Queen Mary University of London

dosreislab.github.io

Comparison of relaxed-clock models using the Bayesian model selection implemented in MCMCtree

Sandra Álvarez-Carretero and Mario dos Reis

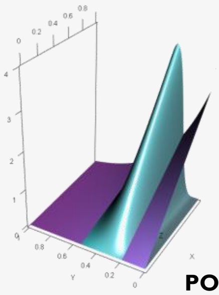
School of Biological and Chemical Sciences. Queen Mary University of London, Mile End Road, London, E1 4NS, UK.

BACKGROUND

MARGINAL LIKELIHOOD ESTIMATION

$$\textit{posterior} = \frac{\textit{prior} \times \textit{likelihood}}{(\textit{marginal likelihood})}$$

$$f(t,r|\mathrm{D}) = \frac{1}{z} f(r|t) f(t) f(\mathrm{D}|t,r)$$



$$? \quad z = \iint f(t) f(r|t) f(\mathrm{D}|t,r)\ drdt$$

MCMC SAMPLING FROM POWER POSTERIOR DISTRIBUTIONS

$$f_\beta(t,r|\mathrm{D}) \propto f(r|t) f(t) f(\mathrm{D}|t,r)^\beta$$

a) Select k β values

b) $0 \leq \beta \leq 1$

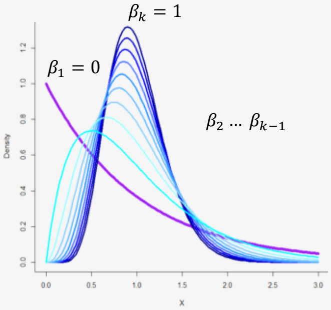
c) Sample $f(\mathrm{D}|t,r)^\beta$

d) Estimate z

- TI[1,2,3] (Thermodynamic integration)
- SS[4] (Steppingstone)

e) Calculate Bayes factors (BFs)

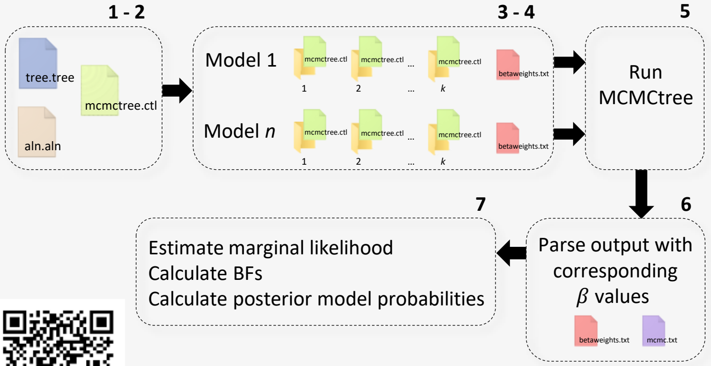
f) Select model



IMPLEMENTATION

mcmc3r[5] (R package) & MCMCtree (PAML[6,7], current v. 4.9h)

1. Prepare **sequence alignment** (`aln.aln`) and **tree** (`tree.tree`) files in `MCMCtree` format for each model to evaluate.
2. Prepare `MCMCtree` **control files** (`mcmctree.ctl`).
3. R Find appropriate ***k β* values** (`betaweights.txt`) to calculate the marginal likelihood using **TI** or **SS** approach.
4. R Create ***k*** **directories** with the corresponding `mcmctree.ctl` file with the appropriate β value appended.
5. Run `MCMCtree` within **each directory created by `mcmc3r`** having the `mcmctree.ctl` with the appropriate β value.
6. R Parse `MCMCtree`'s output to calculate the **marginal likelihood estimates** and **standard errors**.
7. R Compute **BFs** and **posterior model probabilities**.



Estimate marginal likelihood
Calculate BFs
Calculate posterior model probabilities

Parse output with corresponding β values

Complete tutorial with examples here !!
https://dosreislab.github.io/2017/10/24/marginal-likelihood-mcmc3r.html

RESULTS OF A BAYESIAN MODEL SELECTION ANALYSIS

Table 1. Bayesian selection of relaxed-clock model for 10 randomly sampled mammal genes [8]. Three different methods to calculate the Bayes factors and posterior probabilities are used.

	Thermodynamic integration			Steppingstone			Harmonic mean (bad method)		
Gene[(1)]	$P(M_{ILN}\|D)$	$P(M_{GBM}\|D)$	$BF_{ILN,GBM}$	$P(M_{ILN}\|D)$	$P(M_{GBM}\|D)$	$BF_{ILN,GBM}$	$P(M_{ILN}\|D)$	$P(M_{GBM}\|D)$	$BF_{ILN,GBM}$
cds_ENSG00000178691 (SUZ12, *q1g1*)	**1.000**	0.000	25.034	**1.000**	0.000	25.261	**1.000**	0.000	11.678
cds_ENSG00000164066 (INTU, *q1g2*)	0.000	**1.000**	-20.389	0.000	**1.000**	-20.393	0.474	**0.526**	-0.104
cds_ENSG00000108296 (CWC25, *q2g1*)	0.000	**1.000**	-10.999	0.000	**1.000**	10.985	**1.000**	0.000	8.567
cds_ENSG00000164169 (PRMT9, *q2g2*)	**0.883**	0.117	2.022	**0.884**	0.116	2.031	**1.000**	0.000	7.680
cds_ENSG00000182504 (CEP97, *q3g1*)	**1.000**	0.000	11.223	**1.000**	0.000	10.750	**0.998**	0.002	6.392
cds_ENSG00000188266 (HYKK, *q3g2*)	0.002	**0.998**	-6.327	0.002	**0.998**	-6.231	**0.999**	0.001	6.762
cds_ENSG00000164099 (PRSS12, *q4g1*)	0.000	**1.000**	-13.644	0.000	**1.000**	-13.924	**0.975**	0.025	3.672
cds_ENSG00000094963 (FMO2, *q4g2*)	**0.864**	0.136	1.850	**0.887**	0.112	2.067	**0.933**	0.067	2.630
cds_ENSG00000170456 (DENND5B, *q5g1*)	**1.000**	0.000	11.879	**1.000**	0.000	11.775	**1.000**	0.000	9.107
cds_ENSG00000196943 (NOP9, *q5g2*)	0.000	**1.000**	-8.308	0.000	**1.000**	-7.805	**0.995**	0.005	5.322

[(1)] Ensembl identifiers for the genes studied in [8]. Next to them, we have added the gene name and a tag, which indicates the order of the gene sampled and the quintile from where it was sampled. g*X*: gene number X; qY: quantile Y.
- Note: **ILN**: Independent-rates model under log-normal distribution, **GBM**: Autocorrelated-rates model under geometric Brownian motion. The R package `psych` was used to calculate the harmonic mean estimator.

REFERENCES

1. Ogata, Y. (1989) Numer. Math. 55:137
2. Gelman, A. & Meng, X. L. (1998) Stat. Sci. 13(2):163
3. Lartillot, N. & Philippe, H. (2006) Syst. Biol. 55(2):195
4. Xie, W. et al. (2011) Syst. Biol. 60(2):150
5. dos Reis et al. (2018) Syst. Biol. 67(4):594
6. Yang, Z. (2007) Mol. Biol. Evol. 24(8):1586
7. Rannala, B. & Yang, Z. (2017) Syst. Biol. 66(5):823
8. Liu, L. et al. (2017) Proc. Natl. Acad. Sci. U.S.A 114(35):E7282

ACKNOWLEDGEMENTS

We are grateful to the comments and feedback received during the lab meetings at QMUL and UCL.

QUESTIONS? Find us at https://dosreislab.github.io

E-mail: s.alvarez-carretero@qmul.ac.uk