Molecular Phylogenetics and Evolution 90 (2015) 97–103

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Variance estimation for nucleotide substitution models ☆

Weishan Chen, Hsiuying Wang *

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 19 November 2014
Revised 3 May 2015
Accepted 6 May 2015
Available online 15 May 2015

Keywords:
Nucleotide substitution model
Substitution number
Variance estimator

ABSTRACT

The current variance estimators for most evolutionary models were derived when a nucleotide substitution number estimator was approximated with a simple first order Taylor expansion. In this study, we derive three variance estimators for the F81, F84, HKY85 and TN93 nucleotide substitution models, respectively. They are obtained using the second order Taylor expansion of the substitution number estimator, the first order Taylor expansion of a squared deviation and the second order Taylor expansion of a squared deviation, respectively. These variance estimators are compared with the existing variance estimator in terms of a simulation study. It shows that the variance estimator, which is derived using the second order Taylor expansion of a squared deviation, is more accurate than the other three estimators. In addition, we also compare these estimators with an estimator derived by the bootstrap method. The simulation shows that the performance of this bootstrap estimator is similar to the estimator derived by the second order Taylor expansion of a squared deviation. Since the latter one has an explicit form, it is more efficient than the bootstrap estimator.

1. Introduction

A basic process in the evolution of DNA sequences is the substitution of one nucleotide for another during evolution. But since the substitution of one allele for another in a population generally takes thousands of years or longer to complete, the process cannot be directly observed. Thus, to detect evolutionary changes in a DNA sequence, we need to estimate the nucleotide substitution number by comparing two sequences that have descended from a common ancestral sequence.

To estimate the nucleotide substitution number, nucleotide substitution number estimators had been proposed for different substitution models (Felsenstein, 1984, 1981; Hasegawa et al., 1985; Jukes and Cantor, 1969; Kimura, 1980; Tamura and Nei, 1993; Wang et al., 2008). The usual variance estimator for a nucleotide substitution number estimator was derived when this estimator was approximated with the first order Taylor expansion (Tajima and Nei, 1982; Tamura and Nei, 1993; Tateno et al., 1994). In this study, we propose variance estimators based on other Taylor expansions for the important evolutionary models, F81 model (Felsenstein, 1981), F84 model (Felsenstein, 1984), HKY85 model (Hasegawa et al., 1985) and TN93 model (Tamura and Nei, 1993), respectively. In addition to these four models, there are other important evolutionary models such as JC69 model (Jukes and Cantor, 1969) and K80 model (Kimura, 1980). Wang et al. (2008) had provided improved estimators for these two models. Wang (2011) also proposed confidence intervals for these two models.

For the related studies of these four models, Tajima and Nei (1982) proposed a substitution number estimator for the F81 model, Tateno et al. (1994) proposed a substitution number estimator and a variance estimator for the F84 model, and Tamura and Nei (1993) proposed a substitution number estimator and a variance estimator for the TN93 model. These variance estimators were derived when the nucleotide substitution number estimator was approximated with the first order Taylor expansion.

Before giving these existing variance estimators, we first introduce some notations used through the paper. Let d be the nucleotides substitution number of two sequences, p be the proportion of the sites of with different nucleotide between two sequences, S_1 be the proportion of nucleotides substitution between the two pyrimidines (i.e. $T \leftrightarrow C$), S_2 be the proportion of nucleotides substitution between the two purines (i.e. $A \leftrightarrow G$), S be the proportion of the sites of with a transitional difference, V be the proportion of nucleotides substitution between a pyrimidine and a purine. And also let X be the number of the differences between two sequences, Y be the number of the transitional differences between two sequences, W_1 be the number of substitution between the two pyrimidines for two sequences, W_2 be the number of substitution between the two purines for two sequences, W_3 be the number of the transversional differences between two sequences and n

☆ This paper has been recommended for acceptance by N. Saitou.

* Corresponding author.
E-mail addresses: wang@stat.nctu.edu.tw, hsiuyingwang@nctu.edu.tw (H. Wang).

http://dx.doi.org/10.1016/j.ympev.2015.05.003
1055-7903/

be the length of the sequences. Let π_T, π_C, π_A and π_G denote the nucleotide frequencies for T, C, A and G, respectively at a site. Also let $\pi_Y = \pi_T + \pi_C$ and $\pi_R = \pi_A + \pi_G$.

Tajima and Nei (1982) derived an estimator of the nucleotides substitution number for the F81 model

$$\hat{d}_{F81} = -2M\log\left(1-\frac{\hat{p}}{2M}\right) \tag{1}$$

where $M = \pi_T\pi_C + \pi_A\pi_G + \pi_R\pi_Y$ and $\hat{p} = X/n$.

Tateno et al. (1994) derived an estimator of the nucleotides substitution number for the F84 model

$$\hat{d}_{F84} = -2N_1\log\left(1-\frac{\hat{S}}{2N_1}-\frac{N_1-N_2}{2N_1N_3}\hat{V}\right) + 2(N_1-N_2-N_3) \times \log\left(1-\frac{\hat{V}}{2N_3}\right) \tag{2}$$

where $N_1 = \pi_T\pi_C/\pi_Y + \pi_A\pi_G/\pi_R$, $N_2 = \pi_T\pi_C + \pi_A\pi_G$, $N_3 = \pi_Y\pi_R$, $\hat{S} = Y/n$ and $\hat{V} = W_3/n$. The variance estimator of $\hat{d}_{F84}$ is

$$Var\left(\hat{d}_{F84}\right) = \frac{T_1^2\hat{S} + T_2^2\hat{V} - \left(T_1\hat{S}+T_2\hat{V}\right)^2}{n} \tag{3}$$

where

$$T_1 = \frac{N_1N_3}{N_1N_3 - N_3\hat{S}/2 - (N_1-N_2)\hat{V}/2}$$

and

$$T_2 = \frac{N_1(N_1-N_2)}{N_1N_3 - N_3\hat{S}/2 - (N_1-N_2)\hat{V}/2} - \frac{(N_1-N_2-N_3)}{N_3-\hat{V}/2} \text{ (Tateno et al., 1994)}$$

Tamura and Nei (1993) derived an estimator of the nucleotides substitution number for the TN93 model

$$\hat{d}_{TN93} = -2F_1\log\left(1-\frac{\hat{S}_1}{2F_1}-\frac{\hat{V}}{2\pi_Y}\right) - 2F_2\log\left(1-\frac{\hat{S}_2}{2F_2}-\frac{\hat{V}}{2\pi_R}\right) - 2(F_3 - F_1\pi_R - F_2\pi_Y)\log\left(1-\frac{\hat{V}}{2F_3}\right) \tag{4}$$

where $F_1 = \pi_T\pi_C/\pi_Y$, $F_2 = \pi_A\pi_G/\pi_R$, $F_3 = \pi_Y\pi_R$, $\hat{S}_1 = W_1/n$ and $\hat{S}_2 = W_2/n$. The variance estimator of $\hat{d}_{TN93}$ is

$$var(\hat{d}_{TN93}) = \frac{\left(\left(c_1^2\hat{S}_1 + c_2^2\hat{S}_2 + c_3^2\hat{V}\right) - \left(c_1\hat{S}_1 + c_2\hat{S}_2 + c_3\hat{V}\right)^2\right)}{n} \tag{5}$$

where $c_1 = \frac{\partial\hat{d}_{TN93}}{\partial S_1}(\hat{S}_1,\hat{S}_2,\hat{V})$, $c_2 = \frac{\partial\hat{d}_{TN93}}{\partial S_2}(\hat{S}_1,\hat{S}_2,\hat{V})$ and $c_3 = \frac{\partial\hat{d}_{TN93}}{\partial V}(\hat{S}_1,\hat{S}_2,\hat{V})$ (Tamura and Nei, 1993). For the HKY 85 model, by our derivation, we obtain the same form of the nucleotide substitution number estimator as the TN93 model, and therefore, it has the same variance estimator (5) as the TN93 model.

These variance estimators (3) and (5) were derived when the nucleotide substitution number was approximated with the first order Taylor expansion. In this study, we propose using the second order Taylor expansion of the substitution number estimator, the first order Taylor expansion of a squared deviation and the second order Taylor expansion of a squared deviation to derive the variance estimators for these models. From a simulation study, the variance estimator derived by the second order Taylor expansion of a squared deviation is shown to be more efficient than those derived by the other methods. The codes for calculating these estimators can be accessed at http://www.stat.nctu.edu.tw/hwang/Variance.htm.

We also compare these variance estimators with the estimator derived by the bootstrap method (Nei and Kumar, 2000). The simulation study shows that this bootstrap estimator has a similar performance as the estimator derived by the second order Taylor expansion of a squared deviation. Nevertheless, the bootstrap estimator is more time-consuming than this estimator which has an explicit form. Therefore, the estimator derived by the second order Taylor expansion of a squared deviation is commended. However, in other applications when we need variance and covariance estimators for the distance-based phylogenetic inference based on multiple partitions, the bootstrap method is still very useful.

2. Results

We first provide the variance estimators of the four models, F81, HKY85, F84 and TN93. For each model, we give four variance estimators $V_1(\hat{d})$, $V_2(\hat{d})$, $V_3(\hat{d})$ and $V_4(\hat{d})$, where $\hat{d}$ is the estimated nucleotide substitution number for the model. The estimators $V_1(\hat{d})$ and $V_2(\hat{d})$ are obtained by approximating the substitution number estimator $\hat{d}$ with the first order and the second order Taylor expansion, respectively. And the variances $V_3(\hat{d})$ and $V_4(\hat{d})$ are obtained by approximating a squared deviation $(\hat{d}-E(\hat{d}))^2$ with the first order and the second order Taylor expansion, respectively. The estimator $V_1(\hat{d})$ is the existing variance estimator, and the estimators $V_2(\hat{d})$, $V_3(\hat{d})$ and $V_4(\hat{d})$ are derived in this study. The followings are the variance estimators for the four models.

2.1. F81 model

$$V_1(\hat{d}_{F81}) = \frac{\hat{p}(1-\hat{p})}{n\left(1-\frac{\hat{p}}{2M}\right)^2}, \tag{6}$$

$$V_2(\hat{d}_{F81}) = \frac{\hat{p}(1-\hat{p})}{n\left(1-\frac{\hat{p}}{2M}\right)^2} - \frac{\hat{p}(1-\hat{p})(1-2\hat{p})}{2n^2M\left(1-\frac{\hat{p}}{2M}\right)^3} + \frac{\hat{p}^2(1-\hat{p})^2}{8n^2M^2\left(1-\frac{\hat{p}}{2M}\right)^4}, \tag{7}$$

$$V_3(\hat{d}_{F81}) = \frac{\hat{p}^2(1-\hat{p})^2}{16n^2M^2\left(1-\frac{\hat{p}}{2M}\right)^4}, \tag{8}$$

and

$$V_4(\hat{d}_{F81}) = \frac{\hat{p}(1-\hat{p})}{n\left(1-\frac{\hat{p}}{2M}\right)^2} - \frac{\hat{p}^2(1-\hat{p})^2}{16n^2M^2\left(1-\frac{\hat{p}}{2M}\right)^4}. \tag{9}$$

2.2. F84 model

$$V_1(\hat{d}_{F84}) = n\hat{S}(1-\hat{S})\hat{M}_1^2 + n\hat{V}(1-\hat{V})\hat{M}_2^2 - 2n\hat{S}\hat{V}\hat{M}_1\hat{M}_2, \tag{10}$$

$$\begin{aligned}V_2(\hat{d}_{F84}) = {} & n\hat{S}(1-\hat{S})\hat{M}_1^2 + n\hat{V}(1-\hat{V})\hat{M}_2^2 \\ & + \frac{n^2\hat{S}^2(1-\hat{S})^2}{4}\hat{M}_3^2 + \frac{n^2\hat{V}^2(1-\hat{V})^2}{4}\hat{M}_4^2 - 2n\hat{S}\hat{V}\hat{M}_1\hat{M}_2 \\ & + \hat{T}_1\hat{M}_1\hat{M}_3 + \hat{T}_2\hat{M}_2\hat{M}_4 + \hat{T}_3(\hat{M}_1\hat{M}_4 + 2\hat{M}_2\hat{M}_5) \\ & + \hat{T}_4(\hat{M}_2\hat{M}_3 + 2\hat{M}_1\hat{M}_5) + \frac{1}{4}\hat{T}_5\hat{M}_3^2 + \frac{1}{4}\hat{T}_6\hat{M}_4^2 \\ & + \left(\hat{T}_7 + n^2\hat{S}\hat{V}^2(1-\hat{V})\right)\hat{M}_4\hat{M}_5 + \left(\hat{T}_8 + n^2\hat{S}^2(1-\hat{S})\hat{V}\right)\hat{M}_3\hat{M}_5 \\ & + (\hat{T}_9 - n^2\hat{S}^2\hat{V}^2)\hat{M}_5^2 + \frac{1}{2}\left(\hat{T}_9 - n^2\hat{S}(1-\hat{S})\hat{V}(1-\hat{V})\right)\hat{M}_3\hat{M}_4,\end{aligned} \tag{11}$$

$$V_3(\hat{d}_{F84}) = (\hat{M}_6 - \hat{M}_7)^2 \tag{12}$$

and

$$V_4(\hat{d}_{F84}) = (\hat{M}_6 - \hat{M}_7)^2 + n\hat{S}(1-\hat{S})(\hat{M}_1^2 + \hat{M}_3\hat{M}_6 - \hat{M}_3\hat{M}_7) + n\hat{V}(1-\hat{V})(\hat{M}_2^2 + \hat{M}_4\hat{M}_6 - \hat{M}_4\hat{M}_7) - 2n\hat{S}\hat{V}(\hat{M}_1\hat{M}_2 + \hat{M}_5\hat{M}_6 - \hat{M}_5\hat{M}_7), \tag{13}$$

where

$$\hat{M}_1 = \frac{1}{n\left(1 - \frac{\hat{S}}{2N_1} - \frac{(N_1-N_2)\hat{V}}{2N_1N_3}\right)},$$

$$\hat{M}_2 = \frac{\frac{N_1-N_2}{N_3 n}}{1 - \frac{\hat{S}}{2N_1} - \frac{(N_1-N_2)\hat{V}}{2N_1N_3}} - \frac{\frac{N_1-N_2-N_3}{N_3 n}}{1 - \frac{\hat{V}}{2N_3}},$$

$$\hat{M}_3 = \frac{1}{2N_1 n^2\left(1 - \frac{\hat{S}}{2N_1} - \frac{(N_1-N_2)\hat{V}}{2N_1N_3}\right)^2},$$

$$\hat{M}_4 = \frac{\frac{(N_1-N_2)^2}{2N_1N_3^2n^2}}{\left(1 - \frac{\hat{S}}{2N_1} - \frac{(N_1-N_2)\hat{V}}{2N_1N_3}\right)^2} - \frac{\frac{N_1-N_2-N_3}{2N_3^2n^2}}{\left(1 - \frac{\hat{V}}{2N_3}\right)^2},$$

$$\hat{M}_5 = \frac{\frac{N_1-N_2}{2N_1N_3n^2}}{\left(1 - \frac{\hat{S}}{2N_1} - \frac{(N_1-N_2)\hat{V}}{2N_1N_3}\right)^2},$$

$$\hat{M}_6 = -2N_1 \log\left(1 - \frac{\hat{S}}{2N_1} - \frac{(N_1-N_2)\hat{V}}{2N_1N_3}\right) + 2(N_1 - N_2 - N_3) \times \log\left(1 - \frac{\hat{V}}{2N_3}\right),$$

$$\hat{M}_7 = \hat{M}_6 + \frac{n\hat{S}(1-\hat{S})}{2}\hat{M}_3 + \frac{n\hat{V}(1-\hat{V})}{2}\hat{M}_4 - n\hat{S}\hat{V}\hat{M}_5,$$

$$\hat{T}_1 = n\hat{S}(1-\hat{S})(1-2\hat{S}),$$

$$\hat{T}_2 = n\hat{V}(1-\hat{V})(1-2\hat{V}),$$

$$\hat{T}_3 = n\hat{S}\hat{V}(2\hat{V}-1),$$

$$\hat{T}_4 = n\hat{S}\hat{V}(2\hat{S}-1),$$

$$\hat{T}_5 = 3n(n-2)S^4 - 6n(n-2)\hat{S}^3 + n(3n-7)\hat{S}^2 + n\hat{S},$$

$$\hat{T}_6 = 3n(n-2)\hat{V}^4 - 6n(n-2)\hat{V}^3 + n(3n-7)\hat{V}^2 + n\hat{V},$$

$$\hat{T}_7 = 3n^2\hat{S}\hat{V}^2(\hat{V}-1) + n(-6\hat{S}\hat{V}^3 + 6\hat{S}\hat{V}^2 + \hat{S}\hat{V}),$$

$$\hat{T}_8 = 3n^2\hat{S}^2\hat{V}(\hat{S}-1) + n(-6\hat{S}^3\hat{V} + 6\hat{S}^2\hat{V} + \hat{S}\hat{V})$$

and

$$\hat{T}_9 = n^2\hat{S}\hat{V}(3\hat{S}\hat{V} - \hat{S} - \hat{V} + 1) + n\hat{S}\hat{V}(-6\hat{S}\hat{V} + 2\hat{S} + 2\hat{V} - 1).$$

2.3. TN93 model

$$V_1(\hat{d}_{TN93}) = \sum_{i=1}^{3}({}_1\hat{H}_i)\hat{K}_i^2 + 2\sum_{1\leqslant i<j\leqslant 3}({}_2\hat{H}_{ij})\hat{K}_i\hat{K}_j, \tag{14}$$

$$\begin{aligned}V_2(\hat{d}_{TN93}) = {} & \sum_{i=1}^{3}({}_1\hat{H}_i)\hat{K}_i^2 + \sum_{i=1}^{3}({}_3\hat{H}_i)\hat{K}_i\hat{K}_{ii} + \frac{1}{4}\left(({}_6\hat{H}_i) - ({}_1\hat{H}_i)^2\right)\hat{K}_{ii}^2 \\ & + \sum_{1\leqslant i<j\leqslant 3} 2({}_2\hat{H}_{ij})\hat{K}_i\hat{K}_j + \left(({}_8\hat{H}_{ij}) - ({}_2\hat{H}_{ij})^2\right)\hat{K}_{ij}^2 \\ & + \frac{1}{2}(({}_8\hat{H}_{ij}) - ({}_1\hat{H}_i)({}_1\hat{H}_j))\hat{K}_{ii}\hat{K}_{jj} \\ & + \sum_{1\leqslant i,j\leqslant 3, i\neq j}({}_4\hat{H}_{ij})(\hat{K}_i\hat{K}_{jj} + 2\hat{K}_j\hat{K}_{ij}) + ({}_7\hat{H}_{ij} - ({}_1\hat{H}_j)({}_2\hat{H}_{ij}))\hat{K}_{jj}\hat{K}_{ij} \\ & + \sum_{\substack{1\leqslant i\leqslant 3, 1\leqslant j<k\leqslant 3 \\ i\neq j\neq k}} 2({}_5\hat{H}_{ijk})\hat{K}_i\hat{K}_{jk} + ({}_9\hat{H}_{ijk})(\hat{K}_{ii}\hat{K}_{jk} + 2\hat{K}_{ij}\hat{K}_{ik}) \\ & - ({}_1\hat{H}_i)({}_2\hat{H}_{jk})\hat{K}_{ii}\hat{K}_{jk} - 2({}_2\hat{H}_{ij})({}_2\hat{H}_{ik})\hat{K}_{ij}\hat{K}_{ik},\end{aligned} \tag{15}$$

$$V_3(\hat{d}_{TN93}) = (\hat{H}_{10} - \hat{H}_{11})^2 \tag{16}$$

and

$$V_4(\hat{d}_{TN93}) = (\hat{H}_{10} - \hat{H}_{11})^2 + \sum_{i=1}^{3}({}_1\hat{H}_i)(\hat{K}_i^2 + (\hat{H}_{10} - \hat{H}_{11})\hat{K}_{ii}) + 2\sum_{1\leqslant i<j\leqslant 3}({}_2\hat{H}_{ij})(\hat{K}_i\hat{K}_j + (\hat{H}_{10} - \hat{H}_{11})\hat{K}_{ij}) \tag{17}$$

where

$$\hat{K}_1 = \frac{1}{n\left(1 - \frac{\hat{S}_1}{2F_1} - \frac{\hat{V}}{2\pi_Y}\right)},$$

$$\hat{K}_2 = \frac{1}{n\left(1 - \frac{\hat{S}_2}{2F_2} - \frac{\hat{V}}{2\pi_R}\right)},$$

$$\hat{K}_3 = \frac{\frac{F_1}{\pi_Y n}}{1 - \frac{\hat{S}_1}{2F_1} - \frac{\hat{V}}{2\pi_Y}} + \frac{\frac{F_2}{\pi_R n}}{1 - \frac{\hat{S}_2}{2F_2} - \frac{\hat{V}}{2\pi_R}} + \frac{\frac{F_3 - F_1\pi_R - F_2\pi_Y}{F_3 n}}{1 - \frac{\hat{V}}{2F_3}},$$

$$\hat{K}_{11} = \frac{\frac{1}{2F_1n^2}}{\left(1 - \frac{\hat{S}_1}{2F_1} - \frac{\hat{V}}{2\pi_Y}\right)^2},$$

$$\hat{K}_{22} = \frac{\frac{1}{2F_2n^2}}{\left(1 - \frac{\hat{S}_2}{2F_2} - \frac{\hat{V}}{2\pi_R}\right)^2},$$

$$\hat{K}_{33} = \frac{\frac{F_1}{2\pi_Y^2n^2}}{\left(1 - \frac{\hat{S}_1}{2F_1} - \frac{\hat{V}}{2\pi_Y}\right)^2} + \frac{\frac{F_2}{2\pi_R^2n^2}}{\left(1 - \frac{\hat{S}_2}{2F_2} - \frac{\hat{V}}{2\pi_R}\right)^2} + \frac{\frac{F_3 - F_1\pi_R - F_2\pi_Y}{F_3^2n^2}}{\left(1 - \frac{\hat{V}}{2F_3}\right)^2},$$

$$\hat{K}_{12} = 0,$$

$$\hat{K}_{13} = \frac{\frac{1}{2\pi_Y n^2}}{\left(1 - \frac{\hat{S}_1}{2F_1} - \frac{\hat{V}}{2\pi_Y}\right)^2},$$

$$\hat{K}_{23} = \frac{\frac{1}{2\pi_R n^2}}{\left(1 - \frac{\hat{S}_2}{2F_2} - \frac{\hat{V}}{2\pi_R}\right)^2},$$

$$\hat{\mu}_1 = n\hat{S}_1,$$

$$\hat{\mu}_2 = n\hat{S}_2,$$

$$\hat{\mu}_3 = n\hat{V},$$

$${}_1\hat{H}_i = \hat{\mu}_i\left(1 - \frac{\hat{\mu}_i}{n}\right),$$

$${}_2\hat{H}_{ij} = -\frac{\hat{\mu}_i\hat{\mu}_j}{n},$$

$${}_3\hat{H}_i = \hat{\mu}_i\left(1 - \frac{\hat{\mu}_i}{n}\right)\left(1 - \frac{2\hat{\mu}_i}{n}\right),$$

$${}_4\hat{H}_{ij} = \frac{\hat{\mu}_i\hat{\mu}_j}{n}\left(\frac{2\hat{\mu}_j}{n} - 1\right),$$

$${}_5\hat{H}_{ijk} = \frac{2\hat{\mu}_i\hat{\mu}_j\hat{\mu}_k}{n^2},$$

$${}_6\hat{H}_i = \frac{3(n-2)}{n^3}\hat{\mu}_i^4 - \frac{6(n-2)}{n^2}\hat{\mu}_i^3 + \frac{3n-7}{n}\hat{\mu}_i^2 + \hat{\mu}_i,$$

$${}_7\hat{H}_{ij} = \frac{3(n-2)\hat{\mu}_i\hat{\mu}_j^2}{n^2}\left(\frac{\hat{\mu}_j}{n} - 1\right) + \frac{\hat{\mu}_i\hat{\mu}_j}{n},$$

$${}_8\hat{H}_{ij} = \frac{(n-2)\hat{\mu}_i\hat{\mu}_j}{n^2}\left(\frac{3\hat{\mu}_i\hat{\mu}_j}{n} - \hat{\mu}_i - \hat{\mu}_j\right) + \frac{(n-1)\hat{\mu}_i\hat{\mu}_j}{n},$$

$${}_9\hat{H}_{ijk} = \frac{(n-2)}{n^2}\hat{\mu}_i\hat{\mu}_j\hat{\mu}_k\left(\frac{3\hat{\mu}_i}{n} - 1\right),$$

$$\hat{H}_{10} = -2F_1\log\left(1 - \frac{\hat{S}_1}{2F_1} - \frac{\hat{V}}{2\pi_Y}\right) - 2F_2\log\left(1 - \frac{\hat{S}_2}{2F_2} - \frac{\hat{V}}{2\pi_R}\right) - 2(F_3 - F_1\pi_R - F_2\pi_Y)\log\left(1 - \frac{\hat{V}}{2F_3}\right)$$

and

$$\hat{H}_{11} = \hat{H}_{10} + \frac{1}{2}\sum_{i=1}^{3}({}_1\hat{H}_i)\hat{K}_{ii} + \sum_{1\leqslant i<j\leqslant 3}({}_2\hat{H}_{ij})\hat{K}_{ij}.$$

2.4. HKY85 model

The substitution number estimator $\hat{d}$ for the HKY85 model has the same form with the TN93 model. Hence the four variance estimators for the HKY85 model are the same as the TN93 model.

3. Simulation and discussion

We conduct a simulation study to compare the performance of these estimators. In our simulation, the sequence lengths are 500 and 1000. First, for each model, we calculate the true variances of the four models for several sets of parameter values. The method for calculating the true variances is to generate 3000 pairs of sequences for each model under given parameter values to obtain the substitution number $\hat{d}$, and then we can calculate the sample variance of the 3000 substitution numbers and use it to approximate the true variance of each model. In addition to the four variance estimators presented in the Results Section, we also compute the bootstrap variance estimator, denoted by V_5, with 1000 replications. Then we compare the five variance estimators in terms of the average means of the estimators and the mean square errors of the variance estimation. In the simulation, we generate 2000 pairs of sequences to calculate the mean and the mean square errors of each estimator from these 2000 pairs.

Tables 1–8 present the mean square errors and the average mean values of the five estimators for the four models with

Table 1
Comparison of the variance estimators for the F81 model with $n = 500$.

$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{F81})$	$V_2(\hat{d}_{F81})$	$V_3(\hat{d}_{F81})$	$V_4(\hat{d}_{F81})$	$V_5(\hat{d}_{F81})$
(0.25, 0.25, 0.25, 0.25)	0.027260	0.029373[a] (0.000432[b])	0.030281 (0.000661)	0.000575 (0.000722)	0.028799 (0.000331)	0.027627 (0.000168)
(0.2, 0.2, 0.3, 0.3)	0.028096	0.028725 (0.000221)	0.029453 (0.000257)	0.000471 (0.000764)	0.028254 (0.000201)	0.026971 (0.000157)
(0.2, 0.3, 0.3, 0.2)	0.027150	0.029897 (0.000343)	0.030766 (0.000456)	0.000554 (0.000711)	0.029343 (0.000289)	0.027445 (0.000154)
(0.1, 0.4, 0.1, 0.4)	0.025806	0.027612 (0.000281)	0.028467 (0.000345)	0.000513 (0.000641)	0.027098 (0.000248)	0.024940 (0.000128)
(0.1, 0.2, 0.3, 0.4)	0.026820	0.028315 (0.000251)	0.029112 (0.000308)	0.000495 (0.000694)	0.027820 (0.000222)	0.026485 (0.000141)

[a] Average of estimators.
[b] Mean square error.

Table 2
Comparison of the variance estimators for the F84 model with $n = 500$.

(κ, β)	$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{F84})$	$V_2(\hat{d}_{F84})$	$V_3(\hat{d}_{F84})$	$V_4(\hat{d}_{F84})$	$V_5(\hat{d}_{F84})$
(0.2, 0.8)	(0.25, 0.25, 0.25, 0.25)	0.021270	0.022778[a] (0.000168[b])	0.023420 (0.000195)	0.000772 (0.000423)	0.022006 (0.000134)	0.018772 (0.000071)
	(0.2, 0.2, 0.3, 0.3)	0.021682	0.022392 (0.000190)	0.023121 (0.000242)	0.000795 (0.000443)	0.021597 (0.000137)	0.018453 (0.000078)
	(0.1, 0.2, 0.3, 0.4)	0.020492	0.022348 (0.000249)	0.024409 (0.000312)	0.001111 (0.000514)	0.021237 (0.000132)	0.017968 (0.000080)
(0.5, 0.3)	(0.25, 0.25, 0.25, 0.25)	0.002498	0.002599 (2.188×10^{-7})	0.002619 (2.282×10^{-7})	6.390×10^{-6} (62.095×10^{-7})	0.002593 (2.153×10^{-7})	0.002118 (2.665×10^{-7})
	(0.2, 0.2, 0.3, 0.3)	0.002600	0.002611 (2.110×10^{-7})	0.002631 (2.172×10^{-7})	6.507×10^{-6} (67.256×10^{-7})	0.002605 (2.086×10^{-7})	0.002100 (3.679×10^{-7})
	(0.1, 0.2, 0.3, 0.4)	0.002535	0.002542 (2.024×10^{-7})	0.002562 (2.085×10^{-7})	6.358×10^{-6} (63.931×10^{-7})	0.002536 (2.002×10^{-7})	0.002068 (3.393×10^{-7})
(0.3, 0.5)	(0.25, 0.25, 0.25, 0.25)	0.005942	0.006136 (2.238×10^{-6})	0.006204 (2.365×10^{-6})	3.968×10^{-5} (34.834×10^{-6})	0.006096 (2.159×10^{-6})	0.005029 (1.984×10^{-6})
	(0.2, 0.2, 0.3, 0.3)	0.005979	0.006076 (2.070×10^{-6})	0.006144 (2.175×10^{-6})	3.930×10^{-5} (35.275×10^{-6})	0.006037 (2.006×10^{-6})	0.004984 (2.226×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.005889	0.005908 (2.080×10^{-6})	0.005981 (2.191×10^{-6})	3.938×10^{-5} (34.217×10^{-6})	0.005869 (2.016×10^{-6})	0.004850 (2.244×10^{-6})

[a] Average of estimators.
[b] Mean square error.

Table 3
Comparison of the variance estimators for the TN93 model with $n = 500$.

$(\alpha_1, \alpha_2, \beta)$	$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{TN93})$	$V_2(\hat{d}_{TN93})$	$V_3(\hat{d}_{TN93})$	$V_4(\hat{d}_{TN93})$	$V_5(\hat{d}_{TN93})$
(0.8, 0.3, 0.7)	(0.25, 0.25, 0.25, 0.25)	0.007908	0.008732[a] (13.621×10^{-6}[b])	0.009281 (22.395×10^{-6})	0.000448 (56.012×10^{-6})	0.008284 (9.423×10^{-6})	0.006006 (5.645×10^{-6})
	(0.2, 0.2, 0.3, 0.3)	0.008055	0.008621 (19.611×10^{-6})	0.009225 (55.698×10^{-6})	0.000469 (60.376×10^{-6})	0.008152 (9.546×10^{-6})	0.005978 (6.420×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.007853	0.008673 (15.281×10^{-6})	0.009241 (28.198×10^{-6})	0.000453 (55.503×10^{-6})	0.008220 (9.749×10^{-6})	0.005924 (5.882×10^{-6})
(1, 1.2, 0.8)	(0.25, 0.25, 0.25, 0.25)	0.025383	0.027695 (46.960×10^{-5})	0.027042 (51.787×10^{-5})	0.007218 (57.265×10^{-5})	0.016630 (9.518×10^{-5})	0.018154 (11.269×10^{-5})
	(0.2, 0.2, 0.3, 0.3)	0.024834	0.028343 (57.032×10^{-5})	0.027793 (84.672×10^{-5})	0.008005 (64.299×10^{-5})	0.020339 (8.460×10^{-5})	0.017783 (10.697×10^{-5})
	(0.1, 0.2, 0.3, 0.4)	0.022499	0.025107 (30.586×10^{-5})	0.025133 (35.131×10^{-5})	0.006295 (41.339×10^{-5})	0.018812 (6.262×10^{-5})	0.017299 (8.588×10^{-5})
(0.8, 0.5, 1)	(0.25, 0.25, 0.25, 0.25)	0.022383	0.024368 (26.162×10^{-5})	0.031927 (134.922×10^{-5})	0.006040 (41.758×10^{-5})	0.018328 (5.708×10^{-5})	0.016452 (8.175×10^{-5})
	(0.2, 0.2, 0.3, 0.3)	0.023409	0.024260 (23.515×10^{-5})	0.031998 (113.553×10^{-5})	0.006135 (43.205×10^{-5})	0.018125 (6.631×10^{-5})	0.016447 (9.777×10^{-5})
	(0.1, 0.2, 0.3, 0.4)	0.024928	0.023921 (25.041×10^{-5})	0.031636 (114.960×10^{-5})	0.005874 (49.587×10^{-5})	0.018047 (8.901×10^{-5})	0.016194 (12.562×10^{-5})

[a] Average of estimators.
[b] Mean square error.

Table 4
Comparison of the variance estimators for the HKY85 model with $n = 500$.

(α, β)	$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{HKY85})$	$V_2(\hat{d}_{HKY85})$	$V_3(\hat{d}_{HKY85})$	$V_4(\hat{d}_{HKY85})$	$V_5(\hat{d}_{HKY85})$
(0.7, 1)	(0.25, 0.25, 0.25, 0.25)	0.023146	0.023894[a] (18.343×10^{-5})[b]	0.029469 (74.730×10^{-5})	0.005676 (40.355×10^{-5})	0.018218 (5.725×10^{-5})	0.018227 (8.756×10^{-5})
	(0.2, 0.2, 0.3, 0.3)	0.024681	0.024654 (22.182×10^{-5})	0.031218 (97.244×10^{-5})	0.006385 (48.168×10^{-5})	0.018270 (7.331×10^{-5})	0.018319 (10.938×10^{-5})
	(0.1, 0.2, 0.3, 0.4)	0.023708	0.023881 (17.784×10^{-5})	0.029959 (66.218×10^{-5})	0.005738 (41.199×10^{-5})	0.018143 (6.328×10^{-5})	0.017826 (10.124×10^{-5})
(1, 0.8)	(0.25, 0.25, 0.25, 0.25)	0.020755	0.022216 (26.072×10^{-5})	0.024338 (66.278×10^{-5})	0.004487 (38.527×10^{-5})	0.017728 (5.772×10^{-5})	0.016064 (6.704×10^{-5})
	(0.2, 0.2, 0.3, 0.3)	0.019003	0.021987 (27.392×10^{-5})	0.023445 (54.291×10^{-5})	0.004490 (34.844×10^{-5})	0.017497 (5.165×10^{-5})	0.015880 (5.367×10^{-5})
	(0.1, 0.2, 0.3, 0.4)	0.018973	0.020891 (21.725×10^{-5})	0.023263 (58.876×10^{-5})	0.004185 (31.747×10^{-5})	0.016705 (4.977×10^{-5})	0.015637 (5.542×10^{-5})
(0.5, 0.7)	(0.25, 0.25, 0.25, 0.25)	0.007141	0.007060 (3.325×10^{-6})	0.007324 (4.038×10^{-6})	0.000279 (47.117×10^{-6})	0.006781 (2.815×10^{-6})	0.005858 (3.515×10^{-6})
	(0.2, 0.2, 0.3, 0.3)	0.006939	0.007075 (3.538×10^{-6})	0.007357 (4.799×10^{-6})	0.000285 (44.330×10^{-6})	0.006790 (2.768×10^{-6})	0.005861 (3.099×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.006840	0.006961 (3.519×10^{-6})	0.007248 (4.595×10^{-6})	0.000279 (43.088×10^{-6})	0.006682 (2.821×10^{-6})	0.005722 (3.481×10^{-6})

[a] Average of estimators.
[b] Mean square error.

Table 5
Comparison of the variance estimators for the F81 model with $n = 1000$.

$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{F81})$	$V_2(\hat{d}_{F81})$	$V_3(\hat{d}_{F81})$	$V_4(\hat{d}_{F81})$	$V_5(\hat{d}_{F81})$
(0.25, 0.25, 0.25, 0.25)	0.013463	0.013469[a] (1.797×10^{-5}[b])	0.013601 (1.889×10^{-5})	8.871×10^{-5} (1.789×10^{-4})	0.013380 (1.742×10^{-5})	0.012673 (1.931×10^{-5})
(0.2, 0.2, 0.3, 0.3)	0.013341	0.013215 (1.768×10^{-5})	0.013345 (1.856×10^{-5})	8.650×10^{-5} (1.757×10^{-4})	0.013128 (1.716×10^{-5})	0.012607 (2.137×10^{-5})
(0.2, 0.3, 0.3, 0.2)	0.013670	0.013416 (1.868×10^{-5})	0.013551 (1.953×10^{-5})	8.933×10^{-5} (1.844×10^{-4})	0.013327 (1.817×10^{-5})	0.012796 (2.558×10^{-5})
(0.1, 0.4, 0.1, 0.4)	0.011992	0.012680 (1.752×10^{-5})	0.012822 (1.869×10^{-5})	8.736×10^{-5} (1.417×10^{-4})	0.012593 (1.685×10^{-5})	0.012130 (2.228×10^{-5})
(0.1, 0.2, 0.3, 0.4)	0.012915	0.012998 (1.860×10^{-5})	0.013137 (1.960×10^{-5})	8.812×10^{-5} (1.645×10^{-5})	0.012910 (1.802×10^{-5})	0.012380 (2.338×10^{-5})

[a] Average of estimators.
[b] Mean square error.

Table 6
Comparison of the variance estimators for the F84 model with $n = 1000$.

(κ, β)	$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{F84})$	$V_2(\hat{d}_{F84})$	$V_3(\hat{d}_{F84})$	$V_4(\hat{d}_{F84})$	$V_5(\hat{d}_{F84})$
(0.2, 0.8)	(0.25, 0.25, 0.25, 0.25)	0.010013	0.010091[a] (10.258×10^{-6}[b])	0.010192 (10.631×10^{-6})	1.252×10^{-4} (97.780×10^{-6})	0.009966 (9.668×10^{-6})	0.008407 (8.185×10^{-6})
	(0.2, 0.2, 0.3, 0.3)	0.009397	0.010040 (9.916×10^{-6})	0.010151 (10.452×10^{-6})	1.267×10^{-4} (85.948×10^{-6})	0.009913 (9.209×10^{-6})	0.008310 (6.785×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.009407	0.009611 (10.232×10^{-6})	0.009757 (11.033×10^{-6})	1.284×10^{-4} (86.104×10^{-6})	0.009483 (9.527×10^{-6})	0.007971 (7.529×10^{-6})
(0.5, 0.3)	(0.25, 0.25, 0.25, 0.25)	0.001261	0.001291 (2.546×10^{-8})	0.001296 (2.606×10^{-8})	1.544×10^{-6} (158.573×10^{-8})	0.001290 (2.524×10^{-8})	0.001043 (6.359×10^{-8})
	(0.2, 0.2, 0.3, 0.3)	0.001258	0.001279 (2.343×10^{-8})	0.001283 (2.393×10^{-8})	1.524×10^{-6} (157.815×10^{-8})	0.001277 (2.326×10^{-8})	0.001037 (6.531×10^{-8})
	(0.1, 0.2, 0.3, 0.4)	0.001246	0.001251 (2.256×10^{-8})	0.001256 (2.291×10^{-8})	1.504×10^{-6} (154.931×10^{-8})	0.001250 (2.243×10^{-8})	0.001012 (6.910×10^{-8})
(0.3, 0.5)	(0.25, 0.25, 0.25, 0.25)	0.002968	0.002967 (2.255×10^{-7})	0.002983 (2.297×10^{-7})	8.913×10^{-6} (87.575×10^{-7})	0.002959 (2.227×10^{-7})	0.002438 (4.204×10^{-7})
	(0.2, 0.2, 0.3, 0.3)	0.002859	0.002959 (2.493×10^{-7})	0.002975 (2.573×10^{-7})	8.998×10^{-6} (81.245×10^{-7})	0.002950 (2.444×10^{-7})	0.002422 (3.259×10^{-7})
	(0.1, 0.2, 0.3, 0.4)	0.002886	0.002854 (2.143×10^{-7})	0.002871 (2.180×10^{-7})	8.814×10^{-6} (82.761×10^{-7})	0.002845 (2.121×10^{-7})	0.002364 (4.125×10^{-7})

[a] Average of estimators.
[b] Mean square error.

Table 7
Comparison of the variance estimators for the TN93 model with $n = 1000$.

$(\alpha_1, \alpha_2, \beta)$	$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{TN93})$	$V_2(\hat{d}_{TN93})$	$V_3(\hat{d}_{TN93})$	$V_4(\hat{d}_{TN93})$	$V_5(\hat{d}_{TN93})$
(0.8, 0.3, 0.7)	(0.25, 0.25, 0.25, 0.25)	0.003889	0.003967[a] (7.008×10^{-7}[b])	0.004057 (8.044×10^{-7})	8.121×10^{-5} (145.03×10^{-7})	0.003886 (6.308×10^{-7})	0.002904 (12.122×10^{-7})
	(0.2, 0.2, 0.3, 0.3)	0.003776	0.003967 (7.610×10^{-7})	0.004058 (8.942×10^{-7})	8.200×10^{-5} (145.03×10^{-7})	0.003885 (6.678×10^{-7})	0.002884 (10.227×10^{-7})
	(0.1, 0.2, 0.3, 0.4)	0.003914	0.003956 (7.403×10^{-7})	0.004048 (8.431×10^{-7})	8.091×10^{-5} (146.93×10^{-7})	0.003875 (6.730×10^{-7})	0.002836 (13.782×10^{-7})
(1, 1.2, 0.8)	(0.25, 0.25, 0.25, 0.25)	0.010436	0.011759 (9.699×10^{-5})	0.011288 (2.855×10^{-5})	0.001335 (13.32×10^{-5})	0.010424 (2.184×10^{-5})	0.008089 (1.061×10^{-5})
	(0.2, 0.2, 0.3, 0.3)	0.010402	0.011475 (13.255×10^{-5})	0.010720 (20.205×10^{-5})	0.001446 (30.00×10^{-5})	0.010028 (5.280×10^{-5})	0.008010 (1.058×10^{-5})
	(0.1, 0.2, 0.3, 0.4)	0.010008	0.010947 (57.930×10^{-6})	0.010729 (27.386×10^{-6})	0.001201 (108.827×10^{-6})	0.009746 (14.876×10^{-6})	0.007734 (9.690×10^{-6})
(0.8, 0.5, 1)	(0.25, 0.25, 0.25, 0.25)	0.010043	0.010555 (2.640×10^{-5})	0.011833 (14.274×10^{-5})	0.001048 (9.761×10^{-5})	0.009507 (1.209×10^{-5})	0.007429 (1.069×10^{-5})
	(0.2, 0.2, 0.3, 0.3)	0.009836	0.010438 (23.408×10^{-6})	0.011585 (67.759×10^{-6})	0.000973 (83.011×10^{-6})	0.009465 (10.487×10^{-6})	0.007379 (9.460×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.010499	0.010473 (1.767×10^{-5})	0.011542 (3.749×10^{-5})	0.000908 (956.55×10^{-5})	0.009565 (1.054×10^{-5})	0.007151 (1.465×10^{-5})

[a] Average of estimators.
[b] Mean square error.

sequence length 500 and 1000, respectively. To simplify the notations, we use V_i instead of $V_i(\hat{d})$ in the following discussion. First we discuss the performance of the four estimators with an explicit form, V_1, V_2, V_3 and V_4. It shows that V_4 has the smallest mean square errors for all cases and all models, followed by V_1, revealing that V_4 is the best estimator among the four estimators. Although V_1 has smaller mean square errors than the other two estimators for most cases, in a few cases, V_2 is slightly better than V_1. Nevertheless, compared with V_3, the three estimators V_1, V_2 and V_4 have more similar performance because they have significantly smaller mean square errors than V_3 in most cases. Hence for these estimators, the estimator V_4 is more accurate than the other three estimators. And then we compare the bootstrap estimator V_5 with the four estimators, it shows that the performance of the bootstrap estimator is closer to the estimator V_4 than to the other three estimators. In addition, we found that the estimator V_5 frequently underestimates the variance in many cases. Overall, since the estimator V_4 with an explicit form has a better performance than the other three estimators V_1, V_2 and V_3, and has less computational cost than the bootstrap estimator, we recommend using V_4 to estimate the variance.

4. Methods

The differences of the four models are come from their nucleotide substitution rate matrices, which are

$$Q_{F81} = \begin{bmatrix} -(\pi_C + \pi_R) & \pi_C & \pi_A & \pi_G \\ \pi_T & -(\pi_T + \pi_R) & \pi_A & \pi_G \\ \pi_T & \pi_C & -(\pi_G + \pi_Y) & \pi_G \\ \pi_T & \pi_C & \pi_A & -(\pi_A + \pi_Y) \end{bmatrix},$$

Table 8
Comparison of the variance estimators for the HKY85 model with $n = 1000$.

(α, β)	$(\pi_A, \pi_T, \pi_C, \pi_G)$	True variance	$V_1(\hat{d}_{HKY85})$	$V_2(\hat{d}_{HKY85})$	$V_3(\hat{d}_{HKY85})$	$V_4(\hat{d}_{HKY85})$	$V_5(\hat{d}_{HKY85})$
(0.7, 1)	(0.25, 0.25, 0.25, 0.25)	0.010562	0.010440[a] (18.494×10^{-6}[b])	0.011432 (68.996×10^{-6})	0.000970 (98.476×10^{-6})	0.009470 (8.480×10^{-6})	0.008184 (11.108×10^{-6})
	(0.2, 0.2, 0.3, 0.3)	0.010559	0.010297 (13.388×10^{-6})	0.011133 (26.025×10^{-6})	0.000875 (94.953×10^{-6})	0.009422 (8.640×10^{-6})	0.008133 (11.153×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.010698	0.010358 (1.759×10^{-5})	0.011334 (3.841×10^{-5})	0.000930 (9.734×10^{-5})	0.009428 (1.049×10^{-5})	0.007864 (1.262×10^{-5})
(1, 0.8)	(0.25, 0.25, 0.25, 0.25)	0.008531	0.008896 (9.294×10^{-6})	0.009142 (11.334×10^{-6})	0.000547 (64.030×10^{-6})	0.008349 (6.390×10^{-6})	0.007199 (5.239×10^{-6})
	(0.2, 0.2, 0.3, 0.3)	0.008944	0.008970 (9.196×10^{-6})	0.009218 (10.218×10^{-6})	0.000561 (70.539×10^{-6})	0.008409 (6.721×10^{-6})	0.007142 (6.445×10^{-6})
	(0.1, 0.2, 0.3, 0.4)	0.008061	0.008596 (8.195×10^{-6})	0.008875 (10.668×10^{-6})	0.000533 (56.913×10^{-6})	0.008063 (5.492×10^{-6})	0.006911 (4.504×10^{-6})
(0.5, 0.7)	(0.25, 0.25, 0.25, 0.25)	0.003368	0.003351 (3.073×10^{-7})	0.003406 (3.364×10^{-7})	5.869×10^{-5} (109.55×10^{-7})	0.003292 (2.863×10^{-7})	0.002809 (5.096×10^{-7})
	(0.2, 0.2, 0.3, 0.3)	0.003383	0.003329 (3.268×10^{-7})	0.003384 (3.524×10^{-7})	5.825×10^{-5} (110.58×10^{-7})	0.003271 (3.091×10^{-7})	0.002806 (5.274×10^{-7})
	(0.1, 0.2, 0.3, 0.4)	0.003214	0.003296 (3.430×10^{-7})	0.003354 (3.873×10^{-7})	5.799×10^{-5} (99.591×10^{-7})	0.003238 (3.082×10^{-7})	0.002749 (4.234×10^{-7})

[a] Average of estimators.
[b] Mean square error.

$$Q_{HKY85} = \begin{bmatrix} -(\alpha\pi_C + \beta\pi_R) & \alpha\pi_C & \beta\pi_A & \beta\pi_G \\ \alpha\pi_T & -(\alpha\pi_T + \beta\pi_R) & \beta\pi_A & \beta\pi_G \\ \beta\pi_T & \beta\pi_C & -(\alpha\pi_G + \beta\pi_Y) & \alpha\pi_G \\ \beta\pi_T & \beta\pi_C & \alpha\pi_A & -(\alpha\pi_A + \beta\pi_Y) \end{bmatrix},$$

$$Q_{F84} = \begin{bmatrix} -(\gamma_1\pi_C + \beta\pi_R) & \gamma_1\pi_C & \beta\pi_A & \beta\pi_G \\ \gamma_1\pi_T & -(\gamma_1\pi_T + \beta\pi_R) & \beta\pi_A & \beta\pi_G \\ \beta\pi_T & \beta\pi_C & -(\gamma_2\pi_G + \beta\pi_Y) & \gamma_2\pi_G \\ \beta\pi_T & \beta\pi_C & \gamma_2\pi_A & -(\gamma_2\pi_A + \beta\pi_Y) \end{bmatrix}$$

and

$$Q_{TN93} = \begin{bmatrix} -(\alpha_1\pi_C + \beta\pi_R) & \alpha_1\pi_C & \beta\pi_A & \beta\pi_G \\ \alpha_1\pi_T & -(\alpha_1\pi_T + \beta\pi_R) & \beta\pi_A & \beta\pi_G \\ \beta\pi_T & \beta\pi_C & -(\alpha_2\pi_G + \beta\pi_Y) & \alpha_2\pi_G \\ \beta\pi_T & \beta\pi_C & \alpha_2\pi_A & -(\alpha_2\pi_A + \beta\pi_Y) \end{bmatrix}.$$

where $\gamma_1 = (1 + \kappa/\pi_Y)\beta$, $\gamma_2 = (1 + \kappa/\pi_R)\beta$, α, α_1, α_2, β, γ_1, γ_2 and κ are constant. And the column category and row category are in the order of T, C, A and G. That is, for instance, the element Q_{F81} corresponding to the 3th row and the 2th column is the substitution rate of A to C (Yang, 2006).

We give the details of the derivations of the four estimators. To obtain the first variance estimator V_1, we use the first order Taylor expansion $\hat{d}^*$ of $\hat{d}$ to approximate $\hat{d}$, and then approximate $Var(\hat{d})$ by $Var(\hat{d}^*) = E(\hat{d}^{*2}) - (E(\hat{d}^*))^2$. To obtain the second variance estimator V_2, we use the second order Taylor expansion $\hat{d}^{**}$ of $\hat{d}$ to approximate $\hat{d}$, and then approximate $Var(\hat{d})$ by $Var(\hat{d}^{**}) = E(\hat{d}^{**2}) - (E(\hat{d}^{**}))^2$. In addition, we directly approximate variance by using the first order Taylor expansion and the second order Taylor expansion of $(\hat{d} - E(\hat{d}))^2$, and then take the expectation to obtain estimators V_3 and V_4, respectively. The details of the derivation are given in the supplementary material.

Acknowledgment

The authors thank the reviewers for helpful comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.05.003.

References

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.
Felsenstein, J., 1984. Distance methods for inferring phylogenies: a justification. Evolution 38, 16–24.
Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.
Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 21–132.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press.
Tajima, F., Nei, M., 1982. Biases of the estimates of DNA divergence obtained by the restriction enzyme technique. J. Mol. Evol. 18, 115–120.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
Tateno, Y., Takezaki, N., Nei, M., 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. Mol. Biol. Evol. 11, 261–277.
Wang, H., 2011. Confidence intervals for the substitution number in the nucleotide substitution models. Mol. Phylogenet. Evol. 60, 472–479.
Wang, H., Tzeng, Y.H., Li, W.H., 2008. Improved variance estimators for one- and two-parameter models of nucleotide substitution. J. Theor. Biol. 254, 164–167.
Yang, Z., 2006. Computational Molecular Evolution. Oxford University Press.