

## INFERRING SPECIATION RATES FROM PHYLOGENIES

SEAN NEE

Division of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, United Kingdom.  
E-mail: sean.nee@ed.ac.uk

**Abstract.**— It is possible to estimate the rate of diversification of clades from phylogenies with a temporal dimension. First, I present several methods for constructing confidence intervals for the speciation rate under the simple assumption of a pure birth process. I discuss the relationships among these methods in the hope of clarifying some fundamental theory in this area. Their performances are compared in a simulation study and one is recommended for use as a result. A variety of other questions that may, in fact, be the questions of primary interest (e.g., Has the rate of cladogenesis been declining?) are then recast as biological variants of the purely statistical question—Is the birth process model appropriate for my data? Seen in this way, a preexisting arsenal of statistical techniques is opened up for use in this area: in particular, techniques developed for the analysis of Poisson processes and the analysis of survival data. These two approaches start from different representations of the data—the branch lengths in the tree—and I explicitly relate the two. Aiming for a synoptic account of useful theory in this area, I briefly discuss some important results from the analysis of two distinct birth-death processes: the one introduced into this area by Hey (1992) is refitted with some powerful statistical tools.

**Key words.**—Adaptive radiations, macroevolution, phylogenies, speciation rates.

Received June 9, 2000. Accepted November 29, 2000.

Molecular systematics (Hillis et al. 1996) produces phylogenies that may have a temporal dimension, thus containing information about the tempo of the clade's evolution as well as the relationships among the taxa. There is, naturally, interest in extracting this information (Nee et al. 1992, 1995a,b; Harvey and Nee 1993, 1994; Nee and May 1994; Sanderson and Donohue 1994, 1996; Kubo and Iwasa 1995; Purvis et al. 1995; Zink and Slowinski 1995; Sanderson and Wojciechowski 1996; Wollenberg et al. 1996; Mooers and Heard 1997; Paradis 1997, 1998; Baldwin and Sanderson 1998; Johns and Avise 1998).

Baldwin and Sanderson (1998), for example, used molecular phylogenetics to study the rate of diversification of the Hawaiian silversword alliance, a group of 28 plant species exhibiting enormous morphological variation across a broad range of environments. They inferred a speciation rate of  $0.56 \pm 0.17$  species/million years. This rate is remarkably high, exceeding the average rates of continental radiations, and is comparable to the peak rates inferred for radiations with good fossil records, such as the Neogene horses.

To estimate the diversification rate, Baldwin and Sanderson (1998) used a simple stochastic model of clade growth that was introduced by Yule (1924), the Yule process, or the pure birth process. Arguably the simplest model, there nonetheless exists, at the very least, six different ways to derive a confidence interval for the speciation rate of a clade using this model and the likelihood approach to statistical inference. Baldwin and Sanderson identified two in the classical literature and, using the criterion of conservatism, used the wider one. I dissect the reasons for the diversity of rate estimates in the Methods section and derive some new ones; we will see in the Results that they would have been justified in using the narrower interval. Thus, one immediate benefit of the analysis presented here is tighter estimation.

Baldwin and Sanderson (1998) were primarily interested in diversification rate per se and used the Yule process to estimate it. There are many other questions that may be of interest: for example, has the rate of diversification been

changing over time, have some clades diversified faster than others, and so on? Zink and Slowinski (1995), for example, used molecular phylogenetics to examine the hypothesis that a burst of speciation over the Pleistocene epoch produced the high level of contemporary passerine diversity. In fact, they found that diversification rates *decreased* during the epoch in the 11 genera studied.

A main point of this paper is that all questions such as this one about declining rates of diversification can be seen to be variants of the purely statistical question: Is the Yule process model appropriate for my data? A statistical test of, say, the hypothesis of a slowdown in diversification rate is simply a test of a specified type of departure from the Yule process. Another purpose of this paper is to contrast two ways of representing the data in a phylogenetic tree that both allow immediate access to large bodies of existing statistical methods for addressing a great variety of questions of interest. I call these the Poisson process approach and the survival analysis approach, the latter introduced by Paradis (1997). As we will see, Paradis based his analysis on a faulty likelihood function, but that is easily corrected.

### METHODS

The following specification of the probability model relevant to inferences from phylogenies is found, to varying degrees of explicitness, in Thompson (1975), Baldwin and Sanderson (1998), and Nee and May (1994), where it is simply implicit in the derivation of their likelihood equation (20).

We assume that from the time of its origin with two lineages time  $t$  ago, the tree has grown according to a pure birth process with parameter  $\lambda$ : at each point in time each lineage has had the same probability of giving birth to a new lineage and this probability is proportional to  $\lambda$ , which controls the rate of growth of the tree. We wish to use this model to make inferences about  $\lambda$ : in the first instance, we wish to estimate it. Note that we assume that the age of the clade,  $t$ , is not a random variable, but is given by whatever historical circumstance brought about the origin of the clade. However, the

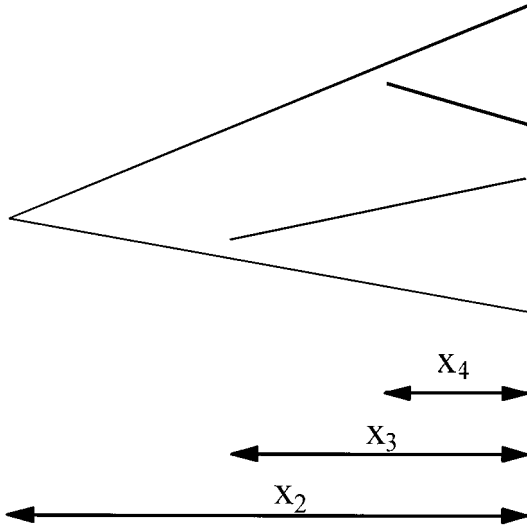


FIG. 1. A phylogenetic tree with four species. The  $x_i$  values are the times from each node to the present. I have broken the tree slightly into its branches to make it clear what, precisely, are the branches referred to in the text.

size that the clade grew to is a random variable. So, if we regrew the tree we would not expect the clade size to necessarily be the same as what we observe. This may seem obvious, but, in fact, these two assumptions will soon be abandoned.

Figure 1 illustrates a tree with four lineages at the present day. It is assumed that our data consist of the length of time from the nodes to the present day; the  $x_i$ . More generally, with  $n$  taxa at the present day, we have the vector  $\mathbf{x}$ ,  $\{x_2, x_3, x_4, \dots, x_n\}$ , which can also be interpreted as the set of branch lengths in the tree. We will also define  $t$  as the age of this monophyletic clade, the time of the last common ancestor:  $t = x_2$ .

There are two quantities that are intuitively natural and appear often in formulae, so they deserve their own symbols:

$$s = 2x_2 + \sum_{i=3}^n x_i, \quad \text{and} \quad (1)$$

$$s_r = \sum_{i=3}^n x_i. \quad (2)$$

$s$  is the sum of all the branch lengths in the tree and  $s_r$  ("reduced sum") excludes the two basal branches.

The informative data for inferences about  $\lambda$  consist of  $\mathbf{x}$  and  $n$ . Using the basic theory of the pure birth process (e.g., Grimmett and Stirzaker 1988) it is a straightforward matter to write down the probability of this data (i.e., the likelihood) given  $\lambda$  and  $t$ :

$$\Pr\{\mathbf{x}, n | \lambda, t\} = (n-1)! \lambda^{n-2} e^{-\lambda s}. \quad (3)$$

The derivation of this, along with some other relevant probability points, is discussed in Appendix 1.

What is the probability density of the  $x_i$ ? We can derive that in a standard way (e.g., Cox and Lewis 1966; Thompson

1975; Rannala 1997). The probability of  $n$  lineages, given  $\lambda$  and  $t$ , is:

$$\Pr\{n | \lambda, t\} = (n-1)e^{-2\lambda t}(1 - e^{-\lambda t})^{n-2} \quad (4)$$

(Feller 1966, XVII.3, eq. 3.5 with  $i = 2$ ). Thus, the probability of  $\mathbf{x}$  given  $n$  is:

$$\Pr\{\mathbf{x} | n, \lambda, t\} = \frac{(n-2)! \lambda^{n-2} e^{-\lambda s_r}}{(1 - e^{-\lambda t})^{n-2}}. \quad (5)$$

This is the probability density function of the order statistics of  $n-2$  independent and identically distributed random variables with truncated exponential distributions. That is, each of the  $x_i$  has the density:

$$\Pr\{X_i = x_i | \lambda, t\} = \frac{\lambda e^{-\lambda x_i}}{1 - e^{-\lambda t}}. \quad (6)$$

The significance of this will soon become apparent as we will see that some approaches to estimating  $\lambda$  assume that the  $x_i$  are exponentially distributed. Yang and Rannala (1997) provide a simulation algorithm for generating such  $\mathbf{x}$  vectors. (In fact, their algorithm is even more general, allowing the simulation of branch lengths from trees grown according to a birth-death process with possibly incomplete sampling.)

#### Confidence Intervals for $\lambda$

A general, approximate approach to constructing confidence intervals is based on the fact that, with a lot of data, the maximum likelihood estimate of  $\lambda$  is approximately normally distributed with mean equal to the true, unknown, value of  $\lambda$  and with variance equal to the inverse of the Fisher information (e.g., DeGroot 1989). As highlighted recently by Baldwin and Sanderson (1998), this fact was first used by Kendall (1949) and Moran (1951) for inferring the parameter of the pure birth process.

The maximum likelihood estimate of  $\lambda$  is, from equation (3):

$$\hat{\lambda} = \frac{n-2}{s}. \quad (7)$$

This has, appropriately, been called the Kendall-Moran estimator by Baldwin and Sanderson (1998) after those who first derived it. Although, as we will discuss below, Kendall and Moran studied somewhat different models, both models have the same maximum likelihood estimate of  $\lambda$ .

We now need the variance of this estimate. Keeping in mind the fact that  $\lambda$  is the *true* (unknown) parameter, Moran's (1951) variance is:

$$\frac{\lambda^2}{n-2}. \quad (8)$$

We will derive a 95% confidence interval from this using what is, in this case, an unnecessarily general language. The purpose of this is to make clear in a simpler context what we are doing when we come to exploit Kendall's variance.

We start with the usual first step in constructing a 95% confidence interval when the normal distribution is involved and write:

$$-1.96 < \frac{\hat{\lambda} - \lambda}{\frac{\lambda}{\sqrt{n-2}}} < 1.96, \tag{9}$$

from which we see that we next need to solve the following equations for  $\lambda$ :

$$\hat{\lambda} = \lambda \left( 1 \pm \frac{1.96}{\sqrt{n-2}} \right), \tag{10}$$

finally yielding our desired confidence interval:

$$\frac{\hat{\lambda}}{1 + \frac{1.96}{\sqrt{n-2}}} < \lambda < \frac{\hat{\lambda}}{1 - \frac{1.96}{\sqrt{n-2}}}. \tag{11}$$

Kendall (1949) derives a different variance:

$$\frac{\lambda^2}{2(e^{\lambda t} - 1)}. \tag{12}$$

Although this is the form we want to construct the confidence interval, to easily see the relationship between this and Moran's variance (eq. 8), we can rewrite it as:

$$\frac{\lambda^2}{E(n) - 2}, \tag{13}$$

where  $E(n)$  denotes the expected population size of the clade, recalling that it began with two ancestral lineages time  $t$  ago.

To construct a 95% confidence interval, we proceed as in the Moran case. However, we need to solve the following equations for  $\lambda$ , corresponding to equation (10):

$$\hat{\lambda} = \lambda \left( 1 \pm \frac{1.96}{\sqrt{2e^{\lambda t} - 2}} \right). \tag{14}$$

These have no analytical solution, so we cannot write down an explicit confidence interval in the general case. However, in any particular case we can solve these equations numerically and thus arrive at the desired confidence interval.

Why do variances (8) and (13) differ? As pointed out in Baldwin and Sanderson (1998), Moran was considering a population of processes that grew until there were exactly  $n$  lineages: so, in the likelihood (3),  $n$  is a predetermined constant and the age of the clade,  $t$  (i.e.,  $x_2$ ) is a variable. In the Moran model, the branch lengths, the elements of  $\mathbf{x}$ , are exponentially distributed. Kendall, however, was considering a population of processes that grew for a fixed time  $t$ , but to variable sizes,  $n$ . In this model, the branch lengths have truncated exponential distributions. Although the maximum likelihood estimate of  $\lambda$  happens to be the same for both populations, we are not surprised that the variances of the estimate differ in the different populations. The Kendall model seems to be the appropriate one for inference in this context, corresponding to our original specifications.

If we were willing to use the Moran model, then it is not necessary to use any approximations: We can easily derive an "exact" confidence interval for  $\lambda$ . (The quotes are intended to reflect our understanding that the confidence interval is only truly exact if the underlying model is exactly true.) We exploit the fact that an exponential random variable with parameter  $1/2$  has a chi-squared distribution with two

degrees of freedom. Let  $c_{2n,\alpha}$  be the upper  $\alpha$  point of the chi-squared distribution with  $2n$  degrees of freedom. Then it is straightforward to show that the following is an exact 95% confidence interval for  $\lambda$  under the Moran model:

$$\frac{c_{2(n-2),0.025}}{2s} < \lambda < \frac{c_{2(n-2),0.975}}{2s}. \tag{15}$$

Strictly speaking, I believe this is new to the literature, because Moran (1951) only presented the approximate confidence interval, not this exact one. However, I am sure that he assumed it would be completely obvious to his audience of statisticians and that the main interest at the time was concentrated on asymptotic likelihood theory. But, given that we now have an exact interval, there seems to be no justification for using the approximate one.

Paradis (1997) has suggested a third choice for the variance:

$$\frac{\hat{\lambda}^2}{n-2} \tag{16}$$

(cf. eqs. 8 and 13) generating another 95% confidence interval for  $\lambda$ :

$$\hat{\lambda} \left( 1 - \frac{1.96}{\sqrt{n-2}} \right) < \lambda < \hat{\lambda} \left( 1 + \frac{1.96}{\sqrt{n-2}} \right). \tag{17}$$

(Technically, Paradis is here using the observed Fisher information, rather than the expected information.)

The analysis of Paradis differs in some more subtle ways from all the others discussed in this paper. He assumes that the branch lengths, the  $x_i$ , are exponentially distributed, as opposed to truncated exponentially distributed, which means that he is studying the same hypothetical population of processes as Moran. However, Paradis offers the following maximum likelihood estimate of  $\lambda$ :

$$\hat{\lambda} = \frac{n-1}{\sum_{i=2}^n x_i}. \tag{18}$$

Here, the numerator is larger by one, and the denominator smaller by  $x_2$ , than the Kendall-Moran estimator. This discrepancy flows from the fact that Paradis, alone among researchers in this area, uses a different likelihood than equation (3). There are some interesting points raised by this discrepancy; however, they are tangential to our purpose here and therefore are discussed in Appendix 1.

Another variant on the Kendall-Moran estimate was suggested by Hey (1992). In our terms, Hey (1992) also independently studied the Moran process and derived a maximum likelihood estimate of  $\lambda$  that is slightly different than equation (7). Hey ignores the length of time between the last node in the tree and the present. This is equivalent to subtracting the quantity  $nx_n$  from  $s$ , the sum of all the branch lengths in the tree. This is an appropriate point to reveal a minor complication. For equation (3) to be the likelihood corresponding to the Moran model, it is necessary to assume that a speciation event occurred at the present day, so that  $x_n = 0$ . If this is not suitable and one wishes to use the Moran model, then Hey's form is correct. For present purposes, I will continue to assume  $x_n = 0$  when discussing the Moran model.

We have considered populations of processes that have either grown to a predetermined size (Moran 1951) or have lasted for a fixed length of time (Kendall 1949). Although the Kendall model is the appropriate one for inference, an exact treatment is not possible. An exact treatment, however, is possible for the Kendall model if we further restrict our attention to the set of processes that have lasted for a fixed time  $t$  and grown to a predetermined size  $n$ . Hence, we discard the information provided by  $n$ . Just as we have been basing our inference on  $s$  in the previous models, we now base our inference on  $s_r$  (they are sufficient statistics for  $\lambda$ ). Setting  $t = 1$ , the probability density of  $s_r$  is:

$$\begin{aligned} \text{PR}\{s_r = y\} &= \frac{\lambda^{n-2} e^{-\lambda y}}{\Gamma(n-2)(1-e^{-\lambda})^{n-2}} \\ &\times \left[ y^{n-3} + \sum_{i=1}^{n-3} (-1)^{n-2+i} \binom{n-2}{i} \right. \\ &\quad \left. \times (y-n+2+i)^{n-3} h\nu(y-n+2+i) \right], \quad (19) \end{aligned}$$

where  $h\nu(\cdot)$  is the unit Heaviside function:

$$h\nu(z) = \begin{pmatrix} 1 & (z \geq 0) \\ 0 & (z < 0) \end{pmatrix}. \quad (20)$$

Because we have set  $t = 1$ , it must be the case that  $s_r \leq n - 2$ . The derivation of the formula starts with the Laplace transform of the truncated exponential distribution and then inverts this transform raised to the power of  $n - 2$ . Although it is not at all practical to exploit equation (19) by hand, it is readily exploited in a system such as Mathematica (Wolfram 1996). Calling the distribution function corresponding to density (19)  $F(\lambda, y)$ , we generate a 95% confidence interval for  $\lambda$  by solving, for  $\lambda$ , the equations  $F(\lambda, y) = 0.025$  and  $F(\lambda, y) = 0.975$ , using the observed value of  $y$ . The add-on package, Calculus LaplaceTransform, enables Mathematica to calculate the necessary integrals involving the Heaviside function, which is actually called the UnitStep function in Mathematica.

So we can, if we wish, discard the information provided by  $n$  and, taking  $s_r$  as our statistic, construct an exact confidence interval for  $\lambda$ .

Although we have completed our discussion of inference over the natural set of processes (fixed number, fixed time, both fixed), this does not even exhaust likelihood based approaches. I will restrict myself to one other.

Denoting the logarithm of the likelihood (3) as  $l(\lambda)$ , we could use the likelihood ratio statistic

$$W(\lambda_0) = 2[l(\hat{\lambda}) - l(\lambda_0)], \quad (21)$$

which, under the null hypothesis, has approximately a chi-squared distribution with one degree of freedom. The 95% confidence interval consists of those  $\lambda$  such that  $W(\lambda) < 3.841$  Cox and Oakes (1990, ch. 3), provide a convenient summary of this, and other, tests. Although explicit solutions of the equation  $W(\lambda) = 3.841$  can be written in terms of the product-log function, this would probably not be widely helpful. The confidence interval can easily be found by graphical or nu-

merical inspection of the  $W(\lambda)$  function. This approach was used in Purvis et al. (1995) and Nee et al. (1995a). (In fact, a somewhat more complicated model of which this is a special case, discussed below, was analyzed with this method.)

Finally, it is possible that the real interest in  $\lambda$  is in the form of a hypothesis like ‘‘ $\lambda$  is less than some particular number.’’ In such a case, a good approach is to compare the observed value of  $s$  with the distribution of  $s$  under the hypothesis, as determined by simulation. One advantage of this is that it is easy to add other concerns into the simulation, for example, uncertainties about the branch lengths or departures from the assumption that speciation is instantaneous (Losos and Adler 1995).

In summary, the Kendall (1949) model corresponds to our original specifications of the correct probability model for inference, but the confidence interval it provides (eq. 14) is an approximation whose accuracy is unknown. We can get an exact interval for this model ( $F[\lambda, y] = 0.025-0.975$ ) but only by discarding the information about  $\lambda$  contained in the clade’s size. The Moran (1951) model provides us with an exact confidence interval (eq. 15) but the model assumes a fixed clade size and a randomly varying clade age, which does not seem appropriate in the present context. And both the Paradis (1997) version of the Moran model and the likelihood ratio test analysis fall outside the natural development of this topic in which we base our analysis on models in which clade size, clade age, or both are progressively fixed.

Because none of these candidate intervals presents an overwhelming case for itself, we will now compare their performances in simulation. We will grow trees for a fixed period of time to varying clade sizes and see both how often the intervals contain the true  $\lambda$  (the hit rate) and, when they do, how wide the intervals are. For 95% confidence intervals, we want a method that generates narrow intervals close to 95% of the time. All analysis was carried out in the Mathematica (Wolfram 1996) programming environment where, for example, equations such as (14) are readily solved.

## RESULTS

The likelihood ratio method ( $W[\lambda] = 3.841$ ) and the Moran method (eq. 15) were found to produce intervals that are almost indistinguishable (quantified in Appendix 2). We will, therefore, drop the former from further consideration.

For a variety of values of  $\lambda$ , 50,000 trees were generated and the various models were applied to the simulated branch lengths. The parameter  $\lambda$  was chosen to generate trees with expected sizes of 10, 20, 50, 100, and 200. The total amount of time,  $t$ , was set to one, so  $\lambda = \log_e(n/2)$ . Table 1 shows the results of this analysis. Hit rate is the proportion of 50,000 trees for which the evaluated confidence intervals for the various models contained the true value of  $\lambda$ . The Paradis (1997) model exhibits a good hit rate, but generates intervals that are considerably larger, on average, than the other two models—particularly so for smaller trees. The Kendall model has the narrowest intervals, but also the worst hit rate. The Moran model is clearly the best, with good hit rates and intervals that are not much wider than those of the Kendall model. In fact, for  $\lambda$  yielding expected clade sizes of 10 and 20, the Moran model has hit rates that are better than or as

TABLE 1. Hit rate is the proportion of 50,000 trees for which the evaluated confidence intervals for the various models contained the true value of  $\lambda$ , and interval width is the average width of the interval on those occasions when there is a hit. Different values of  $\lambda$  were chosen to generate trees of the indicated expected sizes.

	$\bar{n}$				
	10	20	50	100	200
Hit rate					
Moran	0.975	0.968	0.949	0.946	0.943
Kendall	0.981	0.939	0.896	0.883	0.879
Paradis	0.943	0.946	0.949	0.951	0.951
Interval width					
Moran	2.281	2.197	1.918	1.673	1.415
Kendall	1.902	1.85	1.638	1.419	1.179
Paradis	3.66	3.018	2.342	1.924	1.562

good as the Paradis model and also has substantially narrower intervals. The Paradis model is now dropped from further consideration.

The truncated exponential model ( $F[\lambda, y] = 0.025-0.975$ ) was applied to only 1000 simulated trees with expected sizes of 20 as the computations take a long time. Although it had a hit rate of 95%, the mean interval width was 4.8 which is substantially larger than the other three models (Table 1). (I do not understand why.) For this reason, as well as the fact

TABLE 2. (a) One method for constructing an interval is preferred to another if it both contains the true value of  $\lambda$  and is the narrowest. Cells reveal the proportion of trees for which each method was preferred. No. trees is the number of trees, out of 100,000, that satisfied the various m/k ratio criteria and to which the methods were applied.  $\lambda$  was chosen to give an expected clade size of 20. (b) as (a), but  $\lambda$  was chosen to give an expected clade size of 100.

(a)	m/k ratio			
	<0.667	<1	>1	>1.5
$n = 20$				
Preferred				
Moran	0.962	0.964	0	0
Kendall	0.035	0.023	0.913	0.916
Hit rate				
Moran	0.962	0.964	0.969	0.975
Kendall	0.997	0.978	0.913	0.916
Interval width				
Moran	1.492	1.75	2.472	2.796
Kendall	1.977	1.973	1.796	1.795
No. trees	10,961	38,160	61,083	25,715
(b)				
$n = 100$				
Preferred				
Moran	0.948	0.943	0	0
Kendall	0.031	0.023	0.832	0.809
Hit rate				
Moran	0.948	0.943	0.944	0.958
Kendall	0.978	0.951	0.832	0.809
Interval width				
Moran	1.060	1.224	2.000	2.279
Kendall	1.579	1.535	1.332	1.286
No. trees	17,488	41,622	58,584	31,952

that it is less easily applied than the others, it too was dropped from further consideration.

The performances of the Kendall and Moran models were examined in closer detail in the following fashion. As was the case for the data of Baldwin and Sanderson (1998), the ratio of their variances (eqs. 8 and 13) can differ substantially. I compared the performances of the two models when the m/k ratio took on different values, where the m/k ratio is defined as:

$$\frac{\text{Moran variance}}{\text{Kendall variance}} = \frac{2(e^{\lambda} - 1)}{n - 2}. \tag{22}$$

Notice that we are using the maximum likelihood estimate of  $\lambda$  in this formula because this is a quantity that can be calculated in practice. For a criterion such as m/k ratio < 0.66667, as happens to be the case for the data of Baldwin and Sanderson (1998), 100,000 trees of expected sizes 20 and 100 were generated and those trees that satisfied the criterion were then analyzed. A particular model was preferred if it both had the narrowest interval and contained the true  $\lambda$ .

Table 2 shows the results. For m/k ratios less than one, the Moran interval widths are substantially *smaller* than the Kendall widths. Furthermore, the hit rates of both models are good. For ratios greater than one, although the Kendall interval widths become substantially smaller than the Moran widths, this comes at the expense of a rapidly deteriorating hit rate. Thus, overall the Moran model is preferable not only on the grounds of hit rates, which we already saw in Table 1, but, for m/k ratios less than one, it also has narrower intervals. So, Baldwin and Sanderson (1998) would have been very well justified to use the Moran model.

DISCUSSION

An obvious question arises: Is the birth process model a suitable one for my data? One important circumstance in which the model is inappropriate is when  $\lambda$  varies across lineages in the phylogeny. There is considerable evidence that this is often the case (Mooers and Heard 1997).

In fact, questions such as ‘‘Is  $\lambda$  the same across lineages?’’ or ‘‘Is  $\lambda$  really constant or has it been changing over time?’’ may be of primary interest, rather than estimation of  $\lambda$  per se. Such questions can be expressed as simply asking if the birth process model is appropriate for estimating  $\lambda$ . There currently exist two broad approaches to such questions in the context of phylogenies with a temporal dimension, which, for convenience, I will call the Poisson process approach and the survival analysis approach, which share one common feature. They take as their starting point the hypothesis that the branch lengths are exponentially distributed. This hypothesis is not even true for the relevant hypothetical population of pure birth processes. But we have already seen that even the simplest goal, generating a confidence interval for  $\lambda$ , leads to approximate and nonanalytical expressions using the proper theory, and we have also seen that approaches based on the assumption of exponentiality can perform well.

*Poisson Process Approach*

This approach takes the internode distances,  $x_j - x_{j+1}$ , as the data. Assuming these to be exponentially distributed with

parameter  $j\lambda$  (Appendix 1), each such interval is first multiplied by  $j$ . Under this hypothesis, the transformed data are all drawn from an exponential distribution with parameter  $\lambda$ . This means that the growth of the tree, with the time intervals so transformed, can be viewed as a Poisson process, with the events consisting of the births of new lineages.

There are literally books filled with statistical techniques for the analysis of Poisson processes (e.g., *The Statistical Analysis of Series of Events* by Cox and Lewis 1966). This book illustrates illuminating graphical approaches to the preliminary visual inspection of the data; techniques for estimating  $\lambda$ ; tests for the analysis of any hypothesised trends in  $\lambda$  suggested by visual inspection (has it been increasing or decreasing over time); tests of the adequacy of the Poisson process model in the first place (the Markovian pure birth process in our case), in which case  $\lambda$  becomes simply a nuisance parameter that can be eliminated by appropriate conditioning; and techniques for comparing the  $\lambda$  values of different clades. This approach was used in the analysis of a closely related topic in Nee et al. (1995b) and in Zink and Slowinski (1995).

#### *Survival Analysis Approach*

Paradis (1997) recently put forward the survival analysis approach, which takes the branch lengths,  $x_j$ , as the data. Many problems can be looked at statistically from the framework of survival analysis: How long are the branch lengths? How long do my lightbulbs last? How long do my patients survive? Again, there exist book-fulls of statistical techniques for the analysis of such data (e.g., *Analysis of Survival Data* by Cox and Oakes 1990). Clearly, these two approaches are closely related, differing only in their initial perspective on the data themselves. The classic theorem relating the two is presented in Appendix 1. An advantage of this particular approach, as discussed in Paradis (1997), is that it makes for the easy analysis of censored data. For example, suppose we do not know the age of a particular lineage, but we know that it is certainly older than some particular age, then we can readily include this knowledge in our statistical analysis.

#### *Birth-death Processes*

There are currently two birth-death process models in use in this area. The first, proposed in this context by Hey (1992), assumes that a clade has always had the same number of lineages throughout its history and, so, whenever a speciation event occurs it is accompanied by an extinction. This is, if you like, a model of density-dependent cladogenesis. Obviously, estimating the speciation rate is the same as estimating the extinction rate, so perhaps it is best thought of as a turnover rate.

The second model is a natural generalization of the pure birth process and includes a constant probability of lineage extinction. For a clade to exist at all, it is necessary that the speciation rate is larger than the extinction rate. For want of a better name, let's call it the birth-death process.

#### *Hey's model*

This model is made tractable by assuming that the clade reached its constant size long before the age of the last com-

mon ancestor of the contemporary members of the clade. Hey (1992) made the additional assumption that the times between speciation events follows an exponential distribution with parameter  $n\lambda$ , where  $n$ , as usual, is the size of the clade. We take as our data the time intervals between the nodes in the tree,  $(x_j - x_{j+1})$ . Under this model, these time intervals are exponentially distributed with parameter

$$\frac{\lambda i(i+1)}{(n-1)} \quad (23)$$

(Hey 1992). The index  $i$  is understood as follows: there are  $i+1$  lineages in the phylogeny between nodes  $i$  and  $i+1$ .

We now proceed by transforming the data. Multiply each internode time interval by  $i(i+1)/(n-1)$  (Nee et al. 1994). Under the hypothesised model, our data are identically distributed exponential random variables with parameter  $\lambda$ . Denote the sum of these intervals by  $s$ . Then, assuming a speciation/extinction event at the present day, equation (15) is the exact confidence interval for  $\lambda$  under this model with the number of degrees of freedom increased to  $2(n-1)$ .

In fact, once we have transformed the data as described, the arsenal of statistical methodology referred to in the previous section becomes immediately available.

#### *Birth-death process*

Molecular phylogenies of extant taxa obviously do not contain extinct lineages. Nonetheless, extinction can leave a signature in the phylogeny, and we need to know what this is or risk making mistaken inferences. It is even true that although we are considering trees that contain only extant species, it is nonetheless possible, in principle, to estimate extinction rates,  $\mu$ , as well as speciation rates,  $\lambda$ , for clades that have grown according to the birth-death process (Nee and May 1994). We will now discuss the implications of this for the estimation of  $\lambda$ . It is more convenient and natural to refer to new parameters:  $\lambda - \mu$ , the net rate of cladogenesis, or *diversification* rate, and  $\mu/\lambda$ , the magnitude of the extinction rate relative to the speciation rate.

Estimates of the net rate of cladogenesis,  $\lambda - \mu$ , have much greater precision than estimates of  $\mu/\lambda$  (Kubo and Iwasa 1995; Nee et al. 1995a). It is also often the case that even though, on biological grounds, we know perfectly well that extinction has occurred so that, in principle, we should use a birth-death model nonetheless, in practice, the data do not demand a more complicated model than the pure birth process. However, these facts do not justify making the assumption (Paradis 1997) that we can only estimate the net rate of cladogenesis (called diversification rate by Paradis) and, so, can blithely do this using a pure birth process model. The danger inherent in such an approach is illustrated by the following example.

Figure 2 shows the results of a likelihood analysis of a molecular phylogeny of the Plethodontid salamanders constructed by Highton and Larson (1979), adapted from Nee et al. (1995a), based on the likelihood derived in Nee and May (1994) for the birth-death process model. The curves are contour lines of the likelihood surface and correspond, in order of increasing size, to one, two, and three units of support. The 95% confidence region lies somewhere in between the

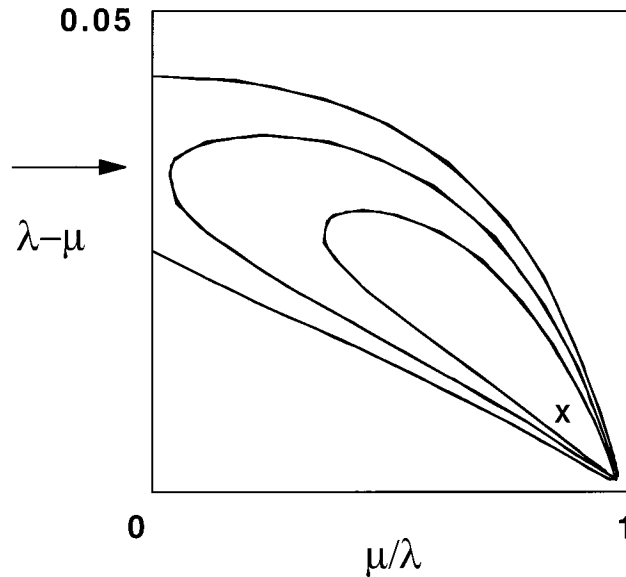


FIG. 2. Likelihood analysis of the phylogeny of the Plethodontid salamanders of Highton and Larson (1979) adapted from Nee et al. (1995a). The maximum likelihood estimate of  $\lambda - \mu$ , the net rate of cladogenesis, and  $\lambda/\mu$ , the ratio of the extinction rate to the speciation rate, is indicated by  $x$ .  $\lambda/\mu$  is, of course, dimensionless and the units of  $\lambda - \mu$  are units of genetic distance. The arrow indicates the point on the vertical axis corresponding to the maximum likelihood estimate of the rate of cladogenesis that would be returned by a pure birth process analysis.

second and third contour line. As we have said, the confidence interval for  $\mu/\lambda$  is very wide. Suppose we confine our attention to the estimation of the net rate of cladogenesis,  $\lambda - \mu$ . If we were to estimate this using the pure birth process model, our estimate would be the point on the vertical axis indicated by an arrow. However, the estimate of  $\lambda - \mu$  provided by the full model, which is the  $\lambda - \mu$  coordinate of the spot marked  $x$ , is much lower and the two estimates have non-overlapping confidence intervals.

To summarize, the adequacy of the pure birth process as a model for estimation is a statistical question, not a free assumption.

#### Booby traps

Apparent trends in diversification rate can arise for two spurious reasons. First, if the tree has grown according to either the birth-death process model or to Hey's model, then a large ratio of  $\mu/\lambda$  ( $\mu/\lambda = 1$  in Hey's model, of course) can create the spurious impression of an accelerating net rate of cladogenesis as we approach the present, even if  $\lambda$  and  $\mu$  are unchanging (Nee and May 1994). Secondly, incomplete sampling of the clade can create the impression of a *decelerating* rate of cladogenesis if the tree has grown according to the birth-death process model, but not Hey's model. This was first described in a closely related context by Slatkin and Hudson (1991) and examples can be found in Nee et al. (1995a). And, as one might have guessed, these two effects can partially cancel each other out if a clade with high  $\mu/\lambda$  is incompletely sampled (fig. 4 in Nee and May 1994). If a tree has grown according to Hey's model of density-depen-

dent cladogenesis, then the only difference incomplete sampling makes is to the width of the confidence interval: less data, less precise inference.

#### Branch length error

We have been assuming that our data consist of the true branch lengths, that is, we have not considered stochastic variation in these times arising from the tree estimation procedure itself. There are different approaches to dealing with this additional source of error. One approach is to assume that the tree (the branch lengths) is true and then do various things to see how robust one's conclusions are in the light of the particular concerns arising from the particular analysis. This can be done by computer simulation or by resampling techniques. Baldwin and Sanderson (1998) provide an illuminating example of the latter.

Another approach is to tackle the problem of estimating the speciation rate simultaneously with the problem of estimating the tree and, so, take the raw molecular data as the data for this problem as well. Interestingly, this snaps us right back to the 1960s and 1970s: In the great work on statistical approaches to tree building, associated with such names as Cavalli-Sforza, Edwards, Felsenstein, and Thompson, the pure birth process was introduced to simplify the problem of tree estimation (discussed in Thompson 1975). Recently, much progress is being made in making this sort of approach tractable for more than just tiny trees (e.g., Yang and Rannala [1997], who, in fact, use the more general birth-death process; Larget and Simon 1999). Although the focus of this literature is on getting the right tree, it is a simple change of perspective to focus instead on the speciation rate.

#### ACKNOWLEDGMENTS

I thank the University of Edinburgh for providing me with a positive research environment and the BBSRC (grant ref. G04872). E. Martins, S. Heard, and an anonymous referee made very valuable comments on a previous version of the manuscript.

#### LITERATURE CITED

- Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* 95:9402-9406.
- Cox, D. R., and P. A. A. Lewis. 1966. *The statistical analysis of series of events*. Methuen, London.
- Cox, D. R., and D. Oakes. 1990. *The analysis of survival data*. Chapman and Hall, New York.
- DeGroot, M. H. 1989. *Probability and statistics*. 2d ed. Addison-Wesley, Reading, MA.
- Edwards, A. W. F. 1972. *Likelihood*. Cambridge Univ. Press, Cambridge, U.K.
- Feller, W. 1966. *An introduction to probability theory and its applications*. Vol. II. John Wiley and Sons, New York.
- Grimmett, G. R., and D. R. Stirzaker. 1988. *Probability and random processes*. Clarendon Press, Oxford, U.K.
- Harvey, P. H., and S. Nee. 1993. New uses for new phylogenies. *Eur. Rev.* 1:11-19.
- . 1994. The molecular evolution of birds. *Biologi Italiani* 24:9-13.
- Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627-640.

- Highton, R., and L. A. Larson. 1979. The genetic relationships of the salamanders of the genus *Plethodon*. *Syst. Zool.* 28:579–599.
- Hillis, D. M., C. Moritz, and B. K. Mable. 1996. *Molecular systematics*. Sinauer, Sunderland, MA.
- Johns, G. C., and J. C. Avise. 1998. Tests for ancient species flocks based on molecular phylogenetic appraisals of *Sebastes* rockfishes and other marine fishes. *Evolution* 52:1135–1146.
- Kendall, D. G. 1949. Stochastic processes and population growth. *J. R. Statist. Soc. B* 11:230–264.
- Kubo, T., and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49:694–704.
- Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–759.
- Losos, J. B., and F. R. Adler. 1995. Stumped by trees? A generalized null model for patterns of organismal diversity. *Am. Nat.* 145:329–342.
- Moore, A. Ø., and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–54.
- Moran, P. A. P. 1951. Estimation methods for evolutive processes. *J. R. Statist. Soc. B* 13:141–146.
- Nee, S., and R. M. May. 1994. The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* 344:305–311.
- Nee, S., P. H. Harvey, and A. Ø. Mooers. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA* 89:8322–8326.
- Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1995a. Estimating extinction from molecular phylogenies. Pp. 164–182 in J. L. Lawton and R. M. May, eds. *Extinction rates*. Oxford Univ. Press, Oxford, U.K.
- Nee, S., E. C. Holmes, A. Rambaut, and P. H. Harvey. 1995b. Inferring population history from molecular phylogenies. *Phil. Trans. R. Soc. Lond. B* 349:25–31.
- Paradis, E. 1997. Assessing temporal variations in diversification rates from phylogenies: estimation and hypothesis testing. *Proc. R. Soc. Lond. B* 264:1141–1147.
- . 1998. Detecting shifts in diversification rates without fossils. *Am. Nat.* 152:176–187.
- Purvis, A., S. Nee, and P. H. Harvey. 1995. Macroevolutionary inferences from primate phylogeny. *Proc. R. Soc. Lond. B* 260:329–333.
- Rannala, B. 1997. Gene genealogy in a population of variable size. *Heredity* 78:417–423.
- Sanderson, M. J., and M. J. Donohue. 1994. Shifts in diversification rate with the origin of angiosperms. *Science* 264:1590–1593.
- . 1996. Reconstructing shifts in diversification rate on phylogenetic trees. *Trends. Ecol. Evol.* 11:15–20.
- Sanderson, M. J., and M. F. Wojciechowski. 1996. Diversification rates in a temperate legume clade: why are there so many species of *Astragalus* (Fabaceae)? *Am. J. Bot.* 83:1488–1502.
- Slatkin, M., and R. R. Hudson. 1991. Pairwise comparison of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–562.
- Thompson, E. A. 1975. *Human evolutionary trees*. Cambridge Univ. Press, Cambridge, U.K.
- Wolfram, S. 1996. *The mathematica book*. Cambridge Univ. Press, Cambridge, U.K.
- Wollenberg, K., J. Arnold, and J. C. Avise. 1996. Recognising the forest for the trees: testing temporal patterns of cladogenesis using a null model of stochastic diversification. *Mol. Biol. Evol.* 13:833–849.
- Yang, Z., and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14:717–724.
- Yule, G. U. 1924. A mathematical theory of evolution based on the conclusions of Dr J. C. Willis. *Phil. Trans. R. Soc. Lond. B* 213:21–87.
- Zink, R. M., and J. B. Slowinski. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. *Proc. Natl. Acad. Sci. USA* 92:5832–5835.

Corresponding Editor: M. Whitlock

## APPENDIX 1

The basic theory of the pure birth process informs us that, if there are  $j$  lineages in the tree at the time of a birth event, the probability of this event is proportional to  $j\lambda$ . Thus, the birth events in the tree of Figure 1, for example, contribute the term  $2\lambda 3\lambda$  to the likelihood. In general, we have the term  $(n-1)\lambda^{n-2}$ . We also know that the probability a lineage does not give birth for a length of time  $\tau$  is  $e^{-\lambda\tau}$ . Apart from the birth events, which are assumed to be instantaneous, the total amount of time that lineages are not giving birth in the tree is simply the summed branch lengths: hence the  $e^{-\lambda s}$  term in equation (3).

Instead of equation (3), the Paradis (1997) likelihood for this problem is:

$$\Pr\{\mathbf{x}, n | \lambda, t\} = \lambda^{n-1} e^{-\lambda(x_2+x_3+\dots+x_n)}, \quad (\text{A1})$$

that is, there is an additional  $\lambda$  and the exponent is missing an  $x_2$  term. What is interesting in this discrepancy is that it seems intuitive that there should only be one  $x_2$  term in the likelihood: after all, we know that there are *two* basal branches of equal length, so once we know what that length is, surely that just represents one item of information.

This is not the case, however. We know more: we know that there is a branch of length  $x_2$ , and we know that there is a *longer* branch, the one connecting the clade to the rest of the tree. Intuitively, two  $x_2$  terms appear in equation (3) because we have these two items of information. (In Paradis's terms, we are guaranteed to have one censored item of information and if we include it, we remove this particular discrepancy.) The discrepancy of the extra  $\lambda$  disappears if we condition (A1) on the fact that there was an initial bifurcation, that is, we are, indeed, looking at a tree. The incorporation of the additional  $\lambda$  in (A1) means that the reference set of birth processes includes those that did not produce a tree at all, only a single lineage. The reference set for likelihood (3) is restricted to actual trees with at least two lineages.

The following famous result concerning exponential distributions allows us to take either the branch lengths (survival analysis approach) or the internode intervals (Poisson process approach) as our data (Feller 1966): if  $X_1, X_2, \dots, X_n$  are independent exponential random variables with parameter  $\lambda$ , and  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  are their order statistics, then the  $n$  variables  $X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}$  are independent and the density of  $X_{(k+1)} - X_{(k)}$  is exponential with parameter  $(n-k)\lambda$ .

Finally, note that it is possible to base inference on  $n$  alone, if information about the branch lengths is lacking (e.g., Sanderson and Donohue 1994).

## APPENDIX 2

For some  $n$  and  $s$ , let  $(a,b)$  be an interval produced by the Moran method (eq. 15) and  $(c,d)$  be one produced by the likelihood ratio method ( $W[\lambda] = 3.841$ ). We wish to quantify the similarity of these intervals. We will express the total width that they do not overlap as a proportion of the width of the Moran interval:

$$\frac{|a-c| + |b-d|}{b-a}. \quad (\text{A2})$$

We divide by the Moran width because this makes this quantity independent of  $s$ , which greatly simplifies the analysis. Why? Writing the expression  $W[\lambda] = 3.841$  in full form, one sees that  $\lambda$  and  $s$  enter it only as the product  $\lambda s$  (which can also be seen, without algebra, by dimensional considerations). Hence, solving it for  $\lambda s$  yields  $\lambda s = F(n)$ , where  $F$  is some function of  $n$  that we need not specify here and, finally,  $\lambda = F(n)/s$ . Hence, all the elements in (A2) contain  $s$  in their denominator, which then cancels.

### Simulation Results

With  $\lambda$  chosen to yield an expected clade size of 20 and 100 species, I simulated 10,000 trees for each case and calculated (A2). In the first case, the mean value was 0.048 and the 0.25, 0.50, and 0.75 quantiles were 0.031, 0.043, and 0.06. In the second case, the mean was 0.021 and the 0.25, 0.50, and 0.75 quantiles were 0.015, 0.019, 0.025. Speaking somewhat loosely, we could summarize the results of this analysis by simply saying that we do not expect the Moran interval to differ from the likelihood ratio interval by more than 5%.