On the phylogenetic comparative analysis of directional evolution by Approximate Bayesian Computation

（近似ベイズ計算を用いた方向性進化の分析について）

Division of Biosphere Science, Graduate School of Environmental Science,

Hokkaido University

September 2019

Yusaku Ohkubo

Acknowledgments

This PhD research was conducted under the support of many researchers.

I thank Assoc. Prof. Itsuro Koizumi, Prof. Takashi Noda, Prof. Hitoshi Suzuki, Prof. Nobuyuki Kutsukake (Graduate University for Advanced Studies), Assoc. Prof. Ryota Morimoto (Health Sciences University of Hokkaido) for many helpful comments and advice not only for this thesis but also for my academic career.

I also would like to express my greatest thanks to all the people those who helped my research life. Especially, Prof. Takenori Takada, Assoc. Prof. Ken Ichiro Shimatani, (Institute of Statistical Mathematics), Prof. Masahiro. Matsuo (Hokkaido University Graduate School of Science), Prof. Makoto Mizunami (Hokkaido University Graduate School of Life Science), Ph. D. Yuta Morii (Massey University, NZ), Ph. D. Tomohiro Harano (Aichi Gakuin University), M. Sc. Hayato Kato (Tokyo University), M. Sc. Natsuki Ogusu (Hokkaido University Graduate School of Letters), M. Sc. Misato Sato (Hokkaido University Graduate School of Life Science), M. D. Tmokatsu Kono (Hokkaido University Graduate School of Science), and Ms. Sanae Watanabe (Sapporo). The short stay of Nitamago (also known as Sasakama) in my home provided me a lot of rest. Finally, I thank all the members of Koizumi Lab. This research would not have been possible without their open-minded culture.

* () indicates affiliation and otherwise affiliate Hokkaido University Graduate School of Environmental Science.

Table of Contents

Chapter 1

General Introduction: Directed Evolution in Phylogenetic Comparison

Phylogenetic comparative methods (PCM) are statistical methods which analyze the evolution of traits in organisms and are applied mainly for two purposes (Losos 2011 for review). One major application is to infer ancestral traits and their evolutionary patterns for species or taxa, given traits of extant organisms. To know the traits of ancestral species is generally difficult since we cannot observe past organisms directly. Trait information might be obtained via fossils records, but this is very limited. By phylogenetic comparative methods, it is possible to estimate a rate of evolutionary changes of traits and its value of ancestral species by using only phenotypic data of living organisms and their molecular phylogenetic trees.

Another major application is to correct bias which is induced by non-independence of data when performing statistical analysis in inter-specific comparisons. Most of today's statistical methods use the assumption that the data are samples from a (conditional) independent and identically distributed (*i.i.d.*) random variables. But, in inter-species comparative studies, trait value tends to be similar (correlated) among closely related species because of the shared evolutional history. Without considering the phylogenetic relationships, any statistical analysis biases the estimates and raises the error probability. PCM is a powerful tool that can examine the relationship between traits and ecological/environmental factors in consideration of phylogenetic correlations.

Since the mid-1980s, when the importance of the phylogenetic information was first pointed out by Felsenstein (1985), phylogenetic comparative methods became an indispensable method in ecology and evolutional biology. Some even argue that there is "nothing in biology makes sense except in the light of phylogeny" (Losos 2011; see also, MacLeod 2001). In recent years, the popularity of the meta-analysis, which integrate the results of primary research, have also boosted the importance of phylogenetic comparative methods, resulting in the number of publications which mention the term "phylogenetic comparative" is exploding (Cooper *et al*. 2016).

Most of the phylogenetic comparative methods impose strong assumptions. The most fundamental assumption is how the traits evolve. In most cases, PCMs have assumed Brownian motion models and, in some cases, the Ornstein-Uhlenbeck model and the Early-Burst model. If these assumptions are not met, it causes a strong bias (Thomas *et al*. 2009, Losos 2011, Cooper *et al*. 2016). In real situation, trait change that does not conform to such a typical model often occurs. For example, one of the most common processes reported in the wild is a directional evolution, where a certain trait value increases (or decreases) with a constant trend. Some researchers have concerned that phylogenetic comparative methods have been used without the validity of these assumptions are not considered (for example, Losos 2011, Cooper *et al*. 2016). Oakley *et. al*. (2000), conducted an experiment with a colony of *E. coli* which we can directly observe the trajectory of evolution in the laboratory and found that phylogenetic comparative methods failed to reconstruct the trait of the ancestral colony when the colony experienced the directional evolution.

Surprisingly, despite the prevalence of directional selection in the wild, statistical methods for evaluating the directional evolution with phylogenetic comparative methods have not been fully developed. As notable exceptions, Pagel (1999) and Hunt (2006) propose a method that assumes all the species evolves in one direction with common speed. But it does not offer a way to evaluate biologically interesting cases of directional evolution; whether and how the strength of directional evolution differs among different clades or branches.

The purpose of the current thesis is to clarify the problems on the analysis of directional evolution in the current phylogenetic comparative methods and to propose a new method. In this chapter, I examine three approaches that are used in the detection of directional evolution: the outlier detection approach, accelerated Brownian Motion model approach, and ABC-PCM. I point out there exists serious philosophical problems on the outlier approach and accelerated Brownian Motion approach. Next, I examine how the ABC-PCM avoids these problems and argue it as the best alternative to analyze directional evolution if some remaining issues are solved. Finally, I show the outline of the remaining chapters.

・Brownian Motion and Stochastic Process

In the first half of this chapter, I will examine two statistical methods, outlier detection approach and accelerated Brownian motion approach for the analysis of directional evolution. Here, the nature of the stochastic process called *Brownian Motion* (or *Wiener process*) plays an important role. So, before examining problems of the two methods, I

will introduce the basic concept of stochastic processes and Brownian motion (Ito 1953; In Japanese).

Let T = [0, ∞) the continuous time. If the probability of an event at $t \in T$ follows the random variable X(t), then {X(t), $t \in T$} is called a *continuous-time stochastic process*. Especially, the continuous-time stochastic process with the following properties is called *Brownian Motion*.

1) X (0) = 0
2) X (t) is an independent incremental process
3) For any $s, t \in T$, the distribution of X ($s + t$) - X (s) is *Normal* (0, $\sigma^2 t$)

Where Normal (x, y^2) is a normal distribution with the mean x and variance y^2. Now, σ^2 is a parameter called the *diffusion coefficient*. As σ^2 becomes larger, the increment of dispersion after the unit time increases. In the context of biological evolution, σ^2 is often referred to as the evolution rate. The property of particular importance in the current thesis is the third; in Brownian Motion, a) the variance of the normal distribution increases in proportion to time difference t and b) expected value of the normal distribution is zero for any t and σ^2. Most of the phylogenetic comparative methods have so far assumed, "traits evolve in the manner of the Brownian motion model, a simple model of trait evolution where *trait variance accrues as a linear function of time*. Researchers have made statistical inferences about the evolution of traits under the assumption (cited from Cooper *et al*. 2016, emphasis added).

・Outlier Detection Approach

Organ introduced the outlier testing approach of posterior predictive distribution by Bayesian statistics to analysis directional evolution (Organ *et al.* 2011*).* Suppose the normally distributed traits of *n* species $\boldsymbol{y} = (y_1, y_2 \ldots y_n)$ and phylogenetic tree of these species are given. And suppose further that these traits have evolved in the Brownian Motion with common evolutional rate. Then, the evolution rate σ^2 can be estimated by the following regression model.

$$\boldsymbol{y} \sim MVN(\mu, \Sigma) - (1.1)$$

Where *MVN* is an *n*-dimensional multivariate normal distribution, and its variance-covariance matrix Σ is an $n \times n$ matrix whose *i, j* element is defined as follows.

$$\Sigma(i,j) = \sigma^2 * \mathrm{bl}_{ij} - (1.2)$$

Here, σ^2 is the evolution rate, which is common to all branches, bl is the branch length of the phylogenetic tree shared by species *i* and *j*. That is, equation (2) formulates the assumption that the magnitude of the correlation between species *i* and species *j* is proportional to the length of time the two species share in the history of evolution. After we have obtained posterior distribution, we can calculate the posterior predictive distribution $p(\hat{\boldsymbol{y}}|\boldsymbol{y}) = \int p(\hat{y}|\mu, \Sigma)\, p(\mu, \Sigma|\boldsymbol{y}) d\mu\, d\Sigma$ for any species *i*. Organ 's proposal was that if the *i*'s actual trait y_i can be said outlier of $p(\hat{y}_i|\boldsymbol{y})$, then this is the evidence that directional evolution has occurred for the trait of species *i*. Organ applied this approach to data on the time length taken by the apes' species to consuming food in a

day and concludes that human food digestion has evolved after diverging from chimpanzees (Organ *et al*. 2011).

・**Accelerated Brownian Motion Model Approach**

Another approach is the accelerated Brownian Motion model approach (or variable rate Brownian motion model; Baker *et al*. 2016, see also Dunbar *et al.* 2018 but Baker & Venditti 2019). This approach also assumes that all branches on the phylogenetic tree have undergone Brownian Motion evolution as similar to outlier detection approach. But, unlike it, accelerated Brownian Motion assume that the evolution rate may be different for each branch in equation (1.2). Then, the evolution rate in each branch is estimated by a Bayesian manner, and if it is twice or more the size of the other branch, it is interpreted that directional evolution has occurred in that branch.

・**Problems of the Previous Methods**

Here in this section, however, I argue that the above two methods using the Brownian motion model suffer from serious problems to interpret that the directional evolution has occurred in the trait of interest. My points are 4-folded.

The first problem is the setting of the threshold. As many researchers have discussed the role of a classical statistical testing and *p*-value (in ecology, Ohkubo & Aiba 2019; in Japanese), some statisticians and scientists argue that there is no objective basis for setting the threshold of significance level $p = 0.05$. Relatedly, some propose we should set a more severe significance level like $p = 0.005$ to reduce false positives

(e.g. Benjamin *et al.* 2018). These arguments are mainly focused on frequentist statistics.

However, the same argument also applies to the analysis of directional evolution by Bayesian phylogenetic comparative methods. In the outlier detection approach, the focal trait was considered as an outlier if the 95% confidence interval of the posterior prediction distribution does not cover the actual trait and take it as the evidence of directional evolution. But the choice of 95% is no less arbitrary than the $p = 0.05$. The accelerated Brownian model concludes that if the estimated evolution rate of a focal branch is more than twice of the other branches (baseline), then this branch has experienced directional evolution. Again, the decision depends on the choice of the threshold of the rate of evolution.

The second problem is both methods rely on an invalid inference called probabilistic Modas Tollens when it reject Brownian Motion Model. Consider a usual consequent denial (Modas Tollens, or MT), one of the rules of deductive inference.

If H is true, then O

Not O

Therefore, not H

Under some axioms, the MT have proven to be valid. However, following probabilistic MT is NOT valid, yet seemingly similar.

If H is true, then the probability of being O is very high

Not O

Therefore, not H

Sober, a defender of likelihood account criticized the significance testing because it is based on probabilistic MT (Sober 2008). Although both the outlier approach and the accelerated Brownian Motion approach are Bayesian methods, the same argument also holds since these methods infer the occurrence of a directional evolution when the Brownian Motion Model was rejected.[1]

The third problem concerns the possibility of a quantitative assessment of the strength of directional evolution. As described above, both outlier detection approach and the acceleration Brownian Motion approach assume the Brownian Motion model for the evolution of focal traits without the parameter representing the strength of the directional evolution is incorporated into the model. It does not offer a way to evaluate the strength of directed evolution and its difference among branches in a quantitative way.

Relatedly, the fourth problem is about the prediction. While an inference and a prediction are closely related task in statistics, an analysis of directional evolution via the previous Bayesian phylogenetic comparative methods lead to paradoxical

[1] It should be noted that the argument is on the form of inference, not the interpretation of the probability. It is not the problem of Frequentist vs Bayesian (and Likelihoodism), I argue.

predictions. Suppose that the average of a trait value has increased Δ for past t years, being undergone positive directional evolution and the outlier approach and/or accelerated Brownian Motion approach succeeded in the inference that directional evolution has occurred. Suppose further this is an idealized situation where the underlying mechanism and environmental factors which have driven this directional evolution past t years are identified and it is known that these factors will keep working for next t years. It is natural to predict that the change of average trait value for next t years will be Δ.

However, both the outlier approach and the accelerated Brownian approach does not predict the trait change will be Δ, but predict exactly zero regardless of the strength of the directional evolution and the length of t. This is because both approaches assume the trait evolves as Brownian Motion, where the expected trait change is always zero as described the property (3) of Brownian Motion above. So even if we admitted the interpretation "the fact that the extant species is the outlier of the posterior predictive distribution is the evidence of the directional evolution" or "the fact that the posterior evolution rate of the focal species is larger than twice of other branches is the evidence of the directional evolution" is not problematic, it causes inconsistent prediction for future evolutions.

In summing up this section, an analysis of directional evolution by the outlier detection approach and the accelerated Brownian motion approach have serious philosophical problems. The following section introduces ABC-PCM, a hopeful alternative to these methods and discuss how it avoids these problems.

Third approach: ABC-PCM

Why so far, phylogenetic comparative analyses of directional evolution have used Brownian Motion model? The major reasons are the ease of mathematical analysis. In the case of Brownian Motion, the likelihood can be easily calculated using the knowledge of stochastic process. However, with respect to directional evolution, it has been considered that it is not possible to write likelihood functions directly.

Kutsukake & Innan (2013, 2014) solved this problem by a novel method to analyze directional evolution based on approximate Bayesian computation (ABC), the widely used methods in population genetics (Tavaré *et al*. 1997, Beaumont *et al*. 2002) and acceptance-rejection sampler (AR). Here in this section, I describe an outline of their method with their application.

The basic idea of ABC is quite simple and intuitive. Let $\pi(\theta)$ be a prior distribution of a parameter of interest $\theta \in \Theta$ and $y \in Y$ is a collected data. Usual Bayesian statistics require a closed form of likelihood function $L(\theta; y)$ to obtain the posterior distribution $\pi(\theta|y)$ but for a complex model or a computationally expensive model, it is not feasible. Then, an approximate inference is possible via an approximated likelihood procedure below.

1. Draw a random sample $\theta' \sim \pi(\theta)$
2. Simulate the data generating process and obtain $y_{sim} \sim f(\theta')$
3. Calculate the discrepancy between the simulated data and the actual data $d(y_{sim}, y)$

3.1 If $d(y_{sim}, y) < \varepsilon$, then accept the θ'

3.2 Otherwise reject θ'

4. Repeat 1-3 until the accepted θ' is accumulated for the desired size

Where ε is a predetermined tolerance and $d()$ is a measure of similarity between the simulated data and the actual data. Note that, $d(y_{sim}, y)$ works as an approximation of negative likelihood. The most popular choice of a discrepancy measure is a summary statistic (e.g. mean, variance).

Even when the distribution of the data generating process cannot be exactly defined, the approximate likelihood can be obtained if the evolutionary process can be simulated. They applied ABC to the phylogenetic comparative method and named it "approximate Bayesian computation phylogenetic comparison method (ABC-PCM)". Here, I introduce a seminal example of its application to empirical study.

The evolution of the brain size in humans (*Homo sapience*) is likely to have undergone directed evolution in the past because its distribution deviates from other apes (see Fig.1, cited from Kutsukake and Innan 2013). For example, the summary statistic of the brain volume of human and 3 other ape is as follows; human (*Homo sapience*); mean = 1321, sd = 123, chimpanzee (*Pan troglodytes*); mean = 348, sd = 46, gorilla (*Gorilla gorilla)*; mean = 467, sd = 46, and orangutans (*Pongo pygmaeus*) ; mean = 334, sd = 52. They assumed that the brain size of the ape was experienced directional evolution only on human and just after human branched with the chimpanzee.

First, the topology and branch length of the phylogenetic trees were obtained in advance by molecular data. Next, based on this tree, the evolution from the common ancestors of 4 species to extant species is simulated.

Let *i* the ancestral species and *j* the descendent species. Then, the evolution from *i* to *j* is simulated by the following procedure.

A-1: Directional Evolution Algorithm of species $i \rightarrow j$

1. Determine the number of mutations by random sample from Poisson distribution.

1.a the number of the positive direction mutations is

$$\mu_{i\rightarrow j}^{+} \sim Poisson(bl * ev * k) \text{- (1.3)}$$

1.b the number of the negative direction mutations is

$$\mu_{i\rightarrow j}^{-} \sim Poisson(bl * ev/k) \text{-(1.4)}$$

Here, *bl* is the branch length, *ev* is the evolution rate, and *k* is the strength of directional evolution. Notice that it becomes a Brownian Motion model when $k = 1$.

2. For each of the positive and negative mutations , determine the amount of phenotype change Δ from the independent exponential distribution exp (θ).

$$\Delta_{m}^{+} \sim \exp(\theta) \ (m = 1,2, \ldots \mu_{i\rightarrow j}^{+}) - \ (1.5)$$

$$\Delta_{n}^{-} \sim \exp(\theta) \ (n = 1,2, \ldots \mu_{i\rightarrow j}^{-}) - (1.6)$$

Where θ is the hyper parameter which controls the trait change per mutation.

3. Add each trait change one-by-one.

$$X_{i\to j} = \sum_{i=1}^{\mu^{+}_{i\to j}} \Delta^{+}_{m} \ -(1.7)$$

$$Y_{i\to j} = \sum_{j=1}^{\mu^{-}_{i\to j}} \Delta^{-}_{n} - (1.8)$$

4. Define the final trait change by $X_{i\to j} - Y_{i\to j}$ $-(1.9)$
5. Add the ancestral trait $y_{sim\,j} = y_{sim\,i} + (X_{i\to j} - Y_{i\to j})$ $-(1.10)$

Finally, the discrepancy is calculated to evaluate how the simulated trait is close to the actual trait. Letting the simulated trait of net species j is $y_{sim\,j}$, and the actual trait $y_j = Normal(\text{mean}_j, \sigma_j{}^2)$, the discrepancy for the species j is defined as follows.

$$d(y_{sim}, y) = Normal\left(y_{sim\,j} \middle|\ \text{mean}_j, \sigma_j{}^2\right) \text{-}(1.11)$$

Where Normal (x | mu, sd^2) is the probability density of normal distribution with mean mu and variance sd^2.

Kutsukake and Innasn (2013) defined the log likelihood of the model as the product of the discrepancy for all extant species.

$$log\ L(\theta) = \sum_{j \in extant\ sp.} Normal\left(y_{sim\,j} \middle|\ \text{mean}_j, {\sigma_j}^2\right) - (1.12)$$

Kutsukake & Innan (2013, 2014) adopted the Acceptance-Rejection (AR) sampler to obtain samples from the posterior distribution (Tavaré *et al.* 1997).

Notice that the method of Kutsukake & Innan (2013, 2014) does not simulate the population of evolved species but simulate the random value from the population. Therefore, the likelihood of the species *i* is the point density at the simulated trait. The validity of this simulation will be discussed later (Chapter 3 and Chapter 4).

One of the major advantages of ABC-PCM is that it does not suffer from the serious philosophical problems mentioned in the previous section. Unlike the outlier detection approach and accelerated Brownian Motion approach, ABC-PCM avoids the arbitrary of threshold choice, the inferential fallacy of probabilistic MT, and it is possible to evaluate the strength of evolution quantitatively since ABC-PCM is not restricted to the Brownian Motion model and offers a direct way to quantitative evaluation parameters. Also, there is no conflict between the posterior distribution and the prediction because, again, simulatable evolution is not restricted to the Brownian Motion model.

Another major advantage of ABC-PCM is its flexibility because various evolutionary models can be incorporated into ABC-PCM as long as its process can be simulated. Garamszegi (2014) reviewed the existing phylogenetic comparative methods and introduced ABC-PCM as a highly flexible approach.

However, despite these philosophical superiorities, ABC-PCM is hardly applied into empirical biology studies. At present (June 2019), six years passed since ABC-PCM was first proposed in Kustukake & Innan (2013), the paper has been cited 24 times. But, of these, the applications to the empirical study are only two: Imai, Ohkubo, *et al*. (2016) and Harano & Kutsukake (2018) (Google Scholar, May 3, 2019)

・Conclusion and the Direction of this Thesis

In this chapter, I have reconsidered methods to analyze directional evolution by phylogenetic comparative methods, with some philosophical arguments. I conclude that the outlier detection approach and the accelerated Brownian motion approach has serious problems in statistical inferences. While simulation-based ABC-PCM avoid them and is a potential alternative, current ABC-PCM is not widely applied to empirical study for both pragmatic and theoretical reasons.

The following chapters are organized as follows. In Chapter 2, I discussed the huge computational cost of ABC-PCM, especially of the simulation of evolution and why it matters in both practical and theoretical reason. I also analyzed the previous procedure of simulation and developed a 1,000-times faster but asymptotically equivalent procedure. In Chapter 3, I conducted simulation experiments to investigate the properties of parameter estimation and hypothesis testing in the existing ABC-PCM. I found that the parameter estimation by ABC-PCM results in an overestimation of the strength of directional evolution, and that the existing hypothesis testing has a low power. In chapter 4, I propose a new simulation method of evolution and define new

likelihood for ABC-PCM to correct the bias and improve the power. I also conducted data analysis and simulation experiment to show its superiority to the previous method. In Chapter 5, I applied the proposed method to meta-analysis of temporal discounting behavior of ape species. Finally, in Chapter 6, I discussed the challenges of ABC-PCM and the possibility of further refinement.

・**Notes on Methodologies**

Before the end of this chapter, I will briefly clarify my philosophical position in this study. In the history of statistics and its philosophy, there have been much of controversies between frequentists, likelihoodists, and Bayesians. From frequentists and likelihoodists, Bayesian statistics have long been criticized on its subjective choice of the prior distribution and the interpretation of results (Mayo 1996, Royall 1997, Sober 2008, *etc*.).

However, in recent years, attempts to justify Bayesian methodologies from a frequentists point of view have attracted attention. For example, Horseshoe prior distribution (Carvalho *et al*. 2010) was first proposed in the context of sparse statistics. But later, it was turned out that it also has excellent properties in frequentist sense (Castillo *et al.* 2015) and applied to ecology (Morii, Ohkubo, Watanabe 2018). Gelman *et al.* (2015) discussed the use of Bayesian statistics while pointing out that modern Bayesian statistics cannot be understood as "the rule of the degree of belief and its update" such as existing philosophers of science have discussed. In the philosophy of statistics, D. Mayo had been the most influential critic of Bayesian statistics but have recently deployed novel arguments to justify the Bayesian statistics from her severe

testing perspective (Mayo 2018). Ohkubo (2019) pointed out the similarity of the structure of argument between “stopping rule problem” developed in the context of criticism against frequentist and the “catchall hypothesis problem” developed in the context of criticism of Bayesian. These trends might indicate it is not always appropriate to dichotomize statistical theory only based on the differences in probability interpretation.

As the name of “approximate Bayesian computation” suggests, the ABC-PCM is mainly founded on the Bayesian framework. However, in this thesis, based on recent trends, I evaluate statistical properties of ABC-PCM from the view of frequency. That is, “when we repeatedly apply a certain statistical method (i.e. parameter estimation or hypothesis testing) of ABC-PCM to a sample obtained from a true distribution, then what kind of phenomenon (distribution of estimator, biased/unbiased, error frequency, etc.) occurs”?

Of course, there might be a case where novel problems arise by “mixing” different statistical schools (Ohkubo & Aiba 2019; In Japanese) but these arguments beyond the scope of the current thesis, i.e. to analyze and develop methods of directional evolution in phylogenetic comparative methods.

- **References**

Baker, J., Meade, A., Pagel, M., & Venditti, C. (2016). Positive phenotypic selection inferred from phylogenetics. *Biological Journal of the Linnean Society*, 118 (1), 95-115.

Baker, J., & Venditti, C. (2019). Rapid Change in Mammalian Eye Shape Is Explained by Activity Pattern. *Current Biology*, 29, 1082-1088.

Beaumont, MA, Zhang, W., & Balding, DJ (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162 (4), 2025-2035.

Benjamin, DJ, Berger, JO, Johannesson, M., Nosek, BA, Wagenmakers, EJ, Berk, R., & Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behavior*, 2 (1), 6.

Carvalho, CM, Polson, NG, & Scott, JG (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97* (2), 465–480. https://doi.org/10.1093/biomet/asq017

Castillo, I., Schmidt-Hieber, J., & Van Der Vaart , A. (2015). Bayesian linear regression with sparse priors. *Annals of Statistics*. https://doi.org/10.214/15-AOS1334

Dunbar, RIM, Mac Carron, P., & Shultz, S. (2018). Primate social group sizes exhibit a regular scaling pattern with natural attractors. *Biology Letters*, 14 (1), 20170490.

Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist,* 125, 1-15.

Gelman, A., & Shalizi, CR (2015). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66* (1), 8–38. https://doi.org/10.1111/j.2044-8317.2011.02037

Garamszegi, LZ (2014). Modern phylogenetic comparative methods and their application in evolutionary biology. In: *Modern* phylogenetic comparative methods and their application in evolutionary *biology*. https://doi.org/10.1007/978-3-662-43550- 2

Harano, T., & Kutsukake, N. (2018). Directional selection in the evolution of elongated upper canines in clouded leopards and sabre-toothed cats. *Journal of Evolutionary Biology*, 1–16. Https://doi.org/10.111 /jeb.13309

Hunt, G. (2006). Fitting and comparing models of physical evolution: random walks and beyond. *Paleobiology*, *32* (4), 578-601.

Imai, S., Kobayashi, K., Ohkubo, Y., Yagi, N., & Hasegawa, E. (2016) Difference in Evolutionary Patterns of Strongly or Weakly Selected Characters Among Ant Populations. *Scientific Reports*. https: //doi.org/10.1038/srep39451

Kutsukake, N., & Innan, H. (2013). Simulation-based likelihood approach for evolutionary models of phenotypic traits on phylogenetics. *Evolution*, *67* (2), 355-367. https://doi.org/10.1111/j .1558-5646.2012.0177.x

Losos, JB (2011). Seeing the Forest for the Trees: The Limitations of Phylogenies in Comparative Biology: (American Society of Naturalists Address). *The American Naturalist,* 177 (6), 709-727.

MacLeod, N. 2001. The role of phylogenetics in quantitative paleobiological data analysis. *Paleobiology* 27: 226-240.

Mayo, DG (1996). Error and the growth of experimental knowledge. University of Chicago Press, Chicago.

Mayo, DG (2018). Statistical Inference as Severe Testing. Cambridge University Press, Cambridge.

*Morii, Y., *Ohkubo, Y., Ando Watanabe, S. (2018). Activity of Invasive Slug *Limax Maximus* In Relation to Climate Conditions Based on Citizen's Observations and Novel Regularization Based Statistical Approaches. *Science of The Total Environment.* (*Equally Contributed) https://doi.org/10.1016/j.scitotenv.2016.04.403

Oakley, TH, & Cunningham, CW (2000). Independent contrasts where successor reconstruction failures in a known bacteria phylogeny. *Evolution*, 54 (2), 397-405.

Ohkubo, Y. (2019). Revisiting the Two Major Statistical Problems, Stopping-Rule and The Catch-All Hypothesis, From the Viewpoint of Neo-Bayesian Statistics. *The 16th Congress of Logic, Methodology and Philosophy of Science and Technology.*

Ohkubo, Y. & Aiba, T. (2019). What was p-value and was not. *Seimeikagaku*, 70 (4), 238-251. *In Japanese.

Organ, C., Nunn, C. L., Machanda, Z., & Wrangham, R. W. (2011). Phylogenetic rate shifts in feeding time during the evolution of Homo. *Proceedings of the National Academy of Sciences*, *108*(35), 14555-14559.

Pagel, M. (1999). Inferring the historical patterns of biological Pagel,on. *Nature*, 401 (6756), 877.

Royall, RM (1997), Statistical Evidence: A Likelihood Paradigm. Chapman and Hall, New York.

Sober, E. (2008). Evidence and evolution: The logic behind the science. Cambridge University Press, Cambridge.

Thomas, GH, Meiri, S., & Phillimore, AB (2009). Body size diversification in Anolis: novel environment and island effects. *Evolution*, 63 (8), 2017-2030.

Tavaré, S., Balding, DJ, Griffiths, RC, & Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145 (2), 505-518.

Venditti, C., Meade, A., & Pagel, M. (2011). Multiple routes to mammalian diversity. *Nature*. https://doi.org/10.1038/nature10516

Ito Kiyoshi (1953). Probability Theory, Iwanami Shoten, Tokyo. *In Japanese.

・**Figures.**

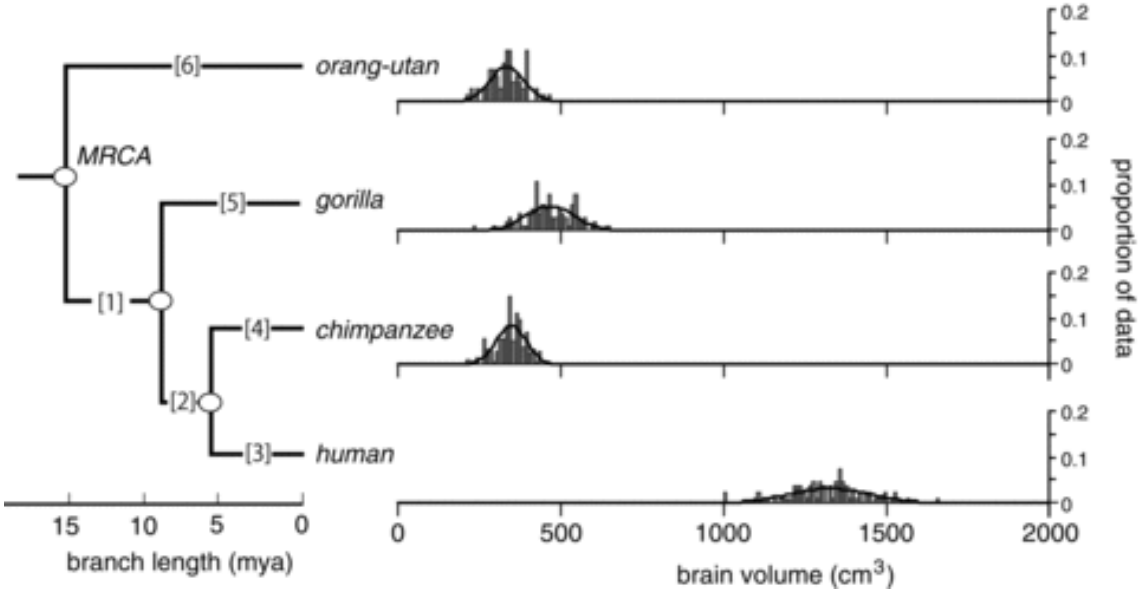

Fig. 1. The phylogenetic tree and the brain volume size of ape species. The left side of the figure shows the phylogenetic tree of 4 ape species where the number embedded in the tree represents the branch number and MRCA indicates the position of the most common recent ancestor. Histograms of the right side show the brain sizes of the corresponding species where the bold line represents estimated trait distribution. (cited from Kutsukake and Innan 2013).

Chapter 2

Simulation Algorithm of Directional Evolution

In Chapter 1, I argued the existing approach for evaluating directional evolution suffers from methodological and interpretational problems of but the simulation likelihood-based method (ABC-PCM), proposed by Kutsukake & Innan (2013), can overcome such problems. However, as mentioned above, ABC-PCM is hardly applied in actual biological studies. One major obstacle is the enormous computational cost that ABC-PCM requires for evolutionary simulation. In this chapter, I first discuss two problems caused by high computational cost and then propose a new algorithm to simulate directional evolution. Finally, applying this algorithm to primate brain size evolution analysis, I confirm the new simulation algorithm achieved 1,000 times faster computation while retaining the accuracy of the posterior distribution.

・The Difficulty of Computation Cost

In this section, I point out the barriers that high computational costs bring to the application of new statistical methods and discuss the significance of computational cost reduction. Note that the calculation cost in this section mainly refers to the time required for simulation procedure and does not consider the time required to learn a new method or coding skills required for implementation. The lack of packaged software that can be easily applied to actual data would be a possible barrier for the application of ABC-PCM into the empirical study. But it will not be discussed in this chapter.

There are two major reasons why the computational cost of ABC-PCM matters. First, it brings a practical burden on the research schedule. When ABC-PCM was first proposed in Kutsukake and Innan (2013), it took approximately one month to get 100,000 posterior samples to analyze brain size data of 4 primates species by 5 workstation PCs with 8 CPU cores (MacPro, OS 10.6.7, 2 × 2.93 GHz Quad-Core Intel Xeon; Kutsukake personal communication). Without such a high-performance PC, as is the case for most empirical researchers, it takes longer time. Further, in recent practices of statistical analysis, it has become common to compare results with multiple statistical models and to expand the structure of the model by doing trial and error (for example, Gelman & Shalizi 2015). It also contributes to bring a novel scientific hypothesis. For ABC-PCM, which can analyze complex evolutional process as long as simulations are possible, the possibility of model expansion is a major advantage over the outlier detection approach and the accelerated Brownian motion approach. However, since it takes one month to obtain reliable results of just one model, it is difficult to evaluate multiple models within a practical research schedule.

Second, more seriously, theoretical problems of the statistical inference arise. When, in statistics, a new method is proposed, it is necessary to show its reliability. In the ideal case, it can be done by mathematical analysis. But, when it is difficult to handle analytically, like ABC-PCM, numerical experiments are often conducted. There, several thousands of artificial data are generated, and the new method is applied to them. But in the case of ABC-PCM, which takes one month to obtain 100,000 samples for a small data set described above, to conduct numerical experiments under different condition is not straightforward. Kutsukake & Innan (2013) conducted simulation experiments on a

power of hypothesis testing of *k*, the strength of directional evolution, but in limited cases. Also, the properties of parameter estimation (e.g. unbiasedness) is not known. No matter how philosophically sounding, applying non-reliable methods to empirical study is a big risk.

Recently, Harano & Kutsukake (2018) and Ohkubo *et al.* (unpublished) introduced MCMC algorithm for ABC-PCM (Casella *et al.* 1992, Marjoram *et al*. 2003), resulting in a certain reduction in computational cost (up to 1,000 times). However, although MCMC reduces the cost of sampling from the posterior distribution, it does not contribute to those of simulation and likelihood calculation. In ABC-PCM, a considerable part of the computation time is spent in the process of simulating the evolution of traits and computation of the likelihood. The relative burden of a simulation grows further when ABC-PCM is applied to hundreds of species or more, as usual inter-species comparison studies do.

In the following section, I focus on simulations of directional evolution in ABC-PCM and devise new algorithms to reduce computational costs. I derived an asymptotic probability distribution which trait that undergoes directional evolution follows, using the elementary law of random variables and the central limit theorem.

・**Methods and Algorithms**

First, I reiterate the simulation algorithm of a directional evolution in the previous ABC-PCM.

A-1: Directional Evolution Algorithm of species $i \to j$

1. Determine the number of mutations by random sample from a Poisson distribution.

1.a the number of the positive direction mutations is

$$\mu_{i\to j}^{+} \sim Poisson(bl * ev * k) \text{- (2.1)}$$

1.b the number of the negative direction mutations is

$$\mu_{i\to j}^{-} \sim Poisson(bl * ev \,/\, k) \text{ -(2.2)}$$

Here, *bl* is the branch length, *ev* is the evolution rate, and *k* is the strength of directional evolution. Notice that it becomes a Brownian Motion model when $k = 1$.

2. For each of the positive and negative mutations, determine the amount of phenotype change Δ from the independent exponential distribution exp (θ).

$$\Delta_m^{+} \sim \exp(\theta) \ (m = 1,2, \ldots \mu_{i\to j}^{+}) - \ (2.3)$$

$$\Delta_n^{-} \sim \exp(\theta) \ (n = 1,2, \ldots \mu_{i\to j}^{-}) - (2.4)$$

Where θ is the hyper parameter which controls the trait change per mutation.

3. Add each trait change one-by-one.

$$X_{i\to j} = \sum_{m=1}^{\mu_{i\to j}^{+}} \Delta_m^{+} \ - (2.5)$$

$$Y_{i\to j} = \sum_{n=1}^{\mu_{i\to j}^{-}} \Delta_n^{-} - (2.6)$$

4. Define the final trait change by $X_{i\to j} - Y_{i\to j}$ $-(2.7)$
5. Add the ancestral trait $y_{sim\,j} = y_{sim\,i} + (X_{i\to j} - Y_{i\to j})$ $-(2.8)$

The simulation algorithm above is configured to reproduce the processes of population genetics step by step (Kutsukake; personal communication). For example, in step 1, the expected value of the number of mutations established in the population is controlled by the parameter *k*, not the mutations that occurred at the gene level. Also, in usual organisms, most mutations cause only a slight change in a trait, but, in rare cases, can cause a large change in a trait (Orr 2005). Such a situation is represented by the exponential distribution in step 2. But, in practices, it is not always necessary to obey the above steps *as par*. A similar evolution can be simulated with high speed if some mathematical analysis and an appropriate approximation is applied.

First, Poisson distribution with mean λ can be approximated by normal distribution if λ is sufficiently large. In general, random samples from a normal distribution are obtained at high speed by the transformation of the uniform distribution by the Box-Muller. But those of Poisson involves condition judgment and repeating processing where the speed is slow (Devroye 1986).

Next, in Step 2.-4, when there are *n* independent random variables $X_1, X_2 \ldots X_n$, all of which is exponential distribution *exp* (θ), the distribution of the sum ($X_1 + X_2 + \ldots + X_n$) is *Gamma* (n, θ) where

$$Gamma(\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}x}$$

$$\mathbb{E}[Gamma(\alpha,\beta)] = \alpha/\beta$$

$$Var[Gamma(\alpha,\beta)] = \alpha/\beta^2$$

Therefore, the distribution of X, the sum of μ^+ positive mutations and the |Y|, the absolute of the sum of μ^- negative mutations is distributed as;

$$X \sim Gamma\ (\mu^+, \theta)\ \text{-(2.9)}$$

$$Y \sim Gamma\ (\mu^-, \theta)\ \text{-(2.10)}$$

Here,

$$\mathbb{E}[X] = \mathbb{E}\left[\frac{k\mu^+}{\theta}\right] = \frac{k * bl * ev}{\theta} - (2.11)$$

$$Var[X] = \mathbb{E}\left[\frac{k\mu^+}{\theta^2}\right] = \frac{k * bl * ev}{\theta^2} - (2.12)$$

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{\mu^+}{k\theta}\right] = \frac{bl * ev}{k\theta} - (2.13)$$

$$Var[Y] = \mathbb{E}\left[\frac{k\mu^+}{k\,\theta^2}\right] = \frac{bl * ev}{k\,\theta^2} - (2.14)$$

Finally, the trait change by evolution is X-Y.

Here, for random variables X and Y,

$$\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y]$$

$$Var[X - Y] = Var[X] + Var[Y] - 2Cov[X, Y]$$

Kutsukake & Innan (2013) assumed independent exponential distribution in step 2, so $Cov[X,Y] = 0$ by definition. Finally, according to the central limit theorem, for *n* independent *Gamma* random variables X_1, X_2, ... X_n,

$$\lim_{\alpha\to\infty}[Gamma(\alpha,\beta)] = Normal(\frac{\alpha}{\beta},\frac{\alpha}{\beta^2})$$

Summarizing the above analysis, an efficient algorithm is obtained below.

A-2: Algorithm of asymptotic directional evolution of species $i \to j$

1. Determine the number of mutations by random numbers from the Normal distribution

1.a the number of the positive direction mutations is

$$\mu^{+}_{i\to j} \sim \text{Normal}(\text{bl} * \text{ev} * \text{k}, \text{bl} * \text{ev} * \text{k}) - (2.15)$$

1.b the number of the negative direction mutations is

$$\mu^{-}_{i\to j} \sim \text{Normal}(\text{bl} * \text{ev/k},\ \ \text{bl} * \text{ev/k}) - (2.16)$$

Where bl is the branch length and ev is the evolution rate, k is the strength of the directional evolution.

2. Determine the change of trait by random numbers

$$(X_{i\to j} - Y_{i\to j}) \sim \text{Normal}(\frac{\mu^{+}_{i\to j} - \mu^{-}_{i\to j}}{\theta}, \frac{\mu^{+}_{i\to j} + \mu^{-}_{i\to j}}{\theta^2}) - (2.17)$$

3. Add the ancestral trait $y_{sim\,j} = y_{sim\,i} + (X_{i\to j} - Y_{i\to j})$

Where θ is the hyper parameter which controls the trait change per mutation.

It should be noted that ABC-PCM does not necessarily follow the above formulation. For example, if it is unsound to assume that the strength of evolution controls the number of mutations, it is possible to formulate *k* as it controls expected change of trait value in (2.3) and (2.4). It derives following simulation.

$$\mu^{+}_{i\to j} \sim Poisson(bl * ev)$$

$$\mu^{-}_{i\to j} \sim Poisson(bl * ev)$$

$$\Delta^{+}_{m} \sim \exp(\theta/k)(m = 1,2,\dots \mu^{+}_{i\to j}) - (2.18)$$

$$\Delta^{-}_{n} \sim \exp(\theta * k)\ (n = 1,2,\dots \mu^{-}_{i\to j}) - (2.19)$$

Here,

$$\mathbb{E}[\exp(\theta)] = 1/\theta$$

$$Var[\exp(\theta)] = 1/\theta^{2}$$

thus

$$\mathbb{E}[\mathrm{X}] = \mathbb{E}\left[\frac{k\mu^{+}_{i\to j}}{\theta}\right] = \frac{k * bl * ev}{\theta} - (2.20)$$

$$Var[\mathrm{X}] = \mathbb{E}\left[\frac{k^{2}\mu^{+}_{i\to j}}{\theta^{2}}\right] = \frac{k^{2} * bl * ev}{\theta^{2}} - (2.21)$$

Therefore, comparing equation (2.11) -(2.12) and the equation (2.20) -(2.21), it can be seen that the evolution is simulated by the almost same algorithm as that of A-2, except that the $\mathrm{Var}[\mathrm{X} - Y]$ is substituted.

Also, when the assumption of independence in the equations (2.9) and (2.10) is unnatural in an actual biological context, X and Y could be replaced by a two-dimensional multivariate random variable. Also, in this case, the simulation can be

performed by the same method except that the variance term of (2.11) decreases as the covariance of X and Y increases. One of the major advantages of ABC-PCM is that it is possible to flexibly incorporate different simulation algorithms into data analysis.

・Simulation Experiments

So far, how to simulate a directional evolution in ABC-PCM with low computational cost have been analyzed using the elementary law of probability and the central limit theorem. The proposed method approximates Poisson distribution and Gamma distribution by Normal distribution, where the accuracy depends on *bl*, the branch length on the phylogenetic tree, *ev*, the evolution rate, and *k,* the strength of the directional evolution (as *bl* * *ev* * *k* decreased, it worsen). It is necessary to evaluate whether the proposed method is accurate. Thus, the difference between the random numbers generated by the two algorithms was examined by simulation experiments. Here, by fixing the length of the branches at 1, I generated 100,000 random samples in various *ev* and *k* and obtained empirical cumulative density. The difference between cumulative density is tested by the method of Kolmogorov=Smirnov (Smirnorv 1939).

Figure 2.1 a-b, shows the empirical cumulative density under different *ev* and *k*. It can be seen that the two lines match well under the tested condition. When *ev* was small and *k* was large, the *p*-value was small suggesting there exist little differences in the two distributions. But as *ev* goes larger, two distribution converged. It is assumed that sufficient approximation has been obtained for the purpose of this chapter. In the next section, I apply the proposed algorithm to the actual data.

・**Data Analysis**

I applied the new simulation algorithm into the brain size data of the apes, mentioned in the previous chapter, and compared the results. The brain size of the most recent common ancestor (*MRCA*), the evolution rate (*ev*), and the strength of directional evolution (*k)* were estimated assuming the brain size of *Homo sapience* experienced a directional evolution after the. speciation of *Homo sapience* and *Pan troglodytes*.

According to Kutsukake & Innan (2013), I set prior distribution of three parameters as follows;

$$MRCA \sim \text{Uniform}\ (330, 1300)$$

$$ev \sim \text{Uniform}\ (1, 10000)$$

$$k \sim \text{Uniform}\ (0, 30)$$

In the above model, I generated 101,000 samples from the posterior distribution by Gibbs sampler and the first 10,000 samples were discarded as a warm-up. The above process is performed by both A-1 and A-2 algorithm, and the posterior distributions of the two are compared.

·**Results**

Posterior distribution

First, trace plots of each parameter were shown to confirm that each chain converged (Fig 2.2). Generally, in the situation where the chain has not converged, the sample values rise and fall like a random walk, but this has not been observed in this result. The

average log likelihoods in both models were in good agreement at -22.57 and -22.55 respectively.

The summary statistics of the posterior distribution (Table 2.1) and the histogram of the sample (Fig. 2.3) are then compared with the two algorithms.

A-1: *MRCA* [333.13 - 444.88], *ev* [28.75 - 4800.56], *k* [1.08 - 26.61]
A-2: *MRCA* [333.14 - 461.15], *ev* [28.59 - 5856.46], *k* [1.07 - 26.01]

Calculation time
The algorithm proposed in this chapter completed the above process in 9.04 seconds, while the calculation by the A-1 algorithm took 164 minutes.

- **Discussion**

In this chapter, I proposed a new algorithm to simulates a directional evolution and it achieved 1,000 times faster computation than the existing one. Also, it was confirmed that the new algorithm can obtain sufficiently a similar posterior distribution.

There are two approximations used in this algorithm; a normal approximation of Poisson distribution to determine the number of mutations, and of the Gamma distribution to determine the sum of trait changes. Approximation accuracy depends on the product of branch length, evolution rate, and the strength of directional evolution. Therefore, if the product is not large enough, it is expected that the approximation deteriorates, and the likelihood value is not accurate. However, in simulation

experiments, it was found that a relatively good approximation was obtained even when the product of branch length and evolution rate was about 10 or so.

The analysis of ape brain size data showed a slight difference in the posterior summary statistics between the two algorithms. However, it is difficult to think that the difference is due to the accuracy of approximation because the estimated evolution rate is above one thousand. Rather, it is supposed to be the stochastic noise of the tail area of the distribution since posterior summary statistics are sensitive to outlier if the distribution is asymmetric.

The algorithm obtained in this chapter is important not only because of the practical convenience but also because large-scale simulation experiments have become possible. In Chapter 3, I generate artificial data under various parameter settings and evaluates the performance of ABC-PCM using this algorithm. At this time, it is important to set true value ranges widely so as not to the accuracy of normal approximation of Poisson and Gamma distribution affects experimental results. The results in this chapter suggest that even if the branch length is 1, the effect of approximation can be almost ignored if the evolution rate is 1000.

Also, in the process of deriving the algorithm, I obtained an approximate distribution of traits changes under the directional evolution. I use this result again in Chapter 4, where further improvement of evolution simulation is introduced.

・**References**

Casella, G., & George, EI (1992). Explaining the Gibbs sampler. *The American Statistician*, 46 (3), 167-174.

Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York.

Gelman , A., & Shalizi , CR (2015). Philosophy and the practice of Bayesian statistics Andrew. *British Journal of Mathematical and Statistical Psychology, 66* (1), 8–38. https://doi.org/10.1111/j .2044-8317.2011.02037

Kutsukake, N., & Innan, H. (2013). Simulation-based likelihood approach for evolutionary models of phenotypic traits on phylogenetics. *Evolution, 67* (2), 355-367. https://doi.org/10.1111/j .1558-5646.2012.0177.x

Harano , T., & Kutsukake , N. (2018). Directional selection in the evolution of elongated upper canines in clouded leopards and sabre-toothed cats. *Journal of Evolutionary Biology,* 1–16. https://doi.org/10.111 /jeb.13309

Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100 (26), 15324-15328.

Orr, HA 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics.* 6: 119.

Smirnov, NV (1939). On the estimation of the discrimination between empirical curves for distribution for two independent samples. *Moscow University Mathematics Bulletin*, 2 (2), 3-14.

・**Tables**.

Table 2.1a : The comparison of the posterior mean between the old and the new algorithm of the directional evolution.

	MRCA	ev	k
A−1	380.2	740.42	6.83
A−2	379.1	698.34	7.18

Table 2.1b : The comparison of the posterior median between the old and the new algorithm of the directional evolution.

	MRCA	ev	k
A−1	374.93	221.1	3.71
A−2	373.97	208.67	3.9

Table 2.1c : The comparison of the MAP between the old and the new algorithm of the directional evolution.

	MRCA	ev	k
A−1	392.55	1606.29	1.24
A−2	334.37	1114.98	1.38

• **Figures.**

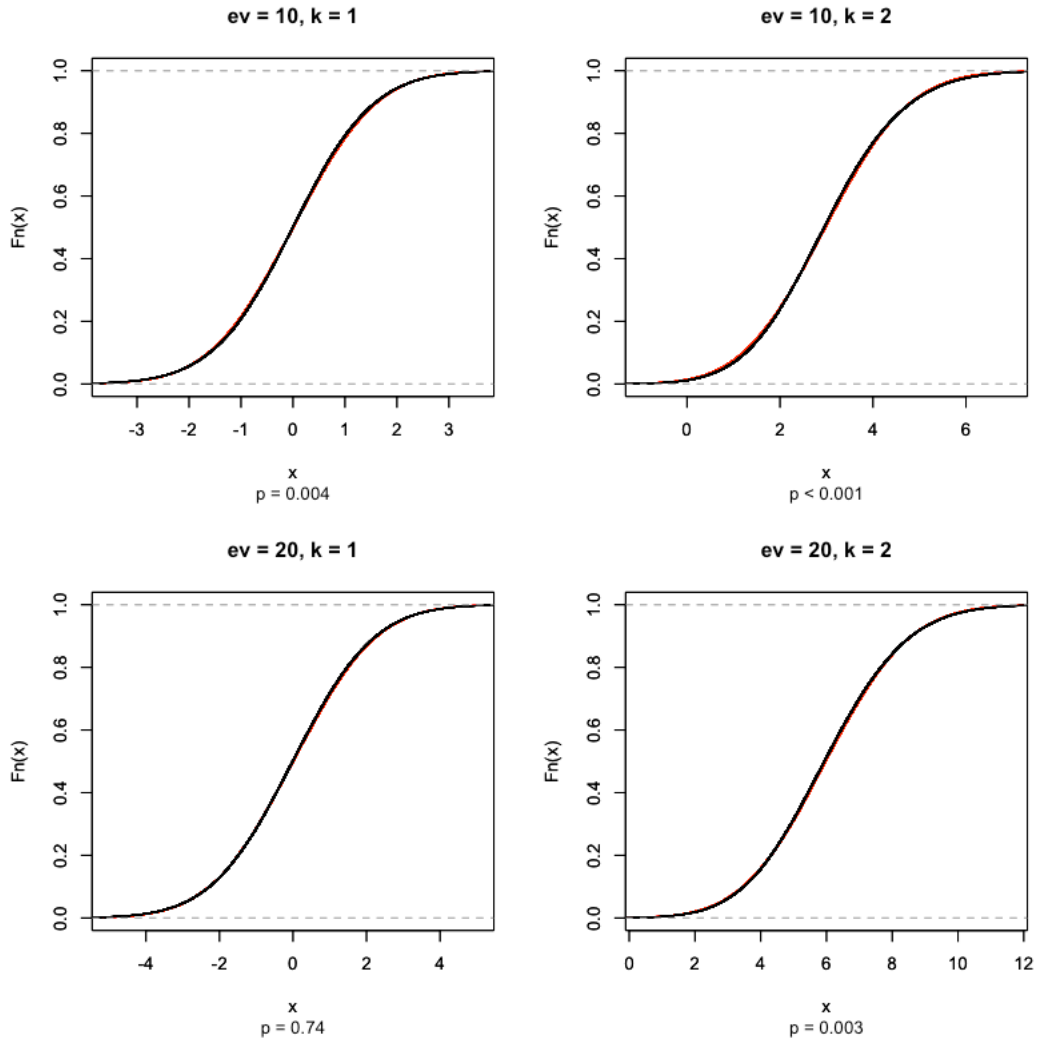

Fig. 2.1a. An empirical distribution of a simulated trait change by the two algorithms (ev = 10, 20). Red line corresponds to the A-1 algorithm and the black one to A-2. The x-axis indicates the trait change and y-axis indicates the cumulative density of the random variable.

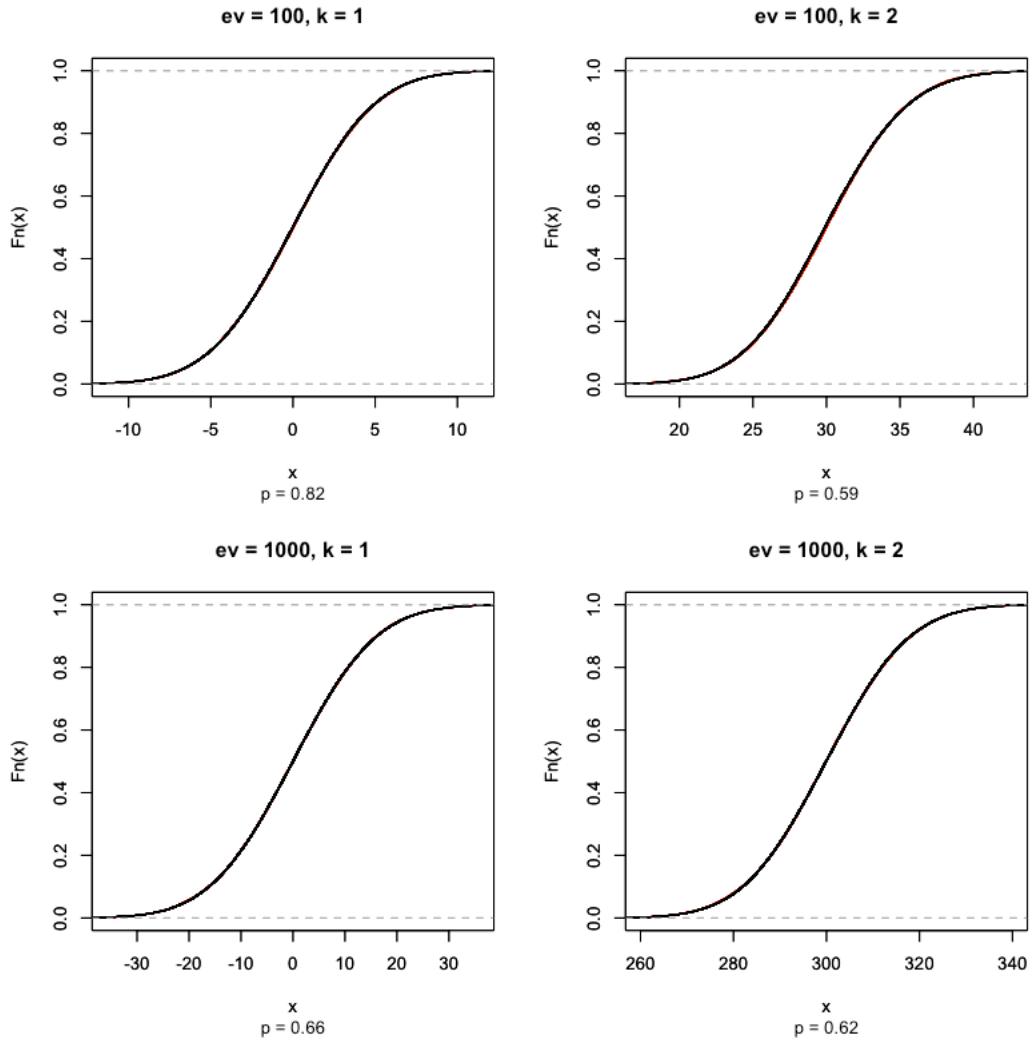

Fig. 2.1b. An empirical distribution of a simulated trait change by the two algorithms (ev = 100, 1000). Red line corresponds to the A-1 algorithm and the black one to A-2. The x-axis indicates the trait change and y-axis indicates the cumulative density of the random variable.

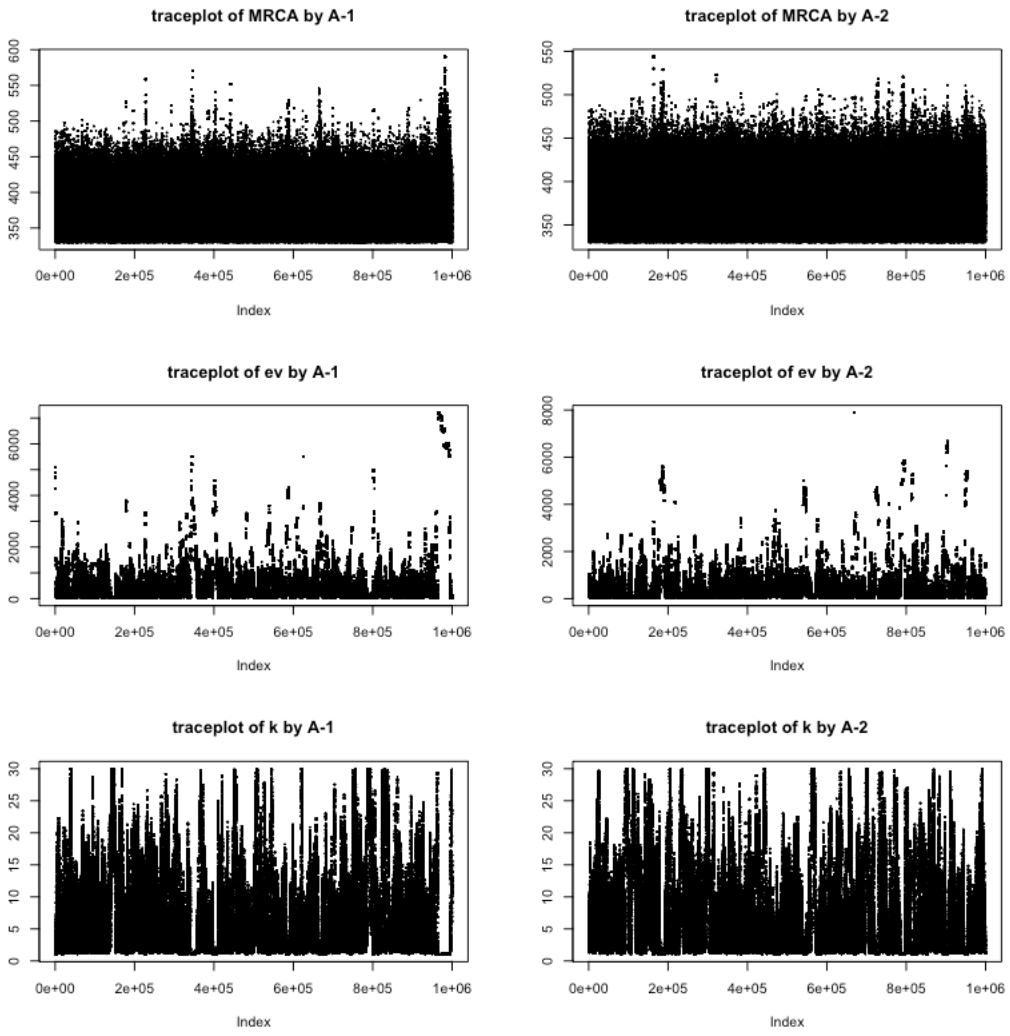

Fig 2.2: Traceplot of the posterior distributions of MRCA, evolution rate and k. Left-side of the figures show the traceplots of A-1 algorithm and right-side of A-2.

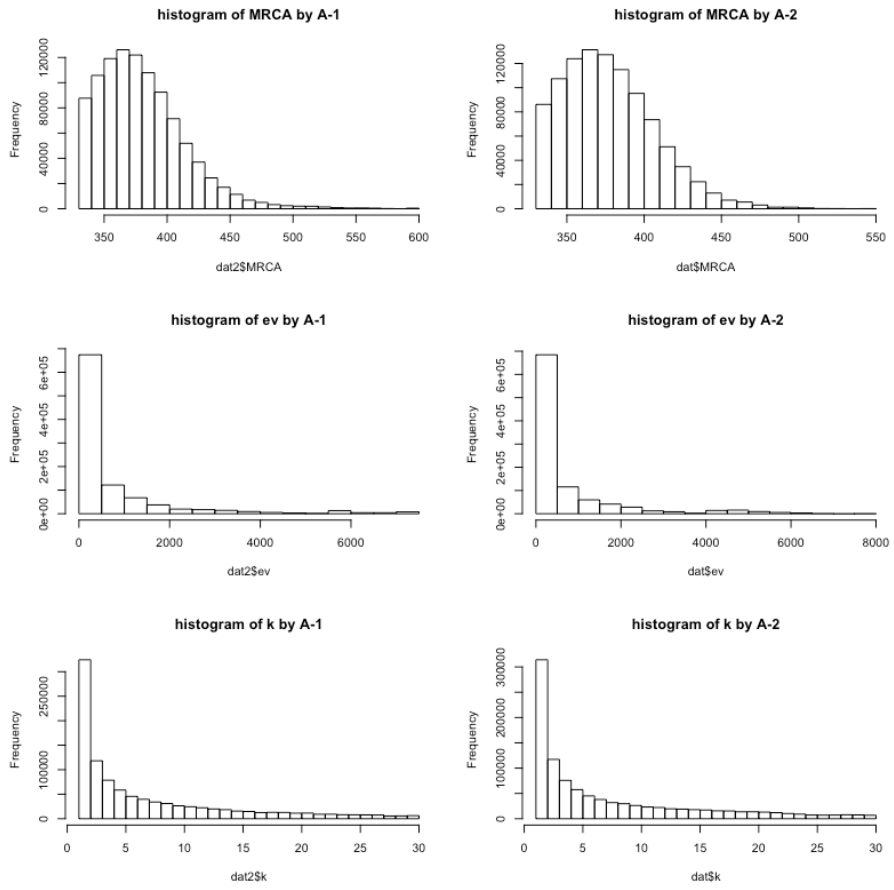

Fig. 2.3: The posterior histogram of MRCA, evolution rate and k. Left-side of the figures show the traceplots of A-1 algorithm and right-side of A-2..

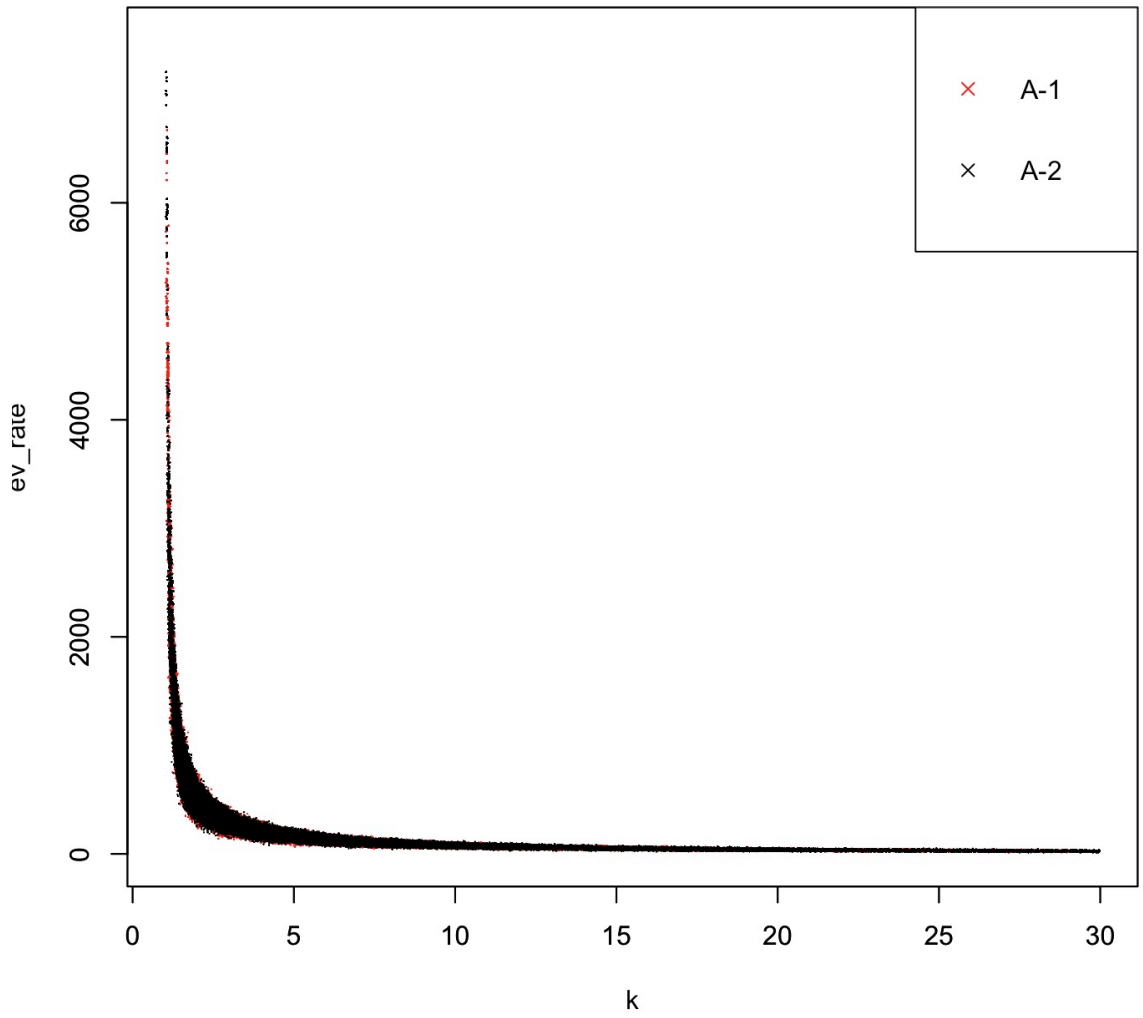

Fig. 2.4: The comparison of the posterior distribution between A-1 and A-2 algorithm.

Chapter 3

Statistical Properties of ABC-PCM

As discussed in Chapter 1, ABC-PCM enjoys excellent flexibility and avoids problems of outlier detection approach and accelerated Brownian models approach in an analysis of directional evolution. However, it has, so far, been hardly known about properties of parameter estimation and hypothesis testing by ABC-PCM. In this chapter, based on the results of Chapter 2, I conducted a simulation experiment, where artificial data were generated under several parameter settings and ABC-PCM was applied to them. I found, a) the strength of directional evolution is overestimated and b) the test for null hypothesis $k = 1$ has low power.

· The Importance of Parameter Estimation by ABC-PCM

Yet parameter estimation is generally the most fundamental problem in any statistical inference, ABC-PCM has more special significance in estimating the strength of directional evolution. So far the two methods, the outlier detection approach and accelerated Brownian motion approach, have been applied to detect the directional evolution in phylogenetic comparative studies. But, as discussed in Chapter 1, these methods suffer from serious philosophical problems since they cannot evaluate the strength of directional evolution quantitatively. One of the major advantages of ABC-PCM is that it offers the direct way to evaluate any parameter in any evolutional model as long as the simulation is possible, avoiding the problems. The possibility of parameter estimation for the strength of the directional evolution is an important advantage of ABC-PCM. However, the possibility of the quantitative evaluation does

not necessarily mean that the quantitative evaluation is reliable since no study has examined the properties of parameter estimation by ABC-PCM. For ABC-PCM to be widely applied to empirical studies, it is extremely important to confirm whether the quantitative evaluation of the strength estimation of directional evolution is reliable.

・**Relationship between parameter estimation and test in ABC-PCM**

Another reason to investigate the properties of parameter estimation in the directional evolution model of ABC-PCM is the preparation for the examination of the hypothesis testing. So far, it has been considered that the decision based on whether 95% confidence interval overlaps the null hypothesis could work as a hypothesis testing with 5% significance level (Kutsukake & Innan 20113, Harano & Kutsukake 2018).

Its interpretation assumes that the alpha% significance testing, the 100-arlpha% credible interval, and the Bayesian 100-alpha% confidence interval is mutually exchangeable. In the linear regression model (LM) or the generalized linear regression model (GLM), the above exchangeability holds for a regression coefficient when the sample size is sufficiently large. It is justified on the nature of the maximum likelihood estimator in LM and GLM (Gelman *et al.* 2004, Watanabe 2018, etc.), but non-trivial for the estimator of the strength of directional evolution obtained by ABC-PCM.

According to the frequentist statistics, an α% significance level testing of a parameter and the 100 - α % credible interval of this parameter is exchangeable as long as the following properties hold; the distribution of maximum likelihood estimators follows the multivariate normal distribution with its mean vector is the true value. That is, when

the process of "obtaining n finite samples from the same probability model, and calculating the maximum likelihood estimate using this n samples" is repeated infinitely, then the estimator holds 1) normality; the distribution of the estimator is the multivariate normal distribution of d dimensions and 2) unbiasedness; the mean vector of the distribution converges to the true value, where, d is the dimension of the parameter vector. In LM and GLM, the regularity of a Fisher Information Matrix guarantees these properties. Also, according to Bayesian statistics, if a posterior distribution is a normal distribution, then 95% credible interval is a good approximation of the 95% confidence interval as long as the maximum likelihood estimator is unbiased.

In order for the 95% confidence interval to be used as an alternative to hypothesis testing in frequentist sense, the unbiasedness of the parameter estimation by ABC-PCM is a necessary condition. But the properties of the parameter estimation have not been investigated. Based on the two backgrounds, this chapter first examines the statistical properties of the parameter estimation of k, the strength of the directional evolution.

In general, there are two major approaches to evaluate a statistical property of a statistical method. One is a mathematical analysis but difficult for ABC-PCM. So, I adopt the numerical approach since a new simulation algorithm, obtained in Chapter 2, brought fast and efficient way to conduct experiments in a broad range of parameter setting. Since there are some differences from the original method when the branch length and evolution rate on the phylogenetic tree is not very large, I set wide range of

true evolution rate to generate artificial data in order to cover areas with different approximation accuracy.

・Simulation Experiment 3.1

I conducted simulation experiments to investigate the properties of parameter estimation of ABC-PCM when applied to directional evolution models. I used a phylogenetic tree of 4 ape spices described earlier and set 6 different *ev* (5, 10, 50, 100, 500, 1000) and 6 different *k* (1, 1.05, 1.1, 1,2, 2.0, 3.0) as true parameter. The trait of a common ancestor (MRCA) is fixed at 100. I repeated the following process.

1. Under the true value (MRCA, common ancestor traits, *ev*, evolution rate, *k*, the strength directed evolution), artificial data was generated by the simulated evolution of a trait. Here, only *Homo sapience* experiences the directional evolution.
2. Set the prior distribution $MRCA \sim (0, 1000)$, $ev \sim$ Uniform (1, 10000), $k \sim$ Uniform (0, 30).
3. 1,010,000 MCMC samples from the posterior distribution are obtained and discard first 10,000 as a warm-up.
4. Get 3 kinds of estimators; posterior mean, posterior median, maximum a posteriori probability value (MAP).

The above process repeated 10,000 times, for each 6 * 6 = 36 settings of the true value.

In order for the samples obtained by the MCMC method to be a good approximation of the posterior distribution, it is necessary to confirm whether the Markov chain has

converged. In the usual practice of Bayesian statistics, visual judgment plays an important role in confirming the convergence but checking 36 sets * 1 million times of Markov chain by trace plot is practically difficult. In the subsequent simulation experiments, the following rule was adopted to ensure the convergence.

1) Make 4 independent Markov chains in each trial and calculate Gelman's Rhat index.
2) If Rhat is above 1.01, discards all samples of the posterior distribution.
3) If the case, regenerate an initial value with random numbers from the prior distribution, and resampling is done.
4) 1) -3) are repeated as long as Rhat is above 1.01.

Note that Rhat indicates "acceptable convergence" if the Rhat is below 1.1 (Gelman *et al*. 2004). I set a bit more conservative rule than usual practice. Also, note that I did not regenerate artificial data when Rhat is above 1.01.

- **Result 3.1**

Table 3.1 a-c shows the results of Experiment 3.1 for each true value. It can be seen that *k* is overestimated regardless of the true value. For example, looking at the posterior mean, it is estimated to be about 4.1 to 4.4 even when the true *k* is 1.0 (Brownian Motion). Overestimation also occurs at the posterior median and the posterior probability maximum estimator (MAP). It is considered that the estimation of *k* by ABC-PCM has no unbiasedness.

・Hypothesis Testing of Strength of Directional Evolution by ABC-PCM

In simulation experiment 2.1, it was revealed that *k* is overestimated in ABC-PCM and that it has no unbiasedness. Thus, any properties based on the multivariate normality of the maximum likelihood estimator would not be validated, including the exchangeability alpha% level hypothesis testing, 100-alpha% credible interval, and 100-alpha% confidence interval. In the second half of this chapter, I will examine the error probability of hypothesis testing for the null hypothesis k=1 (Brownian Motion).

・Simulation Experiment 3.2

I conducted simulation experiments to investigate the properties hypothesis testing of *k* by a 95% credible interval. As Experiment 2.1, I set 6 different *ev* (5, 10, 50, 100, 500, 1000) and 6 different *k* (1, 1.05, 1.1, 1.2, 2.0, 3.0) with fixed MRCA =100. Posterior samples are obtained and calculate whether the 95% confidence interval overlaps the null hypothesis, k=1. I repeated the process 10,000 times for each set of true parameters. The setting of prior and convergence checking rule was the same as Experiment 2.1.

・Result 3.2

Table 3.4 shows the frequency of confidence interval which contains the true value. In all situations tested, the frequency is above 95%. In most setting, there was no interval which is outside the true value, where the 95% confidence interval of the true coverage probability is [0.9996-1.0] (n = 10000).

However, as Table 3.5 shows, the power of detecting k>1 is low. The detection frequency of k > 1 increases as the evolution rate increases, but in the region where

the evolution rate is about 10 to 50, it can hardly be detected even if the true k exceeds 2. Note that since k represents positive mutation occurs k times and negative direction $1/k$ times.

・**Discussion**

In this chapter, I conducted two simulation experiments showing that the estimation of k, the strength of directional evolution is very strongly overestimated, while the hypothesis testing by 95% confidence intervals is overly conservative and power is very weak.

In simulation experiment 2.1 in which the properties of the parameter estimation of ABC-PCM was examined, overestimation occurred in all the conditions tested. While the true value of the evolution rate used in the experiment covered 5 to 1000, overestimation occurred regardless of ev and k. Therefore, we should conclude that overestimation here is not the by-product of the approximation accuracy of the algorithm proposed in Chapter 2.

Experiment 3.2 which evaluated the performance of hypothesis testing by the 95% confidence interval showed the power is low and overly conservative to detect $k > 1$, yet the frequency of the detection becomes larger as the evolution rate increases. The fact that the power depends on the ev is explained as follows; as the true evolution rate increases, the size of the expected mutation increases even when the same branch length and the same strength of directional evolution, making it easier to find the deviation.

Why the overestimation and low detection power coexist? In the following, I discuss this problem. First, I focus on the overestimation of the posterior mean and posterior median. In general, the mean value is not robust estimator being easily influenced when extreme values exist in the data or the true distribution is skewed. As seen in the analysis of the ape brain size data in Chapter 2, the posterior distribution of k has a heavy right tail, so it is natural that the posterior average value is overestimated due to the asymmetry of the posterior distribution. So, it is unsuitable for posterior summary statistics or point estimation. In fact, the posterior mean showed the worst performance among other posterior summary statistics. Since the mean value is sensitive to the extreme value at the tail of the distribution, the posterior average is overestimated almost every time. Yet modest than mean, the median estimator is also dragged to an extreme value or a skewed distribution.

Unlike the mean and the median, MAP estimators are robust against asymmetry of distribution so the above explanation cannot apply.

When ev is large, the expected value of the simulated trait does not change but the variance becomes large from the definition of Brownian motion. So random values from the normal distribution will be highly stochastic for one step of the simulation. Even if the directional evolution is weak, a large trait change may occur and obtain large likelihood with low probability. Conversely, when ev is small, a large likelihood may occur by strengthening the directional evolution. Although the expected likelihood over repeated simulations would change, the maximum likelihood does not change on the curve where the product of ev and k is constant. That is, identifiability is not ensured

between the two parameters. And the likelihood is too high where the average likelihood is low, but the variance of likelihood is large.

For simplicity, suppose that the maximum posterior density on the curve where the two products are constant is completely equal (that is, the MAP estimates on this two-dimensional parameter space are uniformly distributed on this curve). On this curve, when *ev* is large, the corresponding change of *k* will not be large. But, when *ev* is small, *k* changes drastically. Marginalizing *ev*, the expected value of *k* will be sensitive to larger value since the mean value depends on the behavior of the tail area. This is the reason to bias MAP of the strength of the directional evolution by ABC-PCM.

The low power of the hypothesis test by the confidence interval can also be explained from the same viewpoint. Since the likelihood is too high where high evolution rate or strong directional evolution, it is thought that the tail of the posterior distribution becomes wider with its confidence interval.

The numerical experiments in this chapter were carried out based on the A- 2 algorithm of Chapter 2, which adopt some approximation and assumptions. However, the same phenomenon is supposed to occur in a different evolutionary scenario is assumed to the simulation. For example, a model that uses equation (2.18) and equation (2.19) in which the parameter *k* controls the amount of trait change per mutation, not the number of mutations can be applied to the simulation of the evolution. Also, a model with covariances between X, the sum of the trait changes for positive direction and Y, the sum of the trait changes for the negative direction is possible. The only difference from

the adopted simulation is the variance term of normal distribution in step 2 of A-2, the change of the trait variance by the simulation. It does not solve the overestimation of the strength of the directional evolution as long as there is no distinguishability.

One way to solve the problems is to repeat the simulation many times under the same parameter and obtain the average likelihood (Harano, personal communication). Using repeated simulation, MAP is less likely to be sensitive to the stochasticity of likelihood canceling the large likelihood which appears only with low probability (in the case of the current discussion, when the *ev* is large). But this solution comes with a trade-off between the accuracy of likelihood and computational cost. Haruno (personal communication) suggested that 100,000 repeats to obtain an accurate likelihood for a certain set of parameters but it means that the computational cost increases 100,000 times. The reduction of the computational cost achieved in Chapter 2 will be almost offset.

In conclusion, it can be said that the posterior mean, posterior median, and MAP values are all biased estimators in a quantitative assessment of the strength of directional evolution, although the reasons for bias are different for each estimator. The method of hypothesis testing to evaluate the presence of directional evolution based on whether the posterior confident interval crosses the null hypothesis has low type-1 error but has extremely low power. It indicates that ABC-PCM is not reliable for both purposes.

In order to solve the problems identified so far, in the next chapter, I construct a new way of evolution simulations and define new likelihoods for ABC-PCM based on the

results of Chapter 2. There, I show the new method significantly reduces the bias by conducting simulation experiments as in this chapter.

・**References**

Gelman, A., Stern, HS, Carlin, JB, Dunson, DB, Vehtari, A., & Rubin, DB (2004). Bayesian data analysis (The 2nd Edition), Chapman and Hall / CRC, New York.

Harano, T., & Kutsukake, N. (2018). Directional selection in the evolution of elongated upper canines in clouded leopards and sabre-toothed cats. *Journal of Evolutionary Biology*, 1–16. https://doi.org/10.111 /jeb.13309

Kutsukake, N., & Innan, H. (2013). Simulation-based likelihood approach for evolutionary models of phenotypic traits on phylogenetics. *Evolution*, *67* (2), 355-367. https://doi.org/10.1111/j .1558-5646.2012.0177.x

Watanabe, S. (2018). Mathematical Theory of Bayesian Statistics. Chapman & Hall / CRC, New York.

• **Tables.**

Table3.1-a: The true value and the average of the posterior mean by 10,000 simulation.

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	4.43	4.47	4.44	4.48	4.56	4.64
10	4.43	4.44	4.46	4.49	4.69	4.88
50	4.42	4.51	4.60	4.71	6.89	6.18
100	4.41	4.57	4.74	6.41	6.32	7.02
500	4.32	5.05	5.63	6.41	7.79	8.38
1000	4.16	5.52	6.27	6.92	8.22	8.91

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	4.20	4.09	4.21	4.19	4.31	4.15
10	4.24	4.19	4.31	4.33	4.23	4.29
50	4.30	4.24	4.23	4.24	4.85	4.58
100	4.21	4.24	4.30	4.76	4.63	4.98
500	4.18	4.38	4.47	4.76	4.70	6.99
1000	4.12	4.59	4.73	5.25	7.15	7.83

Table3.1-c: The true value and the average of MAP by 10,000 simulation.

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	3.95	3.81	3.88	3.88	4.06	4.04
10	3.90	3.89	4.08	4.05	3.93	4.01
50	3.88	3.93	3.97	3.96	4.64	4.32
100	3.91	4.00	4.02	4.44	4.34	4.68
500	3.90	4.16	4.14	4.47	5.05	6.75
1000	3.92	4.33	4.45	5.01	6.84	7.60

Table3.2-a: The frequency of the confidence interval which contains true k by 10,000 simulation.

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	10000	10000	9999	10000	10000	10000
10	10000	10000	10000	10000	10000	10000
50	10000	10000	10000	10000	10000	10000
100	10000	10000	10000	10000	10000	10000
500	10000	10000	10000	10000	9999	9998
1000	10000	9997	9997	9990	9990	9984

Table3.2-b: The frequency of the detection of k >1 by the 95% confidence interval.

ev/k	1.05	1.1	1.2	2.0	3.0
5	0	0	0	0	0
10	0	0	0	0	18
50	0	0	0	0	18
100	0	0	0	620	9969
500	0	3	6322	10000	10000
1000	85	6592	9991	10000	10000

Chapter 4

A New Framework of ABC-PCM

In Chapter 3, simulation experiments were conducted based on the new computational algorithm introduced in Chapter 2. It was shown that the parameter estimation of the strength of directional evolution does not have unbiasedness (over-estimated) and the hypothesis testing by confidence intervals does not have enough power. In order to solve these problems, this chapter proposes a new simulation algorithm of evolution and defines a new likelihood for ABC-PCM.

・Distribution of a Trait

Before resolving the bias of ABC-PCM, which was clarified in Chapter 3, I show how a predicted trait distribution differs between the accelerated Brownian Motion model and the directional evolution model. It shows the importance of trait variance in phylogenetic comparative methods.

In the accelerated Brownian motion model mentioned in Chapter 1, instead of directly modeling the directional evolution, it was modeled as Brownian Motion with different evolution rates, where the expectation of the change of average trait is always zero. In ABC-PCM, the Brownian Model evolution with acceleration can be represented by the following simulation.

A-3: Asymptotic algorithm for the Accelerated Brownian Motion evolution of species $i \to j$

1. Determine the number of mutation by random numbers from the Normal distribution

1.a the number of the positive direction mutations is

$$\mu_{i\to j}^{+} \sim Normal(bl * ev * k_{acc},\ bl * ev * k_{acc})$$

1.b the number of the negative direction mutations is

$$\mu_{i\to j}^{-} \sim Normal(bl * ev * k_{acc},\ bl * ev * k_{acc})$$

Where *bl* is the branch length, *ev* is the evolution rate, and k_{acc} is the acceleration of baseline evolution rate.

2. Determine the change of trait by random numbers from

$$(X_{i\to j} - Y_{i\to j}) \sim Normal\left(\frac{\mu_{i\to j}^{+} - \mu_{i\to j}^{-}}{\theta}, \frac{\mu_{i\to j}^{+} + \mu_{i\to j}^{-}}{\theta^2}\right)$$

Where θ is the hyper parameter which controls the trait change per mutation.

3. Add the ancestral trait $y_{sim\,j} = y_{sim\,i} + (X_{i\to j} - Y_{i\to j})$

Here, the expected change of mean trait is always zero since $\mathbb{E}\left[\mu_{i\to j}^{+} - \mu_{i\to j}^{-}\right] = 0$. But noting that Step 1, the variance of the trait change in step 2 is proportional to $\mu_{i\to j}^{+} + \mu_{i\to j}^{-}$ when bl and *ev* are given. Since $\mathbb{E}\left[\mu_{i\to j}^{+} + \mu_{i\to j}^{-}\right] = 2 * ev * bl * k_{acc}$, it is assumed that species with accelerated evolution rate have greater trait variance. Of course, this assumption is not necessarily valid in real biological evolution. There is a case that the mean of traits changes drastically but the variance does not.

On the other hand, directional evolution in ABC-PCM can be modeled in which the trait variance does not change drastically.

A-2: Algorithm of asymptotic directional evolution of species $i \to j$

1. Determine the number of mutations by random numbers from the Normal distribution

1.a the number of the positive direction mutations is

$$\mu_{i\to j}^{+} \sim \text{Normal}(\text{bl} * \text{ev} * \text{k}, \text{bl} * \text{ev} * \text{k}) - (2.13)$$

1.b the number of the negative direction mutations is

$$\mu_{i\to j}^{-} \sim \text{Normal}(\text{bl} * \text{ev}/\text{k},\ \text{bl} * \text{ev}/\text{k}) - (2.14)$$

Where bl is the branch length and *ev* is the evolution rate, *k* is the strength of the directional evolution.

2. Determine the change of trait by random numbers

$$(X_{i\to j} - Y_{i\to j}) \sim \text{Normal}(\frac{\mu_{i\to j}^{+} - \mu_{i\to j}^{-}}{\theta}, \frac{\mu_{i\to j}^{+} + \mu_{i\to j}^{-}}{\theta^2}) - (2.15)$$

3. Add the ancestral trait $y_{sim\,j} = y_{sim\,i} + (X_{i\to j} - Y_{i\to j})$

Where θ is the hyper parameter which controls the trait change per mutation.

Here, $\mathbb{E}[\mu_{i\to j}^{+} - \mu_{i\to j}^{-}] = 0$ holds if and only if *k*=1. Therefore, Brownian Motion can be represented as a special case. Similar to accelerated Brownian Motion, the variance of the trait change is proportional to $\mu^{+} + \mu^{-}$when *bl* and *ev* are given. But, unlike the accelerated Brownian Motion, $\mathbb{E}[\mu_{i\to j}^{+} + \mu_{i\to j}^{-}] = ev * bl(k_{dir} + \frac{1}{k_{dir}})$. Letting $(k_{dir}) = k_{dir} + \frac{1}{k_{dir}}$, $f(k_{dir})$ takes its minimum when k_{dir}=1 and $1/k_{dir}$ goes to 0 as $k_{dir} \to \infty$. So, no matter how large k_{dir}, the strength of directional evolution, becomes large, the variance of the trait distribution after directional evolution is at most half of accelerated Brownian Motion. It shows the difference of an evolutionary process is

reflected in the prediction of the trait variance implying the data of trait variance contains quite important information about past evolution.

However, the previous ABC-PCM cannot reflect the information of the trait variance into the likelihood because the simulation of an evolution process is carried out by the random sampling from the trait distribution in step 2 of A-2.

The biases of ABC-PCM revealed in Chapter 3, can be explained by the lack of the trait variance information. For example, when the evolution rate is high, the variance of simulated traits is supposed to be large. Even when k_{dir}, the strength of the directional evolution, is almost 1 (Brownian Motion), the simulated trait change would be large. As a result, the maximum likelihood of the strength of the directional evolution does not lower even when the variance of the simulated trait is drastically different from those of actual trait. Incorporating the information of traits variance is a promising way to obtain an accurate evaluation of models and/or parameters.

I introduce a new simulation algorithm of evolution which involve the information of the trait variance. Also, in order to evaluate the similarity of the simulated distribution and the actual traits, I use a measure called Kullback-Leibler Divergence (KLD) to define a new likelihood.

When there are now two probability distributions, say $p(x)$ and $q(x)$, Kullback-Leibler Divergence (KLD) between the two distributions is defined as follows (Kullback & Leibler 1951).

$$KLD(p(x)||q(x)) \stackrel{\text{def}}{=} \int p(x) log \frac{p(x)}{q(x)} dx$$

Here, KLD ≧ 0, and equality holds if and only if $p(x) = q(x)$. When the actual traits $p(x) = Normal(\mu_2, \sigma_2)$ and simulated trait $q(x) = Normal(\mu_1, \sigma_1)$ are both normal distribution, then

$$KLD(p||q) = \frac{\sigma_2}{\sigma_1} + \frac{{\sigma_1}^2 + (\mu_1 - \mu_2)^2}{2{\sigma_2}^2} + (\text{constant})$$

As is clear from this equation, KLD will be larger if the two distributions have the same mean but the different variances. It is considered that the "similarity" between the distribution of the simulated trait and the distribution of the actual trait can be represented accurately when *ev* or *k* is large. Using KLD, I propose a new trait evolution simulation and definition of the likelihood for analyzing directional evolution with ABC-PCM below.

・**Methods**

First, I formulate a modified algorithm of the simulation of evolution.

A- 4: Algorithm of directional evolution with trait variance of species $i \to j$.

1. Determine the number of mutations by random numbers from the Normal distribution

1.a the number of the positive direction mutations is

$$\mu_{i \to j}^{+} \sim Normal(bl * ev * k_{dir}, bl * ev * k_{dir})$$

1.b the number of the negative direction mutations is

$$\mu_{i\to j}^{-} \sim Normal(bl * ev/k_{dir},\ \ bl * ev/k_{dir})$$

Where bl is the branch length and ev is the evolution rate, k_{dir} is the strength of the directional evolution.

2. If the trait of ancestral species i is $y_{sim\,i} = Normanl(\xi, \sigma^2)$, the evolved trait $y_{sim\,j}$ is simulated as the following rule:

$$y_{sim\,j} = Normanl(\xi + \frac{\mu_{i\to j}^{+} - \mu_{i\to j}^{-}}{\theta}, \sigma^2 + \frac{\mu_{i\to j}^{+} + \mu_{i\to j}^{-}}{\theta^2}) - (3.1)$$

Again, Brownian Motion can be represented as a special case since $\mathbb{E}\left[\mu_{i\to j}^{+} - \mu_{i\to j}^{-}\right] = 0$ holds if and only if *k*=1. Note that, in step 2, instead of simulating evolution with random numbers from a normal distribution, the new algorithm simulates the evolution of the distribution of traits by deterministic way conditional on $\mu_{i\to j}^{+}$ and $\mu_{i\to j}^{-}$.

Assuming that the distribution of the trait of the simulated species *j* is$q_j(x)$ and the actual trait $p_j(x)$, the log likelihood of the parameter θ is defined as follows.

$$log\ L(\theta) = \sum_{j\in extant\ sp.} KLD(p_j(x)||q_j(x))$$

That is, for all the extant species *i*, the log likelihood is the sum of the KLD between the trait distribution of the simulated species *i* and the trait distribution of the actual species *i*.

・**Data Analysis**

I reanalyzed the ape brain size data based on the new likelihood. The model is the same as in Chapter 2. Using the same prior distribution as in Chapter 2, the new likelihood reduced the posterior variance of *k* while of the *ev* increased. It turned out that difficult to capture the overall shape of the posterior distribution, so the prior distribution was changed as follows.

$$ev \sim \text{Uniform}\ (0, 100000)$$

The MAP of each parameter and the 95% confidence interval are as follows. MRCA: 397.9 [335.3-679.0], *k*: 1.084 [1.005-1.115], *ev*: 5330 [3497-42699]. It can be seen that the width of the interval is significantly reduced compared to Chapter 2.

Fig.4.1 shows the 2-dimensional posterior distribution of *ev* and *k*. First, it can be seen that the upper bound of the posterior density is obtained very smoothly with less noise compared to the previous method. Next, the upper limit of the posterior density is significantly reduced in the large *ev* region and the large *k* region, which is in good agreement with previous predictions. That is when *ev* is too large but *k* is too small, the variance of the trait evolved in the simulation is greatly different from the variance of the actual trait resulting in larger KLD (the smaller likelihood).

Another advantage of this new method is computational cost. While it took 8.62 seconds to obtain ten thousand samples by in Chapter 2, the method proposed in this chapter took 1.6 seconds, approximately 5 times faster. It is because the new method requires fewer steps for generating a random number.

The application to actual data revealed new likelihood works well. In the following, I conducted simulation experiments to investigate the properties of parameter estimation, as in Chapter 3.

・**Simulation Experiment 4.1**

Again, I used the phylogenetic tree of 4 ape species and generated artificial trait data. The model and the set of true value is the same to Chapter 3; 6 different *ev* (5, 10, 50, 100, 500, 1000) and 6 different *k* (1, 1.05, 1.1, 1.2, 2.0, 3.0). The trait of the common ancestor is fixed at 100. I repeated generation of artificial data and sampling from the posterior distribution 10,000 times for each pair of the true parameter, where the posterior mean, posterior median, and posterior probability maximum value were recorded for each data. The convergence criterion is also the same as Chapter 3.

・**Result 4.1**

Table 4.1 a-c shows the results of Experiment 3.1 for each true value. It can be seen that any estimator has significantly reduced bias. Although the posterior mean and the posterior median underestimate the true k, MAP estimated was very close to the true value under any condition examined.

・**Simulation experiment 4.2**

As I did in Experiment 3.2 of Chapter 3, I conducted simulation experiments to investigate the error probability of hypothesis testing by a confidence interval. The outline of the experiment is the same as before.

・**Result 4.2**

Table 4.2 a-b shows the frequency of confidence interval containing the true value. In almost all situations, while the error probability is within 95%, the frequency of detecting $k > 1$ when the true k is greater than 1 is drastically improved compared to the method of Chapter 3.

・**Discussion**

In this chapter, I proposed a new method to simulate evolution with the variance information of traits retained. And I defined the new likelihood based on KLD to evaluate the similarity between simulated traits and the actual traits of organisms. While the previous ABC-PCM simulated evolution with the realization of random numbers and calculated the likelihood by a point of trait distribution, the method proposed in this chapter offers a more direct way to evaluate the given parameter.

Incorporating variance information of traits into the simulation, the pair of parameters, *ev*, and *k*, which maximize the likelihood can be determined and indistinguishability was eliminated. Numerical simulation suggested that the almost unbiased estimator can be obtained by MAP.

Harano *et al.* (2018) used a repeated simulation without trait variance directly incorporated and obtained averaged likelihood in ABC-PCM. His motivation was to solve the stacking of MCMC (Harano; personal communication); because the likelihood of ABC-PCM is stochastic and its variance goes large when the evolution rate is high, a posterior sample from high *ev* region generate large likelihood yet with low probability.

But, once the posterior sample from high *ev* region is obtained at iteration t, the proposed value for iteration t+1, which is near the value at t, will be rejected with high probability and the Markov Chain stay there for a long time. It deteriorates the accuracy and efficiency of the sampling. This problem also due to the lack of variance information of traits in the existing ABC-PCM.

To repeat the simulation by random sampling under the same parameters can be seen as an approximate of the method proposed in this chapter, where the integral of KLD is replaced with the finite-sum. However, if there are more branches on the phylogenetic tree, the number of repeating needed for good approximation will be increased, and the computation will be inefficient. Since the proposed method formulates the distribution of traits change, it is possible to simulate the trait variance information with a modest computational cost.

The method of this chapter might be extended further. For a Poisson regression with the conditional mean $\lambda|x$, it is often assumed $log(\lambda|x)$ is normally distributed. The same technique might be also applied to the simulation of ABC-PCM for discrete but countable traits. However, there are more complex cases in biology. The distribution of traits might not follow the normal distribution or evolutionary simulation might be a process other than considered here. If the distribution of the simulated traits is not known, the method of the repeated simulation might be the best alternative. I discuss a broader extension of this study in Chapter 6.

・**References**

Harano, T., & Kutsukake, N. (2018). Directional selection in the evolution of elongated upper canines in clouded leopards and sabre-toothed cats. *Journal of Evolutionary Biology*, 1–16. Https://doi.org/10.111 /jeb.13309

Kullback , S., & Leibler , RA (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics,* 22 (1), 79-86.

• **Tables.**

Table4.1-a: The true value and the average of the posterior mean by 10,000 simulation.

ev/k	**1.0**	**1.05**	**1.1**	**1.2**	**2.0**	**3.0**
5	1.06	1.08	1.10	1.15	1.53	4.64
10	1.02	1.05	1.07	1.12	1.51	2.00
50	1.01	1.02	1.05	1.09	1.49	2.00
100	1.00	1.02	1.04	1.09	1.50	2.28
500	1.00	1.02	1.05	1.10	1.56	2.15
1000	1.00	1.02	1.05	1.11	1.63	2.27

Table4.1-b: The true value and the average of the posterior median by 10,000 simulation.

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	1.02	1.03	1.04	1.07	1.31	1.61
10	1.01	1.02	1.03	1.06	1.30	1.60
50	1.00	1.01	1.03	1.05	1.30	1.62
100	1.00	1.01	1.02	1.06	1.31	1.91
500	1.00	1.01	1.03	1.07	1.38	1.77
1000	0.99	1.02	1.04	1.08	1.45	1.91

Tabl4.1-c: The true value and the average of MAP by 10,000 simulation.

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	1.04	1.09	1.14	1.23	1.98	2.89
10	1.02	1.06	1.11	1.21	1.96	2.90
50	1.00	1.05	1.10	1.19	1.95	2.91
100	1.00	1.05	1.09	1.19	1.96	2.94
500	1.00	1.04	1.09	1.19	1.96	2.94
1000	1.00	1.04	1.09	1.19	1.96	2.93

Table4.2-a: The frequency of the confidence interval which contains true k by 10,000

ev/k	1.0	1.05	1.1	1.2	2.0	3.0
5	10000	10000	10000	9993	9856	9802
10	9999	9998	9996	9976	9877	9841
50	9999	9995	9941	9827	9885	9855
100	10000	9973	9852	9793	9869	9859
500	10000	9779	9632	9632	9778	9747
1000	10000	9586	9430	9469	9639	9551

Table4.2-b: The frequency of the detection of k >1 by the 95% confidence interval.

ev/k	1.05	1.1	1.2	2.0	3.0
5	0	0	2	12	2299
10	2	3	5	406	6616
50	0	1	1	8833	9909
100	0	0	149	9913	9991
500	22	2435	9748	9998	10000
1000	961	9238	9983	10000	10000

Figures.

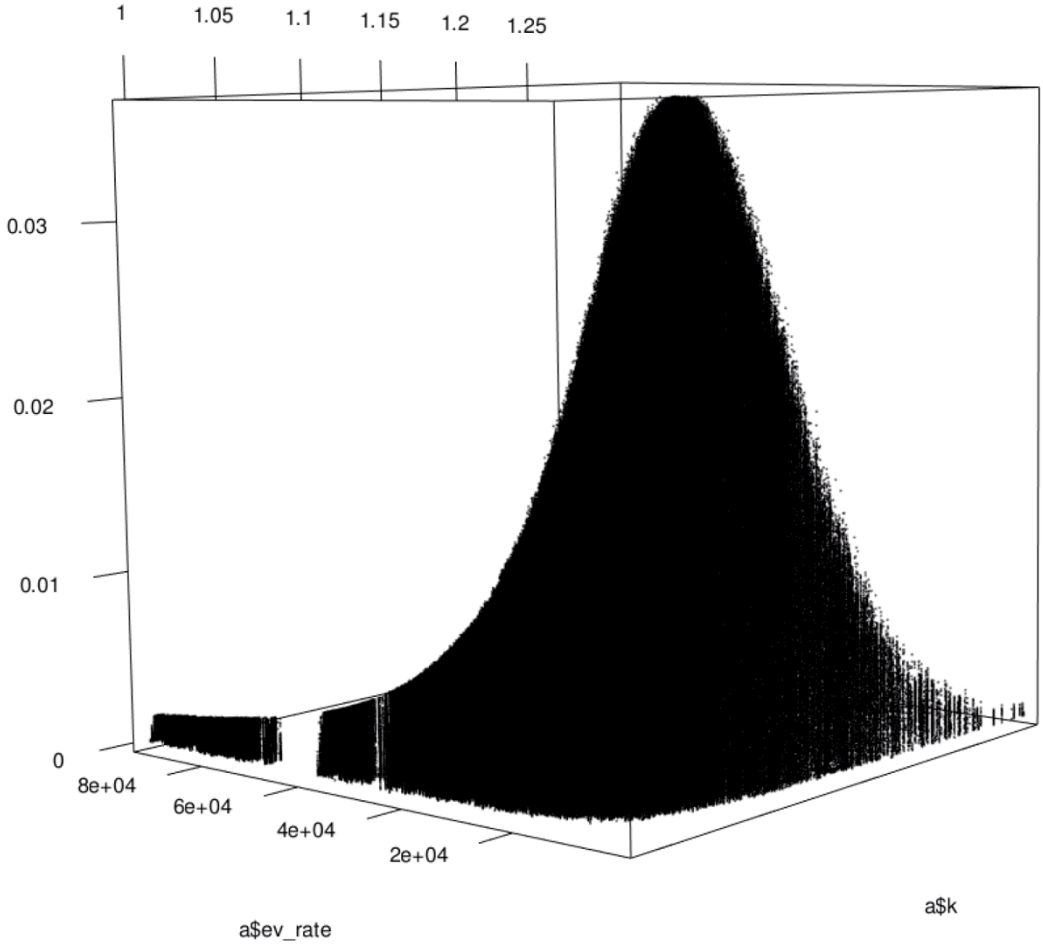

Fig 4.1 The 3-D plot of the posterior distribution, where x-axis is the k, y-axis is the ev, and the z-axis is the posterior density.

Chapter 5

Application: Meta-analysis of Temporal Discounting of Ape Species

Many animals, including humans, prefer a small immediate benefit than a large but temporally delayed benefit (Ainslie 1974). This seemingly irrational behavior is called as "temporal discounting" and has long been studied in various areas including economics (Frederick *et al*. 2002), psychology (Rambaud 2004), and biology (Reynolds 2002). However, why temporal discounting has evolved is not clarified yet. Although some theoretical studies have examined under what conditions temporal discounting will be adaptive (i.e. Sozou 1998, Fawcett *et al.* 2012), they have not empirically been confirmed yet.

Stevens (2014) conducted a meta-analysis using primate data to compare maximum waiting time to a fixed amount of reward and compared three plausible hypotheses why temporal discounting has evolved. One evolutional explanation is the metabolic rate hypothesis. Small animals tend to have high metabolic rates, so the ability of energy storage is limited. It might be adaptive to choose immediate but small reward not to run out of energy. This explanation predicts that the maximum waiting time to rewards is correlated to the allometric body size. Second is the cognitive ability hypothesis. The preference for a small immediate reward might be induced by just the lack of cognitive ability which enables subjects to predict a larger reward of the future or to compare the number of rewards on a long-term perspective. Since a cognitive ability is hard to quantify, Stevens (2014) assumed that the relative brain size is a good proxy of the cognitive ability, predicting that the waiting time negatively correlated with the relative

brain size. The third is the social complexity hypothesis. Previous studies have suggested that the complexity of social structure is one of the key selection pressures on decision making. For example, Amici *et al*. (2008) argued that inhibiting impulsive responses is adaptive to subjects when there are higher ranked members. The measurement of social complexity is also hard to quantify, so Stevens assumed that the average number of group size is a proxy, predicting that waiting time positively correlate with the group size. Stevens (2014) collected 13 ape spices data and employed a phylogenetic comparative method assuming Brownian Motion evolution.

▪Methods

I re-analyzed this dataset by the original ABC-PCM (Kutsukake and Innan 2013) and the proposed method (Capter 4). Because the sample sizes of *Eulemur macaco* and *Varecia rubra* were small, the trait variances were not obtained. Therefore, I substituted them by random samples from a common gamma distribution where the parameter was estimated by other species. I assumed a directional evolution model of maximum waiting time, where k_{dir}, the strength of the directional evolution is proportional to an explanatory variable of interest (e.g. body size and group size). That is;

the number of positive direction mutations is

$$\mu^{+}_{i\to j} \sim Normal(bl * ev * k_{dir}, bl * ev * k_{dir})$$

the number of negative direction mutations is

$$\mu^{-}_{i\to j} \sim Normal(bl * ev/k_{dir},\ \ bl * ev/k_{dir})$$

Where

$$k_{dir} = 1 + \beta x_j$$

Here, x_j is the explanatory variable of spices j and β is the coefficient parameter. Note that I assumed that maximum waiting time evolved following Brownian Motion model for all the extinct species. The number of MCMC iteration was 5,000,000 where the first 500,000 samples were discarded as a warm-up.

▪Results

Table 4.3 shows the posterior model probability and the log marginal likelihood of each model. Note that the comparison of marginal likelihood is meaningless for inter different method since the definition of likelihood is different. While the posterior probability of the original ABC-PCM indicates body size model being the best model, which corresponds to Stevens (2014), the proposed method indicates that the group size model is the best explanatory variable. For the group size model, Fig. 5.1 shows the posterior histogram of β. The 95% confidential interval of β was [0.0005 - 0.02].

▪Discussion

The reason for the different results could be explained by the inter-species difference of the trait variance. For example, while the trait means of *Varecia variegate* (17.9 sec), *Callithrix jacchus* (14.4 sec) and *Macaca mulatta* (19.3 sec) were similar, the standard deviations were drastically different (7.87, 3.41 and 20.9 respectively). This important information cannot be modeled in the original ABC-PCM while the proposed method offers a formal way to capture the difference of trait variance via KLD.

References

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. Current Biology, 18(18), 1415-1419.

Fawcett, T. W., McNamara, J. M., & Houston, A. I. (2012). When is it adaptive to be patient? A general framework for evaluating delayed rewards. Behavioural processes, 89(2), 128-136.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. Journal of economic literature, 40(2), 351-401.

Rambaud, S. C., & Torrecillas, M. J. M. (2004). An analysis of the anomalies in traditional discounting models. International Journal of Psychology and Psychological Therapy, 4(1), 105-128.

Reynolds, B., De Wit, H., & Richards, J. B. (2002). Delay of gratification and delay discounting in rats. Behavioural Processes, 59(3), 157-168.

Sozou, P. D. (1998). On hyperbolic discounting and uncertain hazard rates. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265(1409), 2015-2020.

Stevens, J. R. (2014). Evolutionary pressures on primate intertemporal choice. Proceedings of the Royal Society B: Biological Sciences, 281(1786), 20140499.

Tables.

Table 5.1. the posterior model probability (the log marginal likelihood) of each model.

	Original ABC-PCM	Proposed Method
Null Model	4.3% (-67.12)	9.4% (-17.95)
Body Size Model	90.5% (-64.09)	14.5% (-17.52)
Relative Brain Size Model	1.6% (-68.06)	28.9% (-16.835)
Group Size Model	3.8% (-67.23)	46.9% (-16.35)

Figures.

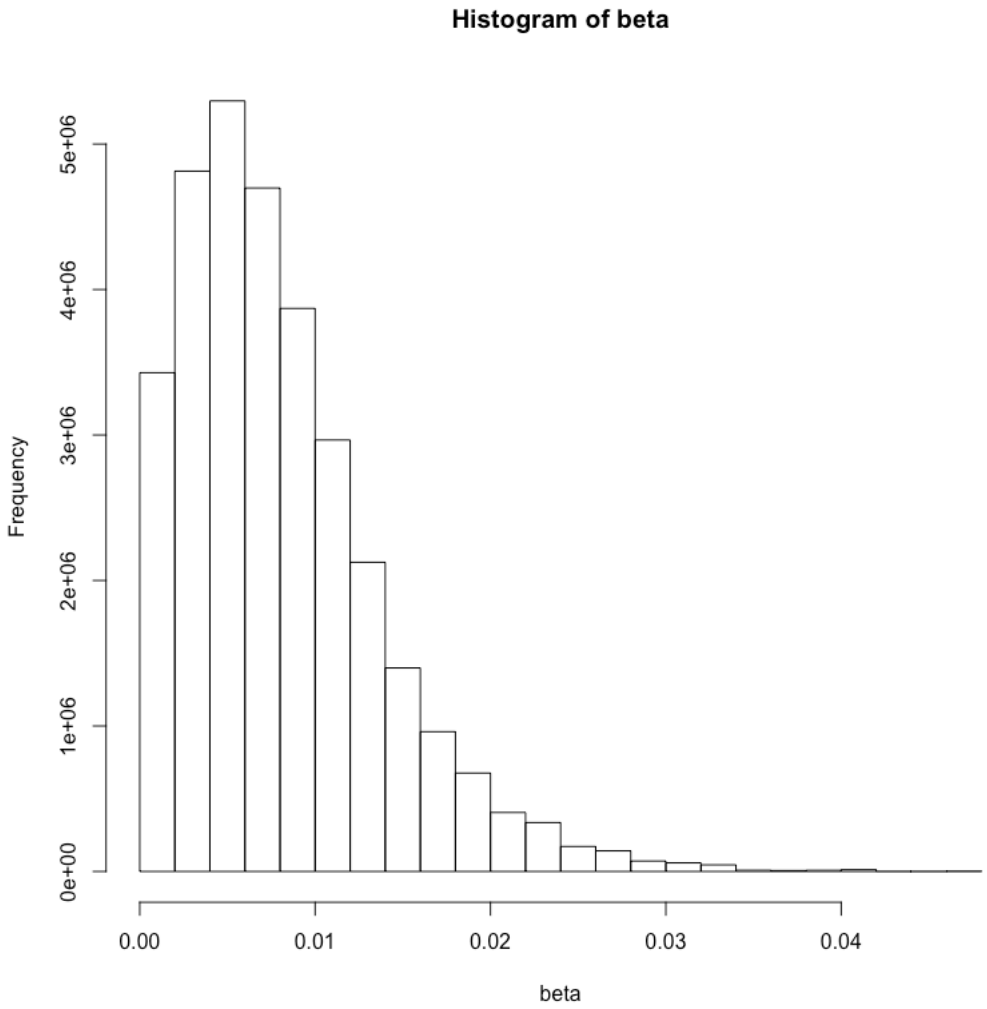

Fig. 5.1 The posterior histogram of β, the coefficient parameter for the strength of directional selection for the group size model.

Chapter 6

General Discussion: Limitations and Extensions

In this thesis, the effectiveness of ABC-PCM to the analyses of directional evolution in phylogenetic comparison has been examined and contributed to the following three points.

1. The calculation speed was improved by mathematical analysis
2. Parameter estimation by the traditional ABC-PCM had a bias there and hypothesis testing had low detection power.
3. Proposed a new likelihood using KLD and greatly reduced the estimation bias

In the final chapter, I discuss the further extensibility of the method proposed in this thesis as well as its limitations.

・Interpretation of Results (1: from a PCM perspectives)

Not only in PCM but also a scientific model in general, how to interpret the structure of the model is a big problem (Weisberg 2012, Evans *et al.* 2013). For example, when ABC-PCM supported for a directional evolution model compared to Brownian motion model and/or an accelerated Brownian motion, is it evidence that there was a process equivalent to directional "selection" in the real biological system?

Unfortunately, it might not be the case. Suppose the following scenario; in the ancestral population, the tail of the trait distribution does not obtain high fitness, but the optimum

trait value in the fitness landscape changed in a certain direction. Then the mean of the trait distribution changed in this way. Even though the unbiasedness the problem of how to interpret the estimated value still remains.

It is not only ABC-PCM that the correspondence problems occur. In the existing PCM, similar arguments have been provoked to the interpretation of Brownian Motion model and the Ornstein-Uhlenbeck model (Hansen *et al.* 1996, O'Meara *et al.* 2006, Uyeda *et al.* 2014). The random variable of the Brownian Motion in a positive direction as well as in a negative direction without constant regularity ("drunken walk"). Even though the trait evolution of an organism fits the Brownian Motion well, it is not possible to distinguish whether the trait experienced a neutral evolution or a stabilization selection where the optimal value of the trait itself was changing.
Also, some might argue that the introduced idealization in the proposed method is too restrictive and not realistic. For example, the proposed method assumes that the expected trait change of directional evolution is a linear function of time but the brain volume size of *Homo sapience* might have experienced *non*linear and burst-like evolution.

In general, a mathematical and statistical model is not aimed to capture all the features of our complex world and intended idealizations (distortion of realities) are introduced as have been pointed out by philosophers for decades (Giere 1988; 2004, Knuuttila 2011, Weisberg 2012). Here, scientists need to decide which aspects in the world to be modeled with high fidelity and which aspects to be distorted or discarded.

Here, one important perspective on a decision is the focus of the study (Weisberg 2012). For this study, the focus has been to detect and quantify the strength of directional evolution. If one would be interested in linear vs burst-like evolution, other models would be needed introducing another idealization.

At least, the proposed method achieved an unbiased estimation of the strength of directional evolution if the supposed model is true, whereas previous ABC-PCM is proven to produce biased estimation *even when* the supposed directional evolution model is known to be true.

PCM is a challenging attempt to infer the evolutional trait changes that have occurred in the past only from the limited information, i.e. the traits of the extant species and the phylogenetic tree. It should be noted that the characteristic of PCM is that it enables us to analyze the past evolution without knowing why or under what mechanisms the trait has changed. In that sense, PCM analysis might be an exploratory practice that proposes further research topics, and it is necessary to search for the actual evolutionary mechanism (Losos 2011). A similar argument could be made about the directional evolution of ABC-PCM.

・**Interpretation of Results (2: from a statistics viewpoint)**

In Bayesian statistics, a posterior distribution and the posterior model probability have been described as expressing "the probability that a certain hypothesis (or model) is true" or "the degree to which a (rational agent) should believe a certain hypothesis is true". However, such explanations often assume that true values (or models) exist in

parameter space and model space, with probability 1 (certainly). If not the case, another interpretation is needed (for example, Bissiri *et al.* 2016). Here in this section, I discuss the meaning of the results of data analysis based on ABC-PCM.

The likelihood of ABC-PCM, introduced in Chapter 4, is based on KLD. According to its definition, it is understood that the MAP estimate of ABC-PCM is "the parameter that minimizes the difference between the distribution of simulated traits and the distribution of actual traits *within the space of the given model and the hyperparameters*". Note that, here, the design of a prior distribution is also included in the model structure. Even if the model structure or the hyperparameter is inappropriate, it is possible to obtain the parameter that minimizes KLD under the given conditions. But if the model or the hyperparameter changes, the result will also change. For example, if we put a larger value on the hyperparameter, which controls trait change per mutation, the same likelihood can be obtained with a small number of mutations. The selection of hyperparameters greatly affect the results. It is risky to interpret the posterior estimates and the confidence intervals literally as "probabilities of truth". Therefore, a method for evaluating the validity of the model structure and hyperparameters is needed.

One possible method to the evaluation is the marginal likelihood, which is widely used in Bayesian statistics and is defined as follows.

$$\int L(\theta)\, p(\theta) d\,\theta$$

Here, $L(\theta)$ is the likelihood of the parameter θ and $p(\theta)$ is its prior distribution. That is, the marginal likelihood is the average of the likelihood, weighted by the prior

distribution. By using this, it is possible to evaluate models and hyperparameters within a set of candidates.

However, this approach alone might be insufficient to evaluate the model. The numerical experiments in this thesis dealt with phylogenetic trees with just four extant species, but other PCMs have already been conducted on a larger scale (i.e. hundreds to thousands of species; Venditti *et al.* 2011). ABC-PCM may also be applied to such a larger-scale problem as high-speed computation have been achieved. Here, for example, suppose that most species (let $p < N$) out of all N species are well fitted to the model, but a small number of N-p species deviate from the model. Since the likelihood of Chapter 4 is defined as the sum of KLD for all the extant species, the average likelihood of the model may not decrease so much even when N-p species does not fit the model well. In the field of machine learning, a similar phenomenon is called as the "class imbalance problem", which plagues many researchers (Buda *et a*l. 2018). Although it is possible to compare multiple models by marginal likelihood, it is unsuitable when all the prepared models are inappropriate or when the actual trait does not fit well, and the model needs improvement.

My proposal in such a case is the outlier detection approach, which I criticized in Chapter 1. The outlier detection approach generates a posterior prediction distribution for each extant species and interpreting that the actual character is an outlier in the posterior prediction distribution as if directional evolution occurred in that species. Major criticisms in Chapter 1 were 4 folded; 1) the inference depends on an arbitrarily chosen threshold to decide the estimated trait value is outlier or not, 2) probabilistic MT

is not a valid inference, 3) outlier detection approach does not offer a quantitative evaluation of the strength and of directional evolution and its difference between branches, and 4) there is incoherence between the inference and the prediction. None of these issues is problematic if the goal is not to inference but only to assess the validity of the current model structure and hyperparameters.

First, regarding 1) and 2), the detection of outliers is not for the final decision but just for the examination if the current model is wrong. Since my proposal here is to search for outliers using a posterior prediction distribution in order to examine improvement points of the model, it is not necessary to make a binary judgment whether it is an outlier or not. In addition, whether the amendment of the model based on the outliers is effective or not can be compared and evaluated, again, by the marginal likelihood. Regarding 3), this problem does not occur in the ABC-PCM because the model is not restricted to the Brownian motion model. Finally, regarding 4), the model evaluation by outlier detection is one of the processes to find out misspecification and to improve the model. It is not a problem that the outlier detection method cannot be used for a prediction.

・**Importance of Trait variance**

Historically, inter-spices study and meta-analysis have mostly focused on the mean values of traits but, in recent years, the significance of variance of traits has been actively discussed (e.g. Nakagawa *et al.* 2015, 2017). The bias of ABC-PCM clarified in Chapter 3 and the usefulness of the new likelihood for ABC-PCM proposed in Chapter 4 might serve a novel viewpoint for the importance of the trait variance.

In the short term, to collect reliable trait variance might be difficult work. In ecology, there have been commonly used libraries for average trait value. It enables practitioners to conduct inter-species meta-analysis without intensive observations in the field. However, there exist comparatively fewer libraries for variance value and it requires a larger amount of samples to obtain a reliable estimation of the variance than average. It could cause troubles for practitioners to apply the proposed method.

One immediate solution is to examine the sensitivity of posterior distribution by the usage of unreliable variance data and/or to examine the required sample size to obtain reliable variance data. Another solution is an extension of the proposed approach to incorporate trait variance uncertainty. I will discuss this point in a later section with other types of data uncertainty. I hope the proposed method would stimulate further discussion about the role of trait variance both between methodologists and practitioners.

Another problem is the trait variance of ancestral species. In previous ABC-PCM, the mean trait of the most recent common ancestor traits can be estimated. But the handling of the variance is not trivial. If the estimation of the variance of the most recent common ancestor traits is not distinguishable, it might be necessary to give as a hyperparameter and could be optimized by marginal likelihood. In the existing phylogenetic comparative methods with Brownian Motion, it is known that fossil information improves the accuracy of the parameter estimation (Slater *et al.* 2012). It would be also useful for ABC-PCM. Further examination is needed for the properties of the trait variance of ancestral traits.

・**Extension of This Research**

In this research, I have considered a limited case where the evolution is caused by Brownian Motion or directional evolution. But the proposed method can be applied to various evolution models including the Ornstein-Uhlenbeck model and the Early Burst model, as long as the evolution can be simulated with trait variance. However, in the simulation experiments conducted in Chapter 4, it is not clear in what evolutionary model the parameter estimation satisfies the unbiasedness and the distinguishability. It is necessary to accumulate further knowledge about ABC-PCM by simulation experiments under various conditions and also mathematical analysis.

In Chapter 4, we have assumed the ideal situation where the trait distribution of an organism follows an exact normal distribution and that there is no uncertainty in its estimation. Of course, such a situation does not necessarily reflect the actual inter-species comparative study. If it is known that the trait follows a distribution other than the normal distribution, and the KLD of that distribution can be calculated, the likelihood of the simulated trait can be calculated in the same manner as in Chapter 4. On the other hand, it is an important task to consider what kind of behavior the estimation and test by ABC-PCM will behave when uncertainty is involved in the trait distribution of the actual organism or an incorrect distribution is assumed to the trait distribution.

There are two possible approaches to deal with uncertainty in the distribution of traits. One is to obtain the posterior estimates of the focal traits by a finite sample from the populations in advance. When we evaluate the KLD between the simulated trait

distribution and the actual trait distribution, it is possible to fluctuate the actual trait according to the uncertainty of the pre-estimated distribution. This method has the advantage that the uncertainty can be considered without increasing the computational cost much. However, if an incorrect assumption is made on the shape of the distribution at the stage of estimating the true trait distribution, it may lead to bias or large variance of the estimator. Using a measure other than KLD might be a promising approach. Recently, Fujisawa and Eguchi (2008) proposed Gamma Divergence and Kanamori and Fujisawa (2014) introduced Holder divergence both of them are more robust than KLD. Wasserstein distance is also known to be useful to evaluate the similarity of the two distribution (Panaretos *et al.* 2018).

Another method is the finite-sum approximation of the integral part of KLD assuming the finite observation of the trait is the random sample from the true trait distribution. The same idea applied to derive the Akaike Information Criterion (Akaike 1998, Konishi & Kitagawa 2008). In this case, there is no need to explicitly describe the probability distribution of trait since it can be approximated by the empirical distribution. It is thought that the above bias caused by assuming an incorrect distribution for the trait distribution can be avoided. However, since it is necessary to calculate the finite sum approximation each time the likelihood is calculated, it is considered that the computation cost increases when the sample size is large.

Uncertainty of a phylogenetic tree is a concern as is the cases of other phylogenetic comparative methods (Cooper *et al.* 2016). Uncertainty of blench lengths would be incorporated into analysis by random sampling from the posterior length distribution.

Properties of the parameter estimation and hypothesis testing found in Chapter 4 might hold as long as the posterior length distribution is unimodal, yet the posterior variance of the evolution rate and the strength of the directional evolution would be larger. But, if topology changes, a posterior distribution of evolution rate and directional evolutionary strength may change drastically. Of course, it is possible to sample a posterior distribution conditional a phylogenetic topology and to obtain an averaged posterior distribution weighted by the posterior probability of the topology. However, the resulting posterior distribution would be a mixed distribution with a complex shape. It might spoil the unbiasedness. It is necessary to further consideration on a phylogenetic uncertainty inside and outside ABC-PCM. I hope this study will help further development of phylogenetic comparative methods and its application into practices of biology.

- **Reference**

Akaike, H. (1998). Information theory and an, extension of the maximum likelihood principle. In: *Selected papers of Hirotugu Akaike* (pp. 199-213). Springer New York, NY.

Bissiri, PG, Holmes, CC, & Walker, SG (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78 (5), 1103-1130.

Buda, M., Maki, A., & Mazurowski , MA (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249-259.

Evans, MR, Grimm, V., Johst, K., Knuuttila, T., De Langhe, R., Lessels, CM, ... & Wilkinson, DJ (2013). Do simple models lead to generality in ecology? *Trends in Ecology and Evolution*, 28 (10), 578-583.

Fujisawa, H., & Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99(9), 2053-2081.

Giere, R.N. (1988). Explaining science: a cognitive approach. Chicago University Press, Chicago

Giere, R. N. (2004). How models are used to represent reality. Philosophy of science, 71(5), 742-752.

Hansen, TF, & Martins, EP (1996). Translation between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50 (4), 1404-1417.

Kanamori, T., & Fujisawa, H. (2014). Affine invariant divergences associated with proper composite scoring rules and their applications. *Bernoulli*, 20(4), 2278-2304.

Konishi, S., & Kitagawa, G. (2008). Information criteria and statistical modeling. Springer Science & Business Media, Berlin.

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. Studies in History and Philosophy of Science Part A, 42(2), 262-271.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, AM (2015). Meta-analysis of variation: Ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6* (2), 143-152. https://Doi.Org/10.1111/2041-210X.12309

Nakagawa, S., Noble, DWA, Senior, AM, & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC Biology*, *15* (1), 1–14. https://doi.org/10.186/s12915-017-0357-7

O'Meara, BC, Ané, C., Sanderson, MJ, & Wainwright, PC (2006). Testing for different rates of continuous evolution using likelihood. *Evolution*, 60 (5), 922-933.

Panaretos, V. M., & Zemel, Y. (2018). Statistical aspects of Wasserstein distances. *Annual review of statistics and its application.*

Slater, GJ, Harmon, LJ, & Alfaro, ME (2012). Integrating fossils with molecular phylogenies improves indifference of trait evolution. *Evolution*, 66 (12), 3931-3944.

Uyeda, JC, & Harmon, LJ (2014). A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic biology*, 63 (6), 902-918.

Weisberg, M. (2012). Simulation and similarity: Using models to understand the world. Oxford University Press, Oxford.

Venditti, C., Meade, A., & Pagel, M. (2011). Multiple routes to mammalian diversity. *Nature*. https://doi.org/10.1038/nature10516

.