BIOINF702 Semester 1, 2019
Sebastian Dunn, sdun067, 4610092

Assignment 4 Results and Discussion

Birth-Death Model Results:

Table 1: Birth-Death model statistics and maximum likelihood parameter estimates.

Model	Free Params	Sample Size	Lambda	Mu	lnL
Birth-Death	2	75	0.256	0.1	58.22
Yule	1	75	0.202	0	57.68

Table 2: Birth-Death model fit statistics.

Model	AIC	AIC Weight	AICc	AICc Weight	BIC	BIC Weight
Birth-Death	-112.44	0.3880	-112.27	0.3747	-107.81	0.1659
Yule	-113.35	0.6121	-113.30	0.6253	-111.04	0.8341

Table 3: Birth-Death model comparison statistics

Parent Model	Nested Model	df	LRT stat	p-value	P < 0.05
Birth-Death	Yule	1	1.088	0.297	FALSE

For my Birth-Death model comparison, I simulated a tree with 75 tips using a birth rate of 0.2 and a death rate of 0.1, using Diversitree's make.bd function. Then I optimised the likelihoods under the Birth-Death (BD) and Yule models.

Interestingly, the BD model gave a higher log likelihood, but was considered a weaker model for my data than Yule, based on the AIC, AICc and BIC. This is because it has fewer free parameters, which these measures favour. The AICc and BIC scale the contribution of the free parameters based on the sample size. Since the same dataset was used to test both models, the increasing favour of the Yule model across these scores reflects their increasing weight of the free parameters.

When the fits of the models were compared directly using the likelihood-ratio test, it was found that there wasn't a significant increase in fit of the BD over the Yule model. This reinforces what the AIC, AICc and BIC described - the higher likelihood measured under the BD model is outweighed by its extra parameter. This makes a lot of sense from an analysis

point of view, but when looked at together with the fact that I know a BD process was used to evolve the data, it becomes more interesting. It's likely that a larger dataset with a similar structure could tease out a significant difference, but for this small dataset a pure birth process is sufficient to describe it.

BiSSE Model Results:

Table 4: BiSSE model statistics and maximum likelihood parameter estimates.

Model	Free Params	Sample Size	lambda 0	lambda 1	mu0	mu1	q01	q10	lnL
BiSSE	6	75	0.364	0.250	0.239	0.249	0.023	0.036	-205.01
Yule-BiSSE	4	75	0.241	0.082	0	0	0.024	0.052	-208.89
Equal Birth	5	75	0.368	0.368	0.244	0.368	0.027	0.052	-205.28
Equal Death	5	75	0.319	0.175	0.175	0.175	0.028	0.051	-205.46
Equal Transition	5	75	0.293	0.168	0.119	0.167	0.026	0.026	-206.06

Table 5: BiSSE model fit statistics.

Model	AIC	AIC Weight	AICc	AICc Weight	BIC	BIC Weight
BiSSE	422.01	0.170	423.25	0.145	435.92	0.057
Yule-BiSSE	425.79	0.026	426.36	0.031	435.06	0.087
Equal Birth	420.55	0.352	421.42	0.361	432.14	0.375
Equal Death	420.93	0.292	421.79	0.300	432.51	0.311
Equal Transition	422.13	0.160	423.00	0.164	433.72	0.171

Table 6: BiSSE model comparison statistics.

Parent Model	Nested Model	df	LRT stat	p-value	P < 0.05
BiSSE	Yule-BiSSE	2	7.775	0.020	TRUE
BiSSE	Equal Birth	1	0.540	0.463	FALSE
BiSSE	Equal Death	1	0.911	0.340	FALSE
BiSSE	Equal Transition	1	2.113	0.146	FALSE
Equal Death	Yule-BiSSE	1	6.864	0.009	TRUE

For my BiSSE model comparison, I simulated a a phylogeny with 75 tips using Diversitree's make.bisse function, with the generating parameters lambda, mu and q of 0.25, 0.2 and 0.02

in the 0 state, and 0.2, 0.12 and 0.01 in the 1 state. You could rightly ask why I didn't simulate a larger dataset to use, and the answer is practical. Even at only 75 tips my full suite of optimisations takes some time to run, and I needed a smaller dataset than my Primates phylogeny to test with.

I decided to simulate data with 6 different rates for a couple of reasons: 1) it more accurately reflects real data, and 2) I wanted to see how the different equal-rate models behave with a dataset that was generated from no equal rates. My intuition was that with a small dataset, the fewer free parameters of an equal-rate model would likely outweigh the increased likelihood of a full BiSSE model.

It turns out this intuition was mostly correct. The BiSSE model optimised to a higher log likelihood value, as expected, but models with equal birth and equal death rates outperformed BiSSE in all of the AIC, AICc and BIC measures. The fewer degrees of freedom outweighed the higher likelihood again. All except the Yule-BiSSE model, where the poor likelihood performance couldn't make up for its few model parameters.

An interesting case was seen with the equal transition rate model - it has a lower likelihood than the equal birth and equal death models, so much so that BiSSE still outperforms it based on AIC. But with AICc and BIC penalising BiSSE's fewer degrees of freedom more harshly, the equal transition rate model shows a better fit in those scores.

Looking at the likelihood ratio test for these models, it reinforces the trend I saw with the Birth-Death models - an extra free parameter can override a somewhat better likelihood when calculating model fits. The likelihood of the BiSSE model was not sufficient to keep it statistically superior to the equal birth, equal death, or equal transition rate models for this dataset. The only model to show a significant difference to any of the others was the Yule model, where the poor likelihood estimation was enough to preference both the BiSSE and equal death rate models over it.

AIC, AICc, and BIC

The information criteria are estimates of the Kullback Leibler information loss, which represents the information lost between our model and the true model. They attempt to trade off between goodness-of-fit and model complexity when creating scores.

Though the AIC, AICc, and BIC are use somewhat interchangeably, they are conceptually different entities. The AIC (and its sample size adjusted form, the AICc) is trying to select the model that most accurately describes an unknown, higher dimensional reality. The BIC, on the other hand, assumes that the "true" model is within the set being tested, and tries to select for that. Practically though, the "true" model is almost never in the model set, and the values that the BIC converges on still give relative estimates of the fits of the different models.

The three measures are usually used next to each other, with the understanding that there are increasing penalties for free parameters moving from AIC to AICc to BIC. With the

weakest penalty, AIC tends to overfit. AICc should be used instead of AIC in almost every case (Burnham and Anderson, 2011), though the real differences between AIC and AICc seem small in practice. A person's choice of model should take these scores into account along with their own experience with the data and their expectations of the processes behind it (Koen, 2018).

For my own analysis, I would use the AICc score as the highest indicator of good model fit. The AIC seemed to prefer the BiSSE model to the equal transition rates model, disagreeing with the AICc and BIC, suggesting some overfitting. However, you could argue that the BIC is perfect in this instance, as I was testing against generated data and could guarantee that the generating model was in my set of models. Unfortunately though, this doesn't help me generate insights.

Assignment Challenges

This time, the biggest challenge I've had is interpreting and analysing the data in front of me. Modifying my existing likelihood functions to create nested models was intuitive once I realised I just needed to restrict the variables my optimisation function has access to, and the actual maths behind the AIC, AICc and BIC is not difficult. The part I struggled with was working out what the outputs were telling me about the models and the data. The different interpretations of the AIC, AICc and BIC in particular were difficult to get my head around. At least what they are meant to represent - not how to use them to rank models.

I overcame this gap in understanding by reading around and finding advice from various sources, including blog posts and stack exchange conversations on the topic. There is a lot to dig into when reading about the advantages and disadvantages of the different information criteria, and I've just touched the surface of the debate.

A Bayesian Alternative

The setup for our maximum likelihood estimation and model comparison would be quite different under a Bayesian process.

Estimating the likelihood first requires us to decide on priors for our data. In our case, we need priors for our tree process (e.g. a coalescent tree prior with constant population size) and priors to draw our model rates from. The likelihood calculation would be carried out by MCMC, where at each time step we pull a set of parameters from our rate priors and calculate a point likelihood. These rates are varied as the MCMC chain progresses, and eventually converge on a set of values with a maximised likelihood.

The outputs from our MCMC likelihood estimations are posterior distributions on our parameters of interest. These distributions and their describing parameters are what we compare in a Bayesian setting, and we are calculating Bayes Factors instead of p-values to compare the likelihoods of our data under the models.

References:

Fitzjohn, R. Diversitree documentation. https://www.rdocumentation.org/packages/diversitree/versions/0.9-11. (Accessed 12th June 2019).

Kenneth P. Burnham and David R. Anderson. (2011). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach.* (2nd ed.).

Blogs and thread I took advice from:

- Koen, 2018. http://www.koenbro.com/information-criteria-aic-aicc-bic/
- https://stats.stackexchange.com/questions/577/is-there-any-reason-to-prefer-the-aic-or-bic-over-the-other