VI. Introduction to Branching Models

1 Overview

1.1 Homogeneous vs. heterogeneous models

1.1.1 Homogeneous models: all taxa governed by same rates at all times.

1.1.2 Simplicity-reality trade-off

1.1.3 Alternative risk models

- Temporal variance
- Variance among taxa
- Dependence on taxon age

**Figure 2.** Four hypothetical taxonomic survivorship curves illustrating modes of extinction. Each diagram shows the decay of a cohort of species through geologic time. A. linear survivorship (perfect adherence to Van Valen’s law). B. survivorship curve indicating early increase in extinction rate through the life of the cohort. C. linear survivorship interrupted by episodes of sudden increase in extinction. D. zero extinction rate punctuated by episodes of high extinction. The dashed lines in B, C, and D are copies of the linear survivorship curve of A.
1.2 Hierarchical vs. non-hierarchical models

1.2.1 Several models are hierarchical in the sense that they track the aggregate behavior of lower-level entities (species) within higher-level entities (clades).

1.2.2 Others more explicitly hierarchical in the sense that they include an additional process that gives rise to new clades.

1.2.3 Because we will allow the origin of new clades from existing clades, the existing clades are rendered paraphyletic—what Raup (1985, *Paleobiology* 11:42-52) referred to as paraclades.
1.3 Pure birth vs. birth-death models

1.3.1 Pure birth unrealistic, but useful for training intuition

1.3.2 For some problems (e.g. dealing with extant clades lacking fossil record) net diversification rate is of greatest interest, and pure birth model is a useful first pass.

1.4 Forward vs. inverse problems

1.4.1 Forward: prediction of system behavior, given model assumptions

1.4.2 Inverse: estimation of model parameters

1.5 Analytical approaches vs. simulation

1.5.1 Analytical approach useful when feasible

1.5.2 Simulation often necessary when model not analytically tractable

2 Pure-birth (Yule) model


2.1 Historic background

2.2 Model structure

2.2.1 Discrete time steps

2.2.2 $p=$Probability of branching per unit time; let $p^t = (1 - p)$

2.2.3 $N_t$ denotes the number of species in the clade at time $t$; $N_0 = 1$ at $t = 0$.

2.3 Model outcome

2.3.1 After one time unit:

- $\Pr\{N_1 = 1\} = p^t$
- $\Pr\{N_1 = 2\} = p$

2.3.2 After two time units:

- $\Pr\{N_2 = 1\} = p^t^2$
• There are two ways to get to $N_2 = 2$:

\[
Pr(N_2 = 2) = pp^2 \quad \text{(if } N_1 = 2) \\
+ p^2p = pp^2(p^t + 1)
\]

\[
Pr\{N_2 = 3\} = 2p^2p^t \quad \text{(if } N_1 = 2)
\]

\[
Pr\{N_2 = 4\} = pp^2 = p^3 \quad \text{(if } N_1 = 2)
\]

2.3.3 We can proceed like this, but it gets tedious. To simplify, switch to continuous time

• Let $p$ be the instantaneous rate per lineage-million-years. If one million years is subdivided into $n$ fine increments of $1/n$ m.y. each, then the probability of multiple events in a fine increment of time will be negligible. Thus, the probability of no events in 1 m.y. will be equal to $(1 - p/n)^n$. As $n \to \infty$, this converges to $e^{-p}$.

• In general, the probability that a process acting at a rate $p$ over a time $t$ will yield zero events is equal to $e^{-pt}$ (Poisson distribution).

• Thus, for the Yule process, the proportion of monotypic taxa after time $t$ is expected to be equal to $e^{-pt}$.

• Applying the same approach to $N = 2, 3, \ldots$ etc., we end up with the probability of having a standing diversity of exactly $n$ species after time $t$ (where $n \geq 0$, since there is no extinction):

\[
P_{n,t} = e^{-pt}(1 - e^{-pt})^{(n-1)}
\]

• This is one of the classic Hollow Curves.

• Note that, for any $p$ and $t$, the maximal probability is that the clade will be monotypic. The probability drops off as $n$ increases, and the curve becomes flatter as $t$ increases.

• Expectation: $E(n, t) = P_{n,t} = e^{pt}$, which is intuitively reasonable.

• Variance: $V(n, t) = e^{pt}(e^{pt} - 1)$.

• Note that the variance grows exponentially with time. This implies that an enormous range of seemingly different diversity trajectories may be probabilistically consistent with a single underlying growth parameter. This same principle holds for birth-death models.
• Note also that the variance increases with the branching rate. Again, this is also true for birth-death models. Clades with higher origination and extinction rates are expected to have greater volatility in diversity (and empirically this has been verified—Gilinsky 1994, *Paleobiology* 20:445-458). Such clades are also more likely to crash to total extinction just by chance (ammonoids as a classic example).

• Because variance (volatility) increases with the rate of the process, it is also expected to be higher at lower taxonomic levels.

From Gilinsky (1994), who measures volatility as average absolute proportional change in diversity per m.y. (see his eqn. 1).
3 Birth-death model

3.1 Model structure

3.1.1 $p$ is the instantaneous branching rate per lineage-million-years (Lmy)

3.1.2 $q$ is the instantaneous extinction rate per Lmy

3.1.3 See Appendix from Raup (1985) (reproduced below) for some of the most important predictions of this model.

(Note that Raup uses $\lambda$ and $\mu$ where we use $p$ and $q$.)

3.1.4 Of most use to us will be equations for lineage survival (A1-A4), survival of a paraclade (A11-A14), paraclade diversity (A15-A18), and total progeny (A28-A30).
Appendix

Some Equations Predicting the Fates of Lineages and Pareclades under the time-Homogeneous Model

Notation:
- \( \lambda \) = speciation events per lineage millions years (Ma)
- \( \mu \) = species extinction events per Ma
- \( t \) = time in millions of years from some arbitrary starting point
- \( s \) = number of existing lineages (species) at time \( 0 \)

If the rates of speciation and extinction are viewed as probabilities, the equations yield only the statistical expectation. The uncertainty of this estimate, in the form of the variance, is given in a few critical cases. In other cases, equations for the variance can usually be found in the source literature or constructed by the user using Monte Carlo techniques. As indicated, derivations of many of the equations can be found in the works of Kendall (1948) and Bailey (1965).

The normal approach is to search for cases where real world data conform to the predictions of the equations. The alternative is to search for genuine departures from the predictions of the model. Where it can be shown that the time-homogeneous model does not apply, constructive arguments can be developed for alternative inhomogeneous models. The inhomogeneous model is thus used as a null hypothesis that should be rejected before it is legitimate to apply the more complex models (see, for example, Raup 1983).

1) Expected fate of a single lineage:

- Probability of survival at least until time \( t \):
  \[
  P_s = e^{-\mu t} \quad (A1)
  \]
  \[
  \text{probability that a lineage will go extinct in a given time interval} \quad (t \leq t + 1):
  \]
  \[
  P_e = e^{-\mu t} - e^{-\mu (t+1)} \quad (A2)
  \]
- Mean lineage duration:
  \[
  T = \frac{1}{\mu} \quad (A3)
  \]
- Median lineage duration (half-life):
  \[
  T_m = -\left(\ln(0.5)\right)/\mu \quad (A4)
  \]

Note.—Three four equations apply whether or not time is measured from the time of origin (speciation) of the lineage. That is, the time designated as \( t = 0 \) can be any point in the history of a lineage. Thus, if Eq. (A4), for example, is applied to a lineage that has already been in existence for several million years, the time remaining for that lineage. Several of these equations have been applied in various forms to paleobiologic data by Raup (1975, 1978), Sabbe (1977), Geller and Ricklefs (1975). Eq. (A1) is the basic equation for species survivability if extinction rate is constant, and this forms the basis for Van Valen's law (Van Valen 1973; Raup 1975).

2) Probability of the ultimate extinction of a lineage (without limits of time):

- When \( \lambda < \mu \):
  \[
  P_s = 1 \quad (Bailly, eq. [8.59])
  \]
- When \( \lambda > \mu \):
  \[
  P_s = (\mu/\lambda)^t \quad (Bailly, eq. [8.59.2])
  \]

Note.—The mean value of these two equations is substantial. Together, they say that unless branching probability is higher than extinction probability, the lineage will inevitably become extinct. Even if the branching probability is higher, the lineage has a nonzero probability of becoming extinct, with the probability being estimated by Eq. (A5) and hinges on the relative magnitudes of \( \lambda \) and \( \mu \) and the number of lineages (s) at the starting point \( t = 0 \).

3) Expected extinction time of a lineage (half-life):

- When \( \lambda = \mu \) and \( s = 1 \):
  \[
  T_s = 1/\mu \quad \text{(Kendall, p. 7)}
  \]
- When \( \lambda = \mu \) and \( s > 1 \):
  \[
  T_s = (1/\lambda)/(2^{1/s} - 1) \quad \text{(Kendall, eq. (283))}
  \]
- When \( \lambda < \mu \) and \( s = 1 \):
  \[
  T_s = \frac{1}{\lambda - \mu} \ln \left(2 - \frac{\mu}{\lambda}\right) \quad \text{(Kendall, eq. (25))}
  \]

where \( \lambda < \mu \) and \( s > 1 \):

\[
T_s = \frac{1}{\mu - \lambda} \ln \left(\frac{2^{1/s} - \mu}{\mu}\right) \quad \text{(Kendall, eq. (29))}
\]

Note.—Eq. (A7) for the half-life of a lineage is algebraically identical to Eq. (A3) for the mean duration of a lineage but this is coincidental; Eq. (A9) has been applied to Paleozoic data by Raup (1978).

4) Probability of extinction at or before time \( t \):

- When \( \lambda \neq \mu \) and \( s = 1 \):
  \[
  P_e = (\lambda t)/(1 + \lambda t) \quad (A11)
  \]
  \[
  \text{(Bailey, eq. [8.53])}
  \]
- When \( \lambda = \mu \) and \( s = 1 \):
  \[
  P_e = \frac{\mu(1 - e^{-\mu t})}{\lambda(1 - e^{-\mu t})} \quad (A12)
  \]
  \[
  \text{(Bailey, eq. [8.58])}
  \]
- When \( \lambda \neq \mu \) and \( s > 1 \):
  \[
  P_e = e^{-\mu t} \quad (A14)
  \]
  \[
  \text{(Bailey, eq. [8.47])}
  \]
  \[
  \text{where} \quad \lambda = \mu \text{from eq. (A13) above.}
  \]

Note.—These four equations can be "algebraically" to provide estimates of the probability \( (P_e) \) of survival of a clad at time \( t \) as \( P_e = 1 - P_s \). As such, they provide basic equations for clad survival and have been applied to various paleobiologic problems by Raup (1978), Ricklefs (1980), and Holm (1985). Also, Raup (1983) used Eq. (A14) to assess the probability that the relapsis of the Paleozoic had values of \( \mu \) and \( \lambda \) typical of all the extinctions but were extincit by chance. Raup's was calculated using \( a = 6000 \) (estimated number of clad lineages living at a point in the Cretaceous), \( r = 20 \) (the remaining species before the end of the Permian), and values of \( \lambda \) and \( \mu \) as estimated for all Paleozoic invertebrates. The calculated value of \( P_e \) was extremely low and provided convincing evidence that the relapsis must have gone extinct by having fundamentally different cladogenetic rates from other invertebrates of the time: either lower speciation rates or higher species extinction rates.

5) Probability of a lineage will have exactly \( s \) species (where \( s \leq 1 \) at time \( t \)):

- When \( \lambda = \mu \) and \( s = 1 \):
  \[
  P_s = \frac{(\lambda t)^{s-1}}{(1 + \lambda t)^s} \quad (A15)
  \]
  \[
  \text{(Bailey, eq. [8.53])}
  \]
- When \( \lambda = \mu \) and \( s > 1 \):
  \[
  P_s = \left(\frac{1}{1 + \lambda t}\right)^s \sum_{j=0}^{s-1} \left(\frac{\lambda t}{j+1}\right)\left(\frac{1}{1 + \lambda t}\right)^{j-1} \quad (A16)
  \]
  \[
  \text{(Fanfani et al., 1991)}
  \]
- When \( \lambda \neq \mu \) and \( s = 1 \):
  \[
  P_s = (1 - a)(1 - \beta^s - 1) \quad (A17)
  \]
  \[
  \text{(Bailey, eq. [8.46])}
  \]
  \[
  \text{where} \quad \alpha = P_s \text{ from Eq. (A13) above;}
  \]
  \[
  \beta = a = \alpha/\beta \quad (A18)
  \]
  \[
  \text{(Bailey, eq. [8.47])}
  \]

Note.—These equations are potentially very useful for estimating an expected frequency distribution of parasite size and that can be used to generate artificial or predicted "bollow curves." By solving any of the equations for all possible values of \( s \), one generates a probability density distribution for parasite size that can, in effect, a prediction of the parasites' frequency distribution.

6) Expected parasite size (standing diversity) at time \( t \), including the possibility of prior extinction:
where \( \lambda = \mu \) and \( \alpha \geq 1 \):

\[
m_i = \alpha
\]  
(Bailey, eq. (8.54))  

\[
\text{variance}(m_i) = 2\alpha p
\]  
(Bailey, eq. (8.54)).

where \( \lambda < \mu \) and \( \alpha \geq 1 \):

\[
m_i = \mu e^{\lambda - \mu i}
\]  
(Bailey, eq. (8.48)).

\[
\text{variance}(m_i) = \frac{\alpha(\lambda + \mu)}{\lambda - \mu} \frac{e^{\lambda i} - e^{\mu i} - 1}{(\lambda - \mu)^2}
\]  
(Bailey, eq. (8.49)).

Notes.—Eq. (A19) indicates that if speciation and extinction rates are identical, the statistical expectation is that the number of species (standing diversity) will not change over time. But note that the variance of this estimate can be extremely high. Eq. (A21) is the fundamental growth equation described in the text.

7) Expected parasite size (standing diversity) at time \( t \), assuming that the parasite survives at least until time \( t \):

\[
m'_i = 1 + \lambda t
\]  
(A23)

\[
\text{variance} (m'_i) = \lambda t(1 + \lambda t)
\]  
(A24)

where \( \lambda = \mu \) and \( \alpha = 1 \):

\[
m'_i = \frac{\mu e^{\lambda - \mu i} - \mu}{\lambda - \mu}
\]  
(A25)

\[
\text{variance} (m'_i) = \frac{\lambda(\lambda - \mu e^{\lambda - \mu i})(1 - e^{\lambda - \mu i})}{(\lambda - \mu)^2 e^{\lambda - \mu i} - 1}
\]  
(A26)

where \( \lambda = \mu \) and \( \alpha = 1 \):

\[
m'_i = \frac{\alpha e^{\lambda - \mu i}}{1 - \alpha e^\mu}
\]  
(A27)

where \( \alpha = P_{\alpha} \) from Eq. (A13) above.

Notes.—These equations are from T. Maruyama (pers. comm. 1976). In estimating expected parasite size after a given passage of time, it is very important to decide whether it is appropriate to assume the basic survival of the parasite (Eq. (A23)- (A27)) or not (Eq. (A19)- (A22)). The answer to this question depends on the nature of the problem.

8) Estimated total "progeny," including all lineages produced (surviving and extinct) up to time \( t \), including the possibility of prior extinction of the whole parasite:

where \( \lambda = \mu \) and \( \alpha = 1 \):

\[
M'_i = \alpha(1 + \beta t)
\]  
(Kendall, p. 10)

where \( \lambda < \mu \) and \( \alpha = 1 \):

\[
M'_i = \alpha \left( \frac{e^{\lambda i} - \lambda e^{\mu i}}{\mu - \lambda} \right)
\]  
(Kendall, p. 10, eq. (45))

\[
\text{variance}(M'_i) = \text{(see Kendall eq. 47B)}
\]  
(A29)

where \( \alpha = P_{\alpha} \) from Eq. (A13) above.

Notes.—Although not yet applied to paleobiologic data, Eq. (A30) has considerable potential in answering the question, How many species have there ever been in a given biologic group? As in the earlier total progeny case (Eq. (A25)- (A27)), it is assumed that \( \alpha - \lambda \) (Eq. (A30)) was developed from a class project by Paul Wellman at the University of Rochester.
4  Discrete-time (“MBL”) simulation of clade histories

4.1 Historically extremely important in consciousness-raising about role of stochastic processes in paleontological data.

4.2 Because variance depends on magnitude of rates, it is essential that simulations be empirically scaled (see Stanley et al. (1981) *Paleobiology* 7:115).

4.3 With discrete-time modeling, time steps should be short enough that one can neglect the probability of multiple events resulting from the corresponding continuous-time process within a time step. (Practically speaking, this means that probabilities per time step should be less than or equal to about 0.1.)

5  Examples of forward problems

5.1 Stochastic survivorship of single clade


- Could trilobites have stochastically drifted to extinction (“Galton extinction”) with high initial diversity if $p = q$?

- Look for conditions that yield intermediate probability of clade extinction (too low, and it is unlikely to have happened; too high, and trilobites should have become extinct much sooner, and most other clades should also have become extinct).

- If not, how much higher would $q$ have to be to account for drift to extinction?

- Note importance of sensitivity analysis: explore wide range of parameters, not just your “best guess”.
where \( a \) is the number of coexisting species at time = 0. It is difficult to estimate \( a \) for Cambrian trilobite species described is known (RAUP, 1976) but this is of little help because (1) it is a composite of all Cambrian forms and thus does not represent standing diversity at a point in time and (2) the number found and described is surely less than the number that actually lived. One approach is to use estimates of total marine invertebrate standing diversity for the Cambrian and calculate the trilobite fraction from this. VALENTINE, et al., (1978) calculated a standing diversity for fossilizable shelled invertebrates for the Cambrian of about 8,000 species. If 75% were trilobites (above), we have an estimate of 6,000 for \( a \) in equation (1). Thus:

\[
P_o(t = 350) = \left[ \frac{\mu t}{1 + \mu t} \right]^a \approx 6.000 \left[ \frac{0.09}{1 + 0.09(350)} \right] = 4 \times 10^{-82}
\]

This result is so near zero that we can conclude with confidence that the time homogeneous model used in this way with these values will not explain the trilobite extinction. That is, the probability is negligible that the trilobites drifted to extinction. Furthermore, the value of \( P_o(t) \) is so low that minor alterations in the constants (such as reducing Cambrian diversity estimate) will not significantly affect the result.

It is instructive, however, to investigate how much the numerical situation would have to be changed to produce a \( P_o(t) \) in a reasonable range. Table 1 explores this. Equation (1) was solved for several values of \( \mu \) (expressed as its reciprocal, mean duration) and several values of \( a \). The time estimate of 350 million years was used throughout. The underlined values of \( P_o(t) \) are those that lie in a reasonable probability range. Values at or near 1.0 (upper right) are excluded in view of the fact that nine other groups were present in the Cambrian and lower diversity and did not go extinct in the Paleozoic. The reader is free to interpret Table 1. It appears to indicate that the time homogeneous model will explain the trilobite extinction only if standing diversity were much lower than has been estimated and/or mean duration of invertebrate species was much less than has been estimated. Both alternatives are conceivable but unlikely in the present state of knowledge.

<table>
<thead>
<tr>
<th>( \text{NUMBER \ OF \ SPECIES} ) AT ( t = 0 )</th>
<th>10</th>
<th>1</th>
<th>0.1</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.97</td>
<td>~1.00</td>
<td>~1.00</td>
<td>~1.00</td>
<td>~1.00</td>
</tr>
<tr>
<td>10</td>
<td>0.75</td>
<td>0.97</td>
<td>~1.00</td>
<td>~1.00</td>
<td>~1.00</td>
</tr>
<tr>
<td>100</td>
<td>0.06</td>
<td>0.75</td>
<td>0.97</td>
<td>~1.00</td>
<td>~1.00</td>
</tr>
<tr>
<td>1000</td>
<td>( 10^{-12} )</td>
<td>0.06</td>
<td>0.75</td>
<td>0.97</td>
<td>~1.00</td>
</tr>
<tr>
<td>6000</td>
<td>( 10^{-74} )</td>
<td>( 10^{-8} )</td>
<td>0.18</td>
<td>0.84</td>
<td>0.98</td>
</tr>
</tbody>
</table>
If we accept that trilobite extinction was not the result of the simple form of GALTON extinction just presented, then we can entertain more seriously the possibility that trilobites were in fact selected against compared with other marine invertebrates of the Paleozoic. The most likely expression of such selection would be a higher than normal extinction probability ($\mu$). Let us assume, therefore, that the trilobite speciation rate was the same as for other organisms ($\lambda = .09$) but that the trilobite $\mu$ was higher. How much higher would it have to have been for selective extinction to be a viable hypothesis? The time homogeneous model can be used to investigate this using the following equation for group extinction probability:

$$P_0(t) = \left[ \frac{\mu (e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} \right]$$

This equation is solved for several values of $\mu$ in Table 2, using the values of $a$ and $t$ employed in the initial calculations (above). The results shown in Table 2 indicate that if the extinction probability for trilobite species was between about 0.105 and 0.125, extinction of the whole group would be plausible. This corresponds to an average species duration which is 14 to 28 percent less than for other Phanerozoic invertebrates. It was noted in earlier analyses (RAUP, 1978a) that generic durations in the Cambrian cohort were less than that of other geologic periods and this may be because of the dominance of trilobites in this cohort.

<table>
<thead>
<tr>
<th>EXTINCTION PROBABILITY ($\mu$)</th>
<th>Equivalent species duration</th>
<th>$P_0(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.095</td>
<td>10.5 my</td>
<td>$10^{-29}$</td>
</tr>
<tr>
<td>0.1</td>
<td>10.0</td>
<td>$10^{-8}$</td>
</tr>
<tr>
<td>0.105</td>
<td>9.5</td>
<td>0.01</td>
</tr>
<tr>
<td>0.11</td>
<td>9.1</td>
<td>0.37</td>
</tr>
<tr>
<td>0.115</td>
<td>8.7</td>
<td>0.81</td>
</tr>
<tr>
<td>0.12</td>
<td>8.3</td>
<td>0.96</td>
</tr>
<tr>
<td>0.125</td>
<td>8.0</td>
<td>0.99</td>
</tr>
<tr>
<td>0.13</td>
<td>7.7</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The exercise just presented (Table 2) illustrates how the time homogeneous model can be used to evaluate the possibility of inter-group differences in extinction probabilities. Table 2 suggests that the trilobite extinction was "caused" by a higher than average species extinction probability for trilobites. The extinction is still a GALTON extinction if one considers the trilobites as a distinct entity with its own value of $\mu$. It should be emphasized that the results shown in Table 2 do not prove that the trilobite extinction occurred in this manner. The calculations only tell us how much the extinction rate for trilobites would have to depart from...
5.2 Stochastic survivorship of multiple clades

5.2.1 Relevant equations: A11-A14 from Raup 1985, plus modifications (see appended text).

5.2.2 Example: thought experiment on multiple origins of life (Raup and Valentine, 1983, PNAS 80:2981-2984).

- Look for conditions that yield intermediate probability of bioclade survival. (Too low and it’s unlikely that life would have survived at all; too high and Phanerozoic life should be polyphyletic.)
- Again note importance of exploring range of parameters.

From Raup and Valentine (1983)

Given the diversity of life that we know to have existed at the start of Phanerozoic time, what is the probability that all species were the descendents of a single founder? That is, what is the probability that all but one bioclade went extinct? Given \( D \), the number of independent origins of life, the probability, \( M_t \), of monophyly (one surviving bioclade) is:

\[ M_t = \frac{D P_x (1 - P_x)^{D-1}}{1 - (1 - P_x)^D} \]  \[ [2] \]

where \( P_x \) is defined in Eq. 1.

By solving Eq. 2 for an array of combinations of the constants, we can estimate the probability of monophyly for the group of lineages existing at some later time. If the later time is 600 Myr B.P. (latest Precambrian), \( t \) in Eq. 2 may be taken to be 3,000 Myr. As an example of the computation, let us assume that five bioclades started at 3,600 Myr B.P. and that subsequent lineages had an average duration of 100 Myr, approximately equivalent to an extinction rate, \( q \), of 0.01. Let us assume further that \( p \) is enough larger than \( q \) so that the expected species diversity in the latest Precambrian period was 1,000; this calls for a \( p \) value of 0.01177 because

\[ S_t = S_0 e^{(p-q)t} \]  \[ [3] \]

where \( S_0 \) is the number of lineages at time = 0, and \( S_t \) is the number at time = \( t \), or

\[ (p - q) = \frac{\ln S_t - \ln S_0}{t} \]  \[ [4] \]

and

\[ p = (p - q) + q. \]  \[ [5] \]

Solving Eq. 2 with these constants yields \( M_t = 0.70 \). This means that the probability is 0.70 that the 1,000 species alive at 600 Myr B.P. descended from one of the five founders of life and, thus, all belong to a single bioclade. But it also indicates a statistical expectation that 30% of the latest Precambrian species belonged to other bioclades.
From Raup and Valentine (1983)

### Table 1. The probability, $M_n$, that all latest Precambrian species belong to a single bioclade, assuming survival of life *

<table>
<thead>
<tr>
<th>Bioclades, starting no.</th>
<th>Lineage duration $(1/q)$, Myr</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>100</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.99</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.93</td>
<td>0.88</td>
<td>0.41</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.99</td>
<td>0.95</td>
<td>0.92</td>
<td>0.90</td>
<td>0.88</td>
<td>0.77</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.96</td>
<td>0.93</td>
<td>0.89</td>
<td>0.86</td>
<td>0.82</td>
<td>0.82</td>
<td>0.68</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>0.99</td>
<td>0.95</td>
<td>0.90</td>
<td>0.86</td>
<td>0.82</td>
<td>0.78</td>
<td>0.73</td>
<td>0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>0.99</td>
<td>0.94</td>
<td>0.88</td>
<td>0.83</td>
<td>0.78</td>
<td>0.73</td>
<td>0.54</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.99</td>
<td>0.93</td>
<td>0.86</td>
<td>0.80</td>
<td>0.74</td>
<td>0.69</td>
<td>0.48</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.99</td>
<td>0.92</td>
<td>0.85</td>
<td>0.78</td>
<td>0.71</td>
<td>0.65</td>
<td>0.42</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.99</td>
<td>0.91</td>
<td>0.83</td>
<td>0.75</td>
<td>0.68</td>
<td>0.62</td>
<td>0.37</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.99</td>
<td>0.90</td>
<td>0.81</td>
<td>0.73</td>
<td>0.65</td>
<td>0.58</td>
<td>0.33</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.98</td>
<td>0.82</td>
<td>0.66</td>
<td>0.53</td>
<td>0.42</td>
<td>0.33</td>
<td>0.10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.96</td>
<td>0.63</td>
<td>0.38</td>
<td>0.22</td>
<td>0.12</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.93</td>
<td>0.42</td>
<td>0.15</td>
<td>0.05</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Latest Precambrian diversity is held constant at 10,000.

### Table 2. Probability that life will survive to the latest Precambrian eon *

<table>
<thead>
<tr>
<th>Bioclades, starting no.</th>
<th>Lineage duration $(1/q)$, Myr</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>100</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.003</td>
<td>0.03</td>
<td>0.06</td>
<td>0.08</td>
<td>0.11</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.06</td>
<td>0.11</td>
<td>0.16</td>
<td>0.21</td>
<td>0.25</td>
<td>0.41</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.09</td>
<td>0.16</td>
<td>0.23</td>
<td>0.29</td>
<td>0.35</td>
<td>0.55</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.11</td>
<td>0.21</td>
<td>0.30</td>
<td>0.37</td>
<td>0.44</td>
<td>0.66</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>0.14</td>
<td>0.26</td>
<td>0.36</td>
<td>0.44</td>
<td>0.51</td>
<td>0.74</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.02</td>
<td>0.17</td>
<td>0.30</td>
<td>0.41</td>
<td>0.50</td>
<td>0.58</td>
<td>0.80</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.02</td>
<td>0.19</td>
<td>0.34</td>
<td>0.46</td>
<td>0.56</td>
<td>0.63</td>
<td>0.85</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.02</td>
<td>0.21</td>
<td>0.38</td>
<td>0.51</td>
<td>0.60</td>
<td>0.68</td>
<td>0.88</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.03</td>
<td>0.24</td>
<td>0.42</td>
<td>0.55</td>
<td>0.65</td>
<td>0.72</td>
<td>0.91</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.03</td>
<td>0.26</td>
<td>0.45</td>
<td>0.59</td>
<td>0.69</td>
<td>0.76</td>
<td>0.93</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.06</td>
<td>0.45</td>
<td>0.70</td>
<td>0.83</td>
<td>0.90</td>
<td>0.94</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.14</td>
<td>0.78</td>
<td>0.95</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.26</td>
<td>0.95</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*It is assumed that $p$ is that value which would predict a final diversity of 10,000 species.
5.3 Expected longevity of clades with variation in starting richness reflecting temporal variation in species-level rates

5.3.1 Relevant equations: A7-A10 and A15-A18 from Raup 1985

5.3.2 Increased longevities of genera that originate right after mass extinctions; may this result from high speciation rate in post-extinction world (Miller and Foote, 2003, Science 302:1030-1032)?

5.3.3 If speciation rate higher right after mass extinction, then genera will accumulate more species and will persist longer even if speciation and species extinction rates are subsequently equal.
Equations for Birth-Death Model

The probability that a genus will have \( n \) species after an elapsed time \( t \), assuming that \( n > 0 \) and \( p > q \), is given by

\[
P(n, t) = (1 - B)^{B(n-1)},
\]

where

\[
B = A \left( \frac{p}{q} \right), \quad A = q \left( \exp[(p - q)t] - 1 \right) \left( \exp[p - q)t] - q \right),
\]

and \( p \) and \( q \) are the per-capita origination and extinction rates. This approach tacitly assumes that the genera in question originate at the start of this initial interval. Given an initial standing diversity of \( n \), and assuming that speciation and species-extinction rates are equal, the median genus duration is equal to

\[
\frac{1}{n} \sum_{n=0}^{\infty} n \cdot P(n, t) = \left( \frac{1}{p} - 1 \right).
\]

Therefore, the expected median duration for a genus, assuming that it originates at the beginning of the initial time interval and is still extant at the end of it, is equal to

\[
\frac{1}{m} \sum_{m=0}^{\infty} m \cdot P(m, t) = \frac{1}{p} - 1.
\]

We focus in this analysis on the median duration because it is analytically more tractable than the mean, but similar reasoning holds for mean durations. For compatibility with our empirical results, only genera that survive to the end of their initial substages are considered.

5.4 Improbability of long-lived clades with few species under homogeneous model

(Strathmann and Slatkin, Paleobiology 9:97-106.)

5.5 Early origins of major biologic groups (Raup, 1983, Paleobiology 9:107-115) (See relevant pages from this paper, reproduced below.)

5.5.1 Branching structure of evolutionary trees predicts many extinct taxa, and just a few living ones; the living ones have deep roots.

5.5.2 Thus, the homogenous branching model predicts that randomly chosen, living species are likely to have a divergence time deep in the past.
From Raup (1983)

over race and elapsed time. The contrast between the trees in Figs. 3–4 suggests that a multitude of relatively short branches leads to a low mean divergence time (Fig. 5). As will be shown later, standing diversity is also an important element determining the divergence time distribution.

Figures 3–6 present some of the characteristics of the divergence time distribution in relation to branching trees, but they can tell us nothing about the real world because they are not scaled to approximate natural rates of turnover or real world diversity. As Stanley et al. (1981) have emphasized, scaling for diversity is often important in the analysis of the dynamics of evolutionary trees.

An Equation for Divergence Time Distribution

Assume that at some time, \( t \), before present there are \( S \) species lineages. In Fig. 3, eight such lineages exist at the bottom of the diagram. During the 90 time units leading to the top of the diagram (“Recent”), all of the starting lineages die out but their progeny persist for varying lengths of time as clusters or clades. In the case of Fig. 3, only the starting lineage on the extreme left has descendants surviving to the Recent. Thus, all living species must have divergence times less than 90 time units. In Fig. 4, on the other hand, several of the starting lineages have living descendants, so that divergence times for many living species pairs will exceed the 90 time units.

The problem is to develop an equation which will predict the expected divergence time distribution for randomly chosen pairs of living species. The relevant constants are \( S_0 \), the standing diversity at time = 0 (“Recent”); \( p \), the average rate of branching, expressed as speciation events per lineage per time unit; and \( q \), the average rate of extinction, expressed as terminations per lineage per time unit. In deriving an equation, I will use the mathematics of random branching processes and assume time homogeneity. That is, it will be assumed that \( p \) and \( q \) do not change through time. This does not prejudge the question of whether speciation and extinction rates are constant; it is only a convenient way to describe differences in branching patterns. The resulting equation could be elaborated to include changing \( p \) and \( q \) values.

Where branching rate is greater than extinction rate (\( p > q \)), standing diversity increases through time. This is the general condition in evolution over long periods of geologic time, and for evolutionary radiations in particular. If we look backward in time from the Recent, standing diversity will appear to decay. Given \( p \) and \( q \) and an assumption of time homogeneity (above), we can compute the expected value of
diversity \( S_t \) at some time \( t \) before present (BP) as follows:
\[
S_t = S_0 e^{-\rho t - \alpha t} \tag{1}
\]
It should be noted that \( S_t \) is an expected value and has a high variance where samples are small. Therefore, the estimate of past diversity, \( S_t \), in the derivation that follows is a reasonable approximation only for situations in which one is dealing with thousands or tens of thousands of species, and the reader should be cautioned that application to cases with smaller samples should be avoided.

Now let us specify two Recent species at random and imagine the evolutionary tree that produced these two (and all other surviving species). Each of the two specified species can be traced back along a unique path from lineage to lineage until the base of the tree is reached. At some point the ancestral paths of the two species will meet, and this point represents the divergence time for the species pair. The two paths will meet at a speciation event, that is, a cross-link in the tree. In each small time interval, the total number of possible cross-links is
\[
S_t C_t = \frac{S_t}{(S_t - 2)/2} - \frac{S_t(S_t - 1)}{2} \tag{2}
\]

The expected number of cross-links (speciation events) that actually occur in a given time interval is the expected diversity times the speciation rate. If we consider a time interval of 1 Myr, the expected number of speciation events, \( N \), is
\[
N = pS_t = pS_0 e^{-(\rho - \alpha)t} \tag{3}
\]

Any one of these speciation events has a probability of successfully connecting the two ancestral paths of interest, and this probability is the reciprocal of the number of possible different cross links (2). And the probability of not making a successful link is one minus the reciprocal of (2). Consequently,
\[
\text{probability of no cross-links in } \frac{n}{n} \text{ tries} = \left[1 - \frac{2}{S_t(S_t - 1)}\right]^n \tag{4}
\]
The probability of at least one successful cross-link in \( t \) successive time units is one minus the product of a series developed from (4). This may be written as follows:
\[
W_t = 1 - \prod_{i=1}^{t} \left[1 - \frac{2}{S_t(S_t - 1)}\right]^n \tag{5}
\]

In this equation, \( S \) is defined as in (1) and \( N \) as in (3). \( W_t \) is the probability that a randomly chosen pair of Recent species has a divergence time less than some time \( t \) before present. The numerical value of \( W_t \) approaches 1.0 as the base of the evolutionary tree is approached (a single ancestor for all Recent species).

Given values of the three constants \( \rho, q, \) and \( S \), the equation can be solved for a series of \( t \) values to develop the distribution of \( W_t \) from which a predicted frequency distribution of divergence times can be developed.

**Divergence Time Distributions**

The \( W_t \) distribution depends on the particular combination of \( \rho, q, \) and \( S \), but some generalizations are possible. All other things being equal, increasing diversity increases mean divergence time. Thus, although the tree in Fig. 3 yields a fairly low mean divergence time (Fig. 5), a tree with the same overall fabric (relatively low \( \rho \) and \( q \) but higher diversity would show a higher mean divergence time. This makes it especially important that divergence time computations be scaled properly for numbers of species. It also means that computer simulation is totally impractical for most real-world situations.

Increasing turnover rate (numerical values of \( \rho \) and \( q \)) decreases mean divergence time, as can be seen from a comparison of the trees in Figs. 3–4. Increasing the rate of growth of diversity (\( \rho - q \)) decreases mean divergence time by causing more of the divergence points to be concentrated higher in the tree.

The divergence time distributions shown in Figs. 5–6 are monotonic and highly skewed, but this is not typical. It is logically true that the frequency of divergence times must drop to zero as the Recent is approached (time = 0) and also go to zero as the ancestor of the entire group is approached (high \( t \)). All complete, ideal distributions are unimodal. This can be seen if simulations such as those in Figs. 3–4 are run repeatedly with the same input constants and the results summed. Whether the frequency distributions are skewed toward or away from the Recent depends on \( \rho, q, \) and \( S \).
FIGURE 3. Hypothetical evolutionary tree based on relatively high rates of species turnover. A few divergence times are indicated, such as 47 for any species pair involving clusters A and C.

FIGURE 4. Hypothetical evolutionary tree based on relatively low rates of species turnover.

FIGURE 5. Frequency distribution of divergence times for all possible pairs of surviving species in Fig. 3. Spikes 41 and 47 reflect concentrations of those divergence times in Fig. 3.

FIGURE 6. Frequency distribution of divergence times for all possible pairs of surviving species in Fig. 4. High frequency of divergence times greater than 50 reflects the fact that many of the common ancestors of surviving species existed at the base of Fig. 4.
Figure 1. First occurrences of 27 modern classes of palaeontologically important marine invertebrates. Stratigraphic range data from Sepkoski (1982).

Figure 2. First occurrences of the modern orders in the 27 classes represented in Fig. 1. Stratigraphic range data from Sepkoski (1982).

Table 1. Predicted proportion of divergence times less than 300 Myr. Equation (5) for $W_t$ is solved with an assumed modern diversity of 150,000 and various combinations of $p$ and $q$. Extinction rate ($q$) is the reciprocal of species duration, and speciation rate ($p$) is determined by the rate of diversity increase, extinction rate, and modern diversity (eq. [7] and [8]).

<table>
<thead>
<tr>
<th>Duration of species (Myr)</th>
<th>Diversity increase factor (Cambrian-Recent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td>20</td>
<td>.000</td>
</tr>
<tr>
<td>10</td>
<td>.001</td>
</tr>
<tr>
<td>5</td>
<td>.002</td>
</tr>
<tr>
<td>1</td>
<td>.004</td>
</tr>
<tr>
<td>0.5</td>
<td>.007</td>
</tr>
<tr>
<td>0.2</td>
<td>.014</td>
</tr>
<tr>
<td></td>
<td>.034</td>
</tr>
</tbody>
</table>

Table 2. Generalized numbers of intraclass species pairs for modern classes of marine shelf invertebrates (limited to groups generally recognized as fossilizable).

<table>
<thead>
<tr>
<th>Classes</th>
<th>Mean specie per class</th>
<th>Total species</th>
<th>Total intraclass pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>25,000</td>
<td>50,000</td>
<td>$6.24 \times 10^8$</td>
</tr>
<tr>
<td>15,000</td>
<td></td>
<td></td>
<td>$2.24 \times 10^8$</td>
</tr>
<tr>
<td>5,500</td>
<td>49,500</td>
<td></td>
<td>$1.36 \times 10^8$</td>
</tr>
<tr>
<td>14,000</td>
<td>36,000</td>
<td></td>
<td>$1.75 \times 12^8$</td>
</tr>
<tr>
<td>31,000</td>
<td>63,000</td>
<td></td>
<td>$9.86 \times 12^8$</td>
</tr>
</tbody>
</table>

Total species pairs (intra- and interclass): $9.32 \times 10^9$

Propportion of intraclass pairs: 0.136
Propportion of interclass pairs: 0.864
From *Principles of Paleontology 3/e* (forthcoming)

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Number of species</th>
<th>Within-phylum pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annelida</td>
<td>11,600</td>
<td>6.72×10⁴</td>
</tr>
<tr>
<td>Arthropoda</td>
<td>50,000</td>
<td>1.25×10⁷</td>
</tr>
<tr>
<td>Brachiopoda</td>
<td>325</td>
<td>5.26×10⁴</td>
</tr>
<tr>
<td>Bryozoa</td>
<td>5,000</td>
<td>1.25×10⁴</td>
</tr>
<tr>
<td>Chordata</td>
<td>45,000</td>
<td>1.31×10⁷</td>
</tr>
<tr>
<td>Cnidaria</td>
<td>9,000</td>
<td>4.05×10⁷</td>
</tr>
<tr>
<td>Echinodermata</td>
<td>6,000</td>
<td>1.30×10⁷</td>
</tr>
<tr>
<td>Mollusca</td>
<td>50,000</td>
<td>1.25×10⁷</td>
</tr>
<tr>
<td>Porifera</td>
<td>5,000</td>
<td>1.25×10⁷</td>
</tr>
</tbody>
</table>

Total species: 181,925
Total within-phylum pairs: 3.56×10⁹
Total species pairs: 1.55×10¹⁰
Total between-phylum pairs: 1.29×10¹⁰

Percent of pairs within phyla: 22%
Percent of pairs between phyla: 78%

**SOURCES OF DATA:** Valentine (2004); Barnes, Calow, and Olive (1993). Species count for arthropods excludes terrestrial arachnids and insects.

### 5.5.3 Within- vs. between-group pairs

- Let there be $g$ groups with $N_i$ species, $i = 1, ..., g$.
- Within-group pairs: $\sum_{i=1}^{g} N_i(N_i - 1)/2$.
- Between-group pairs: $\sum_{i,j(i\neq j)} N_i N_j$.

### 5.5.4 Sensitivity

- Lower rates lead to deeper divergence.
- Higher net diversification rate leads to shallower divergences.
- Thus, single large group with recent radiation (e.g. insects) can greatly skew empirical results; this would represent clear deviation from homogeneous model.