

R^2 s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs

Anthony R. Ives, Department of Integrative Biology, UW-Madison, Madison, WI 53706

arives@wisc.edu

rr2 package: <https://github.com/arives/rr2>

Appendix 1: Details about the implementation of the R^2 s in the R package rr2

This appendix contains the technical details for the implementations of R^2_{ls} , R^2_{ce} , and R^2_{lr} in R. The implementations for LMMs, PGLS, GLMMs, and PLOG are performed for R classes `merMod` {lme4} (Bates *et al.*), `phylolm` {phylolm} (Ho & Ane 2014), `glmerMod` {lme4} (Bates *et al.*), `binaryPGLMM` {ape} (Paradis & Paradis 2012), and `phyloglm` {phylolm} (Ho & Ane 2014).

R^2_{ls} : LMM (class `lmerMod`)

Equation (3) for mSE equals the residual variance of the fitted model, from which the R^2_{ls} is computed using equation (2).

R^2_{ls} : PGLS (class `phylolm`)

Equation (3) for mSE equals the residual variance of the fitted model multiplied by a scaling factor c that equals the sum of branch lengths on the phylogenetic tree. Thus,

$$R^2_{ls} = 1 - \frac{c_f \sigma_f^2}{c_r \sigma_r^2}$$

For reduced models without phylogenetic covariances, the reduced model is calculated from linear regression; in this case, $c = n$, and σ^2 is the maximum likelihood estimate of the variance (with $df = 0$).

R^2_{ls} : GLMM (class glmerMod)

There are two approaches for calculating R^2_{ls} for GLMM and PLOG. For the first (not implemented in the rr2 package), equation (3) for mSE is given by

$$\text{mSE}(\hat{\theta}) = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\hat{\beta})' \mathbf{A}(\hat{\mu})^{-1/2} \mathbf{V}(\hat{\theta})^{-1} \mathbf{A}(\hat{\mu})^{-1/2} (\mathbf{Y} - \mathbf{X}\hat{\beta})$$

where $\hat{\mu} = g^{-1}(\mathbf{X}\hat{\beta})$ are the fitted values of μ , and $\mathbf{V}(\hat{\theta}) = \hat{\sigma}^2 \mathbf{\Sigma}(\hat{\theta}) + \mathbf{I}$ is estimated when fitting the model (Schall 1991; Breslow & Clayton 1993; Ives & Helmus 2011). The matrix \mathbf{A} has diagonal elements given by the variance function $v(\mu)$ for the link function $g()$; for example, a binomial model with a logit link function will have $v(\mu) = \mu(1-\mu)$. This expression for mSE can be interpreted as the variance for the distribution of e in equation (1), with the residuals transformed to be on the level of the link function in which the variation is Gaussian. When $\hat{\sigma}^2 = 0$, the model reduces to a GLM, and this mSE becomes the quasi-likelihood score function that can be used to estimate GLM parameters (McCullagh & Nelder 1989). This approach is not implemented in rr2 because for some datasets, the estimates of μ may be very close to zero or one, and therefore the value of R^2_{ls} will be dominated by these points; this leads to higher variation in the estimates of R^2_{ls} and occasional large, negative values.

The approach implemented for R^2_{ls} for GLMM and PLOG in rr2 follows the approach of Nakagawa & Schielzeth (2013) for R^2_{glmm} . This involves replacing mSE in equation (2) with

$\frac{\sigma_d^2}{\hat{\sigma}_f^2 + \hat{\sigma}_l^2 + \sigma_d^2}$. The estimate of the "residual variance" is the distribution-specific constant $\sigma_d^2 =$

$\pi^2/3$ for logistic models. This variance must be scaled by the variance in the data on the level of the link function in which the variation is Gaussian. The variance explained by the regression coefficients, $\hat{\sigma}_f^2$, is computed as $\text{var}(\mathbf{X}\boldsymbol{\beta})$ where \mathbf{X} is the matrix containing the predictor variables (including ones for the intercept) and $\boldsymbol{\beta}$ is the vector of regression coefficients. For GLMMs, the variance of the random effects is $\hat{\sigma}_l^2$.

R^2_{ls} : PLOG (class binaryPGLMM)

The implementation is the same as for GLMMs. For PLOG, the variance associated with the phylogeny, $\hat{\sigma}_l^2$, is computed as the product of the residual variance $\hat{\sigma}^2$ multiplied by the diagonal elements of the covariance matrix produced from the phylogeny. For binary models, this covariance matrix should be standardized so that all diagonal elements are the same (a contemporaneous or ultrametric phylogenetic tree) (Ives & Garland 2014). In case this is not done, however, the code takes the geometric average of the diagonal elements.

R^2_{ce} : LMM (class lmerMod)

In equation (3), mSE is replaced by the variance between observed and fitted values provided by lmer.

R^2_{ce} : PGLS (class phylolm)

In equation (3), mSE is replaced by the variance between observed and fitted values, where fitted values are computed from equation (4) using the covariance matrix extracted from the fitted phylolm model.

R^2_{ce} : GLMM (class glmerMod)

In equation (3), mSE is replaced by the variance between observed and fitted values provided by lmer.

R^2_{ce} : PLOG (class binaryPGLMM)

In equation (3), mSE is replaced by the variance between observed and fitted values provided by lmer. The fitted values are the values of mu provided in the model output. Note that phyloglm does not provide fitted values on the scale of a link function.

R^2_{lr} : LMM (class lmerMod)

Equation (6) is used. R^2_{lr} is also computed in the MuMIn package of R (Barton 2016).

R^2_{lr} : PGLS (class phylolm)

Equation (6) is used.

R^2_{lr} : GLMM (class glmerMod)

Equation (6) is used. R^2_{lr} is also computed in the MuMIn package of R (Barton 2016).

R^2_{lr} : PLOG (class phyloglm)

Equation (6) is used. Not that binaryPGLMM does not provide a logLik.

References

- Barton, K. (2016) MuMIn: Multi-model inference.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014) lme4: Linear mixed-effects models using Eigen and S4.
- Breslow, N.E. & Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.
- Ho, L.S.T. & Ane, C. (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, **63**, 397-408.
- Ives, A.R. & Garland, T., Jr. (2014) Phylogenetic regression for binary dependent variables. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (ed. L.Z. Garamszegi), pp. 231-261. Springer-Verlag, Berlin Heidelberg.
- Ives, A.R. & Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, **81**, 511-525.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*, 2 edn. Chapman and Hall, London.
- Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133-142.
- Paradis, E. & Paradis, E. (2012) *Analysis of Phylogenetics and Evolution with R Introduction*.
- Schall, R. (1991) Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719-727.