

## CHAPTER 6

### THE LINEAR MODEL

#### 6.1. The “fixed” effects model

##### 6.1.1. *The model*

The model corresponds, in a frequentist context, to a “fixed effects model” with or without covariates. In a Bayesian context all effects are random, thus there is no distinction between fixed models, random models or mixed models. We will describe here the Normal linear model, although other distributions of the data can be considered, and the procedure will be the same. Our model consists in a set of effects and covariates plus an error term. For example, if we measure the weight at weaning of a rabbit and we have a season effect (with two levels) and a parity effect (with two levels) plus a covariate ‘weight of the dam’, the model will be

$$y_{ijkl} = \mu + S_i + P_j + b \cdot A_{ijk} + e_{ijkl}$$

where S is the season effect, P the parity effect and A the age of the dam. As there are several piglets in a litter,  $y_{ijk}$  is the weight of the piglet l of a dam that belongs to the herd i, the piglet was born in the parity j, the dam had an age  $A_{ijk}$  that was the same for all piglets born in the same litter. For example, in the following equations we have two rabbits of the same litter weighting 520 and 430 grams, born in season 1 and the second parity of the dam 111 that weighted 3400 grams. Then we have a rabbit that weighted 480 grams, born in the second season and the first parity of dam 221 that weighted 4200 grams, and finally we have a rabbit weighting 550 grams, born in season 1 and parity 1 of the dam 112 that weighted 4500 grams.

$$520 = \mu + S_1 + P_2 + b \cdot 3400 + e_{1111}$$

$$430 = \mu + S_1 + P_2 + b \cdot 3400 + e_{1112}$$

$$480 = \mu + S_2 + P_1 + b \cdot 4200 + e_{2211}$$

$$550 = \mu + S_1 + P_1 + b \cdot 4500 + e_{1121}$$

In matrix form,

$$\begin{bmatrix} 520 \\ 430 \\ 480 \\ 550 \\ \dots \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 3400 \\ 1 & 1 & 0 & 0 & 1 & 3400 \\ 1 & 0 & 1 & 0 & 1 & 4200 \\ 1 & 1 & 0 & 1 & 0 & 4500 \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} \mu \\ S_1 \\ S_2 \\ P_1 \\ P_2 \\ b \end{bmatrix} + \begin{bmatrix} e_{1111} \\ e_{1112} \\ e_{2211} \\ e_{1121} \\ \dots \end{bmatrix}$$

And, in general form,

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

where  $\mathbf{y}$  contains the data,  $\mathbf{X}$  is a matrix containing the covariates and the presence (1) or absence (0) of the levels of the effects.  $\mathbf{b}$  is a vector of all unknowns and  $\mathbf{e}$  is a vector with the errors. We consider the errors having mean zero and being independently normally distributed, all of them with the same variance,

$$\mathbf{e} \mid \sigma^2 \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$$

as in a Bayesian context there are not fixed effects, the correct way of expressing the distribution of the data is

$$\mathbf{y} \mid \mathbf{b} \sim N(\mathbf{X}\mathbf{b}, \mathbf{I}\sigma^2)$$

$$f(\mathbf{y} \mid \mathbf{b}, \sigma^2) \propto \frac{1}{|\mathbf{I}\sigma^2|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\mathbf{b})' [\mathbf{I}\sigma^2]^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b})}{2}\right] \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b})}{2\sigma^2}\right]$$

In a Bayesian context, to completely specify the model we also need the prior distribution of the unknowns. We will consider below two cases as in the Baby model: flat and conjugated priors.

Our objective is to find the marginal posterior distributions of all unknowns; in our example

$$f(\mu \mid \mathbf{y}), f(S_1 \mid \mathbf{y}), f(S_2 \mid \mathbf{y}), f(P_1 \mid \mathbf{y}), f(P_2 \mid \mathbf{y}), f(\sigma^2 \mid \mathbf{y})$$

or combinations of effects, for example

$$f\left(\frac{\mu + S_1}{\mu + S_2} \mid \mathbf{y}\right)$$

Although there are analytical solutions as in the Baby model, we only will develop the Gibbs sampling procedure. We need to obtain samples of the joint posterior distribution

$$f(\mathbf{b}, \sigma^2 \mid \mathbf{y})$$

We will obtain a matrix of chains, in which each row is a random sample of the joint distribution,

$\mu$	$S_1$	$S_2$	$E_1$	$E_2$	$\beta$	$\sigma^2$
478	-15	10	-45	39	0.21	140
501	-10	2	-87	102	0.12	90
523	3	51	-12	65	0.15	120
...	...	...	...	...	...	...

each column is a random sample of the marginal posterior distribution of each element of  $\mathbf{b}$ , and the last chain is a random sample of the marginal posterior distribution of  $\sigma^2$

### 6.1.2. Marginal posterior distributions via MCMC using Flat priors

To work with MCMC-Gibbs sampling we need the *conditional* distributions of the unknowns  $f(\mathbf{b} | \mathbf{y}, \sigma^2)$  and  $f(\sigma^2 | \mathbf{y}, \mathbf{b})$ . We do not know them, but we can calculate them using Bayes theorem. Using flat priors

$$\begin{aligned} \mathbf{b} &\sim U[\mathbf{b}_1, \mathbf{b}_2] &\longrightarrow & f(\mathbf{b}) = \text{constant} \\ \sigma^2 &\sim U[0, s] &\longrightarrow & f(\sigma^2) = \text{constant} \end{aligned}$$

where  $U$  is the uniform function with its bounds. Now, the posterior distributions are

$$f(\mathbf{b} | \mathbf{y}, \sigma^2) \propto f(\mathbf{y} | \mathbf{b}, \sigma^2) f(\mathbf{b}) \propto f(\mathbf{y} | \mathbf{b}, \sigma^2)$$

$$f(\sigma^2 | \mathbf{y}, \mathbf{b}) \propto f(\mathbf{y} | \sigma^2, \mathbf{b}) f(\sigma^2) \propto f(\mathbf{y} | \sigma^2, \mathbf{b})$$

as we know the distribution of the data, we can obtain both conditionals

$$f(\mathbf{b} | \mathbf{y}, \sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{2\sigma^2}\right]$$

$$f(\sigma^2 | \mathbf{y}, \mu) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{2\sigma^2}\right]$$

Notice that the formulae are the same, but the variable is in red, thus the functions are completely different.

The conditional distribution of  $\mathbf{b}$  is a multinormal distribution (Appendix 6.1).

$$f(\mathbf{b} | \mathbf{y}, \sigma^2) \propto \frac{1}{|\mathbf{X}'\mathbf{X}\sigma^2|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \cdot (\mathbf{b} - \hat{\mathbf{b}})' [(\mathbf{X}'\mathbf{X})^{-1} \sigma^2]^{-1} (\mathbf{b} - \hat{\mathbf{b}})\right] \sim N[\hat{\mathbf{b}}, (\mathbf{X}'\mathbf{X})^{-1} \sigma^2]$$

where  $\hat{\mathbf{b}}$  has the same form as the minimum least square estimator (<sup>1</sup>)

$$\hat{\mathbf{b}} = [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{y}$$

and the conditional distribution of  $\sigma^2$  is an inverted gamma distribution with parameters

---

<sup>1</sup> We do not say that  $\hat{\mathbf{b}}$  is the minimum least square estimator but that it has “the same form” to stress that we are not estimating anything by least squares in a frequentist way, although the first proof of the least square method given by Gauss was a Bayesian one (Gauss, 1809).

$$\beta = \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

$$\alpha = \frac{n}{2} - 1$$

$$f(\sigma^2 | \mathbf{y}, \mathbf{b}) \sim \text{IG}(\alpha, \beta)$$

We have algorithms to extract random samples of both functions, thus we can start with the Gibbs sampler as we have seen in chapter 4 and in the particular case of the baby model in chapter 5.

We start with an arbitrary value for the variance (for example) and then we get a multiple sample value of  $\mathbf{b}$ . We substitute this value in the conditional of the variance and we get a random value of the variance. We substitute it in the conditional distribution of the  $\mathbf{b}$  and we continue the process (figure 6.1), as we have seen in chapter 5 for the baby model.

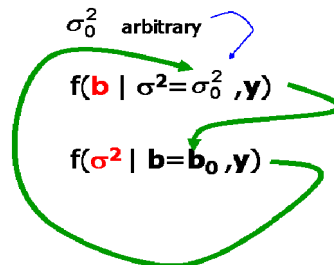


Figure 6.1. Gibbs sampling process for the  $\mathbf{b}$  and the variance of the Normal linear model

#### 6.1.4. Least Squares as a Bayesian Estimator

The least square estimator was developed by Legendre under intuitive bases. Later, Gauss found the first statistical justification of the method, developing least squares first as the mode of the conditional posterior distribution and later under frequentists bases. In a Bayesian context, we have seen that, under flat priors,

$$f(\mathbf{b} | \mathbf{y}, \sigma^2) \sim N[\hat{\mathbf{b}}, (\mathbf{X}'\mathbf{X})^{-1} \sigma^2]$$

As in a Normal distribution the mean, mode and median are the same, the least square estimator can be interpreted as

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \text{mode } f(\mathbf{b} | \sigma^2, \mathbf{y}) = \text{median } f(\mathbf{b} | \sigma^2, \mathbf{y}) = \text{mean } f(\mathbf{b} | \sigma^2, \mathbf{y})$$

Notice that this is a conditional distribution; i.e., the least square estimator needs a value for  $\sigma^2$  to be calculated. In a frequentist context, we estimate  $\sigma^2$ , then we take this as the true value for the variance and we calculate  $\hat{\mathbf{b}}$ . We do not take into account the error committed when estimating the variance. However, in a Bayesian context, we calculate the *marginal* posterior distribution for  $\mathbf{b}$

$$f(\mathbf{b} | \mathbf{y}) = \int f(\mathbf{b}, \sigma^2 | \mathbf{y}) d\sigma^2 = \int f(\mathbf{b} | \sigma^2, \mathbf{y}) f(\sigma^2) d\sigma^2$$

we take into account all possible values of  $\sigma^2$  and multiply by their probabilities, integrating afterwards; i.e., we take into account the error of estimating the variance.

In more philosophical grounds, as we need the *true* value of  $\sigma^2$  to find the least square estimate  $\hat{\mathbf{b}}$ , we know that we never find a real least square estimate because we do not know the true value of  $\sigma^2$ . Bayesian theory does not require true values to work. The Bayesian interpretation of  $\hat{\mathbf{b}}$  is the mean, mode and median of a conditional distribution, in which  $\sigma^2$  is given. We do the same, but at least now we know what we are doing.

## 6.2. The “mixed” model

### 6.2.1. *The model*

The model corresponds, in a frequentist context, to a “mixed model”. As we said before, in a Bayesian context all effects are random, thus there is no distinction between fixed models, random models or mixed models. We will also consider here that the data are normally distributed, although other distributions of the data can be considered, and the procedure will be the same. Our model consists in a set of effects and covariates plus what in a frequentist model is a “random” effect, plus an error term. We can add an individual genetic effect to the model of our former example, and we have in this case

$$y_{ijkl} = \mu + S_i + P_j + b \cdot A_{ijk} + u_{ijkl} + e_{ijkl}$$

where  $u_{ijkl}$  is the individual genetic effect. In matrix form

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where  $\mathbf{y}$  contains the data,  $\mathbf{X}$  is a matrix containing the covariates and the presence (1) or absence (0) of the levels of the effects.  $\mathbf{b}$  is a vector of what in a frequentist context are “fixed effects” and covariates,  $\mathbf{u}$  is a vector with the individual genetics effects that in a frequentist context are considered “random”, and  $\mathbf{e}$  is a vector of the residuals.  $\mathbf{Z}$  is a matrix containing the covariates and the presence (1) or absence (0) of the levels of the individual genetics effects. If all individuals have records,  $\mathbf{Z} = \mathbf{I}$ , but if some individuals do not have records (for example the parental population, or traits in which data are recorded in only one sex),  $\mathbf{Z}$  is not the identity matrix; for example: if individuals 1,3, 4, 5 and 6 have records but individuals 2 has no records,

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z} \mathbf{u} + \mathbf{e}$$

$$\begin{bmatrix} 520 \\ 430 \\ 480 \\ 550 \\ 500 \end{bmatrix} = \mathbf{Xb} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{bmatrix} + \mathbf{e}$$

Thus we will consider that we have 'n' data and 'q' genetic effects to be estimated. We consider the residuals normally independently distributed all of them with mean zero and the same variance  $\sigma^2$ . We will take bounded flat priors for the variance

$$\mathbf{e} \mid \sigma^2 \sim N(\mathbf{0}, I\sigma^2)$$

$$\sigma^2 \sim U[0,p]$$

where  $U[0,p]$  is the uniform distribution between 0 and p, both included, and p is subjectively chosen.

The genetic effects are normally distributed with a variance-covariance matrix that depends on the additive genetic variance  $\sigma_u^2$  and the relationship matrix  $\mathbf{A}$ . This last matrix has the relationship coefficients between genetic effects and it is a known matrix that is calculated according to the parental relationships between individuals, based in Mendel's laws. We will also take a bounded uniform distribution for the genetic variance.

$$\mathbf{u} \mid \sigma_u^2, \mathbf{A} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$$

$$\sigma_u^2 \sim U[0,p]$$

where  $U[0,p]$  is the uniform distribution between 0 and p, both included, and p is subjectively chosen.

We will also consider a uniform distribution for the other effects

$$\mathbf{b} \sim U[\mathbf{0}, \mathbf{w}]$$

where  $\mathbf{0}$  and  $\mathbf{w}$  are vectors and  $\mathbf{w}$  is subjectively chosen.  $U[\mathbf{0}, \mathbf{w}]$  is the uniform distribution between  $\mathbf{0}$  and  $\mathbf{w}$ , both included.

Now the model is completely specified and we can write the distribution of the data

$$\mathbf{y} \mid \mathbf{b}, \mathbf{u}, \sigma^2 \sim N(\mathbf{Xb} + \mathbf{Zu}, I\sigma^2)$$

Notice that this is a simplified notation. We should have written

$$\mathbf{y} \mid \mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u}, \mathbf{A}, \sigma_u^2, \sigma^2, \mathcal{H} \sim N(\mathbf{Xb} + \mathbf{Zu}, I\sigma^2)$$

where  $\mathcal{H}$  is the set of hypothesis we also need to define the data distribution (for example; the hypothesis that the sample has been randomly collected). We simplify the notation because  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{A}$  and  $\mathcal{H}$  are always known and because given  $\mathbf{u}$  we do not need  $\mathbf{A}$  and  $\sigma_u^2$  any more, since the genetic effects are yet determined. We can also write, in simplified notation,

$$\mathbf{y} \mid \mathbf{b}, \sigma_u^2, \sigma^2 \sim N(\mathbf{Xb}, \mathbf{Z}'\mathbf{A}\mathbf{Z}\sigma_u^2 + I\sigma^2)$$

We have now new unknowns and combination of unknowns to estimate. Our objective is to find the marginal posterior distributions of all unknowns

$$f(\boldsymbol{\mu}|\mathbf{y}), f(\mathbf{S}_1|\mathbf{y}), f(\mathbf{S}_2|\mathbf{y}), f(\mathbf{P}_1|\mathbf{y}), f(\mathbf{P}_2|\mathbf{y}), f(\mathbf{u}_1|\mathbf{y}), f(\mathbf{u}_2|\mathbf{y}), \dots, f(\sigma_u^2|\mathbf{y}), f(\sigma^2|\mathbf{y})$$

or combinations of them, for example we can be interested in estimating the marginal posterior distribution of the response to selection, which can be defined as the distribution of the average of the genetic values of the last generation.

### 6.2.2. Marginal posterior distributions via MCMC

To work with MCMC-Gibbs sampling we need the *conditional* distributions of the unknowns.

$$f(\mathbf{b} | \mathbf{u}, \sigma_u^2, \sigma^2, \mathbf{y})$$

$$f(\mathbf{u} | \mathbf{b}, \sigma_u^2, \sigma^2, \mathbf{y})$$

$$f(\sigma_u^2 | \mathbf{u}, \mathbf{b}, \sigma^2, \mathbf{y})$$

$$f(\sigma^2 | \mathbf{u}, \mathbf{b}, \sigma_u^2, \mathbf{y})$$

We will write first the joint distribution

$$f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma^2 | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{u}, \mathbf{b}, \sigma_u^2, \sigma^2) \cdot f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma^2)$$

Some unknowns are not independent “a priori”; for example  $\mathbf{u}$  and  $\sigma_u^2$ . In this case we know by the theory of probability that  $P(A,B)=P(A|B) \cdot P(B)$ , thus

$$f(\mathbf{u}, \sigma_u^2) = f(\mathbf{u} | \sigma_u^2) \cdot f(\sigma_u^2)$$

but if some unknowns are independent “a priori”, for example  $\mathbf{u}$  and  $\mathbf{b}$  we know that

$$f(\mathbf{b}, \mathbf{u}) = f(\mathbf{b}) \cdot f(\mathbf{u})$$

then, assuming independence “a priori” between some unknowns, we have

$$f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma^2 | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{u}, \mathbf{b}, \sigma_u^2, \sigma^2) \cdot f(\mathbf{b}) \cdot f(\mathbf{u} | \sigma_u^2) \cdot f(\sigma_u^2) \cdot f(\sigma^2)$$

This supposition sometimes holds and sometimes not. For example, it is well known that the best dairy cattle farms have the best environment and also buy the best semen, thus their genetic level is higher and this generates a positive covariance between  $\mathbf{b}$  and  $\mathbf{u}$ . In the case of pigs in a model with only season effects, for example, it is not expected that the genetically best pigs come in summer or in winter, thus it seems that we can assume independence between  $\mathbf{b}$  and  $\mathbf{u}$ . It is less clear the independence between the genetic and environmental variances. Usually the literature in genetics offers heritabilities, which is a ratio between the genetic variance and the sum of the genetic and environmental variances, thus it is not easy to have a subjective opinion of the genetic variance independent of our opinion about the environmental variance. However it is even more complicated to assess our opinion about the covariances, thus Bayesian geneticists prefer, with some exception (Blasco et al., 1998) to consider prior independence, hoping that the data will dominate the final result and this assumption will not have any consequence in the results.

Considering the prior distributions we have established in 6.2.1., we have

$$f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma^2 | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{b}, \mathbf{u}, \sigma^2) \cdot f(\mathbf{u} | \mathbf{A}\sigma_u^2) \propto$$

$$\propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})\right] \frac{1}{(\sigma_u^2)^{\frac{q}{2}}} \exp\left[-\frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\right]$$

Now, all the conditionals are based in this function, but changing the red and black parts.

$$f(\sigma^2 | \mathbf{b}, \mathbf{u}, \sigma_u^2, \mathbf{y}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})\right] \frac{1}{(\sigma_u^2)^{\frac{q}{2}}} \exp\left[-\frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\right] \propto$$

$$\propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})}{2\sigma^2}\right]$$

This is an inverted gamma, as we have seen in 6.2.1 and chapter 3, with parameters

$$\beta = \frac{1}{2}(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})$$

$$\alpha = \frac{n}{2} - 1$$

$$f(\sigma_u^2 | \mathbf{b}, \mathbf{u}, \sigma^2, \mathbf{y}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})\right] \frac{1}{(\sigma_u^2)^{\frac{q}{2}}} \exp\left[-\frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\right] \propto$$

$$\propto \frac{1}{(\sigma_u^2)^{\frac{q}{2}}} \exp\left[-\frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\right]$$

This is also an inverted gamma distribution, with parameters

$$\beta = \frac{1}{2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$$

$$\alpha = \frac{q}{2} - 1$$



$$f(\mathbf{b} | \mathbf{u}, \sigma_u^2, \sigma^2, \mathbf{y}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})' (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu}) \right] \exp \left[ -\frac{1}{2\sigma_u^2} \mathbf{u}' \mathbf{A}^{-1} \mathbf{u} \right] \propto$$

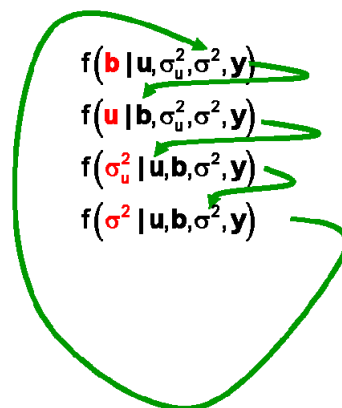
$$\propto \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})' (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu}) \right] \propto \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y}^* - \mathbf{Xb})' (\mathbf{y}^* - \mathbf{Xb}) \right]$$

where  $\mathbf{y}^* = \mathbf{y} - \mathbf{Zu}$ . We have seen in 6.2.1 that this can be transformed in a multinormal distribution.

$$f(\mathbf{u} | \mathbf{b}, \sigma_u^2, \sigma^2, \mathbf{y}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu})' (\mathbf{y} - \mathbf{Xb} - \mathbf{Zu}) \right] \frac{1}{(\sigma_u^2)^{\frac{q}{2}}} \exp \left[ -\frac{1}{2\sigma_u^2} \mathbf{u}' \mathbf{A}^{-1} \mathbf{u} \right]$$

that can also be converted after some algebra in a multinormal distribution.

After having the conditionals identified as functions from which we have algorithms for taking random samples from them, we can start with the Gibbs sampling procedure (Figure 6.2) as we did in the case 6.2.1.



**Figure 6.2.** Gibbs sampling process for the components of the model  $\mathbf{y}=\mathbf{Xb}+\mathbf{Zu}+\mathbf{e}$  and the variance components

### **BLUP as a Bayesian estimator**

In a Bayesian context there are no differences between fixed and random effects, thus we do not need mixed models and we can work with the same model as in 6.1.2. We will use vague priors for  $\mathbf{b}$ . We assume independence between  $\mathbf{b}$  and  $\mathbf{u}$ .

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} = \mathbf{Wt} + \mathbf{e}$$

$$\mathbf{W} = [\mathbf{X} \ \mathbf{Z}] \quad \mathbf{t}' = [\mathbf{b}' \ \mathbf{u}']$$

$$\mathbf{b} | \mathbf{m}_b, \mathbf{S} \sim N(\mathbf{m}_b, \mathbf{S})$$

$$\mathbf{u} | \sigma_u^2 \sim N(\mathbf{0}, \mathbf{A} \sigma_u^2)$$

$$\mathbf{t} | \mathbf{m}, \mathbf{S}, \sigma_u^2 \sim N\left(\begin{bmatrix} \mathbf{m}_b \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \sigma_u^2 \end{bmatrix}\right) \equiv N(\mathbf{m}, \mathbf{V})$$

$$\text{where } \mathbf{m} = \begin{bmatrix} \mathbf{m}_b \\ \mathbf{0} \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \sigma_u^2 \end{bmatrix}$$

$$\mathbf{y} | \mathbf{t}, \sigma^2 \sim N(\mathbf{Wt}, \mathbf{I} \sigma^2)$$

Now we will find the mode of the posterior distribution of  $\mathbf{t}$  given the data, *but also conditioned to the variance components*. Applying Bayes theorem, we have

$$\begin{aligned} f(\mathbf{t} | \mathbf{y}, \mathbf{V}, \sigma^2) &\propto f(\mathbf{y} | \mathbf{t}, \mathbf{V}, \sigma^2) \cdot f(\mathbf{t} | \mathbf{V}) \propto \\ &\propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{Wt})' (\mathbf{I} \sigma^2)^{-1} (\mathbf{y} - \mathbf{Wt})\right] \exp\left[-\frac{1}{2}(\mathbf{t} - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{t} - \mathbf{m})\right] \end{aligned}$$

$$\ln f(\mathbf{t} | \mathbf{y}, \mathbf{m}, \mathbf{V}, \sigma^2) \propto \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{Wt})' (\mathbf{y} - \mathbf{Wt}) + (\mathbf{t} - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{t} - \mathbf{m})$$

$$\frac{\partial}{\partial \mathbf{t}} \ln f(\mathbf{t} | \mathbf{y}, \mathbf{V}, \sigma^2) \propto -\frac{1}{\sigma^2} \mathbf{W}' (\mathbf{y} - \mathbf{Wt}) + \mathbf{V}^{-1} (\mathbf{t} - \mathbf{m})$$

equating to zero this leads to

$$\mathbf{W}' \mathbf{W} \hat{\mathbf{t}} + \sigma^2 \mathbf{V}^{-1} \hat{\mathbf{t}} = \mathbf{W}' \mathbf{y} - \sigma^2 \mathbf{V}^{-1} \mathbf{m}$$

$$\begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix}' \begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} + \sigma^2 \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \sigma_u^2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix}' \mathbf{y} - \sigma^2 \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \sigma_u^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{m}_b \\ \mathbf{0} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} + \sigma^2 \mathbf{S}^{-1} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma^2}{\sigma_u^2} \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} - \sigma^2 \mathbf{S}^{-1} \mathbf{m}_b \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

These equations are very similar to the mixed model equations. In fact, if  $\mathbf{S}^{-1} = \mathbf{0}$  they are identical to the mixed model equations. This condition only holds if the prior variance of  $\mathbf{b}$  is infinite; i.e., if we use unbounded flat priors for  $\mathbf{b}$ . Therefore, in a Bayesian context, the difference between what in a frequentist context are called “fixed” or “random” effects is only the type of prior they have. A “fixed” effect in a frequentist context is just a random effect with an unbounded flat prior in a Bayesian context. The mystery of the difference between fixed and random effects has now been solved.

In a Bayesian context, BLUP is the mode of the joint posterior distribution of  $\mathbf{b}$  and  $\mathbf{u}$ , conditioned not only to the data, but to the values of the variance components, when we use an unbounded flat prior for  $\mathbf{b}$ . Notice that in a Bayesian context BLUP is not biased or unbiased, since there are not repetitions of the experiment. We can be interested in what will happen in repetitions of the experiment, but our inferences are based only in our sample and the priors, not in the information of the sampling space.

#### 6.2.4. REML as a Bayesian estimator

We have seen in 5.2.3 that the mode of the marginal posterior distribution of the variance gives an expression that is the same we obtain in a frequentist context for the REML estimate. The same happens in the linear model, the REML estimators of  $\sigma_u^2$  and  $\sigma^2$  are coincident with the mode of the joint marginal posterior density

$$\begin{aligned} f(\sigma_u^2, \sigma^2 | \mathbf{y}) &= \iint f(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma^2 | \mathbf{y}) d\mathbf{b} d\mathbf{u} = \\ &= \iint f(\sigma_u^2, \sigma^2 | \mathbf{b}, \mathbf{u}, \mathbf{y}) \cdot f(\mathbf{b}) \cdot f(\mathbf{u}) d\mathbf{b} d\mathbf{u} \propto \iint f(\sigma_u^2, \sigma^2 | \mathbf{b}, \mathbf{u}, \mathbf{y}) \cdot f(\mathbf{u}) d\mathbf{b} d\mathbf{u} \end{aligned}$$

when prior values are assumed to be flat for  $\mathbf{b}$  and normal for  $\mathbf{u}$ , as in the case of BLUP. Notice that this is not the best Bayesian solution; we usually will prefer the mean or the median of each marginal distribution for each variance component instead of the mode of the joint distribution of both variance components.

### 6.3. The multivariate model

#### 6.3.1. The model

When several correlated traits are analysed together, we use a multivariate model. Sometimes the models for each trait may be different. For example, when analyzing litter size and growth rate, a dam may have several litters and consequently several records for litter size, but only one data for growth rate. Moreover, many animals will have one data for growth rate but no data for litter size, because they were males or they were not selected to be reproductive stock. We will put an example in which one trait has several records and the other trait only one record: for example, in dairy cattle we have several records for milk production but only one record for type traits. This means that we can add an environmental effect that is common for all lactation records, but we do not have this effect in the type traits because we only have one record for animal. The multivariate model is

$$\mathbf{y}_1 = \mathbf{X}_1 \mathbf{b}_1 + \mathbf{Z}_1 \mathbf{u}_1 + \mathbf{e}_1$$

$$\mathbf{y}_2 = \mathbf{X}_2 \mathbf{b}_2 + \mathbf{Z}_2 \mathbf{u}_2 + \mathbf{W}_2 \mathbf{p}_2 + \mathbf{e}_2$$

where  $\mathbf{b}_1$  and  $\mathbf{b}_2$  are environmental effects (season, herd, etc.),  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are genetic effects,  $\mathbf{p}_2$  is the common environmental effect to all records of trait 2, and  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are the residuals. We assume

$$\mathbf{y}_1 | \mathbf{b}_1, \sigma_1^2 \sim N(\mathbf{X}\mathbf{b}_1, \mathbf{I}\sigma_1^2)$$

$$\mathbf{y}_2 \mid \mathbf{b}_2, \mathbf{u}_2, \mathbf{p}_2, \sigma_2^2 \sim N(\mathbf{X}\mathbf{b}_2 + \mathbf{Z}\mathbf{u}_2 + \mathbf{W}\mathbf{p}_2, \mathbf{I}\sigma_2^2)$$

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \sim \text{Uniform bounded}$$

$$\mathbf{p}_2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$$

$$\sigma_p^2 \sim \text{Uniform, } \geq 0, \text{ bounded}$$

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}_{\text{SORTED BY INDIVIDUAL}} \sim N(\mathbf{0}, \mathbf{G} \otimes \mathbf{A})$$

$$\mathbf{G} = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 \end{bmatrix} \sim \text{Uniform bounded}$$

$$\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}_{\text{SORTED BY INDIVIDUAL}} \sim N(\mathbf{0}, \mathbf{R} \otimes \mathbf{I})$$

$$\mathbf{R} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix} \sim \text{Uniform bounded}$$

when vague priors are used, a multinormal distribution is often used for the priors of  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , and Inverted Whishart distributions (the equivalent to the inverted Gamma for the multivariate case) are used for  $\mathbf{G}$  and  $\mathbf{R}$ .

It is also assumed prior independence between some unknowns. As most priors are constant,

$$f(\mathbf{u}_1, \mathbf{u}_2, \mathbf{b}_1, \mathbf{b}_2, \mathbf{p}, \sigma_p^2, \mathbf{G}, \mathbf{R}) \propto f(\mathbf{u}_1, \mathbf{u}_2 \mid \mathbf{G}) \cdot f(\mathbf{p} \mid \sigma_p^2)$$

This model has the great problem in order to be managed that the design matrixes are different and we have an effect more in trait 2. If all design matrixes would be the same, we could write the model with both traits as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e}$$

where  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{p}$  and  $\mathbf{u}$  are the data and the effects of both traits. The only new distributions we need is

$$\begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}_{\text{SORTED BY INDIVIDUAL}} \sim N(\mathbf{0}, \mathbf{P} \otimes \mathbf{I})$$

$$\mathbf{P} = \begin{bmatrix} \sigma_{p_1}^2 & \sigma_{p_1 p_2} \\ \sigma_{p_1 p_2} & \sigma_{p_2}^2 \end{bmatrix} \sim \text{Uniform bounded}$$

In next paragraph we will see how we can write the multivariate model as if all design matrixes were the same and all traits would have the same effects. This technique is known as “data augmentation”.

### 6.3.2. Data Augmentation

Data augmentation is a procedure to augment the data base filling the gaps until all traits have the same design matrixes. Thus, if some traits have several season effects and one of the traits has only one season effect, new data come with several season effects for this trait until it has the same  $\mathbf{X}$  matrix as the others. If one trait has only one record, new records are added until we have also a common environmental effect for this trait. The conditions that these new records added must follow are:

1. The new records are added to fill the gaps until all traits have the same design matrixes
2. The new records must not be used for inferences, since they are not real records.

The second condition is important. Inferences are only based on the sample  $\mathbf{y}$ , the augmented data must not be used for inferences, and they will not add any information or modify the result of the analyses.

Let us call  $\mathbf{z}' = [\mathbf{z}'_1, \mathbf{z}'_2]$  the vector of augmented data for trait 1 and 2, and let us call the new data vector with the recorded and the augmented data

$$\mathbf{y}_1^* = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{z}_1 \end{bmatrix} \quad \mathbf{y}_2^* = \begin{bmatrix} \mathbf{y}_2 \\ \mathbf{z}_2 \end{bmatrix}$$

the new multivariate model is now

$$\mathbf{y}_1^* = \mathbf{X}\mathbf{b}_1 + \mathbf{Z}\mathbf{u}_1 + \mathbf{W}\mathbf{p}_1 + \mathbf{e}_1$$

$$\mathbf{y}_2^* = \mathbf{X}\mathbf{b}_2 + \mathbf{Z}\mathbf{u}_2 + \mathbf{W}\mathbf{p}_2 + \mathbf{e}_2$$

which can be written as

$$\mathbf{y}^* = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e}$$

and solved as in 6.2.2. Now we should find a way to generate the augmented data to avoid that they will take part in the inference.

Let us call  $\boldsymbol{\theta}$  all unknowns

$$\boldsymbol{\theta} = \mathbf{u}, \mathbf{b}, \mathbf{p}, \sigma_p^2, \mathbf{G}, \mathbf{R}$$

We should generate the augmented data  $\mathbf{z}$  that are also unknown and should be treated as unknowns. As with the other unknowns  $\boldsymbol{\theta}$ , we should estimate the posterior distribution conditioned to the data  $f(\boldsymbol{\theta}, \mathbf{z} \mid \mathbf{y})$ . We do not know this distribution, but we know the distribution of the data and we can apply Bayes theorem. The joint prior, according to the laws of probability, is

$$f(\boldsymbol{\theta}, \mathbf{z} | \mathbf{y}) = f(\boldsymbol{\theta} | \mathbf{y}) \cdot f(\mathbf{z})$$

Applying Bayes theorem, we have

$$f(\boldsymbol{\theta}, \mathbf{z} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\theta}, \mathbf{z}) \cdot f(\boldsymbol{\theta}, \mathbf{z}) = f(\mathbf{y} | \boldsymbol{\theta}, \mathbf{z}) \cdot f(\mathbf{z} | \boldsymbol{\theta}) \cdot f(\boldsymbol{\theta})$$

but, according to the laws of probability, we have

$$f(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) = f(\mathbf{y} | \boldsymbol{\theta}, \mathbf{z}) \cdot f(\mathbf{z} | \boldsymbol{\theta})$$

thus, substituting, we have

$$f(\boldsymbol{\theta}, \mathbf{z} | \mathbf{y}) \propto f(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}) = f(\mathbf{y}^* | \boldsymbol{\theta}) \cdot f(\boldsymbol{\theta})$$

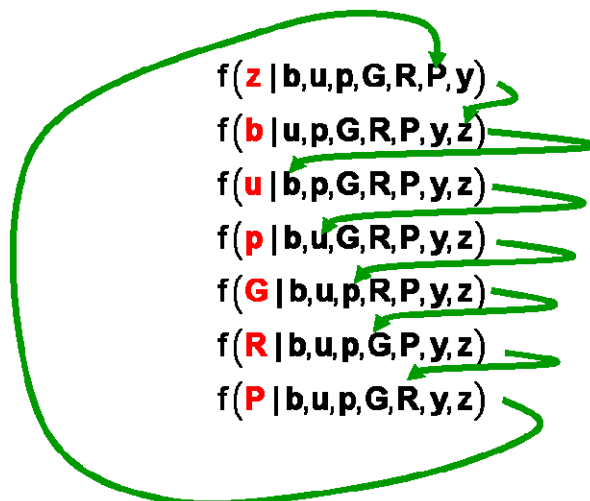
and now we can start with the Gibbs sampling because we know the distribution of  $\mathbf{y}^*$  and the conditionals (Figure 6.3). The only new conditional is the conditional of the augmented data, but the augmented data are distributed as the data, thus

$$\mathbf{z} | \mathbf{b}, \mathbf{u}, \mathbf{p}, \mathbf{R} \sim N(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p}, \mathbf{I}\sigma_2^2)$$

in each case, the data are sampled from the corresponding distribution

$$\mathbf{z}_1 | \mathbf{b}_1, \mathbf{u}_1, \mathbf{p}_1, \sigma_1^2 \sim N(\mathbf{X}\mathbf{b}_1 + \mathbf{Z}\mathbf{u}_1 + \mathbf{W}\mathbf{p}_1, \mathbf{I}\sigma_1^2)$$

$$\mathbf{z}_2 | \mathbf{b}_2, \mathbf{u}_2, \mathbf{p}_2, \sigma_2^2 \sim N(\mathbf{X}\mathbf{b}_2 + \mathbf{Z}\mathbf{u}_2 + \mathbf{W}\mathbf{p}_2, \mathbf{I}\sigma_2^2)$$



**Figure 6.3.** Gibbs sampling process for the components of the multivariate model with augmented data

At the end of the process we will have chains for all unknowns and for the augmented data  $\mathbf{z}$ . We will ignore the augmented data and we will use the chains of the unknowns for inferences. This is a legitimate procedure, because we have sampled the distribution  $f(\boldsymbol{\theta}, \mathbf{z} | \mathbf{y})$ , which depends on the data  $\mathbf{y}$  but not on the augmented data.

In practice, it is more convenient to do a similar process augmenting residuals instead of data. See Sorensen and Gianola (2004) for details.