Flexible Bayesian Inference of Animal Model Parameters Using BUGS Program

G. Gorjanc[*]

Introduction

Bayesian inference of animal model parameters via Markov chain Monte Carlo (McMC) methods is a part of a standard toolkit used in the field of animal breeding and genetics (Sorensen and Gianola, 2002). Several animal model programs are available in public domain for performing such analyses for various kinds of models. Although powerful and extensible, these programs do not provide an environment where various models can be easily constructed and tested. For some programs, source code is available, which provides a possibility for modifications. However, this requires a familiarity with the internal structure of programs, a considerable amount of programming and statistical skills, and time to implement the changes. An alternative is to use general purpose statistical programs. The Bayesian inference Using Gibbs Sampling (BUGS) program (Lunn et al., 2000) is such a program and the aim of this work is to show how animal model parameters can be inferred with this program.

Graphical model representation

This is not the first report on fitting animal model in BUGS. Damgaard (2007) described how to fit animal model in BUGS via Cholesky reparameterization of relationship matrix, while Waldmann (2009) took the advantage of a graphical model perspective - the prime way to describe models in BUGS. However, both reports failed to provide a generic procedure" that can be used independently of the structure of collected data. This is essential for the applied work. Here such a procedure is presented.

Data in Table 1 will serve as an example. Example is small, but contains all the peculiarities that arise in applied work. There are ten individuals in the pedigree among which some have both parents unknown, one parent unknown, or both parents known. For some individuals phenotypic records and group membership are known. Individual 2 has two repeated records. For this data animal model with standard assumptions (1) could be used (Sorensen and Gianola, 2002) - notation for variance priors is omited due to lack of space.

$$\begin{aligned} \mathbf{y} &= \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \\ \mathbf{y}|\mathbf{b},\mathbf{a},\sigma_e^2 &\sim N\left(\mathbf{Xb} + \mathbf{Za}, \mathbf{I}\sigma_e^2\right) \\ \mathbf{b}|\sigma_b^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_b^2\right) \\ \mathbf{a}|\mathbf{A},\sigma_a^2 &\sim N\left(\mathbf{0}, \mathbf{A}\sigma_a^2\right) \end{aligned} \tag{1}$$

---

[*] University of Ljubljana, Biotechnical Faculty, Department of Animal Science, Groblje 3, 1230 Domžale, Slovenia

Table 1: Example data

Id	Father	Mother	Group	Phenotype
1	/	/	/	/
2	/	/	1	103, 106
3	2	1	1	98
4	2	/	2	101
5	4	3	2	106
6	2	3	2	93
7	5	6	/	/
8	5	6	/	/
9	/	/	/	/
10	8	9	1	109

As mentioned, the prime way to fit model in BUGS is to represent it as a graphical model using the concept of Directed Acyclic Graphs (DAG). Using graphs to represent models can be traced back at least to Wright. With DAGs variables are represented as nodes connected with directed arcs based on postulated causal relationships. Nodes that are not connected are assumed conditionally independent. DAG for the example data (Table 1) is shown in Figure 1.

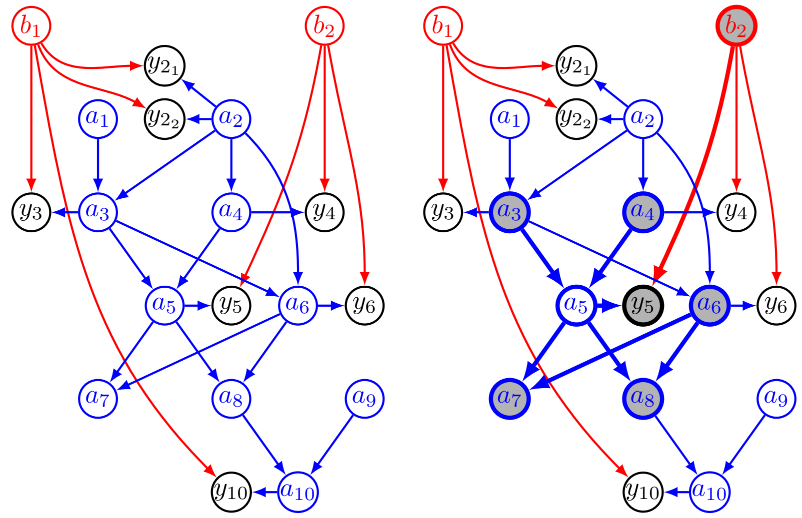


Figure 1: Directed Acyclic Graph representation of animal model (1) for the example data set in Table 1 - variances with associated arcs and priors are omitted to avoid clutter (left) and demonstration of Markov blanket (3) for node a_5 (right)

The following two algebraic expressions are important for DAGs (e.g. Lauritzen, 1996):

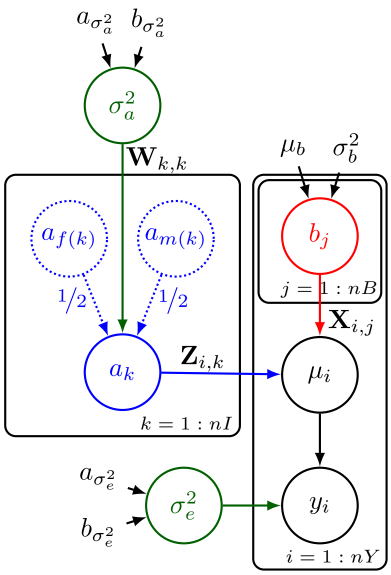
$$p\left(\mathbf{x}\right) = \prod_{x_i \in \mathbf{x}} p\left(x_i | x_{parents(i)}\right), \tag{2}$$

$$p\left(x_i|\mathbf{x}_{-i}\right) \quad \propto \quad p\left(x_i|x_{parents(i)}\right) \prod_{x_j \in x_{children(i)}} p\left(x_j|x_{parents(j)}\right), \tag{3}$$

where (2) describes the joint probability distribution over a graph (model) of nodes (variables) $\mathbf{x}$, while (3) provides a way to automatically construct full conditional distributions needed for McMC procedure. For example, the full conditional distribution of a_5 involves nodes: a_3 and a_4 (graphical parents - $x_{parents(i)}$), a_7, a_8, and y_5 (graphical children - $x_{children(i)}$), and a_6 and b_2 (the corresponding graphical mates) (Figure 1 - right). Use of (3) leads to the equivalent set of full conditionals as the "standard" treatment (Sorensen and Gianola, 2002).

BUGS implementation

DAG representation in Figure 1 is intuitive but notorious for large models. A compact and generic representation using the plate notation (Buntine, 1994) is shown in Figure 2 (left). With plate notation a complete set of model "properties" can be represented easily in a generic way, e.g., priors, dotted line indicates possibly missing parent, etc. In addition, almost direct translation to BUGS model language is possible that can be used with collected data of any structure (pattern) (Figure 2, right). An important peculiarity of BUGS model language is that normal distribution is reparameterized in terms of mean and precision. The later is the reciprocal of variance, which provides a way to use the inverse of numerator relationship matrix directly through the DAG structure $\left(\text{equivalent to } \mathbf{T}^{-1}\right)$ and diagonal values of $\mathbf{W}^{-1}$ matrix, where $\mathbf{A}^{-1} = \left(\mathbf{T}^{-1}\right)^T \mathbf{W}^{-1}\mathbf{T}^{-1}$ (Henderson, 1976; Quaas, 1976).



```
## Precision (= 1 / Variance) priors
tau2e ~ dgamma(0.001, 0.001)
sigma2e <- 1 / tau2e
tau2e ~ dgamma(0.001, 0.001)
sigma2a <- 1 / tau2a

## Additive genetic values
for(k in 1:nI) {
  a[id[k]] ~ dnorm(pa[id[k]], Xtau2a[id[k]])
  pa[id[k]] <- 0.5 * (a[fid[k]] + a[mid[k]])
  Xtau2a[id[k]] <- winv[id[k]] * tau2a
}
a[nU] <- 0 # NULL (zero) holder

## Location priors
for(j in 1:nB) { b[j] ~ dnorm(0, 1.0E-6) }

## Phenotypic values
for(i in 1:nY) {
  y[i] ~ dnorm(mu[i], tau2e)
  mu[i] <- b[x[i]] + a[idy[i]]
}
```

Figure 2: Generic Directed Acyclic Graph representation of animal model (1) (left) and its description using BUGS model language (right)

Prepared data for BUGS for the example are shown in Table 2. Unknown parent(s) are replaced with one dummy individual, whose additive genetic value is set to *a priori* value - zero (Figure 2, right). In addition, diagonal values of $\mathbf{W}^{-1}$ need to be precomputed (Quaas, 1976), while the rest of the data is prepared as for standard models in BUGS (Lunn et al., 2000).

Table 2: Prepared pedigree and phenotype data for animal model

Pedigree “plate”										
`id`	`1`	`2`	`3`	`4`	`5`	`6`	`7`	`8`	`9`	`10`
`fid`	`11`	`11`	`2`	`2`	`4`	`2`	`5`	`5`	`11`	`8`
`mid`	`11`	`11`	`1`	`11`	`3`	`3`	`6`	`6`	`11`	`9`
`winv`	`1`	`1`	`2`	`1.3`	`2`	`2`	`2.5`	`2.5`	`1`	`2.3`
`nI`	`10`									
`nU`	`11`									
Phenotype “plate”										
`idy`	`2`	`2`	`3`	`4`	`5`	`6`	`10`			
`y`	`103`	`106`	`98`	`101`	`106`	`93`	`109`			
`x`	`1`	`1`	`1`	`2`	`2`	`2`	`1`			
`nY`	`7`									
`nB`	`2`									

BUGS model description in Figure 2 assumes gamma prior for precisions, but this can be easily changed, for example to uniform prior for variances or even standard deviations. The same flexibility is also possible for phenotype sampling model so that threshold, Poisson, Weibull, and other models can be accommodated. In addition, BUGS automatically chooses the appropriate variant of algorithm for sampling from full conditional distributions, e.g., Gibbs, Metropolis-Hastings, slice sampler, etc. (Lunn et al., 2000).

Conclusion

Animal model was represented as a generic Directed Acyclic Graph (DAG), which can be easily translated to BUGS model language providing a flexible environment for inferring animal model parameters with McMC methods.

References

Buntine, W. L. (1994). *J. Artif. Intell. Res.*, 2(1):159–225.

Damgaard, L. H. (2007). *J. Anim. Sci.*, 85(6):1363–1368.

Henderson, C. R. (1976). *Biometrics*, 32(1):69–83.

Lauritzen, S. L. (1996). Oxford University Press.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). *Stat. Comput.*, 10(4):325–337.

Quaas, R. L. (1976). *Biometrics*, 32(4):949–953.

Sorensen, D. A. and Gianola, D. (2002). Springer.

Waldmann, P. (2009). *Evolution*, 63(6):1640–1643.