## **Comparison of ADMB-RE**

## and h-GLM fits of a binomial-beta model to the seed germination dataset

Cóilín Minto (Coilin.Minto@gmit.ie)

November 21, 2012

Lee and Nelder (1996) analyze seed germination response from a  $2 \times 2$  factorial experiment with two seed types (073, 075) and two root extracts (Bean, Cucumber). The data consist of 831 observations over 21 plates. The .dat file is:

A binomial GLM with a linear predictor containing the main effects of seed and extract and their interaction returns a residual deviance of 33.28 on 17 degrees of freedom, indicating that the data are over-dispersed relative to a binomial. Lee and Nelder (1996) suggest modelling the extra-binomial variation associated with plate using a binomial-beta model; a beta-distributed random effect. Let  $Y_i$  denote the number of successes per plate, the model used was

$$Y_i|u_i \sim \text{Binomial}(n_i, p_i),$$
 (1)

$$u_i \sim \text{Beta}(\alpha, \alpha),$$
 (2)

where  $u_i$  are beta-distributed random effects. The symmetry constraint on the beta (E( $u_i$ ) = 1/2) resulted from earlier fits by Lee and Nelder (1996) with both shape parameters free. The linear predictor was given by

$$\eta_i = \log(p_i/(1-p_i)) = \beta_0 + \beta_{x_1} + \beta_{x_2} + \beta_{x_1x_2} + v_i$$
(3)

where

$$v_i = \log(u_i/(1 - u_i)).$$
 (4)

Lee and Nelder (1996) used *h*-likelihood to fit this model. Their results will be compared to those of ADMB-RE (Skaug and Fournier, 2006), which uses the probability integral transform to implement

beta-distributed random effects.

The .tpl file I wrote was (note the line break in the arguments for the SEPARABLE\_FUNCTION is for display purposes only, this must be an unbroken line in the .tpl file)

```
DATA SECTION
 init_int nplates;
  init_vector r(1, nplates);
 init_vector n(1, nplates);
  init_vector x1(1,nplates);
  init_vector x2(1,nplates);
PARAMETER SECTION
  init number logalpha
 init_vector beta(1,4);
 random_effects_vector b(1,nplates);
 vector eta(1,nplates);
 vector p(1,nplates);
 sdreport_vector u(1,nplates);
 vector v(1,nplates);
  objective_function_value nll;
PROCEDURE_SECTION
  for(int i=1; i<=nplates; ++i) {</pre>
   seeds_betabin(i,b(i),logalpha,beta);
  // Assign fitted beta-distributed random effects
  // post-estimation
  if (sd phase) {
    dvariable alpha=mfexp(logalpha);
    for(int i=1; i<=nplates; ++i){</pre>
      u(i)=beta_deviate(alpha,alpha,b(i),0.0000001);
  }
  // Note no line break in .tpl file for next line
SEPARABLE_FUNCTION void seeds_betabin(int& i, const dvariable& b_i,
const dvariable& logalpha, const dvar_vector& beta)
  dvariable alpha=mfexp(logalpha);
  // Standard normal random effects (RE) part of likelihood
 nll+=0.5*(log(2.0*M_PI)+square(b_i));
  // Transform to beta-distributed RE via probability integral transform
 dvariable u_i= beta_deviate(alpha,alpha,b_i,0.0000001);
 // Beta-distributed RE on linear predictor scale
  dvariable v_i = \log(u_i/(1.0-u_i));
  // Linear predictor
  dvariable eta_i=beta(1)+beta(2) *x1(i)+beta(3) *x2(i)+beta(4) *x1(i) *x2(i)+v_i;
 // Binomial probabilities on logit scale
 dvariable p_i=1.0/(1.0+mfexp(-eta_i));
  // Data part of the likelihood
  nll = log\_comb(n(i), r(i)) + r(i) * log(p_i) + (n(i) - r(i)) * log(1.0 - p_i);
```

Note that owing to changes made to the source code for ADMB-11, the above code above will not compile for that release version. The latest source code has reverted to the original beta\_deviate code and should run fine on subsequent versions. The latest source code installation instructions can be found here.

Both the default Laplace and Gauss-Hermite quadrature integral approximations were used in ADMB-RE. Gauss-Hermite was implemented by supplying the argument -gh 20 when running the executable. This uses 20 quadrature points, which should be sufficiently accurate (Lesaffre and Spiessens, 2001).

Table 1 displays a comparison of the estimates from the ADMB-RE fits with those of Lee and Nelder (1996) and a WinBUGS fit using diffuse normal ( $N(0, 10^{-6})$ ) priors for the fixed effects and a gamma prior (Gamma(0.001,0.001)) on the shape parameter of the beta-distributed random effects.

Table 1: Comparison of estimated fixed effects and shape parameter  $\alpha$  for a binomial-beta model fit to the seeds data.

			ADMB-RE					
	Lee and Nelder (1996)		Laplace		Gauss-Hermite (20)		WinBUGS	
Variable	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Constant	-0.543	0.187	-0.549	0.166	-0.549	0.166	-0.556	0.193
Seed (073)	0.080	0.303	0.0973	0.278	0.0973	0.278	0.087	0.311
Extract	1.337	0.265	1.337	0.237	1.337	0.237	1.356	0.277
(Cucumber)								
Interaction	-0.882	0.423	-0.810	0.385	-0.810	0.385	-0.829	0.437
$\log \alpha$	3.096	0.431	3.596	0.921	3.597	0.921	3.474	1.132

The Laplace and Gauss-Hermite (20) ADMB-RE estimations produced estimates that were the same to 3 decimal places (same to 2 for shape parameter), minor differences existed beyond this. The scaled deviance (Lee and Nelder, 1996) decreased for this model to 16.4 on 16 marginal degrees of freedom. Note, however, that a straightforward normally-distributed random effect on the linear predictor scale fits these data equally as well with a scaled deviance of 16.45. The marginal likelihoods using Gauss Hermite quadrature were practically equivalent. Parameter estimates and approximate 95% confidence intervals are shown in Figure 1. The results are comparable across the methods. The random effects from ADMB-RE and WinBUGS are also comparable (Figure 2), although there is slightly more shrinkage in the ADMB-RE estimates given the greater estimated shape parameter (Figure 1).

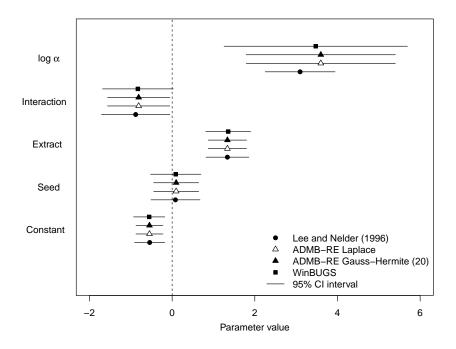


Figure 1: Comparison of parameter estimates for the seeds dataset by method. Note 95% confidence/credible intervals are assumed normal  $(\hat{\theta} \pm 1.96 \text{SE}(\hat{\theta}))$ .

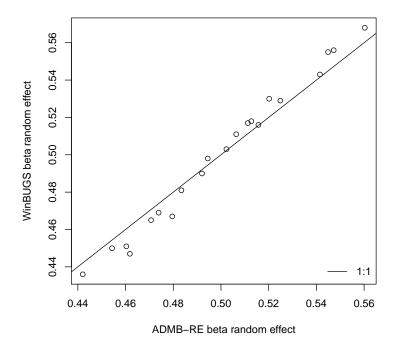


Figure 2: Beta-distributed plate random effects as estimated in ADMB-RE using Gauss-Hermite approximation and in WinBUGS.

## Literature cited

- Lee, Y. and Nelder, J. A. 1996. Hierarchical generalized linear models. *Journal of the Royal Statistical Society B* **58**(4): 619–678
- Lesaffre, E. and Spiessens, B. 2001. On the effect of the number of quadrature points in a logistic random-effects model: an example. *Journal of the Royal Statistical Society. Series C, Applied statistics* **50**(3): 325–335
- Skaug, H. J. and Fournier, D. A. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics & Data Analysis* **51**(2): 699–709