XII. Sampling Models, 3: Confidence limits on stratigraphic ranges

1 Statement of problem

Given observed first (last) appearance, how far below (above) it do we need to go to have desired level of confidence that we have included the time of true origination (extinction)?

2 Uniform sampling probability

2.1 Basic framework

See Fig. 2 of Strauss and Sadler (1989) *Mathematical Geology* 21:411.

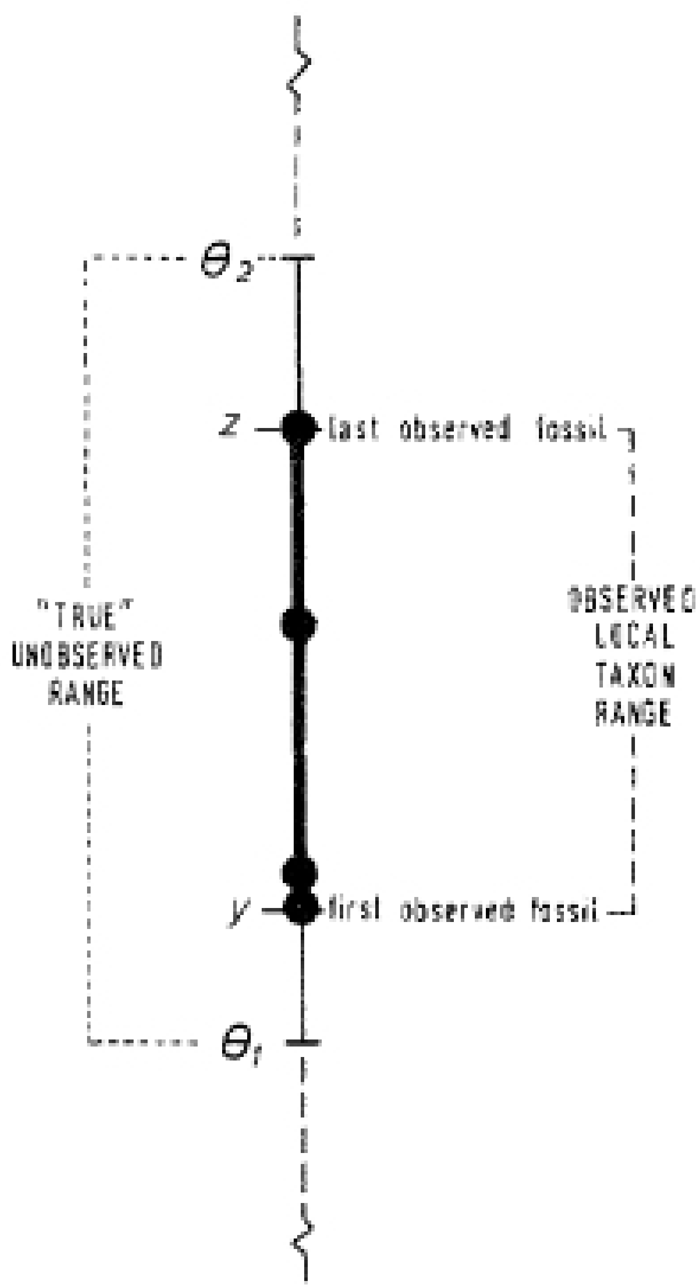


Fig. 2. The relationship between fossil finds (filled circles), ends of observed local range (y and z), and true ends of range (θ_1 and θ_2). Vertical scale is height above an arbitrary datum in the stratigraphic section.

2.1.1 Θ_1 and Θ_2 are times of true origination and extinction.

2.1.2 y and z are times of first occurrence (FO) and last occurrence (LO).

2.1.3 Thus, observed stratigraphic range is $(z - y)$.

2.1.4 n is the number of levels at which taxon is found.

2.1.5 Density of true origination (extinction) declines monotonically below (above) first (last) appearance.

This leads to ML estimates: $\hat{\Theta}_1 = y$ and $\hat{\Theta}_2 = z$, but these *must* be biased if sampling is incomplete.

2.1.6 Unbiased estimates of Θ_1 and Θ_2.

- Note that gap between Θ_1 and y and between z and Θ_2 are drawn from same distribution as gaps between fossil finds.
- Mean gap size is equal to $(z - y)/(n - 1)$.
- Extend FO or LO by mean gap to get unbiased estimator of true origin or extinction:

$$\tilde{\Theta}_1 = y - \frac{(z-y)}{(n-1)} = \frac{(ny-z)}{(n-1)}$$

$$\tilde{\Theta}_2 = z + \frac{(z-y)}{(n-1)} = \frac{(nz-y)}{(n-1)}$$

2.2 Developing confidence interval

2.2.1 Let α be a proportion of observed stratigraphic range, and scale true duration so that $\Theta_1 = 0$ and $\Theta_2 = 1$.

2.2.2 Let Y and Z denote the random variables whose value is the time of first and last appearance.

2.2.3 Consider a prospective "adjustment" of the observed stratigraphic range such that $y_{adj} = Y - \alpha(Z - Y)$.

The larger α is, the greater is the probability that $Y - \alpha(Z - Y) < 0$, i.e. the greater is the probability that the true time of origin lies within the confidence interval $[Y, Y - \alpha(Z - Y)]$.

2.2.4 Denote this probability $p = Pr[Y - \alpha(Z - Y) < 0]$.

2.2.5 Rearrange to yield $p = Pr[Y < \frac{\alpha}{1+\alpha}Z]$.

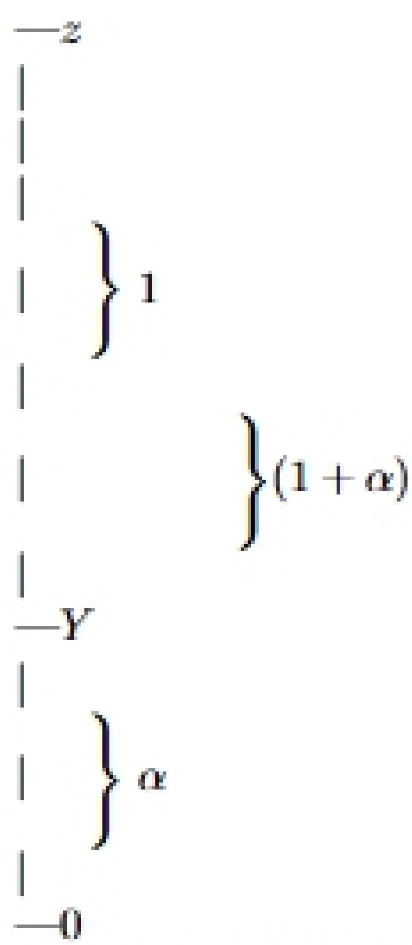
2.2.6 Condition on observed value of z. Then there are $(n-1)$ fossil horizons uniformly distributed on $(0, z)$ (i.e. between true origin and LO).

2.2.7 Since $p = Pr(Y < \frac{\alpha}{1+\alpha}z)$,

$$(1-p) = Pr(Y \geq \frac{\alpha}{1+\alpha}z).$$

(Here we use z rather than Z because we are conditioning on the observed value.)In other words, $(1-p)$ is the probability that all the fossil finds are between $\frac{\alpha}{1+\alpha}z$ and z.

2.2.8 Imagine the distance from 0 to z being divided into a segment of length α (from 0 to Y) and a segment of length 1 (from Y to z).



Then the probability that a randomly dropped point falls between 0 and Y is equal to $\frac{\alpha}{1+\alpha}$, and the probability that it falls between Y and z is equal to $1 - \frac{\alpha}{1+\alpha}$. (Remember, we have conditioned on $Z = z$.) Thus the probability that all $(n-1)$ points fall between Y and z is equal to

$$[1 - \frac{\alpha}{1+\alpha}]^{(n-1)}$$