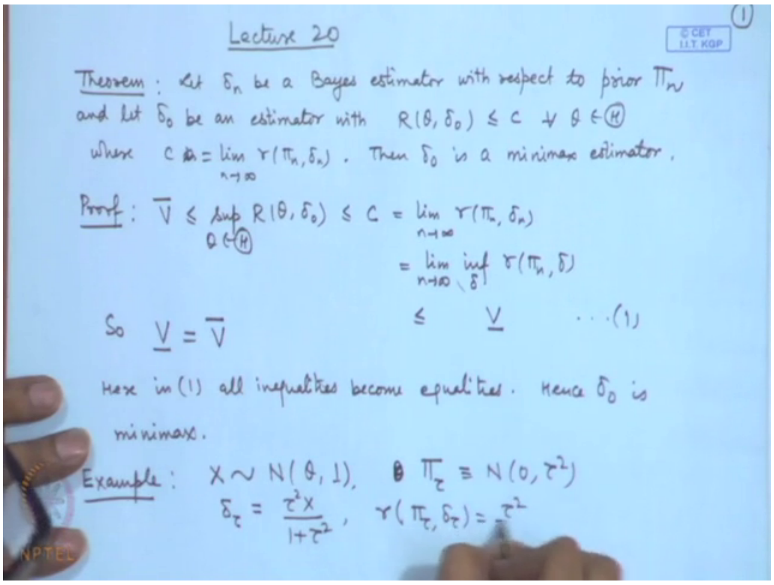
**Statistical Inference**
**Prof. Somesh Kumar**
**Department of Mathematics**
**Indian Institute of Technology, Kharagpur**

**Lecture No. # 20**
**Bayes and Minimax Estimation – III**

In the previous class we have introduced the relationship between the Bayes estimators and the minimax estimators in, in particular I gave a result, that if we can obtain a Bayes estimator with respect to a certain prior and then, if the risk satisfies, say, certain property, that is, there is a relation between the Bayes risk and the risk of the estimator, then the estimator is minimax and the corresponding prior is the least favorable prior. However, in many cases a single prior or a single Bayes estimator may not work, rather we consider a sequence. So, a result, in this direction I will restore it here.

(Refer Slide Time: 00:58)




We have the following theorem, let delta n be a Bayes estimator with respect to prior, say pi n and let delta naught be an estimator with R theta delta naught less than or equal to c for all theta, where c is limit of r pi n delta n as n tends to infinity, then delta naught is a minimax estimator. That means, we need not have actually the least favorable prior, but rather a sequence of priors, which are nice in the sense, that they are Bayes risk converge to a constant and the given estimator has a risk, which is bounded by that, then that estimator will be minimax. Let us look at the proof of this.

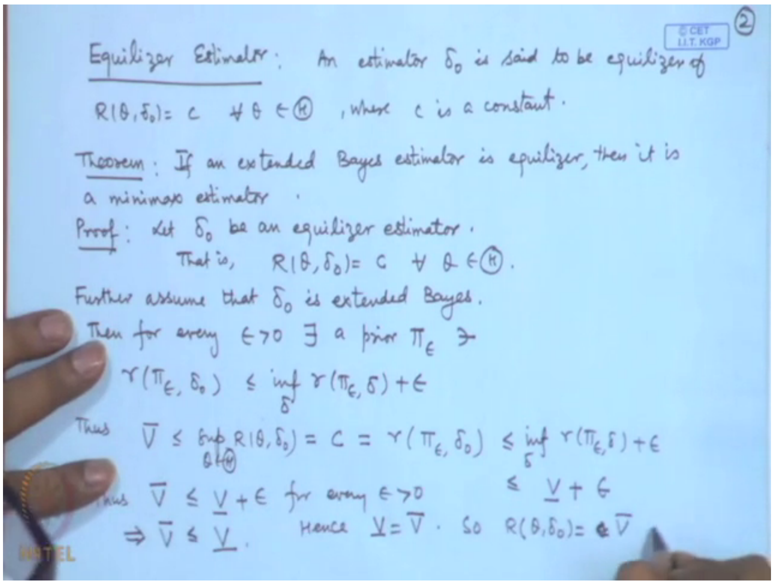
Now, we have already introduced the lower value and the upper value of the game or the lower value and the upper value of the estimation problem. So, using this we have the upper value of the game less than or equal to supremum of r theta delta naught, that is less than or equal to c. Now, that is equal to limit of r pi n delta n as n tends to infinity. Now, this is nothing, but the minimum Bayes risk with respect to prior pi n. Now, this minimum Bayes risk is always less than or equal to V lower bar and therefore, we are getting V lower bar is equal to V upper bar. So, again, here 1, all inequalities become equalities, hence delta naught is minimax.

As an application, let us consider the normal problem once again, x follows normal theta 1 and the prior distributions, that I had taken, pi tau as normal 0 tau square. Now, with respect to this we had seen, the Bayes estimator delta tau was tau square x by 1 plus tau square and the Bayes risk, that we had obtained was tau square divided by 1 plus tau square.

So, now, let us look at the limit of this as tau tends to infinity here, so this tends to 1 as tau tends to infinity. And if I look at the risk of delta star, that is the x, that was 1. So, in this case, this theorem is exactly applicable because 1 is less than or equal to 1 is true. So, delta star that is equal to x is minimax.

Let us summarize the result for the normal distribution. If I am sampling from a normal distribution and here, I have taken the known variance case. In this case what is turning out is that the sample mean is extended Bayes, it is generalized Bayes, it is a limit of Bayes rule, it is a minimax estimator, then the loss function is the squared error. In fact, even if I am considering the loss function to be an increasing function of the distance between the estimator and the estimand, this remains a minimax estimator. However, that, we will talk a little later.

(Refer Slide Time: 06:30)




Let us consider equalizer estimator. An estimator delta naught is said to be equalizer if r theta delta naught is equal to a constant for all theta. So, we have the following result. In fact, if we look at the two previous examples, in the normal example x was an equalizer rule because the risk is 1. In the binomial case, the risk was 1 by n of x by n, so x by n was an equalizer estimator. So, in general, the equalizer estimators are good in the sense, that they may be good candidates for becoming minimax estimators.

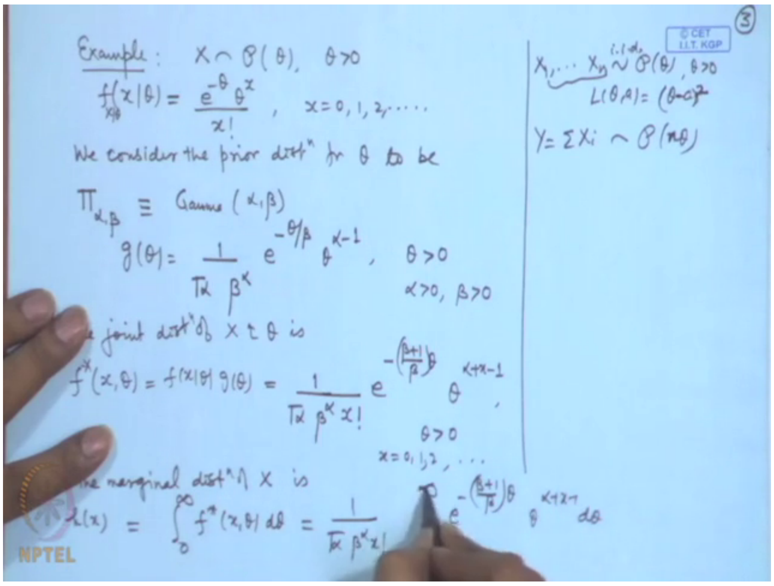
So, if an extended Bayes estimator is equalizer, then it is a minimax estimator. Let us take delta naught to be an equalizer, that is, we are assuming r theta delta naught is equal to c for all theta. Further assume, that delta naught is extended Bayes, then for every epsilon greater than 0 there exists a prior, let us call it pi epsilon such that r pi epsilon delta naught is less than or equal to infimum r pi epsilon delta plus epsilon.

So, now, let us consider the upper value of the game, that is less than or equal to supremum of r theta delta naught, the supremum is over the parameter space. Now, this is equalizer estimator, so this is equal to c and this is same as r pi epsilon delta naught because if the estimator is constant risk, then with respect to any prior, the Bayes risk will also be equal to c. This is, of course, less than or equal to infimum of r pi epsilon delta plus epsilon. This statement is actually a restatement of the definition of the extended Bayes rule. Now, this infimum is less than or equal to the lower value of the estimation problem plus epsilon.

If we skip the intermediate statements, we are actually saying, for every epsilon V upper bar is less than or equal to V lower bar plus epsilon. V upper bar is less than or equal to V lower bar plus epsilon for every epsilon greater than 0, which means, that V upper bar should be less than or equal to V lower bar. Since V lower bar is always less than or equal to V upper bar, we get equality here. So, what we get here, that r theta delta naught, this must be actually equal to c, this must be equal to V upper bar, that is, delta naught is minimax. So, if an extended base rule is equalizer, then it will be minimax.

Now, if we look at the previous example of normal distribution we have proved actually, that x is extended Bayes and it was also equalizer because the risk function was 1. Therefore, this will be minimax from that result also, although here we have used another result to prove minimaxity.

(Refer Slide Time: 11:35)



Let us take another example to prove the minimaxity of an extended Bayes rule. So, I will take a discreet case. Now, let us consider, say x following Poisson theta distribution once again. Let me explain here; see, we could have considered the model like x 1, x 2, x n following Poisson theta. So, we are having a random sample from Poisson theta and we need to make certain inference about, say the mean of the Poisson distribution, that is, theta here. But here, if we see, the sufficient statistics is sigma x i and that will follow Poisson n theta.

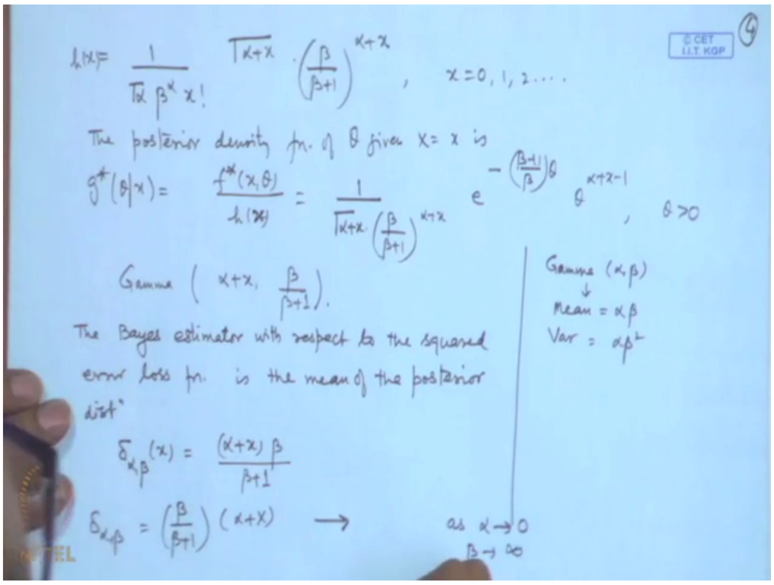
Now, as we gave argument in the normal distribution making an inference about n theta is same as making an inference about theta and here, especially we are considering estimation with respect to the squared error loss function. So, the criteria for Bayesian minimaxity, minimum variance unbiased estimator invariance, etcetera, will not be changed if I replace the problem of estimation of n theta by theta or theta by n theta. So, we can reduce the model to Poisson theta x following Poisson theta, theta greater than 0.

So, here the distribution of x, which we write as the conditional distribution of x given theta is E to the power minus theta, theta to the power x by x factorial x is equal to 0, 1, 2 and so on. We consider the prior distribution for theta to be, let me use a notation pi alpha beta, this is as a gamma distribution with parameters alpha and beta, that means, we are assuming the density function to be 1 by gamma alpha beta to the power alpha, E to the power minus theta by beta, theta to the power alpha minus 1. Here, an alpha and beta are known positive constants here.

Now, here, the distribution of x is discreet and the distribution of theta, that is continuous here. So, the joint distribution of x and theta, it is actually a mixture because x is discreet and theta is continuous, so this is equal to product of f x theta into g theta. So, these are the constant terms, like gamma, alpha, beta to the power alpha, etcetera.

Then, the terms of theta should be combined. So, what we do? We write it as 1 by gamma alpha beta to the power alpha x factorial E to the power minus beta plus 1 by beta theta, theta to the power alpha plus x minus 1. Here, theta is greater than 0 and x is equal to 0, 1, 2 and so on.

(Refer Slide Time: 15:52)




So, the marginal distribution of x can be easily obtained, that is obtained by integrating. Now, this becomes simply a gamma function, so the integral can be easily obtained. So, applying the formula for the gamma function we get the value of this integral as 1 by gamma, alpha, beta to the power alpha x factorial gamma of alpha plus x into beta by beta plus 1 to the power alpha plus x here. Here x takes values 0, 1, 2 and so on.

So, this is the probability mass function of x and therefore, the posterior density function of theta given x is obtained as f star x theta divided by h x. So, when we take the ratio these terms get cancelled out and the remaining quantity we get as 1 by gamma alpha plus x divided by beta by beta plus 1 to the power alpha plus x E to the power minus beta plus 1 by beta theta, theta to the power alpha plus x minus 1. This is nothing, but a gamma distribution with parameters alpha plus x and beta by beta plus 1. In fact, this is a phenomenon, which is like we are saying that the prior distribution is gamma and the posterior distribution is also gamma. Earlier we took the normal distribution as the prior in the normal case and posterior was also normal, these are called conjugate priors.

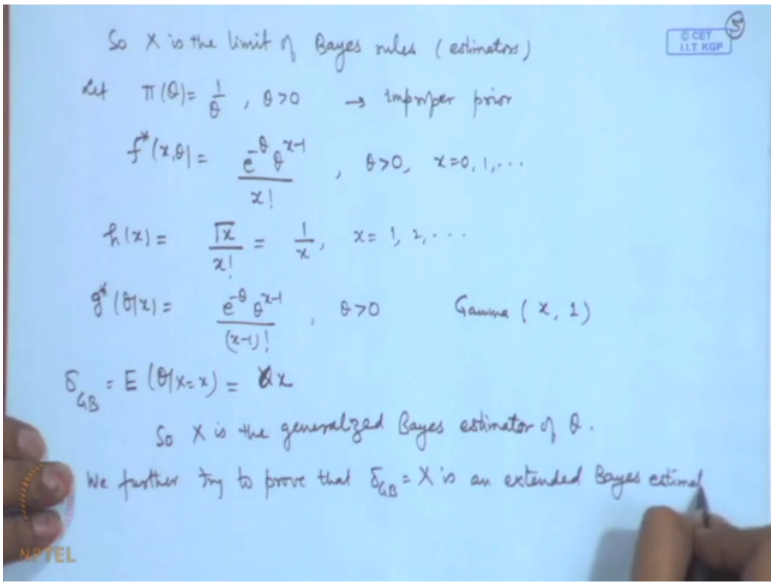
So, when we are dealing with the Poisson distribution, gamma is a conjugate prior for the average of the Poisson distribution. Similarly, for the normal distribution, for the mean the normal distribution itself is a conjugate prior. Similarly, like in a binomial distribution, we get binomial beta distribution as a conjugate prior because there if we take beta, then the posterior will also be beta and similarly, in many other distributions.

It is convenient to work with the conjugate priors in the sense, that easily the form of the base estimators can be derived and secondly, in the limiting sense, in many cases this lead to the extended Bayes rules and also the minimax rules.

So, now, here you can notice, here we have got it as a gamma distribution. Now, if I have, say gamma, say alpha beta distribution, then the mean is alpha beta; the variance is alpha beta square. The 2nd moment, for example, if I say, fine, so, so here the Bayes estimator with respect to the squared error loss function is the mean of the posterior distribution. So, that is equal to, let me name the estimator as delta alpha beta, so that is we got alpha plus x into beta by beta plus 1. So, you can write the Bayes estimator as beta by beta plus 1 into alpha plus x.

Further, you notice here, if I take here the limit as alpha tending to 0 and beta tending to infinity, then this will converge to X. Now, in the case of Poisson distribution the maximum likelihood estimator is x bar and of course, we have shifted the problem. So, X, X is the m L E, it is also the unbiased estimator and this turns out to be the limit of the Bayes rules.

(Refer Slide Time: 20:25)



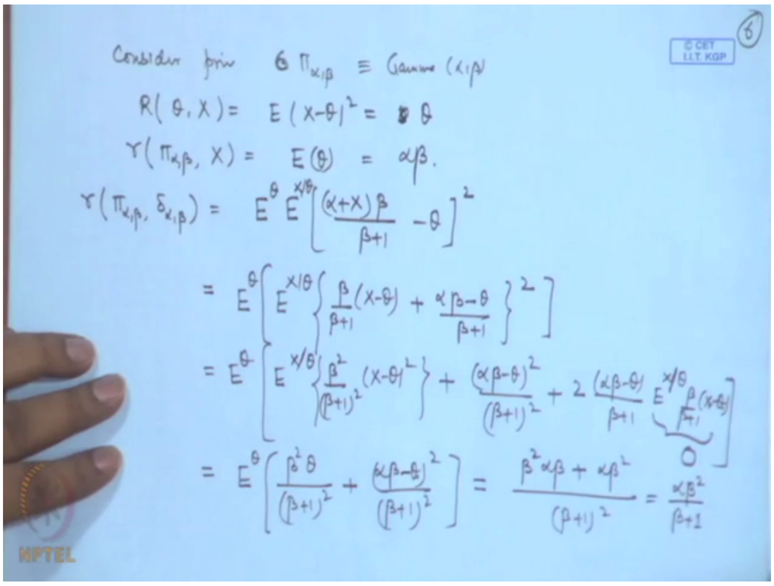

So, so X is the limit of Bayes rules or Bayes estimators. Let us consider, say pi theta is equal to 1 by theta, theta greater than 0. This is an improper prior, if we take this as the improper prior, then the corresponding joint distribution will be equal to E to the power minus theta, theta to the power x minus 1 by x factorial.

Then, the marginal distribution of x, that will be obtained as gamma x by x factorial because that is obtained by integrating this with respect to theta. So, you get gamma x, that is equal to 1 by x and of course, this will be from x equal to 1, 2 and so on because at x equal to 0, this does not make sense here.

So, g star theta given x, that turns out to be E to the power minus theta theta to the power x minus 1 by x minus 1 factorial, that is actually gamma x, 1 distribution. So, if you look at the expectation of theta given x, that is, the generalized Bayes estimator is given by x. So, x is the generalized Bayes estimator.

Further, we will see, whether it is an extended Bayes estimator. We further try to prove, that delta g b, that is equal to x is an extended Bayes estimator. Now, in order to take the extended Bayes estimator we need the sequence of priors. Now, already we have considered the sequence of prior as gamma alpha beta distributions.

(Refer Slide Time: 23:33)



Now, with respect to that let us calculate the infimum and other quantities, minimum Bayes risk, etcetera. So, first of all consider prior, that pi alpha beta, that is, gamma alpha beta. So, risk of X, that is, expectation of X minus theta square, that is equal to 1, sorry, that is equal to theta. Therefore, the Bayes risk of x, that will be expectation of theta, that will be equal to alpha beta.

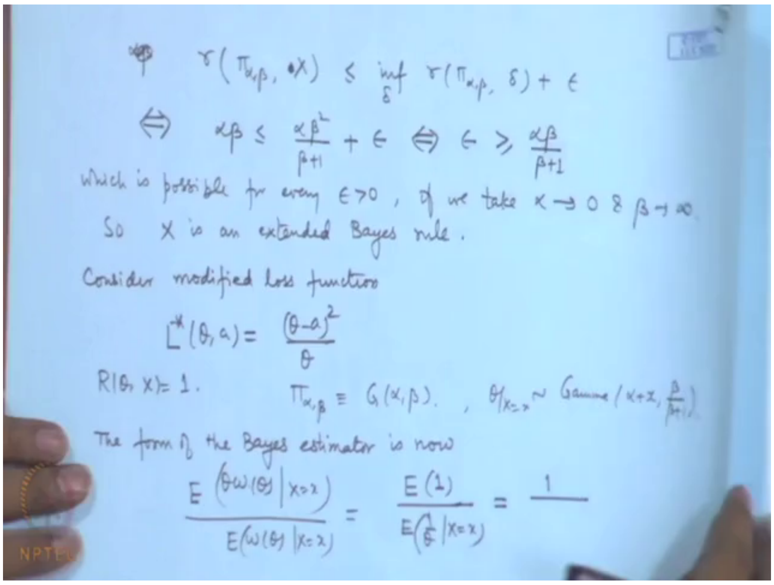
Now, let us consider the Bayes risk of the Bayes estimator, that is, the minimum Bayes risk that is equal to expectation of expectation alpha plus x into beta by beta plus 1 minus theta square. Now, here we can consider any order theta given x and then, x or x given theta and then theta. So, let me take this case here. So, this becomes expectation of theta expectation of x given theta. See, here, x given theta follows a Poisson distribution, so we will make use of that thing here. So, this term we express as beta by beta plus 1 x minus theta plus alpha beta minus theta by beta plus 1.

Now, if I take expectations here, then this term becomes beta square by beta plus 1 square x minus theta square. The 2nd term is constant with respect to x, so this becomes simply alpha beta minus theta square by beta plus 1 square plus twice alpha beta minus theta by beta plus 1 expectation of x given theta beta by beta plus 1 x minus theta.

Now, note here, that when I am considering conditional expectation with respect to x given theta, then the expectation of x is equal to theta. So, this term vanishes, this is constant and here, it will become variance of x. Variance of x is again theta, so this turns out to be simply expectation beta square theta by beta plus 1 square plus alpha beta minus theta square by beta plus 1 square.

Now, here, we have to take the expectation with respect to the distribution of theta, which is gamma alpha beta. So, expectation of theta is alpha beta and this term becomes expectation of theta minus alpha beta square. Now, alpha beta is the mean of theta, so this becomes variance of theta. The variance of the gamma alpha beta distribution is alpha beta square. So, this turns out to be beta square alpha beta plus variance of this, that is, alpha beta square divided by beta plus 1 square. So, here we can take common alpha beta square, so we get alpha beta square by beta plus 1.

(Refer Slide Time: 27:44)




So, now, if we write down the condition alpha beta, that is, r pi alpha beta delta g b, that is, x less than or equal to infimum r pi alpha beta delta plus epsilon. This condition is equivalent to same, alpha beta is less than or equal to alpha beta square by beta plus 1 plus epsilon, which is equivalent to same, epsilon is greater than or equal to alpha beta by beta plus 1, which is possible for every epsilon greater than 0, if we take alpha tending to 0 and beta tending to infinity. So, X is an extended Bayes rule.

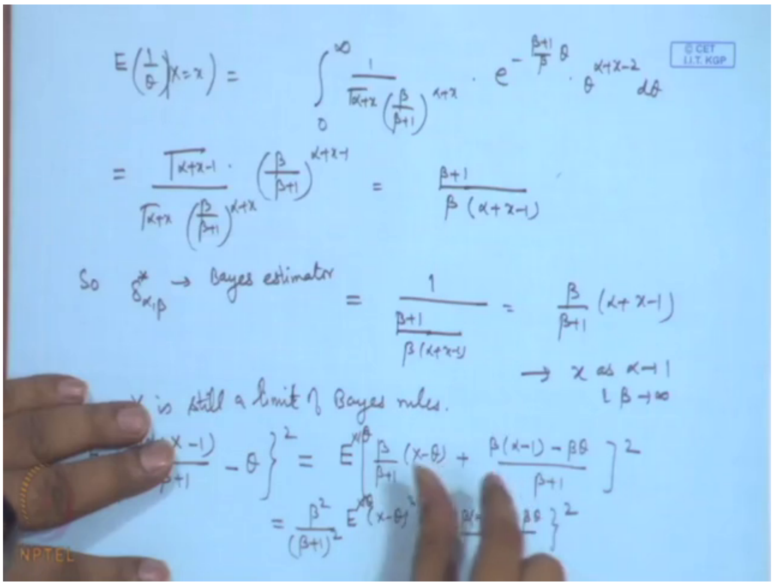
Let us also explore the possibility, that X can be minimax here or not. Then, the 1st problem that we notice here, that X is having the risk equal to theta and if I take the supremum value here, that becomes infinite. Therefore, at least, here X cannot be a Bayes rule. However, if I had taken the loss function where I had divided by theta, then this would have become constant. Of course, in that case the Bayes rule will change, and whether the new rule will remain generalized Bayes or extended Bayes also we have to see.

So, let us try that thing here because in this case the supremum risk is infinite, so certainly this cannot be minimax. So, we will consider modification of the loss function here; consider modified loss function. Let us call it L star, that is equal to theta minus a square by theta.

Let us look at all the other things here. So, r theta x, that becomes 1. Let us still consider the prior pi alpha beta as say, g alpha beta if I take, then posterior has already been calculated, that is, theta given x was following gamma with parameters alpha plus x and beta by beta plus 1.

However, the form of the Bayes estimator will change. Now, the form of the Bayes estimator is, now so, if you apply the formula expectation of theta into w theta by expectation of w theta, then that is equal to, now here w theta is 1 by theta, so this in the numerator you get 1 and in the denominator you get 1 by theta given x is equal x. Now, so the numerator is 1 and expectation of 1 by theta when theta given x follows alpha plus x beta by beta plus 1 gamma density, then that can be calculated here.

(Refer Slide Time: 31:52)

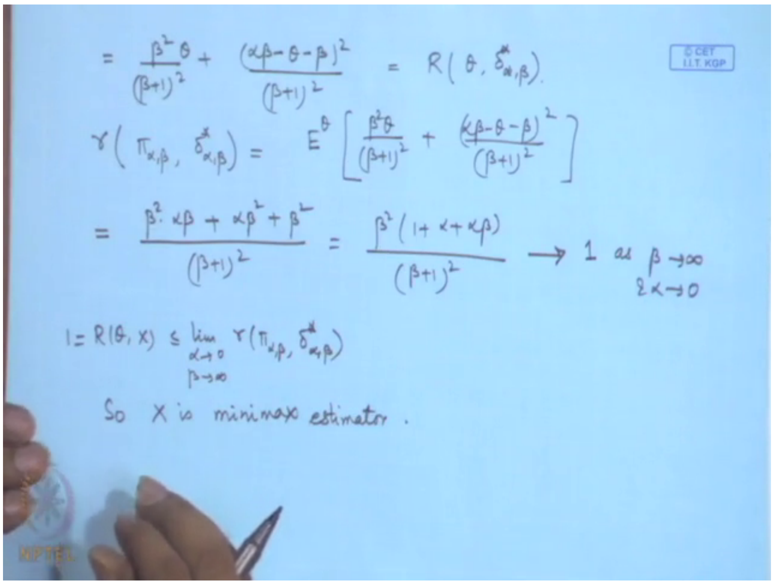



So, expectation of 1 by theta given x is equal to x, that is equal to integral 1 by gamma alpha plus x beta by beta plus 1 to the power alpha plus x e to the power minus beta plus 1 by beta theta and theta to the power alpha plus x minus 2. So, this is again a gamma function and the form is gamma alpha plus x minus 1 into beta by beta plus 1 to the power alpha plus x minus 1 divided by gamma alpha plus x beta by beta plus 1 to the power alpha plus x. So, that turns out to be then beta plus 1 by beta alpha plus x minus 1.

So, the Bayes estimator delta star alpha beta, that is equal to 1 by this quantity, that is beta plus 1 by beta alpha plus x minus 1, that is equal to beta by beta plus 1 alpha plus x minus 1. Now, here also you can see that this will converge to x as alpha tends to 1 and beta tends to infinity. So, x is still a, x is still a limit of Bayes rules and we can also see, whether it is extended Bayes or not.

Let us calculate the risk function. So, expectation of beta alpha plus x minus 1 by beta plus 1 minus theta square, that is equal to expectation of, once again here I will calculate this with respect to the distribution of x. So, this we adjust here, this is equal to beta by beta plus 1 x minus theta plus beta alpha minus 1 minus beta theta by beta plus 1 square. So, that is equal to beta square by beta plus 1 square expectation of x minus theta square plus this constant square, that is, beta alpha minus 1 minus beta theta by beta plus 1 square. Once again the cross product term will become 0 because expectation of x is equal to theta. So, in the cross product this will be out and expectation of this will give 0, so we do not write that term here.

(Refer Slide Time: 35:20)




And we get here, that is equal to beta square by beta plus 1 square into theta plus alpha beta minus theta minus beta. This term is actually equal to theta here. So, this is actually risk of the Bayes estimator delta star alpha beta.

So, if I calculate the Bayes risk with respect to the prior delta alpha beta, pi alpha beta, then it is the risk of this Bayes estimator that is equal to expectation with respect to theta of this quantity. Now, this will involve the mean and the variance of the distribution of theta. The distribution of theta is gamma alpha beta, the mean is alpha beta, the variance is alpha beta square, so we get this term as equal to beta square alpha beta plus.

Now, in the second one, again I expand this, this is variance, that is alpha beta square plus beta square divided by beta plus 1 square. If we take out beta square here, we get 1 plus alpha plus alpha beta divided by beta plus 1 square. Of course, we can notice here, this goes to 1 as beta tends to infinity and alpha tends to 0.

So, so we can see, that here, let us also look at the, we had r theta x is equal to 1 and this is less than or equal to the limit of r pi alpha beta delta star, alpha beta limit is taken as alpha tends to 0 and beta tends to infinity, so x is minimax…

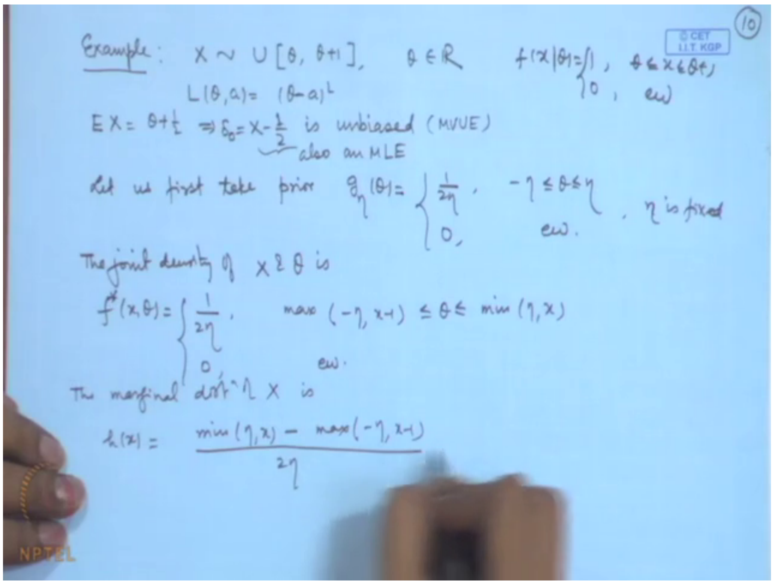
So, here you can see, that by making an equalizer estimator I am able to convert x as a minimax estimator. In the loss function when I considered squared error, then I was getting the risk as theta and therefore, the supremum was infinite, but if I divided by theta here, the risk became constant and as a consequence x has become a minimax estimator here. Of course, one can still prove, that it will remain a, it will remain an extended Bayes rule etcetera, but that we are skipping right now here.

In a similar way we can consider, we can consider the relation of invariance. For example, in order to find out the minimax rule, we can restrict attention to the class of invariant rules. If we can find out the best invariant rules, then it is actually the minimax rule in the class of invariant estimators. Then, under certain conditions this rule will be minimax over all.

Now, some simple cases are, like if I consider a finite group of transformations or a compact group of transformations, in that case the result is straightforward. However, if the group of transformations is not necessarily finite or compact, that means, in general, it could be a locally compact group. Then, there are certain conditions under which it can be proved that the best invariant rule are the minimax invariant estimator is actually minimax over all. In a similar way we have the result, that if we can find out Bayes rule or Bayes estimator within the class of invariant rules, then it will be Bayes over all. We can find out admissible rule within the class of invariant rules, then it will be admissible over all.

So, many of these results are now part of the standard estimation theory, I am not going to discuss this in detail here. I will still consider few more examples of non-conventional cases for deriving the Bayes rules, the limit of Bayes rules, the generalized Bayes rules, etcetera.

(Refer Slide Time: 40:45)



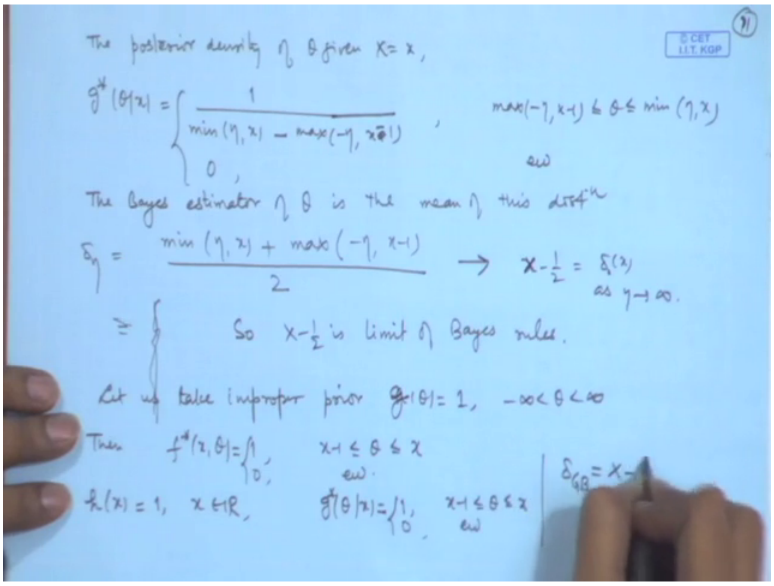
So, let us take another non-conventional example. Let us consider say x following uniform distribution on the interval theta to theta plus 1, where theta is any real number and we are considering the loss function to be squared error.

Firstly, let us understand this problem. Here, if I take expectation of x, that is equal to theta plus 1 by 2, that means, x minus 1 by 2. Let us write it as delta naught, this is unbiased and of course, it will become because here I am taking only one observation. So, this is also MVUE here. And if I consider the maximum likelihood estimator, the maximum likelihood estimator will be any value between x minus 1 to x and that again, we can take the mid value, this is also an MLE, although MLE is not unique in this problem.

Now, let us consider certain things here. Let us first take prior, say g eta theta that is equal to 1 by twice eta, where eta is fixed constant. That means, I am considering the prior distribution for theta to be a uniform distribution on the interval minus eta to eta.

Now, with this one let us calculate here the Bayes rule. So, the joint distribution, the joint density of x and theta is, it is equal to 1 by 2 eta because here the density of x, that will be equal to 1 for x lying between theta to theta plus 1. So, this product becomes same and the reason is here we can say maximum of minus eta x minus 1 less than or equal to theta less than or equal to minimum of eta and x, it is equal to 0 elsewhere. So, the marginal distribution of x, that is, minimum of eta x minus maximum of minus eta x minus 1 divided by twice eta.

(Refer Slide Time: 44:17)




And therefore, the posterior distribution, the posterior density of theta given x, then that is obtained as 1 by minimum of eta x minus maximum of minus eta x minus 1 for theta lying between maximum of minus eta x minus 1 less than or equal to theta less than or equal to minimum of eta x, which is actually a uniformed distribution.

So, here, for this model of uniformed distribution, actually another uniformed distribution is the conjugate prior and therefore, the Bayes estimator of theta is the mean, so that is nothing, but the midpoint here, that is, minimum of eta x plus maximum of minus eta x minus 1 divided by 2, we can actually write it as, that is fine. Now, as eta tends to infinity, this will tend to x minus 1 by 2 as eta tends to infinity.

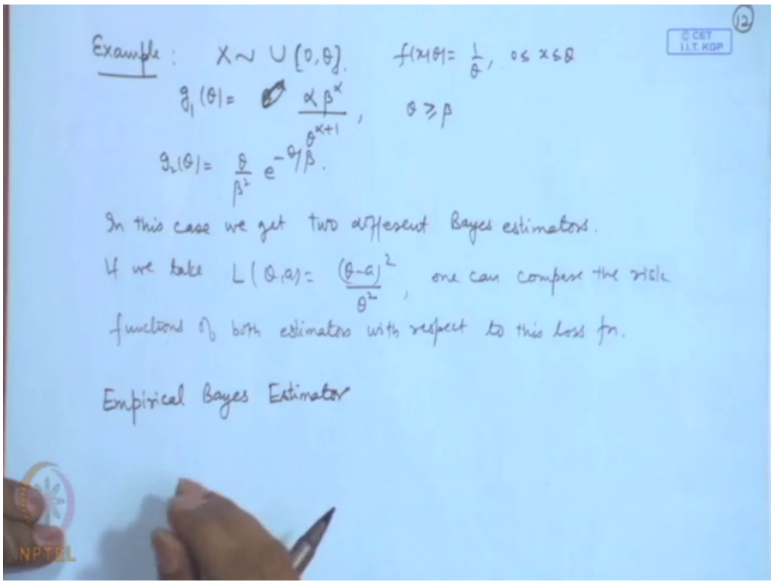
Now, let us also see, so this x minus half, this is limit of Bayes rules. Let us take improper prior, say, h theta is equal to, g theta is equal to 1 for theta belonging to the real line. In that case, f star x theta will be 1 for x minus 1 less than or equal to theta less than or equal to x 0 elsewhere. And the marginal will turn out to be again 1 for x belonging to r and therefore, the posterior will be equal to 1. And therefore, the generalized Bayes estimator is the mean of this distribution, that is, x minus half.

So, in this particular problem also we are able to get the limit of Bayes rules. The generalized Bayes rules, of course, I leave it as an exercise to check whether it will also be an extended Bayes rule or not. The question regarding whether it will be minimax, can also be tried. So, this part I am leaving as an exercise.

Now, now quite often there is a dispute, that whether we have one prior or another prior distribution, then one can actually calculate Bayes estimators with respect to the different priors and then, one can compare the risk functions with respect to a particular loss function and then see, which one performs better.

However, if we consider Bayes estimators with respect to different loss functions, even though you may take the same prior distribution, then you cannot compare them on the same scale. It could be, then with respect to, certainly because with respect to one loss function, one of them is the Bayes. That means, it is having the minimum Bayes risk and the other one is having minimum Bayes risk with respect to another loss function. So, the direct comparison is not possible. However, you may fix one loss function and then compare. In that case, deriving the Bayes estimators with respect to different priors or with respect to different loss functions is only a method of arriving at good estimators. I will illustrate this with respect to one problem here.

(Refer Slide Time: 49:18)




Let us take, say x, following uniform 0 theta once again and here, I consider one prior as g 1 theta, that is equal to theta to the power alpha, sorry, alpha and say, beta to the power alpha divided by theta to the power alpha plus 1. Here, theta is greater than or equal to beta, here f is 1 by theta and g 2 theta. Let us take as, say theta by beta square e to the power minus theta by beta.

Now, in this case, we get two different Bayes estimators. If we take loss function, say theta minus a square by say theta square, one can compare the risk functions of both estimators with respect to this loss function. So, this part I leave as an exercise for you to work out.

Now, what we have discussed so far, how to, we have introduced the optimality criteria, one of them we call Bayes and another one we call minimax. We have shown, that using the different criteria you may arrive at different estimators, but then one of the criteria is utilized to derive the other set of estimators here and therefore, there is a deep interrelationship between these concepts.

The other concept that is of admissibility is also closely related with that I have already introduced, the admissible estimator, that an estimator is said to be admissible if there is no estimator, which is better than the given estimator. Certainly, then the class of all admissible estimators is a desirable class and in the decision theory we call it an essentially complete class. In fact, it is the minimal complete class and therefore, it is desirable in any given estimation problem to restrict attention to the admissible estimators.

Now, proving admissibility or deriving admissible estimators directly is more difficult, so there is a direct interrelationship. In fact, the Bayes rules or the Bayes estimators are shown to be admissible under certain conditions. Conversely, all admissible rules are shown to be either Bayes estimators or limit of Bayes estimators and in that sense, if we consider the Bayes estimators, we are actually dealing with the good estimators.

Now, this is one of the things because of which the Bayesian rules or Bayesian procedures or Bayesian estimators are considered to be very frequently used in present day. And further, with the advent of computational techniques one can easily evaluate the Bayesian procedures because many times the form, the Bayes estimator may be complicated, but because of the computational procedures of level one can evaluate those things.

So, the advanced topics in Bayesian estimation and the minimax estimation relate to these concepts. In Bayesian estimation we have further topics, such as empirical Bayes. For example, the parameters of the prior itself may not be known. In that case, we can calculate the marginal distribution of x.

Now, that marginal distribution of x will include the prior, the parameter of the prior distribution, then we use the data to estimate that and when we consider the form of the Bayes estimator, we substitute the estimator for that parameter of the prior distribution from that data. This is called empirical Bayes estimator. Some of the prominent estimators, such as James Stein estimator, etcetera have been shown to be empirical base estimators.

Then, there is also concept of hierarchical Bayes estimators where we may consider a sequence of priors. For example, a prior distribution has a parameter. Now, that parameter is unknown, therefore we treat this as a random variable. We take further prior for that, so in that case, if we are, these are called hierarchical Bayesian procedures.

So, the Bayesian theory is rich theory and one can look into advanced topics of this. So, I stop my, I close the estimation part of this course today here. From the next lectures we will be considering the testing of hypothesis that is another important component of the statistical inference.