



CONSEL: for assessing the confidence of phylogenetic tree selection

Hidetoshi Shimodaira and Masami Hasegawa

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minatoku, Tokyo 106-8569, Japan

Received on May 30, 2000; revised on June 20, 2000; accepted on July 2, 2000

ABSTRACT

Summary: CONSEL is a program to assess the confidence of the tree selection by giving the p -values for the trees. The main thrust of the program is to calculate the p -value of the Approximately Unbiased (AU) test using the multi-scale bootstrap technique. This p -value is less biased than the other conventional p -values such as the Bootstrap Probability (BP), the Kishino–Hasegawa (KH) test, the Shimodaira–Hasegawa (SH) test, and the Weighted Shimodaira–Hasegawa (WSH) test. CONSEL calculates all these p -values from the output of the phylogeny program packages such as Molphy, PAML, and PAUP*. Furthermore, CONSEL is applicable to a wide class of problems where the BPs are available.

Availability: The programs are written in C language. The source code for Unix and the executable binary for DOS are found at <http://www.ism.ac.jp/~shimo/>

Contact: shimo@ism.ac.jp

The program CONSEL consists of two parts. The first part, *makermt*, reads the matrix of the site-wise log-likelihoods of the candidate trees; the size of the matrix is the number of candidates trees by the sequence length. This matrix is produced by the phylogeny program packages based on the maximum likelihood method (Felsenstein, 1981) such as Molphy (Adachi and Hasegawa, 1996), PAML (Yang, 1997), and PAUP* (Swofford, 1998). Note however that CONSEL is applicable to the other criteria such as the parsimony, the minimum evolution, or the least squares if *makermt* is replaced by a user program as described later. Once the matrix is read, *makermt* generates the bootstrap replicates of the log-likelihoods using the RELL resampling method of Kishino *et al.* (1990) which avoids time-consuming recalculation of the maximum likelihood estimate of the tree parameters, e.g. the edge lengths, for the large number of replicates. The approximation is often accurate enough for the phylogeny analysis as indicated in the simulation of Hasegawa and Kishino (1994) or in Lemma 1 of Shimodaira (2001a), since the sequence length is very large these days.

The second part of the program, *consel*, reads the

bootstrap replicates of the log-likelihoods generated by *makermt*. It calculates several p -values for the candidate trees; the Bootstrap Probability (BP) (Felsenstein, 1985), the Kishino–Hasegawa (KH) test (Kishino and Hasegawa, 1989), the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999; Goldman *et al.*, 2000), the Weighted Shimodaira–Hasegawa (WSH) test (Shimodaira, 1993, 1998; Shimodaira and Hasegawa, 1999; Buckley *et al.*, 2001), and the Approximately Unbiased (AU) test (Shimodaira, 2000, 2001b). The difference in these p -values reflects the assumption on the extent of the selection bias as explained in Shimodaira (2001b). Once you decide which p -value to take, then the confidence set of trees is obtained by collecting the trees with the p -values not smaller than the significance level, say 0.05.

The AU test has been developed recently for general hypothesis testing of regions in the attempt to reduce the bias of the tests (Shimodaira, 2000, 2001b). This test is based on the same theory as that of Efron *et al.* (1996), but the new method comes with higher order accuracy yet with simpler implementation. The theory behind the AU test (Efron, 1985; Efron and Tibshirani, 1998) assumes the smoothness of the boundaries of the hypothesis regions; The singularity of the selection problem leads to the poor performance of the AU test as seen in the special cases, but the AU test works fine in most of the cases. When the AU test breaks down, the SH test and the WSH test are to be used safely. A self-diagnostic statistic is available for the AU test to detect the breakdown of the asymptotic theory.

The p -value of the AU test for a specific tree, or a hypothesis in general, is calculated from the multi-scale bootstrap. This is an idea very similar to the complete-and-partial bootstrap technique of Zharkikh and Li (1995). We generate K sets of bootstrap replicates with the sample sizes (i.e. the sequence lengths) N_1, \dots, N_K , and the numbers of replicates B_1, \dots, B_K . For $k = 1, \dots, K$, the k th set of the replicates is obtained by the usual non-parametric bootstrap resampling, but N_k may differ from the sample size N of the original data. For example, $K = 10$, $N_1/N = 0.5$, $N_2/N = 0.6, \dots, K$, $N_{10}/N = 1.4$, $B_1 = \dots = B_{10} = 10\,000$. Then, we count the number of

times that the tree is selected in each set of the replicates. These counts are denoted C_1, \dots, C_K . The BPs of the tree for the sequence lengths N_1, \dots, N_K are now

$$\frac{C_1}{B_1}, \dots, \frac{C_K}{B_K}.$$

The p -values is calculated from the change in the C_k/B_k along the change of N_k/N . The recipe is described in Shimodaira (2000, 2001b).

consel accepts three types of input file for the AU test:

- The `rmt` file contains the bootstrap replicates of the log-likelihoods generated by `makermt`. The multi-scale bootstrap of $K \geq 2$ is required for the AU test, and $K \geq 3$ is required for detecting the breakdown of the theory.
- The `rep` file contains the bootstrap replicates of the test statistics. The test statistic for tree- i is $T_i = \max\{Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_M\} - Y_i$, where Y_i is the log-likelihood of tree- i and M is the number of the candidate trees. The region defined by $T_i \leq 0$ corresponds to the hypothesis that tree- i is the best. Not only the p -values, but `consel` also calculates the confidence limits of the parameter corresponding to T_i . The `rep` file is produced internally in `consel` from the `rmt` file.
- The `cnt` file contains the counts C_1, \dots, C_K for all the candidate trees; $K \times M$ matrix. The other information such as $N_1/N, \dots, N_K/N$ and B_1, \dots, B_K is also included. The `cnt` file is produced internally in `consel` from the `rep` file.

We can use `consel` for a selection problem with the criteria other than the likelihood by preparing the `rmt` file; it is left for the users to write a program for generating the bootstrap replicates of the appropriate statistics though. Not only the AU test, but also the other p -values are calculated from the `rmt` file. We can also use `consel` for the test of any regions by preparing the `rep` file. If the test statistic is not available, we only have to prepare the `cnt` file; the multi-scale BPs for the hypotheses of interest. Only the p -value for the AU test is calculated from the `rep` file or the `cnt` file. The generality of the input file makes the `consel` applicable to a wide class of problems where the BPs are available.

ACKNOWLEDGEMENTS

This work was supported by Grant-in-Aids from JSPS (12780179 to HS, 12554037 to MH, HS), and from MEXT (11NP0201 to MH).

REFERENCES

- Adachi, J. and Hasegawa, M. (1996) MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monographs* 28. Institute of Statistical Mathematics, Tokyo.
- Buckley, T.R., Simon, C., Shimodaira, H. and Chambers, G.K. (2001) Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (Maoricicada) using multiple-comparison tests of tree topology. *Mol. Biol. Evol.*, **18**, 223–234.
- Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.
- Efron, B. (1985) Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, **72**, 45–58.
- Efron, B. and Tibshirani, R. (1998) The problem of regions. *Ann. Statist.*, **26**, 1687–1718.
- Efron, B., Halloran, E. and Holmes, S. (1996) Bootstrap confidence levels for phylogenetic trees. *Proc. Natl Acad. Sci. USA*, **93**, 13 429–13 434.
- Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–376.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.
- Goldman, N., Anderson, J.P. and Rodrigo, A.G. (2000) Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.*, **49**, 652–670.
- Hasegawa, M. and Kishino, H. (1994) Accuracies of the simple methods for estimating the bootstrap probability of a maximum likelihood tree. *Mol. Biol. Evol.*, **11**, 142–145.
- Kishino, H. and Hasegawa, M. (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.*, **29**, 170–179.
- Kishino, H., Miyata, T. and Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.*, **30**, 151–160.
- Shimodaira, H. (1993) A model search technique based on confidence set and map of models. *Proc. Inst. Stat. Math.*, **41**, 131–147 (in Japanese).
- Shimodaira, H. (1998) An application of multiple comparison techniques to model selection. *Ann. Inst. Stat. Math.*, **50**, 1–13.
- Shimodaira, H. (2000) Another calculation of the p -value for the problem of regions using the scaled bootstrap resamplings. *Technical Report No. 2000-35*. Stanford University.
- Shimodaira, H. (2001a) Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Commun. Stat. Part A—Theor. Meth.*, to appear.
- Shimodaira, H. (2001b) Assessing the confidence of phylogenetic tree selection, submitted.
- Shimodaira, H. and Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.*, **16**, 1114–1116.
- Swofford, D.L. (1998) *PAUP**. *Phylogenetic Analysis Using Parsimony (*and other methods)*, version 4, Sinauer, Sunderland, MA.
- Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS*, **13**, 555–556.
- Zharkikh, A. and Li, W.-H. (1995) Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Mol. Phyl. Evol.*, **4**, 44–63.