INSERT YOU ANSWERS IN THE SPACE ALLOTED FOR EACH QUESTION. DO NOT USE MORE SPACE!

WHEN EXECUTING THE R CODE, THE OUTPUT WILL APPEAR IN THE CONSOLE SCREEN. DO NOT COPY PASTE OUTPUT TO THIS DOCUMENT, SIMPLY REFER TO IT IN YOUR ANSWERS!

MAIL THIS FORM (AND THIS FORM ONLY) TO [H.HOIJTINK@UU.NL](mailto:H.HOIJTINK@UU.NL) WHEN YOU ARE FINISHED.

YOU CAN WORK ON THE EXAM FROM 13.30 UNTIL 16.00.

**NAME:** Nina van Gerwen

**STUDENT NUMBER:** 1860852

**E-MAIL ADDRESS:** n.l.vangerwen@students.uu.nl

**Question 1:**

Explain what the Metropolis-Hastings sampler is: what are its main features, give an example in which it can be used. Note, do not use an example that was discussed in the lectures.

**Use at most 20 lines in this document to provide your answer (do not change letter type or margins).**

A: A Metropolis-Hastings (MH) sampler is an algorithm that through Markov-Chain Monte Carlo algorithms allows us to approximate a posterior distribution instead of having to derive the posterior exactly through Bayes’ theorem. There are multiple types of MH samplers, but they all have the same basis: first, you state an initial value and choose a proposal distribution to sample new values from. After sampling a new value, then you calculate an acceptance ratio through a certain mathematic formula that uses the densities of both your new and previous values for the posterior and proposal density. The new value will only be accepted if the ratio is larger than a randomly sampled value from the uniform distribution 0 – 1 (i.e., you only accept the new value with a certain probability). Then you repeat the sampling new values and acceptance ratios as many times as possible (your iterations) and in the end you will gain an approximation of your posterior distribution!

MH samplers are useful when we do not know (or cannot compute) a (semi) conjugate prior, as you can simply use any proposal distribution. However, the speed/efficiency of the algorithm will depend on your chosen proposal distribution.

MH samplers can be used in any type of situation in which you wish to gain posteriors. For example, if you wish to gain the posteriors of your regression coefficients, you can use a MH sampler for this. An example of a regression could be predicting exam grades through hours studied and IQ.

The two main types of MH samplers are dependent and independent MH samplers. In an independent MH sampler, a proposal density does *not* depend on the most recently accepted value, whereas in a dependent MH sampler it does.

**=======================================================================**

**Question 2:**

Run the R code provided for Question 2 in R Code Exam Rema Bayes 2020.R.

Explain what the command update(modspec,1000) does and why it is necessary to do this.

A: The command update(modspec, 1000) states that we want a burn in period for our specified model (modspec) of 1000 iterations. This means that we will allow the sampler to first ‘warm up’ in 1000 iterations and reach a region where there are good values to accept (i.e., to convergence)

Explain what the command n.chains=2 does and why it is good practice to specify n.chains=2.

A: The command n.chains=2 specifies that we want to parallelly sample two posteriors. The reason this is good practice is because it allows you to inspect whether one of the two posteriors is stuck at a local maximum. If this is the case, the two chains will be apart (when in fact they should be intermingling in order to reach proper convergence).

Explain what can be learned from autocorrelation plots. Which potential problem becomes apparent when you inspect the autocorrelation plot obtained after running the code? How can this problem potentially be solved? Apply your solution, did it work?

A: Autocorrelation plots can help to assess convergence of a model. High autocorrelations indicate a slow mixing of chains (i.e., your model is slow to converge). Furthermore, for a MH sampler, autocorrelations can also be used to evaluate your proposal density.

Inspecting the autocorrelation plot from the provided code, you see that the autocorrelations are slow to reach 0, only after around 28 lags does it reach a proper value. To potentially solve this problem, we can try to center the data and rerun the algorithm.

Now looking at the following autocorrelation plot we see that the autocorrelations already tend to zero at around the 5th lag! In other words, the autocorrelations are much better now.

**Use at most 20 lines in this document to provide your answer (do not change letter type or margins).**

**=======================================================================**

**Question 3:**

Modify and run the R code provided for Question 3 in R Code Exam Rema Bayes 2020.R. In modspec2.txt a model with three predictors is specified. Use the DIC to determine which two-way interaction (added to the three predictors) renders the best model. Present the DIC values computed for each interaction term in your answer. Interpret the DIC values

Run your best model two extra times. Present the DIC values in your answer. What do you see? How can this be explained?

**Use at most 20 lines in this document to provide your answer (do not change letter type or margins).**

A:

First, all predictors were centered in order to prevent issues due to multicollinearity.

Interaction between x1 and x2: DIC = 1751

Interaction between x2 and x3: DIC = 1755

Interaction between x1 and x3: DIC = 1755

Running the best model (interaction between x1 and x2) two extra times: DIC 1 = 1755 , DIC 2 = 1751.

What you see is that the penalized deviance (i.e., the DIC) differs per run. This is because the DIC calculates the effective number of parameters in your model, and this approximation changes every run. Hence, it makes sense that the penalty (i.e., the estimated number of parameters) and in turn the DIC differ per run.

**=======================================================================**

**Question 4.**

Data: a dichotomous dependent variable that is predicted by two continuous variables.

Which model can be used to determine whether or not the dependent variable can be predicted by the two continuous variables.

A: A model that can be used to determine whether the dependent variable can be predicted by the two continuous variables could be a Bernoulli distribution with a different parameter per person.

Imagine, you have available, the posterior predictive distribution of the data corresponding to a null-population in which the dependent variable can not be predicted from the two continuous variables. Imagine, you have also available, the observed data set containing one dichotomous and two continuous variables.

Present two test-statistics (and not discrepancy measures) which can be used to test whether the observed data set comes from the null-population or not. Explain for each test-statistic why it can be used to determine whether or not the predictive power of both continuous variables is zero. Be creative, come up with two different test-statistics and not one being a “variation” of the other.

A: In the null-population, the data should follow a binomial distribution (as everyone has the same probability to be a 1, and it is not predicted by anything). Whereas in the observed dataset, the probability to be a 1 can differ per person (and this is predicted by the two continuous variables).   
  
So to test whether the observed data comes from the null-population or not, we can use the following statistics:

First statistic:  
Divide the one of the two continuous variables into people who score higher than the mean and lower than the mean. Then, you can calculate the proportion of 1s on the dependent variable in the group that scores higher and divide this by the proportion of 1s on the dependent variable in the lower scoring group. Now you have a ratio of probabilities, and because in the null population datasets the probability is not predicted by this, the ratios should be close to 1. Whereas in the observed dataset, the ratio should be either smaller than 1 (if there is a negative relationship) or larger than 1 (if there is a positive relationship). The ratio will also slightly tell you about the strength of the association (the further from 1 the ratio is, the stronger) but not a lot as you dichotomized a continuous variable.

Second statistic:  
Variance of the dependent variable minus the variance of the dependent variable, conditional on the two continuous predictors. In the null-population datasets, this should be close to 0, as the variance should be the same whether you condition on the predictors or not (as they are not predictors of the dependent variable), whereas in the observed dataset, the variance should be lower once you condition on the continuous predictors, as these predict the chance of being a 1.

Again this does not show strength of relationship, only the presence of it.

**Use at most 20 lines in this document to provide your answer (do not change letter type or margins).**