# Introduction to Bayesian Regression

## Video 3 Transcript

Full resource: <https://www.ncrm.ac.uk/resources/online/all/?id=20843>

Oliver Perra: Hello. I am Oliver Perra and this is the third part of my introduction to Bayesian regression. So, in the first part I provided an overview of the approach. In the second presentation I have introduced specifications of linear regression model, so in this third part I’m going to show how Bayesian regression model can be run. And the Bayesian approach uses algorithms to estimate the posterior distribution of the parameter value combinations of interest, so in the linear model we have different parameters and we need to use some algorithms to estimate the plausibility of these different parameter combinations depending conditionally on the data collected and conditionally on the model assumptions, the priors I had specified. And the best algorithms for this task are Markov Chain Monte Carlo algorithms. This can represent the probability distribution of a combination of parameters by taking a series of random walks across the parameter space using a set of rules, and this set of rules basically ensures that these random walks will be more often in the direction of high-probability regions. This also ensures that at the end of these walks, which are called chains, the algorithm should have done a fair bit of exploration that ensures it can provide a reliable high-definition description of the underlying posterior distribution.

 So, this was a very non-technical description of how the algorithms and the Markov Chain works, but you can find more details in the references I have attached with this module. The problem with these algorithms is that they can get stuck in some unrepresentative regions of the parameter space. So, an important issue to check is that the algorithm converged. Another problem is that the solutions of the algorithm may not be stable, so they may provide solutions, the algorithms may provide solutions that are not well defined. Now, a version of the Markov Chain approach that has proved to be very efficient and reliable is the Hamiltonian Monte Carlo. And this is implemented in R through Stan and is particularly approachable using the Rethinking package created by McElrath, and again I refer to the book Statistical Rethinking that provides a lot of details about these methods. An advantage of the Hamiltonian Monte Carlo approach lies in the fact that it will provide error messages when some key problems arise.

 So, here on the left side of the slide you can see the model I had developed in the second presentation, so this is a linear regression model of newborn body weight, so the variable BW here. And newborns’ birth weight is supposed to be normally distributed around an average mu and with a standard deviation sigma. So, this is the likelihood function. The linear model specified in the second line says basically that the average birth weight of newborns is supposed to be a linear function of an intercept and a slope that represents the rate of change in newborns’ birth weight associated with deviations of maternal weight from the general weight. And then I have some prior assumptions about the intercept A, the slope B, and the standard deviation sigma, so the intercept A is supposed to be normally distributed around an average of 3,300 grams. The slope B is supposed to be normally distributed around an average of zero, so I emphasised in the second presentation that this is a sceptical assumption, so I am not taking sides, in a way, I am even equal probability to the slope being negative, positive or null, so with no association between predictor and outcome. And finally, the assumptions for the standard deviation is that any value between 0 and 1kg are equally likely.

 Besides this, on the right side of the panel, you can see how to write this model using the Rethinking package. The Rethinking package is basically specifying an M1 model. I call this model M1, but you can call it with any other name you want. And the option ulam is basically using or invoking a Hamiltonian Monte Carlo approach to estimate this model. So, the other options after a list are basically specifying the likelihood of the model first, so variable BW birth weight is supposed to be approximately normally distributed with mean mu and standard deviation sigma, then I have the linear model where mu(?) is a function of A, of the sum of A and the sum of the product of B by the maternal weight. And then I have priors for A, B and sigma. And finally, you can see that the other options in the right side panel ask the program to use four chains, sorry, six chains, and it is actually recommended to use more than or at least four chains usually to have reliable solutions. And these chains are now distributed across different core processes of the machine. And you can follow the script and the analysis using the script I have provided with the material of this module so you can see how to run this analysis.

[0:07:48]

 To avoid problems in estimation or issues in the length of time for the estimation, it is also advisable to specify (inaudible 0:08:02) or lists of data with only the variables of interest when you’re using ulam options (inaudible 0:08:10). And here you can see the output I have seen running this analysis, and again you can check the script attached. You can invoke more information about the algorithm here and you see that there were no error messages in running the algorithm and also you can have more information about what the algorithm was doing. And finally using the command precis in the Rethinking package, you can look at tables of the posterior distribution of the parameters, but this also provides some (inaudible 0:09:02) for the algorithm, in particular the rhat index here that you can see is 1 for all the three parameters of this model. The rhat statistic is basically an index of convergence of the algorithm compressed between and within chain estimates for model parameters, so if they agree, the ratio is going to be 1, as it is here. So, rhat values larger than 1 actually indicate issues with convergence and the recommendation is that the results of analysis should be used only if rhat is less than 1.05. Here they’re all 1 for all the parameters estimated, so that looks fine.

 The other statistic reported here, ess bulk, is the bulk effective sample size, which is an index of the stability of the algorithm. One recommendation is that bulk ess should be at least 100 for each chain, so in this case I had six chains so I expected to have over 600 in the effective sample size bulk of this parameter, so these results also look good. And authors like Krushke(?) emphasise the importance of reporting diagnostic information about the chains, the Markov chains you run, so this information should be reported in the results section or in appendices. It’s also important to inspect the processes of the algorithm to check for issues and diagnose potential problems and potential concerns. And chapter nine of Statistical Rethinking by McElrath provides clear guidance on these diagnostics and there are scripts I presented in some of the exercises attached with this module so it will provide some examples of other types of diagnostics for the chains and the estimation process.

 So, having ensured that these results are reliable and stable, we can look at the results. So, here the precise command provides some tabular results. For example, the average mean of newborns’ birth weight for women of average weight is estimated to be 2900g with a standard deviation of 29g. And for each kilogram of maternal weight over the average weight, the baby’s birth weight is estimated to increase by 5g, and this effect has a standard deviation equal to 2g. Also, you can see here reported with percentages are the 89% credibility intervals, which basically tell us that, for example, 89% of the time, or there is 89% probability that the slope will range between 1.50 and 9.01. So, you can see how these results also provide a very clear interpretation of these intervals, which is quite different from the confidence intervals reported in other, in the standard approach. Here the credibility intervals, the 89% credibility intervals reported in this table indicate the values that those parameters may take 89% of the time or with 89% probability, so the interpretation of these results is more intuitive and easier to communicate and this is one of the advantages of the Bayesian approach.

 As well as tables of results, plots are also very important in providing information about the results, particularly when the models are complex, for example, they include interactions. And plots also provide a way to informally check on the assumptions of the model, so it’s always recommended to plot the results from the posterior distribution. And in this graph here I presented some of the results. So, the dots represent the scatter of observed data, where on the horizontal axis I reported maternal weight in kilograms, and the vertical axis represents newborns’ weight in grams. The darker line is the slope that presents the average intercept A and the average slope B extracted from the posterior distribution of the plausible combination of parameter values. So, you can follow how I built this graph in the script attached with the material of this module. But one thing to keep in mind, however, is that in Bayesian analysis, this is just one of the most plausible lines that link the two variables, but there are many other lines that the model has ranked for the plausibility, and I have high plausibility as well with this line. So, to make the most of the Bayesian approach we shouldn’t focus our inference only on one value or these average values. And the posterior distribution, remember, considers and ranks every possible regression line and assigns relative plausibility to each of these plausible regulation lines, therefore there may be many other lines that have similar plausibility than the average line shown here, and we need to explore the posterior probability of different combinations of parameters of A and B to have a sense of also how uncertain is the inference from our analysis. So, it’s important to make the most of the richness of Bayesian analysis to also investigate the uncertainty around the estimation of our parameters, and an uncertainty that, as I emphasised, has also a more clear interpretation.

[0:17:01]

 And here in this graph I provide some examples of how to plot information about uncertainty in the distribution of combinations of different parameters according to the posterior distribution estimated depending on the data and the models I had. Here in the graph on the right side, for example, you can see the shaded area that represents the 89% credibility intervals of the average slope. So, that shaded area in light pink represents the values that the slope may take 89% of the time or with 89% probability. And in the script attached with this module you can see how I estimated and created these graphs. The other graph on the right side represents also in the blue shaded area the 89% credibility intervals of the predicted newborns’ birth weight, so it’s basically representing that 89% probability of the birth weights taking different values according to this model, and you can also see in this way plotted against the observed data how the model assumptions fit in an informal way with the data. So, these plots can be very useful in providing more information and more intuition about the model and how the model fits with the data. But at the same time we can use the posterior distribution of the parameters, different combination of the parameters to also report other type of analysis and other information. For example, we may look at how likely birth weights are to be over a certain value depending conditionally on different maternal weights and so on, so there’s a lot of different questions that can be answered by looking at the posterior distribution. Some of the references attached with this module also provide different examples, where Bayesian analysis may be used to test different models and compare different models as well. So, I refer to those references.

 So, to conclude, I wanted to highlight some of the key advantages of a Bayesian approach to regression analysis, in my opinion, and one is that it’s possible to incorporate previous knowledge, which makes for more sensible models and the opportunity really to capitalise on the body of research that exists on a field, and I think this is very important in building better science, in many ways. The other advantage I see is that also assumptions are formalised and therefore, first of all, there are no assumptions usually taken by default or as asserted by default as might happen or is more likely to happen in other approaches. And these assumptions are also easily reported and communicated using the type of notations I have used in these presentations. So, for me, one of the advantages is also the transparency in reporting the model’s assumptions that starting from in the analysis, and as long as they are transparently reported and then checked in sensitivity analysis, it makes sense to have sensible assumptions and more sensible models from which to start.

 The other advantage I see in Bayesian approaches to regression analysis is that every piece of informational data, however small, provide valid estimates, and this is because unlike in the standard approach to statistical hypothesis testing, estimation in a Bayesian approach is not based on assumptions of underlying sampling distributions that are only plausible when samples are large or the so-called asymptotic behaviour. So, however small a sample is, Bayesian estimates have a clear and valid interpretation. Obviously, there is a caveat, and the caveat is that if the sample is small, the estimates are more dependent on the prior assumptions, and if the prior is a bad one, the resulting estimates based on small samples will be misleading.

 And finally, there is an advantage in providing ranking of all the plausibility of all the combinations of parameter values, and as I tried to emphasise, the fact that we have a posterior probability of the different combination of parameter values allows to test and to report answers to different types of questions that are also more intuitive to explain and report to not just other researchers, but lay persons. So, the type of inference and the type of messages that we can provide, it’s also more intuitive, and again it also focuses on issues of uncertainty and probability that allows to focus on the strength of effect, for example, in clearer ways and in ways that are more easily understandable for not just statisticians or researchers.

[0:24:45]

 So, this has been a very quick overview of regression models and there is a lot more to be said and done, but I hope this was useful. And if you have found these presentations useful, please also look at the exercises I provided with this module. And for other material and other resources about research methods, please look at the webpage of the National Centre for Research Methods.

 Thank you very much. Bye.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK