
Fitting Genetic Models Using Markov Chain Monte Carlo Algorithms With BUGS

Stéphanie M. van den Berg, Leo Beem, and Dorret I. Boomsma

Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

Maximum likelihood estimation techniques are widely used in twin and family studies, but soon reach computational boundaries when applied to highly complex models (e.g., models including gene-by-environment interaction and gene-environment correlation, item response theory measurement models, repeated measures, longitudinal structures, extended pedigrees). Markov Chain Monte Carlo (MCMC) algorithms are very well suited to fit complex models with hierarchically structured data. This article introduces the key concepts of Bayesian inference and MCMC parameter estimation and provides a number of scripts describing relatively simple models to be estimated by the freely obtainable BUGS software. In addition, inference using BUGS is illustrated using a data set on follicle-stimulating hormone and luteinizing hormone levels with repeated measures. The examples provided can serve as stepping stones for more complicated models, tailored to the specific needs of the individual researcher.

Currently, the most popular method of fitting variance components models to twin data is maximum likelihood (ML) estimation within the framework of structural equation modeling (SEM) with multivariate normally distributed latent variables. Several software packages can be used, such as Mx, MPlus, EQS, and LISREL. Alternatively, the usual twin models can be formulated as linear mixed effects models and fitted by procedures in, for example, SPSS, Stata, SAS or S-Plus. For continuous univariate and multivariate normal data these methods are often theoretically satisfactory and practically feasible. They are often also reasonable methods even for nonnormal continuous variables.

A natural extension to discrete data are generalized linear mixed models (GLMM), which are generalized linear models with continuous random effects (McCulloch & Searle, 2001), or generalized linear latent variable models (Skron dal & Rabe-Hesketh, 2004). In genetics, discrete data such as dichotomous or ordered categorical data are often fitted by threshold models. Packages that allow the estimation of generalized linear mixed models (e.g., the GLLMM routine in STATA, Rabe-Hesketh & Skron dal, 2005) and most SEM packages can handle threshold models or other models where latent variables are assumed underlying observed dichotomous

or ordered categorical data. The likelihood for such models contains integrals that cannot be evaluated in a closed form (in contrast to the likelihood for linear mixed models). For relatively simple models, it is feasible to directly evaluate those integrals numerically. However, some serious computational challenges must be met when the models become more complex and many latent variables or random effects are assumed, as then the solution of the ML estimation equations require the repeated numerical evaluation of high-dimensional integrals.

For elaborate (hierarchically structured) data with many latent variables, alternative methods have been developed that use random draws of values from the (high-dimensional) distributions involved to evaluate the integrals, rather than direct evaluation of the integrals. The main ingredients of these methods are the Metropolis-Hastings algorithm and in particular its special case, the Gibbs sampler. The Gibbs sampler has its roots in image processing (Geman & Geman, 1984). The algorithms found their application especially in Bayesian models, although they have also been used in classical frequentist models. From a purely formal point of view, the main difference is that in Bayesian models a distribution is specified not only for the random latent variables, but also for parameters which would be considered fixed in a frequentist model. The methods are now often called Markov Chain Monte Carlo (MCMC) methods or algorithms.

The MCMC methodology has been successively applied for twin models with gene-environment (G-E) correlation and G-E interaction (Eaves & Erkanli, 2003), a multivariate threshold model for polytomous data (van den Berg et al., 2006) and twin models that include measurement models based on item response theory (IRT; Eaves et al., 2005, 2004; van den Berg et al., submitted). MCMC estimation can be easily performed using the BUGS family of software programs that is currently freely obtainable

Received 11 November, 2005; accepted 17 January, 2006.

Address for correspondence: S. M. van den Berg, Vrije Universiteit Amsterdam, Department of Biological Psychology, Van der Boechorststraat 1, PO Box 1081 BT, Amsterdam, the Netherlands. E-mail: SM.van.den.Berg@psy.vu.nl

through the Internet. This article introduces some key concepts behind MCMC estimation and provides the genetic researcher with a number of BUGS scripts and example data (see <http://www.psy.vu.nl/mxbib/>). The scripts are for relatively simple models that are as conveniently estimated using ML, but they are merely given for illustrative purposes. Scripts for more elaborate models that are *not* conveniently estimated using ML techniques can then be based on one of these simple scripts. Finally, the practical aspects of parameter estimation using BUGS are then illustrated with a model for bivariate continuous data with repeated measures.

MCMC and Bayesian Statistical Inference

MCMC estimation is most often applied in a Bayesian statistical framework. In order to understand modeling using BUGS software and be able to interpret the output, it is necessary to have a basic understanding of Bayesian statistical inference. To illustrate Bayesian statistics, consider the difference between the frequentist approach to inference, where one is often interested in a p value, the conditional probability of observing the data or more extreme data given a particular hypothesis is true, $P(Y \geq c \mid H_0 \text{ is true})$, and the Bayesian approach to inference, where one is interested to know the conditional probability that a particular hypothesis is true given the observed data, $P(H_1 \text{ is true} \mid Y)$. Suppose that all that is known beforehand is that the average IQ in the Netherlands must be somewhere between 0 and 200, and that after we gather IQ data from a random Dutch sample we find an average IQ of 110; a Bayesian statistician might then want to know, given the assumptions and the data, the probability that the true average IQ in the Netherlands is higher than 100. This might indeed be a more interesting question than asking about the probability of observing an average of 110 or more, given that the population average is 100 (which of course we do not know to be true or not). The conditional probability of a hypothesis being true given the data is termed the *posterior probability*. The *posterior probability density function* is a description of the probabilities of possible values for a parameter or set of parameters given the observed data and forms the basis for Bayesian statistical inference.

More formally, Bayesian inference is based on the posterior density function of the model parameters, $P(\eta \mid Y)$, where η represents the model parameters and Y the observed data. By Bayes' rule (a simple consequence of the identity $P(A \& B) = P(A \mid B) P(B) = P(B \mid A) P(A)$), the density $P(\eta \mid Y)$ is proportional (\propto) to the product of the likelihood of the data given the model parameter $P(Y \mid \eta)$ and the marginal density for η , $P(\eta)$, that is,

$$P(\eta \mid Y) \propto P(Y \mid \eta) P(\eta).$$

The marginal density of η , $P(\eta)$, is termed the *prior* distribution (prior in the sense of before the data have been taken into account), and must be specified by the

analyst, either based on prior research or substantive knowledge. The model provides us with the likelihood function $P(Y \mid \eta)$, and hence the *posterior* density function of η is determined up to a proportionality constant (posterior in the sense of after the data have been taken into account).

As already stated above, the posterior density is a probabilistic description of possible values for η given the observed data and forms the basis for statistical inference. In Bayesian statistics, the posterior density quantifies the belief in the values of the parameter. As our point estimate for η , we may take the mode (analogous to the ML), the mean or the median of this density. Further, the interval between the 2.5th and the 97.5th percentile of the posterior density provides the so-called central 95% credibility region, which, given sufficient data and certain conditions, is analogous to a 95% confidence interval in the ML framework. For more on Bayesian statistics, the reader is referred to the introductions by Box and Tiao (1973) and Gelman et al. (2004).

Sometimes it is easy to determine the posterior density function analytically, but very often this is not possible. One can then use computer simulation to draw a sample of η values from the posterior density function. The mean or median of the posterior density can then be approximated by the mean or median of the sampled η values, and approximate credibility regions can be determined in a similar way. This is termed Monte Carlo. In practice, however, the joint posterior density function of all model parameters is usually quite complicated, too complicated to directly simulate and sample from. Therefore, the complete set of parameters is split up into a number of subsets in such a way that the conditional posterior density of every subset given the data and all other parameters has a tractable form and can be easily sampled from. For instance, one might be interested in the posterior density function $P(\eta, \xi \mid Y)$, with η and ξ indicating subsets of parameters, that by itself is not easy to sample from directly. It may be possible, however, to sample directly from the conditional distributions $P(\eta \mid \xi, Y)$ and $P(\xi \mid \eta, Y)$.

This approach is known as Gibbs sampling (Gelfand & Smith, 1990; Geman & Geman, 1984), which is a special case of the general class of MCMC algorithms. When, however, the conditional posterior density of a subset of the parameters is not easy or even impossible to sample from directly, other MCMC algorithms can be used, where one samples from a similar proposal distribution and uses a decision rule to either accept or reject a sample so that the accepted values can be regarded as randomly sampled values from the target distribution (Robert & Casella, 2004).

An MCMC algorithm starts with some starting values, either randomized or provided by the analyst. Next a number of so-called 'burn-in' iterations are required, which are necessary for an MCMC chain to achieve stationarity (i.e., for approaching the joint posterior or target distribution sufficiently closely). In

each iteration, a new value is sampled from each conditional distribution based on the most recent values for the other parameters in the model. Thus, using the above Gibbs example, in iteration k we first sample $\eta_k \sim P(\eta \mid \xi_{k-1}, Y)$ and then $\xi_k \sim P(\xi \mid \eta_k, Y)$. It can be shown that after a number of iterations, MCMC algorithms generally end up by sampling from the joint posterior distribution. Then the subsequent draws can be regarded as sampled from the joint posterior distribution. The sampling scheme approaches the joint posterior distribution only asymptotically though; it is up to the analyst to decide when the MCMC samplings approach the target distribution sufficiently closely. Note that such a decision (i.e., a stopping rule) is also used with most ML estimation algorithms, although there the algorithms converge to a value, whereas the MCMC algorithm in a Bayesian analysis converges to a distribution. With an adequate number of sampled values for each parameter, the average, the mode or the median of these sampled parameter values can be used as a point estimate.

Fitting Models With BUGS

BUGS refers to a range of general purpose software packages for Bayesian MCMC estimation for different platforms that can be freely downloaded from the Internet (<http://www.mrc-bsu.cam.ac.uk/bugs/>). A large class of models can be fitted by BUGS. The user must specify the data and the model, including the specification of the prior distribution for the unknown model parameters. The software then automatically sets up an MCMC sampling algorithm and the user can then monitor the samples that are iteratively drawn and use these for statistical inference. A BUGS program that runs under Windows is WinBUGS.

A model can be specified using either a script or a graphical representation (see Eaves & Erkanli, 2003, for an example). Here we only use scripts. The way one specifies the model and sets up the script is very much like simulating data sets in R or S-Plus. Some familiarity with linear mixed effects model notation is helpful (see, for example, McCulloch & Searle, 2001). The scripts with example data can be downloaded from <http://www.psy.vu.nl/mxbib/>

The Univariate ACE Model

Let us first start with a simple univariate model for continuous data, which uses the classical twin design to estimate the additive genetic variance, shared environmental variance and nonshared environmental variance (colloquially termed the ACE model). Using Greek letters for fixed effects and Latin letters for random effects, we can parameterize the phenotype of subject j in family i as a linear mixed effects model

$$\text{Phenotype}_{ij} = \mu + a1_i + a2_{ij} + c_i + e_{ij}$$

where μ indicates the population mean, c_i denotes the random effect for being a member of family i , the additive genetic effect is modeled by the random effects $a1_i$

and $a2_{ij}$, and e_{ij} denotes a random (environmental) effect for individual j in family i . All random effects are assumed independent unless their symbol and subscripts are identical. The additive genetic effect is split into $a1_i$ and $a2_{ij}$ in order to model the different genetic correlations among monozygotic (MZ) and dizygotic (DZ) twins (cf. Jinks & Fulker, 1970). The genetic variance is the same for DZ and MZ twins, while the genetic covariance in MZ twins is twice as large as in DZ twins. We now incorporate the restriction $\text{VAR}(a1) = \text{VAR}(a2) = 1/2 \sigma_a^2$, so that the total additive genetic variance becomes σ_a^2 as the random effects are independent in our model. For DZ twins the index j differs (i.e., $j = 1, 2$), whereas it is the same for MZ twins. The genetic covariance of DZ twins is then the genetic between-family variance, $\text{cov}(a1_j, a1_i) = \text{Var}(a1) = 1/2 \sigma_a^2$, and the genetic covariance for MZ twins is $\text{cov}(a1_i + a2_{ij}, a1_i + a2_{ij}) = \text{cov}(a1_j, a1_i) + \text{cov}(a2_{ij}, a2_{ij}) = \text{Var}(a1) + \text{Var}(a2) = \sigma_a^2$, as $a2_{ij}$ is the same for MZ twins. Similarly, $\text{Var}(c)$ can be interpreted as the between-families environmental variance and $\text{Var}(e)$ can be interpreted as the within-family environmental variance. We assume normal distributions for all random effects such that $a1 \sim N(0, 1/2 \sigma_a^2)$, $a2 \sim N(0, 1/2 \sigma_a^2)$, $c \sim N(0, \sigma_c^2)$, and $e \sim N(0, \sigma_e^2)$.

An alternative but equivalent description of the ACE model for twin pairs follows the logic of simulating a data set. For each DZ twin pair i we sample a normally distributed environmental effect that is the same for both twins ('shared'), $c_i \sim N(0, \sigma_c^2)$. We then sample for each pair a genetic effect that is the same for both twins, $a1_i \sim N(0, 1/2 \sigma_a^2)$. For each individual twin j from DZ pair i , we sample a genetic effect that is unique for each twin, $a2_{ij} \sim N(0, 1/2 \sigma_a^2)$ and a unique environmental effect, $e_{ij} \sim N(0, \sigma_e^2)$. Then the phenotypic value for twin j from pair i can be written as the sum of all these effects plus the fixed effect μ , just as shown above.

Yet another way to describe the model is a hierarchical-centred representation. Such a formulation may lead to more efficient MCMC sampling, as it can reduce or sometimes effectively nullify the correlations among parameters in the posterior distribution (Gelfand et al., 1995). In mixed models the random effects are usually specified as having a zero mean (see above). In the centered specifications, the distribution of random effects may contain means that are not equal to zero. Hierarchical centering is fairly unambiguous in completely nested models, but not for models that are not or only partially nested, as for example the ACE model. For the ACE model we may specify for MZ pairs

$$c_i \mid \mu \sim N(\mu, \sigma_c^2), \quad f_i \mid \mu, c_i \sim N(c_i, \sigma_a^2), \\ ph_{ij} \mid \mu, c_i, f_i \sim N(f_i, \sigma_e^2)$$

or the first two specifications may be changed into

$$a_i \mid \mu \sim N(\mu, \sigma_a^2), \quad f_i \mid \mu, a_i \sim N(a_i, \sigma_e^2)$$

In both cases, f_i may be interpreted as the result of the combined effects of shared environment and genotype. Note that the means of the distributions are now replaced by parameters. The first specification may improve convergence if σ_c^2 is relatively large, the latter if σ_a^2 is relatively large. For DZ twin pairs, the only difference is that we need to model the genetic correlation of .5. This is accomplished by adding an extra level $f2_{ij}$ to the model:

$$\begin{aligned} c_i | \mu &\sim N(\mu, \sigma_c^2), & f_i | \mu, c_i &\sim N(c_i, 1/2 \sigma_a^2), \\ f2_{ij} | f_p, \mu, c_i &\sim N(f_p, 1/2 \sigma_a^2), \\ ph_{ij} | \mu, c_p, f_p, f2_{ij} &\sim N(f2_{ij}, \sigma_e^2) \end{aligned}$$

As we are performing a Bayesian analysis we have to specify prior distributions for the parameters. Let us assume we have no prior information about the estimates of interest. Therefore we extend the model by adding relatively noninformative uniform distributions for the square-root of the variance components (the standard deviations, following Gelman et al., 2004), $\sigma_e \sim U(0, 100)$, $\sigma_c \sim U(0, 100)$, $\sigma_a \sim U(0, 100)$, and a locally ‘flat’ normal prior for the population mean, $\mu \sim N(0, 100)$, so that the joint posterior distribution is mainly determined by the likelihood.

It is generally advised to try out different kinds of priors to assess to what degree the estimation depends on them. One can for instance use a uniform prior for the mean, $\mu \sim U(-100, 100)$, or normal priors that are flatter (larger variance) or more pointed (smaller variance). If the results are highly sensitive to the priors, this is an indication that the data provide relatively little information about the parameters.

The ADE Model

When we want to specify a model with both additive (A) and nonadditive (D) genetic effects, apart from nonshared environmental effects (E), we have for MZ twins $a_i \sim N(\mu, \sigma_a^2)$, $f_i | a_i \sim N(a_i, \sigma_d^2)$, $ph_{ij} | a_i, f_i \sim N(f_i, \sigma_e^2)$. In order to model a correlation of .25 in the DZ twins for the nonadditive (dominance) genetic effects we may split up the D component, σ_d^2 , into two parts, $1/4 \sigma_d^2$ and $3/4 \sigma_d^2$ as follows:

$$\begin{aligned} a_i &\sim N(\mu, 1/2 \sigma_a^2), & f_i | a_i &\sim N(a_i, 1/4 \sigma_d^2), \\ f1_{ij} | a_p, f_i &\sim N(f_i, 1/2 \sigma_d^2), \\ f2_{ij} | a_p, f_i, f1_{ij} &\sim N(f1_{ij}, 3/4 \sigma_d^2), \\ ph_{ij} | a_p, f_p, f1_{ij}, f2_{ij} &\sim N(f2_{ij}, \sigma_e^2) \end{aligned}$$

For some estimation problems, an alternative representation of the model may be more convenient to speed up the sampling by reducing autocorrelation in the MCMC chains. High autocorrelation means that a particular draw from the posterior distribution is highly correlated with the previous draw. As a result, the sampler moves only very slowly through the entire posterior distribution, requiring many iterations to get reliable information about the posterior distribution. For the ADE model with continuous normal data it is more effi-

cient to model the sums and differences for each twin pair i instead of modeling the observed phenotypic values (Robert & Casella, 2004, p. 396; cf. Boomsma & Molenaar, 1987): $S_i = Phenotype_{i1} + Phenotype_{i2}$ and $D_i = Phenotype_{i1} - Phenotype_{i2}$. The covariance of the transformed phenotypes is $cov(P1 + P2, P1 - P2) = var(P1) - cov(P1, P2) + cov(P1, P2) - var(P2) = 0$, as $var(P1) = var(P2)$. The zero covariance is the main reason for using this transformation, as this indirectly reduces the autocorrelation. The means for these new phenotypes are $\mu_s = 2\mu$, $\mu_D = 0$ and their expected variances become

$$\begin{aligned} Var(S_{mz}) &= 4 \sigma_a^2 + 4 \sigma_d^2 + 2 \sigma_e^2 \\ Var(D_{mz}) &= 2 \sigma_e^2 \\ Var(S_{dz}) &= 3 \sigma_a^2 + 2.5 \sigma_d^2 + 2 \sigma_e^2 \\ Var(D_{dz}) &= \sigma_a^2 + 1.5 \sigma_d^2 + 2 \sigma_e^2 \end{aligned}$$

First, the data set should be transformed in order to have for each family the sum and the difference of the phenotypes (be careful not to take the absolute differences). We next set up the model as follows. For DZ twin pairs, we have $Sdz_i \sim N(2\mu, 3 \sigma_a^2 + 2.5 \sigma_d^2 + 2 \sigma_e^2)$ and $Ddz_i \sim N(0, \sigma_a^2 + 1.5 \sigma_d^2 + 2 \sigma_e^2)$ and for MZ twin pairs we have $Smz_i \sim N(2\mu, 4 \sigma_a^2 + 4 \sigma_d^2 + 2 \sigma_e^2)$ and $Dmz_i \sim N(0, 2 \sigma_e^2)$.

For incomplete pairs, for which only one twin is available, or for unrelated singletons, we can only model the distribution of the original phenotypes as $Phenotype_{i1} \sim N(\mu, \sigma_a^2 + \sigma_d^2 + \sigma_e^2)$. A more general approach to model reparameterization to reduce the correlations among parameters in the posterior distribution is presented in Appendix A.

Threshold Models

The threshold model (Crittenden, 1961; Falconer, 1965) is often applied in order to estimate the heritability of continuous but categorically measured traits. In this model, an underlying continuous trait is defined, usually termed liability, with a threshold value above which the person is diagnosed. The threshold model is appropriate if, for example, it is reasonable to assume that many independent factors each have a relatively small effect on the liability. Under the central limit theorem, the resulting liability will then have a normal distribution. The above ACE and ADE models must then be extended by introducing an additional level, which links the normally distributed latent liability to an observed dichotomous trait through a step function. For DZ twins and an ACE model we have

$$c_i \sim N(\mu, \sigma_c^2), f_i | c_i \sim N(c_i, 1/2 \sigma_a^2), f2_{ij} \sim N(f_i, 1/2 \sigma_a^2), liability_{ij} | c_i, f_i, f2_{ij} \sim N(f2_{ij}, \sigma_e^2)$$

$$PhenotypeDZ_{ij} = \begin{cases} 1 & \text{if } liability_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that in this parameterization, we fix the threshold to be zero and estimate the average liability in the population. This is equivalent to a parameterization

where we fix the average liability to zero and estimate a threshold. Here we also need the additional constraint that the total variance of the unobserved liability is somehow fixed, for example by constraining $\sigma_e^2 = 1$. For an example of a multivariate model for polytomous data, see van den Berg et al. (2006).

An Application

For an illustration of how statistical inferences can be made using BUGS we used data on hormone levels from the Netherlands Twin Registry (Boomsma et al., 2002). Follicle-stimulating hormone (FSH) and luteinizing hormone (LH) levels were based on morning urine samples from 9-year-old MZ (37 pairs) and DZ twins (44 pairs) collected twice, 2 days apart. Many LH levels (46% of the cases) could not be detected ($< .1$ U/L) and these censored data were included into the analysis by assigning these a value of .05 U/L. Next, LH and FSH levels in units (U) per mmol Creatinine (Cr) showed positively skewed distributions and were therefore transformed by taking the natural logarithm before statistical analysis.

Prior to applying genetic models, it is wise to first get an idea of the twin correlations. This can be accomplished by applying a simple model $Phenotype_{ij} = f_i + e_{ij}$ such that $f_i \sim N(\mu, \sigma_{between}^2)$ and $Phenotype_{ij} \sim N(f_i, \sigma_{within}^2)$, to obtain an estimate for the twin correlation, namely $r = \sigma_{between}^2 / (\sigma_{between}^2 + \sigma_{within}^2)$ for a certain phenotype. This can be done for MZ and DZ twins separately. For multivariate datasets, see van den Berg et al. (2006). This approach can be applied where one wishes to estimate MZ and DZ correlations on latent traits, if for example the latent trait is linked to a number of dichotomous or polychotomous items such as in an IRT model (van den Berg et al., submitted).

In this case, DZ correlations were about half the MZ correlations (not shown), so we specified a bivariate AE model in order to estimate genetic and environmental correlations between FSH and LH levels. For the j th individual of the i th MZ pair we modeled the hormone levels as resulting from genetic and environmental influences that are the same for LH and FSH levels (common), and genetic and environmental influences that are hormone specific:

$$\begin{aligned} A_common_i &\sim N(0, \sigma_{ac}^2) \\ A_FSH_i &\sim N(A_common_i, \sigma_{aFSH}^2) \\ A_LH_i &\sim N(A_common_i, \sigma_{aLH}^2) \\ E_common_{ij} &\sim N(0, \sigma_{ec}^2) \\ FSH_{ij} &\sim N(\mu_{FSH} + A_FSH_i + E_common_{ij}, \sigma_{eFSH}^2) \\ LH_{ij} &\sim N(\mu_{LH} + A_LH_i + E_common_{ij}, \sigma_{eLH}^2) \end{aligned}$$

Moreover, since we have two measures for each hormone level, we specify that we have two random samples, $k = (1, 2)$, from each individual's 'true' level as follows:

$$\begin{aligned} FSH_measure_{ijk} &\sim N(FSH_{ij}, \sigma_{errorFSH}^2) \\ LH_measure_{ijk} &\sim N(LH_{ij}, \sigma_{errorLH}^2) \end{aligned}$$

with $\sigma_{errorFSH}^2$ and $\sigma_{errorLH}^2$ representing measurement variance. The resulting variance decomposition can now be based on the estimated 'true' FSH and LH levels, thereby excluding unsystematic measurement error. Consequently, the nonshared environmental variance component will be estimated to be smaller than if the analysis were based on only one urine sample.

The model for the hormone measurements in DZ twin pair i is as follows, where the additive genetic components are again split into between and within family components:

$$\begin{aligned} A_between_common_i &\sim N(0, 1/2 \sigma_{ac}^2) \\ A_between_LH_i &\sim N(0, 1/2 \sigma_{aLH}^2) \\ A_between_FSH_i &\sim N(0, 1/2 \sigma_{aFSH}^2) \\ A_common_{ij} &\sim N(A_between_common_i, 1/2 \sigma_{ac}^2) \\ A_LH_{ij} &\sim N(A_between_LH_i, 1/2 \sigma_{aLH}^2) \\ A_FSH_{ij} &\sim N(A_between_FSH_i, 1/2 \sigma_{aFSH}^2) \\ E_common_{ij} &\sim N(0, \sigma_{ec}^2) \\ LH_{ij} &\sim N(\mu_{LH} + A_LH_{ij} + A_common_{ij} + E_common_{ij}, \sigma_{eLH}^2) \\ FSH_{ij} &\sim N(\mu_{FSH} + A_FSH_{ij} + A_common_{ij} + E_common_{ij}, \sigma_{eFSH}^2) \\ LH_measure_{ijk} &\sim N(LH_{ij}, \sigma_{errorLH}^2) \\ FSH_measure_{ijk} &\sim N(FSH_{ij}, \sigma_{errorFSH}^2) \end{aligned}$$

The total variance for each of the hormone level measurements can now be decomposed into additive genetic variance shared with the other hormone level, hormone-specific additive genetic variance, environmental variance shared with the other hormone level, hormone specific environmental variance, and variance due to random measurement error.

$$\begin{aligned} \sigma_{LH}^2 &= \sigma_{ac}^2 + \sigma_{aLH}^2 + \sigma_{ec}^2 + \sigma_{eLH}^2 + \sigma_{errorLH}^2 \\ \sigma_{FSH}^2 &= \sigma_{ac}^2 + \sigma_{aFSH}^2 + \sigma_{ec}^2 + \sigma_{eFSH}^2 + \sigma_{errorFSH}^2 \end{aligned}$$

When interested in the genetic background of the 'true' individual hormone levels (i.e., LH_{ij} and FSH_{ij} in the model), we may ignore the measurement variance so that the expected genetic correlation on the basis of this model is $r_g = \sigma_{ac}^2 / [\sqrt{(\sigma_{ac}^2 + \sigma_{aLH}^2)} \sqrt{(\sigma_{ac}^2 + \sigma_{aFSH}^2)}]$. Similarly, the expected correlation between the environmental influences other than random measurement error is $r_e = \sigma_{ec}^2 / [\sqrt{(\sigma_{ec}^2 + \sigma_{eLH}^2)} \sqrt{(\sigma_{ec}^2 + \sigma_{eFSH}^2)}]$. The expected heritability coefficients for expected or 'true' LH and FSH level are

$$\begin{aligned} h_{LH}^2 &= (\sigma_{ac}^2 + \sigma_{aLH}^2) / (\sigma_{ac}^2 + \sigma_{aLH}^2 + \sigma_{ec}^2 + \sigma_{eLH}^2), \text{ and} \\ h_{FSH}^2 &= (\sigma_{ac}^2 + \sigma_{aFSH}^2) / (\sigma_{ac}^2 + \sigma_{aFSH}^2 + \sigma_{ec}^2 + \sigma_{eFSH}^2) \end{aligned}$$

The model was implemented in WinBUGS using locally uniform and independent priors for the square-roots of the variance component parameters and normal (but locally uniform, thus uninformative) priors for the average hormone levels (see <http://www.psy.vu.nl/mxbib/> for the script). After the data were loaded, two independent Markov chains

Table 1

Means, Standard Deviations and Quantiles of the Posterior Distributions for Several Model Parameters

BUGS name	Model name	Mean	SD	2.5th percentile	50th percentile	97.5th percentile
mu1	μ_{LH}	-4.80	0.07	-4.95	-4.80	-4.66
mu2	μ_{FSH}	-1.24	0.07	-1.38	-1.24	-1.10
re	r_e	0.06	0.05	0.00	0.05	0.17
rg	r_g	0.48	0.07	0.34	0.48	0.62
var_A1	σ_{eLH}^2	0.31	0.08	0.18	0.30	0.49
var_A2	σ_{eFSH}^2	0.30	0.07	0.18	0.29	0.45
var_AC	σ_{ac}^2	0.28	0.06	0.18	0.27	0.42
var_E	σ_{bc}^2	0.01	0.01	0.00	0.00	0.02
var_E1	σ_{eLH}^2	0.09	0.04	0.03	0.08	0.17
var_E2	σ_{eFSH}^2	0.09	0.03	0.04	0.08	0.16
var_rep1	$\sigma_{errorLH}^2$	0.22	0.02	0.18	0.22	0.27
var_rep2	$\sigma_{errorFSH}^2$	0.10	0.01	0.08	0.09	0.12

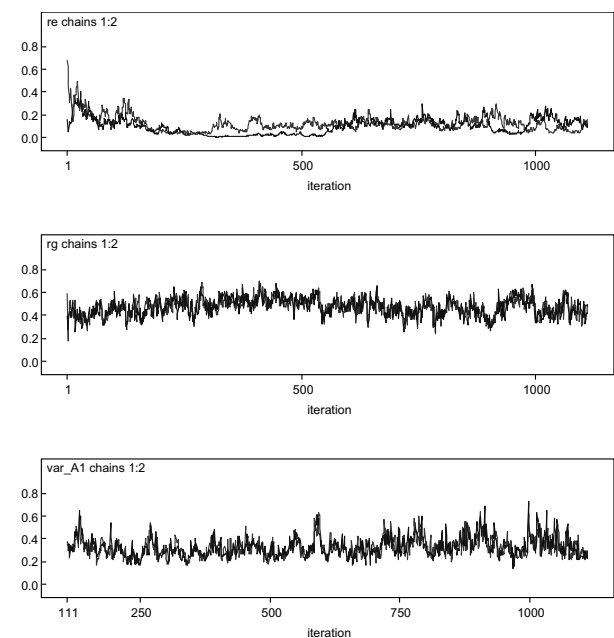
Note: Results based on two independent Markov chains, each consisting of 8000 iterations after a burn-in phase of 1000 iterations.

were initialized using randomized starting values. By putting the equations of interest (e.g., genetic and environmental correlations) already in the script, one is able to directly monitor these ‘nodes’ (as BUGS calls parameters or functions of them) while sampling from the joint posterior distribution. The randomized starting values produced rather dissimilar initial values for the environmental correlation (re). However, as can be seen from Figure 1, over the first few draws the values for the independent draws draw nearer. After that, the chains seem to converge to the same sampling distribution: lines are continuously crossing each other and show more or less the same variance. In order to be sure that the chains have approached the target posterior distribution, one needs to assess whether the other parameters are converging too. Note that the chains for the environmental correlation look a bit odd: at times the chains seem to get stuck close to zero. This behavior is often seen with estimates of variance components that are zero or close to zero. In addition, the chains show high autocorrelation.

Apart from whether independent Markov chains with different sets of starting values seem to converge by visual inspection, there are other ways to assess sufficient convergence. For example, BUGS can plot running quantiles. Figure 2 shows that the two chains show largely the same distribution (left panel) and this is also indicated by the Gelman-Rubin statistic (G-R; right panel). A G-R statistic close to 1 indicates that the chains are sampling from the same distribution, which might be the target distribution.

When we decide that the sampling chains approach the joint posterior distribution sufficiently closely, we let BUGS run a number of extra iterations to sample from in order to obtain point estimates for the parameters of interest. In this case, the relatively high autocorrelation for the environmental correlation forces us to use a rather large number of iterations. Here we first sampled from 5000 iterations, so that

we have (number of chains * number of iterations equals) 10,000 sampled values for each parameter. Figure 3 shows kernel density plots (smoothed histograms) for the genetic and environmental correlations indicating the shape of their marginal posterior distributions. The distribution of the genetic correlation looks rather smooth and suggests a nice approximation to the true posterior density. The distribution for the environmental correlation looks somewhat less pretty and rather bumpy, probably due

**Figure 1**

Sampled values up until the 1110th iteration for the environmental correlation (first panel), the genetic correlation (second panel), and the LH-specific additive genetic variance (third panel).

Note: The two differently coloured lines represent independent Markov chains.

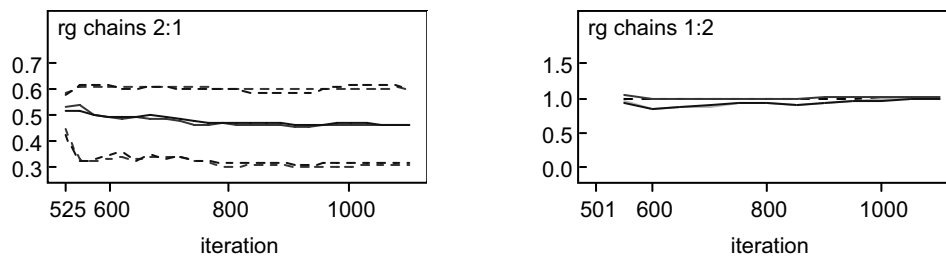


Figure 2
 Convergence indicators: plot of running 2.5th, 50th and 97.5th quantiles (left panel) and the Gelman-Rubin statistic (right panel) for the genetic correlation.
 Note: The two different lines represent independent Markov chains.

to high autocorrelation. One could run more iterations, trying to make up for this high autocorrelation. Table 1 presents means and quantiles of the marginal posterior distributions for several parameters, based on 16,000 iterations. On the basis of this analysis, one can conclude that the heritability of ‘true’ (perfectly measured) FSH level is around 85% and around 86% for LH level. The genetic correlation is estimated at .48 (95% central credibility region: .34–.62), and the environmental correlation is negligible. Since we used relatively noninformative priors, these estimates would not differ much from ML estimates and confidence intervals.

If one wishes to know whether a model without ‘common’ environmental influences on FSH and LH levels would fit the data as well (common in the sense that these influences affect both LH and FSH levels) — in other words, a model with zero environmental correlation — one could request BUGS to print the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), a parsimony-based fit index, that could be compared to the DIC index for a model where the common environmental variance is fixed to zero (i.e., delete all references to this component in the script). The model with the lowest DIC value is then the to-be-preferred model in terms of parsimony. In this case, the DIC index for the original model, including common environmental variance, was 842.19, while the DIC index for the restricted model, postulating no environmental variance common to both phenotypes, was 839.00. Thus, we may conclude that there is no

correlation between the environmental effects that influence LH and FSH levels.

Note that this model assumes that the daily fluctuations in measurements are not correlated across the two hormones, that is, that daily fluctuations in LH and FSH levels result from independent processes, and this is not necessarily the case. A full analysis should take this possibility into account, as well as dealing appropriately with the data censoring (e.g., by defining a threshold) and sex differences in means and variance components. Model fitting can be accomplished by either making use of the DIC index, or by using the 95% central credibility regions. If a hypothesized value, say zero for the difference between male and female hormone levels, is not included in the 95% central credibility region, the hypothesis that males and females show the same average level can be rejected. Note that the latter method cannot be applied with variance components since these are constrained to be zero or larger through the priors, so that the central 95% credibility interval will never include zero.

Conclusion

Using a set of simple examples, we have attempted to illustrate the use of BUGS in estimating the most common genetic twin models. Once accustomed to the language of mixed effects models, these simple models can easily be extended to more complex models, including covariates, repeated measures, a longitudinal structure, measurement models, extended family data,

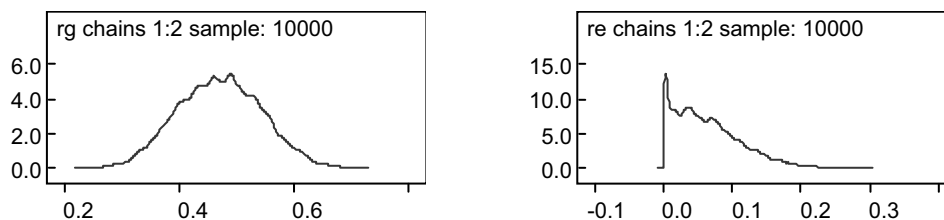


Figure 3
 Kernel density plots based on 10,000 sampled values from the posterior distributions of the genetic correlation (left panel) and the environmental correlation (right panel), based on two independent Markov chains with a burn-in phase of 1000 iterations.

sex limitation, mixtures, latent classes, and much more. It is with the more complex models that MCMC estimation can prove its value in comparison with ML estimation techniques, as ML estimation time increases exponentially with increasingly complex dependency structures. In contrast, MCMC estimation time increases at a much slower rate as a function of the number of levels in the model. The sample of WinBUGS scripts and example data at <http://www.psy.vu.nl/mxbib/> can function as a stepping stone for more complex models tailored to the specific needs of the genetic researcher. In addition, the freely obtainable BUGS software is accompanied by a large set of example scripts for many other, nongenetic types of data models that may serve as sources of inspiration.

MCMC estimation has been observed only sporadically in twin studies (but see Eaves & Erkanli, 2003; Eaves et al., 2005, 2004; van den Berg et al., submitted, 2006). MCMC algorithms allow the fitting of highly complex models easily and swiftly, particularly models with highly complex dependency structures, such as latent growth models for repeated measures on traits that are measured using a limited set of items (see, for example, Burton et al., 2005; van den Berg et al., submitted). Future work should, however, focus on alternative parameterizations for specific types of models to speed up MCMC estimation and on the assessment of model fit.

Acknowledgments

This work was supported by a grant from the Netherlands Organisation for Scientific Research, NWO 051-02-060, 668-772.

References

- Boomsma, D. I., & Molenaar, P. C. M. (1987). Using LISREL to analyze genetic and environmental covariance structure. *Behavior Genetics*, *16*, 237–250.
- Boomsma, D. I., Vink, J. M., van Beijsterveldt, C. E. M., de Geus, E. J., Beem, A. L., Mulder, E. J., Derks, E. M., Riese, H., Willemsen, G. A., Bartels, M., van den Berg, M., Kupper, N. H., Polderman, T. J., Posthuma, D., Rietveld, M. J., Stubbe, J. H., Knol, L. I., Stroet, T., & van Baal, G. C. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research*, *5*, 401–406.
- Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MS: Addison-Wesley.
- Burton, P. R., Scurrah, K. J., Tobin, M. D., & Palmer, L. J. (2005). Covariance components models for longitudinal family data. *International Journal of Epidemiology*, *34*, 1063–1077.
- Crittenden, L. B. (1961). An interpretation of familial aggregation based on multiple genetic and environmental factors. *Annals of the New York Academy of Sciences*, *91*, 769–780.
- Eaves, L. J., & Erkanli, A. (2003). Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and $G \times E$ interaction. *Behavior Genetics*, *33*, 279–299.
- Eaves, L. J., Erkanli, A., Silberg, J., Angold, A., Maes, H. H., & Foley, D. (2005). Application of Bayesian inference using Gibbs sampling to item-response theory modeling of multi-symptom genetic data. *Behavior Genetics*, *35*, 765–780.
- Eaves, L. J., Silberg, J., Foley, D., Bulik, C., Maes, H., Erkanli, A., Angold, A., Costello, E. J., & Worthman, C. (2004). Genetic and environmental influences on the relative timing of pubertal change. *Twin Research*, *7*, 471–481.
- Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics*, *29*, 51–71.
- Gelfand, A. E., Sahu, S. K., & Carlin, B. P. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika*, *82*, 479–488.
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman and Hall.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.
- Jinks, J. L. & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, *73*, 311–349.
- McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.
- Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, *64*, 583–616.
- van den Berg, S. M., Glas, C. A. W., & Boomsma, D. I. (submitted). *Variance decomposition using an IRT measurement model*. Manuscript submitted for publication.
- van den Berg, S. M., Setiawan, A., Bartels, M., Polderman, T. J. C., Van der Vaart, A. W., & Boomsma, D. I. (2006). Individual differences in puberty onset in girls: Bayesian estimation of heritabilities and genetic correlations. *Behavior Genetics* [Epub ahead of print].

Appendix A

High autocorrelation in the MCMC chains can sometimes in part be accounted for by the high correlation in the data. This can be remedied by not modeling observed twin phenotypes, that are correlated in twin pairs, but some functions of these phenotypes that, even though they convey the same information, are no longer correlated. One solution is to use the sum and the difference of the phenotypes from a twin pair: given the usual assumptions of normal distributions and homoscedasticity, the expected sum of two twins is independent of their difference.

The transformation to sums and differences is a case of an (unscaled) Helmert transformation. The same result, uncorrelated data (i.e., a diagonal covariance matrix), can be obtained by applying the Helmert transformation to families with an arbitrary number of genetically nonidentical regular offspring (DZ twins and/or other regular offspring). The Helmert matrix H of order $n \times n$ is defined in such a way that the sum of elements in the first row equals \sqrt{n} , the elements in the other rows sum to zero, and $HH' = I$ (since H is a square matrix H' is the inverse of H , hence also $H'H = I$). For example, for $n = 3$, the three rows are $(1, 1, 1) / \sqrt{3}$, $(1, -1, 0) / \sqrt{2}$, $(1, 1, -2) / \sqrt{6}$, where l_i is the sum of the squared numbers in row i .

The rows of H contain the Helmert transformations (or contrasts), which are defined in many textbooks on analysis of variance. The values of the elements in the matrix depend only on n . To show why the Helmert transformations achieve zero covariances, we write $\sigma_1^2 = \sigma_a^2$, $\sigma_2^2 = \sigma_b^2$, $\sigma_3^2 = \sigma_c^2$, $\sigma_4^2 = \sigma_e^2$ and corresponding constants $g_1 = 1/2$, $g_2 = 1/4$, $g_3 = 1$, $g_4 = 0$. Let 1_n be a vector with all n elements equal to 1 and $J = 11'$. Then the expected covariance matrix V for a family with n offspring under the ADCE model can be written as $V = \sum_i [g_i \sigma_i^2 J + (1 - g_i) \sigma_i^2 I]$. If for the original phenotypes $y \sim N(\mu, V)$, then the transformation $z = Hy$ gives $\mu_z = H\mu$ and $V_z = HVH'$. Now $V_z = H \sum_i [g_i \sigma_i^2 J + (1 - g_i) \sigma_i^2 I] H' = \sum_i [g_i \sigma_i^2 HJH' + (1 - g_i) \sigma_i^2 I]$. Because $J = 11'$, $HJH' = H11'H'$. But $H1$ is a vector whose elements are equal to the row sums of H . Therefore the first element is \sqrt{n} and all other elements are zero. Thus HJH' is a matrix with its leading element equal to n and all other elements zero. Therefore the covariance matrix V_z becomes a sum of diagonal matrices and hence is diagonal. This transformation is especially attractive because H and its inverse are simple and do not depend on unknown parameters.

If a family contains both an MZ pair and other offspring, the covariance matrix cannot be written in the simple form $\sum_i [g_i \sigma_i^2 J + (1 - g_i) \sigma_i^2 I]$ as the covariances among the offspring are not all identical. Although the covariance matrix associated with a particular σ_i^2 may still be diagonalized independently of unknown parameters, this seems in general not to be possible for all matrices simultaneously. However, depending on the model and data it may still be useful to apply the Helmert transformation. Alternatively, preliminary estimates can be used to select a different kind of transformation.

For this type of transformation, transforming y and simultaneously redefining the covariance matrix does not change the likelihood. The parameters of the normal likelihood are contained in the expressions $(y - \mu)'V^{-1}(y - \mu)$ and the determinant $|V|$. For family data, these expressions reduce to similar expressions for each family. That is, for k families

$$(y - \mu)' V^{-1} (y - \mu) = \sum_k (y_k - \mu_k)' V_k^{-1} (y_k - \mu_k) \text{ and } |V| = \prod_k |V_k|$$

Now apply the Helmert transformation to the data for each family, which may contain a different number of offspring. Dropping the subscript k , we get for each family $z = Hy$, with $\mu_z = H\mu$, $V_z = HVH'$ and $|V_z| = |H||V||H'|$. Because H' is the inverse of H , $|H| = 1/|H'|$ and therefore $|V_z| = |V|$. For the other expression we get

$$(z - H\mu)' (HVH')^{-1} (z - H\mu) = (z - H\mu)' HV^{-1} H'(z - H\mu) = (y - \mu)' V^{-1} (y - \mu)$$

So this transformation leaves the likelihood unchanged.