VI. Some applications of homogeneous sampling models


1. Explore alternative models of descent:
   a. Bifurcation (ancestor gives rise to two distinct descendants)
   b. Budding (ancestor persists and gives rise to one descendant)
   c. Phyletic transformation (ancestor gives rise to one descendant, does not persist)

2. Use homogeneous branching and sampling models in discrete time to predict the proportion of species with at least one preserved descendant, direct or indirect.

3. Calibrate empirically to estimate the probability of finding ancestors given $p$, $q$ and $R$.


1. Use homogeneous branching model in continuous time to predict $S$, the sum of missing lineage durations, implied by: 1. postulated extinction rate, 2. gap between clade origin and first appearance in fossil record, 3. minimum clade diversity at time of first fossil appearance, and 4. assumed diversification model.

2. Use homogeneous sampling model in continuous time to calculate the probability that $S$ could go unobserved for assumed value of sampling rate $r$: $Pr(\text{no fossils}) = e^{-rS}$.

3. Use observed stratigraphic ranges to estimate $r$ empirically, and compare this with value required for missing clade history to be plausible.

** Note the importance of conditioning probabilities on survival of clade to some specified time. Standard exponential growth model is an expectation (mean) which averages together zero values for numerous clades that do not survive. Much higher diversities are expected, especially early in clade history, if we condition on survival of the clade. In general, it is important to think carefully about whether conditioning is appropriate for a given question.


2. Degrade stratigraphic ranges of species within higher taxa by dropping sampling events randomly in time.

3. Compare size-frequency distributions of pristine and degraded clades.

4. Principal results (with respect to monotypy):
   a. Most truly monotypic taxa are not sampled at all.
   b. Most apparently monotypic taxa were not in fact monotypic before sampling.
c. However, the entire size-frequency distribution is affected in such a way that the overall proportion of taxa that are monotypic is similar in the pristine and degraded distributions.

d. Likewise, the proportion of species in the largest genus is also accurately represented (although this proportion is highly variable in both the raw and sampled simulations).