Merrill Rudd

FSH 507 – Spatio-temporal models

April 13, 2016

Homework 2 – Generalized linear mixed models in Template Model Builder

To estimate the predicted counts, I created four estimation models with different assumptions about site effects and overdispersion. The models took the following four forms:

1. 
2. 
3. 
4. 

where  is the predicted counts for observation *i,*  is the log-mean for expected counts,  is the random effect representing among-site variability, where *si* is the site of observation *i*, and  is the random effect representing overdispersion of observations.

Each estimation model assumed a Poisson distribution to calculate the likelihood of the data given the specified model.

From 100 replicates of generated data, the mean estimate of  becomes closer to the truth when the estimation model accounts for an overdispersion effect (Table 1; Figure 1). The proportion of replicates where the true value is within the estimated confidence interval also improves drastically when the overdispersion effect is included in the model (Table 1; Figure 1).

When the model is mis-specified, confidence interval coverage falls below 100%. According to the definition of a 95% confidence interval, 95% of replicated experiments will have intervals that contain the true value of the parameter. Therefore, when the model is correctly specified (accounting for among-site variability and overdispersion), the interval coverage (100%) is actually higher than the theoretical expectation (95%). The same is true when only the overdispersion effect is accounted for in the estimation model (ignoring the among-site variability), where the interval coverage is 99%. However, when only the among-site variability is accounted for, ignoring the effect of overdispersion, the interval coverage falls far below 95% (down to 13%). Without considering either of these effects, the standard errors on the estimates of are very small, and none of the 95% confidence intervals for the 100 generated datasets cover the true value of.

The main reason I would expect there to be lower confidence interval coverage when the model is mis-specified is because the observed counts are arising from natural processes (site variability and patterns of dispersion within the sites) that are not accounted for. The calculations of standard error are wrong when the model is mis-specified. This would explain why the standard error is estimated to be much different between the fixed effects model (average SE = 0.027 across 100 iterations) and the correctly specified full model (average SE = 0.197). The lower and inaccurate estimate of standard error results in very poor interval coverage when the model is misspecified. The coverage improves as more of these “true” processes are accounted for in the estimation model.

Some of this lack of interval coverage could be due to the bias in parameter estimates. The estimates of  are biased high for all models. Estimates from mixed-effects models can be biased if the random effects are not exchangeable, but because this is simulated data, it is not the case in this example (i.e. we know that the random variables are independent and identically distributed, conditional on the underlying distribution from which we generated them).

The most bias exists in estimates of  when overdispersion of counts is ignored. These models are more “wrong” than the models that account for overdispersion. Accounting for these processes as random effects allows for the flexibility of true heterogeneity in expected counts between sites.

Table 1. Mean estimate of the log-mean of expected counts,, and the proportion of replicates where the true value of  is within the estimated 95% confidence intervals (interval coverage), from 100 generated datasets for each of the four estimation models.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | | Mean estimate μ | Interval coverage |
| No site or overdisperson effects | 2.584 | | 0.00 |
| Among-site variability | 2.417 | | 0.13 |
| Overdispersion | 2.112 | | 0.99 |
| Among-site variability and overdispersion | 2.117 | | 1.00 |

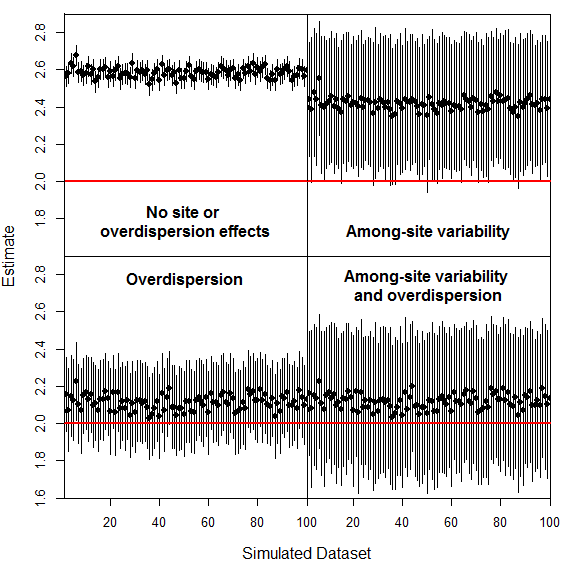


Figure 1. Estimates of the log-mean of expected counts,(solid points) with their 95% confidence intervals (black lines) for each 100 replicates of generated data and each of the four estimation models. The red line indicates the true value of.