Bayesian Mixed Models in Ecology

Iordan Ganev

1 Introduction

In many ecological studies, researchers construct a theoretical model in order to understand relationships and interactions between several natural phenomena. Ideal models explain the variability of the data and generate precise predictions. Among the many types of models that appear in ecological literature, a *mixed model* is extensively used because it groups experimental units into "blocks" under the presumption that units within the same block respond more consistently than units from different blocks. In this way, the model accounts for certain variability among the units. The blocks that appear in ecology are often different populations of the same species, of groups of organisms experiencing different treatments. Some blocking factors – so-called *fixed effects* – may be controlled by the researcher and are usually repeatable. *Random effects*, on the other hand, are blocking factors whose levels are not repeatable or are considered a random sample from an essentially infinite set of possibilities [15].

For example, in a study of whether invasive plants are larger than native ones, seeds from each of two plant species were collected from several different populations around the world and compared [6]. The experimental unit is each seed and "population" is a blocking factor. Each seed experiences a fixed effect based on whether it inhabits a native or exotic habitat, as well as a random effect of population, since not all native or exotic populations are alike. We imagine that there is a distribution of possible levels for the effect of population and the level of this effect for each population in the study is a random draw from this distribution. In fact, several seeds were collected from the same plant, so "individual plant" can be treated as a random effect nested within population.

Researchers ordinarily wish to estimate the fixed effects of mixed models, not the random effects. However, estimation of fixed effects relies directly on estimation of the variance of the random effects, also known as the variance component [8]. What is the best statistical approach to estimate variance components? Some studies have compared classical and Bayesian methods of variance component estimation in mixed models [16], but the simulated data sets are limited and not representative of realistic biological data. In particular, they do not address cases with few replicates per block. Such data sets are common in ecology [3, 1] and a comparison of classical and Bayesian inference will indicate the best way to estimate variance components.

This honors thesis conducts such a comparison. In addition, it gives a background to the theory behind mixed models (section 2) and demonstrates the formulation of a mixed model for a real ecological study (section 3). Examples of other uses of mixed models in ecological research are given in section 4. An overview of Bayesian inference in section 5 prepares for the simulation study and comparison in section 6.

During my year-long independent study and summer research mentored by Dr. Doug Noe (Statistics) and Dr. Hank Stevens (Botany), I have explored how mixed models and Bayesian inference are used in recent ecological research. Books on mathematical statistics, scientific papers, simulations using statistical software, and discussions with my advisers have all contributed to my appreciation for the development of these tools and their impact in ecology. Much of what I have learned is described here. This project has allowed me to achieve numerous goals, including

- to gain first-hand experience of interdisciplinary research as an undergraduate

- to understand the formulation and use of mixed models

- to grasp the distinctions between Bayesian and classical methods of estimation

- to develop an intermediate level of fluency in statistical software such as R and BUGS [20, 13]

- to read scientific papers about the application of these tools

- to contribute to the understanding of mixed models by resolving an important question that permeates both statistics and ecology

This project has been a valuable and successful journey of intellectual discovery and I hope to turn this honors thesis into a research paper for publication in a journal related to ecology or evolution.

2   Mixed Models

Mixed models (also known as hierarchical or multilevel models) are linear models whose parameters may be constants or random variables [8]. The mixed model can be notated as

$$Y = X\beta + Z\gamma + \epsilon \tag{1}$$

where X and Z are known matrices of covariates, β is a vector of fixed effects, and γ is a vector of random effects. In addition, $\gamma \sim N(0, D)$, $\epsilon \sim N(0, R)$, $\text{Cov}(\gamma, \epsilon) = 0$. Note that

$$V = \text{Cov}(Y) = \text{Cov}(Z\gamma + \epsilon) = ZDZ' + R.$$

Estimation of the fixed effects (i.e., the parameters in β) depends on knowledge of the variance component V since

$$X\hat{\beta} = X[X'V^{-1}X]^- X'V^{-1}Y.$$

The variance-covariance matrix for the random effects, D, is often assumed to be diagonal with each entry on the diagonal measuring the among-group variability of the group from which the observation was taken. In general, there may be r different groups, that is, r different levels of the random effect. Let $q(i)$ and σ_i^2 denote, respectively, the number of observations in and the variance of the ith group. Then D is a block diagonal matrix, denoted

$$D = \text{Blk diag}[\sigma_i^2 I_{q(i)}], \qquad i = 1, 2, ..., r.$$

Similarly, we assume $R = \sigma_0^2 I$ [8]. Each σ_i^2 can be thought of as a measure of among-group variability while σ_0^2 reflects within-group variance [17].

Christensen [8] gives several approaches to estimating the variance component V. The maximum likelihood estimates of β and σ_i^2 are solutions to

$$X'V^{-1}X\beta = X'V^{-1}Y$$

$$tr(V^{-1}Z_iZ_i') = (Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}(Y - X\beta).$$

where Z_i consists of the columns of Z that correspond to the ith group. These equations are derived directly from the log likelihood for β and V given Y. In most cases, iterative computation is required to find solutions of these equations. The method of restricted maximum likelihood (REML) relies on the distribution of residuals to estimate variance components. However, since V is unknown, the estimates must be derived through an iterative procedure. First an initial guess for V is taken. This guess is updated by solutions for σ_i^2 from the equations

$$\sum_{j=0}^{r} \sigma_j^2 tr[Z_jZ_j'V^{-1}(I - A_0)Z_iZ_i'V^{-1}(I - A_0)] = Y'(I - A_0)'V^{-1}Z_iZ_i'V^{-1}(I - A_0)Y \quad (2)$$

for $i = 0, ..., r$, where $A_0 = X(X'V^{-1}X)^-X'V^{-1}$. The updated guess is substituted back into the equations and a new estimate is found. These steps are repeated until each σ_i^2 converges.

Minimum norm quadratic unbiased (translation invariant) estimation (MINQUE) also relies on an iterative procedure and gives the same estimates as REML. However, MINQUE allows the user to adopt weights w_i^2 for $i = 0, 1, ..., r$ and a good choice of initial weights saves many iterations. Given these weights, define

$$V_w = \sum_{i=0}^{r} w_i Z_i Z_i' \quad \text{and} \quad A_w = X(X'V_w^{-1}X)^-X'V_w^{-1}.$$

The equations used in MINQUE iterations are

$$\sum_{j=0}^{r} \sigma_j^2 tr[Z_jZ_j'V_w^{-1}(I - A_w)Z_iZ_i'V_w^{-1}(I - A_w)] = Y'(I - A_w)'V_w^{-1}Z_iZ_i'V_w^{-1}(I - A_w)Y.$$

In these equations, V_w is known, whereas V is unknown in the REML equations (2). Also, the choice of weights in MINQUE may require fewer iterations to obtain satisfactory estimates of the σ_i^2's. Common selections include $w_i = 1$ for all i or $w_0 = 1$ and $w_i = 0$ for all $i \neq 0$. In the latter case, $V_w = I$ and A_w is the perpendicular projection matrix onto the column space of X; the estimates obtained are frequently referred as MINQUE(0) estimates.

The final method Christensen describes is Henderson's Method 3. Let P be the perpendicular projection operator onto the column space of the matrices X and Z, that is $C(X, Z_1, Z_2, ..., Z_r)$. Then an unbiased estimate of σ_0^2 is

$$\hat{\sigma}_0^2 = \frac{Y'(I - P)Y}{tr(I - P)}.$$

To estimate, say, σ_r^2, let P_0 be the perpendicular projection operator onto $C(Z, Z_1, Z_2, ..., Z_{r-1})$. Then

$$\hat{\sigma}_r^2 = \frac{Y'(P - P_0)Y - \hat{\sigma}_0^2 tr(P - P_0)}{tr[(P - P_0)Z_rZ_r']}.$$

In general, an estimate for σ_i^2 is found the same way, but with P_0 defined to be the perpendicular projection matrix onto the column space of the X and all the Z_j's except Z_i.

3

3 Lady Beetle Example

Schaalje *et al.* [18] claim that the number of sensillae (Y) on a particular antenna segment of a lady beetle depends on the sex of the beetle and the segment number. This example will be used to demonstrate the formulation of a mixed model. The data was not available, so no variance component estimation has been conducted.

3.1 Fixed Effects

First, suppose μ is the mean number of sensillae ($E\{Y\}$) on the tenth segment of a male lady beetle. Although this choice of sex and segment is somewhat arbitrary, it saves notation and elucidates the model's construction. For the males, segment ten, we may write

$$Y_k = \mu + \epsilon_k. \tag{3}$$

The index k for individuals is used for future convenience. We may assume $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Let α_i denote the fixed effect of sex, with $i = 1$ corresponding to females and $i = 2$ corresponding to males. In order to adapt this parameter into (3), α_2 is defined to be 0, while α_1 is the difference between the mean number of sensillae on segment 10 of females and the mean number of sensillae on segment 10 of males; that is, $\alpha_1 = E\{Y_1\} - E\{Y_2\}$ (segment 10). The experiment includes six individuals, 3 females and 3 males. Since sex in nested in individual, let $i(k) = 1$ for the females $k = 1, 2, 3$ and $i(k) = 2$ for the males $k = 4, 5, 6$. Considering only the influence of sex, our model becomes

$$Y_k = \mu + \alpha_{i(k)} + \epsilon_k \qquad k = 1, 2, 3, 4, 5. \tag{4}$$

Now we ignore sex and focus on the fixed effect of segment number. Define β_j to be difference between the mean number of sensillae on segment j of a lady beetle and the mean number of sensillae on segment 10 of a lady beetle. In symbols, $\beta_j = E\{Y_j\} - E\{Y_{10}\}$ for $j = 1, 2, ..., 10$. Note that $\beta_{10} = 0$. The influence of segment number is given by:

$$Y_{jk} = \mu + \beta_j + \epsilon_k \qquad j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. \tag{5}$$

In order to combine the two fixed effects, we may simply add (4) and (5):

$$Y_{jk} = \mu + \alpha_{i(k)} + \beta_j + \epsilon_k.$$

However, we have assumed that the two effects are independent – a justified claim in some cases. If it were true, the difference in $E\{Y\}$ for, say, segments 5 and 8 on a particular lady beetle is the same if the beetle is males or if it is female. In this case, we may not assume independence and assign new set of parameters to account for an interaction between these variables. Let γ_{jk} be the difference of the segment j effect for sex i and the segment j effect for sex 2 (males). That is,

$$\gamma_{ij} = E\{Y_{ij}\} - E\{Y_{i,10}\} - (E\{Y_{2,j}\} - E\{Y_{2,10}\})$$

Equivalently, γ_{ij} represents the difference of sex i effect for segment j and and the sex effect for segment 10. Note that $\gamma_{2,j} = 0$ for all 10 choices of j and $\gamma_{i,10} = 0$ for the two choices of i. Consequently, only 9 of the 20 γ terms are nonzero. The more general combined fixed effects model becomes

$$Y_{jk} = \mu + \alpha_{i(k)} + \beta_j + \gamma_{i(k),j} + \epsilon_k.$$

4

3.2 Within Group Variability

Perhaps there are differences between the variability of sensillae count in males and that in females. We can accommodate for these differences by introducing a random effect term for sex at the individual level. Define $\sigma_{s_1}^2$ and $\sigma_{s_2}^2$ as the variance of sensillae count in females and males, respectively. For $k = 1, ...6$ let $S_k \sim N(0, \sigma_{i(k)})$ represent how much the different individuals are influenced by sex. The S_k's are assumed to be independent from each other and from ϵ_k. Now we have:

$$Y_{jk} = \mu + \alpha_{i(k)} + \beta_j + \gamma_{i(k),j} + S_{i(k)} + \epsilon_k.$$

Differences in the variability of sensillae count among the 10 antenna segments are represented similarly. Let $\sigma_{b_j}^2$ be the variance of sensillae count in segment j. Now, let $B_{j,k} \sim N(0, \sigma_{b_j}^2)$. The $B_{j,k}$ are assumed to be independent from each other and from different the S_k's and ϵ_k's. Thus the model becomes:

$$Y_{jk} = \mu + \alpha_{i(k)} + \beta_j + \gamma_{i(k),j} + S_{i(k)} + B_{j,k} + \epsilon_k.$$

There are 32 parameters in this model: α_1, β_1 through β_9, 9 γ's, 12 random effect variances, and σ_ϵ^2.

3.3 Vector and Matrix Notation

Matrix notation often simplifies a model's notation and the inferences associated with parameters. In terms of equation (1),

$$\beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \vdots \\ \beta_1 0 \\ \gamma_{1,1} \\ \vdots \\ \gamma_{2,10} \end{bmatrix}, \qquad \gamma = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_6 \\ B_{1,1} \\ \vdots \\ B_{10,6} \end{bmatrix},$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix} = [J, I_2 \otimes J_{30}, J_6 \otimes I_{10}, J_6 \otimes I_{10}], \text{ and}$$

$$Z = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix} = [I_6 \otimes J_{10}, I_6 \otimes I_{10}]$$

5

Recall that we have assumed that the effect of sex, segment, and individual are independent. Therefore, $R = \text{Cov}(\epsilon) = \sigma_\epsilon^2 I_{60}$, and

$$
D = \text{Cov}(\gamma) = \left[\begin{array}{cccccc|cccc}
\sigma_{s_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma_{s_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_{s_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma_{s_2}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_{s_2}^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_{s_2}^2 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \sigma_{b_1}^2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{b_1}^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{b_{10}}^2
\end{array}\right].
$$

Thus, with combinations listed from female segment 1 through male segment 10,

$$
V = \text{Cov}(Y) = ZDZ' + R
$$

$$
= \left[\begin{array}{ccc}
\sigma_{s_1}^2 + \sigma_{b_1}^2 + \sigma_\epsilon^2 & \sigma_{s_1}^2 & \cdots \\
\sigma_{s_1}^2 & \sigma_{s_1}^2 + \sigma_{b_2}^2 + \sigma_\epsilon^2 & \cdots \\
\vdots & \vdots & \ddots \\
\sigma_{s_1}^2 & \sigma_{s_1}^2 & \cdots \\
\sigma_{b_1}^2 & 0 & \cdots \\
0 & \sigma_{b_2}^2 & \cdots \\
\vdots & \vdots & \ddots \\
0 & 0 & \cdots
\end{array}\right.
$$

$$
\left.\begin{array}{cccccc}
\cdots & \sigma_{s_1}^2 & \sigma_{b_1}^2 & 0 & \cdots & 0 \\
\cdots & \sigma_{s_1}^2 & 0 & \sigma_{b_2}^2 & \cdots & 0 \\
\ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\cdots & \sigma_{s_1}^2 + \sigma_{b_{10}}^2 + \sigma_\epsilon^2 & 0 & 0 & \cdots & \sigma_{b_{10}}^2 \\
\cdots & 0 & \sigma_{s_2}^2 + \sigma_{b_1}^2 + \sigma_\epsilon^2 & \sigma_{s_2}^2 & \cdots & 0 \\
\cdots & 0 & \sigma_{s_2}^2 & \sigma_{s_2}^2 + \sigma_{b_2}^2 + \sigma_\epsilon^2 & \cdots & 0 \\
\cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\cdots & \sigma_{b_{10}}^2 & \sigma_{s_2}^2 & \sigma_{s_2}^2 & \cdots & \sigma_{s_2}^2 + \sigma_{b_{10}}^2 + \sigma_\epsilon^2
\end{array}\right]
$$

$$
= \left[\begin{array}{cc}
\sigma_{s_1}^2 & 0 \\
0 & \sigma_{s_2}^2
\end{array}\right] \otimes J_{10}^{10} + J_2^2 \otimes \left[\begin{array}{cccc}
\sigma_{b_1}^2 & 0 & \cdots & 0 \\
0 & \sigma_{b_2}^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_{b_{10}}^2
\end{array}\right] + \sigma_\epsilon^2 I_{60}.
$$

Another way V can be derived is as follows. Any pair of observations of Y will fall into one of four cases and our assumption if independence among sex and segments is used to calculate the covariance between the observations in all four cases:

6

	Covariance
1. Different sex, different segment	0
2. Same sex (i), different segment	$\sigma^2_{s_i}$
3. Same segment (j), different sex	$\sigma^2_{b_j}$
4. Same sex (i), same segment (j)	$\sigma^2_{s_i} + \sigma^2_{b_j} + \sigma^2_{\epsilon}$

4 Examples of Mixed Models in Ecological Literature

Ecology research abounds with examples of mixed models. In a literature search, I became familiar with the specifics of several studies that considered mixed models. I was also interested in what methods were used to estimate parameters in the models and whether there were potential problems with the data, e.g., skewed distributions, few replicates per block, few blocks.

Note: not all examples will be included in the final version

4.1 Population Studies

With data from North American Breeding Bird Survey, Link and Sauer have used mixed models in order to estimate the population size of bird species [12]. Fixed effects are stratum and novice observer effects, while the influences of year, route, and overdispersion are considered random effects. The authors employ a Bayesian approach (using BUGS) with uniformed prior distributions for "hyperparamters".

In the text, the model is formulated as follows. $Y_{i,j,t}$ denotes the counts of a bird species in stratum i year t. The index j indicates a unique combination of route and observer. The authors use λ to represent the expected value of Y, i.e., $E(Y_{i,j,t}) = \lambda_{i,j,t}$. The fixed effects parameters are

- stratum effect $= S_i$

- year effect $= \beta_i$

- first-year observer effect $= \eta$.

All have uninformed $N(\mu = 0, \sigma = 1000)$ distributions. The first year observer effect is multiplied by an indicator variable $\mathbf{I}(j,t)$ which is 1 for the observer's first year and 0 otherwise. With the random effects, the parameters themselves are not as crucial as their variances:

- route effect $= \omega_j \sim N(0, \sigma^2_\omega)$

- overdispersion effect $= \epsilon_{i,j,t} \sim N(0, \sigma^2_\epsilon)$

- stratum-year interaction $= \gamma_{i,t} \sim N(0, \sigma^2_{\gamma,i})$ (a different variance for each strata).

The distribution of $\sigma_{\gamma,i}^2$ is an inverse gamma with mean 1 and variance 1000. The model is stated in the paper as:

$$\log(\lambda_{i,j,t}) = S_i + \beta_i(t - t^*) + \omega_j + \eta\mathbf{I}(j,t) + \gamma_{i,t} + \epsilon_{i,j,t}.$$

where t^* is a baseline year. This is a generalized linear model, hence the logarithmic component and the overdispersion effect. These data may have zero-inflation and non-normal data.

There are numerous models for population dynamics, and Barrowman *et al.* [4] estimated parameters of four such models using nonlinear mixed-effects models with two random effects: log slope at the origin — $\log(\alpha)$ — and log asymptotic level — $\log(2K)$. Although I was not able to formulate the mixed-model, this paper reveals that mixed models can be used within more complex models. The mixed-model results were compared with Bayesian analyses for robustness. In the Bayesian analyses, the priors distributions for the variance components were always approximately uninformed (gamma with shape = .0001 and scale = 1000). Markov chain Monte Carlo and BUGS were used. Bayesian results matched well with mixed-models estimate. Fully Bayes predictive distributions were slightly wider than the random-effects distribution estimate given by the mixed-models approach.

4.2 Ecosystem Dynamics

A mixed model can be used to represent the effect of species richness, disturbance, and the species richness-disturbance interaction on three ecological processes in streams: biofilm respiration (BR), gross primary production (GPP), and particulate organic material flux ($POMF$) [7]. Individual stream mesocosms constituted blocking factors and mixed model ANOVAs were used in the model. Schwarz's Bayesian criteria specified co-variance structures.

We may employ, for example, the following notation in the model: For $p = BR$, GPP, $POMF$, let Y_p be the response variable that corresponds to process p. For example, when $p = BR$, Y_{BR} represents concentration of O_2. Let $\alpha_{p,i}$ denote the fixed effect of species richness on process p. Then the values of α_i when $i = 1, 2$, or 3 correspond to the effects when there are 1, 2, or 3 caddisfly species present in the mesocosm, respectively. For example, $\alpha_{GPP,1}$ is the effect of one caddisfly species on GPP (gross primary production). Similarly, let $\beta_{p,j}$ for $j = 1, 2$ denote the fixed effect of no disturbance and + disturbance, respectively, on process p. The species richness-disturbance interaction effect is given by $\gamma_{p,i,j}$. Let $S_{p,k}$ denote the effect of each block on process p for $k = 1, 2, ..., 9$. In each block, there are 2 replicates for each combination of species richness and disturbance and these replicates are represented by $l = 1, 2$. Then we have the three separate models combined into one:

$$Y_{p,i,j,k,l} = \alpha_{p,i} + \beta_{p,j} + \gamma_{p,i,j} + S_{p,k} + \epsilon_{p,i,j,k,l}.$$

4.3 Plant Development

Mixed-effects models have been used to measure how the environment affects the development of plants. In one study, researchers estimated the fixed effects of shade, rain, bare ground, grass cover, density, year, other species, and damage on the vegetative growth,

probability of flowering, growth of flowering stems, fruit production, sucker production, and survival change of *H. perforatum*[6]. Random effects emerged only through "individual experimental units drawn at random from a population." These included the nested spatial variables site - treatment - block - quadrat - plant. This model allowed researchers to determine the correlations within grouped data.

Linear mixed-effects models (LME) was used where random effects vary over a cluster of observations units and nested random effects may exist. For instance, the full model describing the growth of vegetative stems includes "a selection of fixed effects fit as explanatory variables," which are denoted here as $\alpha(1), ..., \alpha(n)$ and indexed $i(1), ..., i(n)$ respectively. That is, if there are 4 levels for the first fixed effect $\alpha(1)$, then $i(1) = 1, 2, 3, 4$. In addition, the model features "the full range of nested spatial random effects" (site / treatment / block / quadrat / plant) denoted here as $\beta(1), ..., \beta(5)$ and indexed $j(1), ..., j(5)$. Since these are nested, the level of any β depends the levels of all the higher random effects. The nested structure requires that a random effect index to include indices of all higher random effects. Therefore, we may write the

$$
\begin{aligned}
\log(Y_{i(1),...,i(n),j(1),...j(5)}, j) = & \alpha(1)_{i(1)} + \alpha(2)_{i(2)} + ... + \alpha(n)_{i(n)} \\
& + \beta(1)_{j(1)} + \beta(2)_{j(1),j(2)} + \beta(3)_{j(1),j(2),j(3)} \\
& + \beta(4)_{j(1),j(2),j(3),j(4)} + \beta(5)_{j(1),j(2),j(3),j(4),j(5)} \\
& + \epsilon_{i(1),...,i(n),j(1),...,j(5),k}.
\end{aligned}
$$

A similar model describes the growth of flowering stem. An F-test of likelihood ratios for linear mixed-effects models compared the full model to depleted models that exclude explanatory variables. Maximum likelihood and restricted maximum likelihood methods were also used to compare such models and obtain estimated coefficients. Akaike's information criterion (AIC) determined the choice of the best model.

Binary variables were interpreted with generalized linear mixed models (GLMM), in which random effects were nested at only one level. Variables of interest here included the probability of flowering and the probability of survival. For other analyses, different models such as generalized linear models (GLM) were used.

Human activity has increased the number of invasive species in the modern world, and it has been hypothesized that invasive plants produce larger seeds than native ones of the same species [5]. For each of the two species, *C. scoparius* and *U. europeus* , seeds collected from several sites around the world followed the nested spatial random effect structure country / site / population / individual. The species were native to some of the countries, but exotic in others; this contrast was a fixed effect. The study included eleven countries, 20 sites total (various number of sites per country), approximately 15 plants within a site, and 30 seeds per individual plant. The weights of theses seeds (Y), was fitted to the model:

$$
Y_{i,j,k,l,m,n} = \alpha_i + C_j + S_{j,k} + P_{j,k,l} + V_{j,k,l,m} + \epsilon_{i,j,k,m,n}
$$

where α_1 = native fixed effect, α_2 = exotic fixed effect; and C_j , $S_{j,k}$, $P_{j,k,l}$, and $V_{j,k,l,m}$ are the nested country, site, population, and individual random effects, respectively.

In the analysis, multilevel LME from Pinheiro and Bates [17] was used. "The significance of each scale in the spatial hierarchy was determined by comparing models with

each scale individually deleted while preserving the nested structure." The authors note that several populations and sites had only no replicates and thus no within-block variance is stated.

4.4 Tracing Evolution

One way evolutionary biologists measure rates of a population's genetic evolution is through the parameter Q_{ST}, which is defined as the proportion of total genetic variance in the individuals of a species that is attributed to the population level [11]. In notation,

$$Q_{ST} = \frac{\sigma_B^2}{\sigma_B^2 + 2\sigma_W^2}$$

where σ_B^2 is between-population variance and σ_W^2 is within-population variance. In mixed-models terminology, σ_B^2 is the variance of the random effect of the population in question, while σ_W^2 is the variance of the error term. The genes considered in Q_{ST} control quantitative traits and are presumed to be influenced by natural selection. By contrast, another parameter F_{ST}, which is not defined in this paper, gives a similar measure of genetic differentiation, but at "neutral marker loci" – those not influenced much by environmental selection pressures. To determine rates of genetic evolution, Q_{ST} is compared to F_{ST}. $Q_{ST} > F_{ST}$ indicates divergent or disruptive selection, while $Q_{ST} < F_{ST}$ indicates stabilizing selection.

The model of interest in Hall *et al.* is stated as

$$z_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \epsilon_{ijkl}$$

"where z_{ijkl} is the phenotype of the lth individual in the kth block from the jth clone from the ith population.... μ denotes the grand mean and ϵ_{ijkl} is the residual error term." Random effects emerge from the population (α_i), the clone (β_j), and the block (γ_k). The between-population variance σ_B^2 is the variance of the population effect and is estimated through the α's. The within-population variance σ_W^2 is the variance of the clone effect and is estimated through the β's. Q_{ST} is defined to be $\frac{\sigma_B^2}{\sigma_W^2 + \sigma_B^2}$. A Bayesian approach was used with noninformative priors to find 95% confidence intervals for Q_{ST}.

Banta *et al.* [2] were also concerned with differences between F_{ST} and Q_{ST} as evidence for evolution. The researchers used method of moments approach in analyzing mixed-model ANOVA. The 7 response variables were bolting time (Y_1), number of rosette leaves (Y_2), rosette diameter (Y_3), height (Y_4), number of branches (Y_5), time to ripening (Y_6), and number of fruits (Y_7). For any Y_p for $p = 1, ..., 7$ the fixed effects are:

- region; $\alpha(1)_{p,i(1)}$, $i(1) = 1, ..., 5$,

- growth chamber; $\beta(1)_{p,j(1)}$, $j(1) = 1, 2, 3$,

- shelf within growth chamber; $\beta(2)_{p,j(1),j(2)}$, $j(2) = 1, 2$,

- tray within a shelf; $\beta(3)_{p,j(1),j(2),j(3)}$, $j(3) = 1, ..., 25$.

Random effects:

- population (nested in region); $\alpha(2)_{p,i(1),i(2)}$, $i(2) = 1, ..., 16$,

- family (nested in population); $\alpha(3)_{p,i(1),i(2),i(3)}$, $i(3) = 1, ..., 41$.

Interaction effects:

- region-by-growth chamber; $\gamma_{p,i(1),j(1)}$,

- population-by-growth chamber; $\delta_{p,i(2),j(1)}$,

- family-by-growth chamber; $\zeta_{p,i(3),j(1)}$.

Supposedly, then, that the full model takes the form:

$$
\begin{aligned}
Y_{p,i(1),i(2),i(3),j(1),j(2),j(3)} =& \alpha(1)_{p,i(1)} + \alpha(2)_{p,i(1),i(2)} + \alpha(3)_{p,i(1),i(2),i(3)} \\
&+ \beta(1)_{p,j(1)} + \beta(2)_{p,j(1),j(2)} + \beta(3)_{p,j(1),j(2),j(3)} \\
&+ \gamma_{p,i(1),j(1)} + \delta_{p,i(2),j(1)} + \zeta_{p,i(3),j(1)} + \epsilon_{p,i(1),i(2),i(3),j(1),j(2),j(3)}.
\end{aligned}
$$

In the analysis, the authors jackknifed the statistic $Q_{ST} - F_{ST}$ at the population level and then performed a one-sided t-test.

4.5 Changes in Body Mass

Bårdsen *et al.* [3] investigated how differences in winter food supply affect reindeer body mass through three different comparisons. For several years before the experiment, reindeer herd #1 had received supplemental food during the winter (February - May) while reindeer herd #2 depended entirely on natural pastures during the winter. In January 2003 and again in January 2004, twenty females from each herd were moved to the other herd. These were treatment units identified by their herd of origin. At the same time, twenty other females per herd served as controls and remained in their herd of origin. All 80 females were weighted in January, July, and September 2003 and 2004.

A similar model for female body mass (Y) is applied to all three comparisons:

$$
Y_{i,j,k,l} = \alpha_i + \beta_j + \gamma_{i,j,k} + \delta_{j,l} + S_{i,k.} + \epsilon_{i,j,k,l}.
$$

The effects common to the three models are: year, β_j, $j = 2003, 2004$ (fixed); initial female body mass, $\gamma_{i,j,k}$, $k = 1, ..., 20$ (fixed); season, $\delta_{j,l}$, $l = 1$ (January), 2 (September), and 3 (July) (fixed); individual, $S_{i,k}$ (random); and random error $\epsilon_{i,j,k,l}$. Although each of these effects appears in the three models and is represented by the same symbol and index, its level varies across model. The crucial effect in each comparison is given by α_2.

Comparison A: control from herd #1 vs. control from herd #2; $\alpha_1 = 0$ is the fixed herd effect for herd #1, α_2 is that for herd #2.

Comparison B: control from herd #1 vs. treatment from herd #2 (improved winter conditions); $\alpha_1 = 0$ is the fixed herd effect for herd #1, α_2 is that for herd #2.

Comparison C: control from herd #2 vs. treatment from herd #1 (reduced winter conditions); $\alpha_1 = 0$ is the fixed herd effect for herd #2, α_2 is that for herd #1.

It seems that a potential problem in this data set is there are no replicates for herd - year - season - female combination (hence there is no need for a separate index for random error) and only three measurements per block, i.e. individual female. This is because females were measured three times a year and potentially different females were chosen each year.

While Bårdsen *et al.* were interested in body mass changes during the winter, Alonso-Alvarez *et al.* traced body mass changes associated with reproduction [1]. In the latter experiment, 46 randomly formed pairs of zebra finches were kept each in a separate cage. "Thirty-six pairs bred and their brood size was manipulated as to create two groups: (i) pairs raising a two-chicks brood ($n = 20$); (ii) pairs raising a six-chicks brood ($n = 16$)." Ten pairs did not breed. A model for change in body mass during the period of study (Y) is:

$$Y_{i,j,k} = \alpha_i + \beta_j + S_k + \epsilon_{i,j,k}$$

where α_i is the fixed effect of number of offspring ($i = 1$ for no chicks, $i = 2$ for two chicks, and $i = 3$ for six chicks), β_j is the fixed effect of sex ($j = 1$ for females and $j = 2$ for males), and S_k the random effect of cage identity for $k = 1, ..., 46$.

As with [3], there are few measurements within blocks: two per cage, in this case. Equivalently, there was only one replicate per block. The experimental structure of this paper was simulated – see section 6.

5   Bayesian Inference

The philosophical differences between classical and Bayesian methods of inference stem in part from conflicting notions of probability [9]. How does one define *probability*? Perhaps it is a measure of the relative frequency of certain events and can be determined with greater accuracy as more events are observed. This frequentist view of probability is the basis of classical methods of estimation. On the other hand, Bayesians consider probability as a subjective assignment or expectation. When dealing with parameters in models, a Bayesian view of probability reflects uncertainty about those parameters. One consequence of the subjective notion of probability is that researchers may incorporate previous beliefs in an analysis.

Practical, rather than philosophical, differences between the two methods have motivated the adoption of Bayesian inference by many ecologists [9]. Recent research indicates that Bayesian methods may improve interpretation of data, reduce confidence intervals, and readily accommodate complexity [14, 19]. The major innovation of Bayesian approaches is that parameters are considered to be random variables with probability distributions.

5.1   General approach

Consider a model with a vector of parameters θ and a vector of observations y. The model determines a distribution for y given θ, i.e. $p(y|\theta)$, which is the *likelihood function*. In classical inference, θ is considered fixed, while Bayesians assign a probability distribution $p(\theta)$[1] to the

---

[1]Following the convention of Gelman *et al.* [10], "[d]ifferent distributions in the same equation (or expression) will each be denoted by $p(\cdot)$ .... Although an abuse of standard mathematical notation, this

parameter θ. This so-called *prior* distribution is 'known' and can assigned based on previous information about θ or can be a flat, 'non-informative' distribution [10]. Even though θ has a distribution associated with it, in many cases a simple Bayesian method "shares with the frequentist approach the assumption that the parameter itself is a fixed constant" [9].

Bayes' Rule gives a formula for the *posterior* distribution of θ, that is, the distribution of θ given the observations y:

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

Since y has been observed and $p(y)$ is considered a constant [10], we have

$$p(\theta|y) \propto p(\theta)p(y|\theta). \tag{6}$$

In this way, the posterior distribution combines prior expectations with the model and the data. Although the prior distribution $p(\theta)$ and the likelihood $p(y|\theta)$ may be straightforward distributions, the posterior $p(\theta|y)$ is often difficult to find. Instead, techniques such as Markov chain Monte Carlo – the basis of BUGS [13] – sample repeatedly from the prior and likelihood functions in order to compile a histogram of the posterior density [9]. The posterior density is also used to make predictions about an 'unknown observable', $\tilde{y}$. The *posterior predictive distribution* is the distribution of $\tilde{y}$ given the known observations y:

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y)d\theta \\ &= \int p(\tilde{y}|\theta, y)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \end{aligned} \tag{7}$$

since y and $\tilde{y}$ are assumed to be conditionally independent.

5.2 Example

For instance, suppose we flip a coin n times and observe y heads. We wish to estimate θ, the probability of heads in this binomial experiment. The likelihood function is

$$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}. \tag{8}$$

A convenient and reasonable prior distribution for θ is a Beta(α, β) distribution. A non-informative prior could be Beta(1, 1) which is equivalent to a uniform $[0, 1]$ distribution. Alternatively, we may impose the expectation that $\theta = 1/2$ upon our prior and have it be, for example, Beta(2, 2). In general,

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{9}$$

---

method is compact and similar to the standard practice of using $p(\cdot)$ for the probability of any discrete event"

Hence,

$$p(\theta|y) \propto \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}$$
$$\propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$$
$$\propto \text{Beta}(\alpha+y, \beta+n-y). \tag{10}$$

Our estimate for θ is $E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n}$ and our 95% confidence interval for θ is between the $.025^{th}$ and $.975^{th}$ quantiles of $\text{Beta}(\alpha+y, \beta+n-y)$. To predict $\tilde{y}$, the number of heads in a new set of $\tilde{n}$ tosses, we refer back to (7):

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$
$$\propto \int_0^1 \theta^{\tilde{y}}(1-\theta)^{\tilde{n}-\tilde{y}}\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}d\theta$$
$$= \int_1^0 \theta^{\tilde{y}+\alpha+y-1}(1-\theta)^{\tilde{n}-\tilde{y}+\beta+n-y-1}d\theta.$$

In order to demonstrate the procedure of Markov chain Monte Carlo and BUGS, suppose we didn't recognize the posterior distribution in (10) as a Beta. Suppose further that we wish to compare the relative probabilities for the values $\theta = 0.1$, 0.5, and 0.7. Then, using (8), we would find $p(0.1)$, $p(0.5)$, and $p(0.7)$, and by (9) evaluate $p(y|0.1)$, $p(y|0.5)$, and $p(y|0.7)$. Finally, we multiply the respective values to get $p(0.1|y) \propto p(0.1)p(y|0.1)$ and so forth. It may turn out that $p(0.7|y) > p(0.5|y) > p(0.1|y)$, so we can conclude that it is more likely that $\theta = 0.7$ than 0.1 or 0.5, though the assumption of a fixed θ prevents probabilistic claims. BUGS performs this procedure with a multitude of values and an estimate, as well as approximate 95% confidence interval can be determined from these samples.

5.3   Bayesian Estimation in Linear Models

The following discussion is based on Christensen [8, Section 2.9]. A linear model can be parameterized to take the form

$$Y = Z\gamma + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

where Z is of full rank r. In this case, (6) becomes

$$p(\gamma, \sigma^2|Y) \propto f(Y|\gamma, \sigma^2)p(\gamma|\sigma^2)p(\sigma^2).$$

The likelihood function $f(Y|\gamma, \sigma^2)$ is normal with mean γ and variance σ^2. One convenient and largely uninformative choice for the prior distributions is uniform $p(\gamma|\sigma^2)$ and exponential for $p(\sigma^2)$. Another way to specify the prior distribution for $\gamma|\sigma^2$ is to consider $r = \text{rank}(Z)$ potential observations. Each potential observation corresponds to a vector $\tilde{z}_i$ of covariates and has expected value $\tilde{z}_i'\gamma$. Suppose it is expected that an observation with this particular level of covariates will have response $\tilde{y}_i$. Suppose further that this guess is considered equivalent to $1/\tilde{w}_i$ observations, that is, including this guess is the same as augmenting

our data set with $1/\tilde{w}_i$ identical observations $(\tilde{z}_i'\gamma, \tilde{y}_i)$. These expectations are equivalent to the assumption that $\tilde{z}_i'\gamma \sim N(\tilde{y}_i, \tilde{w}_i\sigma^2)$. As for σ^2's prior distribution, suppose the guess is $\tilde{\sigma}^2$ and has the weight of $\tilde{v}$ observations. The options for specifying a prior are not very well understood, but one possibility is an inverse gamma with parameters $a = \tilde{v}/2$ and $b = \frac{\tilde{v}}{2\tilde{\sigma}^2}$ [8, 12].

Let $\tilde{Y} = (\tilde{y}_1, ..., \tilde{y}_r)'$, $\tilde{Z}_{r \times r} = (\tilde{z}_1, ..., \tilde{z}_r)'$, and $D(\tilde{w})$ be a diagonal matrix with $\tilde{w}_i$ in the ith diagonal position. To summarize the model and assumptions,

$$Y|\gamma, \sigma^2 \sim N(Z\gamma,\ \sigma^2 I)$$
$$\gamma|\sigma^2 \sim N(\tilde{Z}^{-1}\tilde{Y},\ \tilde{Z}^{-1}\sigma^2 D(\tilde{w})\tilde{Z}^{-1'})$$
$$\sigma^2 \sim InvGa(a, b)$$

assuming $\tilde{Z}$ is nonsingular. It can be shown that

$$p(\gamma, \sigma^2|Y) \propto (\sigma^2)^{-(n+r)/2} p(\sigma^2)$$

$$\times \exp\left\{\frac{-1}{2\sigma^2}\left[(\gamma - \hat{\gamma})'(Z'Z + \tilde{Z}'D^{-1}(\tilde{w})\tilde{Z})(\gamma - \hat{\gamma})\right]\right\}$$

$$\times \exp\left\{\frac{-1}{2\sigma^2}\left[(Y - Z\hat{\gamma})'(Y - Z\hat{\gamma}) + (\tilde{Y} - \tilde{Z}\hat{\gamma})'D^{-1}(\tilde{w})(\tilde{Y} - \tilde{Z}\hat{\gamma})\right]\right\}, \quad (11)$$

where

$$\hat{\gamma} = \left(Z'Z + \tilde{Z}'D^{-1}(\tilde{w})\tilde{Z}\right)^{-1}\left[Z'Y + \tilde{Z}'D^{-1}(\tilde{w})\tilde{Y}\right]. \tag{12}$$

More useful than the posterior distribution in this case are the marginal distribution for estimable functions $\lambda'\gamma$, the marginal distribution of σ^2, and the predictive distribution for new observations y_0. In order to arrive at these, we combine the observed and prior information into one generalized least squares model

$$\begin{bmatrix} Y \\ \tilde{Y} \end{bmatrix} = \begin{bmatrix} Z \\ \tilde{Z} \end{bmatrix}\gamma + \begin{bmatrix} e \\ \tilde{e} \end{bmatrix}, \quad \begin{bmatrix} e \\ \tilde{e} \end{bmatrix} \sim N\left(\begin{bmatrix} 0_{n \times 1} \\ 0_{r \times 1} \end{bmatrix}, \sigma^2 \begin{bmatrix} I_n & 0 \\ 0 & D(\tilde{w}) \end{bmatrix}\right). \tag{13}$$

The generalized least squares estimate of γ in this model is precisely (12), or, in terms of Bayesian inference, $\hat{\gamma}$ is the expected value of γ given Y. From the model (13), we may find the mean squared error, using the methods of generalized least squares given in [8, Sections 2.2 & 2.7]. Since we perform a Bayesian analysis, the mean squared error will be denoted $BMSE$ and has degrees of freedom $BdfE$. Consequently, the following distributions hold:

$$\frac{\lambda'\hat{\gamma} - \lambda'\gamma}{\sqrt{BMSE\lambda'\left(Z'Z + \tilde{Z}'D^{-1}(\tilde{w})\tilde{Z}\right)^{-1}\lambda}} \sim t(BdfE)$$

$$\frac{(BdfE)BMSE}{\sigma^2} \sim \chi^2(BdfE)$$

$$\frac{y_0 - z_0'\hat{\gamma}}{\sqrt{BMSE\left[1 + z_0'\left(Z'Z + \tilde{Z}'D^{-1}(\tilde{w})\tilde{Z}\right)^{-1}z_0\right]}} \sim t(BdfE).$$

In the frequentist framework, γ and σ^2 are fixed, while $\hat{\gamma}$ and $BMSE$ are random. The Bayesian view is the opposite: $\hat{\gamma}$ and $BMSE$ are fixed, whereas γ and σ^2 are random. Frequentist confidence intervals for $\lambda'\gamma$ and σ^2, and the prediction intervals for y_0 are equivalent to Bayesian posterior probability intervals or predictive intervals for the respective quantities. The generalized least squares approach in this paragraph is often more practical than finding the joint posterior distribution (11) since fewer than r potential observations $(\tilde{z}_i'\gamma, \tilde{y}_i)$ need to be specified. If there are no priors, than (13) reduces to the usual linear model.

6   Alonso-Alvarez Simulations

Alonso-Alvarez *et al.* [1] have conducted an experiment that generated data with two replicates per block. The framework of their study will be used here as a representative of such data in order to evaluate the two methods of estimating variance compoenents. We first produce confidence intervals for the variance component of simulated Alonso-Alvarez data using the method maximum likelihood available in the `lme4` library of R [20, 17]. Next, for the same simulations we derive Bayesian credible intervals [9] for the variance component through Bayesian inference using Gibbs Sampling (BUGS). Although there are many methods that implement classical and Bayesian inference, we use maximum likelihood and BUGS as a representatives of their respective types of inference because both are widely used and available free of charge.

Smaller credible intervals than confidence intervals will indicate that Bayesian inference should be used when there are few replicates per block; larger credible intervals lead to the opposite conclusion.

6.1   Model

When an organism increases the allocation of energy and resources into reproduction, it likely diverts energy and resources from other functions. One way to measure the cost of reproduction is a change in the organism's body mass. Alonso-Alvarez *et al.* compared changes in the body mass of zebra finches with varying litter sizes when they randomly formed 46 pairs of zebra finches and kept each pair in a separate cage. "Thirty-six pairs bred and their brood size was manipulated as to create two groups: (i) pairs raising a two-chicks brood ($n = 20$); (ii) pairs raising a six-chicks brood ($n = 16$)." Ten pairs did not breed. Different brood sizes and sex are considered fixed effects that explained change in body mass, while cage identity is a random effect and only two measurements within a block (cage).

A model for change in body mass during the period of study (Y) is:

$$Y_{i,j,k} = \alpha_i + \eta_j + S_k + \epsilon_{i,j,k}$$

where α_i is the fixed effect of number of offspring ($i = 1$ for no chicks, $i = 2$ for two chicks, and $i = 3$ for six chicks), η_j is the fixed effect of sex ($j = 1$ for females and $j = 2$ for males), and S_k the random effect of cage identity for $k = 1, ..., 46$. It is assumed that $S_k \sim N(0, \sigma_s^2)$ and $\epsilon \sim N(0, \sigma^2)$. In matrix notation,

$$Y = X\beta + Z\gamma + \epsilon \tag{14}$$

where

$$
X_{92\times5} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \eta_1 \\ \eta_2 \end{bmatrix}, \quad Z_{92\times46} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix}.
$$

The random vector $\gamma \sim N(0, \sigma_s^2 I)$ has dimension 46, while $\epsilon \sim N(0, \sigma^2 I)$ is a 92-dimensional random vector.

6.2 Maximum likelihood

As described, the study includes two replicates per cage. To simulate these data in R, first we define the design matrices (See AA.Sim 2 rep.R):

```
> X = cbind(  c( rep(1, 20 ), rep(0, 72) ),
+  c( rep(0, 20 ), rep(1, 40), rep(0, 32)),
+  c( rep(0, 60 ), rep(1, 32)),
+  c( rep(c(1,0), 46)),
+  c( rep(c(0,1), 46))
+  )
> X
       [,1] [,2] [,3] [,4] [,5]
 [1,]    1    0    0    1    0
 [2,]    1    0    0    0    1
 [3,]    1    0    0    1    0
 [4,]    1    0    0    0    1
...
[92,]    0    0    1    0    1
> Z = matrix( c(1, 1, rep(0, 92)), nrow = 92, ncol = 47 )[,-47]
> Z
       [,1] [,2] [,3] ... [,44] [,45] [,46]
 [1,]    1    0    0 ...     0     0     0
 [2,]    1    0    0 ...     0     0     0
 [3,]    0    1    0 ...     0     0     0
 [4,]    0    1    0 ...     0     0     0
...
[90,]    0    0    0 ...     0     1     0
[91,]    0    0    0 ...     0     0     1
[92,]    0    0    0 ...     0     0     1
```

Next, we chose somewhat arbitrary values for the effects of brood size and sex, assign variances, and select the random effect and error levels:

```
> beta <- c(0, -10, -20, 0, -4)
> sigma_s = 4
> gamma <- rnorm(46, mean = 0, sd = sigma_s)
> sigma_e = 1
> e <- rnorm(92, mean = 0, sd = sigma_e)
```

Having selected all the information of the right-hand-side of (14), we find Y:

```
> Y <- X %*% beta + Z %*% gamma + e
```

In order to use the `lme()` command (part of the `nlme` library) for a linear mixed model in R, our explanatory variables must be given as factors and combined with Y in a data frame. This is done easily as follows:

```
> broodsize <- factor( X %*% c(1, 2, 3, 0, 0) )
> sex <- factor( X %*% c(0, 0, 0, 1, 2) )
> cage <- factor( Z %*% 1:46 )
> dat <- data.frame( Y, broodsize, sex, cage )
> ML.mod <- lme(Y ~ broodsize:sex - 1, data = dat, random = ~ 1 | cage)
```

The last command uses the method of maximum likelihood to find estimates and confidence intervals of the parameters, i.e. $\beta = $ beta, $\sigma_s = $ sigma_s, and $\sigma = $ sigmasq_e. The next set of lines finds the width of the 95% confidence interval for the variance component $\sigma_s = $ sigma_s. If this interval captures the "true" value, i.e, the value assigned in this simulation (4), then `ML.capture = 1`. If not, then `ML.capture = 0`.

```
> (ML.lo <- intervals(ML.mod)$reStruct$cage$lower)
[1] 1.433
> (ML.hi <- intervals(ML.mod)$reStruct$cage$upper)
[1] 2.327
> (ML.width <- ML.hi - ML.lo)
[1] 0.8937
> ML.capture <- 1
> if(ML.hi < sigma_s ) { ML.capture <- 0 }
> if(ML.lo > sigma_s ) { ML.capture <- 0 }
> ML.capture
[1] 1
```

In order to assess the average length of these intervals, the commands for one simulation and interval can be placed in a function, which will be called here `AA.ML.1.Int(beta, sigma_s, sigma_e)`. The output of this function is `c("ML.width" = ML.width, "ML.rate" = ML.capture)` and its default values are those used in the simulation above. Another function, `AA.ML.n.Int(beta, sigma_s, sigma_e, n)`, finds the average length of intervals generated from `n` simulations and the proportion of those intervals that enclose the assigned value of `sigma_s` $= \sigma_s$. Of course, we expect that this proportion is 0.95.

```
> AA.ML.n.Int <- function(beta = c(0, -10, -20, 0, -4),
+   sigma_s = 4, sigma_e = 1, n = 10){
+   rowMeans(sapply(1:n,
+   function(i) AA.ML.1.Int(beta, sigma_s, sigma_e) ) )
+ }
> AA.ML.n.Int(n = 100)
ML.width  ML.rate
   1.756    0.950
```

Thus, the 95% maximum likelihood confidence intervals, our representative of classical methods, have a width of approximately 1.756 and capture our assigned value 95% of the time for the specifications of these simulations. They are also accurate.

6.3 BUGS

Estimates and sampling with BUGS were performed through R using the `arms` library. We first investigate an uninformative prior distribution for σ_s: Unif(0,1000). All other priors are similarly uninformative. The model for the Alonso-Alvarez simulation is notated for BUGS as

```
model{
for(i in 1:10){
s[i] ~ dnorm(0, tau.s)
mu[2*i - 1] <- s[i]
mu[2*i] <- eta2 + s[i]
y[2*i - 1] ~ dnorm(mu[2*i - 1], tau.e)
y[2*i] ~ dnorm(mu[2*i], tau.e)
}

for(i in 11:30){
s[i] ~ dnorm(0, tau.s)
mu[2*i - 1] <- alpha2 + s[i]
mu[2*i] <- alpha2 + eta2 + s[i]
y[2*i - 1] ~ dnorm(mu[2*i - 1], tau.e)
y[2*i] ~ dnorm(mu[2*i], tau.e)
}

for(i in 31:46){
s[i] ~ dnorm(0, tau.s)
mu[2*i - 1] <- alpha3 + s[i]
mu[2*i] <- alpha3 + eta2 + s[i]
y[2*i - 1] ~ dnorm(mu[2*i - 1], tau.e)
y[2*i] ~ dnorm(mu[2*i], tau.e)
}

alpha2 ~ dnorm(0, 1.0E-6)
```

```
alpha3 ~ dnorm(0, 1.0E-6)
eta2 ~ dnorm(0, 1.0E-6)
tau.e <- pow(sigma.e, -2)
tau.s <- pow(sigma.s, -2)
sigma.e ~ dunif(0, 1000)
sigma.s ~ dunif(0, 1000)
}
```

Note that $\alpha_1 =$ `alpha_1` and $\eta_1 =$ `eta_1` are excluded from this model statement in order to decrease the variable dependencies and improve the BUGS assessment of the model.

We simulate data `Y` from the Alonso-Alvarez framework and set the initials values for the parameters we wish to estimate: $\alpha_2 =$ `alpha_2`, $\alpha_3 =$ `alpha_3`, $\eta_2 =$ `eta`, $\sigma_s =$ `sigma_s` = `sigma.s`, and $\sigma =$ `sigma_e` = `sigma.e`. Using the `bugs()` function, these parameters are estimated through a Markov chain in BUGS.

```
> beta = c(0, -10, -20, 0, -4)
> sigma_s = 4
> sigma_e = 1
> gamma <- rnorm(46, mean = 0, sd = sigma_s)
> e <- rnorm(92, mean = 0, sd = sigma_e)
> X <- cbind( c( rep(1, 20 ), rep(0, 72) ),
+ c( rep(0, 20 ), rep(1, 40), rep(0, 32)),
+ c( rep(0, 60 ), rep(1, 32)),
+ c( rep(c(1,0), 46)),
+ c( rep(c(0,1), 46)))
> Z <- matrix( c(1, 1, rep(0, 92)), nrow = 92, ncol = 47 )[,-47]
> Y <- X %*% beta + Z %*% gamma + e
> y <- as.vector(Y)
> inits <- function()
+ list(alpha2 = runif(1, 0, 100),
+ alpha3 = runif(1, 0, 100),
+ eta2 = runif(1, 0, 100),
+ sigma.e = runif(1, 0, 100),
+ sigma.s = runif(1, 0, 100) )
> parameters <- c( "alpha2", "alpha3", "eta2", "sigma.e", "sigma.s")
> BUGS.data <- list("y")
> BUGS.mod <- bugs( BUGS.data, inits, parameters, "AA2.bug",
+ n.chains = 3, n.iter = 1000)
```

The file "AA2.bug" is where the model is stored. Now for the confidence intervals. For all the parameters that were estimated, these intervals are found in:

```
> BUGS.mod$summary
          mean      sd      2.5%      25%      50%      75%    97.5%  Rhat n.eff
alpha2  -10.339  0.8176 -11.8405 -10.880 -10.355  -9.791  -8.740 1.022  1500
alpha3  -19.454  1.0329 -21.1500 -20.250 -19.570 -18.718 -17.359 1.010   200
```

```
eta2      -4.019  0.2248  -4.4665  -4.166  -4.017  -3.869  -3.584 1.001  1500
sigma.e    1.102  0.1145   0.9054   1.021   1.090   1.173   1.354 1.002  1500
sigma.s    4.030  0.4677   3.2434   3.702   3.990   4.318   5.023 1.002  1500
deviance 277.807 14.0030 254.4000 267.700 277.100 286.125 309.952 1.001  1500
```

But we're only interested in the 95% confidence interval for `sigma.s`, so we label its width and encode whether the assigned value of `sigma.s` is in the interval:

```
> (BUGS.width <- BUGS.mod$summary[5,7] - BUGS.mod$summary[5,3])
[1] 1.780
> BUGS.capture <- 1
> if(BUGS.mod$summary[5,7] < sigma_s ) { BUGS.capture <- 0 }
> if(BUGS.mod$summary[5,3] > sigma_s ) { BUGS.capture <- 0 }
> BUGS.capture
[1] 1
```

As above for maximum likelihood, a function `AA.BUGS.1.Int(beta, sigma_s, sigma_e)` envelopes these commands and returns `c("BUGS.width" = BUGS.width, "BUGS.rate" = BUGS.capture)`. Then, the function, `AA.BUGS.n.Int(beta, sigma_s, sigma_e, n)`, finds the average length of intervals generated from `n` simulations and the proportion of those intervals that enclose the assigned value of `sigma_s` $= \sigma_s$. The default values for this function are those used throughout these simulations.

```
> AA.BUGS.n.Int(n = 20)
BUGS.width  BUGS.rate
    1.790     0.8261
```

Thus, for the specifications of these simulations, the 95% BUGS confidence intervals, our representative of Bayesian methods, have a width of approximately 1.790 – wider than the maximum likelihood intervals – and capture our assigned value 82% rather than the claimed 95% of the time.

So far we have considered a relatively uninformative prior distribution for σ_s, i.e. uniform on [0, 100]. However, in many cases the ecological background of the model may lead the researcher to specify an informative prior. For example, in the Alonso-Alvarez study, the change in body mass of the zebra finches is limited by the range of reasonable body masses. The variance component for the model will similarly be restricted by this range. Suppose this range is about 50 mass units. This prior information suggest the prior $\sigma_s \sim \text{Unif}(1, 50)$, or in BUGS model notation `sigma.s` $\sim$ `dunif(0, 50)` (See Figure 1). Iterative simulations and estimations are defined similar to above. The function `AA.BUGS.Inf.Unif.1.Int(beta, sigma_s, sigma_e)` returns the width of the BUGS confidence interval for σ_s with this uniform prior and whether it captures the assign value. Iteration results are averages in the function `AA.BUGS.n.Int(beta, sigma_s, sigma_e, n)`.

```
> AA.BUGS.Inf.Unif.1.Int()
BUGS.Unif.width  BUGS.Unif.rate
           2.0             1.0
> AA.BUGS.Inf.Unif.n.Int(n = 10)
```
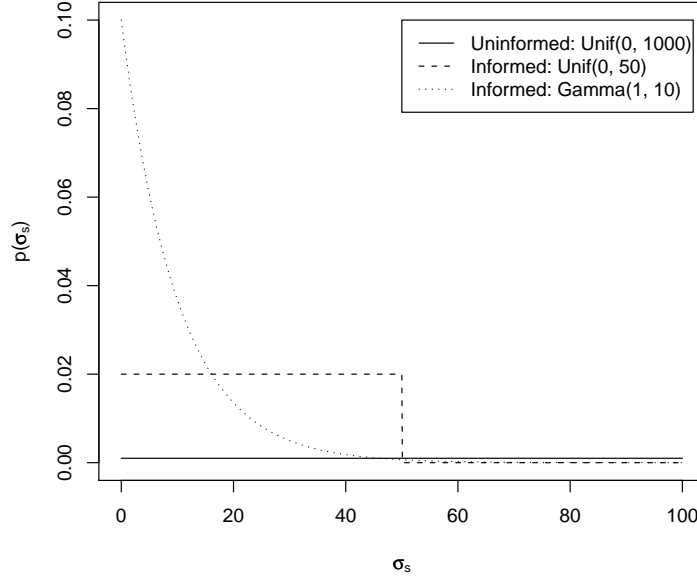
Figure 1: Some prior distributions for σ_s.

```
BUGS.Unif.width  BUGS.Unif.rate
          1.8              1.0
```

Another expectation researchers may have σ_s is that small values are more likely than larger values. A Gamma(1, 10) prior distribution for σ_s reflects this assumption. This prior is notated in BUGS model notation as `sigma.s` $\sim$ `dgamma(1, .01)`. Functions comparable to the uniform prior give the following results:

```
> AA.BUGS.Inf.Gamma.1.Int()
BUGS.Inf.Gamma.width  BUGS.Inf.Gamma.rate
                 2.0                  1.0
> AA.BUGS.Inf.Gamma.n.Int(n = 10)
BUGS.Inf.Gamma.width  BUGS.Inf.Gamma.rate
                 2.5                  1.0
```

6.4   Combining Maximum Likelihood and BUGS

A more conclusive comparison of maximum likelihood and the uninformed and informed Bayesian inference assesses estimates from the same set of simulated data. The function `AA.MLBUGS.1.Int(beta, sigma_s, sigma_e)` combines `AA.ML.1.Int()`, `AA.BUGS.1.Int()`, `AA.BUGS.Inf.Unif.1.Int()`, and `AA.BUGS.Inf.Gamma.1.Int()` into one function that returns the widths of the four confidence intervals and their "capture rate". The default values are those used throughout these simulations. As before, multiple simulations are averaged by one function `AA.MLBUGS.n.Int(beta, sigma_s, sigma_e, n)`. These functions are central to the current study, so they are reproduced here in full:

```
AA.MLBUGS.1.Int <- function( beta = c(0, -10, -20, 0, -4),
sigma_s = 4, sigma_e = 1, debug = FALSE )
{ gamma <- rnorm(46, mean = 0, sd = sigma_s)
e <- rnorm(92, mean = 0, sd = sigma_e)
X = cbind(  c( rep(1, 20 ), rep(0, 72) ),
c( rep(0, 20 ), rep(1, 40), rep(0, 32)),
c( rep(0, 60 ), rep(1, 32)),
c( rep(c(1,0), 46)),
c( rep(c(0,1), 46) ) )
Z = matrix( c(1, 1, rep(0, 92)), nrow = 92, ncol = 47 )[,-47]
Y <- X %*% beta + Z %*% gamma + e
broodsize <- factor( X %*% c(1, 2, 3, 0, 0) )
sex <- factor( X %*% c(0, 0, 0, 1, 2) )
cage <- factor( Z %*% 1:46 )
ML.data <- data.frame( Y, broodsize, sex, cage )

### Maximum Likelihood interval ###
ML.mod <- lme(Y ~ broodsize:sex - 1, data = ML.data, random = ~ 1 | cage  )
ML.hi <- intervals(ML.mod)$reStruct$cage$upper
ML.lo <- intervals(ML.mod)$reStruct$cage$lower
ML.width <- ML.hi - ML.lo
ML.capture <- 1
if(ML.hi < sigma_s ) { ML.capture <- 0 }
if(ML.lo > sigma_s ) { ML.capture <- 0 }

### BUGS -- Set up ###
parameters <- c( "alpha2", "alpha3", "eta2", "sigma.e", "sigma.s")
y <- as.vector(Y)
BUGS.data <- list("y")

### BUGS -- Uninformed Prior ###
inits <- function()
list( alpha2 = runif(1, 0, 100),
alpha3 = runif(1, 0, 100),
eta2 = runif(1, 0, 100),
sigma.e = runif(1, 0, 100),
sigma.s = runif(1, 0, 100) )
BUGS.mod <- bugs( BUGS.data, inits, parameters, "AA2.bug",
n.chains = 3, n.iter = 2000, debug)
BUGS.width <- BUGS.mod$summary[5,7] - BUGS.mod$summary[5,3]
BUGS.capture <- 1
if(BUGS.mod$summary[5,7] < sigma_s ) { BUGS.capture <- 0 }
if(BUGS.mod$summary[5,3] > sigma_s ) { BUGS.capture <- 0 }

### BUGS -- Unif(0,50) Prior ###
```

```r
inits.Unif <- function()
list( alpha2 = runif(1, 0, 100),
alpha3 = runif(1, 0, 100),
eta2 = runif(1, 0, 100),
sigma.e = runif(1, 0, 100),
sigma.s = runif(1, 0, 50 ))
BUGS.Inf.Unif.mod <- bugs( BUGS.data, inits.Unif, parameters,
"AA.Inf.Unif.bug", n.chains = 3, n.iter = 2000, debug)
BUGS.Inf.Unif.width <- BUGS.Inf.Unif.mod$summary[5,7] - BUGS.Inf.Unif.mod$summary[5,3]
BUGS.Inf.Unif.capture <- 1
if(BUGS.Inf.Unif.mod$summary[5,7] < sigma_s ) { BUGS.Inf.Unif.capture <- 0 }
if(BUGS.Inf.Unif.mod$summary[5,3] > sigma_s ) { BUGS.Inf.Unif.capture <- 0 }

### BUGS -- Gamma(1, 1/10) Prior ###
inits.Gamma <- function()
list( alpha2 = runif(1, 0, 100),
alpha3 = runif(1, 0, 100),
eta2 = runif(1, 0, 100),
sigma.e = runif(1, 0, 100),
sigma.s = rgamma(1, 1, 10) )
BUGS.Inf.Gamma.mod <- bugs( BUGS.data, inits.Gamma, parameters,
"AA.Inf.Gamma.bug", n.chains = 3, n.iter = 2000, debug)
BUGS.Inf.Gamma.width <- BUGS.Inf.Gamma.mod$summary[5,7] - BUGS.Inf.Gamma.mod$summary[5,3
BUGS.Inf.Gamma.capture <- 1
if(BUGS.Inf.Gamma.mod$summary[5,7] < sigma_s ) { BUGS.Inf.Gamma.capture <- 0 }
if(BUGS.Inf.Gamma.mod$summary[5,3] > sigma_s ) { BUGS.Inf.Gamma.capture <- 0 }

### RETURN ###
return <- t(c("ML.width" = ML.width, "ML.rate" = ML.capture,
"BUGS.width" = BUGS.width, "BUGS.rate" = BUGS.capture,
"BUGS.Unif.width" = BUGS.Inf.Unif.width,
"BUGS.Unif.rate" = BUGS.Inf.Unif.capture,
"BUGS.Gamma.width" = BUGS.Inf.Gamma.width,
"BUGS.Gamma.rate" = BUGS.Inf.Gamma.capture))
t(return)
}
AA.MLBUGS.n.Int <- function(n = 10, beta = c(0, -10, -20, 0, -4),
sigma_s = 4, sigma_e = 1, debug)
{ colMeans(sapply(1:n, function(i)
AA.MLBUGS.1.Int( beta, sigma_s, sigma_e) ) )
}
>
```

7 Conclusion

Mixed models significantly impact many sectors of ecological research because they include random effects in order to explain variability among groups of experimental units. The fixed effects, which are ordinarily of primary interest to researchers, can only be inferred through estimation of the variance of the random effects – the variance component. There are several techniques for variance component estimation, including maximum likelihood (which employs "classical" or "frequentist" inference) and BUGS (an application of Bayesian inference). Data with few replicates per block is common in ecology but poses a statistically inconvenience to these methods of estimation. The theoretical basis of Bayesian inference has led to the development of new statistical tools that at once warrant the re-evaluation earlier data and expand the frontier of ecology.

References

[1] C. Alonso-Alvarez, S. Bertrand, G. Devevey, J. Prost, B. Faivre, and G. Sorci. Increased susceptibility to oxidative stress as a proximate cost of reproduction. *Ecology Letters*, 7:363 – 368, 2004.

[2] J. A. Banta, J. Dole, M. B. Cruzan, and M. Piglucci. Evidence of local adaptation to coarse-grained environmental variation in *Arabadopsis thaliana*. *Evolution*, 61(10):2419 – 2432, 2007.

[3] B. Bårdsen, P. Fauchald, T. Tveraa, K. Langeland, N. G. Yoccoz, and R. A. Ims. Experimental evidence of a risk-sensitive reproductive allocation in a long-lived mammal. *Ecology*, 83:829 – 837, 2008.

[4] N. J. Barrowman, R. A. Myers, R. Hilborn, D. G. Kehler, and C. A. Field. The variability among populations of coho salmon in the maximum reproductive rate and depensation. *Ecological Applications*, 13(3):784 – 793, 2003.

[5] Y. Buckley, D. T. Briese, and M. Rees. Demography and management of the invasive plant species *Hypericum perforatum*. I. Using multi-level mixed-effects models for characterizing growth, survival and fecundity in a long-term data set. *Journal of Applied Ecology*, 40:481 – 493, 2003.

[6] Y. M. Buckley, P. Downey, S. V. Fowler, R. Hill, J. Memmot, H. Norambuena, M. Pitcairn, R. Shaw, A. W. Sheppard, C. Winks, R. Wittenberg, and M. Rees. Are invasives bigger? A global study of seed size variation in two invasive shrubs. *Ecology*, 84:1434 – 1440, 2003.

[7] B. J. Cardinale and M. A. Palmer. Disturbance moderates biodiversity-ecosystem function relationships: Experimental evidence from caddisflies in stream mesocosms. *Ecology*, 83(7):1915 – 1927, 2002.

[8] Ronald Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models.* Springer-Verlag, 3 edition, 2002.

[9] J. S. Clark. Why environmentalists are becoming Bayesians. *Ecology Letters*, 8:2 – 14, 2005.

[10] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 2 edition, 2004.

[11] D. Hall, V. Luquez, V. M. Garcia, K. St. Onge, S. Jansson, and P. K. Ingvarsson. Adaptive population differentiation in phenology across a latitudinal gradient in european aspen (*Populus termula,* L.). *Evolution*, 61(12):2849 – 2860, 2007.

[12] W. A. Link and J. R. Sauer. A hierarchical analysis of population change with application to Cerulean Warblers. *Ecology*, 83(10):2833 – 2840, 2002.

[13] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337, 2000.

[14] M. A. McCarthy and P. Masters. Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology*, 42:1012 – 1019, 2005.

[15] J. A. Newman, J. Bergelson, and A. Grafen. Blocking factors and hypothesis tests in ecology: Is your statistics test wrong? *Ecology*, 78:1312–1320, 1997.

[16] R. B. O'Hara and J. Merila. Bias and precision in qst estimates: Problems and some solutions. *Genetics*, 171:1331 – 1339, 2005.

[17] J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-Plus*. Springer-Verlag, 2000.

[18] G. B. Schaalje, J. B. McBride, and G. W. Fellingham. Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics*, 7:512 – 524, 2002.

[19] P. A. Stephens, S. W. Buskirk, and C. Martínez del Rio. Inference in ecology and evolution. *Trends in Ecology and Evolution*, 22:192 – 197, 2006.

[20] R Development Core Team. R: A language and environment for statistical computing, 2008.