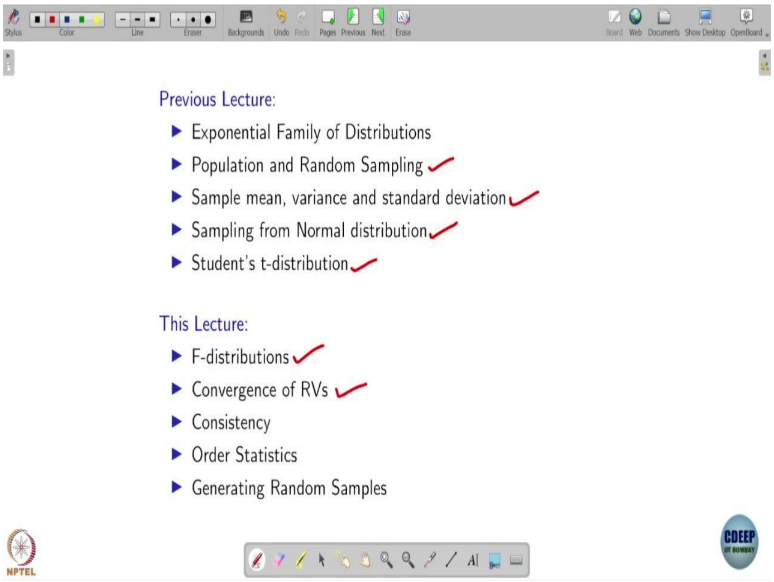
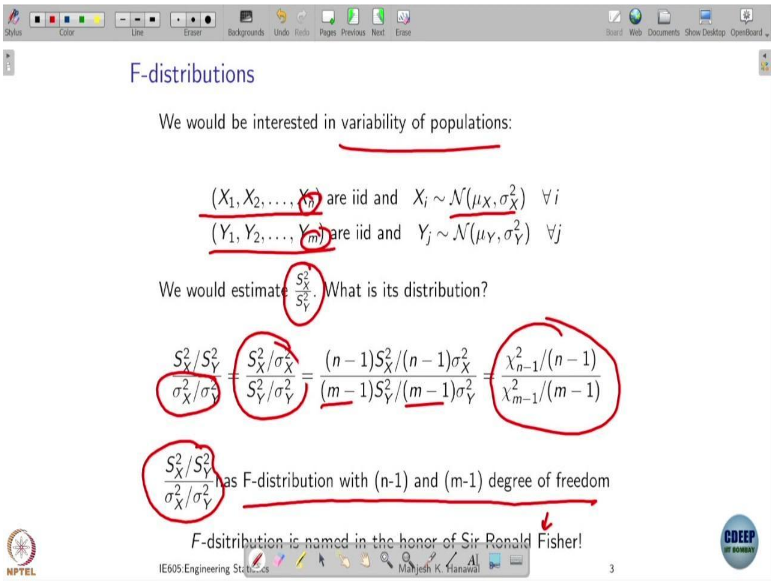
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operational Research
Indian Institute of Technology, Bombay
Lecture - 24
F-distribution and its properties

(Refer Slide Time: 0:21)



So, we just talked about population and random sampling, sample mean, sample variance, standard deviations, sample standard deviation, sampling, we are now currently focusing on sampling from normal distribution. And we just talked about student t distribution. And now we are going to talk about some more distribution today called F-distribution. And we will talk a little bit about convergence of random variables is a convergence of random variables done in IE 621, convergence and probability, convergence and distribution, convergence and expectation, nothing of that sort. Let us see how much we can do that.

(Refer Slide Time: 1:05)



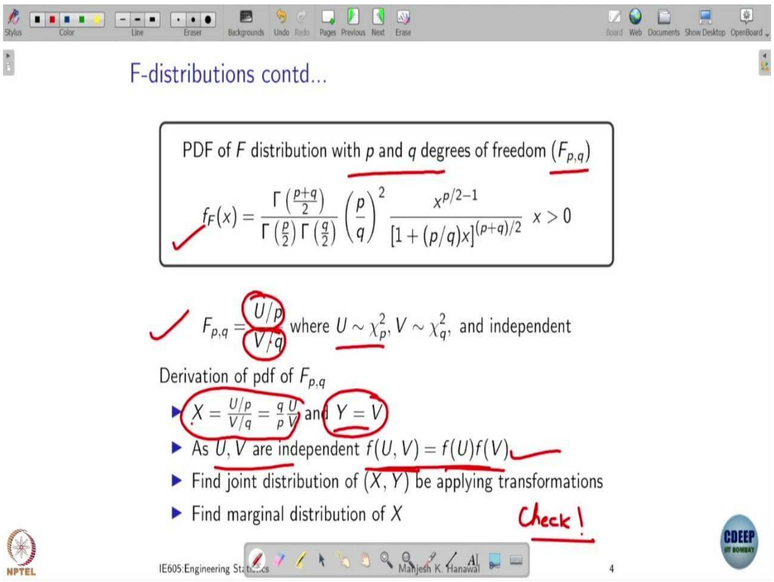
Now let us say we have actually two population denoted as normal with mean mu and variance sigma X square. Another one normal with mean mu Y and variance sigma Y squared. These are two different populations. From one population, I have the samples i.i.d. And another population I have these random samples. Now suppose let us say I am interested in the ratio of their sample variances. So, what are these capture? I want to basically look into the variability of the populations. That is how does a variance of one population compare with that or the other. And since I do not know the variance, I would replace that by their sample variances and then look at their ratios.

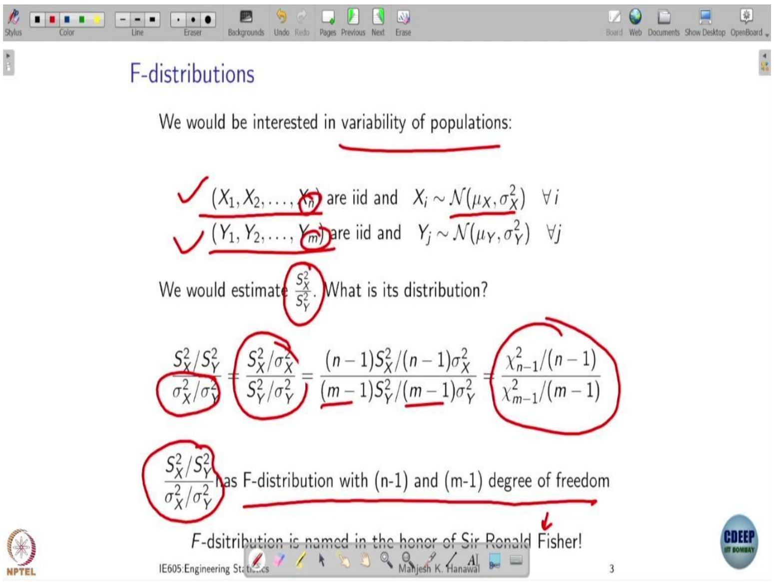
Now let us see how to compute a distribution. I will ask you, find out the distribution of this ratio of the sample variances, how to go about it? Now, you want to again, appeal to the Gaussian distribution properties. Let us see how to do that. This XX squared, XY square I will divide it by this quantity, the actual variances and if I simplify this, I will get this quantity. And the numerator I am multiplying and dividing by n minus 1, denominator, I am multiplying and dividing by m minus 1. But by the way, notice that the population this first populations have n samples and a second population have m samples. n and m are not the same.

Now, if you focus on the numerator here this we know it to be chi square distribution with n minus 1 degrees of freedom. That is what we said. And the denominator is again a chi square distribution with the n minus 1 degrees of freedom. And this quantity, we are going to call this the distribution of this we are going to call it as F-distribution with n minus 1 and m minus 1 degrees of freedom or alternatively, the ratio of chi square distribution with n minus degrees of freedom and m minus 1 degrees of freedom. We are going to call it as F-distribution with n minus 1 and m minus 1 degrees of freedom. So, we have another distribution here.

Again, this distribution is kind of popularized by this Ronald Fisher who is another famous statistician and F-distribution comes from his first letter F. What we are basically going through is different statistics we are interested in that, sample mean, sample variance, they themselves are random variables. And now, we have tried to basically look into what kind of distributions they will have and that has led us to find out this t-distribution and F-distributions.

(Refer Slide Time: 5:17)





And again, you can go back and compute the actual PDF of this F-distribution and we will end up with this PDF function which is PDF with p and q degrees of freedom and we are going to denote an F-distribution with p and q degrees of freedom like this. Now, the question is how this PDF comes? Again you can go back to your classical method of finding the joint distributions from the known distributions. Now, if you notice F-distribution is nothing but ratio of two chi square distribution is the numerator this chi square distribution in the numerator and the chi square distribution in the denominator, are they independent of each other?

Student: Yes.

Professor: Why?

Student: Population is different.

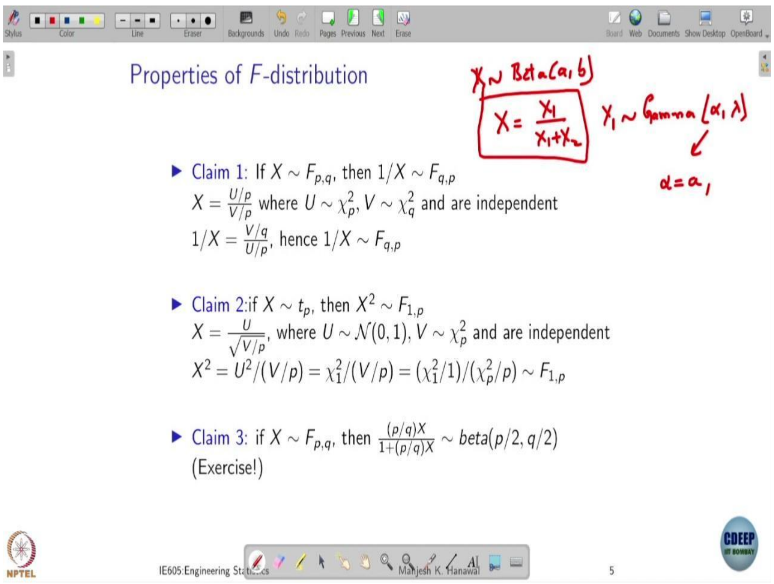
Professor: They are coming from different populations which are assumed to be independent because the numerator is coming from this population and the denominator is coming from this top and they are independent. So, now, what I will do is I have now represented this F p, q has nothing but U by p, V by q, where U is chi square distributed with p degrees of freedom and V is chi square distribution with q degrees of freedom and are appropriately divided by p and q. So, am I right in saying that? F-distribution and p and q degrees of freedom can be written like this, if that is the case. And you know that U and V are independent, then things are easy for us.

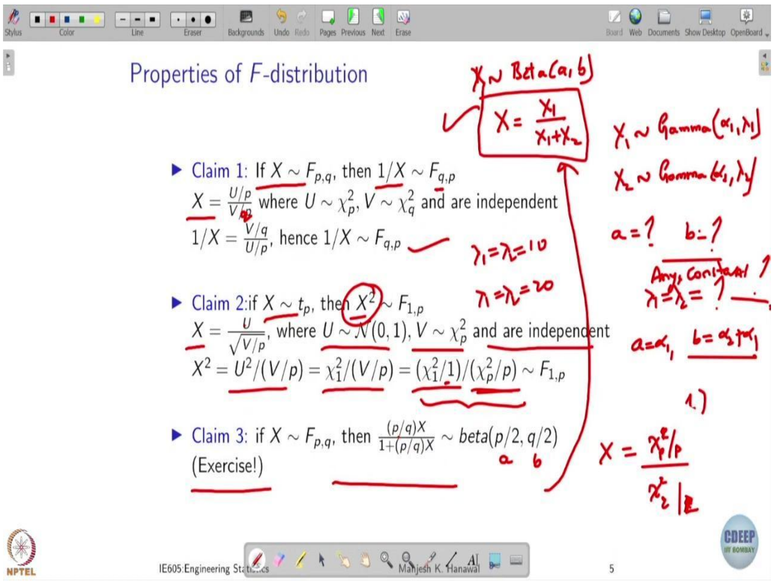
We can again go back and appeal to your computation of joint distribution functions of joint distributions. So, now to find out what we will do is again, I am going to define one random variable like this X, which is of my interest. And I am going to denote another random variable. Whatever of my interest, I am going to call it as X and I have denoted another random variable V. And are X and Y are independent? Why X and Y has to independent? Because this depends on U, V, X also depends on V. So, X and Y need not be independent. But we know that U and V are independent. And we know that is why joint distribution of U and V is nothing but product of each one of them.

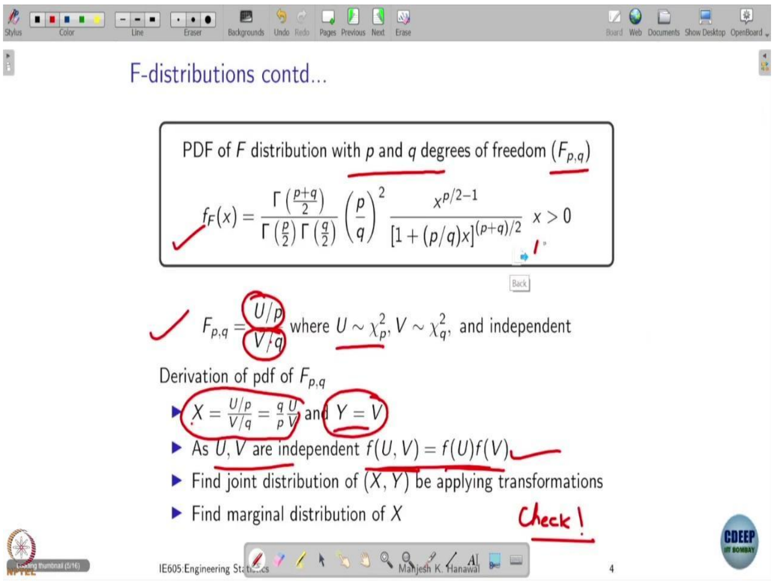
And we know do we know the CDF of U? U means what, chi square distribution. We know that chi square distributions PDF. So, we can write the distribution of joint distribution of U and V. Now that is it. If you know this distribution, now, you can find out the joint distribution of X and Y. Is not it?

Again what you have to do is use your standard Jacobian method and then find a marginal of X and I will again leave it to you to complete all these standard steps, that is again nothing you have to do, try yourself, works out better otherwise just refer to the books that I will post it in which all the details calculations are given. But at least try yourself if you are able to reproduce that result.

(Refer Slide Time: 9:38)







Now see, like we have enlarge our scope of many distributions. It is just like now going from Bernoulli binomial Poisson, Gaussian exponential. Now, we talked about beta distributions, gamma distributions. Then we talked about t -distributions and F-distributions, but as you see gamma and beta distributions they are somehow related to my basic distributions. What was the relation between gamma distributions and exponential distribution, was there any relation?

Student: Gamma is sum of exponential distributions.

Professor: Gamma distribution was nothing but a summation of n independent exponential distributions with parameter lambda. So, gamma and lambda was like that. And what was the relation between beta distribution and other distributions?

Student: (()) (10:48)

Professor: (()) (10:48) other than that anything.

Student: Beta 1 1 is (()) (10:51).

Professor: That is right beta 1 1 is.

Student: Ratio of gamma (()) (10:57)

Professor: Ratio of gamma distribution is beta, did you say that?

Student: Beta distribution is bigger this is the exponential distribution wise, sampling minus beta normally. X is the completed this.

Professor: Now, this beta distribution is the ratio of the gamma distribution, can you tell like if I say beta distribution with parameter a and b, now, can you express in terms of the gamma distribution? Let us say my X is beta a, b, now, you want to write X as X1 by X2.

Student: X1 by X1 plus X2.

Professor: X1 plus X1 by X2, now, what is X1?

Student: Gamma.

Professor: Gamma, what parameters, what is n? You have to know tell me in terms of a and b.

Student: a comma lambda

Professor: What is lambda now? You have to tell me everything now, you are trying to write express beta a, b in terms of gamma distribution. Only thing you have is a, b.

Student: Sir b and b would be a (()) (12:31) of gamma distribution and gamma distribution should have a common shape.

Professor: Tell me that.

Student: X 1 (()) (12:43)

Professor: Now, n, what is n? So, gamma has to be n n lambda so, we said gamma is now alpha and lambda we said. Tell me how is alpha related to a and b?

Student: Alpha equal to (()) (13:01), alpha equal to (()) (13:03).

Professor: Alpha equal to.

Student: a power X1 and X2 of b minus (()) (13:12).

Professor: Let us take this is X1, alpha equals to a and what was lambda?

Student: Lambda is same for (()) (13:20).

Professor: But what is it? Any lambda, how can it be any lambda. It has to be dependent on a and b. I am not sure. You check this. Your claim is if I give you two gamma distributions, you are able to get a beta distribution by writing those gamma distribution in this format. Either if I tell you let us say I will tell you X1 is gamma with parameter let us say alpha 1 and lambda 1 and I will say X2 to be gamma alpha 2 lambda 2 and now your claim is X is beta distributed with a, b, now tell me what is a and what is b?

Student: Sir, here lambda 1 and lambda 2 should be equal.

Professor: Lambda 1 and lambda 2 is equal to what?

Student: Sir, some constant value which is greater than zero.

Professor: That has to be that is fine. And you want to set lambda 1 lambda 2 be the same. What about alpha 1 and alpha 2.

Student: Sir, a is equal to alpha 1 and b is equals to alpha minus alpha 1.

Professor: And lambda 1, lambda 2?

Student: Alpha 2 (()) (14:56).

Professor: No, you cannot be just greater, greater than zero means, I can take a 10, 20, 30, 40 like you said a equals to alpha 1, you fixed it. I cannot take anything now. When you say b equals to this you fixed it. Now fixed me lambda 1 and lambda 2.

Student: Sir, b is equals to alpha 1 plus alpha 2.

Professor: b is equals to alpha 1 plus alpha 2. Lambda?

Student: Lambda is (()) (15:34).

Professor: Cannot be then let us say I take a lambda 1 equals to lambda 2 equals to 10 then this will give me some value and if I take lambda 1 equals to lambda 2 equals to 20 cannot be same. It will give me something else.

Student: It can may be when we derives the median (()) (16:03).

Professor: Just you throw up or you want to invert, then tell me.

Student: Like the both lambda is same (()) (16:13).

Professor: Then what is its value?

Student: Any constant.

Professor: Any constant will do?

Student: Greater than this.

Professor: I do not know about any constant, then it is an exercise for all of you. If you do not get to catch these two people who are making this claim. Their claim is you should take two distributions gamma 1 and gamma 2 with parameters alpha 1 lambda 1, alpha 2 lambda 2 where alpha 1 is same as alpha 2 and any value then you will get a beta distribution when you express like this. X is like a beta distribution where a equals to alpha 1 b equals to alpha 1 plus alpha that is their claim. So, I may ask you to prove or disprove this. So, verify this.

Now, let us see some simple properties of this F-distribution. Suppose you have X to be given to be F-distribution with degrees of freedom p and q and you may be interested in taking the reciprocal of that 1 by X then my claim is it will have a F-distribution with degrees of freedom q and p. So, p and q has become q and p now, why is that that? That observation is obvious because let us say you know that X can be written as U by p, V by q this should have been q where U is chi square distribution p degrees of freedom and U is chi square distribution with q degrees of freedom.

Now, what is 1 by X in this case? 1 by X is now just. Now, the numerator is chi square distribution q degrees of freedom and the denominator is chi square distribution with p degrees of freedom then by definition this should be F q, p just again verify. Now suppose what is the connection between student t-distribution and F-distribution? Now, suppose x is student t-distribution with p degrees of freedom then it so happens that the square of that is

F-distribution with parameters 1 and p. So, why is that? Again you can verify this pretty straightforward.

Let us take X to be student t-distribution and if it is a student t-distribution I know that that could be represented as a ratio of two random variables U and V actually U and square root of V by p where U is normally distributed and V is chi square distribution with p degrees of freedom that we have discussed. And we also said they are independent. This U and V are independent. Now, if you just take the square of this, you just take the square now this is a U squared divided by V by p.

Now, we have already it discussed that if U is normal then what is the distribution of U square? Chi squared with what degrees of freedom? 1 degree of freedom that is what we have discussed some time back. This is going to be 1 degrees of freedom. Now what I have done, now if you carefully look into this now, I have this like chi square numerator I can divide it by 1 nothing changes and the denominator is now this V is chi square distribution with p degrees of freedom.

Now notice that now I am able to write this as ratio of 2 chi square distribution where the numerator is of 1 degrees of freedom and denominator is of p degrees of freedom and appropriately normalized. And by definition what is this? F 1 p because the numerator has 1 degrees of freedom, denominator has p degrees of freedom, so you people are talking about this third example.

Student: Today (()) (21:22).

Professor: Let us look into that. Suppose let us say X is F-distribution with parameters p and q if I write like this X divided by 1 by X but X is being multiplied by 1 plus q X, so what is this actually? Can I write it in this form? Let us take it in this form and now we are saying that this is nothing but beta distribution with parameters p by 2 and q by 2, now let us try to solve your problem, can we use this result suppose this is true?

Student: You can take X as chi square be with that.

Professor: You are talking about this X, X is.

Student: Chi square

Professor: You want to say X is what, this is my right this is F p, q. Now, how is this I have two ratio take like this how this becomes beta and this is not even gamma, I do not see gamma coming. Gamma is not directly related to F p q,

Student: Gamma is not directly related but the chi square is related to gamma.

Professor: Chi square is related to gamma that is right. We are now try to fill in the gap like how now can you do this using this?

Student: Lambda 1 equals to lambda 2 (()) (23:50).

Professor: So, here this is my a and this is b. According to you, it is not clear what you are saying, it does not come? Anyway, you this requires some computation this is not, so the claim 1 and 2 are straightforward. We can just apply the definition but this requires little more thinking, how to derive this relation between F-distributions and beta distribution. So, work out this as an exercise. So, and then try to see that if you can connect with this result if and what if it is indeed correct. Let us stop here today.