



university of  
 groningen

faculty of science  
 and engineering

mathematics and applied  
 mathematics

# Modeling the Transmission of Intelligence within a Bayesian Framework

Bachelor's Project Mathematics

July 2018

Student: G. Nolle

First supervisor: dr. M.A. Grzegorzczuk

Second assessor: dr. W.P. Krijnen

## Abstract

Psychological studies have shown that intelligence clusters within families. This similarity between parents and their children is the result of a mixture of genetic and environmental factors. In this paper it is described how we can infer a model of transmission of intelligence within a Bayesian framework. The Metropolis Hastings MCMC algorithm has been applied for model inference. A data set containing scores of parents and twins who completed the RPM has been used. The RPM is a standardized test for quantifying the intelligence of an individual. After an description of how we can use the Metropolis Hastings algorithm to generate samples from the marginal posterior distributions of the parameters of our model, several plots of these samples are shown. The trace plots show that the Markov chain has converged within 1.000.000 iterations for all model parameters. The density plots show that the marginal posterior distributions of the parameters have the shape of a normal distribution. Finally, two methods have been provided that can determine if the transmission of intelligence of an individual is dominated by genetic factors or environmental factors. The first approach showed that for approximately one half of the twins the transmission of intelligence is dominated by genetic factors. The transmission of intelligence of the other part of the twins is dominated by environmental factors. Further research is necessary to determine the number of twins for which the transmission of intelligence is dominated by genetic factors and the number of twins for which the transmission of intelligence is dominated by environmental factors in the second approach.

## Acknowledgements

First of all, I would like to thank my first supervisor dr. Marco Grzegorzcyk for thinking along with me, answering all my questions and taking plenty of time to do so. This project has made me feel more confident about programming and your faith has definitely contributed to this, so I want to thank you for that. I would also like to thank my second supervisor dr. Wim Krijnen for taking the time to read and evaluate this project. Furthermore, I would like to acknowledge prof. dr. Van den Berg for providing a data set for my research. This project would not have been as interesting without it. Finally, I would like to say thank you to V.A. Bernal Arzola for running my code on a cluster and helping me to fix a software bug in R in the last week of my project.

# Contents

|           |   |           |
|-----------|---|-----------|
| <b>1</b>  | <b>Introduction</b>                           | <b>4</b>  |
| <b>2</b>  | <b>Background theory</b>                      | <b>6</b>  |
| 2.1       | Markov chain Monte Carlo methods . . . . .    | 8         |
| 2.1.1     | Metropolis Hastings Algorithm . . . . .       | 9         |
| 2.1.2     | Gibbs sampling . . . . .                      | 10        |
| <b>3</b>  | <b>Data</b>                                   | <b>11</b> |
| <b>4</b>  | <b>Parent-offspring model</b>                 | <b>12</b> |
| 4.1       | The measurement model . . . . .               | 12        |
| 4.2       | The genetic model . . . . .                   | 15        |
| 4.3       | The complete parent-offspring model . . . . . | 18        |
| <b>5</b>  | <b>Metropolis Hastings algorithm applied</b>  | <b>20</b> |
| <b>6</b>  | <b>Results</b>                                | <b>23</b> |
| 6.1       | Descriptive statistics . . . . .              | 23        |
| 6.2       | Convergence of Markov chains . . . . .        | 24        |
| 6.3       | Transmission of intelligence . . . . .        | 27        |
| <b>7</b>  | <b>Discussion</b>                             | <b>30</b> |
| <b>8</b>  | <b>Conclusion</b>                             | <b>31</b> |
| <b>9</b>  | <b>Bibliography</b>                           | <b>32</b> |
| <b>10</b> | <b>Appendix</b>                               | <b>33</b> |

# 1 Introduction

Psychological studies have shown that intelligence clusters within families. This similarity in intelligence is a result of a mixture of genetic and environmental transmissions. Genetic transmission is defined as the transfer of genetic information from parent to offspring. So, parents pass on the genes that express intelligence to their children. On the other hand, cultural transmission takes place. This means that the phenotype of offspring depends on their environment and on the way that their parents pass on information to them [8]. An example of cultural transmission is that parents with a high intelligence quotient (IQ) may be able to provide a better education for their offspring than parents with a lower IQ are able to, as they usually have a higher income. This can result in a difference in IQ in their offspring: the offspring of parents with a higher IQ will have a higher IQ than the offspring of parents with a lower IQ.

In this study, it will be investigated if the transmission of intelligence is dominated by genetic factors or environmental factors. The foundation and research design of this study are based on the Master thesis of Otermann [18] from the University of Utrecht. Her research differs from previous studies by taking into account the effect of assortative mating and possible correlations between genotype and environment.

Assortative mating is a form of non-random mating in which pair bonds are established on the basis of phenotype [9]. Examples are choosing a partner based on ethnic preferences or professional interests. One form of assortative mating is social homogamy. This means that partners choose each other because of a similar environment. Some studies showed that there is evidence for this form of assortative mating [2]. Another form of assortative mating is phenotypic assortment. This is the tendency for individuals to select their partner on the basis of observed phenotype [9]. In the context of this study, it means that partners choose each other based on intelligence. This may result in a transmission of not only genes, but also in a transmission of environment from parents to their children. Therefore, phenotypic assortment may lead to a correlation between genotype and environment. In this study, it will be assumed that phenotypic assortment might occur.

Otermann made use of a parent-twin design in her research. A twin design allows to make a distinction between genetic effects and environmental effects, since monozygotic (identical) twins share all of their DNA. Dizygotic (fraternal) twins share on average 50% of their segregating genes [4]. In this study only data from monozygotic twins has been used. Variations of a twin design that have been adopted in previous twin studies use models that take into account data of twins together with data of their parents, partners, siblings and children [1, 9]. Examples of models that evaluate such a family-twin design are the Cascade and Stealth model [16].

In these classical twin studies, it has been assumed that random mating occurs instead of assortative mating and that there is no correlation between genotype and environment. This may lead to biased results [5]. If random mating between parents is assumed while there is non-random mating, for example, this will decrease the estimation of genetic influences and increase the estimation of the influence of a similar environment. Fortunately, assortative mating and some forms of genotype-environment correlation can be assessed by extending the classical twin design by including the twins' parents in the research design [5]. This has been done in this study. So, assortative mating and the correlation between genotype and environment have been taken into account and it can be tested to which extent they influence the similarity in intelligence between parents and their children.

Since merging a genetic model and an Item Response Theory (IRT) measurement model leads to the most precise inference of the parameters of the genetic model [18], this will be done in this project. The Metropolis Hastings Markov chain Monte Carlo algorithm will be used for model inference. This algorithm is a Bayesian inference method. The data that Otermann used in the measurement model was obtained by means of the Raven Progressive Matrices test (RPM). However, since this original data set is strictly confidential, it will be worked with a synthetic parent-offspring data set in this research instead. This data set is representative of the original data set and has been provided by Prof. dr. Van den Berg from the University of Twente, who was the first supervisor of Otermann.

The main goal of this bachelor project is to describe the parent-offspring model comprehensively and statistically. Moreover, the aim is to implement the MCMC algorithm. It will be worked with the R programming environment. In upcoming sections, more explanation will be given about Bayesian statistics and MCMC methods first of all. Then, the details of the measurement model and the genetic model will be given successively, which together make up the complete parent-offspring model of transmission of intelligence. Thereafter, it will be explained how the Metropolis Hastings algorithm is applied to this model to generate samples from the marginal posterior distributions of its parameters. Convergence plots of these samples are included in the successive chapter, followed by an analysis with respect to the transmission of intelligence and a discussion of the results. We conclude with some final remarks and further research possibilities.

## 2 Background theory

A Bayesian inference method has been used to infer the genetic model and the (IRT) measurement model simultaneously. This section will provide a short introduction to Bayesian statistics and Markov Chain Monte Carlo (MCMC) methods.

In Bayesian statistics, a parameter  $\theta$  is treated as a random variable. The goal is to infer the distribution of  $\theta$ . This procedure works as follows [14]: First of all, a prior distribution for  $\theta$  is chosen. This prior reflects the a-priori probability of the parameter  $\theta$  and is denoted by  $p(\theta)$ . Then, a probability density function  $p(y|\theta)$  is chosen, which is called the likelihood. It describes our belief about the data  $y$  if we know that  $\theta$  is true. Once we have the data  $y$ , the last step is to update our belief about  $\theta$ . This is reflected in the posterior distribution  $p(\theta|y)$ . The posterior describes our belief that  $\theta$  is the true value, having observed the data set  $y$ . It is given by the following formula, which is also known as Bayes' rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}.$$

It can now be observed that the term in the denominator does not depend on  $\theta$ . Therefore, it follows that the posterior density is proportional to the likelihood times the prior:

$$p(\theta|y) \propto p(\theta)p(y|\theta). \quad (1)$$

The prior is called conjugate if the posterior is from the same distribution family as the prior [14]. We can determine the posterior distribution of the parameter analytically from a conjugate prior and the likelihood of the data. This is shown in the following example.

### Example 1

Suppose we have  $n$  data points that are all independently and identically Poisson distributed with parameter  $\theta$ :

$$Y_1, \dots, Y_n \sim Poi(\theta).$$

Then the likelihood of each data point is given by the following formula:

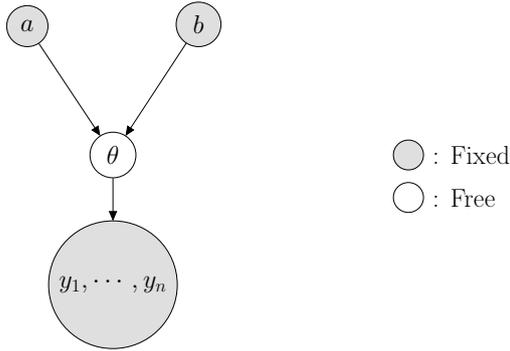
$$p(y_i|\theta) = \frac{e^{-\theta}\theta^{y_i}}{y_i!}, \quad \text{where } i = 1, \dots, n \text{ and } i \in \mathbb{N},$$

since this is the probability density function of a Poisson distribution. Moreover, we let the parameter  $\theta$  have a Gamma distribution with parameters  $a$  and  $b$ . Then, the prior density is given by

$$p(\theta|a, b) = \frac{b^a}{\Gamma(a)} \cdot \theta^{a-1} \cdot e^{-b\theta},$$

which is the probability density function of a Gamma distribution.

It is convention to depict this situation in a graphical model as in figure 1, where the parameters that have to be inferred (i.e., the free parameters) are represented by white circles and the data and the fixed parameters (i.e., the observed parameters) are represented by grey circles. Such a graphical representation of the model makes it easier to recognize how the parameters depend on each other and therefore, to see what the posterior distributions of the parameters are proportional to. This will be useful when we generate samples from the posterior distributions of our parent-offspring model, as will become clear in an upcoming chapter.



**Figure 1: Graphical representation of the situation in example 1. Parameters that have to be inferred (i.e., the free parameters) are represented by white circles. The data and the fixed parameters (i.e., the observed parameters) are represented by grey circles.**

As can be seen from equation 1, the posterior distribution is proportional to the prior times the likelihood. We can simply fill in these distributions and arrive at the following:

$$\begin{aligned}
 p(\theta|y_1, \dots, y_n) &\propto p(\theta|a, b) \cdot p(y_1, \dots, y_n|\theta) \\
 &= \frac{b^a}{\Gamma(a)} \cdot \theta^{a-1} \cdot e^{-b\theta} \cdot \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \\
 &\propto \theta^{\sum_{i=1}^n y_i + a - 1} \cdot e^{-(n+b)\theta},
 \end{aligned} \tag{2}$$

since  $\frac{b^a}{\Gamma(a)}$  and  $\prod_{i=1}^n y_i!$  do not depend on  $\theta$ . We can recognize that the last expression of 2 is proportional to the probability density function of a Gamma distribution. We conclude that the posterior is Gamma distributed with parameters  $\sum_{i=1}^n y_i + a - 1$  and  $n + b$ :

$$p(\theta|y_1, \dots, y_n) \sim \text{Gamma}\left(\sum_{i=1}^n y_i + a - 1, n + b\right).$$

Since the prior and posterior distributions both have a Gamma distribution, the prior is conjugate. In this case, we are able to determine certain quantities of interest of the posterior distribution analytically. Examples are the mean and the standard deviation. However, if there is more than one parameter or non-conjugate priors are used, it is not possible to determine the posterior distribution of a parameter analytically. We can then use a so-called Markov chain Monte Carlo (MCMC) method. In the following section, a short introduction to MCMC inference methods will be given.

## 2.1 Markov chain Monte Carlo methods

Sometimes it is difficult or impossible to obtain exact values for quantities of the posterior distribution. This problem can be overcome by an Markov chain Monte Carlo (MCMC) method.

Two properties of a MCMC method appear in the name, namely the Markov property and the Monte Carlo property. The Markov property of an MCMC method relates to the fact that the random samples are formed by a special sequential process. A Markov chain describes a sequence of possible events, the random samples. Each random sample is used to generate the next random sample in the chain. The special property of this chain, the Markov property, is that new samples do not depend on any samples before the previous one. So, each sample only depends on the one before it. The Markov chain will eventually converge to a stationary distribution, which will always be the same if the number of samples is large enough. The initial value of the chain does not matter. Secondly, there is the Monte-Carlo property of an MCMC method. This property relates to the idea that random samples from the Markov chain are examined. The properties of the stationary distribution can then be estimated with these samples. [19]

To summarize, the following is happening in an MCMC method. First of all, we generate a chain of samples from the stationary distribution by applying an algorithm in which the Markov property is reflected. Then, we can use these new samples to estimate the quantity of interest by means of a Monte Carlo approximation. It is important to note that an analytic computation is always preferable. However, this is not always feasible. It is then appropriate to use an MCMC method.

In order to draw samples from the stationary distribution, an MCMC method starts with an initial guess. This is a value that could have been drawn from the distribution. Thereafter, MCMC is used to produce a chain of new samples from this initial guess. The process of proposing a new sample consists of two steps [19]:

1. A proposal for the new sample is created. We add a small perturbation to the old sample in order to do this.
2. This new proposal is either accepted as the new sample or rejected. We retain the old sample in the last case.

There are several ways to create a proposal for the new sample in the first step. One way is by means of the Metropolis-Hastings algorithm. This algorithm generates samples from the posterior distribution of a parameter. When parameters are very strong correlated, it can be beneficial to use Gibbs sampling instead. This MCMC algorithm that generates samples from the posterior distribution of a parameter as well. The Metropolis Hastings algorithm will be explained in the following section. [19]

### 2.1.1 Metropolis Hastings Algorithm

As mentioned above, one way to construct a Monte-Carlo chain that converges to the posterior is by means of the Metropolis Hastings MCMC algorithm (MH-MCMC). This algorithm consists of a few steps, which in pseudo-code look as follows [11]:

1. Initialization: set  $\theta^{(1)} = \theta \in \Theta$ . Here,  $\theta^{(1)}$  is the initial guess and  $\Theta$  denotes the parameter space.
2. MCMC iterations: for  $t = 1, \dots, 2T$ 
  - (a) Propose to move from the current state  $\theta^{(t)}$  to  $\theta^{(*)}$ . This new state is proposed with proposal probability  $Q(\theta^{(t)}, \theta^{(*)})$ .
  - (b) Accept the new state  $\theta^{(*)}$  with probability  $A(\theta^{(t)}, \theta^{(*)})$ . Compute  $A(\theta^{(t)}, \theta^{(*)})$ .
  - (c) Draw a random number  $u$  from a uniform distribution on  $[0, 1]$ .
    - If  $u \leq A$ , then set  $\theta^{(t+1)} = \theta^{(*)}$  (i.e., "accept the candidate").
    - If  $u > A$ , then set  $\theta^{(t+1)} = \theta^{(t)}$  (i.e., "reject the candidate" and leave the state unchanged).
  - (d) Do the next iteration.

Each MCMC iteration generates one new sample in the chain by this algorithm. A chain consisting of  $N$  samples from the posterior distribution in the one parameter case is denoted by:

$$\theta^{(1)}, \dots, \theta^{(N)} \sim p(\theta|y_1, \dots, y_n).$$

In principle, the proposal probabilities  $Q(\theta^{(t)}, \theta^{(*)})$  in step 2a can be chosen arbitrarily, as long as two conditions are fulfilled:

1.  $Q(\theta, \theta^{(*)}) > 0 \Leftrightarrow Q(\theta^{(*)}, \theta) > 0$ .
2. Every parameter  $\theta^{(*)} \in \Theta$  must be reachable from every parameter  $\theta \in \Theta$  in a finite number of steps.

In this research, the proposed state  $\theta^{(*)}$  is created in the following way. Given  $\theta^{(t)}$ , we sample a random number  $v$  from a uniform distribution over the interval  $[-\epsilon, \epsilon]$  and propose:

$$\theta^{(*)} = \theta^{(t)} + v. \quad (3)$$

Then, the acceptance probability  $A(\theta^{(t)}, \theta^{(*)})$  in step 2b is given by the following expression:

$$A(\theta^{(t)}, \theta^{(*)}) = \min\left\{1, \frac{p(\Theta^{(*)}|y_1 \dots y_n)}{p(\Theta^{(t)}|y_1 \dots y_n)} \cdot \frac{Q(\theta^{(*)}, \theta^{(t)})}{Q(\theta^{(t)}, \theta^{(*)})}\right\},$$

where  $\Theta^{(t)}$  denotes the vector of parameters in iteration  $t$  with parameter  $\theta$  having the value of the current state  $\theta^{(t)}$ . Likewise,  $\Theta^{(*)}$  denotes the vector of parameters in iteration  $t$  with parameter  $\theta$  having the value of the proposed state  $\theta^{(*)}$ . The term  $\frac{Q(\theta^{(*)}, \theta^{(t)})}{Q(\theta^{(t)}, \theta^{(*)})}$  is called the Hastings Ratio. This choice of  $A(\theta^{(t)}, \theta^{(*)})$  ensures that the MH-MCMC algorithm outputs a sample from the posterior.

Because of how  $\theta^{(*)}$  is created in equation 3, the Hastings Ratio is equal to 1 [12]. Therefore, the acceptance probability can be computed by the following formula:

$$A(\theta^{(t)}, \theta^{(*)}) = \min\left\{1, \frac{p(\Theta^{(*)}|y_1 \dots y_n)}{p(\Theta^{(t)}|y_1 \dots y_n)}\right\}. \quad (4)$$

Again, this acceptance probability ensures that the Markov chain converges to samples from the posterior distribution by the Metropolis Hastings algorithm.

If  $\Theta$  is a  $K$ -dimensional parameter space (i.e.,  $\Theta = (\theta_1, \dots, \theta_K)$ ), we can move from the current state  $(\theta_1^{(t)}, \dots, \theta_k^{(t)}, \theta_{k+1}^{(t)}, \dots, \theta_K^{(t)})$  to  $(\theta_1^{(t+1)}, \dots, \theta_k^{(t+1)}, \theta_{k+1}^{(t)}, \dots, \theta_K^{(t)})$  by  $k$  successive sub-steps. This can be seen as follows:

Step 1: propose to go from  $(\theta_1^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_K^{(t)})$  to  $(\theta_1^{(*)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_K^{(t)})$   
Step 2: propose to go from  $(\theta_1^{(t+1)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_K^{(t)})$  to  $(\theta_1^{(t+1)}, \theta_2^{(*)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}, \dots, \theta_K^{(t)})$   
 $\vdots$   
Step  $k$ : propose to go from  $(\theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, \theta_k^{(t)}, \dots, \theta_K^{(t)})$  to  $(\theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, \theta_k^{(*)}, \dots, \theta_K^{(t)})$

where

$$\theta_k^{(t+1)} = \begin{cases} \theta_k^{(t)}, & \text{if the } k\text{-th move is rejected} \\ \theta_k^{(*)}, & \text{if the } k\text{-th move is accepted} \end{cases}$$

It can be observed that the  $k$ -th sub-step proposes to move from  $\Theta_A$  to  $\Theta_B$ , where

$$\begin{aligned} \Theta_A &:= (\theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, \theta_k^{(t)}, \theta_{k+1}^{(t)}, \dots, \theta_K^{(t)}) \\ \Theta_B &:= (\theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, \theta_k^{(*)}, \theta_{k+1}^{(t)}, \dots, \theta_K^{(t)}) \end{aligned} \quad (5)$$

Eventually, this algorithm generates dependent sequences of vectors:

$$\begin{aligned} \Theta^{(1)} &= \{\theta_1^{(1)}, \dots, \theta_K^{(1)}\} \\ \Theta^{(2)} &= \{\theta_1^{(2)}, \dots, \theta_K^{(2)}\} \\ &\vdots \\ \Theta^{(2T)} &= \{\theta_1^{(2T)}, \dots, \theta_K^{(2T)}\} \end{aligned}$$

We have then obtained a chain from the posterior  $p(\theta_1, \dots, \theta_K | y_1, \dots, y_n)$ , where the chain "walks" through  $2T$  states and the parameter space consists of  $K$  parameters. From here, we can approximate what we are interested in by means of a Monte Carlo approximation.

If it is possible to compute the full conditional distribution of the  $k$ -th parameter, than the  $k$ -th sub-step (as in equation 5) can be replaced by a more efficient Gibbs move.

### 2.1.2 Gibbs sampling

As mentioned before, Gibbs sampling is another method to construct a Markov chain that converges to samples from the posterior distribution. Although Gibbs sampling can not be used for inference of the model in this research, a full explanation of the procedure can be found in Appendix E, since it is an efficient method in case it can be applied.

### 3 Data

As mentioned in the introduction, the Raven Progressive Matrices test (RPM) was used to measure the level of intelligence of an individual in the thesis by Otermann. The twins were tested with the Standard Progressive Matrices (SPM) and the parents with the Advanced Progressive Matrices (APM). The SPM contains 60 items in total, distributed over 5 sets ( $A - E$ ) with 12 items per set. The APM contains only 36 items. The difficulty parameters  $\beta_k$  and  $\beta_{\bar{k}}$  of the SPM and APM, respectively, can be found in the Appendix A.

Since the data set that Otermann used in her thesis is strictly confidential, it has been worked with a synthetic parent-offspring data set in this research instead. Prof. dr. Van den Berg from the University of Twente has provided this data set. It contains the score on each item of the APM and SPM for parents and twins from 48 families. The twins are all monozygotic. The data of the participants on the SPM and APM is dichotomous, since the items on these tests are multiple choice questions. A fraction of the data set can be found in figure 2 to give an idea of what the data set looks like. Note that this figure contains only  $5 \cdot 36 = 180$  data points, whereas the complete data set contains  $(2 \cdot 36 \cdot 48) + (2 \cdot 60 \cdot 48) = 9216$  data points.

As can be seen in figure 2, not every score on an item is available in the data set. This is indicated with the letters NA (non available). Possible explanations are a divorce or the passing away of a parent, so that all scores of this parent have become unavailable. Occasionally, there is an unavailable score of a twin in the data set as well. It might be the case that a twin simply did not understand the particular question or forgot to make it, since NA scores only occur sporadically for twins.

|    | 1_parent1 | 2_parent1 | 3_parent1 | 4_parent1 | 5_parent1 |
|----|-----------|-----------|-----------|-----------|-----------|
| 1  | NA        | NA        | NA        | NA        | NA        |
| 2  | 0         | 1         | 1         | 1         | 1         |
| 3  | 0         | 1         | 1         | 1         | 1         |
| 4  | 0         | 1         | 1         | 1         | 1         |
| 5  | 1         | 1         | 1         | 1         | 1         |
| 6  | 1         | 1         | 1         | 1         | 0         |
| 7  | NA        | NA        | NA        | NA        | NA        |
| 8  | 1         | 1         | 1         | 0         | 0         |
| 9  | 1         | 1         | 1         | 1         | 1         |
| 10 | 1         | 1         | 0         | 1         | 1         |
| 11 | 1         | 1         | 1         | 1         | 1         |
| 12 | 1         | 1         | 1         | 1         | 0         |
| 13 | 1         | 1         | 1         | 1         | 1         |
| 14 | 1         | 1         | 1         | 1         | 1         |
| 15 | 1         | 1         | 0         | 1         | 0         |
| 16 | 1         | 1         | 1         | 1         | 1         |
| 17 | 1         | 1         | 1         | 1         | 1         |
| 18 | 1         | 1         | 1         | 1         | 0         |
| 19 | 1         | 1         | 1         | 1         | 1         |
| 20 | 1         | 1         | 1         | 0         | 0         |
| 21 | 1         | 1         | 1         | 1         | 1         |
| 22 | 1         | 1         | 0         | 1         | 0         |
| 23 | NA        | NA        | NA        | NA        | NA        |
| 24 | NA        | NA        | NA        | NA        | NA        |
| 25 | 1         | 1         | 1         | 1         | 1         |
| 26 | 1         | 1         | 1         | 1         | 1         |
| 27 | 1         | 1         | 1         | 1         | 1         |
| 28 | 1         | 0         | 1         | 0         | 0         |
| 29 | 1         | 1         | 1         | 1         | 1         |
| 30 | 1         | 1         | 1         | 1         | 1         |
| 31 | 1         | 1         | 1         | 1         | 1         |
| 32 | 1         | 1         | 1         | 1         | 0         |
| 33 | 1         | 1         | 0         | 1         | 0         |
| 34 | 1         | 1         | 0         | 1         | 1         |
| 35 | 1         | 1         | 1         | 1         | 1         |
| 36 | NA        | NA        | NA        | NA        | NA        |

Figure 2: Fraction of the synthetic data set that is provided by Prof. dr. Van den Berg. This fraction shows the scores on all 36 items of the SPM that are answered by the fathers from the first 5 families. A "1" indicates a correct answer, whereas a "0" indicates a wrong answer. "NA" indicates that the score on that particular item is not available.

## 4 Parent-offspring model

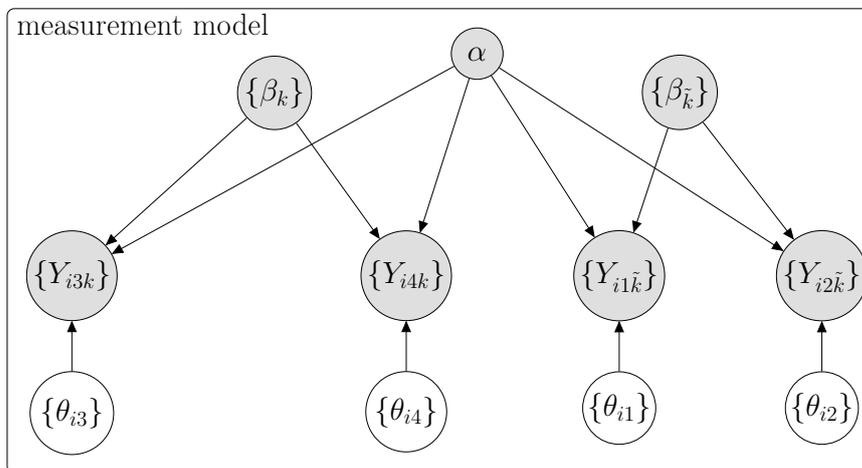
Now that the theoretical part of this research has been covered and the data has been explained, this chapter will describe the study design. The complete parent-offspring model that describes the transmission of intelligence consists of two parts: the measurement model and the genetic model. These models will be described successively in the following sections.

### 4.1 The measurement model

The first part of the complete parent-offspring model is the measurement model. We use the Rasch model as our measurement model in this research, which is a standard model from Item Response Theory (IRT) model for dichotomous data [18]. IRT describes the relationship between item responses from an individual and the characteristics of the test [15].

Local independence is one of the assumptions of the Rasch model. This means that the probability of a correct answer is explained by the ability score and the difficulty of that item instead of by previously answered items or other persons who answered the items. Therefore, response data of individuals can be compared independent of individuals and tests that are used. Moreover, the Rasch model separates the influence of item difficulty and ability level. Figure 3 provides a graphical representation of the measurement model.

As can be seen in this graphical representation, the measurement model consists of several variables that influence the response variables. Each response variable  $Y_{ijk}$  represents the score on the  $k$ -th question from the test that is answered by the  $j$ -th member from family  $i$ , where  $i = 1, \dots, N$ . In this research, the father and mother of a family are indicated by  $j = 1$  and  $j = 2$ , respectively. Moreover, the first born twin is indicated by  $j = 3$  and called twin 1, whereas the second born twin is indicated by  $j = 4$  and called twin 2. Since the RPM is a test that consists only of multiple choice questions, each  $Y_{ijk}$  takes on a value of either 0 or 1, corresponding to a wrong or a right answer, respectively.



**Figure 3:** Graphical representation of the measurement model, where  $Y_{ijk}$  are the item scores of each individual,  $\beta_k$  are the difficulty values,  $\alpha$  is the scale parameter and  $\theta_{ij}$  are the ability values. Parameters that have to be inferred (i.e., the free parameters) are represented by white circles. The data and the fixed parameters (i.e., the observed parameters) are represented by grey circles.

Moreover, the subscript  $k$  denotes the  $k$ -th question on the Raven Progressive Matrices test (RPM) in this study. Recall that the twins were tested with the Standard Progressive Matrices and the parents with the Advanced Progressive Matrices. Because of these two different test versions, we use two subscripts in our model to denote the  $k$ -th question on the test. We use  $k$  to indicate an item on the SPM. Likewise, we use  $\tilde{k}$  to indicate an item on the APM.

Furthermore, we see from the graphical representation of the measurement model that the response variables are influenced by the parameters  $\beta_k$ ,  $\alpha$  and  $\theta_{ij}$ . Here,  $\beta_k$  represents the difficulty of the  $k$ -th item on the test. The parameter  $\alpha$  is a scale parameter and is fixed to 1. Finally, the parameter  $\theta_{ij}$  indicates the ability score of the  $j$ -th member from family  $i$ .

An overview of all subscripts and parameters of the measurement model can be found in tables 1 and 2, respectively.

| Subscript   | Description   |
|-------------|---|
| $i$         | Family  |
| $k$         | Item from the SPM, where $k = 1, \dots, 60$         |
| $\tilde{k}$ | Item from the APM, where $\tilde{k} = 1, \dots, 36$ |
| $j = 1$     | Father  |
| $j = 2$     | Mother  |
| $j = 3$     | First born twin                                     |
| $j = 4$     | Second born twin                                    |

**Table 1: Descriptions of the subscripts of the measurement model, as in figure 3.**

| Parameter           | Description  |
|---------------------|--|
| $Y_{i1\tilde{k}}$   | The score of the father from family $i$ on the $\tilde{k}$ -th item of the APM |
| $Y_{i2\tilde{k}}$   | The score of the mother from family $i$ on the $\tilde{k}$ -th item of the APM |
| $Y_{i3k}$           | The score of twin 1 from family $i$ on the $k$ -th item of the test            |
| $Y_{i4k}$           | The score of twin 2 from family $i$ on the $k$ -th item of the test            |
| $\beta_k$           | Difficulty of the $k$ -th item of the SPM                                      |
| $\beta_{\tilde{k}}$ | Difficulty of the $\tilde{k}$ -th item of the APM                              |
| $\theta_{i1}$       | Ability score of the father from family $i$                                    |
| $\theta_{i2}$       | Ability score of the mother from family $i$                                    |
| $\theta_{i3}$       | Ability score of the first born twin from family $i$ (twin 1)                  |
| $\theta_{i4}$       | Ability score of the second born twin from family $i$ (twin 2)                 |
| $\alpha$            | Scale parameter. It is fixed to 1.   |

**Table 2: Descriptions of the parameters of the measurement model, as in figure 3.**

Since each item on the RPM is a multiple choice question, the response variables  $Y_{ijk}$  are Bernoulli distributed. The probability of a correct response is modeled as follows:

$$p_{ijk} = p(Y_{ijk} = 1 | \theta_{ij}, \beta_k, \alpha) = \frac{e^{\alpha\theta_{ij} - \beta_k}}{1 + e^{\alpha\theta_{ij} - \beta_k}}, \quad (6)$$

which is a logistic function. As a consequence, the probability of answering the item correctly is 50% in case the ability score of an individual is equal to the difficulty parameter of the item and  $\alpha$  is equal to 1. Also, the probability of answering the item correctly is smaller than 50% when  $\alpha$  is equal to 1 and  $\theta_{ij} < \beta_k$ , so when the ability of an individual is smaller than the difficulty of an item. Similarly, the probability of answering the item correctly is greater than 50% when  $\alpha$  is equal to 1 and  $\theta_{ij} > \beta_k$ , so when the ability of an individual is larger than the difficulty of an item.

Since the response variables  $Y_{ijk}$  are Bernoulli distributed, their probability mass function - which is the probability of an item being answered correctly or not - is given by the following expression:

$$p(Y_{ijk} = y) = p_{ijk}^y \cdot (1 - p_{ijk})^{1-y}, \quad \text{where } y \in \{0, 1\}.$$

It follows that the full likelihood, so the probability of the complete data set given our parameters, is equal to the following expression in our measurement model:

$$p(\{Y_{ijk}\} | \theta_{ij}, \beta_k, \alpha) = \prod_{i=1}^N \prod_{j=3}^4 \prod_{k=1}^K \left( \frac{e^{\alpha\theta_{ij} - \beta_k}}{1 + e^{\alpha\theta_{ij} - \beta_k}} \right)^{y_{ijk}} \cdot \left( 1 - \frac{e^{\alpha\theta_{ij} - \beta_k}}{1 + e^{\alpha\theta_{ij} - \beta_k}} \right)^{1-y_{ijk}} \cdot \prod_{i=1}^N \prod_{j=1}^2 \prod_{\tilde{k}=1}^{\tilde{K}} \left( \frac{e^{\alpha\theta_{ij} - \beta_{\tilde{k}}}}{1 + e^{\alpha\theta_{ij} - \beta_{\tilde{k}}}} \right)^{y_{ij\tilde{k}}} \cdot \left( 1 - \frac{e^{\alpha\theta_{ij} - \beta_{\tilde{k}}}}{1 + e^{\alpha\theta_{ij} - \beta_{\tilde{k}}}} \right)^{1-y_{ij\tilde{k}}}. \quad (7)$$

As before,  $k$  indicates one of the 60 items from the SPM and  $\tilde{k}$  indicates one of the 36 items from the APM. Hence,  $k$  runs from 1 to 60 and  $\tilde{k}$  runs from 1 to 36.

So, this is the expression of the likelihood that is used in the Metropolis Hastings algorithm that corresponds to our model. This likelihood applies to the data that has been described in the previous chapter.

## 4.2 The genetic model

Now that the first part of the complete parent-offspring model has been described, an overview of the second part will be given, which is the genetic model. It will be based on the thesis by Otermann.

Classical studies with a parent-twin design assume that the phenotype of an individual (denoted by  $P$ ) is dependent on additive genetic effects (denoted by  $A$ ) and environmental effects (denoted by  $E$ ). This can be expressed as follows:

$$P = h \cdot A + e \cdot E,$$

where  $h$  and  $e$  are the factor loadings for the  $A$  and  $E$  components, respectively. Since we are modeling the transmission of intelligence from parents to their children, the phenotype of an individual corresponds to his or her level of intelligence in this study. The variation in level of intelligence between individuals is a result of genotypic and phenotypic components. This is expressed as follows [18]:

$$\begin{aligned} \text{Var}(P) &= \text{Var}(hA) + \text{Var}(eE) + 2\text{Cov}(hA, eE) \\ &= h^2\text{Var}(A) + e^2\text{Var}(E) + 2he\text{Cov}(A, E). \end{aligned} \quad (8)$$

As mentioned in the introduction, cultural transmission might take place in the transmission of intelligence. We denote this cultural transmission by  $z$ . The genetic effects and the environmental effects will be correlated because of cultural transmission. The covariance between these variables is denoted by  $s$ . So,  $s = \text{Cov}(A, E)$ . Moreover,  $\text{Var}(A) = \text{Var}(E) = 1$ . It follows that equation 8 can be written as

$$\text{Var}(P) = h^2 + e^2 + 2hes.$$

Also, it has been shown that  $s$  can be written in terms of  $\rho$ ,  $z$  and  $h$  in previous studies [8]:

$$s = \frac{(1 + \rho)zh}{1 - (1 + \rho)ze}, \quad (9)$$

where  $\rho$  denotes the phenotypic correlation between parents. This phenotypic correlation is a result of the assumption of phenotypic assortment. In turn, this leads to a correlation between the genotypes and the environmental factors. We denote the correlation in genotype between parents by  $\gamma$ . It has been shown that this correlation is equal to  $\rho(h + se)^2$  [6]. Moreover, the correlation between environmental factors of parents can be denoted by  $\epsilon$ . The correlation between the genotype of the first parent and the environmental factors of the second parent of a couple can be denoted by  $\delta$ .

The parameter  $\rho$  can be used to model the parental phenotypes. The phenotypes of the individuals, i.e., their level of intelligence (ability score), correspond to the  $\theta_{ij}$  parameters that have already been introduced in the measurement model. So, these parameters form the link between the measurement model and the genetic model. The parental phenotypes are again given by  $\theta_{i1}$  and  $\theta_{i2}$ . All other notations - such as subscripts - are similar to the notations in the measurement model as well. The parental phenotypes, together with the parameters  $h$ ,  $s$  and  $e$ , can then be used to predict the parental genotypes, which are given by  $A_{i1}$  and  $A_{i2}$ :

$$\begin{aligned} \eta_i &\sim N(\mu_p, \rho) \\ \theta_{i1} &\sim N(\eta_i, (1 - \rho)) \\ \theta_{i2} &\sim N(\eta_i, (1 - \rho)) \\ A_{i1} &\sim N((h + se)\theta_{i1}, \sigma_{Ares}^2) \\ A_{i2} &\sim N((h + se)\theta_{i2}, \sigma_{Ares}^2) \end{aligned}$$

where  $i = 1, \dots, 48$ ,  $\eta_i$  denotes the parental average phenotype of family  $i$  and  $\sigma_{Ares}^2$  is equal to  $1 - (h + se)^2$ .

| Parameter     | Description   | Distribution   |
|---------------|---|--|
| $A_i$         | Genotype of the twins from family $i$                 | $N((0.5 \cdot (A_{i1} + A_{i2}) + \mu_0), \sigma_{MS}^2)$    |
| $E_{i3}$      | Environmental factor of twin 1 from family $i$        | $N(z \cdot (\theta_{i1} + \theta_{i2}), \sigma_{E_{res}}^2)$ |
| $E_{i4}$      | Environmental factor of twin 2 from family $i$        | $N(z \cdot (\theta_{i1} + \theta_{i2}), \sigma_{E_{res}}^2)$ |
| $A_{i1}$      | Genotype of the father from family $i$                | $N((h + se)\theta_{i1}, \sigma_{A_{res}}^2)$                 |
| $A_{i2}$      | Genotype of the mother from family $i$                | $N((h + se)\theta_{i2}, \sigma_{A_{res}}^2)$                 |
| $\theta_{i1}$ | Ability score of the father from family $i$           | $N(\eta_i, (1 - \rho))$                                      |
| $\theta_{i2}$ | Ability score of the mother from family $i$           | $N(\eta_i, (1 - \rho))$                                      |
| $\eta_i$      | Parental average phenotype of family $i$              | $N(\mu_p, \rho)$   |
| $\mu_p$       | Mean component of the parents                         | $N(0, \sqrt{10})$  |
| $\mu_0$       | Mean component of the twins                           | $N(0, \sqrt{10})$  |
| $h$           | Loading factor of the genetic effects component       | $U(0, 4)$  |
| $e$           | Loading factor of the environmental effects component | $U(0, 4)$  |
| $\rho$        | Phenotypic correlation between parents                | $U(0, 1)$  |
| $z$           | Cultural transmission                                 | $U(-1, 1)$   |

**Table 3: Distributions of the free variables in the parent-offspring model for cultural and genetic transmission, as in figure 4. In  $N(\mu, \sigma^2)$ ,  $\mu$  indicates the mean and  $\sigma^2$  indicates the variance. In  $U(a, b)$ ,  $a$  indicates the lower bound and  $b$  indicates the upper bound.**

Then, the parental genotypes are used to predict the genotypes of their children. This genotype is given by  $A_i$  for both twins, since the twins are monozygotic.

$$A_i \sim N((0.5 \cdot (A_{i1} + A_{i2}) + \mu_0), \sigma_{MS}^2).$$

Here,  $i = 1, \dots, 48$ ,  $\mu_0$  denotes the individual mean contribution of the genotypes of the offspring that is not transmitted from parent to child and  $\sigma_{MS}^2$  is equal to  $0.5 - 0.5\gamma$  [18]. Moreover, the parental phenotypes are used to predict the environmental factors of the offspring (given by  $E_{i3}$  and  $E_{i4}$ ):

$$E_{ij} \sim N(z \cdot (\theta_{i1} + \theta_{i2}), \sigma_{E_{res}}^2)$$

where  $j = 3, 4$ ,  $i = 1, \dots, 48$  and  $\sigma_{E_{res}}^2$  is equal to  $1 - 2(z^2 + pz^2)$ .

Moreover, the phenotypes of the twins are given by  $\theta_{i3}$  and  $\theta_{i4}$ . These parameters are given by the following equation:

$$\theta_{ij} = h \cdot A_i + e \cdot E_{ij} \quad (10)$$

where  $j = 3, 4$  and  $i = 1, \dots, 48$ . Finally, the following prior distributions have been used:

$$\begin{aligned} \rho &\sim U(0, 1) \\ z &\sim U(-1, 1) \\ e &\sim U(0, 4) \\ h &\sim U(0, 4) \\ \mu_p &\sim N(0, \sqrt{10}) \\ \mu_o &\sim N(0, \sqrt{10}) \end{aligned}$$

An overview of all free and fixed parameters together with their description and distribution can be found in table 3 and table 4, respectively. Moreover, a graphical representation of the genetic model is provided in figure 4.

| Parameter         | Description  | Definition                        |
|-------------------|--|-----------------------------------|
| $\theta_{i3}$     | Ability score of the first born twin from family $i$ (twin 1)  | $h \times A_i + e \times E_{i3}$  |
| $\theta_{i4}$     | Ability score of the second born twin from family $i$ (twin 2) | $h \times A_i + e \times E_{i4}$  |
| $\sigma_{Eres}^2$ | Variance of $E_{i3}$ and $E_{i4}$                              | $1 - 2(z^2 + \rho z^2)$           |
| $\sigma_{Ares}^2$ | Variance of $A_{i1}$ and $A_{i2}$                              | $1 - (h + se)^2$                  |
| $s$               | Covariance between genetic effects and environmental effects   | $\frac{(1+\rho)zh}{1-(1+\rho)ze}$ |
| $\gamma$          | Correlation in genotype between parents                        | $\rho(h + se)^2$                  |
| $\sigma_{MS}^2$   | Variance of $A_i$  | $0.5 - 0.5\gamma$                 |

Table 4: Definitions of the variables in the parent-offspring model for cultural and genetic transmission that depend deterministically on their parents nodes, as in figure 4.

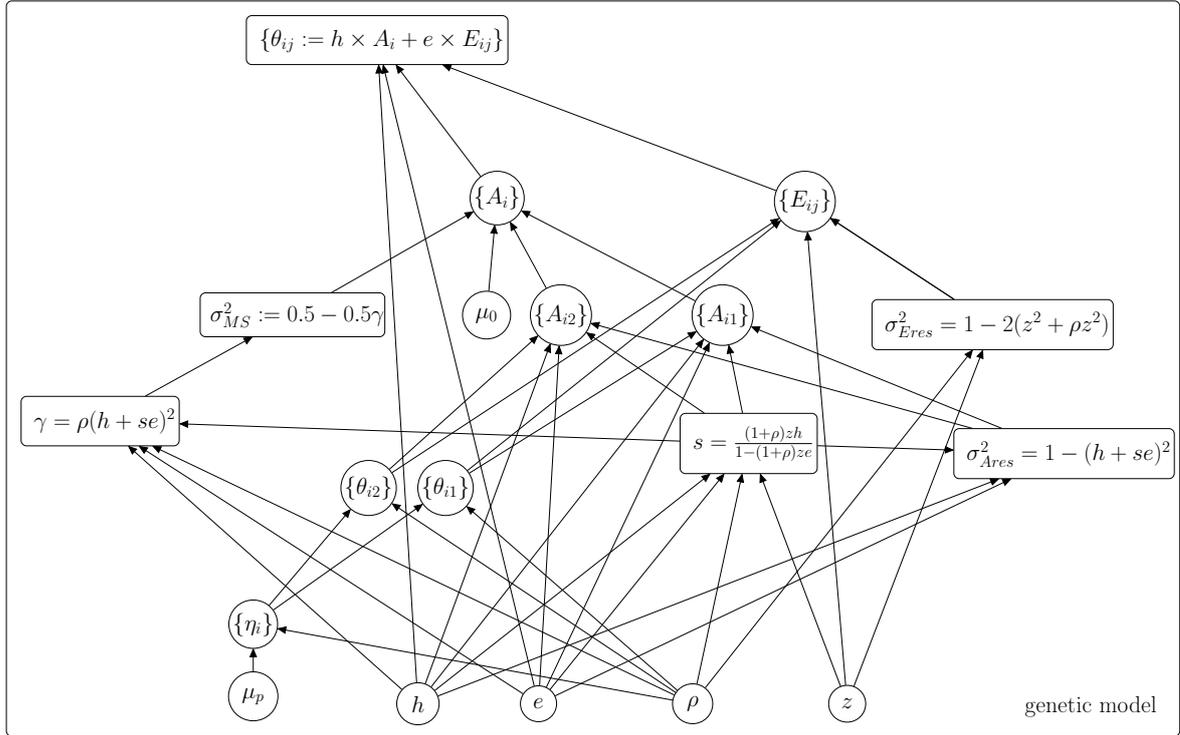


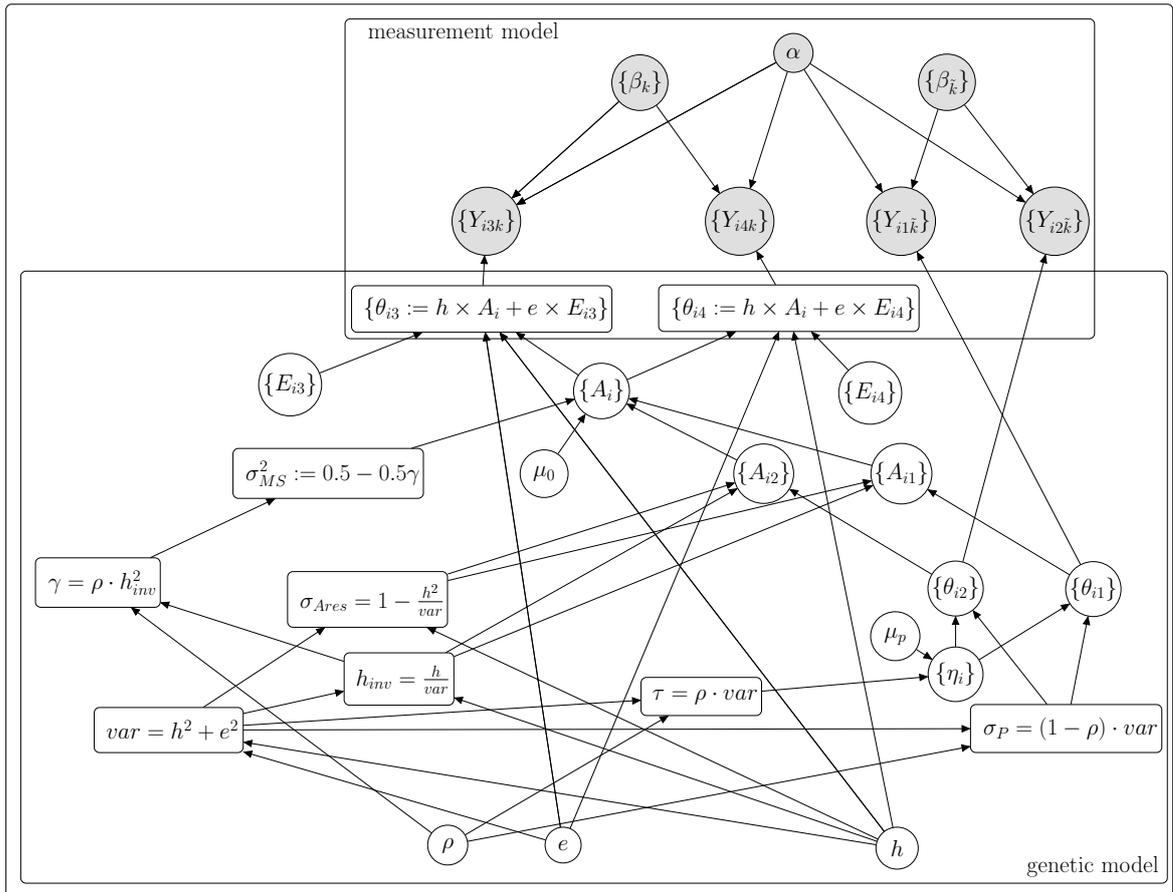
Figure 4: Graphical representation of the genetic model. Parameters that have to be inferred (i.e., the free parameters) are represented by white circles. The data and the fixed parameters (i.e., the observed parameters) are represented by grey circles. The six rectangles indicate definitions, which deterministically depend on the parent nodes. The symbol  $\times$  indicates a multiplication.

### 4.3 The complete parent-offspring model

By merging both parts of the model we obtain our complete parent-offspring model. In the thesis by Otermann, different versions of this model have been compared. One of them assumes phenotypic assortment and no cultural transmission. This is the model that will be worked with in the remainder of this current research. It will be referred to as model 1 in the remainder of this research. A graphical representation is provided in figure 5. Note that model 1 consists of two parts, namely the measurement model and the genetic model.

As model 1 does not include cultural transmission, the parameter  $z$  is fixed to zero. It follows from equation 9 that the parameter  $s$  becomes zero as well. Moreover, it turns out that there is an identification problem of the model. Therefore, some parameters and variance components needed to be fixed [18]. This leads to a slight adjustment of the genetic part of the model that has been introduced in the previous section. A graphical representation of the genetic part of model 1 can be found in figure 19 in Appendix C. The measurement model remains the same. Hence, model 1 differs only from the most general model in the genetic part. It has been derived by analyzing the JAGS code for the model that has been included in the thesis by Otermann.

An overview of all free and fixed parameters in model 1 together with their description and distribution can be found in table 5 and table 6, respectively. The MCMC simulations in the upcoming chapter will be based on this model.



**Figure 5: Graphical representation of model 1.** Parameters that have to be inferred (i.e., the free parameters) are represented by white circles. The data and the fixed parameters (i.e., the observed parameters) are represented by grey circles. The nine rectangles indicate definitions, which deterministically depend on the parent nodes. The symbol  $\times$  indicates a multiplication.

| Parameter     | Description   | Distribution  |
|---------------|---|---|
| $A_i$         | Genotype of the twins from family $i$                 | $N((0.5 \cdot (A_{i1} + A_{i2}) + \mu_0), \sigma_{MS}^2)$ |
| $E_{i3}$      | Environmental factor of twin 1 from family $i$        | $N(0, 1)$   |
| $E_{i4}$      | Environmental factor of twin 2 from family $i$        | $N(0, 1)$   |
| $A_{i1}$      | Genotype of the father from family $i$                | $N(h_{inv} \cdot \theta_{i1}, \sigma_{Ares}^2)$           |
| $A_{i2}$      | Genotype of the mother from family $i$                | $N(h_{inv} \cdot \theta_{i2}, \sigma_{Ares}^2)$           |
| $\theta_{i1}$ | Ability score of the father from family $i$           | $N(\eta, \sigma_P^2)$                                     |
| $\theta_{i2}$ | Ability score of the mother from family $i$           | $N(\eta, \sigma_P^2)$                                     |
| $\eta_i$      | Parental average phenotype of family $i$              | $N(\mu_p, \tau)$  |
| $\mu_p$       | Mean component of the parents                         | $N(0, 10)$  |
| $\mu_0$       | Mean component of the twins                           | $N(0, 10)$  |
| $h$           | Loading factor of the genetic effects component       | $U(0, 4)$   |
| $e$           | Loading factor of the environmental effects component | $U(0, 4)$   |
| $\rho$        | Phenotypic correlation between parents                | $U(0, 1)$   |

**Table 5: Distributions of the free variables in model 1, as in figure 5. In  $N(\mu, \sigma^2)$ ,  $\mu$  indicates the mean and  $\sigma^2$  indicates the variance. In  $U(a, b)$ ,  $a$  indicates the lower bound and  $b$  indicates the upper bound.**

| Parameter         | Description                                 | Definition                       |
|-------------------|---|----------------------------------|
| $\theta_{i3}$     | Ability score of twin 1 from family $i$     | $h \times A_i + e \times E_{i3}$ |
| $\theta_{i4}$     | Ability score of twin 2 from family $i$     | $h \times A_i + e \times E_{i4}$ |
| $\sigma_P^2$      | Variance of $\theta_{i1}$ and $\theta_{i2}$ | $(1 - \rho) \cdot var$           |
| $\tau$            | Variance of $\eta$                          | $\rho \cdot var$                 |
| $h_{inv}$         |   | $\frac{h^2}{var}$                |
| $var$             |   | $h^2 + e^2$                      |
| $\sigma_{Ares}^2$ | Variance of $A_{i1}$ and $A_{i2}$           | $1 - \frac{h^2}{var}$            |
| $\gamma$          | Genotypic correlation between parents       | $\rho \cdot h_{inv}^2$           |
| $\sigma_{MS}^2$   | Variance of $A_i$                           | $0.5 - 0.5\gamma$                |

**Table 6: Definitions of the variables in model 1 that depend deterministically on their parents nodes, as in figure 5.**

## 5 Metropolis Hastings algorithm applied

Now that the relevant theory has been explained and the parent-offspring model has been given that will be worked with, it will be described how the Metropolis Hastings algorithm has been applied to generate samples from the posterior distributions of this model.

Since the distributions of all free variables in figure 5 have been given, Bayesian statistical modeling can be used to generate samples from the parameters of model 1. It is not possible to determine the posterior distributions of these parameters analytically, so this must be done by means of an MCMC inference method.

The way in which an MCMC method has been applied in this research differs from the way in which it has been applied in the research that has been conducted by Otermann. Whereas the Metropolis Hastings algorithm has been applied in this paper, Otermann made use of the JAGS package in R. JAGS (Just Another Gibbs Sampler) can compute the full conditional distributions of the parameters and produce the formula of the posterior distribution when this is possible. It is automatically switched to Metropolis Hastings sampling when the full conditional distribution can not be computed [7]. However, since the JAGS package functions as a black-box and it is not possible to derive the full conditional distributions for the parameters in our model, we have focused on a direct implementation of the Metropolis Hastings algorithm in this paper instead.

The implementation of the Metropolis-Hastings algorithm of our model starts with the initial values of the free parameters:

$$\Theta^{(1)} = \{\{A_i^{(1)}\}, \{E_{i3}^{(1)}\}, \{E_{i4}^{(1)}\}, \{A_{i2}^{(1)}\}, \{A_{i1}^{(1)}\}, \{\theta_{i1}^{(1)}\}, \{\theta_{i2}^{(1)}\}, \{\eta_i^{(1)}\}, \mu_0^{(1)}, \mu_p^{(1)}, h^{(1)}, e^{(1)}, \rho^{(1)}\}.$$

These initialization values can be chosen arbitrarily. Then, we move in 13 sub-steps from  $\Theta^{(t)}$  to  $\Theta^{(t+1)}$ , where

$$\Theta^{(t)} := \{\{A_i^{(t)}\}, \{E_{i3}^{(t)}\}, \{E_{i4}^{(t)}\}, \{A_{i2}^{(t)}\}, \{A_{i1}^{(t)}\}, \{\theta_{i1}^{(t)}\}, \{\theta_{i2}^{(t)}\}, \{\eta_i^{(t)}\}, \mu_0^{(t)}, \mu_p^{(t)}, h^{(t)}, e^{(t)}, \rho^{(t)}\},$$

since there are 13 free parameters for each family  $i$ . We go through these sub-steps for  $t = 2, \dots, 2T$ . In each sub-step, it is proposed for only one parameter, say  $\mu_p$ , to move from  $\mu_p^{(t)}$  to  $\mu_p^{(*)}$ , as in equation 5. Recall that the acceptance probability of this move is given by

$$A(\mu_p^{(t)}, \mu_p^{(*)}) = \min\left\{1, \frac{p(\tilde{\Theta}^{(*)}|y_1 \dots y_n)}{p(\tilde{\Theta}^{(t)}|y_1 \dots y_n)}\right\},$$

where

$$\begin{aligned} \tilde{\Theta}^{(t)} &= (A_i^{(t+1)}, \dots, \mu_0^{(t+1)}, \mu_p^{(t)}, h^{(t)}, \dots, \rho^{(t)}) \\ \text{and } \tilde{\Theta}^{(*)} &= (A_i^{(t+1)}, \dots, \mu_0^{(t+1)}, \mu_p^{(*)}, h^{(t)}, \dots, \rho^{(t)}). \end{aligned}$$

Therefore, we must know what the joint posterior distribution is proportional to in order to compute the ratio of the posterior distributions in the formula of the acceptance probability.

Since the posterior is proportional to the likelihood times the priors, it follows from figure 5 that:

$$\begin{aligned}
& p(\{A_i\}, \{E_{i3}\}, \{E_{i4}\}, \mu_0, \mu_p, \{\eta_i\}, \{A_{i2}\}, \{A_{i1}\}, \{\theta_{i2}\}, \{\theta_{i1}\}, h, e, \rho | \{Y_{i1\tilde{k}}\}, \{Y_{i2\tilde{k}}\}, \{Y_{i3k}\}, \{Y_{i4k}\}) \\
& \propto \prod_{i=1}^N \prod_{j=3}^4 \prod_{k=1}^K p(Y_{ijk} | A_i, E_{ij}, h, e) \cdot \prod_{i=1}^N \prod_{j=1}^2 \prod_{\tilde{k}=1}^{\tilde{K}} p(Y_{ij\tilde{k}} | \theta_{i1}, \theta_{i2}) \cdot \prod_{i=1}^N p(A_i | \mu_0, A_{i2}, A_{i1}, h, e, \rho) \\
& \cdot \prod_{i=1}^N p(E_{i3}) \cdot \prod_{i=1}^N p(E_{i4}) \cdot \prod_{i=1}^N p(A_{i2} | \theta_{i2}, h, e) \cdot \prod_{i=1}^N p(A_{i1} | \theta_{i1}, h, e) \cdot \prod_{i=1}^N p(\theta_{i2} | \eta_i, \rho, h, e) \\
& \cdot \prod_{i=1}^N p(\theta_{i1} | \eta_i, \rho, h, e) \cdot p(\eta_i | \mu_p, \rho, h, e) \cdot p(\mu_p) \cdot p(\mu_0) \cdot p(h) \cdot p(e) \cdot p(\rho),
\end{aligned} \tag{11}$$

where  $K = 60$ ,  $\tilde{K} = 36$  and  $N = 48$ .

The term on the right hand side of the  $\propto$  sign is called the score-function. We will use this score function instead of the posterior distribution in the computation of the acceptance probability in the Metropolis Hastings algorithm, since the constant terms in the posteriors cancel and only the scores remain.

Moreover, since the probabilities of the response variable  $Y_{ijk}$  as well as the probabilities of a response variables  $Y_{ij\tilde{k}}$  are between 0 and 1 (conform equation 6), the logarithms of the scores are used in the computation of the acceptance probability in the Metropolis Hastings algorithm. It follows that

$$\begin{aligned}
& \log \left[ p(\{A_i\}, \{E_{i3}\}, \{E_{i4}\}, \mu_0, \mu_p, \{\eta_i\}, \{A_{i2}\}, \{A_{i1}\}, \{\theta_{i2}\}, \{\theta_{i1}\}, h, e, \rho | \{Y_{i1\tilde{k}}\}, \{Y_{i2\tilde{k}}\}, \{Y_{i3k}\}, \{Y_{i4k}\}) \right] \\
& = \sum_{i=1}^N \sum_{j=3}^4 \sum_{k=1}^K \log(p(Y_{ijk} | A_i, E_{ij}, h, e)) + \sum_{i=1}^N \sum_{j=1}^2 \sum_{\tilde{k}=1}^{\tilde{K}} \log(p(Y_{ij\tilde{k}} | \theta_{i1}, \theta_{i2})) + \sum_{i=1}^N \log(p(A_i | \mu_0, A_{i2}, A_{i1}, h, e, \rho)) \\
& + \sum_{i=1}^N \log(p(E_{i3})) + \sum_{i=1}^N \log(p(E_{i4})) + \sum_{i=1}^N \log(p(A_{i2} | \theta_{i2}, h, e)) + \sum_{i=1}^N \log(p(A_{i1} | \theta_{i1}, h, e)) \\
& + \sum_{i=1}^N \log(p(\theta_{i2} | \eta_i, \rho, h, e)) + \sum_{i=1}^N \log(p(\theta_{i1} | \eta_i, \rho, h, e)) + \sum_{i=1}^N \log(p(\eta_i | \mu_p, \rho, h, e)) \\
& + \log(p(\mu_p)) + \log(p(\mu_0)) + \log(p(h)) + \log(p(e)) + \log(p(\rho)) + c,
\end{aligned} \tag{12}$$

where  $c$  is constant.

The logarithm of the score function is equal to the term on the right hand side of equation 12 minus this constant. This is what has been implemented in the code of the acceptance probability of the Metropolis Hastings algorithm.

One part of this element of the code is the likelihood function. It has already been given in equation 7. The probability of only one data point is given by

$$p(Y_{ijk} | \theta_{ij}, \beta_k, \alpha) = \left( \frac{e^{\alpha \theta_{ij} - \beta_k}}{1 + e^{\alpha \theta_{ij} - \beta_k}} \right)^{y_{ijk}} \cdot \left( 1 - \frac{e^{\alpha \theta_{ij} - \beta_k}}{1 + e^{\alpha \theta_{ij} - \beta_k}} \right)^{1 - y_{ijk}},$$

for  $k = k$  and  $k = \tilde{k}$ , so for twins as well as parents.

By applying logarithm rules and setting  $\alpha$  equal to 1, it follows that

$$\begin{aligned}
\log(p(Y_{ijk}|\theta_{ij}, \beta_k, \alpha)) &= \log\left(\left(\frac{e^{\alpha\theta_{ij}-\beta_k}}{1+e^{\alpha\theta_{ij}-\beta_k}}\right)^{y_{ijk}} \cdot \left(1-\frac{e^{\alpha\theta_{ij}-\beta_k}}{1+e^{\alpha\theta_{ij}-\beta_k}}\right)^{1-y_{ijk}}\right) \\
&= y_{ijk} \cdot \log\left(\frac{e^{\theta_{ij}-\beta_k}}{1+e^{\theta_{ij}-\beta_k}}\right) + (1-y_{ijk}) \cdot \log\left(1-\frac{e^{\theta_{ij}-\beta_k}}{1+e^{\theta_{ij}-\beta_k}}\right) \\
&= y_{ijk} \cdot \left(\log(e^{\theta_{ij}-\beta_k}) - \log(1+e^{\theta_{ij}-\beta_k})\right) + (1-y_{ijk}) \cdot \log\left(\frac{1}{1+e^{\theta_{ij}-\beta_k}}\right) \\
&= y_{ijk} \cdot (\theta_{ij} - \beta_k - \log(1+e^{\theta_{ij}-\beta_k})) + (1-y_{ijk}) \cdot \left(\log(1) - \log(1+e^{\theta_{ij}-\beta_k})\right) \\
&= y_{ijk} \cdot (\theta_{ij} - \beta_k - \log(1+e^{\theta_{ij}-\beta_k})) - (1-y_{ijk}) \cdot \log(1+e^{\theta_{ij}-\beta_k}).
\end{aligned}$$

This function is called the log-likelihood function.

As observed in chapter 3, not every score on an item is available in the data set. We only take into account the response variables  $Y_{ijk}$  with value 0 or 1 in our MCMC inference. The response variables with NA value are skipped, so unanswered questions are simply left out and we loop the MCMC simulations only over the available data points. We can do this, because we can assume that the data are missing at random. Another approach to deal with the missing data points would have been to treat them as unknown parameters and sample their posteriors.

Eventually, after applying the process of moving from  $\Theta^{(t)}$  to  $\Theta^{(t+1)}$  for  $t = 1, \dots, 2T$ , a dependent sequence is generated:

$$\begin{aligned}
\Theta^{(1)} &= \{\{A_i^{(1)}\}, \{E_{i3}^{(1)}\}, \{E_{i4}^{(1)}\}, \{A_{i2}^{(1)}\}, \{A_{i1}^{(1)}\}, \{\theta_{i1}^{(1)}\}, \{\theta_{i2}^{(1)}\}, \{\eta_i^{(1)}\}, \mu_0^{(1)}, \mu_p^{(1)}, h^{(1)}, e^{(1)}, \rho^{(1)}\} \\
\Theta^{(2)} &= \{\{A_i^{(2)}\}, \{E_{i3}^{(2)}\}, \{E_{i4}^{(2)}\}, \{A_{i2}^{(2)}\}, \{A_{i1}^{(2)}\}, \{\theta_{i1}^{(2)}\}, \{\theta_{i2}^{(2)}\}, \{\eta_i^{(2)}\}, \mu_0^{(2)}, \mu_p^{(2)}, h^{(2)}, e^{(2)}, \rho^{(2)}\} \\
&\vdots \\
\Theta^{(2T)} &= \{\{A_i^{(2T)}\}, \{E_{i3}^{(2T)}\}, \{E_{i4}^{(2T)}\}, \{A_{i2}^{(2T)}\}, \{A_{i1}^{(2T)}\}, \{\theta_{i1}^{(2T)}\}, \{\theta_{i2}^{(2T)}\}, \{\eta_i^{(2T)}\}, \mu_0^{(2T)}, \mu_p^{(2T)}, h^{(2T)}, e^{(2T)}, \rho^{(2T)}\}
\end{aligned}$$

Hence, the chain walks through  $2T$  states, where each  $\Theta^{(t)}$  is one state. The parameter space consists of 5 parameters without index  $i$  and of 8 parameters with index  $i$ , where  $i = 1, \dots, 48$ . Therefore, we have obtained the marginal posterior distributions of  $(8 \cdot 48) + 5 = 389$  parameters from the joint posterior distribution

$$p(\{A_i\}, \{E_{i3}\}, \{E_{i4}\}, \mu_0, \mu_p, \{\eta_i\}, \{A_{i2}\}, \{A_{i1}\}, \{\theta_{i2}\}, \{\theta_{i1}\}, h, e, \rho | \{Y_{i1\bar{k}}\}, \{Y_{i2\bar{k}}\}, \{Y_{i3k}\}, \{Y_{i4k}\}).$$

The samples from the marginal posterior distribution of the parameters can be used to approximate the quantity of interest of a specific parameter, such as the mean of some function of that parameter. Such Monte Carlo approximations will be done in an upcoming section to determine if the level of intelligence of an individual is dominated by genetic factors or environmental factors.

The complete R code of the implementation of the Metropolis Hastings algorithm can be found in Appendix D.

## 6 Results

### 6.1 Descriptive statistics

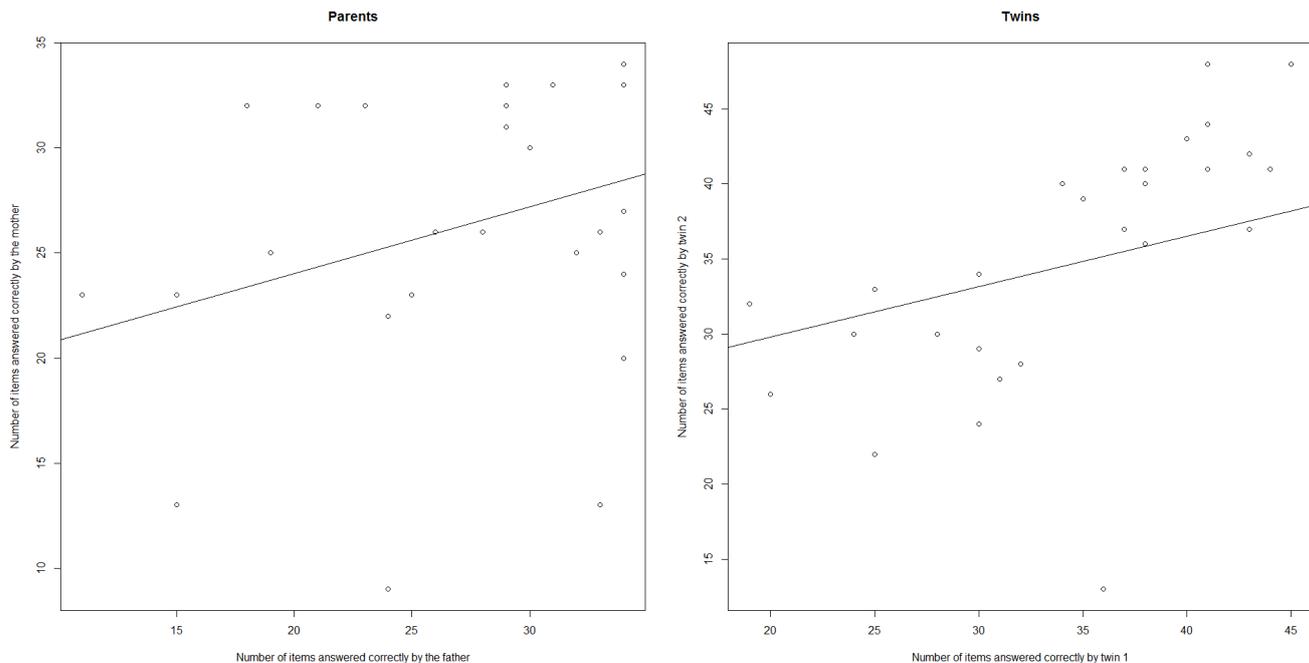
Some descriptive statistics of the synthetic data set are presented in table 7. Unavailable numbers have not been taken into account. The total score on the APM was unavailable for parents from 5 families and the total score on the SPM was unavailable for twins from 6 families. Hence,  $N = 43$  for both parents and  $N = 42$  for both twins.

|        | N  | Min | Max | Mean  | SD   |
|--------|----|-----|-----|-------|------|
| Father | 43 | 11  | 34  | 25.47 | 6.61 |
| Mother | 43 | 9   | 34  | 24.53 | 7.48 |
| Twin 1 | 42 | 17  | 45  | 34.07 | 6.86 |
| Twin 2 | 42 | 13  | 48  | 34.29 | 9.51 |

**Table 7: Descriptives for all individuals of the score on the RPM, where parents completed the APM (maximum score = 36) and twins received the SPM (maximum score = 60)**

Moreover, the population correlation coefficient between parents is computed from those 43 families with available item scores only. This coefficient corresponds to  $\rho$  in our parent-offspring model. It is equal to 0.36. This indicates a low positive to moderate positive correlation between the scores on the APM between parents [17] and corresponds reasonably to our assumption of phenotypic assortment.

Similarly, the population correlation coefficient between twins is computed from those 42 families with available item scores only. This correlation coefficient is equal to 0.51. This indicates that there is a moderate positive correlation between the scores on the SPM between twins [17]. The correlation coefficients are reflected in the graphs in figure 6.



**Figure 6: Scatter plots of the number of items answered correctly by the parents of a family and twins of a family. The straight lines are the regression lines.**

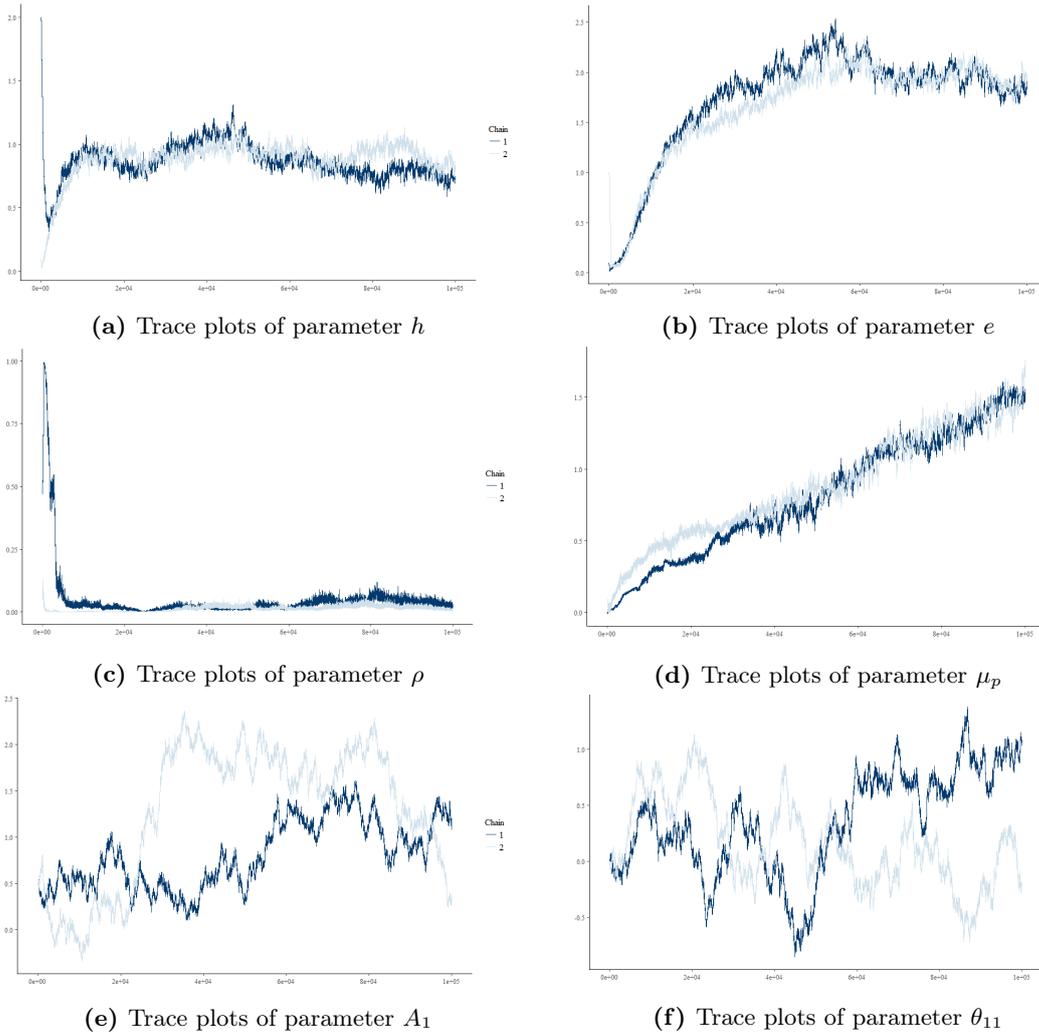
Other visualizations of the data set can be found in the Appendix B.

## 6.2 Convergence of Markov chains

As described in previous chapters, the Metropolis Hastings algorithm generates samples from the marginal posterior distributions of the parameters of a model. Trace plots are useful to assess how quickly the Markov chain that consists of those samples converges to the stationary distribution. They show the sampled values of a parameter over time.

Since we only want to use those samples from the converged part of the chain, i.e., the samples from the stationary distribution, it is common to withdraw its first states. This process is called burning in. So, when we perform Monte Carlo approximations with the sampled values, we only use those samples after the burn-in phase. A trace plot is useful to determine which samples we should withdraw and which samples we should take into account.

In figure 7, trace plots are shown of parameters  $h$ ,  $e$ ,  $\rho$ ,  $\mu_p$ ,  $A_1$  and  $\theta_{11}$  for 100.000 iterations. It follows from this figure that the Markov chains have not converged to a stationary distribution yet. This is evident in the trace plots of  $\mu_p$ , for example. The Markov chains are systematically increasing up to and including the 100.000-th state, which indicates that more iterations are necessary to reach the desired stationary distribution. It can also be observed from the trace plots of  $\theta_{11}$  that the Markov chains have not converged for 100.000 iterations. The chains do not take approximately the same path after a certain number of iterations lower than 100.000.

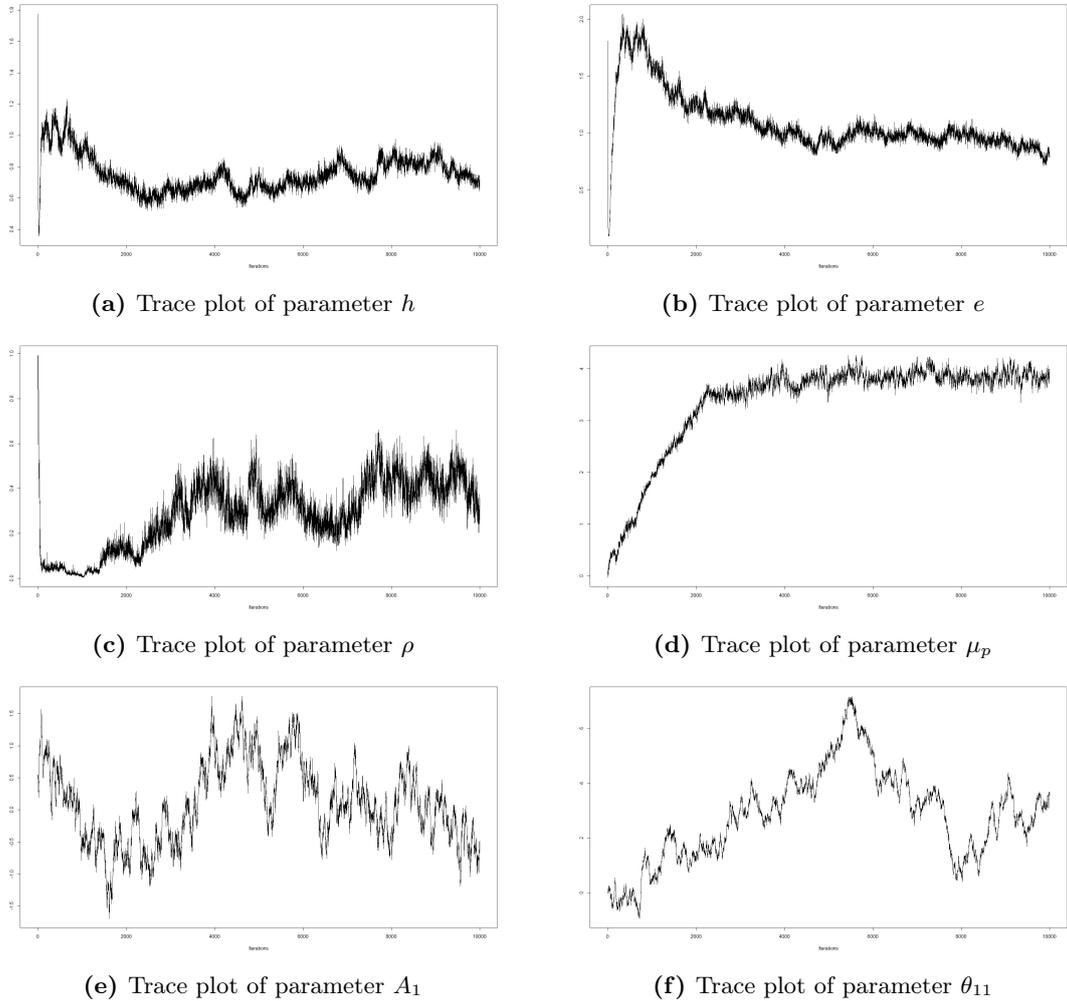


**Figure 7: Trace plots of various model parameters. The plots show two MCMC chains for 100.000 iterations.**

Since the Markov chains of a significant number of the 389 parameters have not converged yet for 100.000 iterations, we want to examine their behavior for more iterations. The aim was to run the code on a cluster of ERIBA for 1.000.000 iterations, as it took approximately 3 hours on a university computer to run it for 100.000 iterations. However, since the code was still running on the cluster after a few days, a high-end laptop has been used to realize the execution.

Samples from only one Markov chain have been collected for 1.000.000 iterations due to time constraints. It would have been favorable to compare samples from at least two chains to check if they converge to the same stationary distribution. For now, we will rely on the correctness of the convergence of the chains.

Moreover, since the output of the code takes a substantial amount of memory for 1.000.000 iterations, we only keep every 100-th state and discard all other samples. This process is known as thinning the Markov chain. The thinning factor is 100 en since we run the code for 1.000.000 iterations, only the samples from 10.000 states remain. In figure 8, trace plots are shown of parameters  $h$ ,  $e$ ,  $\rho$ ,  $\mu_p$ ,  $A_1$  and  $\theta_{11}$  for 1.000.000 iterations with a thinning factor of 100.

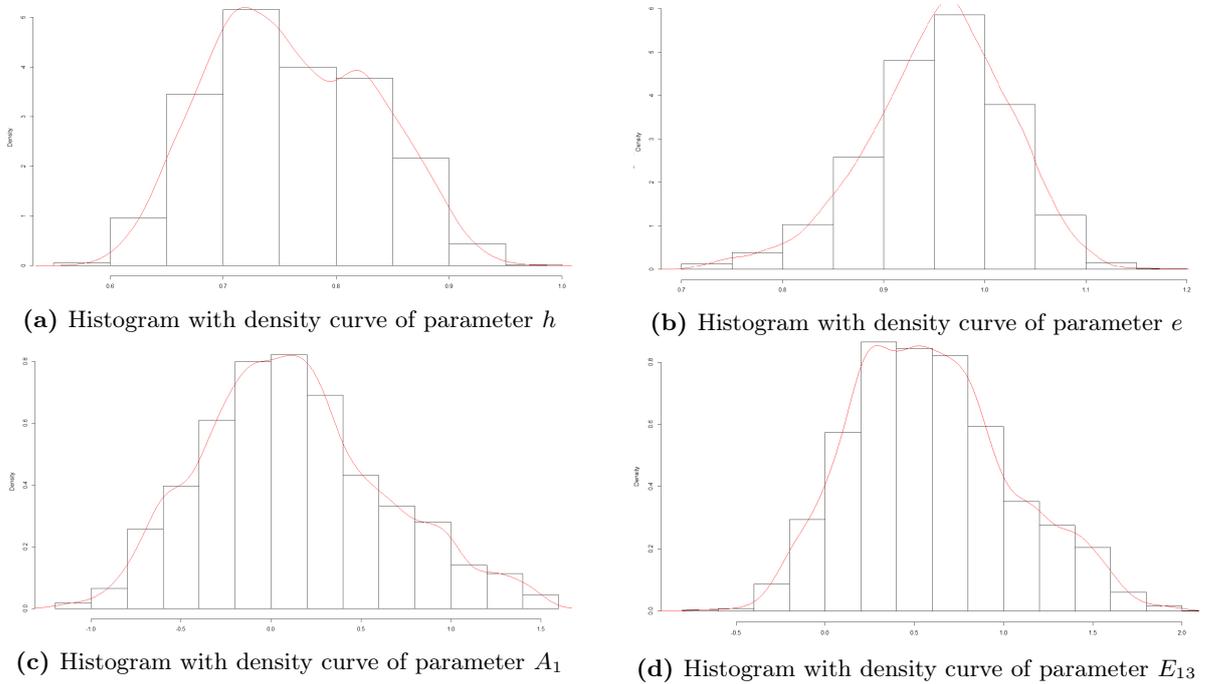


**Figure 8: Trace plots of various model parameters. The plots show one MCMC chain for 1.000.000 iterations with a thinning factor of 100.**

We see from figure 8 that the Markov chain has converged to the stationary distribution within 1.000.000 iterations for parameters  $h$ ,  $e$ ,  $\rho$ ,  $\mu_p$ ,  $A_1$  and  $\theta_{11}$ . Therefore, it was effective to run the code again for 1.000.000 iterations. Figure 7d showed that the Markov chains were still systematically increasing for any state less than 100.000, for example. In figure 8d, it can be observed that the Markov chain has converged after approximately 300.000 iterations. We could also have been misled by the trace plots of some parameters up until 100.000 iterations. It appeared that samples from the marginal posterior of  $\rho$  take on a value of circa 0, for example. This follows from figure 7c. However, we see in figure 8c that the chain starts to increase after approximately 150.000 iterations. Thereafter, it reaches the stationary distribution around 400.000 iterations.

Now that we have obtained samples from the marginal posteriors of the parameters, we can assess the distribution of these samples by means of a density plot. This is the smoothed histogram of the values in the trace plot. So, it gives the distribution of the values of the parameters in the chain. Note that the burn-in phase needs to be taken into account. Since the Markov chain has reached the stationary distribution after 500.000 iterations for each parameter in figure 8, we will take this as our burn-in phase. The histograms with a fitted density curve can be found in figure 9 for parameters  $h$ ,  $e$ ,  $A_1$  and  $E_{13}$ .

We can recognize the shape of a normal distribution in the density curves of all parameters in figure 9. The marginal posterior distribution could not be investigated for all parameters in the model, since there are 389 parameters in total. Nonetheless, the density curves of all other parameters that have been examined also have the shape of a normal distribution. Moreover, the Markov chain has reached the stationary distribution after 500.000 iterations for all other parameters that have been examined as well. The trace plot, histogram and density curve of some of these parameters are included in Appendix F.



**Figure 9: Histograms with a fitted density curve of various model parameters. The plots are based on one MCMC chain of 1.000.000 iterations with a thinning factor of 100, after a burn-in phase of 500.000 iterations.**

### 6.3 Transmission of intelligence

Now that we have obtained samples from the marginal posteriors of the parameters of model 1, we can use these samples to determine if transmission of intelligence is dominated by genetic factors or environmental factors for each twin. Note that the components that define the level of intelligence of an individual also represent the extent to which the transmission of intelligence is determined by genetic factors or environmental factors. So, when we know which part of the level of intelligence can be explained by genetic factors, we also know which part of the transmission of intelligence can be ascribed to genetic factors.

The level of intelligence of a twin depends on the genetic component  $A$  together with its loading factor  $h$  and on the environmental component  $E$  together with its loading factor  $e$ , as given in equation 4.2:

$$\theta_{ij} = h \times A_i + e \times E_{ij},$$

where  $j = 3, 4$  and  $i = 1, \dots, 48$ . So, in order to determine to which extent the level of intelligence of twin 1 from family 1 is determined by genetic factors and environmental factors, we need to compute which parts of the total level of intelligence  $\theta_{13}$  amount to those factors. One way to do this is by means of computing the fraction. If we want to know which part of the level of intelligence is determined by genetic factors, we need to compute the following:

$$\text{genetic part} = \frac{|h \cdot A_1|}{|h \cdot A_1| + |e \cdot E_{13}|}, \quad (13)$$

where  $h$ ,  $A_1$ ,  $e$  and  $E_{13}$  denote the means of the samples from the marginal posterior distributions of the parameters after the burn-in phase. Table 8 provides an overview of these mean values. Note that they coincide with the mean of the probability density functions in figure 9.

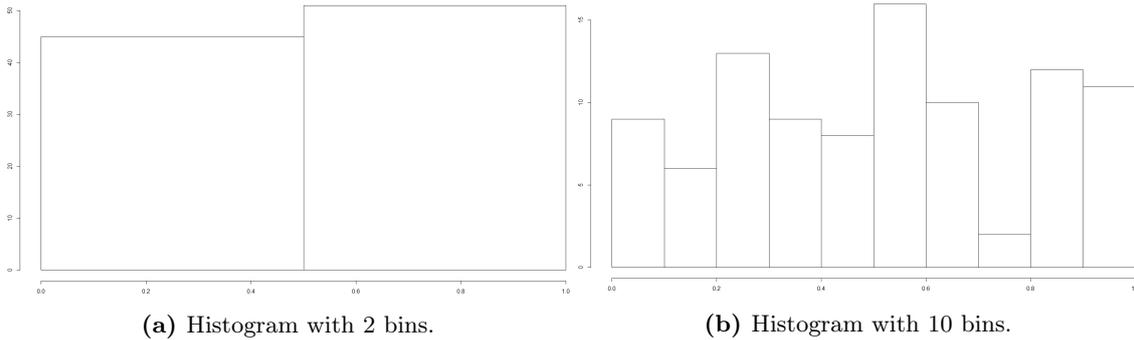
We can now use those values from table 8 to compute the fraction in equation 13:

$$\text{genetic part} = \frac{|h \cdot A_1|}{|h \cdot A_1| + |e \cdot E_{13}|} = \frac{0.76 \cdot 0.13}{(0.76 \cdot 0.13) + (0.95 \cdot 0.59)} = 0.150$$

Therefore, we can conclude that the level of intelligence of twin 1 from family 1 is for 15,0% determined by genetic factors and for 85,0% determined by environmental factors. We can do the same computation for all other 95 twins. Histograms that show which fraction of the total level of intelligence is determined by genetic factors for how many twins are given in figure 10.

| Parameter | Mean |
|-----------|------|
| $h$       | 0.76 |
| $A_1$     | 0.13 |
| $e$       | 0.95 |
| $E_{13}$  | 0.59 |

**Table 8:** Mean values of the samples from the marginal posterior distribution of parameters  $h$ ,  $A_1$ ,  $e$  and  $E_{13}$  after the burn-in phase.



**Figure 10: Histograms of the fractions of the level of intelligence that amount to genetic factors for the twins. The horizontal axis indicates the fraction of the level of intelligence that is determined by genetic factors. The vertical axis indicates the number of twins.**

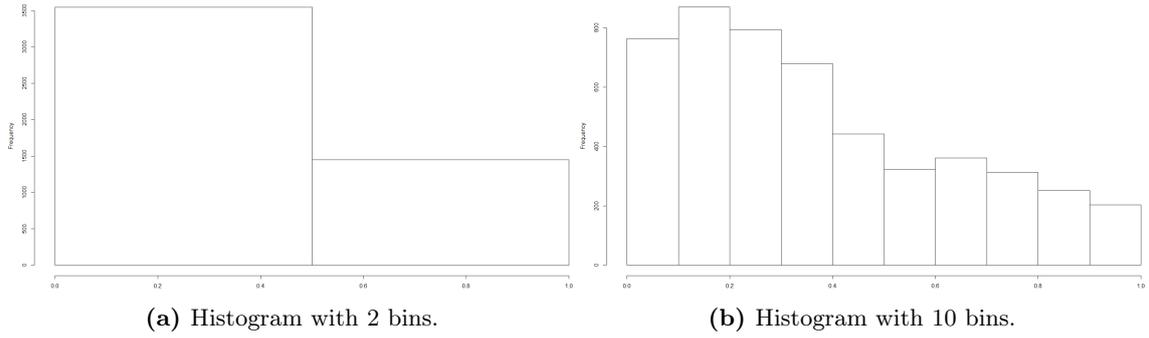
The level of intelligence of an individual is dominated by genetic factors if the fraction is bigger than 0.5. Likewise, the level of intelligence of an individual is dominated by environmental factors if the fraction is smaller than 0.5. It follows from the histogram in figure 10a that the level of intelligence is dominated by genetic factors in case of approximately half of the twins. The level of intelligence is dominated by environmental factors for the other part of the twins. A histogram with 10 bins has been included in figure 10b. This histogram shows more precisely to which extent the level of intelligence is determined by genetic factors for how many twins.

In the previous approach, we took the mean values of the samples from the marginal posteriors of the parameters in order to compute the part of the total level of intelligence that was determined by genetic factors for each individual. However, samples from the marginal posterior of a parameter may differ significantly from the mean value. Therefore, it is more precise to compute the fraction for each state of the Markov chain after the burn-in phase for each twin.

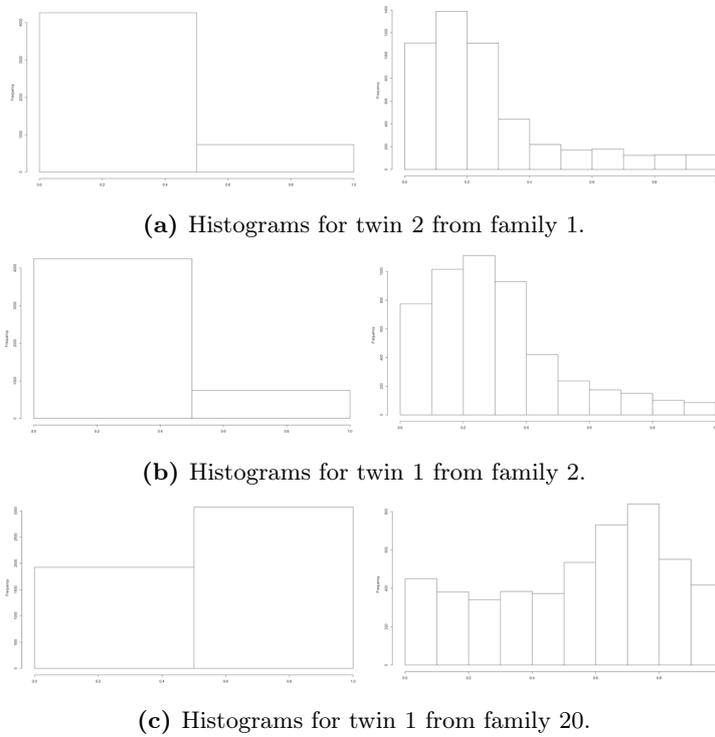
Histograms that show which fraction of the level of intelligence of twin 1 from family 1 is determined by genetic factors for how many states of the Markov chain after the burn-in phase can be found in figure 11. We see from the histogram in figure 11a that the fraction is smaller than 0.5 in most states of the Markov chain. Hence, we can conclude from this method that it is most likely that the level of intelligence of twin 1 from family 1 is dominated by environmental factors. A possible explanation of the fact that the level of intelligence is dominated by genetic factors in some states is that samples of a parameter in the Markov chain are constantly adapting to samples of other parameters from the previous state. It may take a few states before a sample has reached a probability region that is high enough for the parameters  $h$ ,  $e$ ,  $A_1$  and  $E_{13}$  to give an accurate indication of how the level of intelligence is composed.

Note that the result of the second approach corresponds to the result of the first approach, i.e., that the level of intelligence of twin 1 from family 1 is dominated by environmental factors. Histograms of the fractions of the level of intelligence that amount to genetic factors for each state of the Markov chain after the burn-in phase for more twins can be found in figure 12.

Moreover, the composition of an interesting quantity that relates to the transmission follows from the output of the Metropolis Hastings algorithm. This quantity is the variance in the level of intelligence between twins. It follows from equation 8 that  $\text{Var}(P) = h^2 + e^2$ , where  $P$  represents the level of intelligence of a twin. We can read table 8 and see that the mean values of the samples from the marginal posteriors of  $h$  and  $e$  are 0.76 and 0.95, respectively. Therefore, we can conclude that the variance in level of intelligence between the 96 twins in this research exists mainly because of environmental factors. This variance is equal to 1.48.



**Figure 11: Histograms of the fractions of the level of intelligence that amount to genetic factors for each state of the Markov chain after the burn-in phase for twin 1 from family 1. The horizontal axis indicates the fraction of the level of intelligence that is determined by genetic factors. The vertical axis indicates the number of states.**



**Figure 12: Histograms of the fractions of the level of intelligence that amount to genetic factors for each state of the Markov chain after the burn-in phase for 3 twins. The horizontal axis indicates the fraction of the level of intelligence that is determined by genetic factors. The vertical axis indicates the number of states.**

## 7 Discussion

A description of the process of inferring a model that describes the transmission of intelligence from parents to offspring by means of the Metropolis Hastings algorithm has been given in previous chapters and was followed by convergence plots of its output and its interpretation. A discussion of these results will be presented in the present chapter.

Our 2 initial Markov chains consisted of only 100.000 states, corresponding to 100.000 iterations. These Markov chains did not reach the stationary distribution yet. Thereafter, it followed from the trace plots of the samples from the marginal posteriors of the parameters that the Markov chain did reach the stationary distribution within 1.000.000 iterations. However, because we have only plotted one chain for 1.000.000 iterations, we can not be completely certain that it has reached a state of convergence. It might be that the chain reaches only temporary plateaus within 1.000.000 iterations, after which it systematically decreases or increases again. We would be misled in the same way as if we would not have looked at the trace plot of  $\rho$  for 1.000.000 iterations, but only at the trace plot of  $\rho$  for 100.000 iterations. For now, we will rely on the fact that the chain has reached the stationary distribution for all parameters of the model. Hence, it appears that the Metropolis Hastings algorithm was a valid method to infer our parent-offspring model for 1.000.000 iterations.

Two methods to interpret the output from the Metropolis Hastings algorithm and investigate if the transmission of intelligence is dominated by genetic factors or environmental factors have been given in the last subsection of the previous chapter. The first method was to take the mean values of the samples from the Markov chain after the burn-in phase of parameters  $h$ ,  $e$ ,  $A_i$ ,  $E_{i3}$  and  $E_{i4}$ , for  $i = 1, \dots, 48$ , and to compute which part of the total level of intelligence is determined by genetic factors for each twin from all 48 families. We plotted these fractions in a histogram with 2 bins and observed that genetic factors dominate the level of intelligence for approximately half of the twins. Histograms with more than 2 bins show more specifically to which extent these factors have an impact on the transmission of intelligence for how many twins. However, since this method uses the mean values of the parameters instead of the exact values in each state, this method is not as precise as the second method.

The second method was to compute the part of the total level of intelligence that is determined by genetic factors for each state of the Markov chain after the burn-in phase for a twin. Since the total number of twins in this research is 96, it was not possible to examine histograms of these fractions for all of them. This was due to time constraints and space constraints. The histograms of 4 twins have been included in figure 11 and 12 on page 29. These histograms are representative of the other circa 15 histograms that have been examined but have not been included in this paper. The reason that this selection of histograms is representative is that most of them (i.e., 3 out of 4) show that it is most likely that the level of intelligence of that twin is dominated by environmental factors, as is the case for most examined histograms. This agrees with previous twin studies that showed that 20 to 50 percent of the level of intelligence of an individual is determined by genetic effects and the other part is determined by environmental effects [1, 9].

So, we can conclude that the first approach shows that the number of twins for which transmission of intelligence is dominated by genetic factors and the number of twins for which transmission of intelligence is dominated by environmental factors is approximately the same. The second approach shows that the transmission of intelligence of most twins is probably dominated by environmental factors. However, the histograms of all 96 twins must be taken into account before we can determine the exact numbers of twins for which the transmission of intelligence is dominated by genetic factors and environmental factors.

## 8 Conclusion

Unfortunately, not all histograms of the 96 twins could be examined in this research because of time limitations. This is necessary to be able to conclude for how many twins the transmission of intelligence is dominated by genetic factors and for how many twins the transmission of intelligence is dominated by environmental factors. This might be done in a successive study. Additionally, there is still a lot more possible for further research. Some ideas are presented in this section, as well as some final remarks.

First of all, it might be useful to run the code of the Metropolis Hastings algorithm again for 1.000.000 iterations to check that this chain converges to the same values of the marginal posterior distributions. An even more appropriate approach would be to study the behavior of two Markov chains for more iterations to confirm our results.

Secondly, the interpretation of more parameters might be considered. This research has only focused on the transmission of intelligence, i.e., on the composition of the ability levels of the twins. The parent-offspring model contains more interesting variables, such as the correlation between intelligence levels of parents.

Moreover, different parent-offspring models that describe the transmission of intelligence might be compared with each other. In this research, only one model has been inferred. This model assumed phenotypic assortment and no cultural transmission. However, there might exist a model that fits the data better. Model inference will then lead to a more accurate analysis of the transmission of intelligence and the outcome with respect to the transmission of intelligence will be more valid. Two improvements of the model that has been inferred in this study can be considered, for example. The first improved model contains different genetic influences for parents and twins and the second improved model contains different test circumstances for parents and twins. Another possible model to infer is the model that includes phenotypic assortment as well as cultural transmission.

These models can be compared by looking at their complexity and the fit. The complexity of the model is the number of free parameters and there exist several methods to check whether a model fits the data. The deviance information criterion (DIC), for example, is a well-known model comparison method [3]. This method uses samples from the marginal posterior distributions of the parameters that are obtained by the Metropolis Hastings algorithm. Another method to check the model fit is to use posterior predictive checks. This involves simulating data under the fitted model and then comparing these to the observed data [10]. It can be concluded that the model fits the data well if the simulated data looks much like the data from the data set.

Furthermore, there are possibilities to improve the R code for the Metropolis Hastings algorithm, so that the algorithm is faster and computationally less expensive. One way is to use the specific formula of the acceptance probability to move to a different state for each parameter, instead of using the complete score functions as has been done in this research. Many terms will cancel in the ratio of the scores. An example can be found in Appendix G.

Finally, a few final remarks will be stated. First of all, the parent-twin design of this study is a good way to model the transmission of intelligence from parents to offspring, as has been shown in previous studies. The Metropolis Hastings algorithm and its output seem to be valid as well, as the Markov chain reached the stationary distribution for all parameters. The reliability might be tested in further research to check if the results will be confirmed. One method of this study showed that for circa one half of the twins their level of intelligence is dominated by genetic factors. The level of intelligence of the other part of the twins is dominated by environmental factors. It must still be determined for most twins which factors dominate their level of intelligence in the other approach. The cases for which this has been investigated hint to a domination of environmental factors.

## 9 Bibliography

- [1] F. Baker, S. Kim (2004). *Item response theory: Parameter estimation techniques*. New York, N.Y.: Dekker.
- [2] L. Baker, N. Pedersen, C. Reynolds (2000). *Multivariate models of mixed assortment: Phenotypic assortment and social homogamy for education and fluid ability*. Behavior Genetics, 30(6), 455-476. doi:10.1023/A:1010250818089
- [3] N. Best, B. Carlin, D. Spiegelhalter, A. Van der Linde (2002). *Bayesian measures of model complexity and fit (with discussion)*. Journal- Royal Statistical Society Series B Statistical Methodology, 64, 583-640.
- [4] D. Boomsma, A. Busjahn, L. Peltonen (2008). *Classical twin studies and beyond*. Nature Reviews Genetics, 3(11), 872-882. doi:10.1038/nrg932
- [5] D. Boomsma, S. Van den Berg, M. Van Leeuwen (2008). *A twin-family study of general IQ*. Learning and Individual Differences, 18(1), 76-88. doi:10.1016/j.lindif.2007.04.006
- [6] C. Cloninger (1980). *Interpretation of intrinsic and extrinsic structural relations by path analysis: Theory and applications to assortative mating*. Genetical Research, 36(02), 133-133. doi:10.1017/S0016672300019765
- [7] G. Coro (2013). *A lightweight guide on Gibbs sampling and JAGS*. ISTI-CNR, Technical Report cnr.isti/2013-TR-036, Pisa, Italy. Retrieved from: [https://www.researchgate.net/publication/258630806\\_A\\_Lightweight\\_Guide\\_on\\_Gibbs\\_Sampling\\_and\\_JAGS](https://www.researchgate.net/publication/258630806_A_Lightweight_Guide_on_Gibbs_Sampling_and_JAGS)
- [8] L. Eaves, H. Eysenck, N. Martin, R. Jardine (1989). *Genes, culture and personality : An empirical approach*. London etc.: Academic Press.
- [9] J. DeFries, D. Fulker (1983). *Genetic and environmental transmission in the colorado adoption project: Path analysis*. British Journal of Mathematical and Statistical Psychology, 36(2), 175-188. doi:10.1111/j.2044-8317.1983.tb01123.x
- [10] A. Gelman, J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models (Analytical methods for social research)*. Cambridge: Cambridge University Press.
- [11] M. Grzegorzcyk (2016). *MCMC simulations*. Lecture notes, Statistical Reasoning WISR-11 University of Groningen, delivered 05 October 2016.
- [12] M. Grzegorzcyk (2016). *MCMC simulations 2*. Lecture notes, Statistical Reasoning WISR-11 University of Groningen, delivered 07 October 2016.
- [13] R. Hambleton, H. Swaminathan (1985). *Item response theory: Principles and applications (Evaluation in education and human services, [7])*. Boston: Kluwer-Nijhoff.
- [14] P. Hoff (2009). *A first course in Bayesian statistical methods (Springer texts in statistics)*. New York: Springer. doi:10.1007/978-0-387-92407-6
- [15] P. Holland, H. Wainer, (1993). *Differential item functioning*. Hillsdale, NJ etc.: Lawrence Erlbaum.
- [16] M. Keller, S. Medland (2009). *Modeling extended twin family data II: power associated with different family structures*. Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies, 12(1), 19-25. doi:10.1375/twin.12.1.19
- [17] M. Mukaka (2012). *Statistics corner: A guide to appropriate use of correlation coefficient in medical research*. Malawi Medical Journal: The Journal of Medical Association of Malawi, 24(3), 69-71.
- [18] B. Otermann (2014). *Bayesian Assessment of Parent-Offspring Models using Item Data*. (Unpublished master's thesis). Utrecht University, Utrecht, The Netherlands.
- [19] D. Ravenzwaaij, P. Cassey, S. Brown (2016). *A simple introduction to markov chain monte-Carlo sampling*. Psychonomic Bulletin and Review 25(1), 143-154. doi:10.3758/s13423-016-1015-8

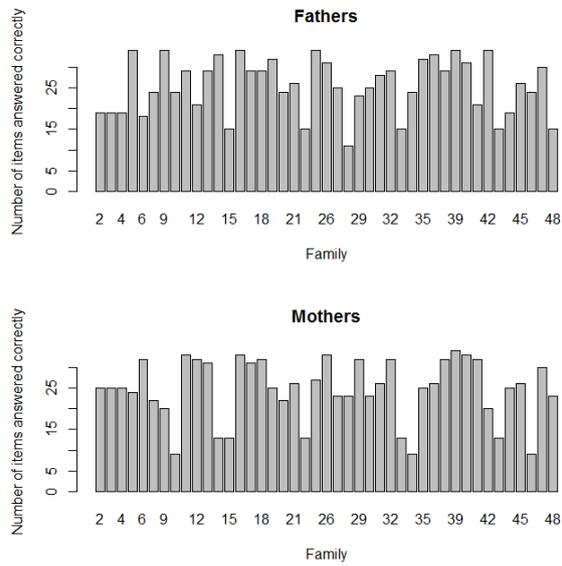
## 10 Appendix

### Appendix A

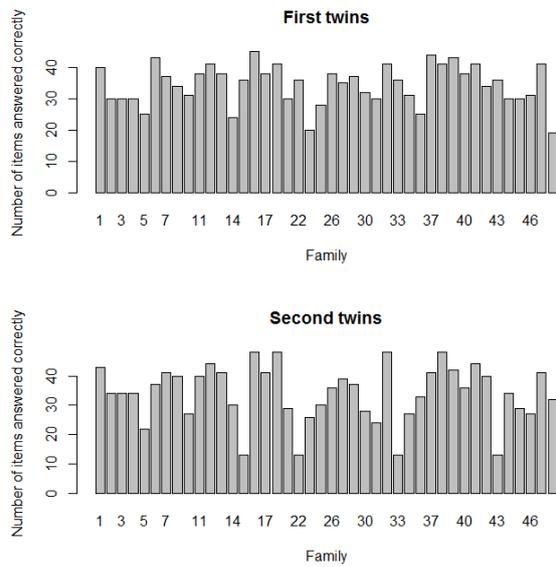
|       |       |       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| B1    | A2    | C1    | A3    | A4    | A1    | A5    | A6    | B3    | D1    |
| -6.94 | -6.94 | -5.87 | -5.82 | -5.52 | -5.52 | -5.09 | -5.09 | -4.93 | -4.45 |
| B2    | A9    | B4    | A10   | B5    | A7    | C2    | A8    | D5    | B6    |
| -3.94 | -3.65 | -3.59 | -2.73 | -2.69 | -2.44 | -2.34 | -1.92 | -1.92 | -1.64 |
| C5    | D2    | C3    | D3    | B10   | D6    | C4    | D4    | B7    | B9    |
| -1.47 | -1.30 | -1.29 | -1.12 | -0.87 | -0.86 | -0.80 | -0.74 | -0.69 | -0.65 |
| A11   | C7    | C9    | C6    | B11   | 1     | D8    | B8    | 2     | D7    |
| -0.51 | -0.44 | -0.29 | -0.26 | -0.22 | -0.21 | -0.17 | -0.10 | -0.01 | 0.06  |
| 6     | D9    | D10   | C8    | 11    | E2    | A12   | E1    | 3     | E3    |
| 0.17  | 0.30  | 0.36  | 0.39  | 0.47  | 0.55  | 0.56  | 0.66  | 0.72  | 0.78  |
| 8     | 7     | 9     | 10    | 15    | 5     | B12   | 4     | 14    | E4    |
| 0.82  | 0.92  | 1.11  | 1.19  | 1.27  | 1.27  | 1.29  | 1.34  | 1.34  | 1.35  |
| 12    | E5    | C10   | E6    | E7    | C11   | 16    | E8    | D11   | 17    |
| 1.42  | 1.45  | 1.58  | 1.77  | 2.03  | 2.12  | 2.33  | 2.33  | 2.35  | 2.46  |
| 19    | 13    | 20    | 21    | 31    | 30    | C12   | 18    | 26    | E9    |
| 2.50  | 2.61  | 2.61  | 3.00  | 3.16  | 3.21  | 3.25  | 3.29  | 3.34  | 3.40  |
| 23    | 24    | E10   | 27    | D12   | 22    | 28    | E12   | E11   | 25    |
| 3.46  | 3.50  | 3.58  | 3.71  | 3.74  | 3.78  | 3.87  | 3.87  | 3.88  | 4.05  |
| 34    | 33    | 29    | 35    | 32    | 36    |       |       |       |       |
| 4.09  | 4.12  | 4.53  | 4.65  | 4.74  | 5.95  |       |       |       |       |

**Table 9:** Sorted linked  $\beta$  values of the items of the APM (1-36) and the SPM (A1:E12)

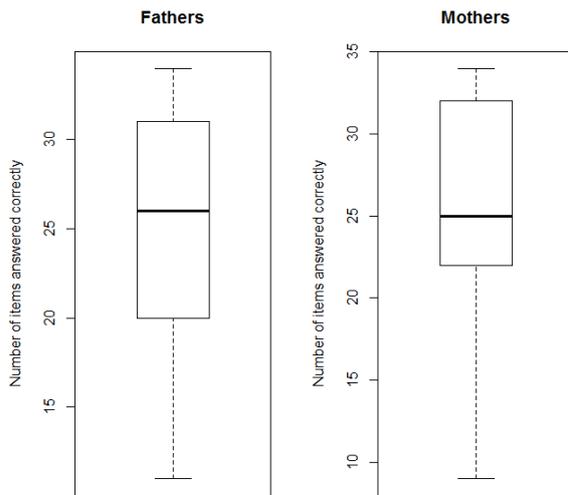
## Appendix B



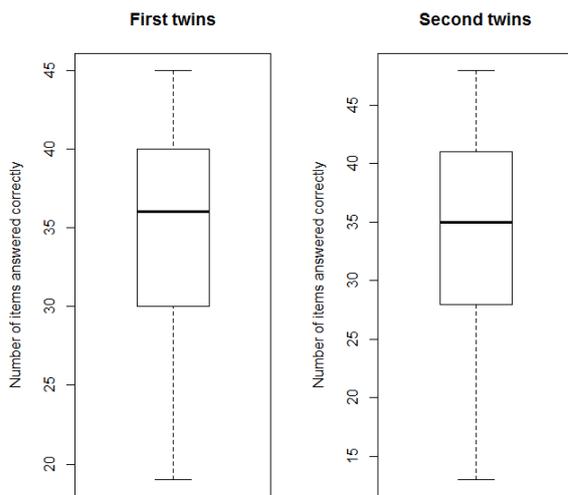
**Figure 13: Histogram of the number of items answered correctly by each father of a family.**



**Figure 14: Histogram of the number of items answered correctly by each mother of a family.**



**Figure 15:** Box plot of the number of items answered correctly by the fathers of the families.



**Figure 16:** Box plot of the number of items answered correctly by the mothers of the families.

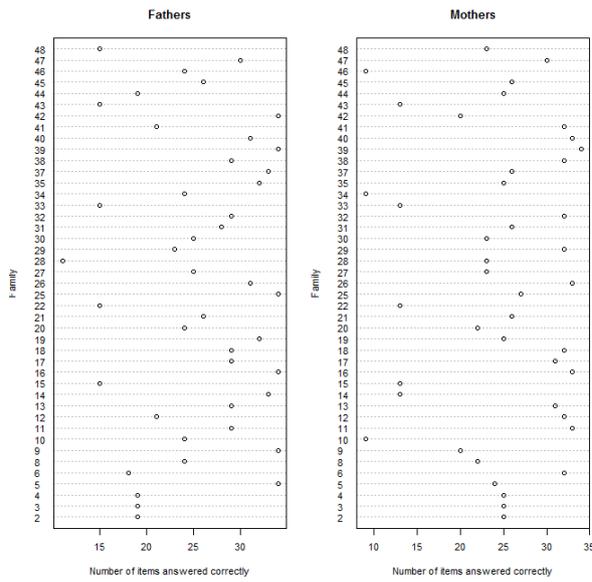


Figure 17: Dot charts of the number of items answered correctly by the parents of a family.

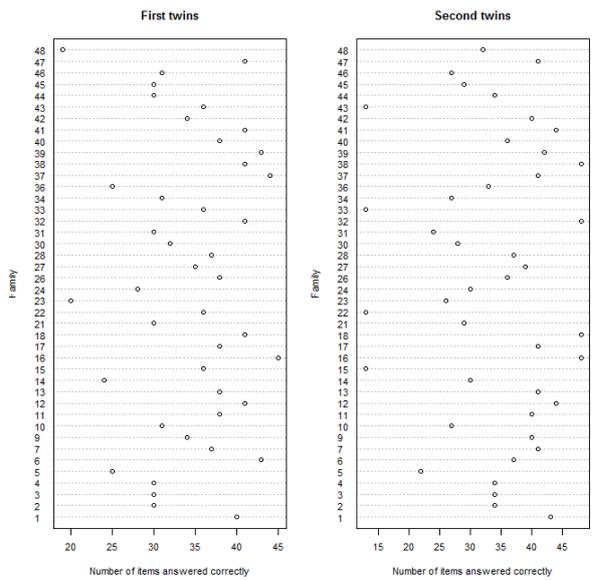
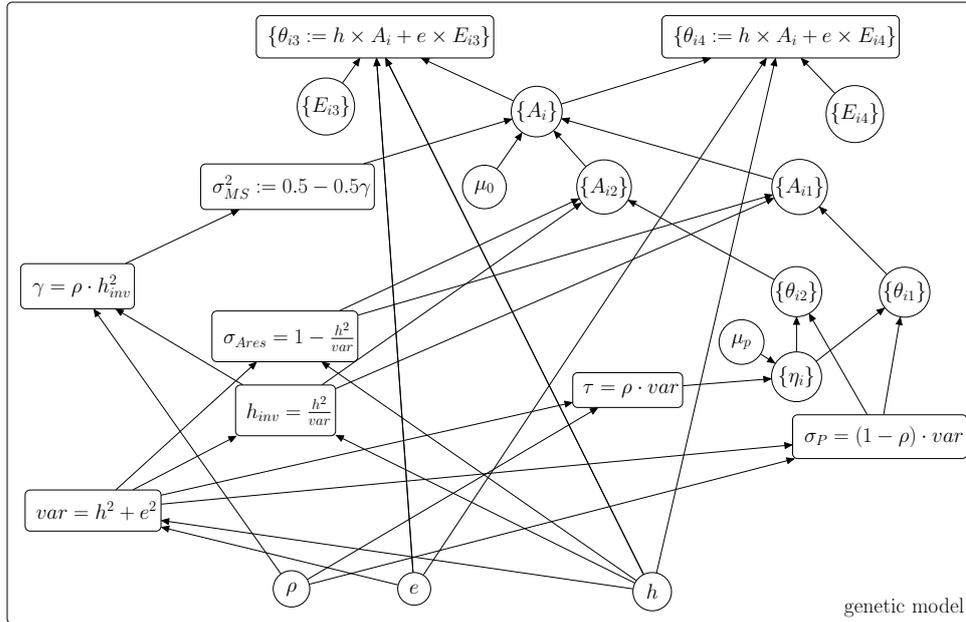


Figure 18: Dot charts of the number of items answered correctly by the twins of a family.

## Appendix C



**Figure 19: Graphical representation of model 1. Parameters that have to be inferred (i.e., the free parameters) are represented by white circles. The data and the fixed parameters (i.e., the observed parameters) are represented by grey circles. The nine rectangles indicate definitions, which deterministically depend on the parent nodes. The symbol  $\times$  indicates a multiplication.**

## Appendix D

```

listbeta = data.frame(beta)
beta3 = listbeta[c(1:36),] #vector with 36 entries: beta values of the parents
beta4 = listbeta[c(37:96),] #vector with 60 entries: beta values of the twins
beta = c(beta3,beta3,beta4,beta4)

Y = as.matrix(MZ_item_GN)
APM = 36
SPM = 60

#logarithm of the score (cf. equation 10)
logscore = function(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1,theta2,h,e,rho){

  loglikelihood1 = 0
  loglikelihood2 = 0
  loglikelihood3 = 0
  loglikelihood4 = 0
  problogA = 0
  problogE3 = 0
  problogE4 = 0
  problogA1 = 0
  problogA2 = 0
  problogtheta2 = 0
  problogtheta1 = 0
  problogeta = 0

  #variables that depend deterministically on their parents nodes
  var = (h^2)+(e^2)
  tau = rho * var
  sigmaP = (1-rho) * var
  hinv = (h*h)/var
  gamma = rho * (hinv)
  sigmaMS = 0.5 - 0.5*gamma
  sigmaAres = 1 - ((h^2)/var)

  theta3 = rep(0,dim(Y)[1])
  for (i in 1:dim(Y)[1]){
    theta3[i] = h * A[i] + e * E3[i]}

  theta4 = rep(0,dim(Y)[1])
  for (i in 1:dim(Y)[1]){
    theta4[i] = h * A[i] + e * E4[i]}

  #log-likelihood function
  for (i in 1:dim(Y)[1]){

    beta1 = beta[which(is.na(Y[i,1:APM])==0)]
    loglikelihood1 = loglikelihood1 + (Y[i,which(is.na(Y[i,1:APM])==0)] %%%
    (theta1[i] - beta1 - log(1+exp(theta1[i]-beta1))) - (1-Y[i,which(is.na(Y[i,1:APM])==0)]) %%%
    log(1 + exp(theta1[i]-beta1)))

    beta2 = beta[APM + which(is.na(Y[i,(APM+1):(APM*2)])==0)]
    loglikelihood2 = loglikelihood2 + (Y[i,APM + which(is.na(Y[i,(APM+1):(APM*2)])==0)] %%%
    (theta2[i] - beta2 - log(1+exp(theta2[i]-beta2))) - (1-Y[i,APM + which(is.na(Y[i,(APM+1):(APM*2)])==0)]) %%%
    log(1 + exp(theta2[i]-beta2)))
  }
}

```

```

beta5 = beta[2*APM + which(is.na(Y[i,(2*APM + 1):(2*APM + SPM)])==0)]
loglikelihood3 = loglikelihood3 + (Y[i,2*APM + which(is.na(Y[i,(2*APM + 1):(2*APM + SPM)])==0)] %%
(theta3[i] - beta5 - log(1+exp(theta3[i]-beta5))) -
(1-Y[i,2*APM + which(is.na(Y[i,(2*APM + 1):(2*APM + SPM)])==0)]) %% log(1 + exp(theta3[i]-beta5)))

beta6 = beta[2*APM + SPM + which(is.na(Y[i,(2*APM + SPM + 1):(2*APM + 2*SPM)])==0)]
loglikelihood4 = loglikelihood4 + (Y[i,2*APM + SPM + which(is.na(Y[i,(2*APM + SPM + 1):
(2*APM + 2*SPM)])==0)] %% (theta4[i] - beta6 - log(1+exp(theta4[i]-beta6))) -
(1-Y[i,2*APM + SPM + which(is.na(Y[i,(2*APM + SPM + 1):(2*APM + 2*SPM)])==0)]) %%
log(1 + exp(theta4[i]-beta6)))
}

#the sum of the logarithms of the priors for each parameter with subscript i
problogA = sum(log(dnorm(A[1:dim(Y)[1]], 0.5*(A1[1:dim(Y)[1]]+A2[1:dim(Y)[1]])+mu0, sqrt(sigmaMS) )))
problogE3 = sum(log(dnorm(E3[1:dim(Y)[1]],0,1)))
problogE4 = sum(log(dnorm(E4[1:dim(Y)[1]],0,1)))
problogA1 = sum(log(dnorm(A1[1:dim(Y)[1]], hinv * theta1[1:dim(Y)[1]], sqrt(sigmaAres) )))
problogA2 = sum(log(dnorm(A2[1:dim(Y)[1]], hinv * theta2[1:dim(Y)[1]], sqrt(sigmaAres) )))
problogtheta2 = sum(log(dnorm(theta2[1:dim(Y)[1]], eta[1:dim(Y)[1]], sqrt(sigmaP) )))
problogtheta1 = sum(log(dnorm(theta1[1:dim(Y)[1]], eta[1:dim(Y)[1]], sqrt(sigmaP) )))
problogeta = sum(log(dnorm(eta[1:dim(Y)[1]], mup, sqrt(tau) )))

#the logarithm of the prior for each parameter without subscript i
problogmup = log(dnorm(mup,0,sqrt(10)))
problogmu0 = log(dnorm(mu0,0,sqrt(10)))
problogh = log(dunif(h,0,4))
probloge = log(dunif(e,0,4))
problogrho = log(dunif(rho,0,1))

x3 = c(loglikelihood1,loglikelihood2,loglikelihood3,loglikelihood4, problogA, problogE3, problogE4,
problogA1, problogA2, problogtheta2, problogtheta1, problogeta, problogmup, problogmu0, problogh,
probloge, problogrho)
x4 = sum(x3)
return(x4)
}

#the acceptance probability function
Accep = function(LL_new, LL_old){
  HR = 1
  R = exp(LL_new-LL_old + log(HR))
  p2 = min(1,R)
  return(p2)
}

#to store the sampled values in the chains
outA = list()
outE3 = list()
outE4 = list()
outA1 = list()
outA2 = list()
outtheta1 = list()
outtheta2 = list()
outeta = list()
outmup = list()
outmu0 = list()
outh = list()
oute = list()
outrho = list()

```

```

#initializations
A = rep(0.5,dim(Y)[1])
E3 = rep(0.5,dim(Y)[1])
E4 = rep(0.5,dim(Y)[1])
A1 = rep(0.5,dim(Y)[1])
A2 = rep(0.5,dim(Y)[1])
theta2 = rep(0,dim(Y)[1])
theta1 = rep(0,dim(Y)[1])
eta = rep(0,dim(Y)[1])
mup = 0
mu0 = 0
h = 2
e = 2
rho = 0.5

outA[[1]] = A
outE3[[1]] = E3
outE4[[1]] = E4
outA1[[1]] = A1
outA2[[1]] = A2
outtheta2[[1]] = theta2
outtheta1[[1]] = theta1
outeta[[1]] = eta
outmup[1] = mup
outmu0[1] = mu0
outh[1] = h
oute[1] = e
outrho[1] = rho

eps = 0.01

#MCMC iterations
for (t in 2:100000){

  #update vector A
  v = runif(dim(Y)[1],-eps,eps)
  A_new = A + v

  LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A_new,A1,A2,theta1,theta2,h,e,rho)
  LL_old = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1,theta2,h,e,rho)
  Acc = Accep(LL_new, LL_old)

  u = runif(1,0,1)
  if (u<Acc)
  {A = A_new
  LL_old = LL_new}
  outA[[t]] = A

  #update vector E3
  v = runif(dim(Y)[1],-eps,eps)
  E3_new = E3 + v

  LL_new = logscore(Y,beta,E3_new,E4,mu0,mup,eta,A,A1,A2,theta1,theta2,h,e,rho)
  Acc = Accep(LL_new, LL_old)
}

```

```

u = runif(1,0,1)
if (u<Acc)
{E3 = E3_new
  LL_old = LL_new}
outE3[[t]] = E3

#update vector E4
v = runif(dim(Y)[1],-eps,eps)
E4_new = E4 + v

LL_new = logscore(Y,beta,E3,E4_new,mu0,mup,eta,A,A1,A2,theta1,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{E4 = E4_new
  LL_old = LL_new}
outE4[[t]] = E4

#update vector A1
v = runif(dim(Y)[1],-eps,eps)
A1_new = A1 + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1_new,A2,theta1,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{A1 = A1_new
  LL_old = LL_new}
outA1[[t]] = A1

#update vector A2
v = runif(dim(Y)[1],-eps,eps)
A2_new = A2 + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2_new,theta1,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{A2 = A2_new
  LL_old = LL_new}
outA2[[t]] = A2

#update vector theta2
v = runif(dim(Y)[1],-eps,eps)
theta2_new = theta2 + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1,theta2_new,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{theta2 = theta2_new
  LL_old = LL_new}
outtheta2[[t]] = theta2

```

```

#update vector theta1
v = runif(dim(Y)[1],-eps,eps)
theta1_new= theta1 + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1_new,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{theta1 = theta1_new
 LL_old = LL_new}
outtheta1[[t]] = theta1

#update vector eta
v = runif(dim(Y)[1],-eps,eps)
eta_new = eta + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta_new,A,A1,A2,theta1,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{eta = eta_new
 LL_old = LL_new}
outeta[[t]] = eta

#update mup
v = runif(1,-eps,eps)
mup_new = mup + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup_new,eta,A,A1,A2,theta1,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{mup = mup_new
 LL_old = LL_new}
outmup[t] = mup

#update mu0
v = runif(1,-eps,eps)
mu0_new = mu0 + v

LL_new = logscore(Y,beta,E3,E4,mu0_new,mup,eta,A,A1,A2,theta1,theta2,h,e,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{mu0 = mu0_new
 LL_old = LL_new}
outmu0[t] = mu0

#update h
v = runif(1,-eps,eps)
h_new = h + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1,theta2,h_new,e,rho)
Acc = Accep(LL_new, LL_old)

```

```

u = runif(1,0,1)
if (u<Acc)
{h = h_new
 LL_old = LL_new}
outh[t] = h

#update e
v = runif(1,-eps,eps)
e_new = e + v

LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1,theta2,h,e_new,rho)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{e = e_new
 LL_old = LL_new}
oute[t] = e

#update rho
v = runif(1,-eps,eps)
rho_new = rho + v

#to prevent that rho gets out of boundary [0,1]
if ((rho_new < 1) & (rho_new > 0)){
LL_new = logscore(Y,beta,E3,E4,mu0,mup,eta,A,A1,A2,theta1,theta2,h,e,rho_new)
Acc = Accep(LL_new, LL_old)

u = runif(1,0,1)
if (u<Acc)
{rho = rho_new
 LL_old = LL_new}
}
outrho[t] = rho
}

```

## Appendix E

In some models with multiple parameters it is easy to sample from the full conditional distribution (fcd) of each parameter. The distribution  $p(\theta_k | \theta_{-k}, Y_1, \dots, Y_n)$  - where  $\theta_{-k}$  denotes all parameters of the parameter space except for the  $k$ -th one - is called the fcd of  $\theta_k$ , since it is the probability of parameter  $\theta_k$  given everything else [14]. In such cases, posterior approximations can be made with the Gibbs sampler. Like the Metropolis Hastings algorithm, this is an iterative algorithm that constructs a Markov chain of parameter values whose distribution converges to the joint posterior distribution.

So, if it is possible to compute the fcd  $p(\theta_k | \theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_K, y_1, \dots, y_n)$ , then the  $k$ -th MH-MCMC move can be replaced by a more efficient Gibbs-MCMC move. Here, the  $k$ -th sub-step is as in equation 5. Thus, instead of the MH-MCMC proposal move

$$\theta_k^{(*)} = \theta_k^{(t)} + v,$$

we can apply a Gibbs move and sample the new candidate  $\theta_k^{(*)}$  from its full conditional distribution.

By rewriting the expression for the Hastings Ratio it can be shown that the acceptance probability  $A(\Theta_A, \Theta_B)$  is equal to 1 in any case:

$$\begin{aligned} \frac{Q(\Theta_B, \Theta_A)}{Q(\Theta_A, \Theta_B)} &= \frac{Q(\theta_k^{(*)}, \theta_k^{(t)})}{Q(\theta_k^{(t)}, \theta_k^{(*)})} \\ &= \frac{Q(\theta_k^{(t)})}{Q(\theta_k^{(*)})} \\ &= \frac{p(\theta_k^{(t)} | \theta_{-k}^{(t)}, Y_1, \dots, Y_n)}{p(\theta_k^{(*)} | \theta_{-k}^{(t)}, Y_1, \dots, Y_n)} \\ &= \frac{\frac{p(\theta_k^{(t)}, \theta_{-k}^{(t)}, Y_1, \dots, Y_n)}{p(\theta_{-k}^{(t)}, Y_1, \dots, Y_n)}}{\frac{p(\theta_k^{(*)}, \theta_{-k}^{(t)}, Y_1, \dots, Y_n)}{p(\theta_{-k}^{(t)}, Y_1, \dots, Y_n)}} \\ &= \frac{p(\theta_k^{(t)}, \theta_{-k}^{(t)}, Y_1, \dots, Y_n)}{p(\theta_k^{(*)}, \theta_{-k}^{(t)}, Y_1, \dots, Y_n)} \\ &= \frac{p(\Theta_A, Y_1, \dots, Y_n)}{p(\Theta_B, Y_1, \dots, Y_n)} \\ &= \frac{p(\Theta_A | Y_1, \dots, Y_n)}{p(\Theta_B | Y_1, \dots, Y_n)} \cdot \frac{p(Y_1, \dots, Y_n)}{p(Y_1, \dots, Y_n)} \\ &= \frac{p(\Theta_A | Y_1, \dots, Y_n)}{p(\Theta_B | Y_1, \dots, Y_n)}. \end{aligned}$$

We can then plug this expression into the formula for the acceptance probability:

$$\begin{aligned} A(\Theta_A, \Theta_B) &= \min\left\{1, \frac{p(\theta_B | y_1 \dots y_n)}{p(\theta_A | y_1 \dots y_n)} \cdot \frac{Q(\Theta_B, \Theta_A)}{Q(\Theta_A, \Theta_B)}\right\} \\ &= \min\{1, 1\} \\ &= 1. \end{aligned}$$

It follows that every sample  $\theta_k^{(*)}$  is accepted in Gibbs sampling, since we accept the random number  $u$  from a uniform distribution on  $[0, 1]$  when  $u \leq A$  and  $A = 1$ . This is the reason that Gibbs-MCMC sampling is more efficient than MH-MCMC sampling.

If we let  $\Theta$  be a  $K$ -dimensional parameter space and  $\Theta^{(1)} = (\theta_1^{(1)}, \dots, \theta_K^{(1)})$  be the starting point, then the Gibbs sampler generates  $\Theta^{(t+1)}$  from  $\Theta^{(t)}$  as follows:

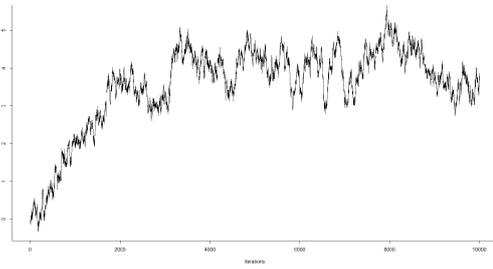
Step 1: sample  $\theta_1^{(t+1)} \sim p(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_K^{(t)})$   
 Step 2: sample  $\theta_2^{(t+1)} \sim p(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_K^{(t)})$   
 $\vdots$   
 Step  $K$ : sample  $\theta_K^{(t+1)} \sim p(\theta_K | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{K-1}^{(t+1)})$ .

This algorithm generates dependent sequences of vectors:

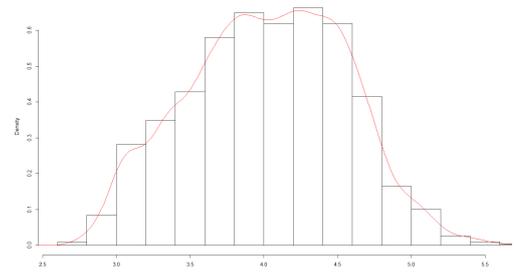
$$\begin{aligned} \Theta^{(1)} &= \{\theta_1^{(1)}, \dots, \theta_K^{(1)}\} \\ \Theta^{(2)} &= \{\theta_1^{(2)}, \dots, \theta_K^{(2)}\} \\ &\vdots \\ \Theta^{(2T)} &= \{\theta_1^{(2T)}, \dots, \theta_K^{(2T)}\}. \end{aligned}$$

We have then obtained a chain from the posterior  $p(\theta_1, \dots, \theta_K | y_1, \dots, y_n)$ , where the chain "walks" through  $2T$  states and the parameter space consists of  $K$  parameters. Again, we can approximate what we are interested in by means of a Monte Carlo approximation from here.

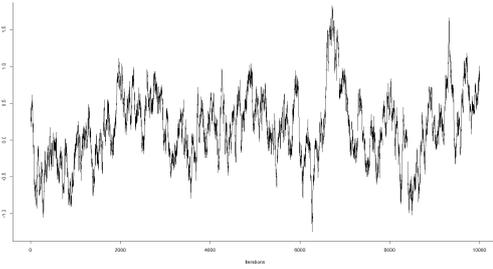
## Appendix F



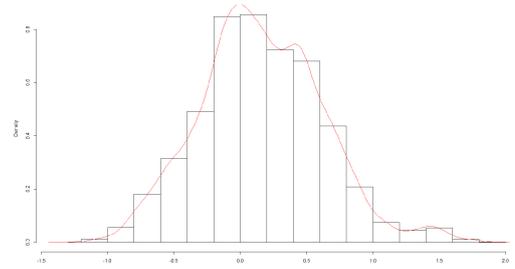
(a) Trace plot of parameter  $\eta_{42}$



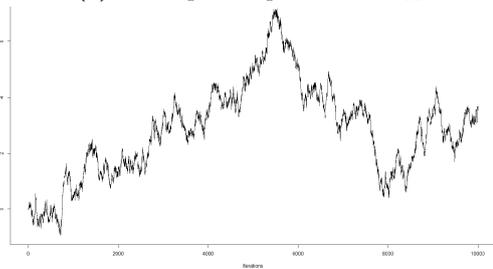
(b) Histogram with density curve of parameter  $\eta_{42}$



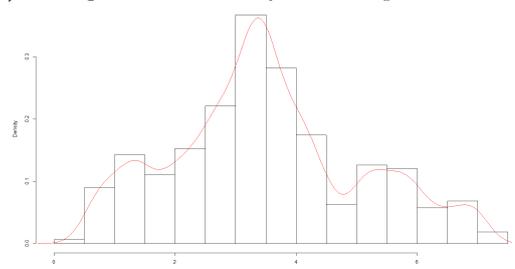
(c) Trace plot of parameter  $E_{11}$



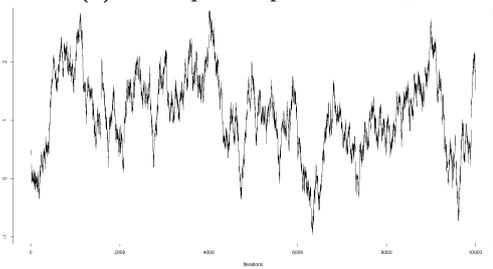
(d) Histogram with density curve of parameter  $E_{11}$



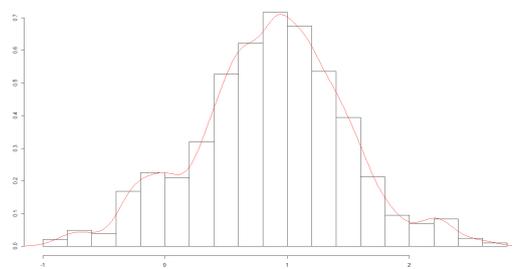
(e) Trace plot of parameter  $\theta_{11}$



(f) Histogram with density curve of parameter  $\theta_{11}$



(g) Trace plot of parameter  $A_{30,1}$



(h) Histogram with density curve of parameter  $A_{30,1}$

**Figure 20: Trace plots and histograms with a fitted density curve of various model parameters. The plots are based on one MCMC chain of 1,000,000 iterations with a thinning factor of 100. The histograms correspond to samples after a burn-in phase of 500,000 iterations.**

## Appendix G

One way to improve the R code for the MH-MCMC algorithm is to use the specific formula of the acceptance probability to move to a different state for each parameter. For example, consider the sub-step that proposes to update the parameter  $\mu_p$  from  $\Theta^{(t)}$  to  $\Theta^{(*)}$ , where

$$\begin{aligned}\Theta^{(t)} &= (A_i^{(t+1)}, \{E_{i3}^{(t+1)}\}, \{E_{i4}^{(t+1)}\}, \{A_{i2}^{(t+1)}\}, \{A_{i1}^{(t+1)}\}, \{\theta_{i1}^{(t+1)}\}, \{\theta_{i2}^{(t+1)}\}, \{\eta_i^{(t+1)}\}, \mu_0^{(t+1)}, \mu_p^{(t)}, h^{(t)}, e^{(t)}, \rho^{(t)}) \\ \Theta^{(*)} &= (A_i^{(t+1)}, \{E_{i3}^{(t+1)}\}, \{E_{i4}^{(t+1)}\}, \{A_{i2}^{(t+1)}\}, \{A_{i1}^{(t+1)}\}, \{\theta_{i1}^{(t+1)}\}, \{\theta_{i2}^{(t+1)}\}, \{\eta_i^{(t+1)}\}, \mu_0^{(t+1)}, \mu_p^{(*)}, h^{(t)}, e^{(t)}, \rho^{(t)})\end{aligned}$$

Then, it follows from equation 11 that the acceptance probability of moving from  $\Theta^{(t)}$  to  $\Theta^{(*)}$  is given by

$$\begin{aligned}A(\Theta^{(t)}, \Theta^{(*)}) &= \min\left\{1, \frac{p(\Theta^{(*)}|y_1 \dots y_n)}{p(\Theta^{(t)}|y_1 \dots y_n)}\right\} \\ &= \min\left\{1, \frac{\prod_{i=1}^N \prod_{j=3}^4 \prod_{k=1}^K p(Y_{ijk}|A_i^{(t+1)}, E_{ij}^{(t+1)}, h^{(t)}, e^{(t)}) \dots \dots p(\eta_i^{(t+1)}|\mu_p^{(*)}, \rho^{(t)}, h^{(t)}, e^{(t)})}{\prod_{i=1}^N \prod_{j=3}^4 \prod_{k=1}^K p(Y_{ijk}|A_i^{(t+1)}, E_{ij}^{(t+1)}, h^{(t)}, e^{(t)}) \dots \dots p(\eta_i^{(t+1)}|\mu_p^{(t)}, \rho^{(t)}, h^{(t)}, e^{(t)})} \cdot \frac{p(\mu_p^{(*)}) \dots \dots p(\rho^{(t)})}{p(\mu_p^{(t)}) \dots \dots p(\rho^{(t)})}\right\} \\ &= \min\left\{1, \frac{p(\eta_i^{(t+1)}|\mu_p^{(*)}, \rho^{(t)}, h^{(t)}, e^{(t)}) \cdot p(\mu_p^{(*)})}{p(\eta_i^{(t+1)}|\mu_p^{(t)}, \rho^{(t)}, h^{(t)}, e^{(t)}) \cdot p(\mu_p^{(t)})}\right\}\end{aligned}$$

Computing the acceptance probability for each parameter in a similar way will result in a code that is computationally less expensive.