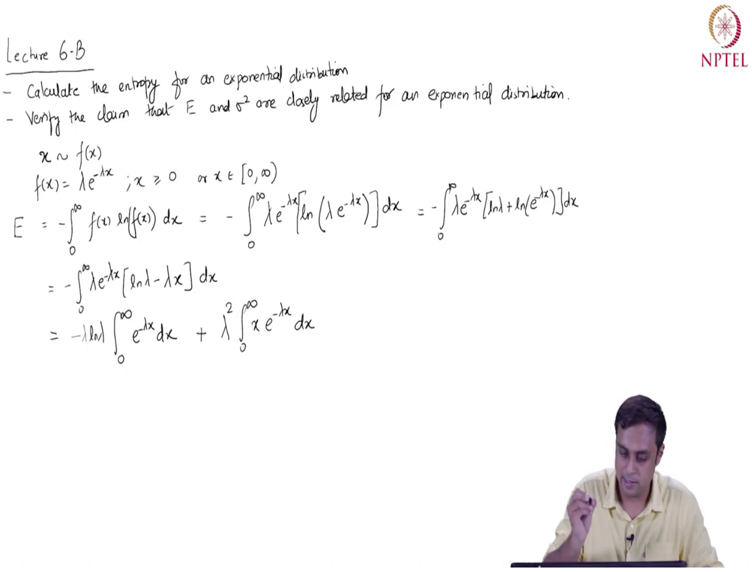
Spatial Statistics and Spatial Econometrics

Prof. Gaurav Arora
Department of Social Science and Humanities
Indraprastha Institute of Information Technology, Delhi

Lecture - 6B
Entropy - II

Hello everyone. Welcome back, to the 2nd part of lecture 6. So, we ended the previous lecture, where we wanted to calculate the entropy of an exponential distribution.

(Refer Slide Time: 00:28)



So, a random process that is characterized by the exponential distribution. And finally, we wanted to verify the claim, that the entropy measure and the variance of this exponential, exponentially distributed random process are closely related. So, they are kind of sort of measuring the same thing which is the variation, the variability of the random process, right.

So, let us begin by the let us get to the calculation part of entropy. So, we have *x* which is distributed *f of x*, and *f of x* is given as lambda e to the power minus lambda *x* such that *x* should be greater than or equal to 0 or we can write x belongs to the set start begins close 0 and ends on you know, open at infinity, right, ok.

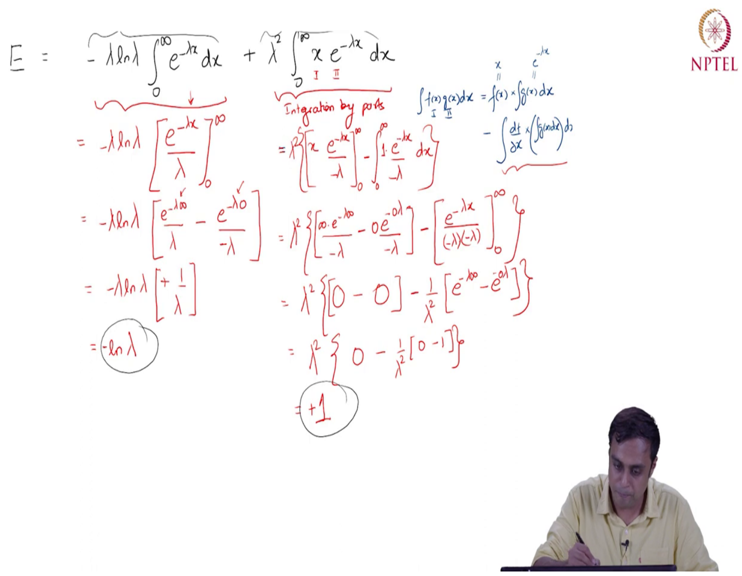
So, entropy measure by definition is 0 to infinity, a negative sign in the front, f of x ln f of x d x, right. This can be written as minus 0 to infinity, I am going to put the value of f of x now, lambda e to the power minus lambda x, log of the same thing, lambda e to the power minus lambda x whole thing times d x.

We can write this further, open we can simplify this further as, minus 0 to infinity lambda e to the power minus lambda x ln lambda minus or let us say plus ln of e to the power minus lambda x d x. So, this finally, turns out to be, minus integration from 0 to infinity, lambda e to the power minus lambda x, sorry about that, lambda x ln lambda minus lambda x d x, ok.

So, the entropy is given as, the following, lambda ln lambda e to the power lambda x minus. So, we can break these things into 2 entities, 2 different integration, entities minus let us say, plus lambda integration x e to the power minus lambda x d x. In the first integrand, I mean we can simply because lambda and ln lambda are constants.

So, we can simply take them out. So, I am just going to; I am just going to do that here. I am going to, first write minus lambda log lambda and then write integration 0 to infinity e to the power minus lambda dx times plus lambda x e to the power minus lambda x dx, ok. So, we have the, we have the measure with us now. So, we go into the next page, we will write this down again.

(Refer Slide Time: 04:03)



So, we have e equals minus lambda ln lambda 0 to infinity e to the power minus lambda x dx plus lambda 0 to infinity x e to the power minus lambda x dx, ok. So, this is going to be lambda square, I am quickly going to go back and check that my calculations are correct, ok, sorry. So, I have minus 1 lambda ln lambda 0 to infinity, ok. This is correct, the first one is good. The second one is, ok, I should have a square here.

So, plus lambda square, x into e to the power minus x dx. And, this looks good as well. So, what we are going to do is, we are going to now integrate these 2 entities separately. So, I am going to now first solve out the first entity which looks much simpler. This is going to be minus lambda ln lambda, which is just which are just constants. When I integrate exponential to the power minus lambda x, I will simply end up with the following. And these have to be evaluated from bounds 0 to infinity which are the bounds of x.

So, I can say minus lambda ln lambda, e to the power minus lambda times infinity over lambda minus e to the power minus lambda 0 over lambda, ok. So, I am multiplying lambda with 0 in the exponent. Now, e to the power minus lambda times infinity is nothing but, e to the power minus infinity here. So, this will just reduce to 0. So, this is just going to be 0. And the next the second component is e to the power 0 which is 1, right.

So, what I have is minus lambda ln lambda times minus 1 over lambda, which just simplifies to ln lambda, ok. The second integrand is a little bit more complicated. So, here I am working with, a situation where I must use integration by parts. So, I am going to call x as my first part and e to the power minus x as my second part.

So, Integration by parts: let us write down the formula. So, we have when we have f of x times g of x being interpret being, you know, integrated as a multiplier, that is given as this multiplication or integration by parts is given as f of x is taken as 1st, integrand g of x is the 2nd, this is given as f of x integral g of x dx, ok. So, this is multiplied here minus integration del f by del x times integration gx dx whole integrated with respect to dx, ok. So, this is a integration by parts formula we are going to apply it to this situation.

So, in applying integration by parts, I have f of x which in my example is x, and g of x which in my example is exponential to the power minus lambda x. So, I am simply going to follow the steps in the formula. So, we have, lambda squared we have lambda squared times f of x which is just x times integration of e to the power minus lambda x which we have seen in the previous you know case here, right here, we can write this as e to the power minus lambda x by minus lambda and this is bounded by 0 to infinity.

So, there is no upper bound of course, but there is a lower bound which is 0. So, we must evaluate this part like that. And we have that then we have the second component of integration by parts which is this. So, I have minus integration d f by d x which in this case will simply be 1, right. So, del x by del x is 1 times, integral of e to the power minus lambda x which we have now learnt x by minus lambda dx, ok.

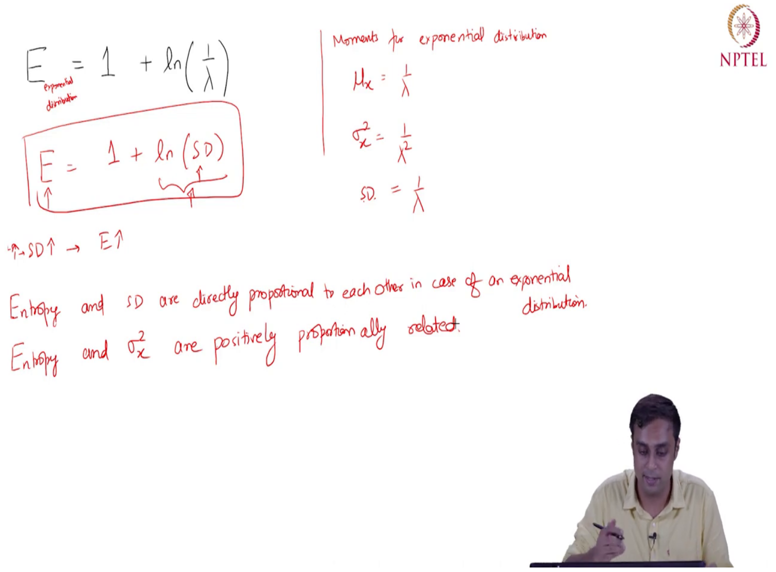
So, we are going to simply follow this through, we are going to solve for the first part, for which we are already aware of the bound. So, we can evaluate it. So, we have lambda squared applies to both entities. So, lambda squared times infinity e to the power minus lambda infinity over minus lambda minus 0 e to the power minus 0 lambda over minus lambda, ok.

So, this is part 1, the first component of integration by parts. In the second component, I have a negative sign and then I am integrating e to the power minus lambda x over minus lambda. So, minus lambda is simply as constant. So, this will turn out to be e to the power minus lambda x over minus lambda into minus lambda. This whole thing evaluated with the bounds between 0 and infinity, ok.

So, we have and curly brackets closed. So, we have lambda squared. So, in this part you have an exponent with a negative infinity sitting on it. So, the first you know portion of the first integral will be 0 minus is 0 obviously. So, this is my first part will become 0 here, minus we have 1 over lambda squared times e to the power minus lambda infinity minus e to the power minus 0 lambda.

So, I have lambda squared times 0 minus 1 over lambda squared multiplied by minus 1. So, I am left with just plus 1 as the solution to the second component, alright.

(Refer Slide Time: 11:09)



So, coming back to the fact that my entropy now, is nothing but, 1 plus log of 1 over lambda, we are going to toggle to previous screen and make sure that we have it correct on this page. So, expo, the entropy measure is equal to the first component plus the second component. So, this first component plus the second component, the first component was solved to be ln lambda, right. And the second component is solved to be plus 1. So, I have here 1 plus log 1 over lambda, ok. So, I have done a little bit of a fumble.

So, we have we are going to have negative for minus lambda ln lambda on the previous slide you can check my calculations again, ok. So, that is it. So, we have arrived at a entropy measure, for the exponential distribution. So, remember this is entropy for exponential distribution. And if you now, go back and look at the different moments for the exponential distribution.

So, moments for exponential distribution that we had figured out in the previous you know the first part of lecture 6, they were, in the following order. So, mu of x, where x is distributed exponential lambda is 1 over lambda. Sigma squared x was 1 over lambda squared and standard deviation was 1 over lambda. Hence, the entropy measure can be written as, 1 plus log of the standard deviation.

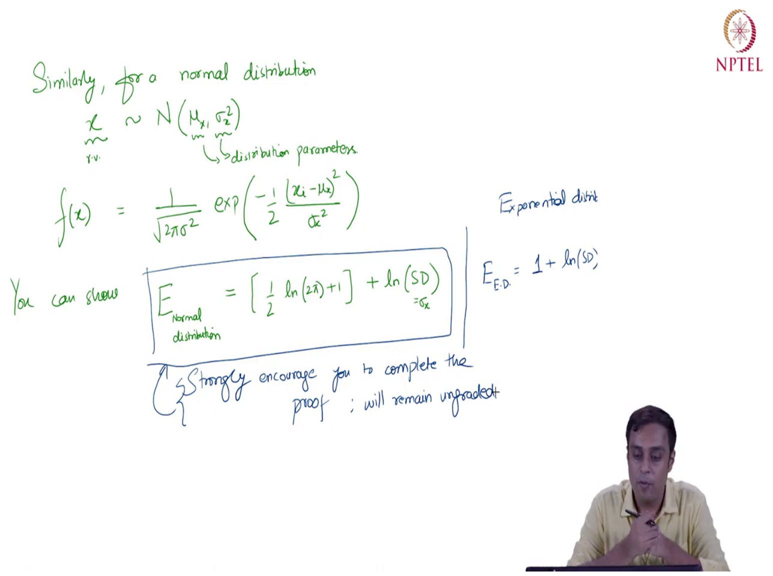
Standard deviation which is just the square root of variance is a measure for variation in the system. And the entropy measure turns out to be a you know a monotonic transformation of the standard deviation; that means, we are applying a log to the standard deviation measure.

So, if standard deviation is higher, then, the log of standard deviation is higher and 1 plus of you know this entity is also higher.

So, when SD increases, E increases and SD increases if sigma square increases, right. So that means, that higher variance is simply an information that is subsumed into the entropy measure as well. Therefore, what we are saying is, that entropy and standard deviation are directly proportional to each other, to each other in case of an exponential distribution, ok.

Now, as a sort of extension, because, you know SD is simply you know square root of sigma squared, we can also say that entropy and sigma squared x, are positively related or positively proportionally related, ok. Therefore, you know, we can say that entropy is an alternative measure of you know variability in a random process that sigma squared x provides us, a value for, ok.

(Refer Slide Time: 15:08)



Now, similarly you know, you can also sort of we can say similarly, for a normal distribution. So, what is a normal distribution? Well, you can have a random variable x, it is distributed normally, with mean, mu and variance sigma squared, we can call them mu x and sigma squared x. Now, here you know, x itself is a random variable mu x and sigma squared x these are the parameters of the distribution parameters.

And, you know f of x which is the density function for a normal distribution is given as, I am sure, many of you might be aware of this, is 1 over square root 2 pi sigma squared, exponential of minus half x i minus mu the whole square divided by sigma squared x, ok. And, you can show; you can show that exponential the, sorry, the entropy measure for the normal distribution entropy for normal distribution entropy for normal distribution, is given as half log of 2 pi plus 1 plus log of the standard deviation for you know, the standard the normal distribution.

Standard deviation is nothing but, sigma, right, sigma x which is the square root of the variance. Now, interestingly, you know, even though the variance of a normal distribution and the shape of a normal distribution looks very different, the formulation of a normal distribution looks very different from the exponential distribution. Very interestingly, the formulation where entropy and standard deviation are linked are very very similar, right.

Remember for the case of exponential distribution, we showed, right, we showed a proof that entropy for exponential distribution was log, sorry, 1 plus log of the standard deviation.

So, in case of normal distribution, we have very similar formulation and hence we can infer that entropy and standard deviation are directly proportionally related. And also, that, the information that is contained in sigma squared x, which is a traditional variance measure, which provides us the variability of the states of nature that are comprised in a normally distributed random process, can very well be represented by an entropy measure as well, ok.

I am not providing you a proof of the exponential, sorry, the entropy measure for a normal distribution. But I encourage you strongly, to go ahead and you know and complete this proof at your own time. This is a strongly encouraged assignment, but it will not be graded. So, this is going to be left to you, to you know complete, it will be a lot of it will give you a lot of clarity to go through the proofs.

So, in case of mathematical statistics you know, I have said that you know we should always work with examples and we should always look at how the data are behaving and the concepts are not sufficient. But on the other hand, it is also very very important to conduct analytics, to go through proofs and go through the steps of proofs.

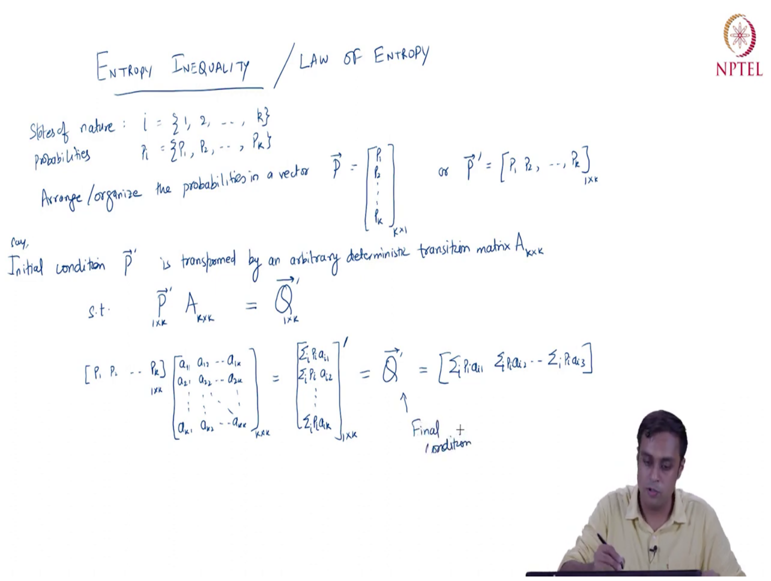
Because sometimes, our understanding of the real-world phenomena, enhances or increasing when we go through these proofs. We are able to understand, what is this measure really trying to get us at, right. The process of getting to the final step, the final closed form solution is very useful, if you want to apply an understanding of entropy or adapt it to a certain real-world situation.

So, I strongly encourage you to complete the proof, with the rider, then it will be it will remain ungraded. So, that is it, ok. So, now, having established having established, we have established now that entropy is you know we can provide an interpretation to entropy, which is quite similar to the interpretation that we have for the second moment of any distribution which is the variance, right, the variability the range with which a random variable can vary in a probabilistic environment.

We should ask, why are we studying entropy to begin with? I mean why do we come, if I mean variance could have been sufficient, right. I mean why do we need entropy? Well, the understanding of entropy, comes you know provides us a very useful philosophy of technology, you know to begin with, right.

It provides us an answer, why would we need technology, and why do as why do we as human race as government systems invest so much in technological development through times. It provides us a statistical, mathematical understanding of the same, right.

(Refer Slide Time: 21:22)



So, that is comprised in what is called as the "entropy inequality". This entropy inequality also is known as the "law of entropy", ok. Now, what is the law of entropy and how does it sort of you know signify the importance of this measure, even though we had an alternative like sigma squared x that is the variance, something that we are more used to as statisticians. And why should we even you know care to learn you know things like entropy?

So, you know let us go back to our example, where we had where we had k states of nature, states of nature given by i, that goes from 1 to k. And we have probabilities attached to each state of nature, Pi, which goes from P1, P2 to Pk, ok. Now, let us arrange these probabilities in a vector, ok. So, we will arrange or organize the probabilities in a vector P.

So, P is nothing but, a k by 1 vector, P1 to Pk that contains this information of all different probabilities that, you know, the states of nature can take. So, P is quite comprehensive. It first of all identifies the indices, i 1 to k, I mean that is an index on the probability. So, we know that we are working with k states of nature. And we also know the probabilities attached to each state of nature.

So, P by itself is pretty comprehensive, in describing the random process that we are working with. I mean we can write P, we can take a transpose of P and we will then get a row vector, which goes from P1, P2 to P k, right, which will be a 1 k vector, 1 row and k columns.

And now let us say, this initial condition say, initial condition P prime, that is the original states of you know probabilities of the states of nature, the original state of the world of the universe that we are working with is transformed by an arbitrary its arbitrary deterministic transition matrix A, which has the dimension k by k. So, what we are doing is, we are transitioning the world that we are in, to a new world of possibilities of the these k states of nature.

So, what I am going to say is, that I am going to transition transform P by a matrix A by simply you know such that, I take P which is a 1 by k matrix and multiply it by this k by k deterministic matrix and I get a new matrix, let us say Q. Q by itself, will has will have the dimension 1 by k. So, Q is nothing but a new vector of probabilities. So, I have new vector of probability.

So, state 1, state of nature 1 will have a new probability after being transformed from the previous state. So, I am applying a transformation or a transition matrix onto P and getting a new state of the world or the state of universe called Q, right. So, to see this process mathematically if you are interested, I mean you can look at, you can actually conduct, some matrix operations and it becomes a little clearer.

So, I am writing out P and then I can write out a, a will have you know its constituents a11, a12, a1k, a21, a22, a2k keep going, ak1, ak2, akk, ok. So, this transition, matrix is a k by k matrix. The matrix that I am transforming is a 1 by k matrix. So, what am I going to get? I am going to get a 1 by k matrix. How do I get this 1 by k matrix, I want the first element?

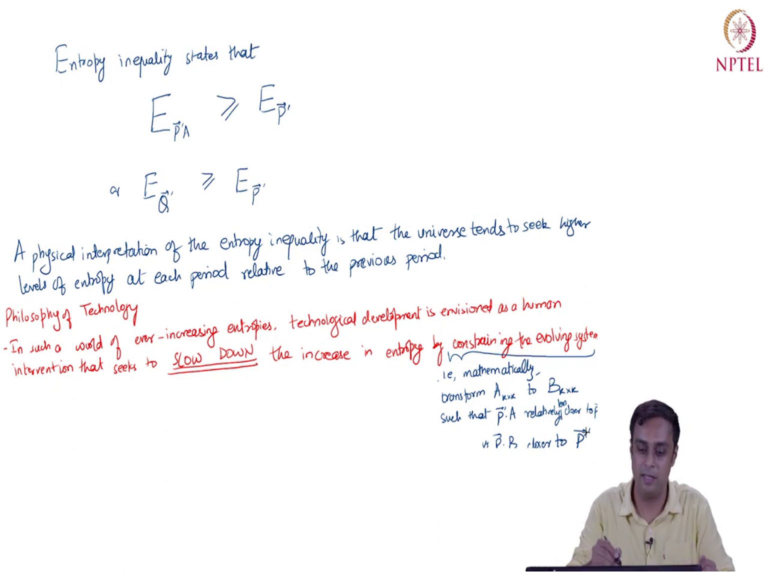
I will take the first row, that is the row vector and multiply it with the first column. So, I am going to get summation i Pi ai1, second element summation i Pi 2, sorry Pi ai2. So, what am I doing? I am simply taking a weighted average of all the probabilities and getting a new probability measure. I require this probability measure to be between 0 and 1. So, each element of this row vector must be between 0 and 1. And secondly, when I sum all the elements of this new row vector Q, they must all sum to 1.

So, I am transitioning from universe 1 to a universe 2 which are delineated by the probabilistic nature of how these states will occur. For my example for let us say, prices of gold, if I am looking at a commodity price environment and I have k different price levels that gold can take in universe 1. I have probabilities P attached to each state of or each price value or each price level that gold can take. I transition from this world to another world, where the prices price levels remain the same.

So, gold still takes those k price levels, but with new probability measures given by the vector Q, ok. So, let us complete this you know vector Q, right. So, this is Q prime. So, by the way I should have had a 1 by k. So, I am I should have I am writing a transpose of the column vector, right. So, this is really just summation i Pi ai1 i Pi ai2 summation i Pi ai3, ok, alright.

So, what does the law of entropy suggest? So, I have a state of world P, a state of world Q. We have started the initial condition was P, the final condition is Q. So, this is the final condition.

(Refer Slide Time: 28:57)



So, the law of entropy or the entropy inequality states, that, entropy inequality states that, the entropy measure of state P prime A, we know how to calculate the entropy, which is a description of the variability of this random process, is at least as greater as the probability or the entropy or the variability of the initial condition of the probabilistic universe that we are studying, right.

We can also write this as entropy of Q is greater than or equal to entropy of P, right. So, physical interpretation a physical interpretation, of the entropy inequality is that the universe tends to seek higher levels of entropy at each period or instance at each period, relative to the previous period, ok.

So, as we move through time or through different universes, you know, the universe that follows will have will we tend to have higher variance higher variability higher possibilities, you know a greater degree of possibilities than the universe that we have simply where we have just you know arrived from, right.

So, in such a world of ever enhancing ever increasing entropies, technology is envisioned as an as a human intervention, to slow down the process of how you know uncertain the future becomes to present, right. So, let us write it down. So, you know. So, I am going to write down what is now, I am going to call it the philosophy of technology, a philosophy, its a philosophy of technology which comes from the idea of you know universe being explained as a random process, right, it comes from that from the statistical understanding of the world.

So, I am going to say in such a world, of ever-increasing entropies, technology or technological development is envisioned as a human intervention that seeks to slow down, ok, that seeks to slow down the increase in entropy, by constraining by constraining the evolving systems, ok. So, the idea is when, I am trying to put this constraint you know.

So, mathematically if you want to understand what does it mean by constraining the evolving systems, you know what I am trying to say is that is mathematically, if you want to go back to our you know transition matrices and all that, you know when we transform you know basically you know this transition matrix which was A k by k to a matrix B k by k with this human intervention of technology.

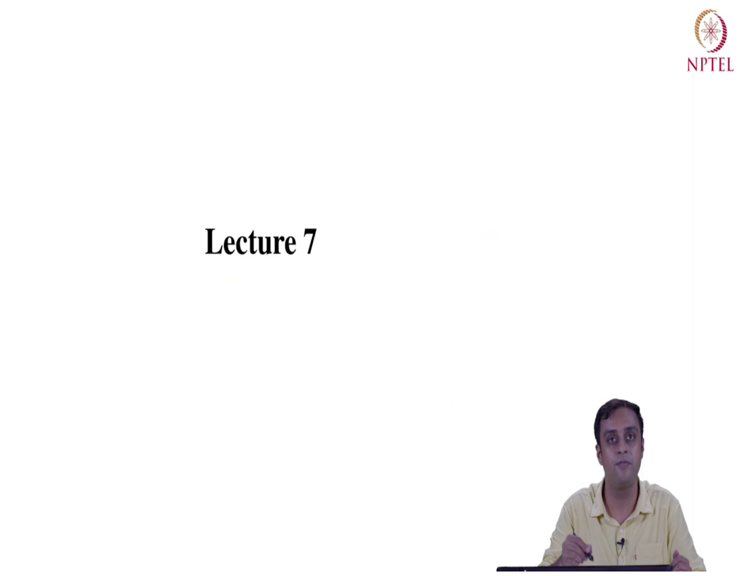
So, B is the transition matrix with a human intervention, A was the transition matrix which would have appeared naturally. This transition would be such that, the new state of the world which is P times A versus P times B, what will happen is that P times B will be closer to P relative to what P times A was, ok. So, P times B is closer to P.

So, we are closer to the previous universe that we are transitioning from, with human intervention versus you know when what we were, if there were no human intervention. So, that is that provides you a philosophy of technology you know, which can provide a rationale for why you know the human race or the governments all across the world sort of are always trying to invest in better technology or you know higher degree of technological advancements at any given point in time, right.

So, this is a philosophy you know this explains the world with a perspective, right. I mean the that is statistical perspective is not the only perspective out there, but it provides an explanation, irrational for what type of you know happenings we see around us you know in the world, ok.

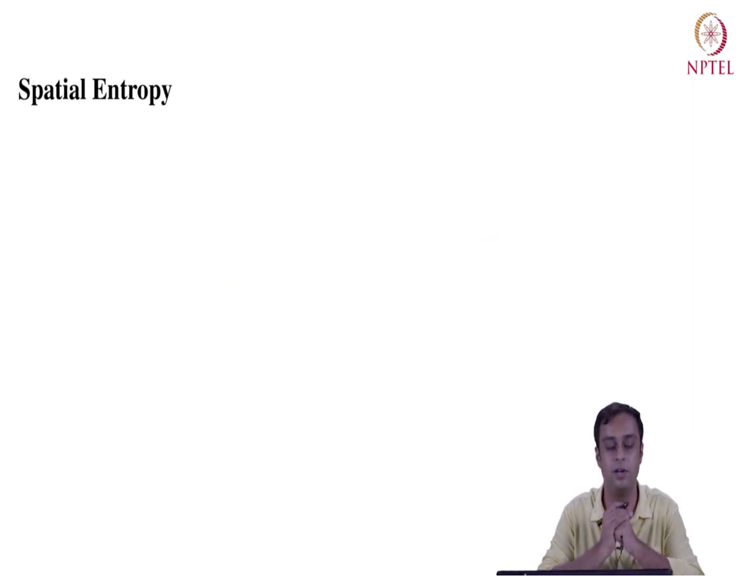
So, with that, I mean there is one thing that you know although technology can slow down, one pointer I want to give here is that, technology can although slow down the increase there are mathematical proofs that it can never reverse the process, right. So, it can only slow down the process of ever-expanding you know entropies or the degree of variability that we experience in the world, ok.

(Refer Slide Time: 35:39)



So, that is about it for this lecture. But I want to sort of end the lecture with this idea of what we are going to cover in the next lecture, which is spatial entropy.

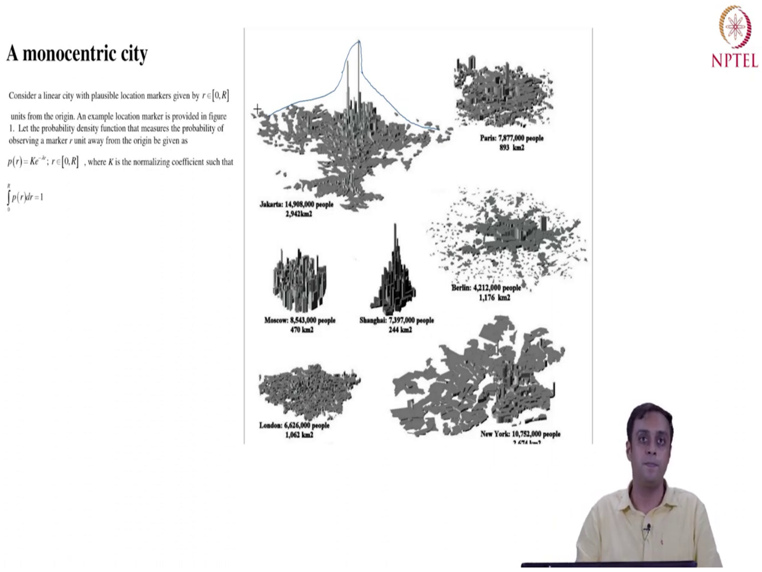
(Refer Slide Time: 35:42)



So, now we have established that entropy is a measure of variability of a random process. What would spatial entropy be? Right. You can you know you can try and think in your head, what would, how would entropy be then understood as spatial entropy? Well, one of the things that will happen is that, we will start to add this dimension of locationally delineated entropy. So, entropy with a weir component.

So, you know location wise, different differentiated variability of a random process would be you know spatial entropy.

(Refer Slide Time: 36:27)



So, where we will begin next time, is a model for a monocentric city, you know. So, here what we are going to try and do is use the concept of concept of entropy, to explain how cities are organized around us. So, there is often a city center, the city center has quite a high level of activity going on, the population density is higher, prices are higher and so on and so forth.

And as we start to move away from this center, you know the population density becomes lower, the prices become lower and so on and so forth, right. So, the cost of living the real estate prices, the population density and so on and so forth. So, spatial entropy will provide us a measure and a you know, an opportunity to model how cities are often organized around us, alright. So, that is all for today’s lecture. Thank you very much for your attention and I look forward to you know having you over in the next lecture.

Thank you.