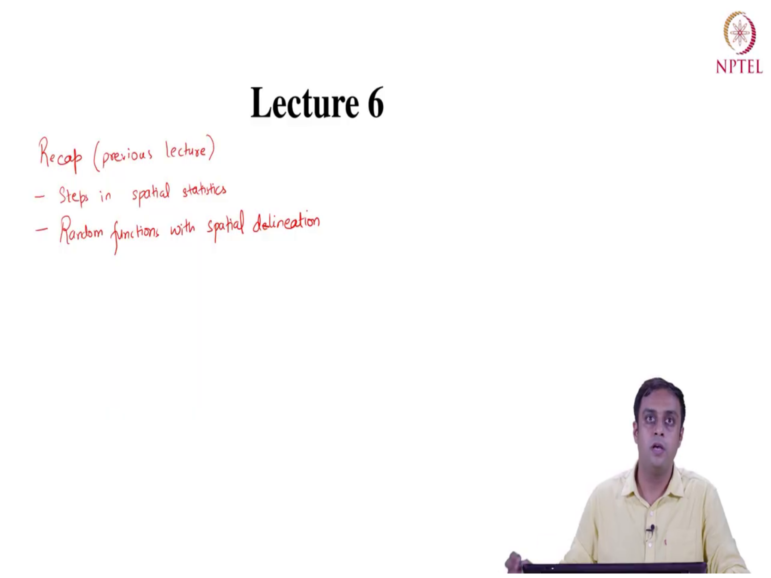
**Spatial Statistics and Spatial Econometrics**

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

**Lecture - 6A**
**Entropy**

Hello everyone, my name is Gaurav Arora and welcome back to the 6th lecture of Spatial Statistics and Spatial Econometrics, before we go on to this lecture let us do a little recap of what we covered in the last lecture.

(Refer Slide Time: 00:37)



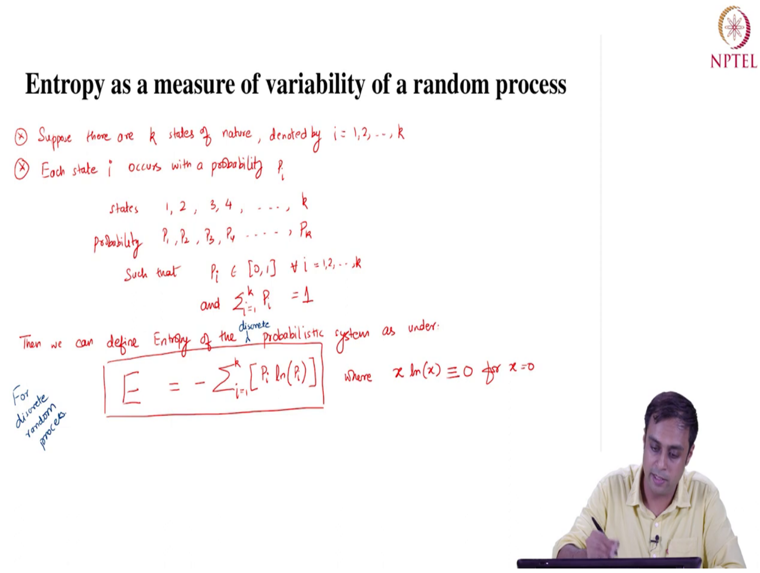

As you can see on your screen we broadly covered two topics, in the previous lecture where we formally sort of transitioned from data understanding of you know of spatial delineated products to statistical formulations and statistical you know modeling of those data.

So, the first thing we did was broad steps in spatial statistics. So, you know the first step was you know what is spatial statistics. Well, it is a science of uncertainty of spatial nature. So, the first step was to measure or quantify disorder of spatial nature right. So, to quantify spatial disorder and we talked about variance inter-quartile range and entropy being the measures alternative measures that we will look at.

The second step was mostly about modeling or measuring spatial dependence, the third step was from moving from correlation that is what of location delineated statistics to why in explaining, why do we see the type of spatial trends we see, why do we see the kind of you know spatial dependent structures that we see. And we ended the lecture with this understanding of random functions with spatial delineations right.

So, we looked at what a random variable is and how does an understanding of random variables, then translates into these jointly distributed random functions, where you have random variables located at different locations in space and then those are moving together bound by a density function called as the joint CDF alright.

(Refer Slide Time: 02:28)




So, in today's lecture we are going to look at a measure of spatial disorder or spatial variation known as entropy.

So, here on this slide I have a heading or a title called as, entropy, as a measure of variability of a random process. So, we are look going to study entropy in general, from those of you coming from you know statistics background and econometrics background, economics background, might not have heard of entropy or might not have really studied it in detail.

However, those coming from engineering and physical sciences might have actually studied entropy formally in your studies right. So, it will be interesting to first review entropy as a measure of disorder and then take that understanding for a general entropy measure of variation to entropy as a measure for spatial variation what we call as spatial entropy alright. So, let us get started.

So, in order to understand what is entropy, let first suppose there are $k$ states of nature alright. So, these can be indexed by or denoted by $i$ goes from 1, …, $k$ alright and each state i each state $i$ occurs with a probability $P_i$ ok. So, we have a probability measure for each state occurring in this system a random process or random system that we are trying to study.

So, basically you know what we have is states 1, 2, 3, 4 all the way to $k$ and then we also have a probability measure attached to each state that is $P_1$, $P_2$, $P_3$, $P_4$ all the way till $P_k$ right. And given that these are probability measures they should you know they should follow some properties that we know from our you know previous training in probability and statistics, but we are going to still look at them.

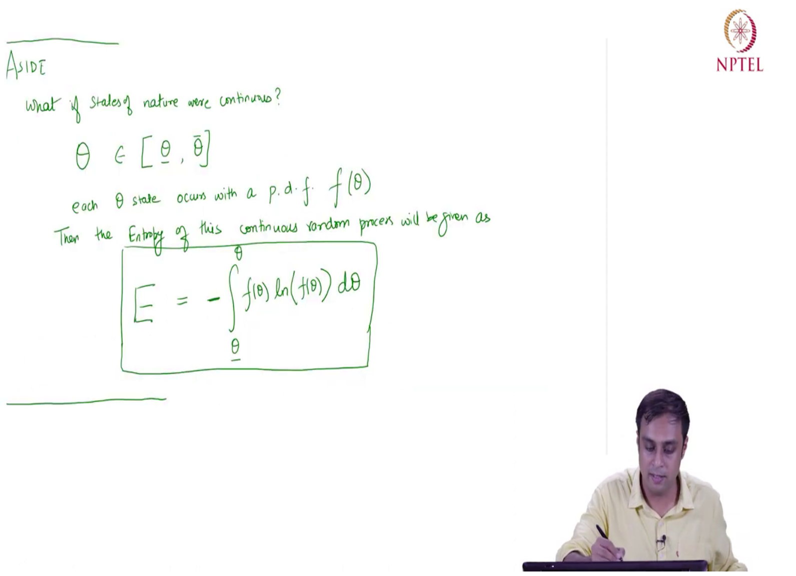
Each probability i must lie between 0 and 1 right, for all i's that is 1, 2, $k$ right and the sum of these probability measures for all k states must be equal to 1 right. So, the probabilities must sum to 1 and each probability entity must be between 0 and 1 right. Then having learnt that having learnt that we can define entropy of the probabilistic system as under; so, we have $E$ as the entropy measure for the system equals minus summation $i$ equals 1 to $k$, $P_i$ that is the probability measure times the log of $P$.

$$E = -\sum_{i=1}^{k} P_i \, ln(P_i)$$

So, it is a natural log that I am taking of $P_i$ and you know times the probability itself right. So, $P_i$ times log $P_i$ summed across all $k$ states of nature and then multiplied by minus 1 is equal to entropy this is the definition of entropy right. We have to be a little bit careful here first of all, you know look $P_i$ can take a value of 0, but then log of 0 is not defined right. So, we say where. So, to account for that we say where x log x is defined to be 0 for x equals 0.

So, wherever you have probability being 0, that does not contribute to the entropy of the system right. So, the state that does not occur does not contribute to the entropy of the system that is the interpretation of the physical world that we get here. So, one thing we you know we have defined entropy formally, but the system that we have worked with is a discrete probabilistic system right. So, what we are working with is a discrete random process right. So, this is a entropy definition for discrete random process ok, what if we had a nature where states were continuous let us look at that.

(Refer Slide Time: 08:00)




So, let us call this a side let us call this a little aside so, that we can define entropy for the case where the states of nature were continuous. So, now, you know the states of nature are denoted by θ instead of $i$, θ is a continuous random variable which goes from theta lower bar to a upper bar. So, there is a lower bound and upper bound and each state of theta right each theta will appear or occur with probability density function f of theta right.

So, each theta state occurs with a probability density function denoted by $f(\theta)$ which is nothing but also the frequency with which each theta value will appear. Then, the entropy of this continuous random process will simply be given by $E$ equals minus. So, now, the summation in case of continuous system becomes integral, integral that ranges from the lowest to the highest possible value of theta, times $f(\theta)$ which is nothing but $P_i$ in case of you know the discrete case and times the natural log of $f(\theta)d(\theta)$ ok.

$$E = -\int_{\theta_{lower}}^{\theta} f(\theta) ln(f(\theta)) d(\theta)$$

(Note: $\theta_{lower}$ in the above equation implies the lowest value of θ, in the slide θ "lower bar".)

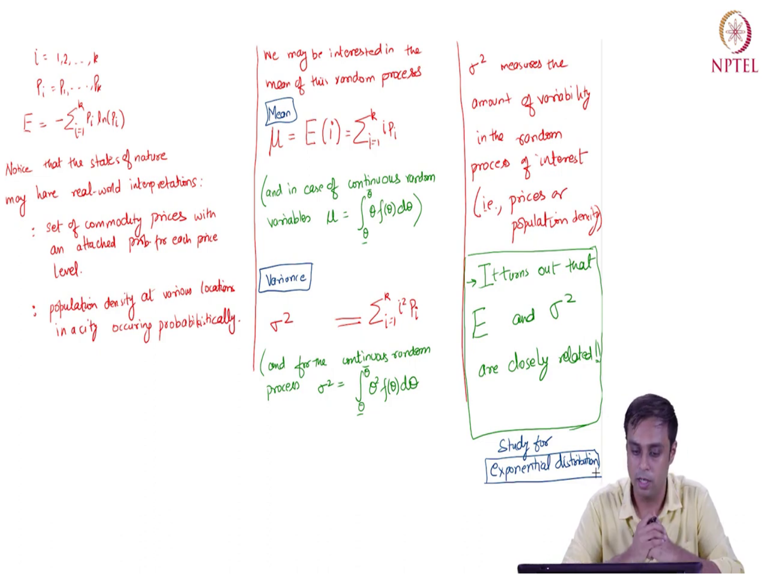
So, now I have a formal definition of entropy for the case when we are working with a continuous you know states of nature we are working with a set of continuous states of nature. So, this is an aside we will still sort of you know work with this formal setting that we began with. So, we are working with suppose there are $k$ states of nature, each state given by a index $i$, $i$ goes from 1 to $k$ right.

And for each state we have a probability you know of occurrence of each of that state $i$ given by $Pi$ and for that then we define the entropy as minus summation i equals 1 to $k$ $P_i$ ln $P_i$ ok.

$$E = -\sum_{i=1}^{k} P_i ln(P_i)$$

So, with that understanding let us come back you know to our discrete case.

(Refer Slide Time: 11:03)




So, first of all when I have this idea that $i$ could be 1, 2, $k$, $P_i$ could be $P_1$ to $P_k$ and I have an entropy measure which is minus summation $i$ equals 1 to $k$ $P_i$ ln $P_i$ right.

$$E = -\sum_{i=1}^{k} P_i ln(P_i)$$

Notice, that the states of nature states of nature may have real world interpretations right. So, it is not necessarily a mathematical entity these could be real world entities as well right. So, what could be those examples I mean. So, $i$'s could be a set of commodity prices right. So, you could have a commodity let us say let us say gold or you could have a commodity like wheat or rice and for that commodity you have $k$ possible price levels that appear in the real world and each price level $i$ can appear with the probability $P_i$ right.

So, we can adapt this abstract mathematical notation to a real world understanding of what might $i$ be you know representing right. So, we will have set of $k$ commodity prices with an attached probability for each price level. These may you know as well, you know as a second example represent population density, population density you know at various locations in a city. So, I have kept the city to be constant.

Let us say we talk about national capital territory of Delhi or we can talk about Mumbai, we can talk about Chennai, talk about Kolkata or any other city that you may be interested in.

Now, the population density can be considered as a random variable, you could have $k$ different possibilities of population densities for that given city which can be sort of you know where do these $k$ different possibilities come from. Well, it depends what location are you know drawing this population density from, if you are drawing it from a you know less part sparsely populated area.

Let us say near the airport you know an international airport you might not have that much population density, but if you go to the city center you might have very high population density right. So, all these population densities you know may occur with attached probability right. So, they can occur let us say probabilistically it could be probabilistic ok.

Now, in that case, in that case right, we may be interested, may be interested in you know the mean of this random process right. So, you know mean is the first moment of a random process, if I have a range of prices for gold occurring with different probabilities, I might want to know what is the average price of gold that I can expect right or like I might be interested in the second moment you know what is the variance you know can I give a measure of variance to the price of gold.

If I am comparing two different commodities, you know I might want to know which commodity has higher variability in prices before I for example, I am trying to you know invest in those commodities right. So, in that case you know the mean is given by mu which is the first moment.

So, that would be expectation of $i$ which is given as summation $i$ equals 1 to $k$, $i$ times $P_i$ right:

$$\mu = E(i) = \sum_{i=1}^{k} iP_i$$

and if we and in case of you know use a different pen and in case of continuous random variable *mu* will be integration integral of theta *f* of theta t theta where the integration is done from the lower bound to the upper bound of theta right.

$$\mu = E(i) = \int_{\theta_{lower}}^{\theta} \theta f(\theta) d(\theta)$$

(Note: $\theta_{lower}$ in the above equation implies the lowest value of θ, in the slide θ "lower bar".)

So, this is a standard deviation of mean right, this is something we are aware from the basic statistics you know a exposure right. The second moment that you may be interested in as I talked about in a min a minute ago is called as the variance. So, variance sigma squared is the second moment. So, it is given by expectation of $i$ squared which is then given as $i$ equals 1 to $k$ $i$ squared $P_i$ right.

$$\sigma^2 = Var(i) = \sum_{i=1}^{k} i^2 P_i$$

And for the continuous case for the corresponding you know continuous case and in case for the continuous random process, we can say sigma squared is nothing but integration theta lower bar theta upper bar theta squared *f* of theta d theta ok.

$$\sigma^2 = Var(i) = \int_{\theta_{lower}}^{\theta} \theta^2 f(\theta) d(\theta)$$

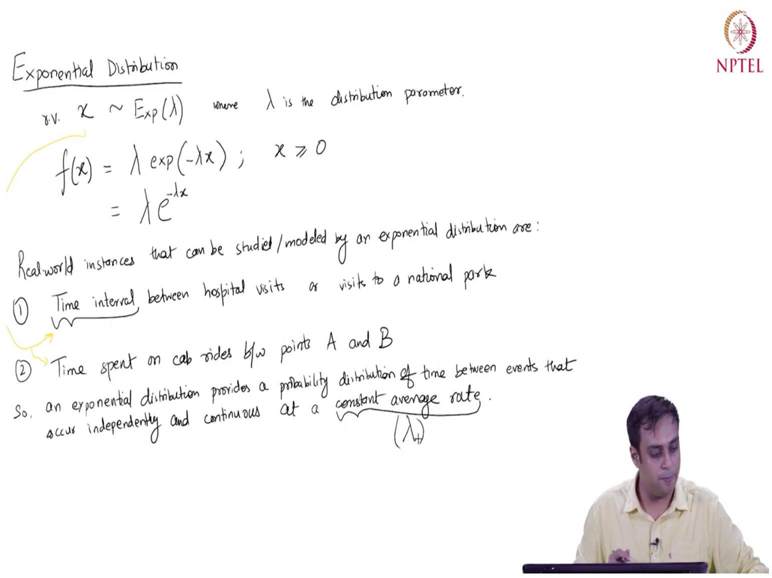
(Note: $\theta_{lower}$ in the above equation implies the lowest value of θ, in the slide θ "lower bar".)

Now, ok. So, ok just rely on the definition here ok. Having understood that now we know that sigma squared measures sigma squared measures the amount or extent of variability amount or extent of variability in the random process of interest ok. And of course, you know random process of interest could be these examples like prices or you know population density, whatever you deem to be you know interesting for your own research depends on the analyst ok.

Now, the thing is that the interesting thing is that it turns out it turns out that E entropy and variance sigma squared are closely related. So, what I am claiming the claim that I am giving you here is that, entropy is an alternative measure of sigma squared if we speak mathematically and you know more substantially you know entropy provide us provides us a measure of variability of a random process just like variance would do that right.

And we will see this. So, we will see how this key information piece plays out for a for an example distribution. So, I am going to go on to you know study this study for exponential distribution in the next slide ok. So, let us talk about exponential distribution now and then figure out how you know sigma squared and $E$ might be closely related ok.

(Refer Slide Time: 19:49)




So, as a next step we are studying exponential distribution. So, what is exponential distribution? So, we have a random variable $x$ right, we always begin with a random variable *x* and we say that this *x* is distributed with exp(lambda), what is lambda? Lambda is the distribution parameter.

$$x \sim e^{\lambda}$$

So, lambda where lambda is the distribution parameter ok and the density function *f(x)* which is the frequency or the probability of occurrence of any given value of *x*, which is distributed by an exponential distribution is given as lambda times exponential minus lambda *x*, such that; *x* is greater than or equal to 0 and I have already said that lambda is the distribution parameter. I mean alternatively I could have written this as lambda e to the power minus lambda x right.

$$f(x) \; = \; \lambda e^{-\lambda x} \quad s.t \;\; x \geq 0$$

You are aware that these are we are aware that these are equivalent expressions mathematically. So, the real world instances so, how what good is exponential distribution, why do we even care about it right. So, the real world instances that can be studied that can be studied or modeled by an exponential distribution are entities like you know time interval between hospital visits ok or you know visits to you know to a national park you know or visits to a national park ok, or you know the visits for vacation right.

So, how frequently an individual takes vacation, an individual needs to you know visit a hospital for getting some treatment or how frequently do they visit a given national park. So, if you live around New Delhi we have Jim Corbett National Park and we can be interested in modeling what is the time interval between two, you know visits between two visits to the Jim Corbett National Park right.

That gives us an understanding of how frequently an individual is undertaking recreation and then moreover you know whenever they take recreation undertaken recreation what is the time interval you know with which they visit a particular you know national park right. So, that to be able to model this time interval between visits for a particular kind or two events is you know is modeled by the exponential distribution.

So, *x* per say, *x* per say this random variable *x* is nothing but the you know the time interval between two events of interest right. We can have different interpretations here right. So, it could also be time spent on cab rides between points A and B right. So, you can have two interest points A and B, you can have it from let us say the railway station to your work place, from the airport to the to the city center, something like that and during the day you have different during different time you have different parameters like heat, pollution, traffic congestion or availability of rides right and so on and so forth.

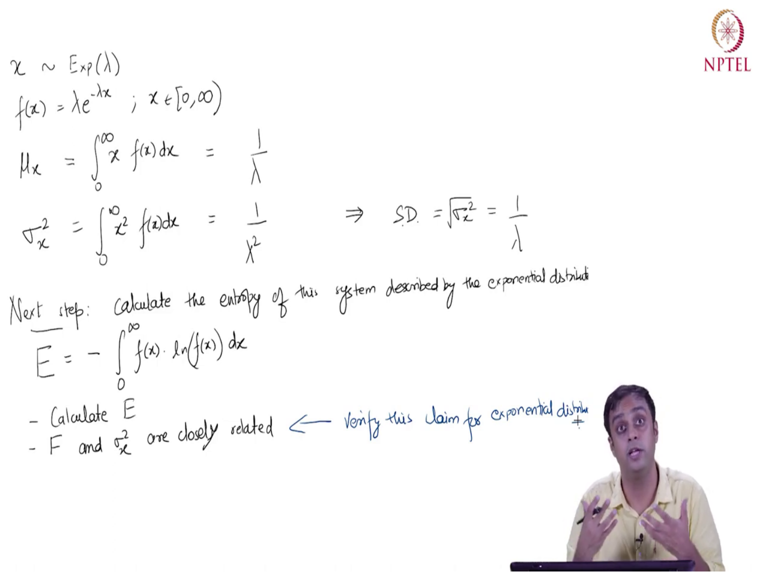
Which can you know basically deliver different time periods that I will take from me to go from the city cities airport to the city center right. How do we model this time spent you. So, these the time spent can take a range of values and each value that this time spent takes has a given probability distribution, in order to model this we have this exponential distribution you know device at our disposal right.

So, to define finally, so, an exponential distribution an exponential distribution provides a probability distribution of time between events that occur independently. So, two events like for example, rides from the airport to the city center they have to be independent right. So, it is not it is not that you know I go to the airport and then somehow my second right is dependent on what I did in my first right. So, the two rides are independent of each other right and these are continuous events.

So, I must be able to you know observe them happening frequently right, it is not that they only happen once. So, if it happens only once in a while or it only happens once forever then you know exponential distribution will not be able to model it and they happen at a constant average rate. So, there is some rate at which these rides are taken or the hospital wages visits are taken or visits to national parks occur and this constant average rate is modeled by the parameter lambda of our distribution.

So, this is a measure for the average rate at which this happens is given by is given by you know lambda ok.

(Refer Slide Time: 26:39)




So, I mean what do we have eventually I mean we have *x* is distributed as exponential lambda and; that means, that *f(x)* is lambda e to the power minus lambda x, such that; x is between 0 to infinity bound open.

$$f(x) = \lambda e^{-\lambda x} \quad s.t \quad x \in [0, \infty)$$

So, *x* is greater than or equal to 0 and then we can write mu x which is nothing but the first moment of this distribution which is nothing the mean of x.

$$\mu = \int_{0}^{\infty} x f(x) d(x) = \frac{1}{\lambda}$$

So, if you have time spent on a cab right from airport to city center, then mu x will be interpreted as the average time that is spent on this you know on this ride from the airport to the city center. This is given as 0 to infinity that is the that is the range of *xf(x)dx* this value turns out to be 1 over lambda. Similarly, I can figure out the second moment 0 to infinity $x^2 f(x)dx$ is given as 1 over lambda squared.

$$\sigma^2 = \int_0^{\infty} x^2 f(x) d(x) = \frac{1}{\lambda^2}$$

And you know that will imply that the standard deviation of *x* is nothing but the square root of variance that is 1 over lambda.

$$SD = \sqrt{\sigma^2} = \frac{1}{\lambda}$$

So, the one of the properties or exponential distribution is that the mean is exactly equal to you know standard deviation and they are both inverse of the average rate at which this event is happening. So, the next step that we will do now, next step that we will undertake is to calculate the entropy of you know of this system described by the exponential distribution ok. So, now the entropy we know is negative of integration from 0 to infinity which is the range of *x*, *f(x)* times *ln(f(x))dx* ok.

$$E = -\int_0^{\infty} f(x) ln(f(x)) d(x)$$

So, as a next step we want to first calculate this entropy. So, calculate *E* and after that we will be showing we want to show at you know *E* and sigma squared x are closely related right. So, what we are trying to do is, we are trying to verify this claim, verify this claim, for exponential distribution ok. So, what we will do is, we will stop here in this lecture and we will go to the next lecture and we will start from calculation of entropy and showing or verifying the claim that entropy is indeed closely related to sigma squared *x*.

Thank you.