Seeds: Random effect logistic regression

This example is taken from Table 3 of Crowder (1978), and concerns the proportion of seeds that germinated on each of 21 plates arranged according to a 2 by 2 factorial layout by seed and type of root extract. The data are shown below, where r_i and n_i are the number of germinated and the total number of seeds on the i th plate, i =1,...,N. These data are also analysed by, for example, Breslow: and Clayton (1993).

seed O. aegyptiaco 75						*seed O. aegyptiaco 73*					
Bean			Cucumber			Bean			Cucumber		
r	n	r/n	r	n	r/n	r	n	r/n	r	n	r/n
10	39	0.26	5	6	0.83	8	16	0.50	3	12	0.25
23	62	0.37	53	74	0.72	10	30	0.33	22	41	0.54
23	81	0.28	55	72	0.76	8	28	0.29	15	30	0.50
26	51	0.51	32	51	0.63	23	45	0.51	32	51	0.63
17	39	0.44	46	79	0.58	0	4	0.00	3	7	0.43
			10	13	0.77						

The model is essentially a random effects logistic, allowing for over-dispersion. If p_i is the probability of germination on the i th plate, we assume

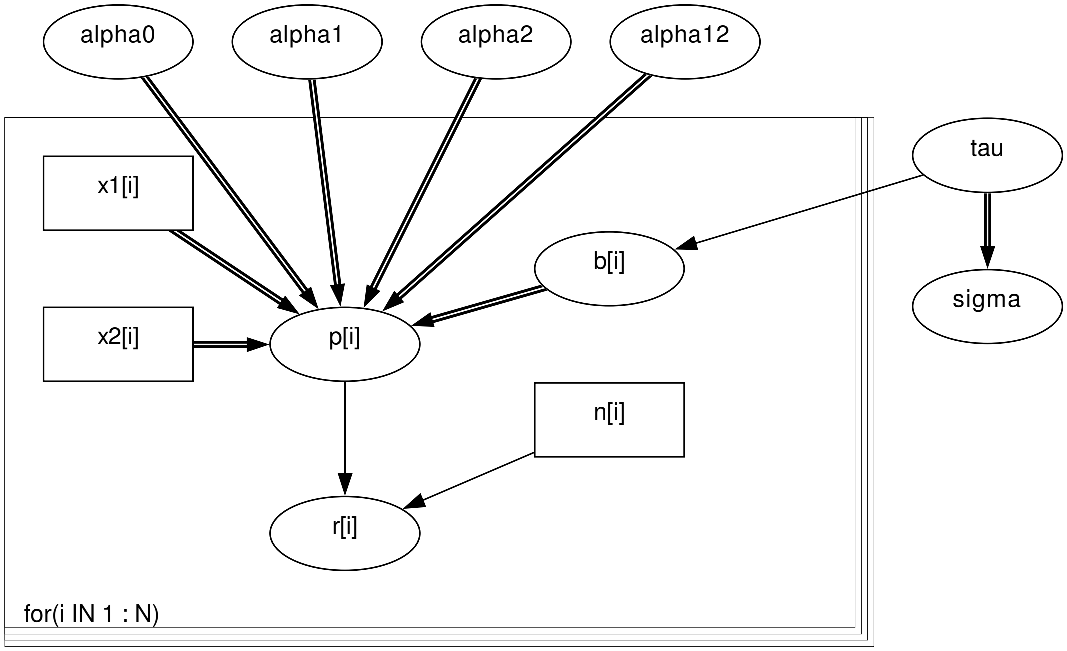
$$r_i \sim \text{Binomial}(p_i, n_i)$$

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{12} x_{1i} x_{2i} + b_i$$

$$b_i \sim \text{Normal}(0, \tau)$$

where x_{1i} , x_{2i} are the seed type and root extract of the i th plate, and an interaction term $\alpha_{12} x_{1i} x_{2i}$ is included. α_0 , α_1 , α_2 , α_{12} , τ are given independent "noninformative" priors.

Graphical model for seeds example



BUGS *language for seeds example*

```
model
{
   for( i in 1 : N ) {
      r[i] ~ dbin(p[i],n[i])
      b[i] ~ dnorm(0.0,tau)
      logit(p[i]) <- alpha0 + alpha1 * x1[i] + alpha2 * x2[i] +
         alpha12 * x1[i] * x2[i] + b[i]
   }
   alpha0 ~ dnorm(0.0,1.0E-6)
   alpha1 ~ dnorm(0.0,1.0E-6)
   alpha2 ~ dnorm(0.0,1.0E-6)
   alpha12 ~ dnorm(0.0,1.0E-6)
   tau ~ dgamma(0.001,0.001)
   sigma <- 1 / sqrt(tau)
}
```

Data (click to open)

Inits (click to open)

Results

A burn in of 1000 updates followed by a further 10000 updates gave the following parameter estimates:

	mean	**sd**	**MC_error**	**val2.5pc**	**median**	**val97.5pc**	**start**	**sample**
alpha0	-0.5525	0.1852	0.00402	-0.9312	-0.5505	-0.1879	1001	10000
alpha1	0.08382	0.3031	0.005803	-0.5238	0.09076	0.6794	1001	10000
alpha12	-0.8165	0.4109	0.008128	-1.671	-0.8073	-0.0287	1001	10000
alpha2	1.346	0.2564	0.00553	0.8501	1.34	1.881	1001	10000
sigma	0.267	0.1471	0.007996	0.03842	0.2552	0.5929	1001	10000

We may compare simple logistic, maximum likelihood (from EGRET), penalized quasi-likelihood (PQL) Breslow and Clayton (1993) with the *BUGS* results

variable	Logistic regression β	SE	maximum likelihood β	SE	PQL β	SE
α_0	-0.558	0.126	-0.546	0.167	-0.542	0.190
α_1	0.146	0.223	0.097	0.278	0.77	0.308
α_2	1.318	0.177	1.337	0.237	1.339	0.270
α_{12}	-0.778	0.306	-0.811	0.385	-0.825	0.430
σ	---	---	0.236	0.110	0.313	0.121

Heirarchical centering is an interesting reformulation of random effects models. Introduce the variables

$$\mu_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{12} x_{1i} x_{2i}$$

$$\beta_i = \mu_i + b_i$$

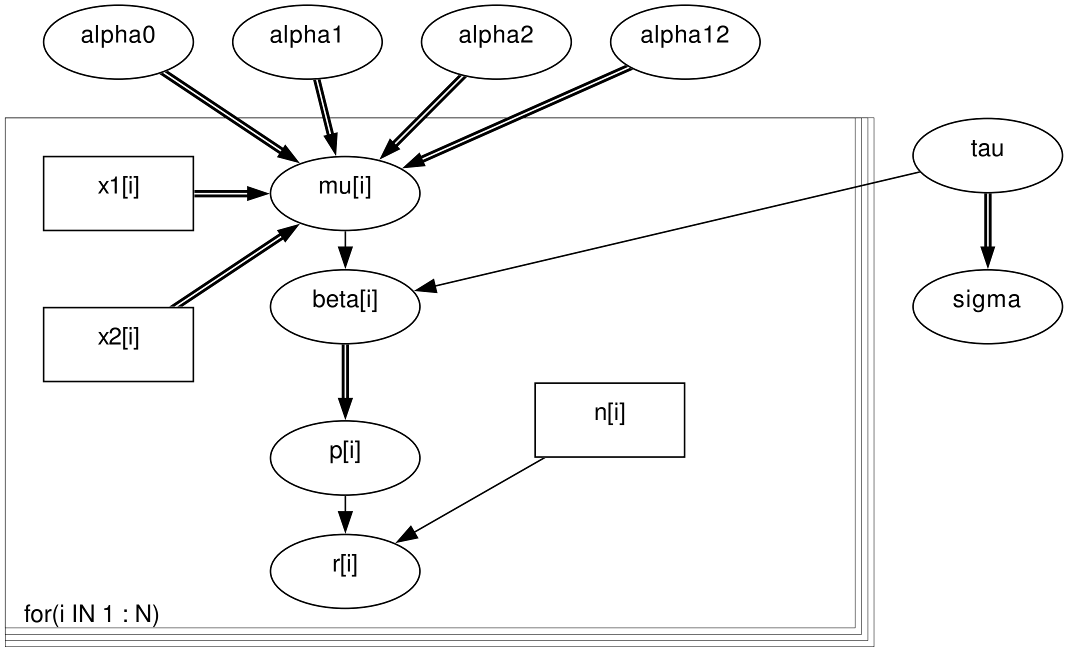
the model then becomes

$$r_i \sim \text{Binomial}(p_i, n_i)$$

$$\text{logit}(p_i) = \beta_i$$

$$\beta_i \sim \text{Normal}(\mu_i , \tau)$$

The graphical model is shown below



This formulation of the model has two advantages: the squence of random numbers generated by the Gibbs sampler has better correlation properties and the time per update is reduced because the updating for the α parameters is now conjugate.