Predictive Analytics - Regression and Classification

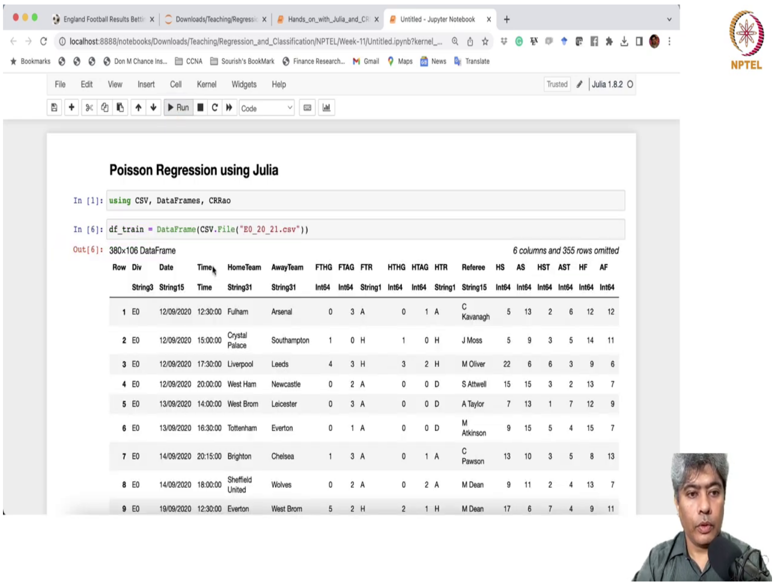
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 61
Hands on with Julia_Bayesian Poisson Regression with Horse Shoe English Prior_League Data

Hi all, in this video, I am going to do Poisson regression or count regression using Julia. And in this video, I am going to show how we can use the UK football data or English prior league data, because we have seen in the previous videos that English prior league data, the home teams code the number of goal. And if we want to model number of goal, it will be either through a Poisson regression or negative binomial regression.

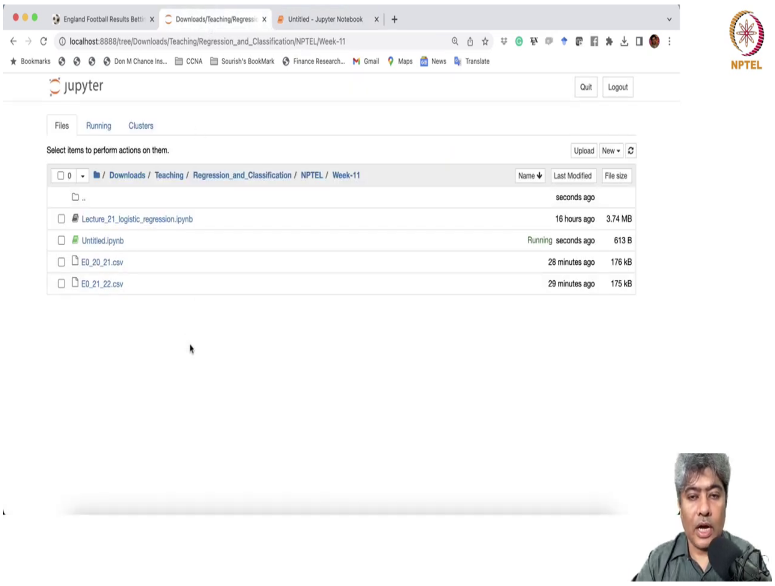
So, what I am going to do in this video, I am going to show you how to do Poisson regression using Julia, ok.

(Refer Slide Time: 01:06)



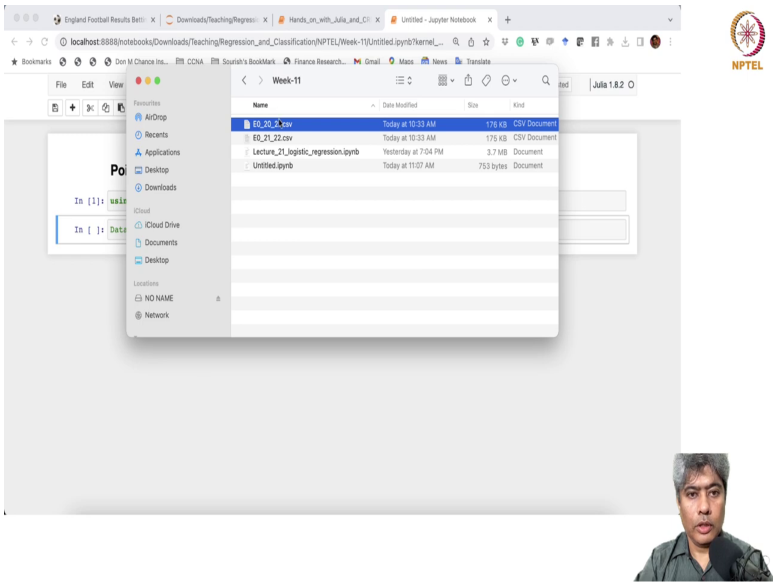
So, first in my Jupyter, I am going to start our Jupyter Notebook. So, first I will write give a title to my Notebook, say Poisson Regression using Julia, ok. And then the first thing I am going to do, I am going to call CSV DataFrames and CRRao is 3 for sure. Let me run this.

(Refer Slide Time: 01:50)



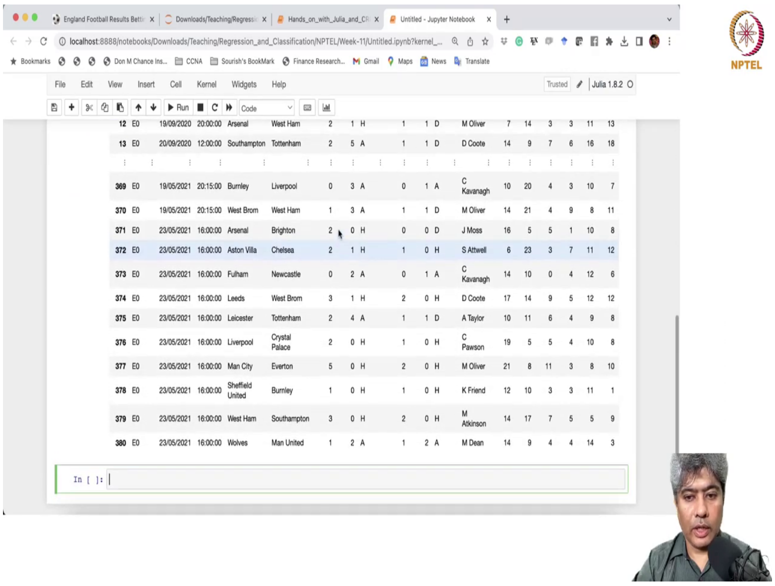
Now, in my the same folder where you are starting your Jupyter Notebook, I have downloaded 2021 data and 21 22 data. So, I am going to use 2021 data, ok.

(Refer Slide Time: 02:18)

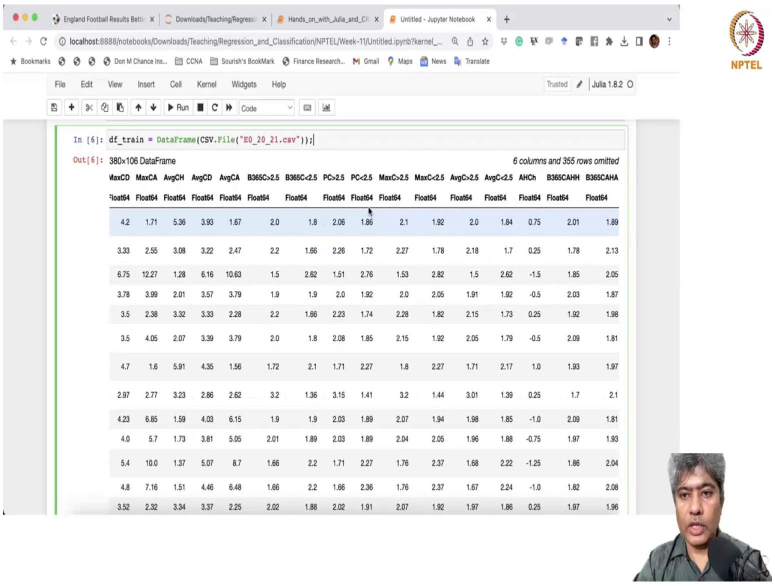


So, what I will do DataFrame CSV dot File quote unquote, I just need the name. So, file name, you keep the file name in the same folder. So, that will be helpful, ok.

(Refer Slide Time: 02:36)

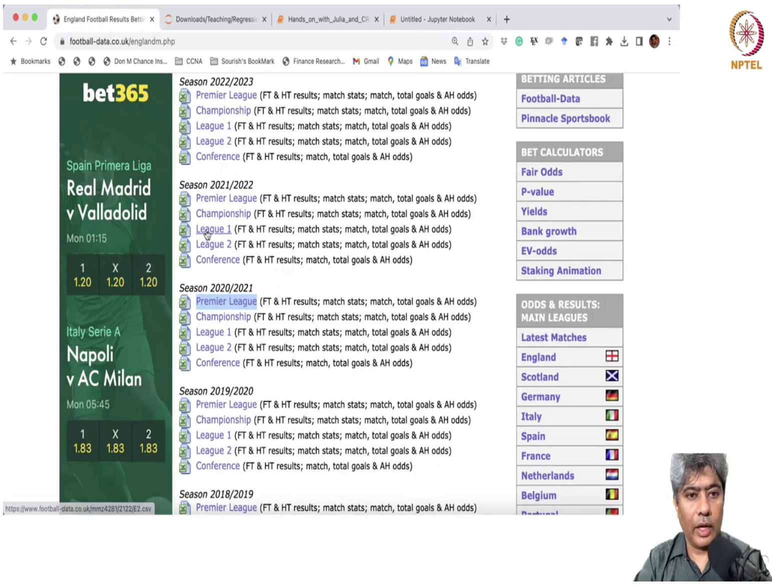


(Refer Slide Time: 02:50)

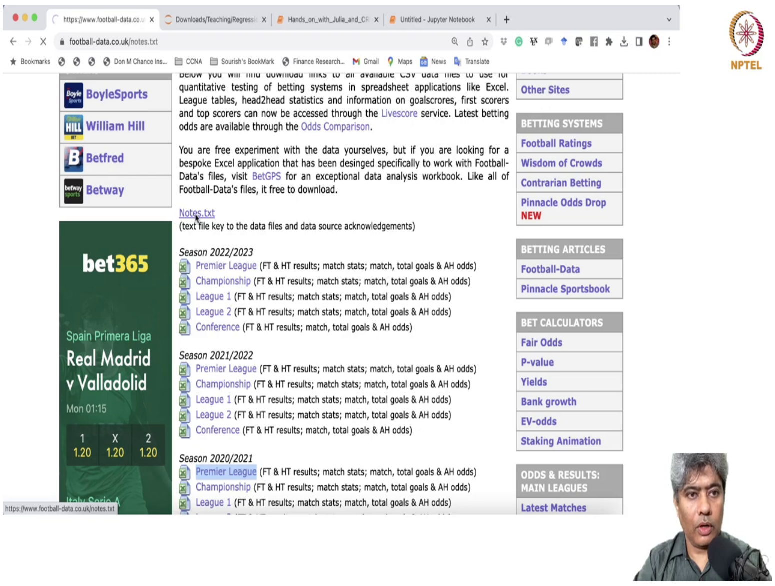


So, this gives me the data set that I am looking for. So, there are 380 gaming are being played, ok. And there are 106 columns are there. So, there is lots of columns are there, ok.

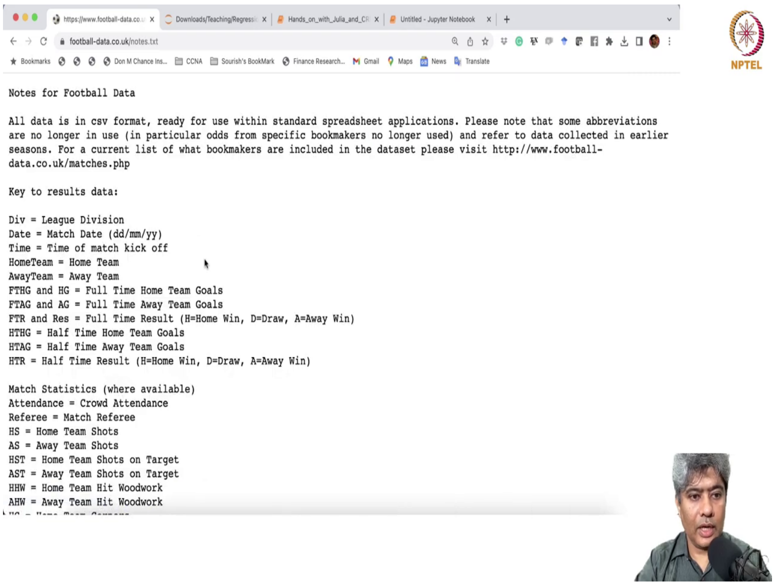
(Refer Slide Time: 03:04)



(Refer Slide Time: 03:06)

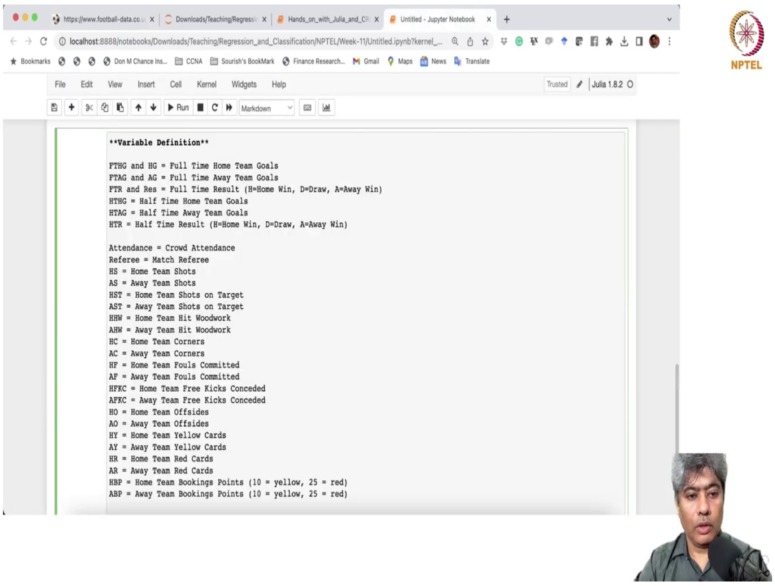


(Refer Slide Time: 03:07)



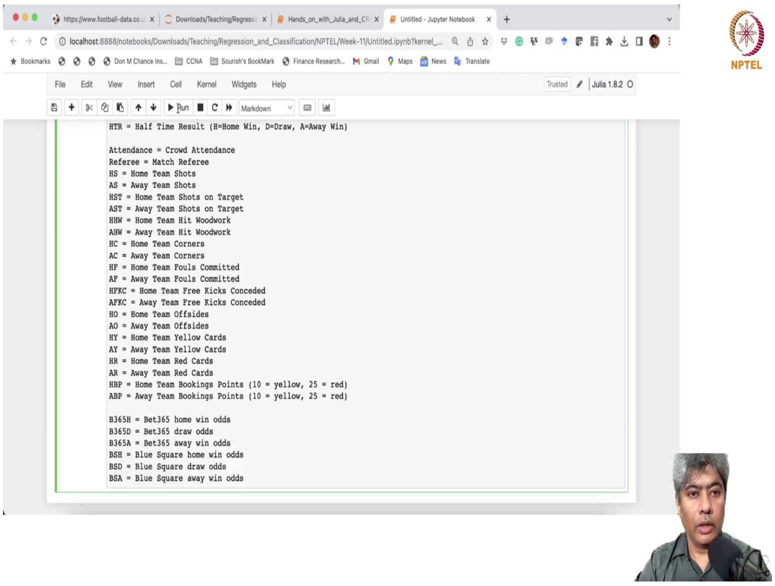
And now my what I am going to do, I want to model number of goals scored by. So, if I just go back there and you know, if I Note dot txt click on that. So, I want to come full-time home team goal that is FTHG. This is what I want to model, ok. So, what I want to model is let us the model is, ok.

(Refer Slide Time: 03:30)



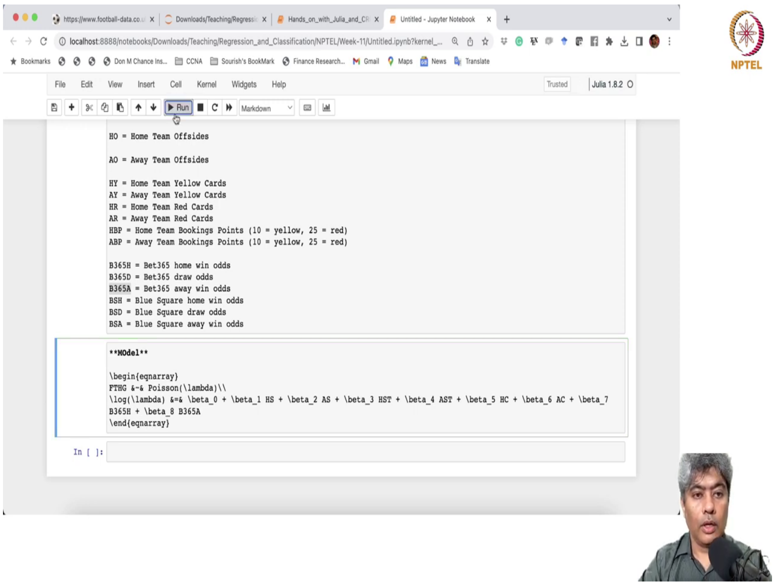
First is let me just at least write down few things FTHG. These are the basic definitions. Let me just copy it down, ok. These are the variable definition, ok. Variable Definition, ok. And then I will just copy few more things. Can there is all these information's so I am going to, ok.

(Refer Slide Time: 04:23)



And some of the betting statistics as well, ok. Opsee; alright so, let me just stop here. So, these are the, so, now, FTHG is the one of the variable here FTHG. This is the home teams code basically full-time. How many goals scored by the home team?

(Refer Slide Time: 05:04)

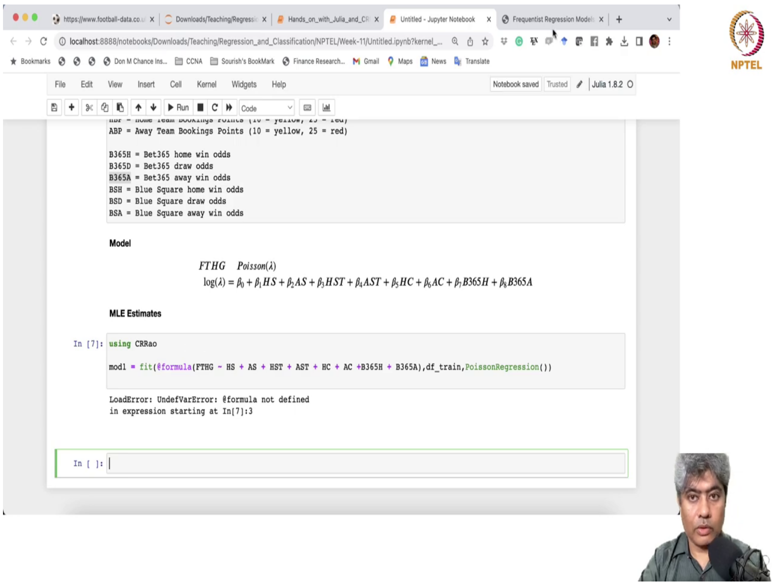


So, this is what the, what we want to model. And we want to model as a function of what? We want to model it as a function of maybe, you know, home team, how many shots were taken by the home team? And like HS. And how many shots were taken by away team? Ok. And then how many home team shots on target? Ok. And then away team shots on target. Then home team corners. How many corners were taken by home team away team corners?

And then maybe BETrix 65, the score by BETrix 365. And away teams score all these things we want to model. This is the model that we want to model. But the home team is. So, model this is my model. This is the model that we want to fit, ok. So, this is, so, effectively this will be like home Poisson, lambda Poisson lambda. And then log of lambda BDA lambda equals to beta naught plus beta 1 HS so, this way beta 2, beta 3, beta 4, beta 5, beta 6, beta 7 and beta 8.

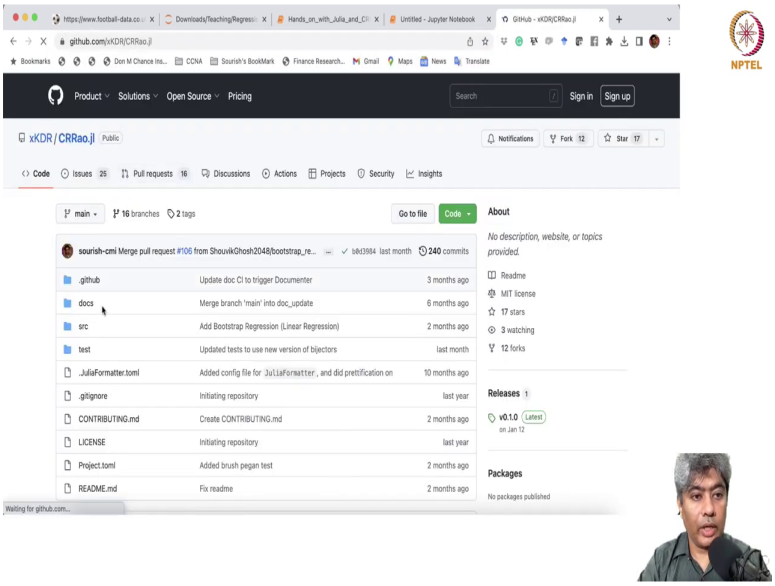
So, this is pretty much what we have, ok. So, let me not take dollar here. So, maybe, I will just put it like this. I will put it like this. Let me run it, ok. Yeah so, maybe I will just put, yeah. If I just do this, probably this is, we will make, ok. So, maybe I will just put begin equation array, e q n a r r a y and end eqnarray. And it will work.

(Refer Slide Time: 08:41)

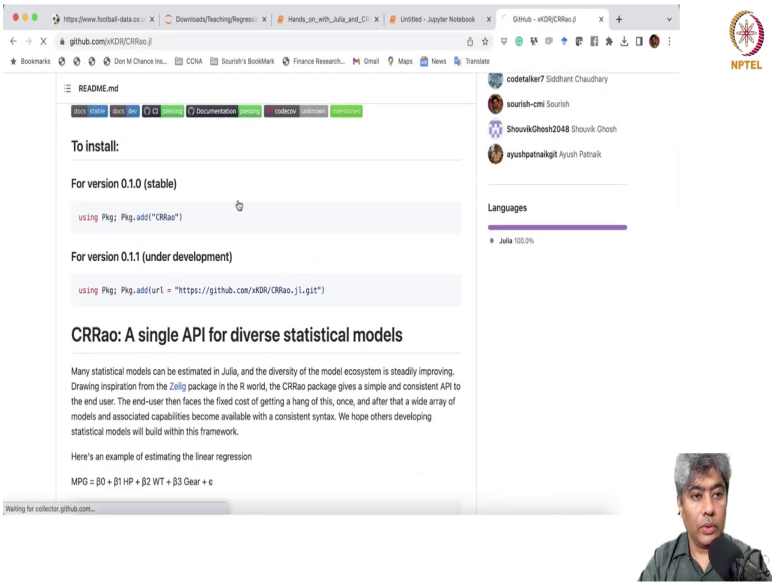


Yeah. So, this is what the, this is what the model I want to fit. And so, for that, I am going to call CRRao. And the first model that we are going to fit is fit, ok. At the rate formula, ok and first is FTHG, FTHG and then HS plus AS plus HST plus AST plus HC plus AC plus B365, opsee B365H plus B365A. Now, after that, I have to give the name of the data set df train, ok. And then I have to give, I have to say the class of the thing, but X.

(Refer Slide Time: 10:18)

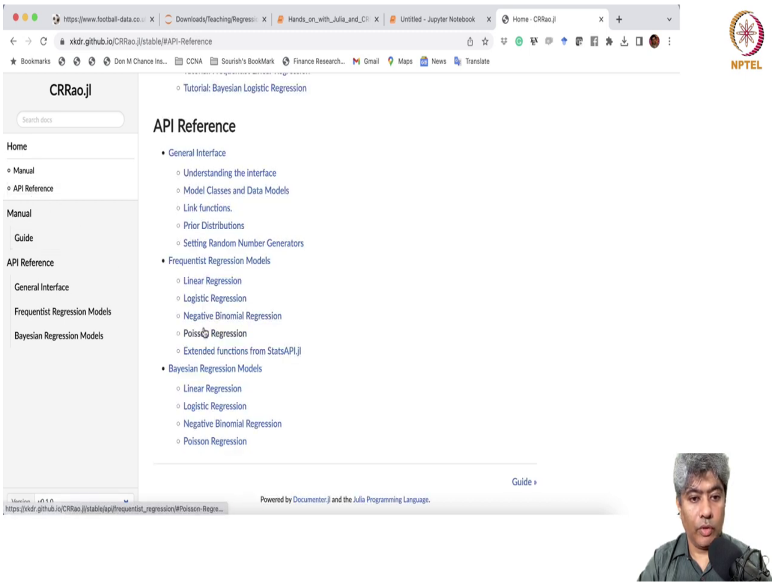


(Refer Slide Time: 10:22)

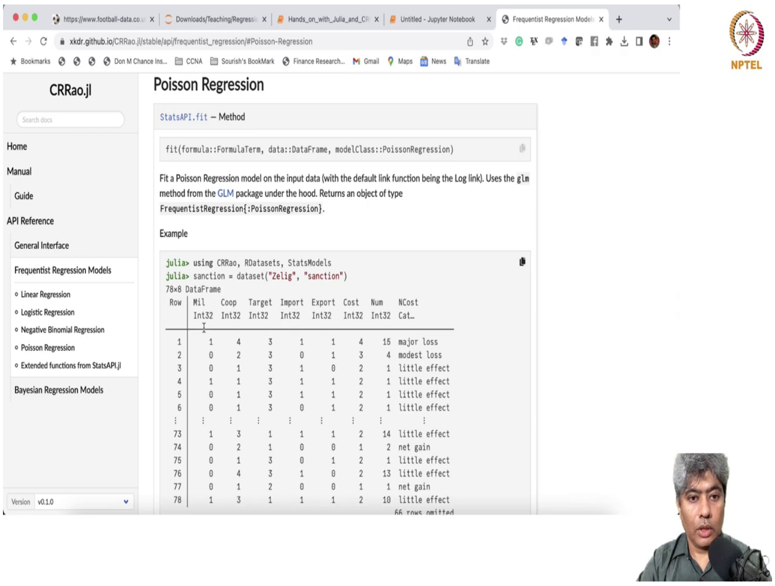


So, maybe I will just CRRao package. I will just go to the stable, doc package.

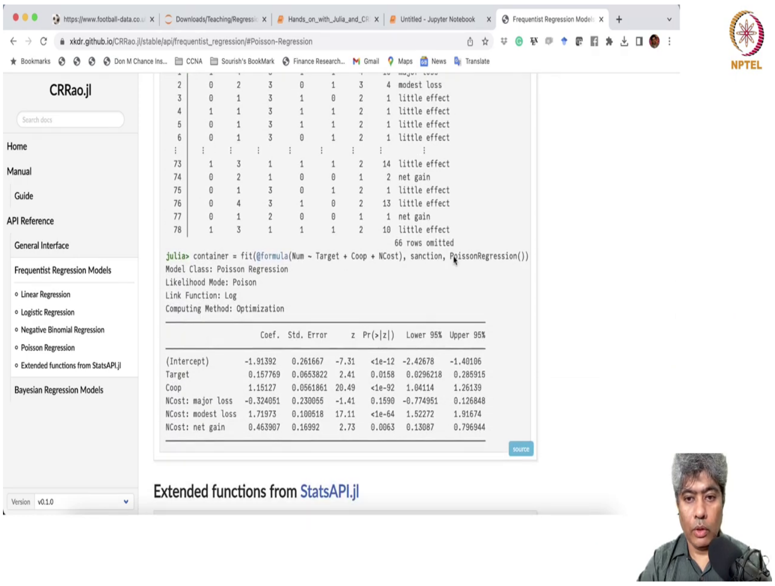
(Refer Slide Time: 10:32)



(Refer Slide Time: 10:35)



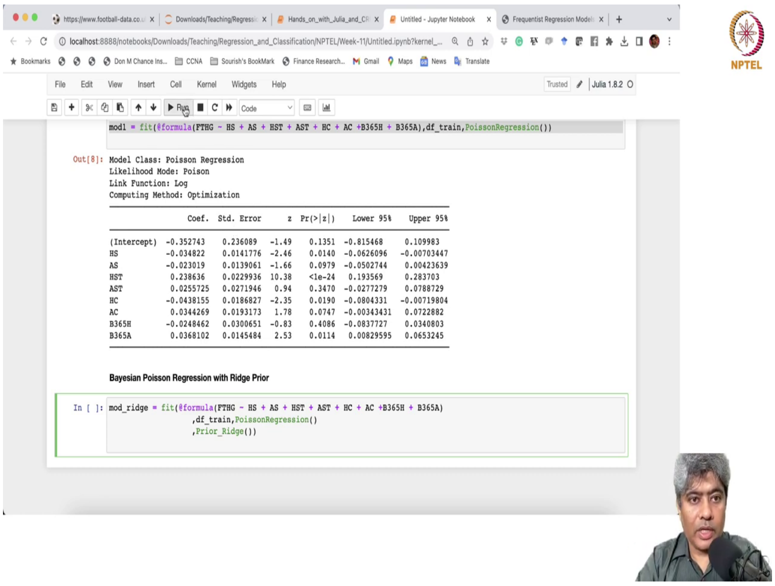
(Refer Slide Time: 10:38)



So, API reference Poisson regression, what I will do is what I have to do. Yeah. So, Poisson regression, that is what we have to do here. So, PoissonRegression. So, if I just, if you write POI and then tab, it will fill it up for itself. Now, if you just do this, this will fill a, set a fit a maximum likelihood estimate, ok.

Maximum likelihood estimate or MLE estimates, ok. Let me just put it up there. Yeah, and now let me just run this. Why it is not formula? Ok. So, I have to, maybe I have to call a StatsModels. Yeah StatModels, I need. Yeah so, yeah. And I think now it is fine, ok.

(Refer Slide Time: 11:50)

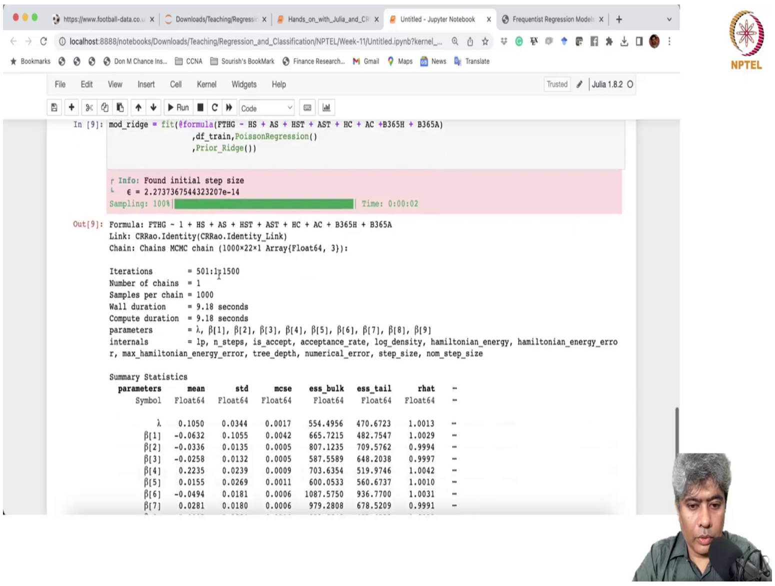


So, if I look into it, HST home, number of shot on target by home team is either going to have a very strong effect, positive effect. Obviously more shot you have on the target, the chance of having score is I. And then this is another home team, now shot and has a effect. And then this is also home team number of goal, number of corner by home team. And what is the weight 365s, hours odds for the away team? That also is the important role.

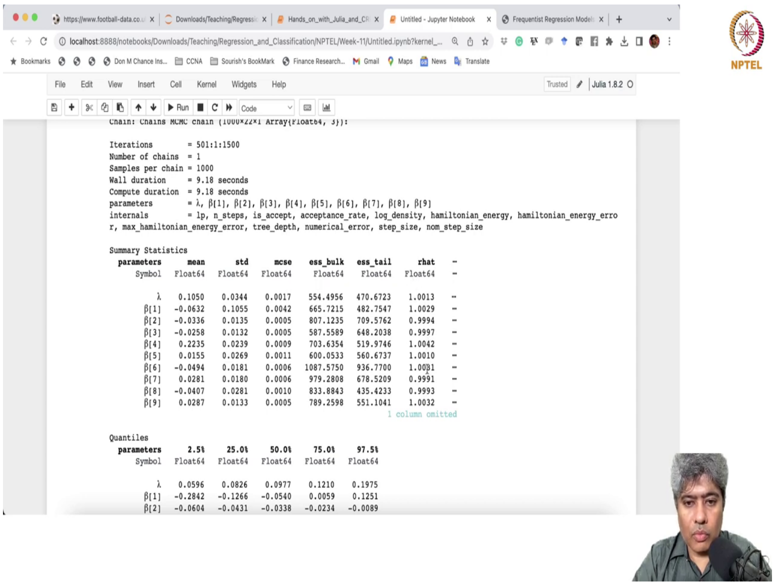
And you can see that this coefficient has a positive effect alright interesting phenomena that we are getting. Now, suppose we want to predict, suppose we want to fit Bayesian regression model, Bayesian Poisson regression model with, say ridge prior Bayesian Poisson let me Bayesian Poisson Regression with Ridge Prior, R i d g e, Ridge Prior. So, if I just run this. So, the if you just go there, all I have to do is essentially just copy this thing. And; obviously, now I want to change the name, maybe I will say ridge.

And after the Poisson regression, I would say Prior Ridge. And if you just Prior underscore R, and then if you just tab, it will take the rest of the thing, ok. So, you do not have to really worry about that. Let me just put it here. So, that you know it is kind of aligned. And you can see the entire model.

(Refer Slide Time: 14:19)

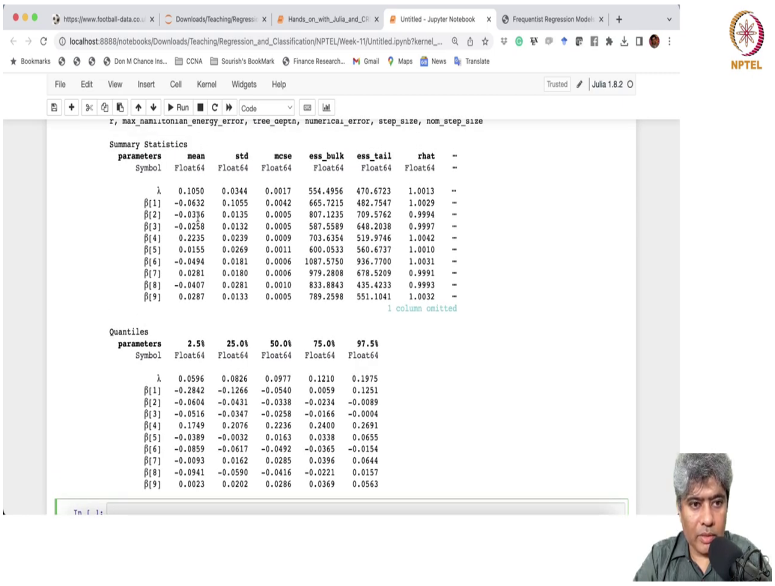


(Refer Slide Time: 14:26)



So, let me just run it. So, if you look into the, by default, it always simulate 1000 samples. And if you look into the rhat, you see the rhats are all close to 1. So, it took 500 burning. And after that, it simulated 1000 samples after burning. It is very fast because it uses Hamiltonian Monte Carlo method.

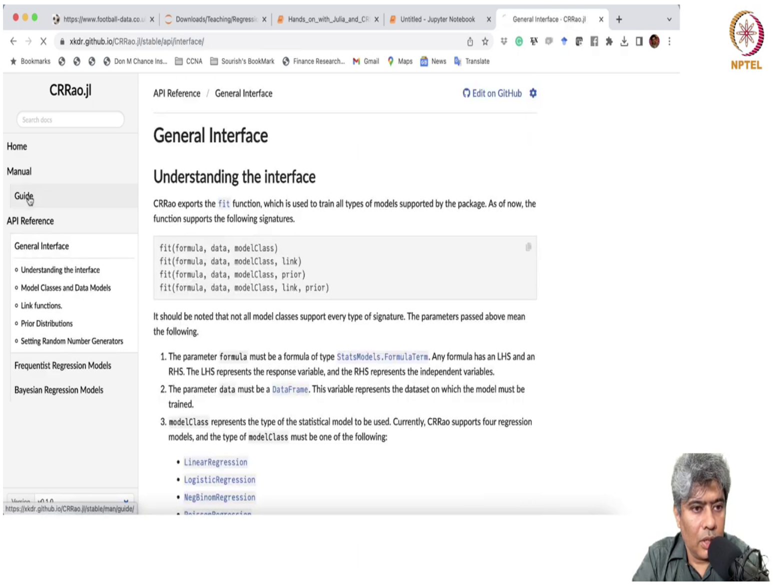
(Refer Slide Time: 14:50)



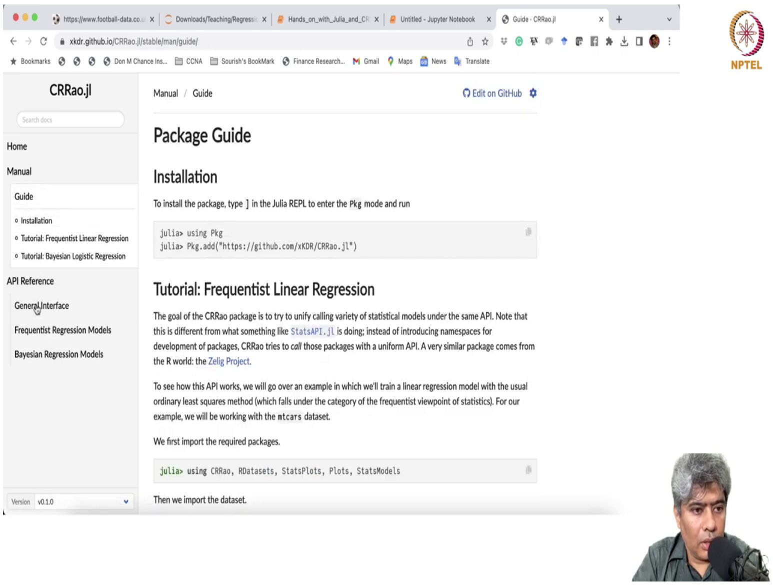
And these are the coefficients estimates. So, here you have 1, 2, 3, 4, 5, 6, 7, 8, 8 coefficients and intercept and now here also. So, total 9. So, you can see there are 9 coefficients, including first one is the intercept and lambda is the scale parameter that is required, ok.

So, intercept is negative 0.35 in MLE. And we are getting slightly off here, negative 0, 6. Others are like negative 0, 3 for HS, negative 0 3, negative 0 2, 4, yeah. So, the other value that looks like 0.238 for HST, which is 2, 3, 4th coefficient 0.22, yeah. So, the coefficients looks like similar to close to that of MLE. But it is very simple.

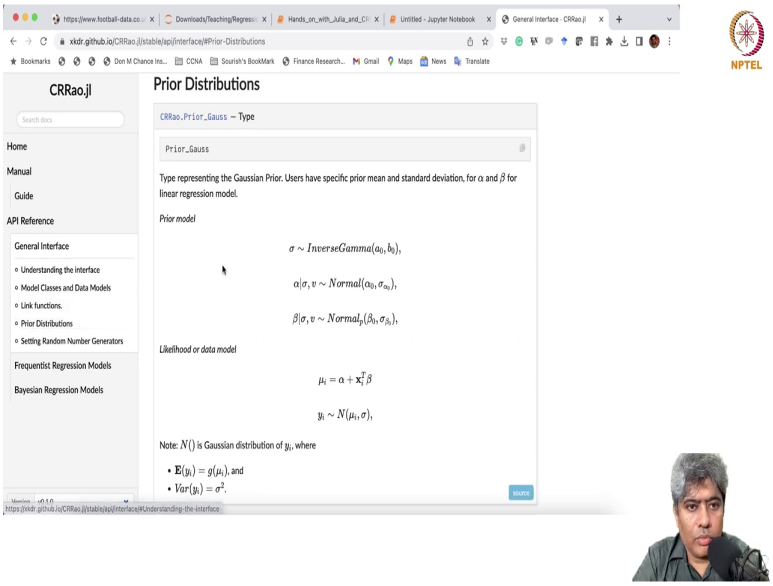
(Refer Slide Time: 16:17)



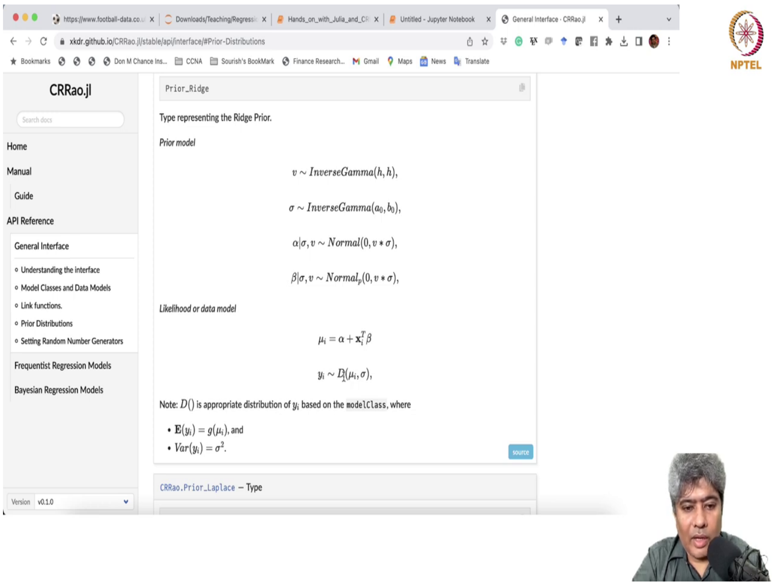
(Refer Slide Time: 16:19)



(Refer Slide Time: 16:22)



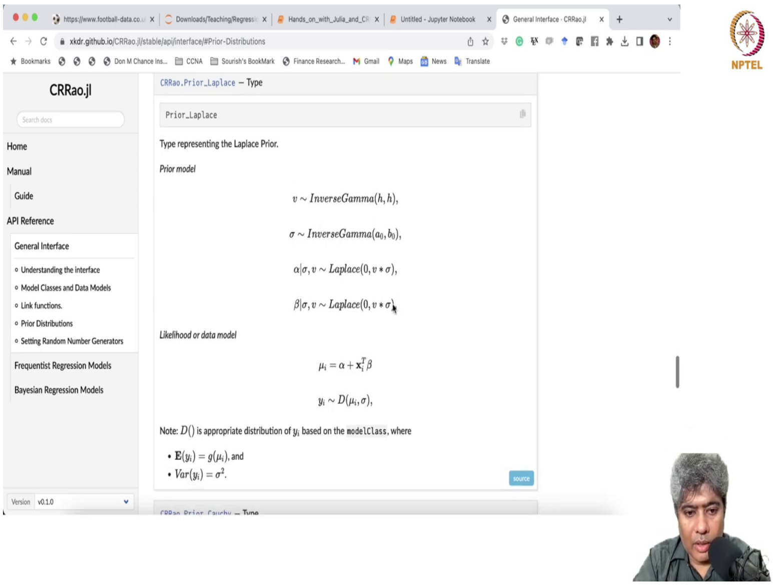
(Refer Slide Time: 16:28)



If you want to see the detail of the Ridge Prior, I think you can go to the right, API References of General Interface General Interface and then Prior Distribution. So, the first thing you can say prior Gauss and then here is the ridge prior and all the definition of the ridge prior.

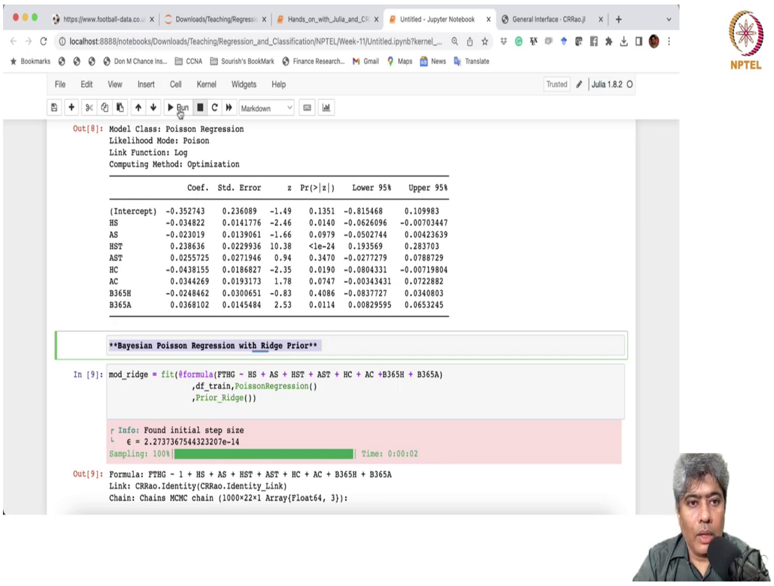
The way it is written, we have taken a appropriate distribution means Poisson, then you just take the Poisson and automatically it will do the rest of the thing. So, this is the ridge prior, the way it is defined in this in CRRao.

(Refer Slide Time: 16:48)

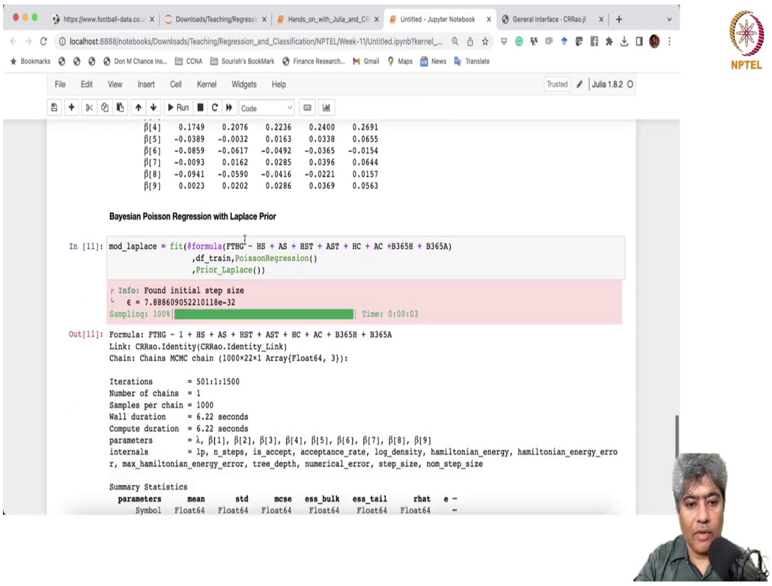


Next, we can try Laplace priors on the beta, Laplace distribution being imposed and inverse gamma prior is being imposed on the Poisson regression.

(Refer Slide Time: 17:08)



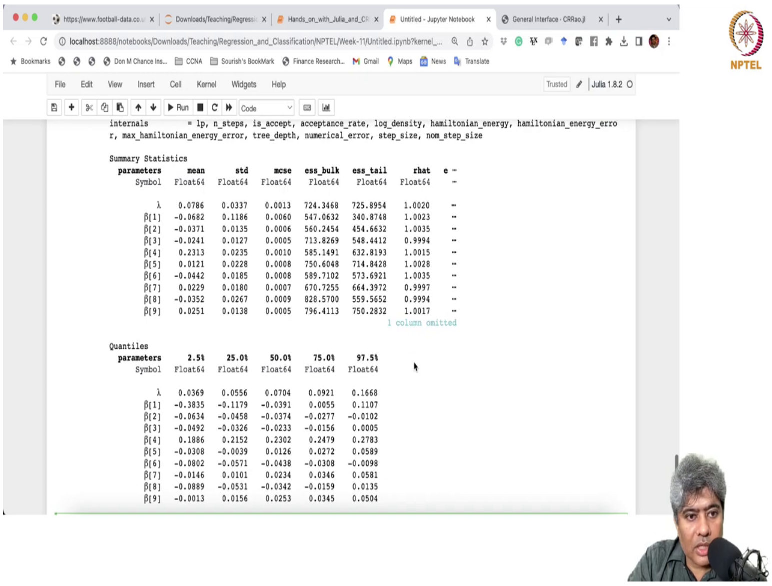
(Refer Slide Time: 17:13)



So, if I want Poisson regression with Poisson regression with Laplace priors, you can plot very easily using CRRao. So, Laplace Prior, all you have to do Laplace Prior, opsee, sorry, I have to do a mark down, ok. Now, it is fine. So, Bayesian Regression with Laplace Prior so, all you have to do is just copy this guy.

So, now on the coefficient, we are applying Laplace Prior, Laplace Prior distribution. On the coefficient, this is typically. So, you just change the name of the prior Laplace and it should and run. And CRRao will understand automatically that, ok, this is Laplace Prior that the user want to apply and boom, it runs and it gives you the all the estimates. So, by default, it will simulate 1000 samples after 500 burning. So, from 501 to 1500 iteration is being reported, rhat are all close to 1.

(Refer Slide Time: 18:24)

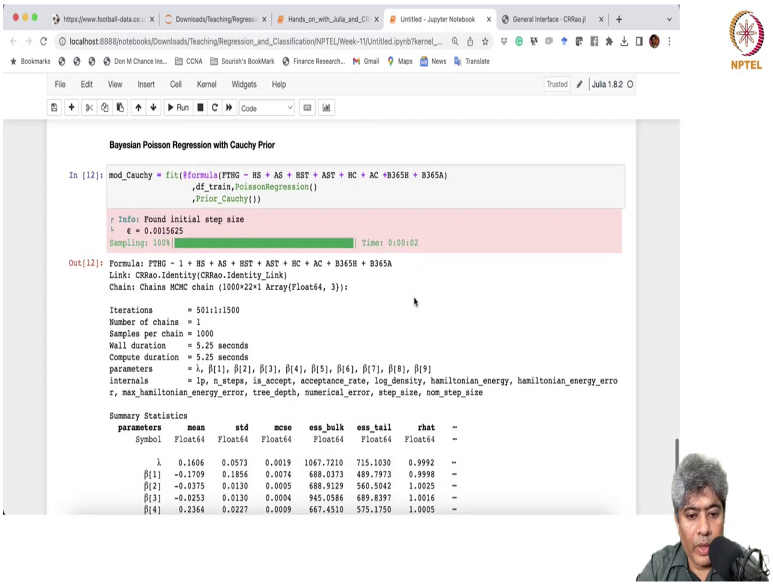


So, this is a good news; that means, the chain has converged and you can see similar kind of things behavior you can see, ok. So, this second coefficient plus H is effective, this is strong effective, the fourth coefficient, both coefficient was my 1, 2, 3, if HST, HST is strongly have strong effect. And with the last one here, it is saying that the odds does not have any effect, the odds does not have any effect.

The 6th one is, I think it was corner, I think it was corner 1, 2, 3, 4, 5, 6, yeah, corner does have had effect yeah. So, Laplace is saying that in the MLE, we got 3 in MLE method, we it got 3 at a 5 percent level, we can say home shot, number of shot taken, number of shot on target and the bet365s odd for awaiting. But it was surprising that I would not expect that it will awaiting, so, it will have no effect on number of goals code and the Bayesian methods is actually rejecting this method.

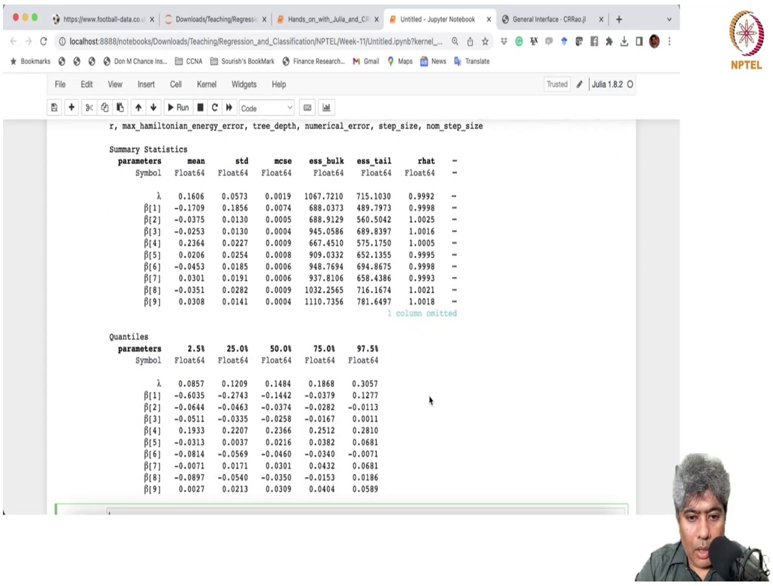
I mean, saying though ridge prices, yes, it did indeed, but Laplace Prior is saying no, no, we are not. So, this is important that you apply different kind of models and then you compare the results. And then next we can try Cauchy Prior, ok.

(Refer Slide Time: 20:23)



So, again, very simple fitting Cauchy Prior, first we will give a name, first we will put it as a mark down and then say instead of Laplace, we see first Cauchy prior and then what we are going to do is simply plot, just copy this guy. Cauchy prior and instead of Laplace, we just let C and then tab, it will fill it up by itself and then run. So, while, ok, it just, it was very fast, it was very fast. So, and it is indeed all converged as usual; the rhat is all close to 1.

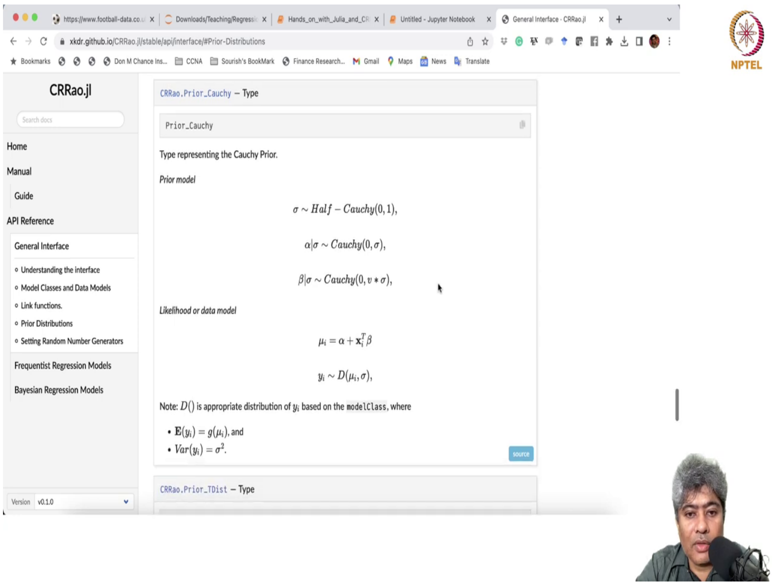
(Refer Slide Time: 21:17)



So, the this one has a effect, strong effect, this one has a strong effect, the entire 95 percent confidence bill does not include 0 and the last one also has a strong effect, this does not include 0, 6th one, the corner also has a effect.

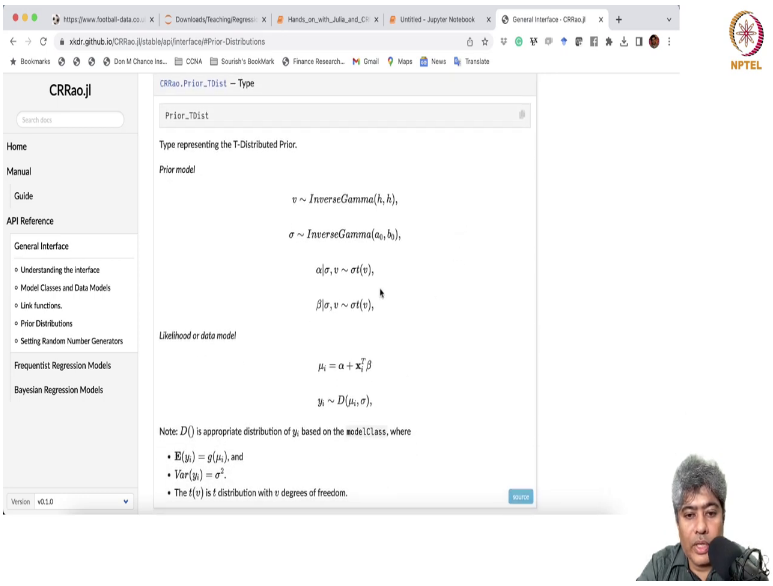
So, corner number of shots, number of shots on target, number of shots, number of shots on target, number of corners, these are going to have an effect, the model is again and again saying and the odds of the Away Team by Bet365 seems like having an effect, not maybe strong or something, but it seems like it does have an effect.

(Refer Slide Time: 22:04)



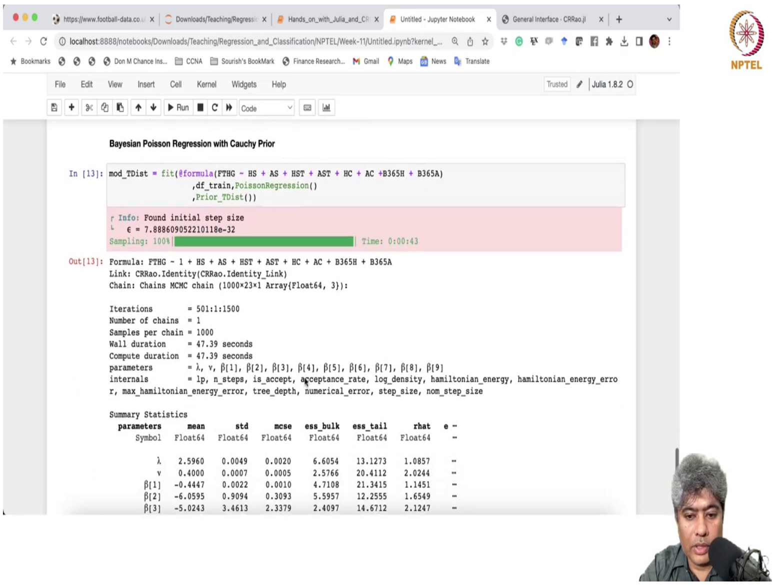
So, what is Cauchy prior? Let us see it will Cauchy prior in the everything of the model is same, but on the beta and alpha, on the coefficient we put Cauchy prior with a scale parameter have a half Cauchy distribution. So, this is very robust prior and next what we will try, we will try T-distributed prior, ok.

(Refer Slide Time: 22:27)



So, here we are going to scaled T-distribution, distributed prior on the beta and alpha and on the scale, we implement it InverseGamma prior. So, very variety of prior options we have in the CRRao.

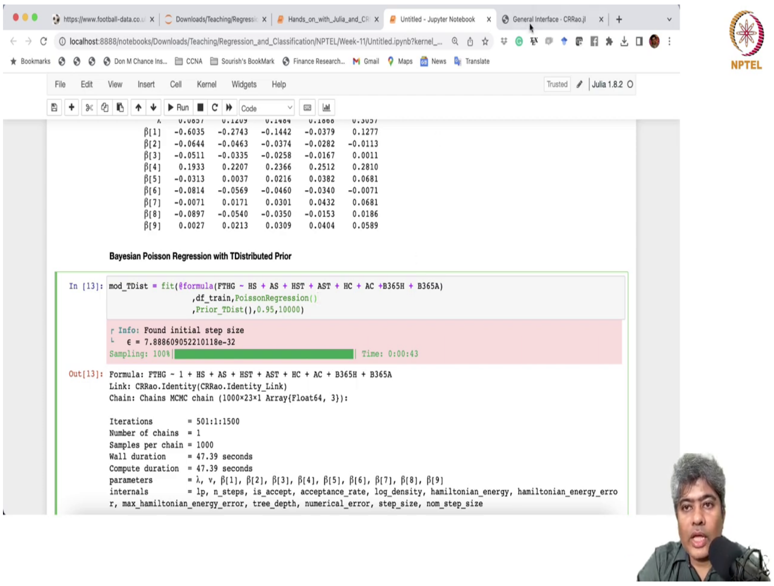
(Refer Slide Time: 22:58)



So, what we are going to do next is T-distributed prior, we are going to implement the T-distributed prior, ok. T-Distributed, ok. Just you write T and press tab and that should be fine. It is very fast looks like, (Refer Time: 23:41) sometimes it is it is bit slow, but the T-distributed might take a little bit time, ok.

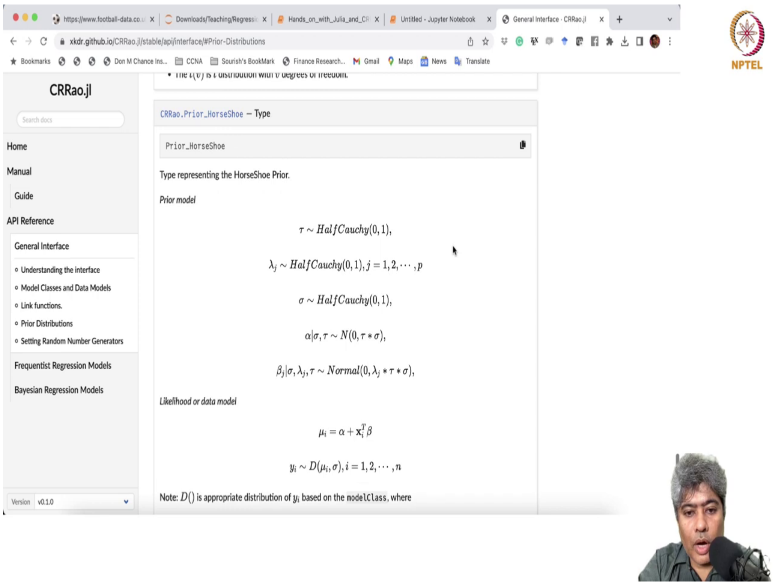
Let us see r you see T-distributed prior dies not did not able to converge you can see the r hat r 2 near 2. So, that means, definitely these the algorithm quantum integral did not converge by the 1000 samples, we need to increase the number of samples. So, and you saw it is slow.

(Refer Slide Time: 24:50)

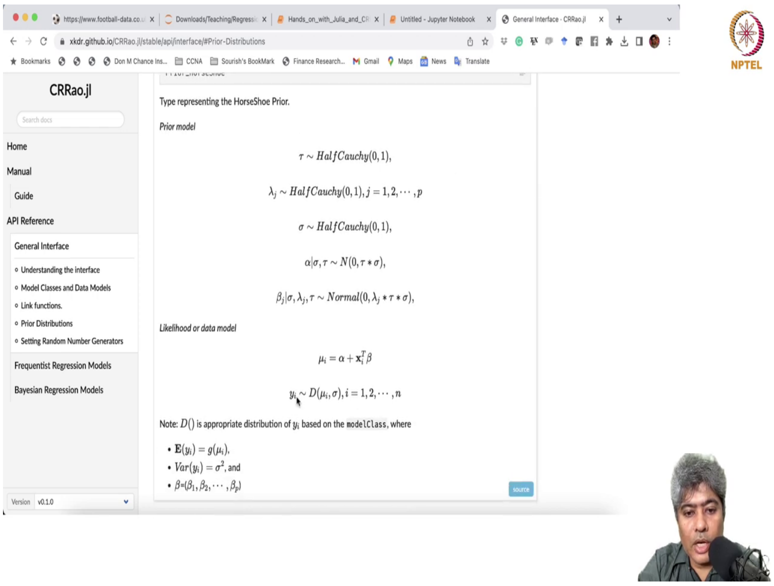


So, I am not going to I am not going to again increase the samples and run it because it will just take time. It will, all you have to do just increase the sample. I can show you all you have to do just probably you just paid 0.95 and then you have to instead of 1000 you have to just set 10,000 and run it. But I am not going to do that because it will just take longer time. So, you try it yourself, but I am going to try another prior, which is very popular prior called HorseShoe Prior, ok.

(Refer Slide Time: 25:23)

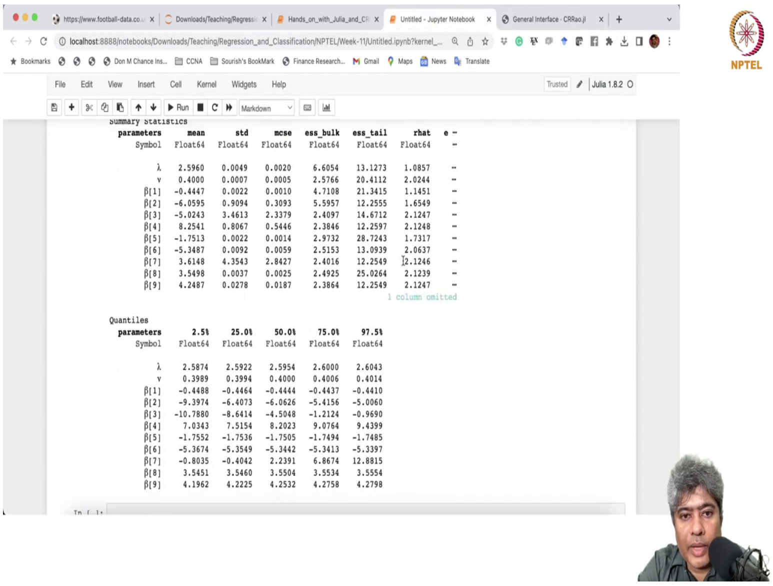


(Refer Slide Time: 25:32)

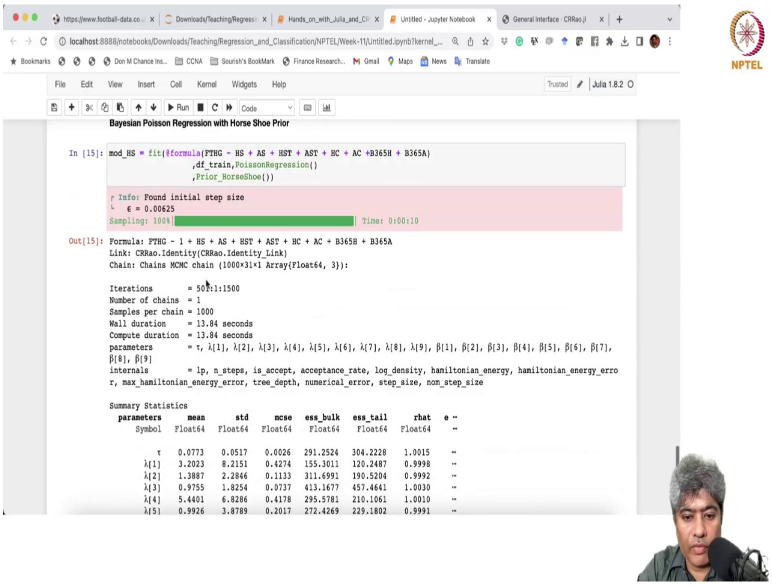


So, let me try HorseShoe Prior and HorseShoe Prior is let me show you how the HorseShoe Prior works. So, this is the T Dist y follow proper distributed typically in natural exponential family. In the our case it will be Poisson with mu i and. So, mu i will be alpha plus x i transpose beta. And on the beta and alpha there will be a conditional normal distribution and on the scale parameters of the conditional normal distributions we will have half cauch, HalfCauchy distribution.

(Refer Slide Time: 26:11)



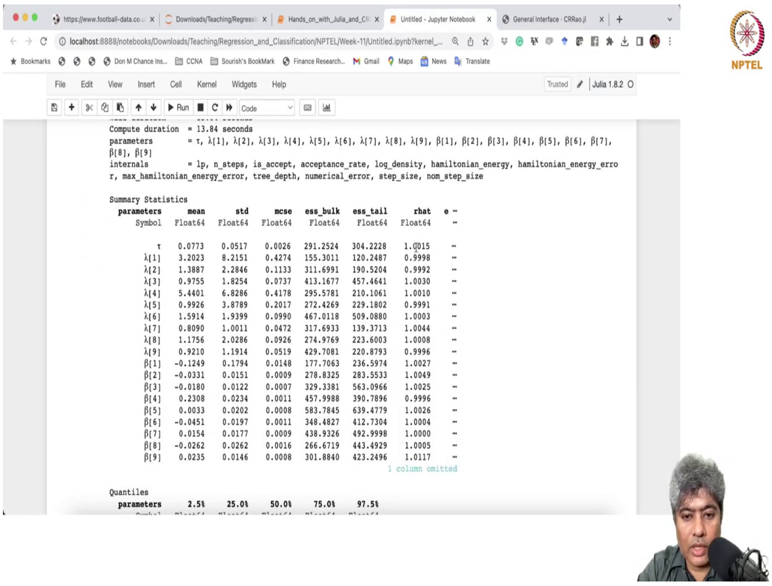
(Refer Slide Time: 26:15)



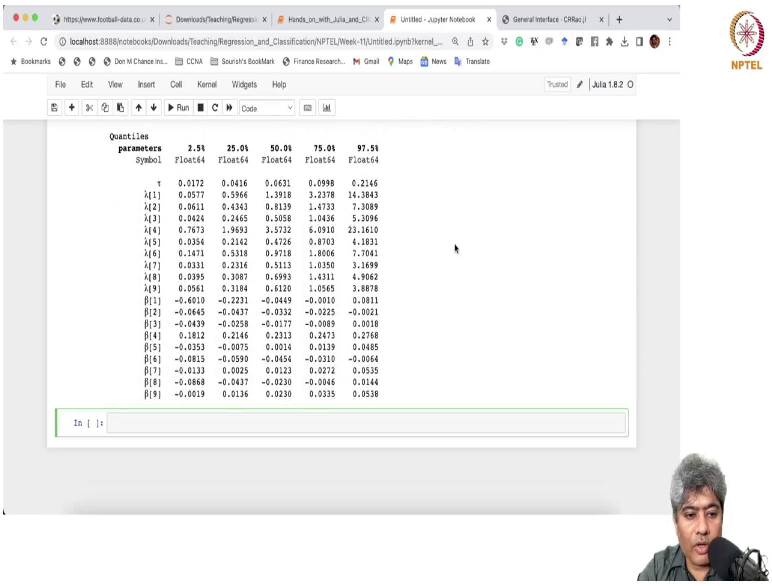
So, it is interesting it is a scale mixture distribution and it is very nice. It is very popular in the Bayesian literature. It has lot of good properties lot of good HorseShoe Prior, ok. So, let me just run it. I have to call it markdown and then it will run yeah. And then now what I will do, I will just copy this.

So, we what we found the T distributed prior will always require a bit of a more samples and then. So, HS I will just say and so, I do not need I will just use the default and H and then HorseShoe will work. Let me just run now ok, of course, this may be ok. Let me see if it was converged.

(Refer Slide Time: 27:15)



(Refer Slide Time: 27:21)



Quantiles

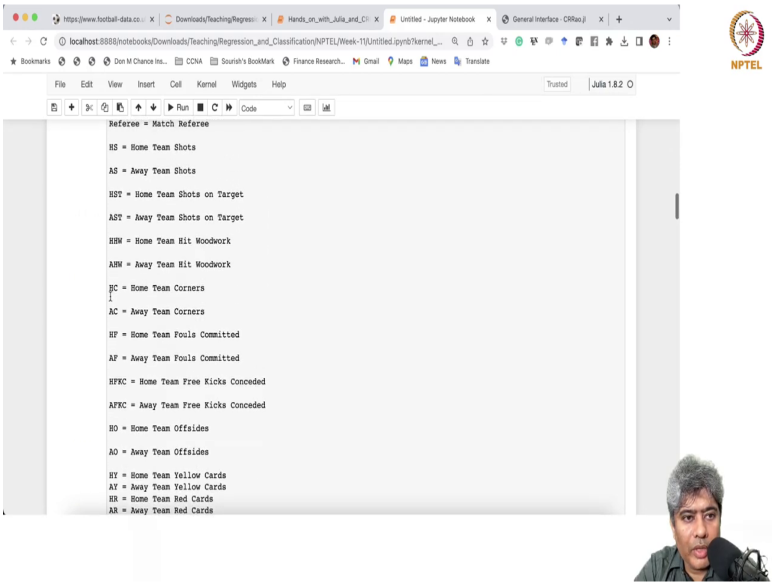
parameters Symbol	2.5% Float64	25.0% Float64	50.0% Float64	75.0% Float64	97.5% Float64
τ	0.0172	0.0416	0.0631	0.0998	0.2146
λ[1]	0.0577	0.5966	1.3918	3.2378	14.3843
λ[2]	0.0611	0.4343	0.8139	1.4733	7.3089
λ[3]	0.0424	0.2465	0.5058	1.0436	5.3096
λ[4]	0.7673	1.9693	3.5732	6.0910	23.1610
λ[5]	0.0354	0.2142	0.4726	0.8703	4.1831
λ[6]	0.1471	0.5318	0.9718	1.8006	7.7041
λ[7]	0.0331	0.2316	0.5113	1.0350	3.1699
λ[8]	0.0395	0.3087	0.6993	1.4311	4.9062
λ[9]	0.0561	0.3184	0.6120	1.0565	3.8878
β[1]	-0.6010	-0.2231	-0.0449	-0.0010	0.0811
β[2]	-0.0645	-0.0437	-0.0332	-0.0225	-0.0021
β[3]	-0.0439	-0.0258	-0.0177	-0.0089	0.0018
β[4]	0.1812	0.2146	0.2313	0.2473	0.2768
β[5]	-0.0353	-0.0075	0.0014	0.0139	0.0485
β[6]	-0.0815	-0.0590	-0.0454	-0.0310	-0.0064
β[7]	-0.0133	0.0025	0.0123	0.0272	0.0535
β[8]	-0.0868	-0.0437	-0.0230	-0.0046	0.0144
β[9]	-0.0019	0.0136	0.0230	0.0335	0.0538

Yes, you see it did converge. It did converge and the way it has been HorseShoe prior works is for each beta i there will be a local scale distributed. So, now you can see that. So, this is so; that means, for each beta 1 there will be a lambda 1 beta 2 there will lambda 2 and beta 9 there will lambda 9. And then these are the local shrinkage parameter and tau is the global shrinkage parameter.

So, all these parameters makes interesting local shrinkage and global shrinkage make the HorseShoe prior very very attractive to the Bayesian community. And what happens is what we see let us this is the intercept then next is the HS the home team shot on target. We see this is fully negative all the entire 95 percent confidence interval. Then this is number of shots on target is completely positive and this is number of corners completely negative.

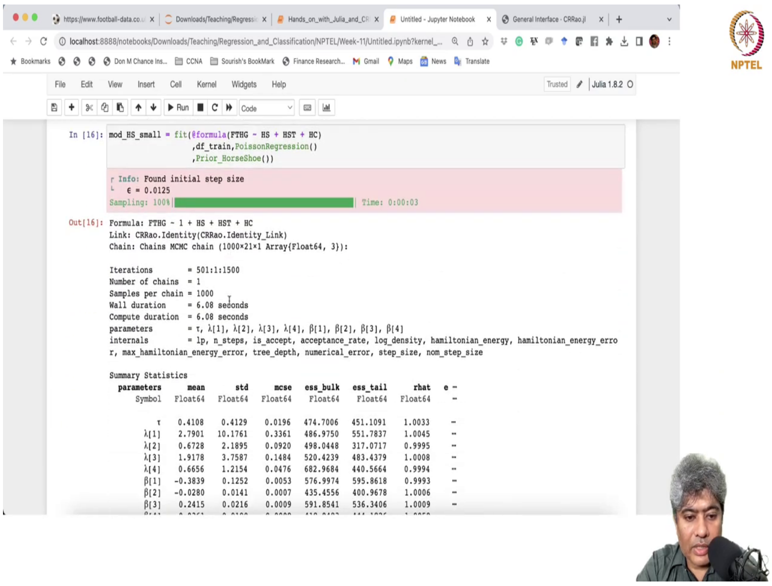
If you get too many corners your number of goals score will be less somehow corner has to do with the you know negative something to do with the negative. And what we are seeing that away teams and the Bet365 does not have the odds does not have an effect by HorseShoe prior.

(Refer Slide Time: 29:06)



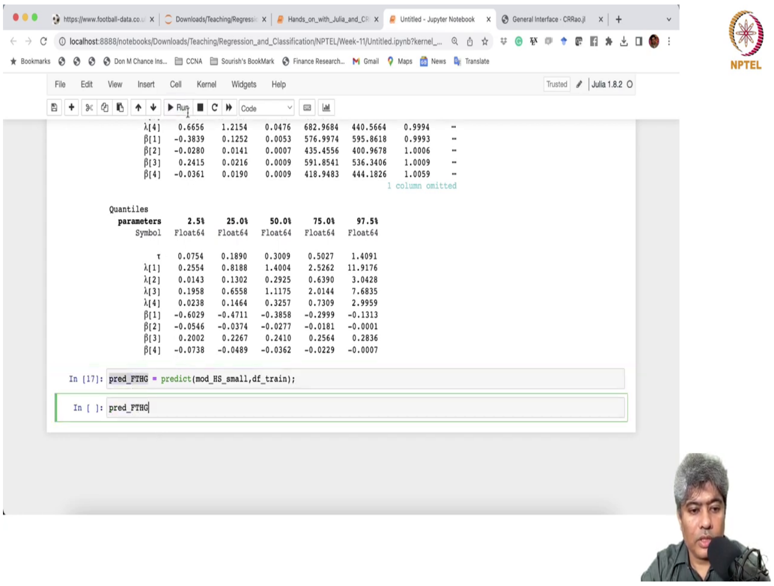
So, what we are seeing according to the model of. So, we can safely say HS which is home teams home team home team shot HST home team shots on target. HC HC that number of corners by home team. These are the three are the sort of sure for sure kind of model.

(Refer Slide Time: 29:32)



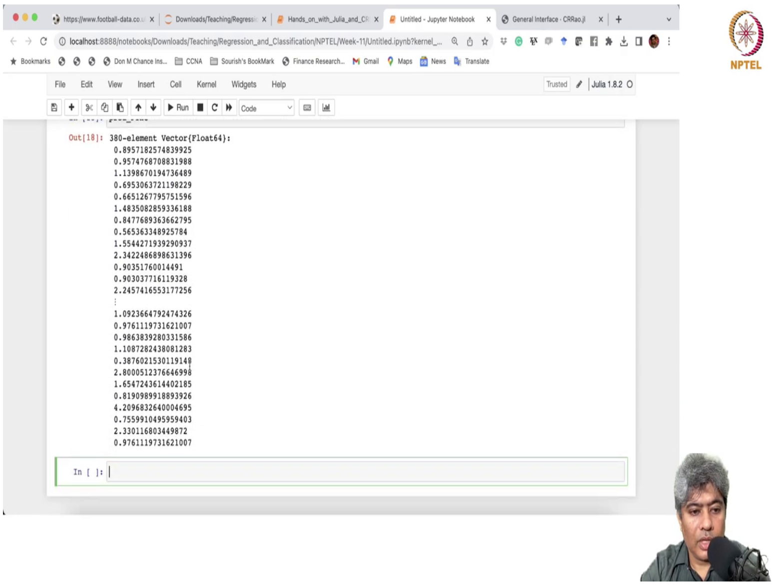
So, we can what we will try; we will try a slightly smaller model let us try a smaller model and home just home team shot on target home team corner number of, ok. So, that is how so, now, we have a shorter model and let us run this.

(Refer Slide Time: 29:56)



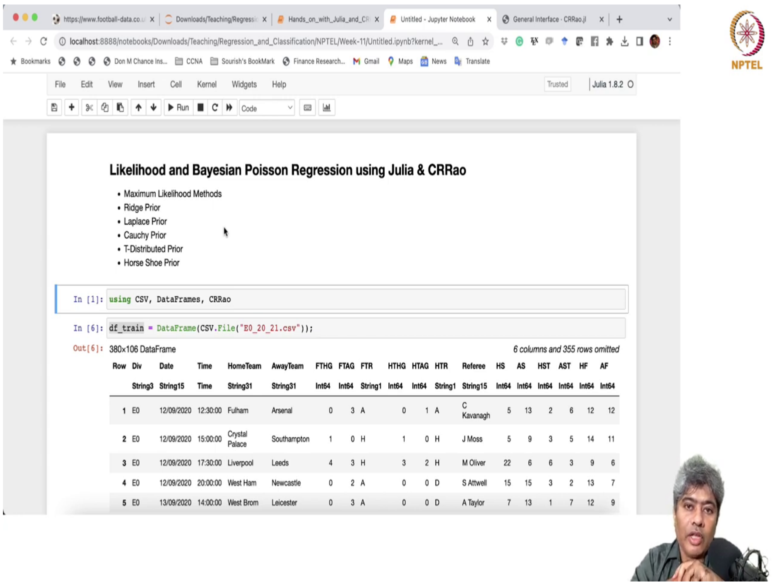
So, it will be faster equals faster yeah. So, we can see that each of these parameters were wet fast. And now what I am going to do we are going to just say predict if I just say predict you can just do use these models to predict also. And then you give the model and give the data set name. So, the data set name is say here I am using train you can have a test data also. If you just give say predict FTHG right FTHG.

(Refer Slide Time: 30:53)



And if you can see so, you can see that they are giving you the prediction also. So, this is one of the advantage of CRRao we do not have to when I mean you can fit any models and you can get the prediction also right away using just all the predict functions and it will be done. So, this is how you implement Poisson regression using you know Bayesian prior. So, Poisson regression both likelihood and the methods and the Bayesian implementation of the Poisson regression with Julia.

(Refer Slide Time: 31:34)



So, let me just write Likelihood and Bayesian Poisson Regression using Julia the number of implementation that we have done. Likelihood Maximum Likelihood Maximum Likelihood Methods so, actually we use Julia and CRRao ok. And then we used Ridge Prior, Bayesian Ridge Prior then we used Laplace Prior, Laplace Prior, then we used Cauchy Prior, Cauchy Prior, then we used T Distributed Prior, but it was not very successful it did not converge nicely.

We had to run more samples, T Distributed Prior and finally, HorseShoe Prior, HorseShoe Prior. So, all these we implemented very nicely and using CRRao in Julia. So, I hope you enjoyed this video and we will continue this in the next video. This next video we will be doing negative binomial regression.

Thank you very much.