XI. Sampling Models, 3: Introduction to time-varying models (forward problems)

1 Basic framework

(see Foote 2000, *Paleobiology* Supplement to 26(4):74-102, Foote 2001, *Paleobiology* 27:796 [erratum], and Foote 2003, *Journal of Geology* 111:125-148, 752-753 [erratum])

1.1 Let there be n time intervals, each characterized by a set of origination, extinction, and sampling rates: p_i, q_i, and r_i.

1.2 Use time series of p and q to predict true time series of N_{bL}, N_{Ft}, N_{FL}, and N_{bt} (and thus N_b and N_t).

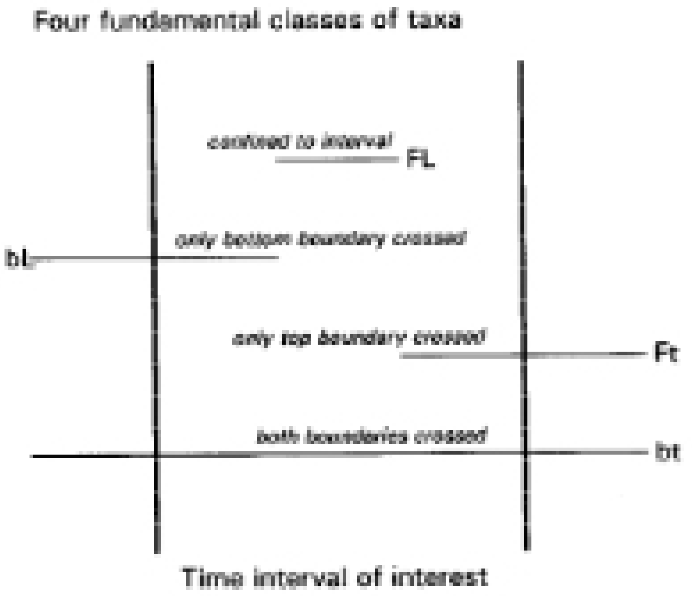

FIGURE 1. Illustration of four fundamental classes of taxa present during a stratigraphic interval. N_{FL} is the number of taxa confined to the interval, N_{bL} is the number that cross the bottom boundary only, N_{Ft} is the number that cross the top boundary only, and N_{bt} is the number that cross both boundaries.

1.2.1 These quantities all scale to N_{b_i}.

1.2.2 N_b for continuous-turnover model (q.v.) in which origination and extinction occur at constant per-capita rate within an interval:

1. Let the age of the bottom boundary be at time $t = x$.

2. Let $N_0 = 1$ at $t = 0$.

3. Let p_t and q_t be time-specific rates.

4. Then
$$N_x = \exp\big[\int_0^x (p_t - q_t)\, dt\big].$$

5. Or, if we divide time into intervals and assume constant p and q *within an interval* (while still varying among intervals):
$$N_{b_i} = \exp\big[\sum_{j=1}^{i-1}(p_j - q_j)\big],$$
if the rates p_i and q_i are expressed per lineage per interval. If instead they are expressed per lineage-million-years, and interval durations are given by Δt_i, then we have:
$$N_{b_i} = \exp\big[\sum_{j=1}^{i-1}(p_j - q_j)\Delta t_j\big].$$

1.2.3 N_b for pulsed-turnover model (q.v.) in which originations are all at start of interval and extinctions at end of interval (so all lineages extend throughout the interval):

1. In this model, P is the number of new lineages produced per lineage extant at the start of the interval. (So number of originations is equal to $N_b \cdot P$ and total interval diversity is $N_b[1 + P]$.)

2. In this model, Q is the extinction probability. (So number of extinctions is equal to $N_b[1 + P]Q$).

3. Thus
$$N_{b_i} = \prod_{j=1}^{i-1}(1 + P_j)(1 - Q_j).$$

1.3 Use time series of p, q, and r to determine sampling probabilities for given time intervals.

Probability of being sampled in a given interval of time (either before a reference point, after a reference point, or between two reference points) is obtained as the integral, over all possible durations, of the probability of having a certain duration multiplied by the probability of being sampled given that duration.

Table 1. Expressions Used to Calculate Survivorship Probabilities

Quantity/model	Expression
N_{bt}:	
•C	$N_b e^{-q}$
•P	$N_b(1-q)$
N_{bL}:	
•C	$N_b(1-e^{-q})$
•P	$N_b q$
N_{Ft}:	
CC	$N_b e^{p-q}(1-e^{-p})$
PP	$N_b p(1-q)$
CP	$N_b e^{p}(1-q)(1-e^{-p})$
PC	$N_b p e^{-q}$
N_{FL}:	
CC[a]	$N_b(e^{-q}+p-1)$ if $p=q$,
	$N_b\dfrac{qe^{(p-q)}+(p-q)e^{-q}-p}{p-q}$ if $p\neq q$
PP	$N_b pq$
CP	$N_b q(e^{p}-1)$
PC	$N_b p(1-e^{-q})$
$P_A(i)$:	
•C	$\sum_{k=i+1}^{n}\left(\left[(e^{-\sum_{m=i+1}^{k-1}q_m})(1-e^{-q_k})\right]\{1-(e^{-\sum_{m=i+1}^{k-1}r_m})[1-P_{D\mid bL}(k)]\}\right)+(e^{-\sum_{m=i+1}^{n}q_m})\{1-(e^{-\sum_{m=i+1}^{n}r_m})[1-P_A(n)]\}$
•P	$\sum_{k=i+1}^{n}\left(\left(\prod_{m=i+1}^{k-1}1-q_m\right)(q_k)\{1-(e^{-\sum_{m=i+1}^{k-1}r_m})[1-P_{D\mid bL}(k)]\}\right)+\left[\prod_{k=i+1}^{n}(1-q_k)\right]\{1-(e^{-\sum_{m=i+1}^{n}r_m})[1-P_A(n)]\}$
$P_B(i)$:	
C•	$\sum_{k=1}^{i-1}\left(\left[(e^{-\sum_{m=k+1}^{i-1}p_m})(1-e^{-p_k})\right]\{1-(e^{-\sum_{m=k+1}^{i-1}r_m})[1-P_{D\mid Ft}(k)]\}\right)+(e^{-\sum_{k=1}^{i-1}p_k})\{1-(e^{-\sum_{k=1}^{i-1}r_k})[1-P_B(1)]\}$
P•	$\sum_{k=1}^{i-1}\left(\left(\prod_{m=k+1}^{i-1}\frac{1}{1+p_m}\right)\left(\frac{p_k}{1+p_k}\right)\{1-(e^{-\sum_{m=k+1}^{i-1}r_m})[1-P_{D\mid Ft}(k)]\}\right)+\left(\prod_{k=1}^{i-1}\frac{1}{1+p_k}\right)\{1-(e^{-\sum_{k=1}^{i-1}r_k})[1-P_B(1)]\}$
$P_{D\mid bt}$:	
••	$1-e^{-r}$
$P_{D\mid bL}$:	
•C[b]	$\dfrac{[r+qe^{-(q+r)}]/(q+r)-e^{-q}}{1-e^{-q}}$
•P	$1-e^{-r}$
$P_{D\mid Ft}$:	
C•[c]	$\dfrac{[r+pe^{-(p+r)}]/(p+r)-e^{-p}}{1-e^{-p}}$
P•	$1-e^{-r}$
$P_{D\mid FL}$:	
CC[d]	$\dfrac{N_b p}{N_{FL}}\left\{\dfrac{r}{p+r}-\dfrac{1-e^{-p}}{p}-\dfrac{p[1-e^{-(p+r)}]}{(p+r)^2}\right\}$ if $p=q$,
	$\dfrac{N_b}{N_{FL}}\left\{\dfrac{pr(e^{(p-q)}-1)}{(q+r)(p-q)}+\dfrac{pqe^{-(p+r)}(e^{(p+r)}-1)}{(p+r)(q+r)}-e^{-q}(e^{p}-1)\right\}$ if $p\neq q$
PP	$1-e^{-r}$
CP[e]	$\dfrac{r(e^{p}-1)+p(e^{-r}-1)}{p(r+p)(e^{p}-1)}$
PC[f]	$\dfrac{r(e^{q}-1)+q(e^{-r}-1)}{q(r+q)(e^{q}-1)}$

Note. In the two-character code for model, the first character denotes origination and the second extinction; C = continuous, P = pulsed. A bullet means the expression applies to either model for the corresponding process. N_b is the true standing diversity at the start of the interval; because all relevant numbers scale to N_b, this can be arbitrarily set to unity.

[a] Foote 2000*a*, eqq. (6b) and (6c).

[b] Foote 2000*a*, eq. (27b).

[c] Foote 2000*a*, eq. (28b).

[d] Foote 2000*a*, eqq. (29b) and (29c).

[e] Let z represent time within an interval of duration t, where $z=0$ and $z=t$ are the beginning and end of the interval, respectively. By assumption, there is no extinction until the end of the interval. Thus, the density of origination at time z is equal to $e^{pz}/(e^{pt}-1)$ (cf. Foote 2001*a*, eq. [3]). Because all lineages originating within the interval extend to the end, the probability of preservation, given origin at z and extinction at t, is equal to $1-e^{-r(t-z)}$. It is necessary to integrate the density of origination times the probability of preservation over all values of z. Thus, $P_{D\mid FL}=[1/(e^{pt}-1)]\int_0^t e^{pz}[1-e^{-r(t-z)}]dz$, which is equal to the expression in the table once t is set to unity.

[f] Derived as in the foregoing footnote, with origination and extinction reversed.