IX. Sampling Models, 1: Introduction to homogeneous sampling models

1 Basic framework for homogeneous sampling models

1.1 r is the per-capita rate of sampling per lineage-million-years.

1.2 "Sampling" means the joint incidence of preservation, exposure, collection, identification, etc.

1.3 Generally consider the relevance only of extinction rate q (tacitly assume $p \approx q$ and truncation effects negligible).

This restriction can be fairly easily relaxed.

1.4 Assume all rates temporally and taxonomically homogeneous (for now).

1.5 It will often be useful to deal with "dimensionless rates" and "dimensionless time".

Let $q = 1$, express r in terms of multiples of q, and express time in terms of multiples of $1/q$ (expected mean duration).

2 Long-term average properties of a set of taxa

(See Foote, 1997, *Paleobiology* 23:278-300.)

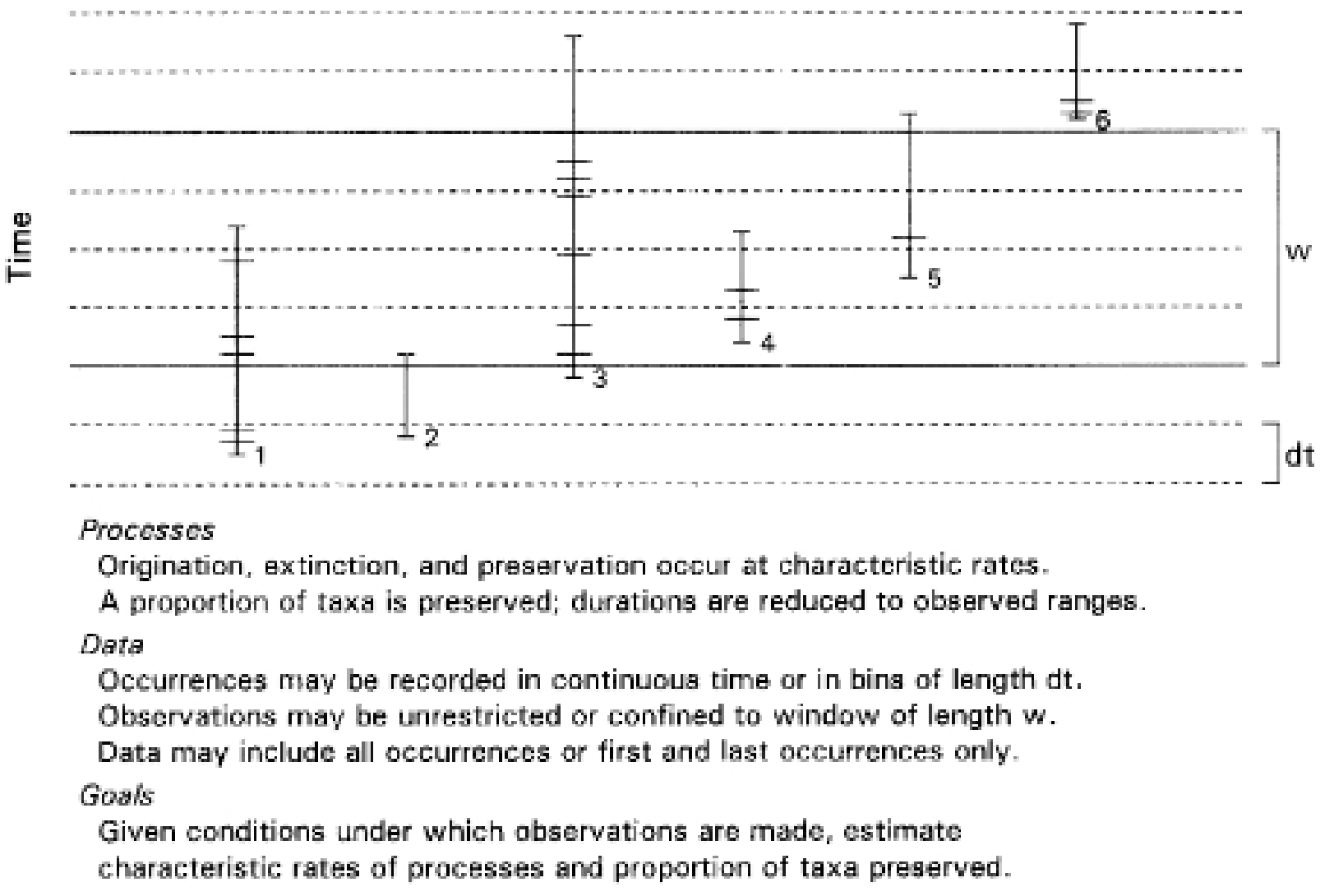

FIGURE 1. Hypothetical examples of taxonomic durations, stratigraphic ranges, and nature of data. Various combinations of the three aspects of data require different approaches for estimating taxonomic rates, preservation rate, and proportion of taxa preserved (see text). For each of six lineages, short bars denote true times of origination and extinction and long bars denote fossil occurrences. Lineage 1 originates and is first preserved prior to window of observation, terminates and is last preserved during window, has five occurrences recorded in continuous time, and has three occurrences recorded in discrete time. Lineage 2 originates before window, terminates during window, and is never preserved. Lineage 3 originates before window, terminates after window, is preserved during window only, has six occurrences recorded in continuous time, and has four occurrences recorded in discrete time. Lineage 4 originates and terminates during window and has two occurrences recorded in both continuous and discrete time. Lineage 5 originates during window, terminates after window, and is recorded as a single hit during window (in both continuous and discrete time). Lineage 6 originates and terminates after window, has two occurrences recorded in continuous time, but has only one occurrence recorded in discrete time.

2.1 For a lineage with duration T, model sampling as a Poisson process with parameter rT.

Thus:

- Pr(lineage never sampled)$= e^{-rT}$ (Poisson probability of zero successes.)
- Pr(lineage sampled at least once)$= 1 - e^{-rT}$

- Pr(lineage sampled exactly once)= rTe^{-rT} (Poisson probability of exactly one success.)
- Pr(observed range=t [$t > 0$])= $r^2(T-t)e^{-r(T-t)}$.This is actually a density, not a probability, and is derived as follows:

$$f(t|T) = \int_0^{T-t} [re^{-ry}][re^{-r(T-y-t)}]\,dy,$$

where

  - the first term within the integral is the density function for the first sampling event from the beginning of the duration
  - the second term is the density function for the first event from the end of the duration ($[T-y-t]$ is the distance from the end working backward)

    and
  - the upper limit of integration is the longest that a true duration (T) can be if the range is t.

- If sampling rate varies, substitute $e^{-\int_0^T r_x\,dx}$ for e^{-rT}.

2.2 These probabilities are then integrated over the entire distribution of durations (exponential or otherwise).

2.2.1 Example with untruncated exponential distribution

- The probability that a randomly chosen lineage will never be sampled is equal to: $\int_0^\infty qe^{-qT} \cdot e^{-rT}\,dT = q/(r+q)$.
- Pr(sampled at least once)= $\int_0^\infty qe^{-qT} \cdot (1-e^{-rT})\,dT = r/(r+q)$. This is the expected prortion of lineages sampled, which has also been referred to as *paleontological completeness*.
- Pr(sampled exactly once)= $\int_0^\infty qe^{-qT} \cdot rTe^{-rT}\,dT = qr/(r+q)^2$. These are the **singletons**.
- Pr(observed range=t [$t > 0$])= $\int_0^\infty qe^{-qT} \cdot r^2(T-t)e^{-r(T-t)}\,dT = qr^2e^{-qt}/(r+q)^2$. (NB: This is actually a density.)
- By the number of times sampled, we mean the number of distinguishable stratigraphic horizons at which it is sampled, not the actual number of fossils or localities. Thus, a singleton may be sampled numerous times, but if these are all, within the limits of resolution, at a single time horizon, this counts as being sampled once.