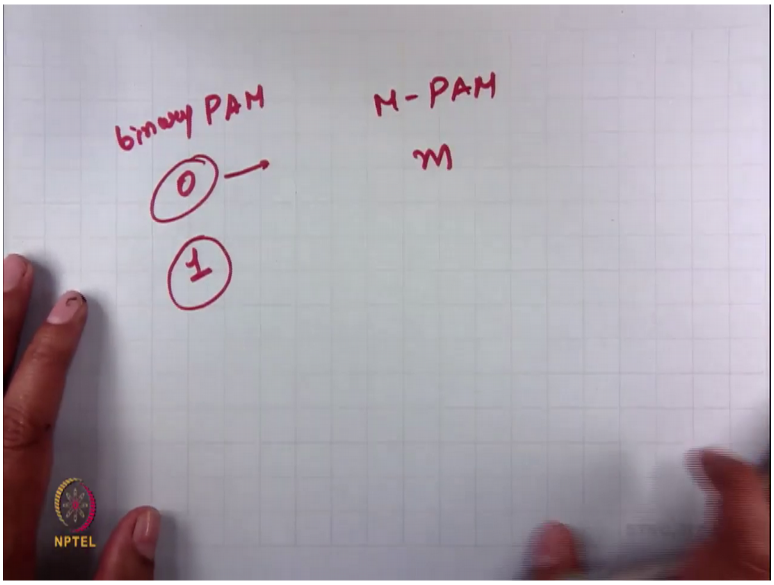
The receiver design objective is to maximize the probability of guessing correctly. So, based on our observation you have to maximize the probability with which you guess correctly and this will minimize the probability of errors. And, the first detector that we will like to look into is a MAP detector. MAP stands for Maximum Aposteriori Probability detector, these words will become clear in a while. But this is basically the detector with which we want to start. So, MAP detector uses MAP rule and what is the MAP rule we will come to that in a while.

(Refer Slide Time: 26:52)



Let us first look at this a posteriori probability what is that. So, a posteriori probability deals with the probability of a hypothesis. So, H corresponds to a hypothesis, j given that random variable V has taken a numerical value small v. What is hypothesis? Hypothesis means the symbols that you are transmitting. For example, if we have M-PAM system, there are m possible symbols and each possibility is a hypothesis ok.

(Refer Slide Time: 27:47)



Let us write down this. So, for example, let us take the case of a binary PAM. So, either you can transmit 0 or you can transmit 1. So, this is one hypothesis, this is the second hypothesis. The two hypothesis in a binary PAM, in M-PAM similarly there are m possible hypothesis; that means, transmitter might have transmitted one of those m symbols. Transmission of each symbol corresponds to one hypothesis. So, what we are finding out in here is what is the probability of a hypothesis j given that this random variable V has taken in a numerical value small v.

That means, we are trying to find out this probability of a hypothesis based on an observation; that means, you have conducted out an experiment. You have seen the result of an experiment, the result of this experiment is the numerical value v. And, based on that numerical value v you are trying to predict the probability of a hypothesis j. And that is why this is known as a posteriori probability; that means, the probability after an experiment has finished, after you have made an observation.

So, once you have carried out this experiment of transmitting a symbol and receiving the numerical value of a random variable, based on that observation which is the numerical value of that random variable we are trying to find the probability of a hypothesis j and that is why this is a posteriori probability, probability based on an observation, probability based on an outcome of an experiment ok. And, MAP rule says that I want to find out the hypothesis j for which this a posteriori probability is maximum that is the

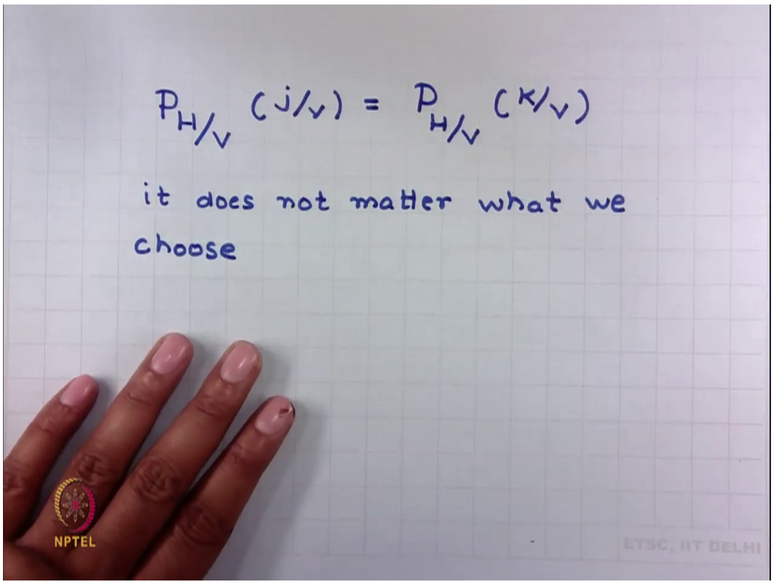
MAP rule. This is a posteriori probability, I want to select the hypothesis as the hypothesis for which this a posteriori probability is maximum. So, arg max simply means that I want to find out the argument for which this a posteriori probability is maximum.

So, the outcome of this would be a hypothesis which maximizes the a posteriori probability, this is the MAP rule. We have already introduced the word hypothesis, but I want to remind you again. So, whenever we are saying the hypothesis; that means, there are a finite number of possibilities, finite number of distinct possibilities if we have to be precise. So, in digital communication we have finite number of distinct possibilities. If we have infinite number of possibilities for example, if H is drawn from an uncountable set, this MAP rule cannot be used because then this a posteriori probability is 0. Because, there are infinite possibilities to choose from and then instead of detection problem we are actually going into estimation problem and that is the main difference between detection and estimation.

In detection we are trying to identify a possibility from finite possibilities, in estimation we are dealing with infinite number of possibilities and this probability of a possibility is 0 and thus, this MAP rule does not make sense there. So, in estimation problems we rather talk about the distance and such matrix. So, this MAP rule applies to detection which is same as hypothesis testing; that means, the number of hypotheses is finite and that is also how it is defined. If the number of possibilities is infinite actually it is not hypothesis, it is really an estimation problem.

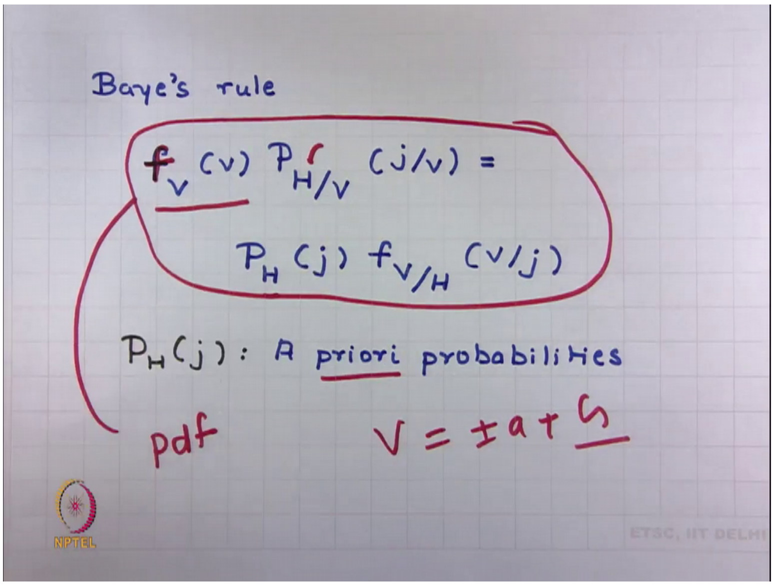
We are not dealing with estimation in this course, we are only confining ourselves to hypothesis testing alright. So, what is MAP rule again it is important, first we are talking about a posteriori probability which is the probability of a hypothesis given an observation has happened. And MAP rule simply says that choose the hypothesis which maximizes this a posteriori probability and that is why the named MAP: Maximum Aposteriori Probability.

(Refer Slide Time: 33:12)



What happens if two arguments give the same posteriori probability? You can choose any one them, it would not impact the probability of error. So, it can also happen that two hypothesis gives the same a posteriori probability and then we can choose any one of them. Let us go to this Baye's rule which is the first thing or probably one of the first thing that we study when we study probability theory.

(Refer Slide Time: 33:35)

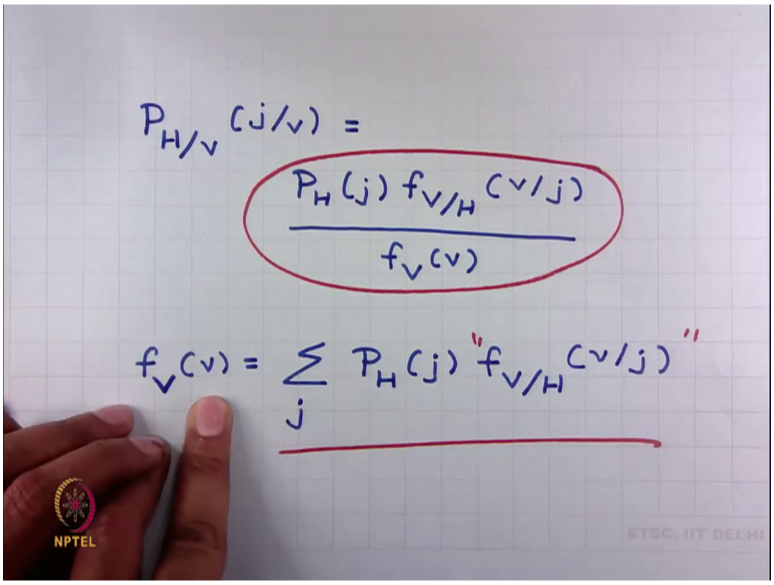


So, Baye's rule says that marginal probability density function so, f in this course denotes the pdf or probability density functions. So, V as a random variable, it is an analogue random variable; that means, V can take any value. It is not confining itself to a discrete set. Why because, what is V? It is plus or minus a plus G, it is a Gaussian random variable and this Gaussian random variable comes because of a Gaussian noise, noise can have any amplitude labels. And, this V is an analogue random variable or it is a continuous random variable as we have said and thus, we have the probability density function in there.

So, probability density function of random variable V times the probability of hypothesis H given V is same thing as probability of a hypothesis H j times the probability density function of V given H. This formula comes from the Baye's rule. If you do not know this it is a good idea to have a look at it ok. We assume that you have done a basic course on probability. What are these P H of j; that means, this is a probability of hypothesis j. These are a priori probabilities, this is the probability with which a transmitter sends in hypothesis and this is before an experiment actually happens and hence the name a priori; before an experiment happens this is the probability of a hypothesis.

So, just in this detection what we have to get used to as this notation, once you are comfortable with this notation detection theory is really straightforward. So, based on Baye's rule we have got a link between a posteriori probability and a priori probability.

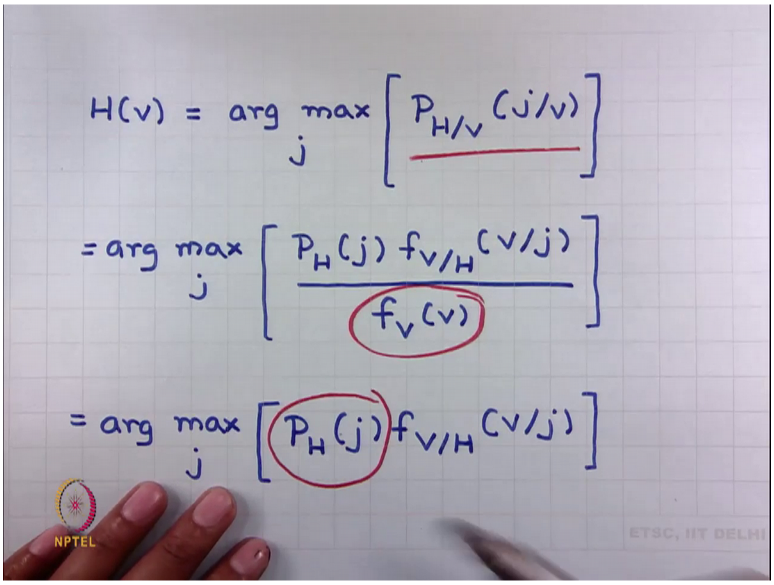
(Refer Slide Time: 36:06)



Let us rearrange this basic equation by bringing down this probability density function of random variable V in here. We have got a posteriori probability of a hypothesis H equals to j can be thought in this way and you can also easily compute the probability density function of this random variable V in this way. So, this is the probability of a hypothesis j and once this hypothesis j has happened; what is the probability density function of random variable V taking a numerical value small v.

So, what you are doing is actually in this part you are trying to find out the probability density function for a random variable V given an hypothesis as happened and then you have average this out for various hypothesis. You multiply this thing with this probability of hypothesis j to happen and then you sum this up for all j's, that will give me probability density function of random variable V.

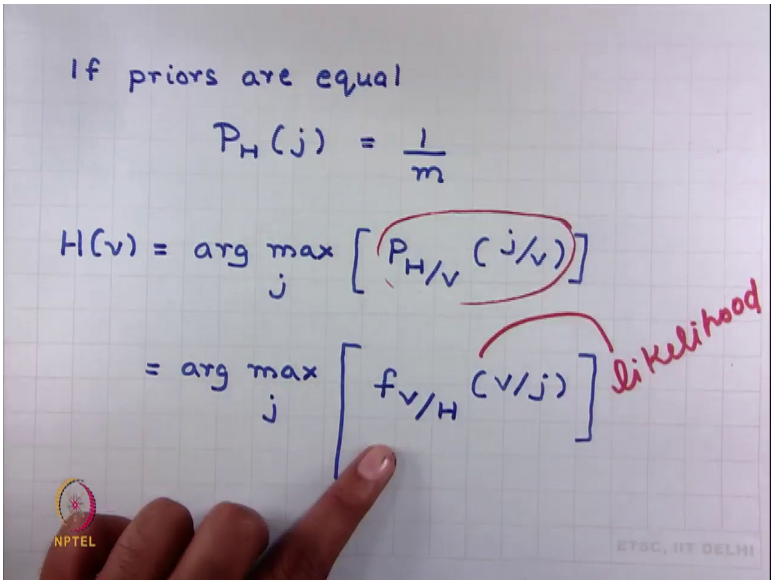
(Refer Slide Time: 37:28)



If I look at the MAP rule, MAP rule says that I am trying to find out the argument j for which this a posterior probability is maximum. We have seen that this a posteriori probability can be arranged in this way. Now, this probability density function of random variable V does not depend upon j, it is independent of j. Hence, if you try to maximize this quantity, this is the same thing as trying to maximize this quantity because this is not a function of j.

And, hence MAP rule can also be understood in this way. Now, let us see something more. If these probabilities are equal for various hypothesis that means; all hypothesis are equally likely before an experiment has happened; that means, a priori probabilities of all hypothesis is same or we say that if priors are equal.

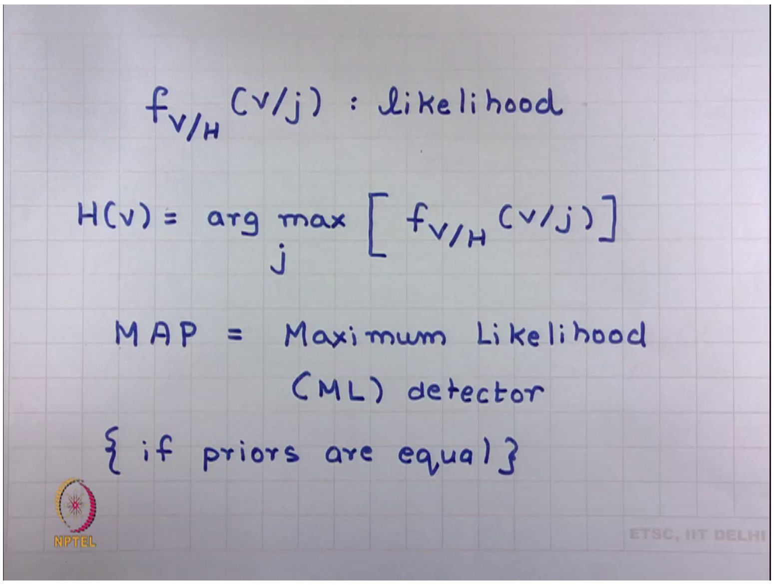
(Refer Slide Time: 38:46)



If priors are equal means a priori probabilities is equal then probability of hypothesis H being j is 1 by m where, m is the number of hypothesis. And, this a priori probability also does not remain a function of j, it is also becoming independent of j. And, hence if you are trying to maximize this quantity and if this thing is also not a function of j; this quantity would be maximized if you simply maximize this; and what is this? This is known as likelihood.

So, what we are saying is if priors are equal instead of trying to maximize the a posteriori probability which is this and we have seen that from here we can go to here, get rid of this. Because, not a function of j, get rid of this because, this is not a function of j; in case priors are equal, maximizing this is same as maximizing this and this is the likelihood function. What is a likelihood function? Is the conditional probability density of random variable V given an hypothesis. So, it is a conditional pdf ok, this is also known as likelihood.

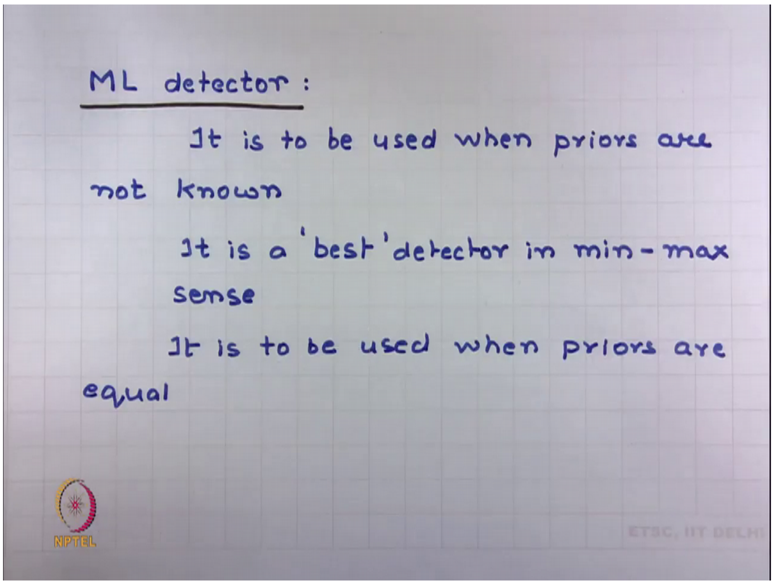
(Refer Slide Time: 40:28)



So, then the MAP rule becomes the rule which chooses an argument j which maximizes the likelihood and thus the MAP detector becomes a maximum likelihood detector or ML detector. So, maximum likelihood detector because it is maximizing the likelihood function.

So, MAP is same as ML detector that is the second kind of detector that we deal with and luckily we only confine ourselves to these two kinds of detectors. So, either we have MAP detector which is a generalized detector, in the case when the priors are not equal and once the priors are equal the MAP detector is same thing as maximum likelihood detector. Because, maximizing a posteriori probabilities is same thing as maximizing likelihood. So, ML detector is to be used when these priors are equal or it can also be used when these priors are not known.

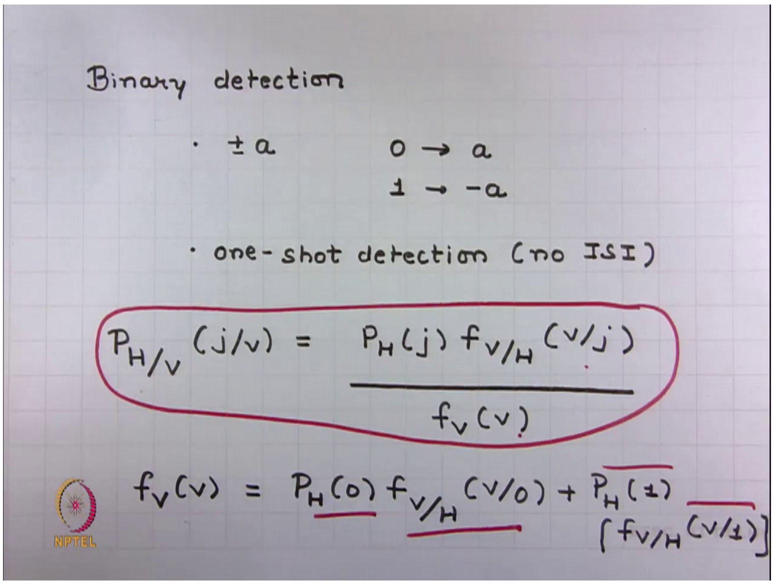
(Refer Slide Time: 41:34)



Sometimes you do not have the probabilistic model of the source. What do you do then? You choose an ML detector because, you do not know the priors and it is also a best detector in min max sense. For example, if you have for some reason chosen the wrong source model or for some reason if you have chosen the wrong probabilistic model of the transmitter.

Or, for some reason if you have chosen the wrong priors then ML detector will minimize your chances of getting hit with a poor performance. ML detector is best in these three cases either the priors are not known or the priors are equal or you want to be safe; you are not very sure about the probabilistic model of the transmitter. So, once we have understood detector let us try to extend this theory for binary detection case.

(Refer Slide Time: 42:39)

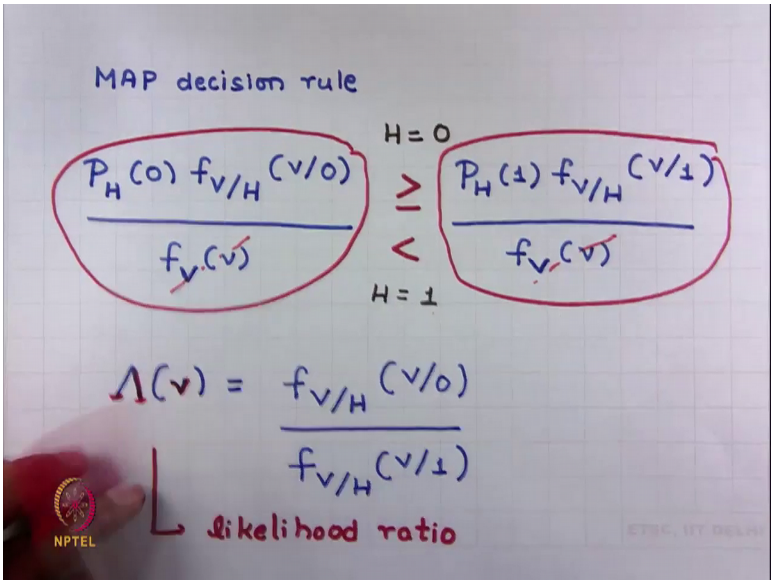


So now, we are confining ourselves to one shot detection problem and we are also confining ourselves to binary detection; that means, either we are transmitting a symbol a. We transmit it when you have a bit 0 coming in or you transmit a symbol minus a you have this when you have a bit 1 coming in and we assume one shot detection. That means, we only transmit one symbol and then we bring down our communication system. We do not worry about what happens when you have another symbol in there and this one shot detection approach helps us in not worrying about inter symbol interference at all because, there is this one symbol.

So, we are separating the issues of inter symbol interference from the issues of detection. When we studied these issues of inter symbol interference in lecture 25, there we separated it from the issues of noise. We said that we consider the modulator when signal to noise ratio is pretty large so that noise does not bother us and here we are taking the approach that we assume that there is no inter symbol interference by just confining ourselves to the transmission of one symbol at a time. And, then we will try to mix these two things together and we will see that there is no different, the theory remains same, performance remain same alright.

So, this is the MAP rule, we have already seen that MAP works on the posterior probabilities and this posterior probabilities can be thought in terms of priors, likelihoods divided by the probability density function of random variable V. And, this probability density function of random variable V for the binary case because, there are only two hypotheses can be thought like this. So, you have a likelihood given that 0 has been transmitted multiplied by the probability that 0 has been transmitted plus the likelihood given 1 has been transmitted multiplied with the probability that 1 has been transmitted and this will give you the pdf of random variable V. So, what is the MAP rule saying? MAP rule says that choose j for which this quantity is maximum and when is this quantity maximum, when this quantity is maximum.

(Refer Slide Time: 45:24)



$$\frac{P_H(0) f_{V/H}(v/0)}{\cancel{f_V(v)}} \underset{H=1}{\overset{H=0}{\gtrless}} \frac{P_H(1) f_{V/H}(v/1)}{\cancel{f_V(v)}}$$

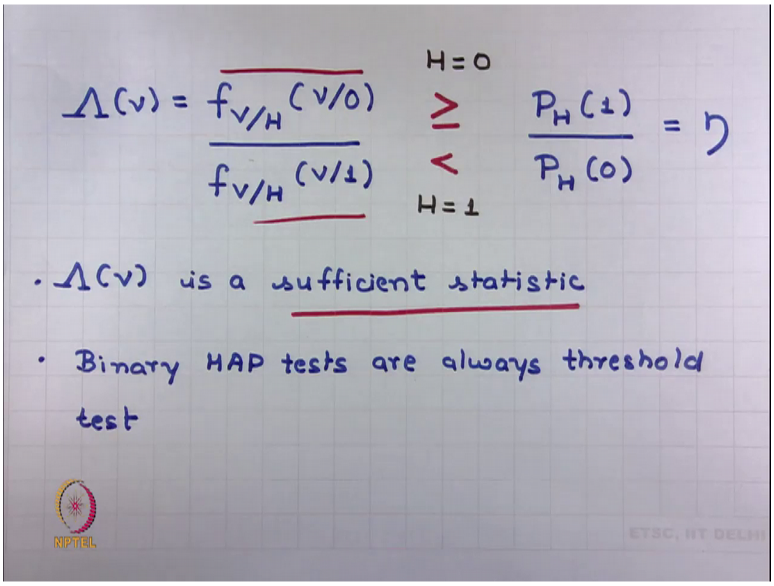
$$\Lambda(v) = \frac{f_{V/H}(v/0)}{f_{V/H}(v/1)}$$

So, let us say if we choose hypothesis to be 0, this is the posterior probability of hypothesis to be 0. This is the posterior probability for hypothesis to be 1 and this is notation and you have to become comfortable with this. This says that if this posterior probability is greater than this posterior probability then choose hypothesis to be 0; because hypothesis 0 maximizes the posterior probability because this posterior probability is larger and when you compare this posterior probability with this posterior probability and you find out that this posterior probability is smaller than this posterior probability, then you choose hypothesis to be 1 and that is the simple MAP rule.

So, you compare the posterior probability of 0 with posterior probability of 1, a posterior probability of 0 is less than posterior probability of 1; choose the hypothesis to be 1. If posterior probability of 0 is greater than posterior probability of 1, choose the hypothesis to be 0. And, we can simplify these things a little bit more because this and this does not depend upon the hypothesis we have said this before so, this is really redundant. So, based on this we can define a very popular ratio known as likelihood ratio. And, the notation we use for likelihood ratios this is likelihood ratio given that you have observed a numerical value of random variable V as a small v.

And this is the likelihood for hypothesis 0 divided by likelihood of hypothesis 1. You can also define the likelihood ratio in a different way, namely that you could have defined the likelihood ratio as the likelihood of hypothesis 1 divided by likelihood of a hypothesis 0 does not matter; the idea remains same. So, here we are defining the likelihood ratio as the likelihood of a hypothesis 0 divided by likelihood of a hypothesis 1 and then I can simply bring this to this side and priors to this side.

(Refer Slide Time: 48:02)



And, this MAP rule can be converted into a rule like this. So, this simply says that first you find out the likelihood ratio which is this thing by this thing. So, I am just pulling this to this side and P H 1 is here. So, I am just saying if this likelihood ratio is greater than this ratio; then this ratio I also called as eta. Then you choose the hypothesis to be 0 and if this likelihood ratio is smaller than this ratio, then you choose hypothesis to be 1. The idea is same; it is just we are using a different notation that is it. Then we say that this likelihood ratio is a sufficient statistic. Once you have obtained this likelihood ratio, these likelihoods themself are not relevant. So, if a detector has this likelihood ratio, it can simply decide on this hypothesis based on this likelihood ratio. These likelihoods are not relevant, the ratio for a likelihood is a sufficient statistic for a detector to decide on the hypothesis.

And, this does not look like a great saving at this moment because, this is one number, this is another number; you are saying that instead of dealing with two numbers I have; I am dealing with one number. But, this will be great saving when we discuss waveforms. A waveform with a bandwidth b and a time duration t naught is actually to be t naught real numbers. And, there we will see that instead of taking a decision based on these two b t naught real numbers or based on numbers proportional to this p t naught, we can simply take a decision based on one number and that would be a great saving. So, what does the detector do? It calculates this likelihood ratio which is just a number obtained by the ratios of these two likelihoods and then it compares this likelihood ratio with this threshold.

And, decides whether it should choose for hypothesis to be 0 or to be 1 and it is a sufficient statistic. We can also see that this binary MAP tests are always threshold test. So, you have obtained a likelihood ratio and then you compare it with the threshold and threshold is eta here. So, this is basically what is in the theory of this detectors, we have seen the two kinds of detectors: a MAP detector and ML detector. And, we have seen that this MAP detector maximizes posteriori probability whereas, this ML detector maximizes the likelihoods and these two detection approaches the same once we are dealing with equal priors. So, from the next lecture we will make this detection problem more interesting. For example, considering how do we detect real vectors or complex vectors and we will also see some examples of trying to find out this probability of errors for some modulation schemes.

Thank you.