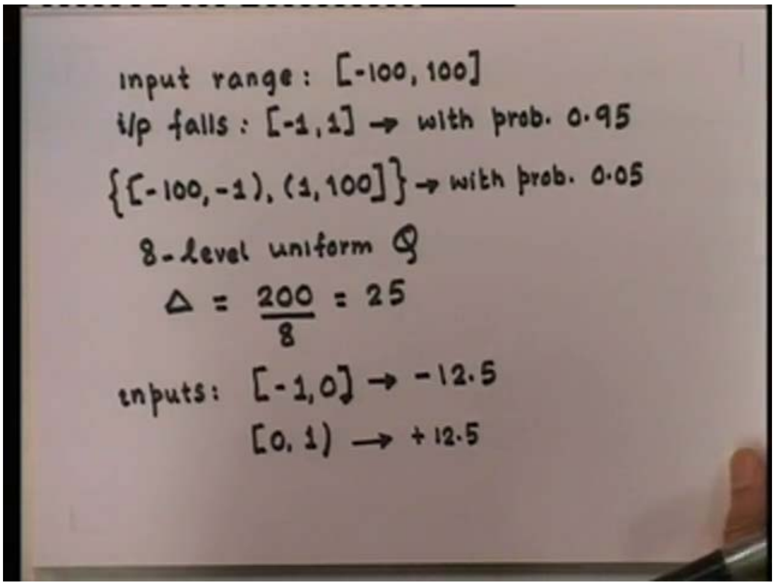
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 36
Lloyd-Max Quantizer

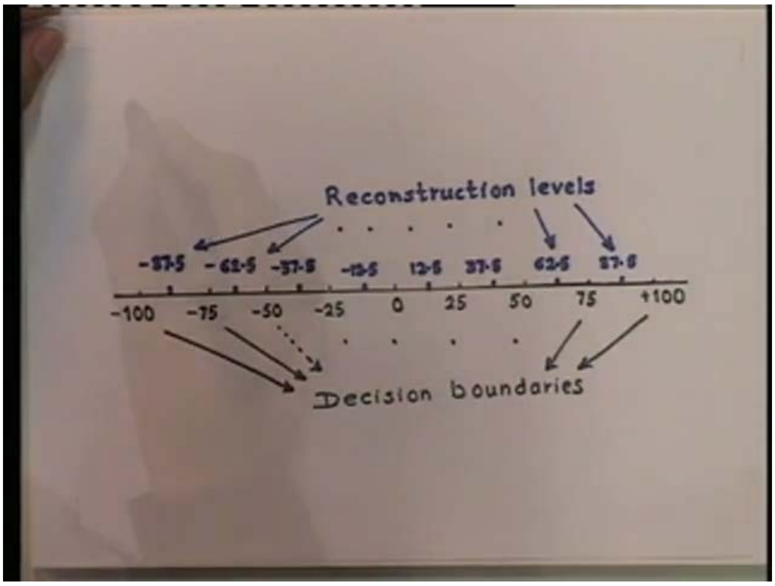
In the previous class, we have learnt about quantization. We have also studied the design process for a uniform quantizer, when the input source has uniform probability density function. However, in practice quite often we come across input sources, which have non uniform probability density function. In such a case, if we still design the simplicity of a uniform quantizer, and obtain the step size based on the earlier procedure. That is dividing the input range by the number of levels of quantizer, then this design is not a very good one. In order to understand this concept clearly, let us take an example.

(Refer Slide Time: 02:20)



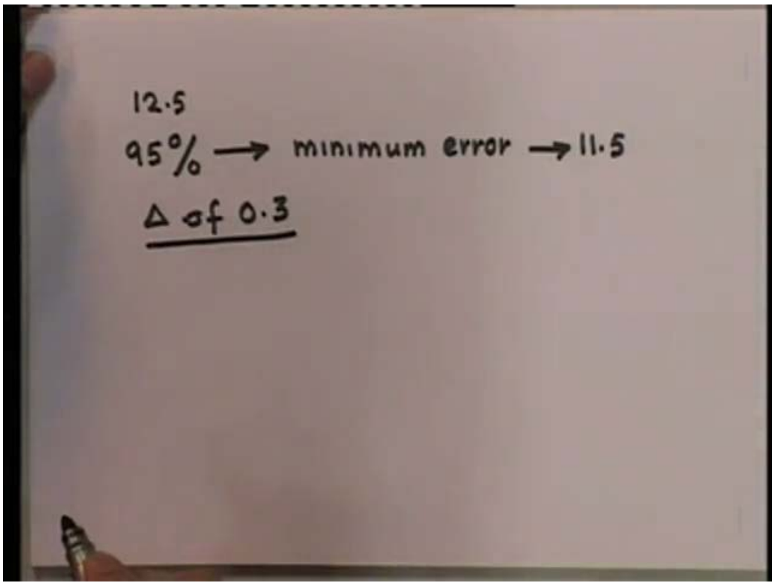
Let us assume that input range is minus 100 to plus 100, and let us also assume that the input falls in the interval minus 1 to plus 1 with probability 0.95 and it falls in the interval minus 100 minus 1 1 100 with probability 0.05. Now, let us design 8 level uniform quantizer. Now, if we follow the procedure which we discussed in the earlier class, and obtain the step size that is delta as the input range. That is 200 divided by the number of levels of the quantizer that is 8, we get equal to 25. What this implies that the inputs in the interval between minus 1 to 0 will be represented by the value minus 12.5 and the input in the range between 0, and 1 would be represented by plus 12.5.

(Refer Slide Time: 05:01)



So, on the real axis, we can denote the decision boundaries and the reconstruction levels as shown here. Now, the blue ones are the reconstruction levels and the black ones are the decision boundaries. So, the maximum quantization error that can be incurred is.

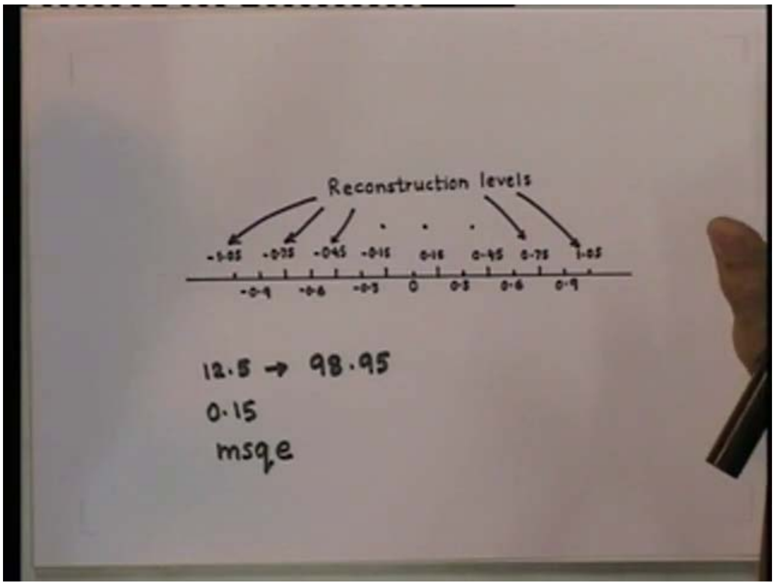
(Refer Slide Time: 05:27)



12.5 However, at least 95 percent of the time the minimum error that will be incurred will be 11.5 because the input lies between minus 1 to plus 1 with the probability of 0.95.

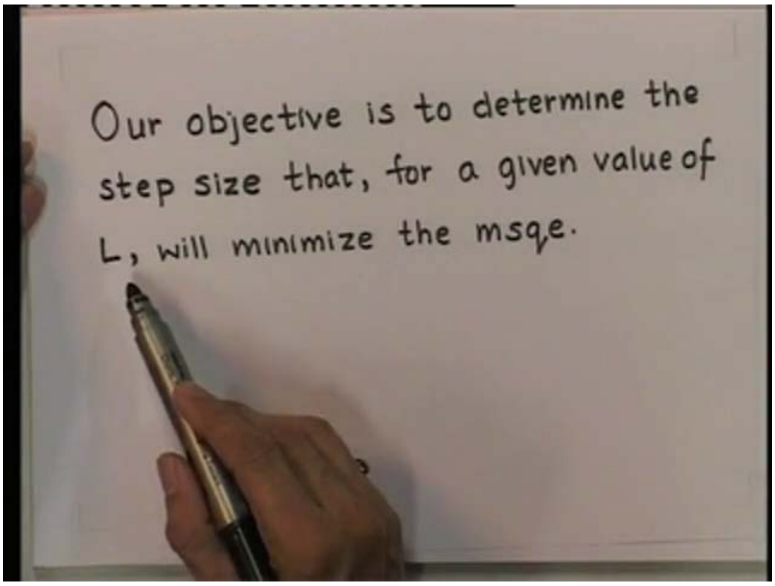
Now, obviously this is not a very good design. So, a much better approach would be to use a smaller step size, which would result in better representation of the values in the range minus 1 to plus 1 interval, given if it is meant a larger maximum error, so based on this idea suppose we pick step size of 0.3. Now, based on this step size the reconstruction levels would be as follows minus 1.05 minus 0.75 up to plus 1.05. Now, and this are the decision boundaries, where the extreme ones are minus 100 and plus 100. Now, in this case the maximum quantization error goes from.

(Refer Slide Time: 07:24)



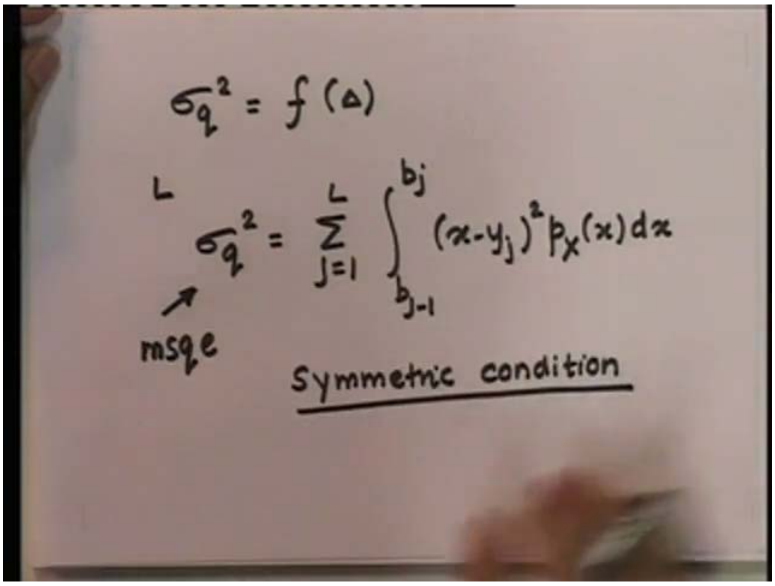
12.5 to 98.95 that is when the input is plus hundred and the reconstruction level is 1.05. However, 95 percent of the time the quantization error will be less than 0.15. Therefore, the average distortion, or the mean square quantization error for this quantization would be substantially less, than the earlier quantizer. So, the conclusion is that when the source PDF is not uniform, it is not good idea to obtain the step size by simply dividing the input range by the required number of quantization levels. This approach becomes impractical when we model our sources with probability density functions that are unbounded such as Gaussian PDF. Therefore, somehow we must include PDF in the design process of our quantizer.

(Refer Slide Time: 09:11)



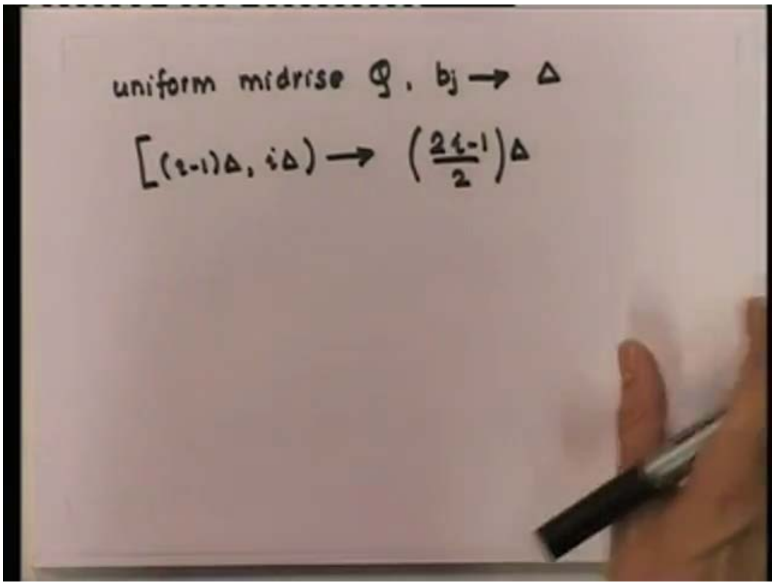
So, our objective is to determine the step size that for a given value of L that is number of levels of quantizer will minimize the mean squared quantization error. Now, the simplest way to achieve this is to write the distortion as a function of the step size.

(Refer Slide Time: 09:42)



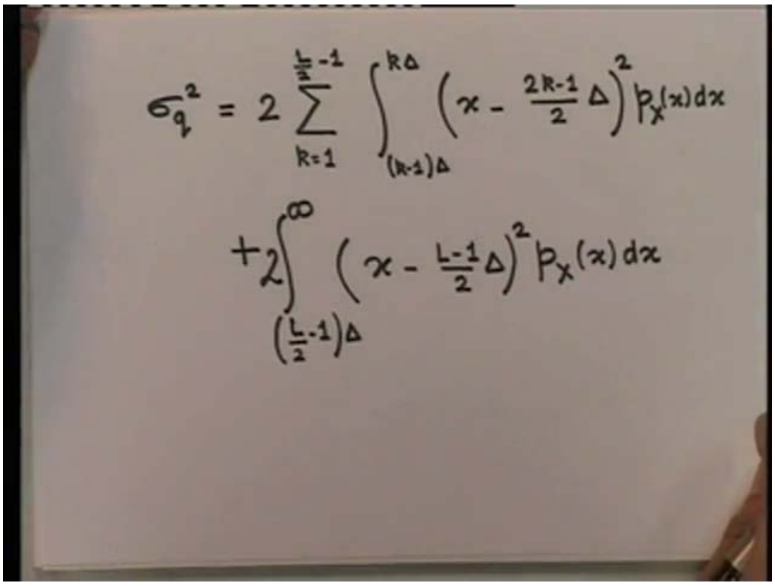
So, we should able to write the distortion as a function of step size and then minimize this function with respect to the step size. Now, an expression for the minimum squared quantization error for L level uniform quantizer, as a function of delta can be found by replacing the boundary, and the reconstruction levels in the following equation. Now, this is the mean squared quantization error as the function of reconstruction level, and decision boundary. So, we should replace the decision boundary that is d j s and the reconstruction levels y j by functions of delta. Now, as we are dealing with a symmetric condition, we need only compute the mean square quantization error for positive values of x. The mean square quantization error for negative values of x will be the same.

(Refer Slide Time: 11:40)



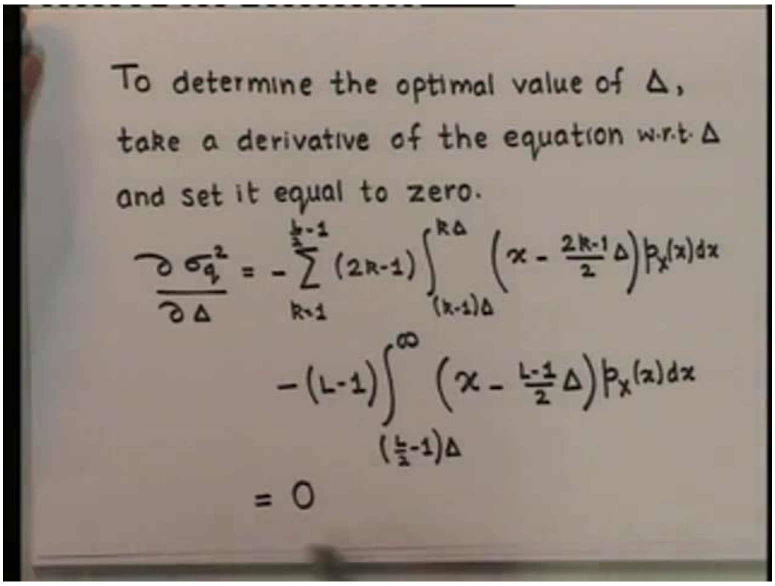
So, for a uniform mid rise quantizer the decision boundaries are integral multiples of the step size delta. And the representation level for the interval i minus 1 delta to i delta is given by 2 i minus 1 upon 2 times delta.

(Refer Slide Time: 12:32)



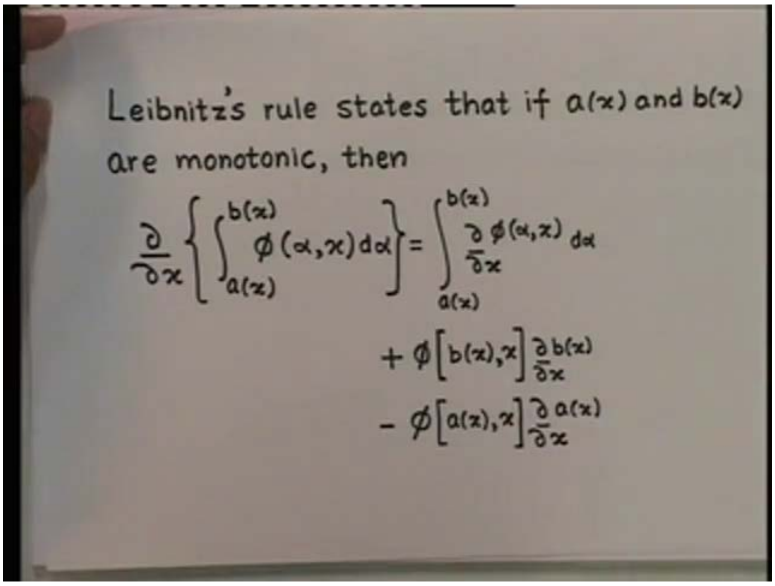
Therefore, based on this the minimum squares quantization error can be rewritten as shown here. Now, this expression is function of delta. Therefore, to determine the optimal value of delta, we can take a derivative of this equation with respect to delta and set it equal to 0. So, if we do that.

(Refer Slide Time: 13:12)



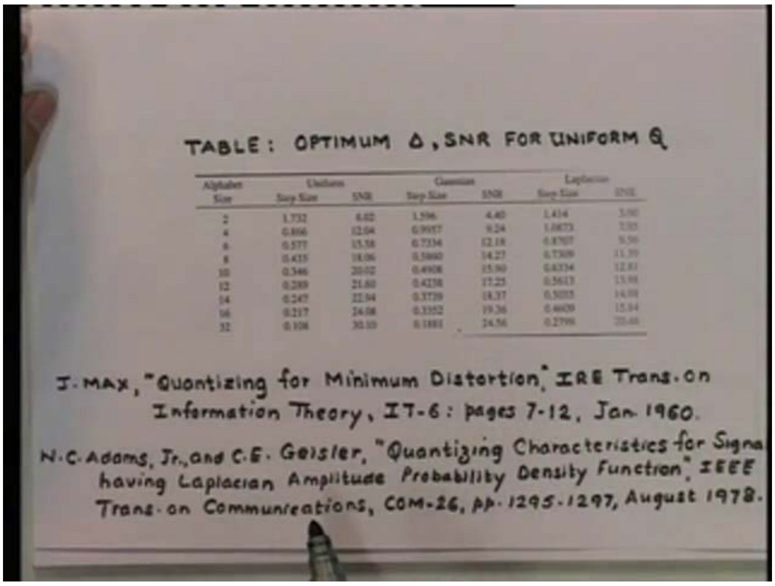
This is the expression, which we get. Now, in order to derive this expression, we have to use the Leibnitz's rule which states.

(Refer Slide Time: 13:36)



That if a x and b x are monotonic then differentiation of this expression is given on the right hand side, using this Leibnitz's rule we get the following expression. Now, this is a rather messy looking expression, but given the probability density function that is PDF it is easy to solve this equation using any one of a number of numerical techniques.

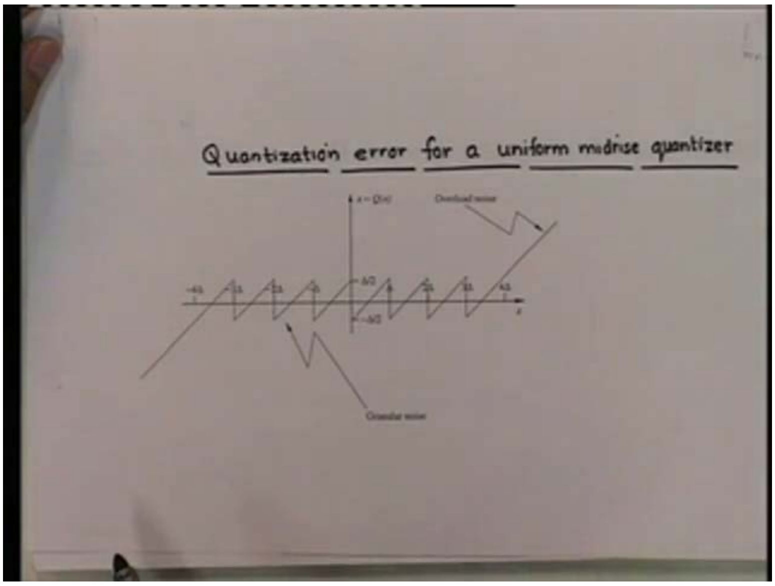
(Refer Slide Time: 14:17)



The following table list step sizes found by solving the previous expression, for nine different alphabet sizes and three different probability density functions. Now, the results given in this table are taken from this two papers, one is by max and the other is by

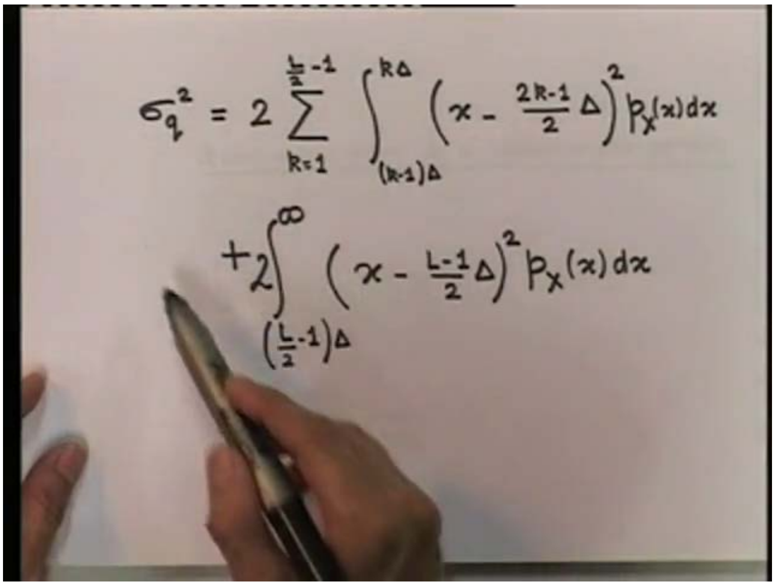
Adam’s and others, which had appeared in IRE transmission on information theory and IEEE transmission on communications. We will return back to the discussion on the results depicted in this table, but before that it is important to know that in practical situations the inputs are always bounded, and it is only a mathematical convenience that we model the non uniform sources with unbounded support. Now, if the input is unbounded this implies that the quantization error is no longer bounded either. So, the quantization error as a function of input is shown in the following figure.

(Refer Slide Time: 15:56)



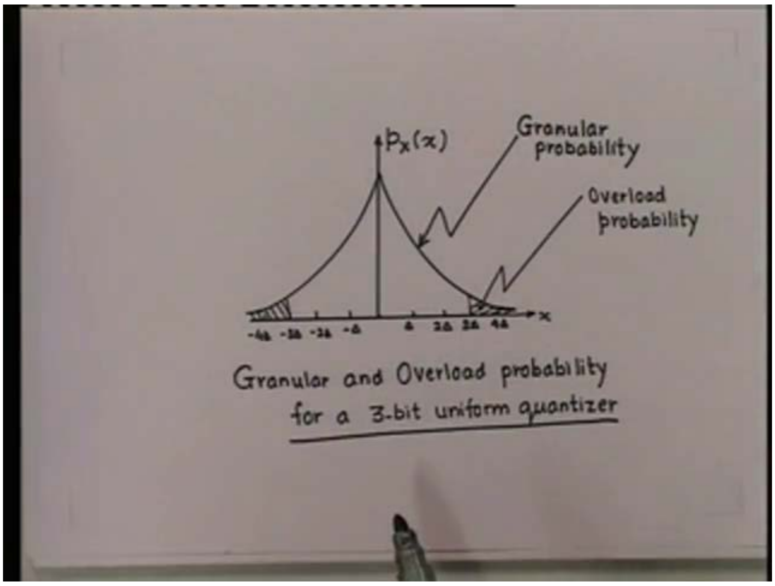
Now, from this figure we see that in the inner intervals, the error still bounded by delta by 2. However, the quantization error in the outer intervals is unbounded. Now, this two types of quantization error are given different names, the bounded error is called granular error or granular noise. While the unbounded error is called over load error, or overload noise.

(Refer Slide Time: 16:47)



In the expression for the minimum squared quantization error, the first term represents the granular noise and the second term represents the overload noise. Now, the probability that the input will fall into the overload region is called the overload probability.

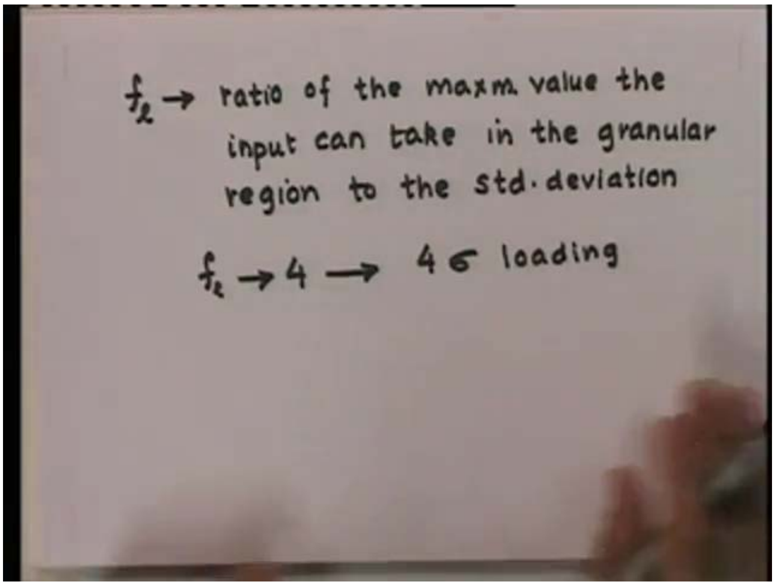
(Refer Slide Time: 17:15)



So, if we assume that the input has probability density function as depicted here, then this region denotes the overload probability and the region here denotes the granular probability. Now, as depicted here the non uniform sources usually, have probability density function that is PDF that have a general peaked at 0, and decay as we move away from the origin. Therefore, the overload probability is much smaller than the probability of the input falling in the granular region. Now, from this equation it is clear that if we increase the step side delta.

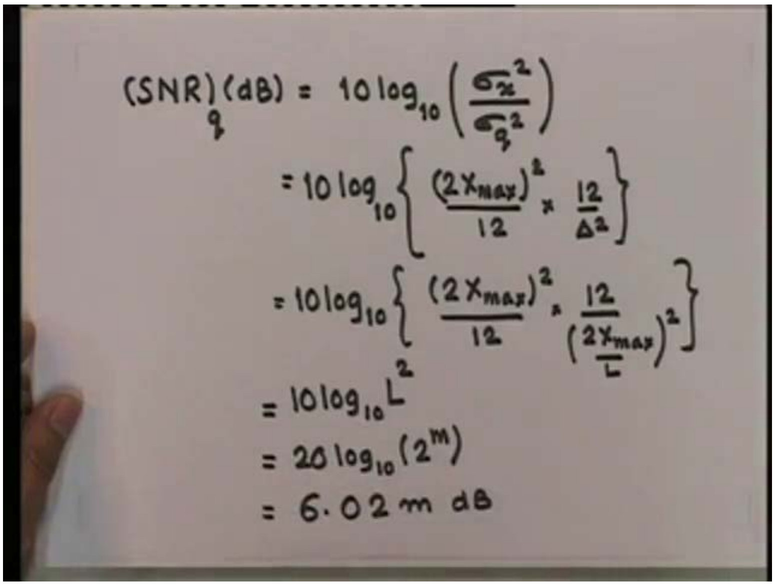
This will result in a increase in the value of l by 2 minus 1 times delta, which in turn will result in decrease in the overall overload probability, therefore the overload noise. However, an increase in delta will also increase the granular noise, which is the first term in this equation. So, the design process for a uniform quantizer is a balancing of these two effects, an important parameter that describes this trade off is the loading factor.

(Refer Slide Time: 19:13)



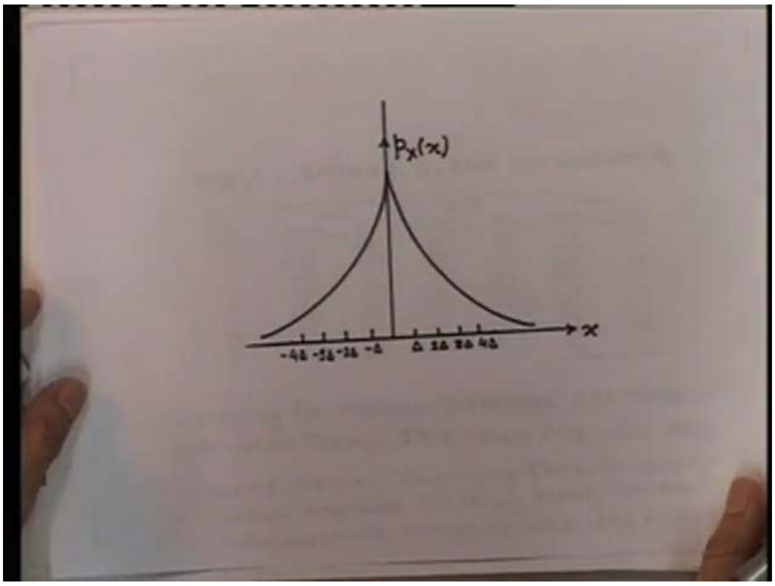
F l which is defined as the ratio of the maximum value, the input can take in the granular region to the standard deviation. A common value of the loading factor is 4 and this is also referred to as 4 sigma loading. Now, we have also studied that when quantizing an input with a uniform PDF the S N R and the bit rate are related by this equation.

(Refer Slide Time: 20:35)



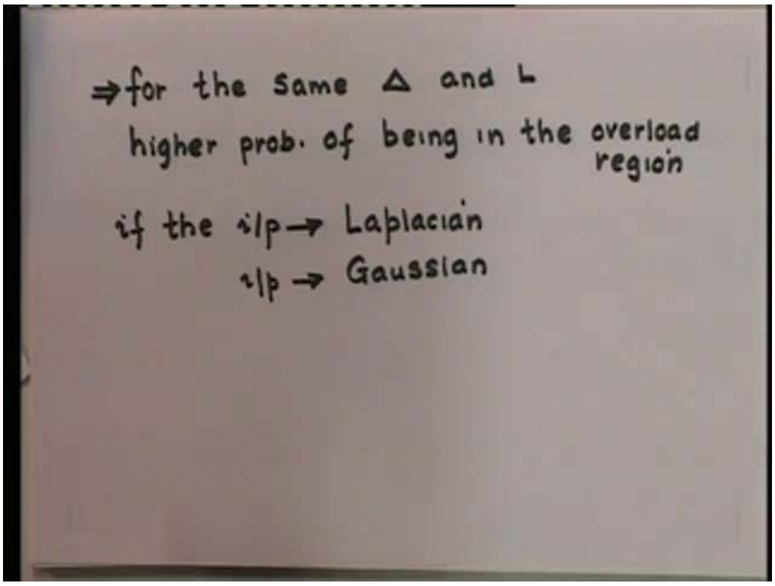
What this equation states is that for each bit increase in the rate, there is an increase of 6.02 d B in the S N R. Now, coming back to the earlier table, this table shows that although the S N R for the uniform PDF follows, the rule of a of 6.02 d B increase in the S N R for each additional bit, this is not true for the other PDF ’s. And it is also evident from this table that the more peaked a PDF is that is the further, away from the uniform PDF it is, the more it seems to deviate from the 6.02 d B rule. We also concluded that the selection of the step size is a balance between the overload and granular noise.

(Refer Slide Time: 21:57)



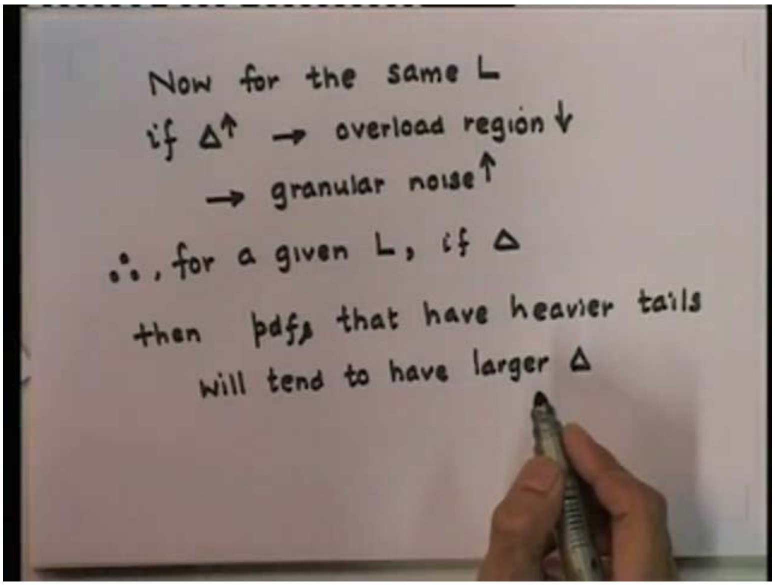
Now, this a PDF for the Laplacian, the Laplacian PDF has more of its probability mass away from the origin in its tail, than the Gaussian PDF .

(Refer Slide Time: 22:16)



So, what this implies is that for the same step size and the number of levels of quantizer that is L, there is higher probability of being in the overload region, if the input has a Laplacian PDF than if the input has a Gaussian PDF.

(Refer Slide Time: 23:26)

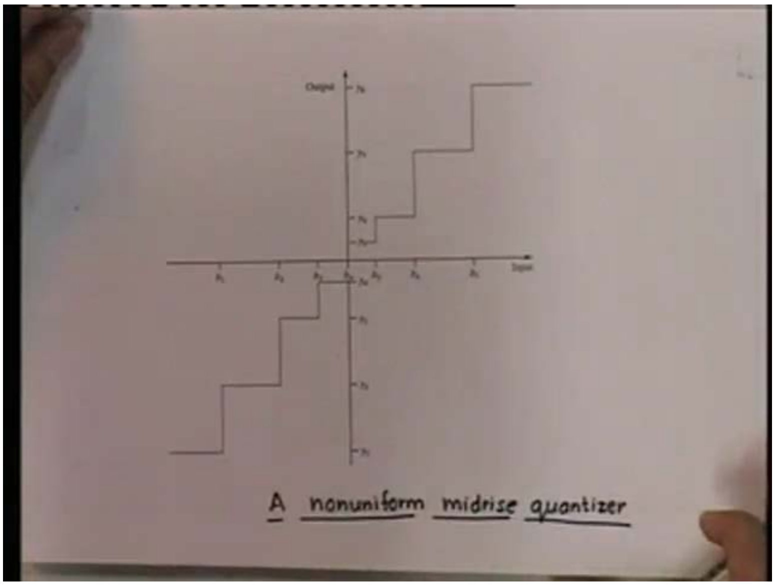


So, this implies that for the same number of quantization levels, if we increase the step size then the size of the overload region is reduced. Hence, the overload probability decreases at the expense of increase in the granular noise. Therefore, what it implies that for a given L, if we calculate the step size to balance the effects of granular band over load noise, then PDF ’s that have heavier tails will tend to have larger step sizes. Based on this argument this table shows this effect. For example, for 16 levels the step size for the uniform PDF is 0.217.

Whereas, for the same alphabet size the step size or the Gaussian PDF is 0.3352, whereas, for the Laplacian PDF it is still larger it is 0.4609. So, if the input distribution has more mass near the origin then the input is most likely to form in the inner levels of the quantizer. Now, if we recall our design procedure for lossless compression scheme, where we had tried to minimize the bit rate or the average number of bits per input symbol, we had followed a special way of assigning a shorter code word to symbols, that occurred with higher probability. And longer code words to the symbols that occurred with lower probability.

So, in a similar manner in order to decrease the average mean square quantization error, we can approximate the input better in regions of higher probability, perhaps at the cost of most approximation in regions of lower probability. Now, in order to achieve this we can apply smaller quantization intervals, in those regions that have more probability mass. So, if the source distribution is like the distribution shown here, we would have smaller intervals near the origin. So, if we wanted to keep the number of intervals constant, this would mean that we have larger intervals away from the origin. So, a quantizer that has non uniform intervals is called a non uniform quantizer.

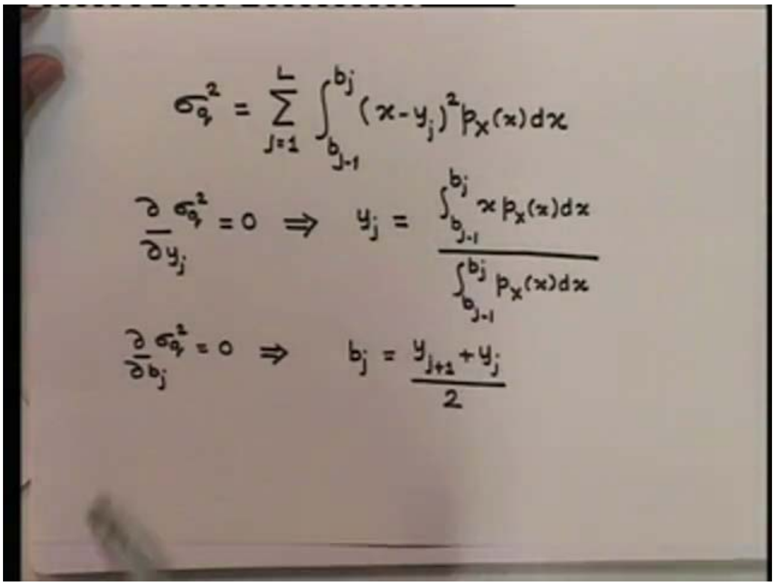
(Refer Slide Time: 27:40)



An example of a non uniform quantizer is shown in this figure. Notice that the intervals closer to 0 are smaller hence, the maximum value that the quantizer error can take is also smaller in this region and that is why there's a better approximation. Now, we pay for this improvement in accuracy at lower input levels by incurring larger errors, when the input falls in the outer intervals.

Now, since the probability of getting smaller input values is much higher than getting larger signal values, on the average the distortion will be lower than if you had a uniform quantizer. So, while a non uniform quantizer provides, lower average distortion the design of a non uniform quantizer is also somewhat more complex. So, the basic idea is quite straight forward, find the decision boundaries and the reconstruction levels that minimize the mean squared quantization error. Now, let us look at a design of such a non uniform quantizer in more detail.

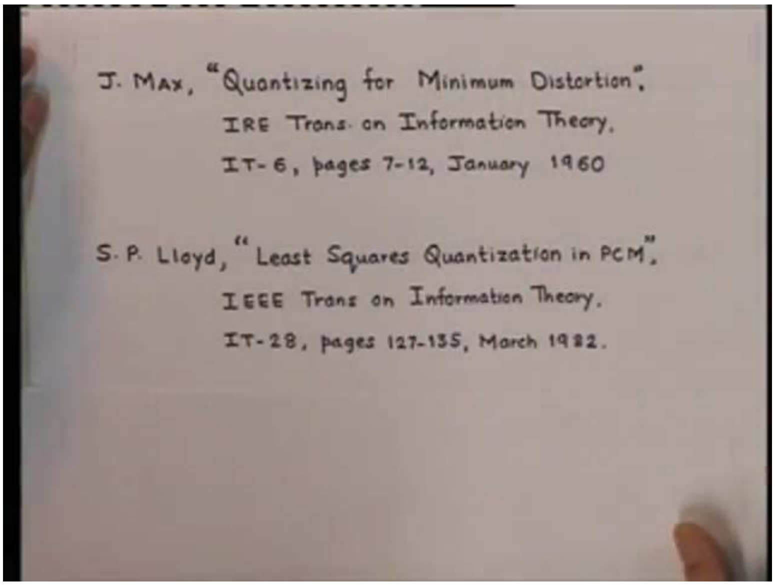
(Refer Slide Time: 29:18)



So, if we have a probability model for the source then a direct approach for designing the best non uniform quantizer is to find the b j's, and y j 's that minimize this equation. Now, to obtain b j's and y j's we can follow the following procedure setting the derivative of this equation with respect to y j to 0, and solving for y j we will get the following expression. What this expression suggests is that the output points for each quantization interval is the centroid of the probability mass in that interval.

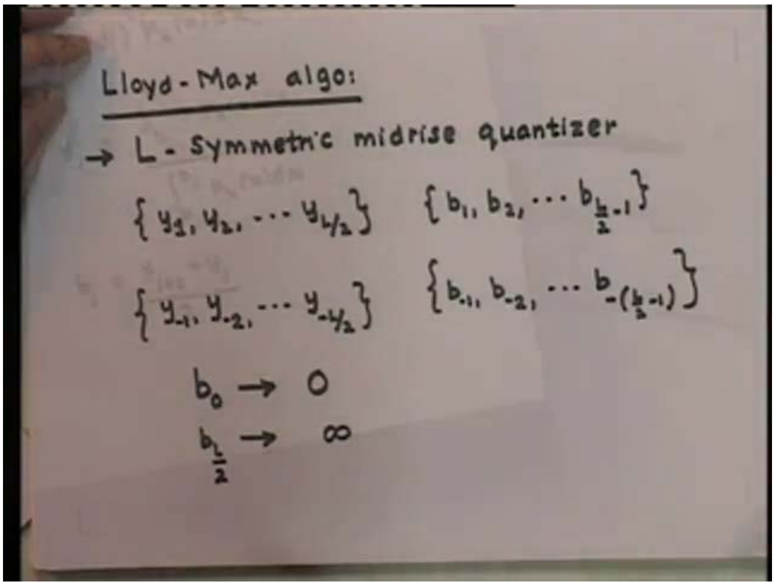
Similarly, taking the derivative of mean squared quantization error with respect to b j, and setting it equal to 0 we get an expression for b j as follows and this expression suggest that the decision boundary is simply the midpoint of the two neighboring reconstruction levels. Solving this equation will give us the values for the reconstruction levels and decision boundaries that minimize the mean squared quantization error. Unfortunately to solve for y j, we need the values of b j and b j minus 1. To solve for b j we need the values of y j plus 1 and y j.

(Refer Slide Time: 31:27)



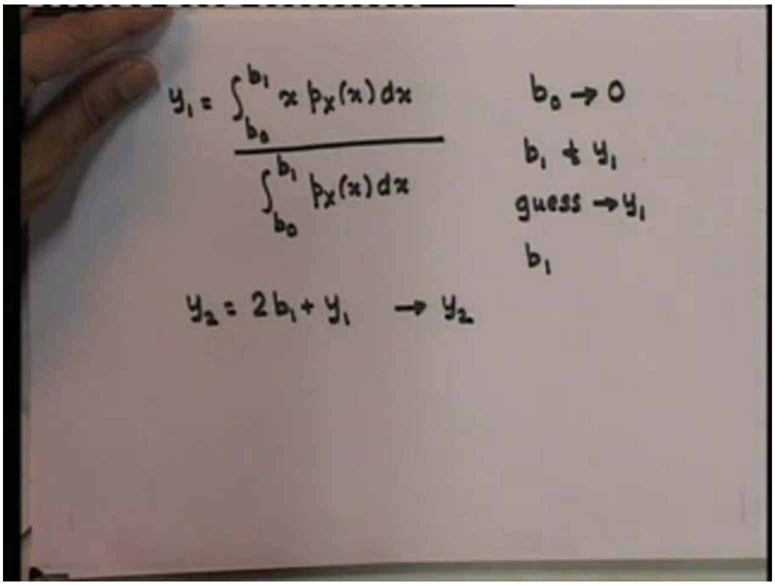
In 1960 Joel max showed how to solve the equations iteratively, in this paper. Now, the same approach was discussed by Stuart P Lloyd in 1957 internal dell labs memorandum, and was published little later in 1982. So, both the approaches are same therefore, the algorithm which is used to solve the two equations iteratively is known as Lloyd max algorithm.

(Refer Slide Time: 32:16)



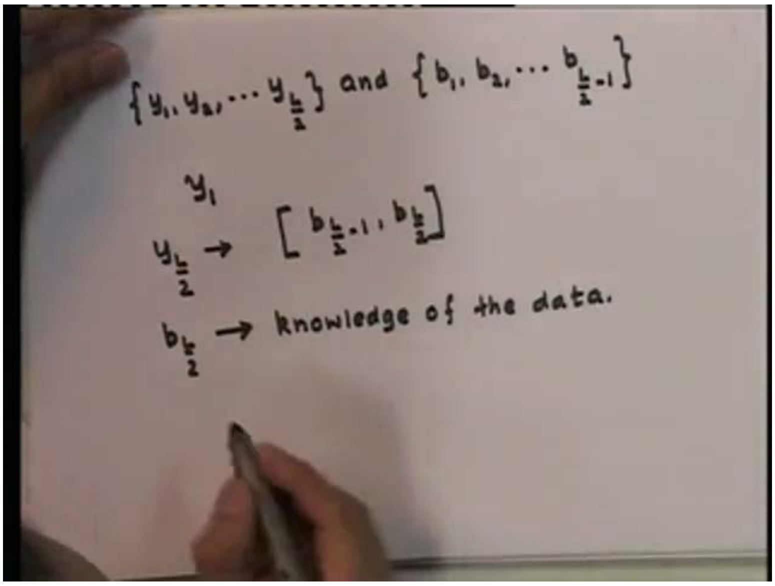
So, Lloyd max algorithm works as follows suppose, you want to design an L level symmetric mid rise quantizer. Now, to define our symbols which will be using for derivation, let us use the following figure, from the figure we observe that the design of this quantizer needs to obtain the reconstruction levels y 1, y 2 up to y L by 2. So, and the decision boundaries b 1, b 2 up to b L by 2 minus 1, the example here is for L equal to 8. The reconstruction levels y minus 1, y minus 2 up to y minus L by 2 and the decision boundaries b minus 1, b minus 2 up to b minus L by 2 minus 1 are obtained via symmetry the decision boundary b 0 is 0. And the decision boundary b L by 2 is simply the largest value the input can take, and for the unbounded input this would be infinite. So, in this equation let us set j equal to 1, if we set j equal to 1.

(Refer Slide Time: 34:58)



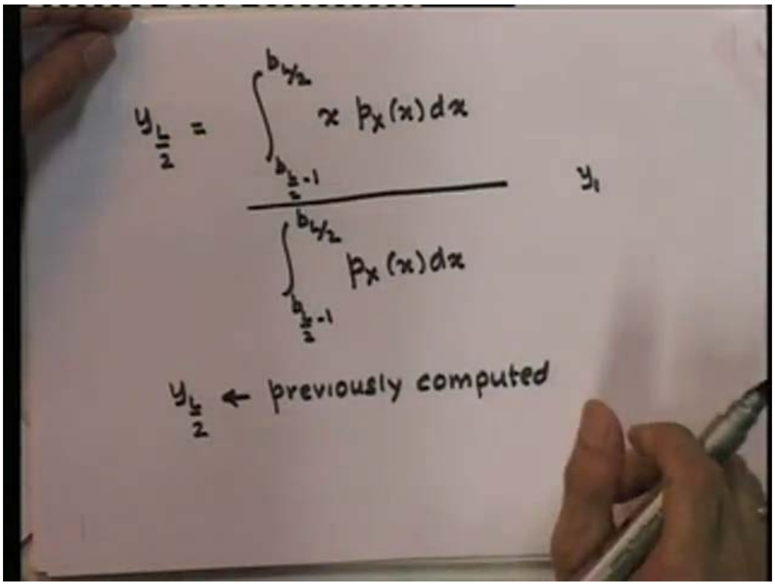
We get y 1 is equal to b b 0 x p x x d x by integral over b 0 to b 1 p x x dx as b 0 is known to be 0. We have two unknowns in this equation b 1 and y 1. Now, we make a guess at y 1 and later, we try to refine this guess using this guess in this equation we numerically find the value of b 1 that satisfies this equation. Now, setting j equal to 1 in this equation, and rearranging the term slightly we get y 2 is equal to 2 b 1 plus y 1. So, from this we can compute y 2. Now, this value of y 2 can be used in this equation with j equal to 2 to find b 2 and which in turn can be used to find y 3.

(Refer Slide Time: 37:18)



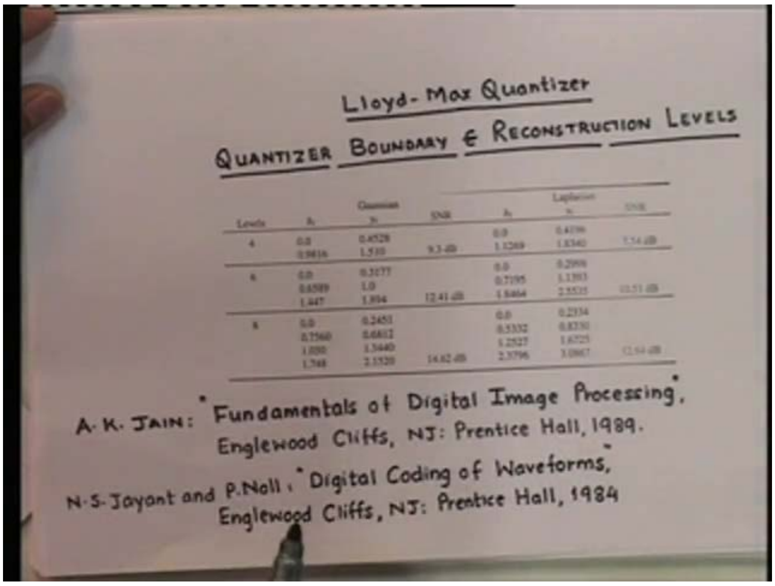
We continue this process until we obtain a value for Y 1, y 2 up to y L by 2 and decision boundaries b 1, b 2 up to b L by 2 minus 1. Now, accuracy of all the values obtained to this point depends on the quality of the initial estimate of y 1. Now, we can check this by noting that y L by 2 is the centroid of the probability class of the interval b L by 2 minus 1 and b L by 2. We know b L by 2 from our knowledge of the data. Therefore, we can compute the integral.

(Refer Slide Time: 38:34)



Y L by 2 is equal to and compare it with the previously computed value of y L by 2. Now, if the difference is less than some tolerance threshold we can stop otherwise, we estimate the value of y 1 in the direction indicated by the sign of the difference and repeat the whole procedure.

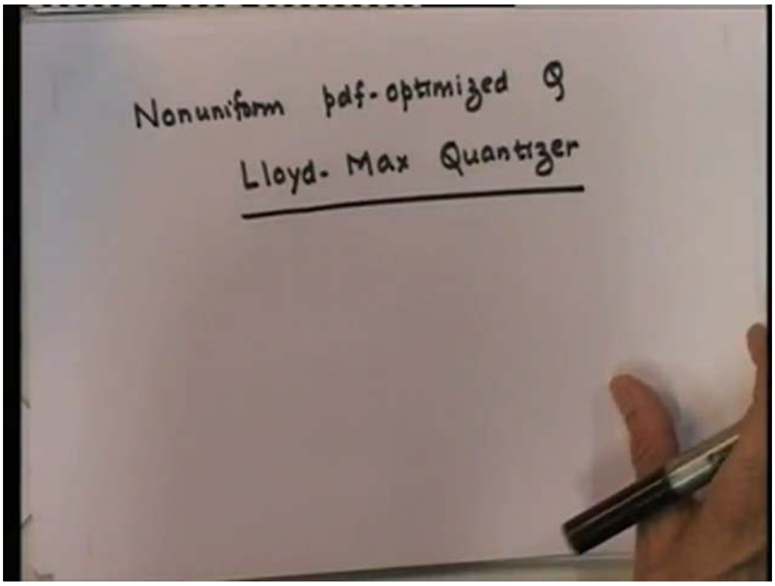
(Refer Slide Time: 39:46)



Now, this table provides decision boundaries and reconstruction levels for various probability density functions, and number of levels generated using the procedure just discussed. Notice that the input distributions that have heavier tails also, have larger outer step sizes. However, the same quantizer have smaller inner step sizes because they are more heavily peaked. The S N R this table shows the decision boundaries and the reconstruction levels, for various probability density functions and number of levels generated using the procedure just discussed.

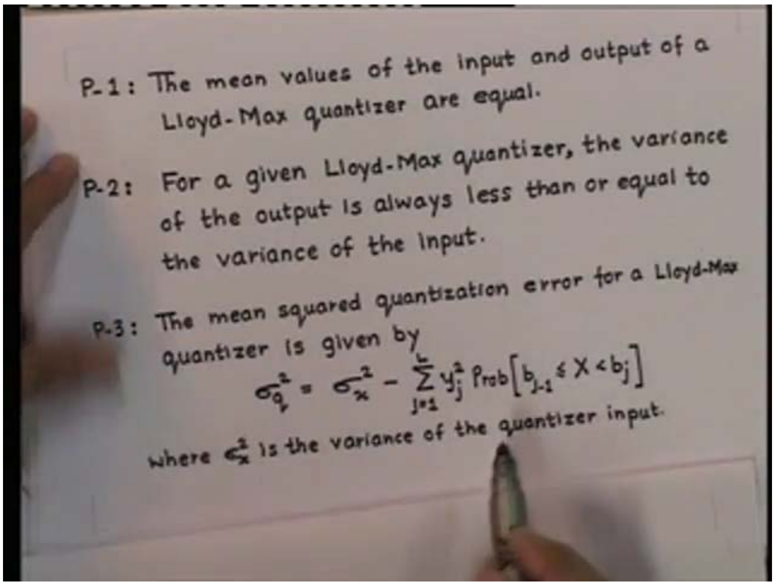
Notice that the distributions that have heavier tails also have larger outer step sizes. However, the same quantizers have smaller step sizes because they are more heavily peaked the S N R for this quantizers is also listed in this table. Now, if we compare this values with those for the PDF optimized uniform quantizer, we can see a significant improvement. Specifically, for distributions further away from the uniform distribution.

(Refer Slide Time: 41:45)



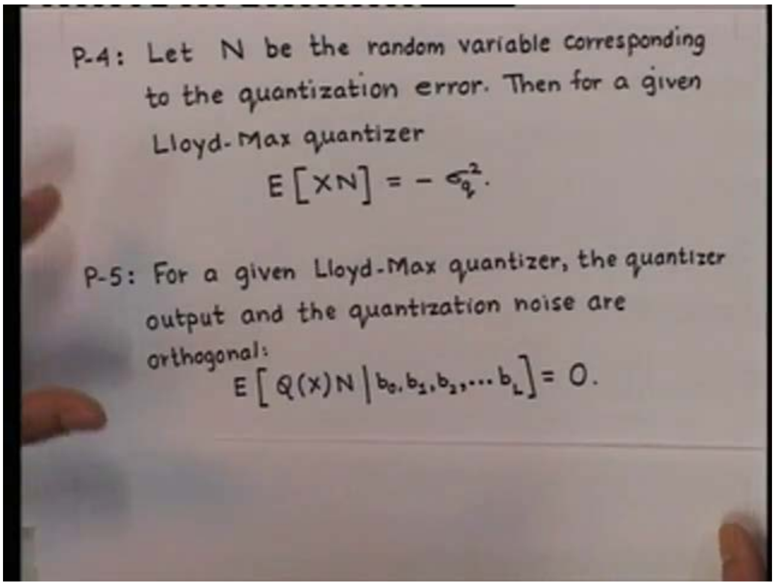
Now, this non uniform PDF optimize quantizer is also known as Lloyd max quantizer. This quantizer has a number of interesting properties, we will not go into proof of these properties, but for the sake of completion we are stated here.

(Refer Slide Time: 42:33)



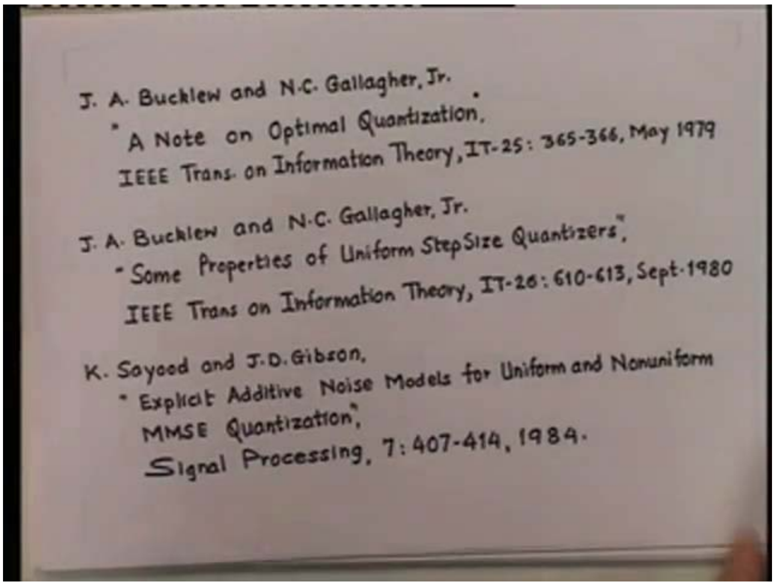
The first property says that the mean values of the input and output of a Lloyd max quantizer are equal. The next property says that for a given Lloyd max quantizer, the variance of the output is always less than or equal to the variance of the input. Third property says that the mean square quantization error, for a Lloyd max quantizer is given by the following expression, where the first term that is sigma x square is the variance of the quantizer input and the second term is the second moment of the output. So, if the input has 0 mean, then this is the variance of the output.

(Refer Slide Time: 43:49)



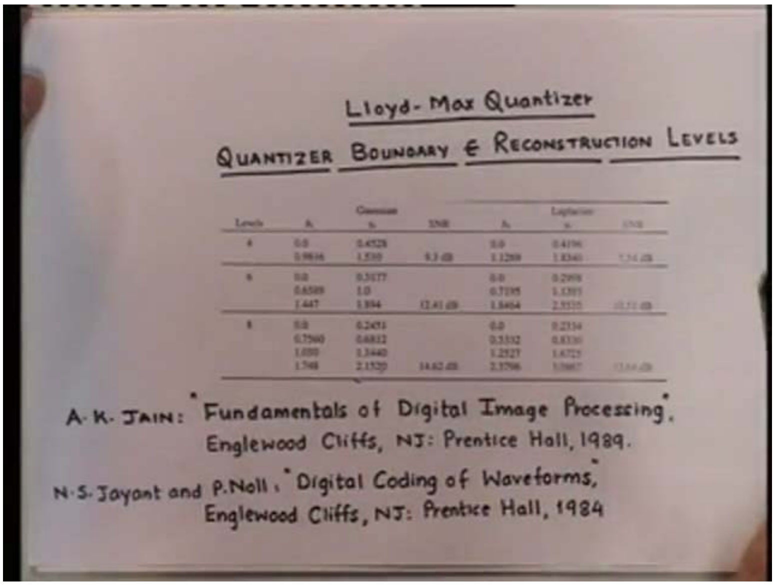
The fourth property states that if N denotes the random variable corresponding to the quantization error. Then for a given Llyod max quantizer expectation of the random variable X and N is equal to minus sigma q square. The finally, for a given Lloyd max quantizer the quantizer output and the quantization noise are orthogonal. What it means that expectation of the quantizer output and the quantizer noise, given the Lloyd max quantizer that is the decision boundaries, this is equal to 0. Now, the proofs for these properties can be found in the following references.

(Refer Slide Time: 44:54)



The first two references had appeared in the I triple E transition on information theory, and last references is in the Eurosif journal of signal processing. It is important to remember that while designing both uniform, and non uniform PDF optimized quantizers. It is essential to know the input source probability density function. The parameters associated with this PDF, for example, in this table, which we saw earlier.

(Refer Slide Time: 45:38)



The values were obtained under the assumption of either Gaussian PDF , or Laplacian PDF . It was also assumed that the input PDF have 0 mean and unit variance, but now in a practical scenario, it is strictly not necessary that the actual input source behaves as the one assumed, while designing the quantizer. Now, the two most important types of mismatch, which can occur are as follows. First the actual input probability density function matches with the PDF assumed during the design process.

But the variance of the actual input process does not match, with the variance assumed in the design process, or the actual input source probability density function itself does not match with the PDF assumed in the design process. And the variance of the actual input source, matches with the variance assumed in the design process. Now, there are different methods to overcome this problems and this have been extensively discussed in the literature under the topic of adaptive quantization, which we will not discuss in this class.

Now, the Lloyd max quantizer is an optimum quantizer as far as minimization of mean squared quantization error is concerned, but the implementation of a Lloyd max quantizer is far more complex than a uniform quantizer because of unevenly spaced decision boundaries. The very basic principle on which the design of a Lloyd max quantizer is based is to provide small size quantization intervals, in the regions where the input has high probability. Therefore, another approach to the design of a quantizer which exploits the input PDF would be to expand the region in which lands with high probability, in proportion to the probability with which the input lands in this region. We will investigate this approach in much more detail in the next class.