Random Walks, Evolutionary Trends, and Branching Models of Diversification

GEOL 305 - PALEOBIOLOGY

One of the aims of paleobiology is to document the history of life on Earth and determine the processes that influenced that history. This lab focuses on the identification (or not) of evolutionary trends in the fossil record. Trends reflect change in a consistent direction through time and are often interpreted as indicators that some driving mechanism was influencing how organisms evolved. Identifying trends is an important part of understanding changes in past life on Earth.

In order to test for the statistical significance of trends, paleobiologists generate artificial patterns using random numbers. If real patterns cannot be distinguished from randomly generated patterns, then perhaps scientists should focus their efforts on explaining those aspects of the history of life that are more statistically robust. At this point, it is necessary to make clear that randomness in this context does not mean that any particular example of evolution does not reflect biological processes: apparent randomness in a time series of data points can arise as a result of the complex interaction of many processes. In addition, a trend may be random in one aspect (e.g., directionality) but non-random in another (e.g., autocorrelation) and these may not have been distinguished when the pattern was originally described.

The aim of this lab is to give you an idea of how paleobiologists make use of stochastic or Monte Carlo (i.e., random) models to test for the significance of pattern in the fossil record. The first part is a general exploration of random walks and time series, and the second part uses a stochastic branching model to test for randomness in the diversity history of post-Paleozoic echinoids.

Part 1. Random Walks and Trends

In order to get a handle on the nature of randomness, you will start with a very simple random walk. Here are the rules for generating a random walk through time: At the start, let your variable have a value of 0. From this point, let’s say that the value can either go up by 1 or down by 1 with equal probability in each timestep. To figure out which way it goes, use a random number table. A random number table is simply a list of numbers (usually four digits long); they are random because the value of each consecutive number cannot be predicted based on knowledge of the previous values. Consult the table: If the next random number in the list is between 0000 and 4999 inclusive, let the variable decrease by 1. If the next random number in the list is between 5000 and 9999 inclusive, let the variable increase by 1. This means that the value has an equal chance of going up or down at any timestep. If you repeat this procedure for many timesteps, you will generate a random walk – i.e., a record of how the value of the variable changed through time.

Imagine that the variable was a measurement of the average size of an appendage on a trilobite or the number of ribs on a brachiopod shell and that each timestep represents 100,000 or 1,000,000 years. If a real trend in the fossil record resembles a random walk, then it may not be prudent to infer the action of an evolutionary process.

In order to explore the degree to which random walks can resemble directional trends, you will generate 11 random walks of 80 steps. Your first walk should use the random numbers in Table1. Plot your random walk on a piece of graph paper as you generate it. For the next 10 random walks, you will use the last four digits of telephone numbers from the South Central Texas Region telephone book. (Although one can make an argument that the people listed in a telephone book are not a random cross-section of society, it is hard to make the case that the order of numbers in a telephone listing is not random.) Be careful to avoid obviously non-random strings of numbers like the list of extensions for a business (these are often sequential and therefore not random). Plot all your random walks on one sheet of paper. You should distinguish them with different symbols or use a different colored pencil for each one.

Table 1. 80 Random Numbers

8449	8340	9934	9614	2134	7772	4540	2925
5834	1543	3832	4077	8509	5750	1118	4620
1058	2534	959	7961	7885	2741	4497	7943
9298	1962	1454	2725	3868	8748	5080	4697
5576	1504	7666	2864	149	3373	1799	2923
2223	9341	267	9998	5323	3196	6823	3824
8842	5271	296	460	6751	4391	1338	5578
7178	8556	7165	739	1033	8299	1206	6218
992	9391	4688	2306	9527	8568	4774	9890
5655	2842	2948	5565	2146	6339	1583	1117

Use your results to answer the following questions:

1) Do all random walks end in the same place (more or less)? What is the range of final values for your random walks?
2) Are the final values for the random walks also the most extreme (minimum or maximum) values? Is there a limit on how extreme the variable can be in any given timestep?

Part 2. Branching Models of Diversity

The procedure you used in the first part is a little too simple for most biological systems, but should have given you the idea of how stochastic modeling can be used to generate paleontological trends. The model for the second part is based on time-homogeneous, branching diversification. Here's the idea: biotic lineages (species or higher taxa) split and end through time. Splitting represents speciation or origination; ending represents extinction (Fig. 1).

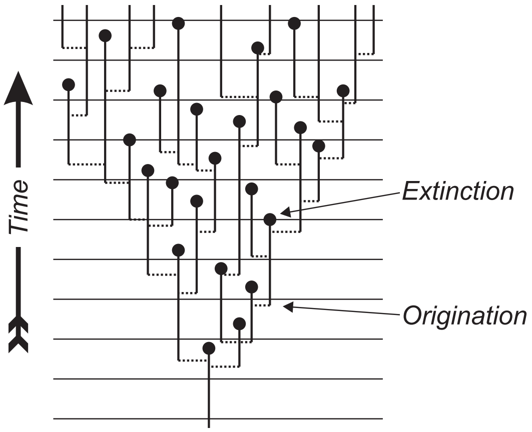

Figure 1.

Assume that all lineages have an equal chance of splitting in each increment of time ("time-homogeneous" means that rates do not change through time). The same goes for ending. If we know the number of lineages, and the rates of origination and extinction, then we can predict the number of taxa at any given point in time by using the exponential growth equation (Fig. 2):

$N_t = N_o e^{(\lambda-\mu)t}$

N_t = number of taxa at time t
λ = origination rate constant
μ = extinction rate constant
N_o = initial number of taxa.

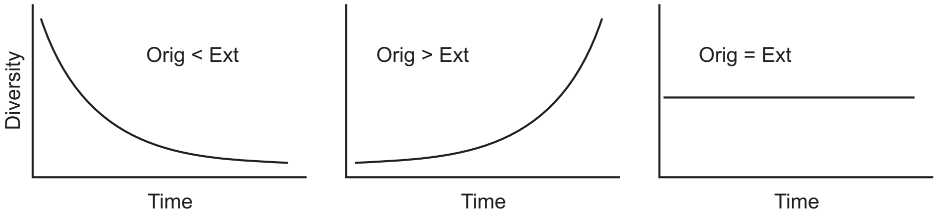

Figure 2.

Deviations from this pattern indicate changes in rates of origination and extinction through time.

However, comparison of data to the real world requires that we have some sort of error bars, which the deterministic equation does not provide. To get these, we will use random numbers to build a set of hypothetical trees using the time-homogeneous assumptions. If the real world lies outside the range of variation, then we must reject the time-homogeneous assumptions and start searching for other influences on the rates of evolution.

The tree shown in figure 3 represents the phylogenetic history of Euechinoid orders (echinoids are sea biscuits, sand dollars, and their relatives). *Diversity* is simply the number of lineages present at any one time. Draw the diversity curve of this tree using the times marked along the side. When counting diversity, include only the solid lines that cross time lines (the dotted lines are inferred and therefore do not represent data describing the real world). If a lineage starts at a time line, <u>do</u> count it. The curve generated through this exercise shows the diversity of sea urchins and their relatives through time.

From the real tree, we can count the number of extinctions and the number of originations and work out the probability of an event at a moment in time. This is where the telephone book comes into play. Start with one founding species (exactly like the real tree at the Permian-Triassic boundary). If the last four digits of a phone number go from 0000 to 0295 then an extinction event has occurred in a 10 Myr interval in a single lineage (this is equivalent to a lineage ending 2.95 out of every 100 times tried). If the last four digits of the number go from 0296 to 1405 then an origination event occurs (11.09 times out of every 100). If the number is greater than 1405, then nothing happens to that lineage in the 10 Myr interval examined. Do this for each lineage in each 10 Myr time increment from the beginning to the end. If all your lineages die out before the present day, start again until you produce a tree that has at least one survivor to the Recent (this is not likely after the first one or two timesteps). Draw your tree as a branching diagram and then calculate the equivalent diversity curve.

Repeat this procedure 10 times and plot all 10 diversity curves on the same graph as your true diversity curve. When all 10 diversity curves are drawn on the same sheet of paper, you get a messy cloud of intertwined lines. This is good. Comparing a real diversity curve to a single stochastic curve is of limited value – maybe the difference is just random. However, because you can generate many artificial curves, you can explore a range of random possibilities. If a real curve falls outside the cloud, we can say that it is significantly different from the artificial lines. In this case, we this can say with 90% confidence because with 10 artificial curves, we have a

1/10 chance that at least one will resemble the real curve. Since the model trees were generated based on the null hypothesis (i.e. constant rates of extinction and origination through time), if the real Euechinoid curve falls outside the cloud, we can reject the null and something more must have been going on to influence the observed diversity history of the group.

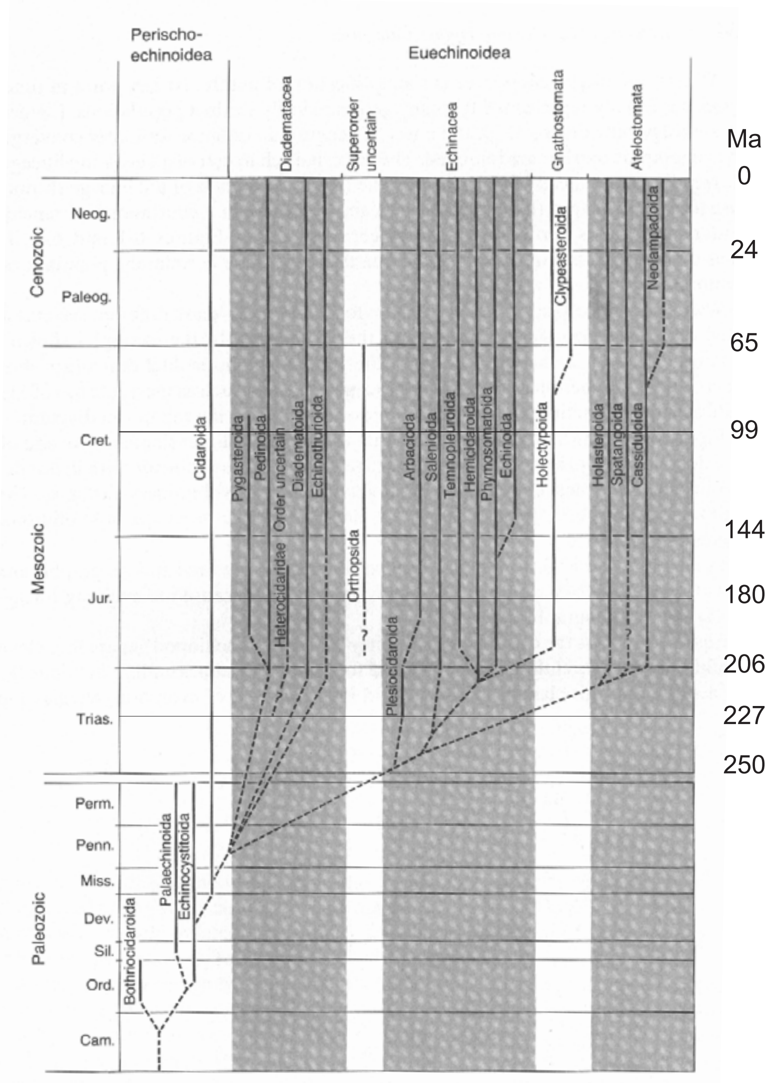

FIGURE 6-2
Phylogeny of the class Echinoidea, based on ranges in the stratigraphic record of specimens of the various orders and on inferred evolutionary relationships. Gap above Permian indicates a change in the vertical scale. (From Durham, 1966.)

Figure 3.

Hand in your branching diagrams.
Hand in a chart of all the diversity curves.
Use your results to answer the following questions:

1) Do you accept or reject the null hypothesis? What is the basis for your decision?
2) What are the similarities and differences between the real diversity curve and the artificial ones?
3) If you reject the null, how would you modify its assumptions in the next round of hypothesis testing?