

Estimating extinction with the fossil record

Peter J. Wagner and S. Kathleen Lyons

19.1 Introduction

Many ecological and palaeontological studies focus on extinction. The fossil record is particularly important for studying long-term patterns in extinction: although analyses of extant phylogenies can estimate extinction rates (e.g. Alfaro et al. 2009) and even suggest mass extinctions (e.g. Crisp & Cook 2009), they cannot imply trilobites ever existed or that sphenodonts (now represented only by the tuatara) were once as diverse as lepidosaurs (lizards and snakes). However, workers also use the fossil record to test ideas about the pace of major extinction events that use methods similar to those that conservation biology might use. Here we will review current palaeobiological methods for inferring extinction patterns, spanning 'traditional' methods using stratigraphic ranges to methods using more exact information about distributions of finds within stratigraphic ranges.

Palaeontological studies of diversity patterns also focus extensively on standing richness and origination rates. We do not focus on either of these parameters for their own sake. However, standing richness (usually referred to simply as 'diversity' in palaeontological studies) is an important parameter when discussing extinction as extinction metrics necessarily rely on changes in standing richness. Origination rates also can be an important parameter as the distribution of originations within time spans (i.e. evenly throughout the interval vs concentrated in the beginning) have small but important effects on the predictions of extinction rate hypotheses. Moreover, palaeontological techniques for estimating origination rates

are essentially identical to those for estimating extinction, save that they are done 'in reverse'. Thus, discussions of *how* one estimates extinction rates using fossil data can almost double for discussion of how one estimates origination from fossil data.

The other parameter that will be important throughout this review is preservation (i.e. sampling) rate. Just as ecologists know that incomplete sampling affects sampled richness (e.g. Hurlbert 1971) and implied extinction (e.g. Solow 1993a), palaeontologists know that incomplete sampling affects the first and last appearances of fossils, both in local sedimentary sections (Signor & Lipps 1982) and globally (Sepkoski 1975). Workers have expressed concern that major extinctions are exaggerated by or possibly even illusions of intervals of poor preservation (Raup 1979; Smith et al. 2001; Peters & Foote 2002). Because sampling intensity is of interest to both fields, we shall also discuss how palaeontological studies address sampling when estimating diversity and diversity dynamics.

19.2 State of the field

19.2.1 Basic metrics

Palaeontological studies of richness and diversity dynamics (i.e. extinction and origination) date back to the 19th century (Phillips 1860). Traditional palaeontological studies use *synoptic databases* (e.g. Sepkoski 1982, 2002), which catalogue the first and last appearances of taxa. These almost always are binned into chronostratigraphic units: typically stages and substages, but sometimes as fine as

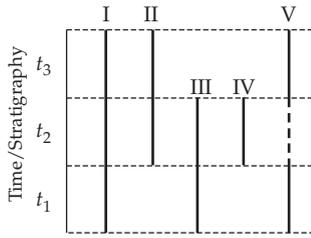


Figure 19.1 Chronostratigraphic ranges of five hypothetical taxa. Each t gives a separate chronostratigraphic unit, with t_1 the oldest. Dashed vertical line for V indicates that taxon V is not actually found in unit t_2 . Modified from Foote (2000), Fig. 1.

faunal zones. (We use ‘intervals’ to refer chronostratigraphic units because of the correspondence between units of time and strata; see Gradstein et al. (2005)). For numerous reasons, synoptic studies typically use supraspecific taxa, with genera being the most common taxon since the mid 1980s.

Fig. 19.1 presents a simple example. We use last appearances in each unit t to infer rates of extinction. Two types of metrics commonly are used: per taxon (pt) and per capita (pc; see Foote (2000)). Per-taxon metrics measure the proportion of standing richness, S , that have their last appearance in an interval. Per-taxon extinction thus is:

$$\mu_{\text{pt}} = \frac{S_L}{S} \quad (19.1)$$

where S_L is the number of taxa last appearing in an interval. In Fig. 19.1, taxa III and IV last appear in t_2 , so $S_L = 2$. S is not the sum of taxa sampled in an interval, but the sum of taxa with chronostratigraphic ranges spanning the interval. Taxon V in t_2 illustrates the difference: because it first appears before t_2 and last appears after t_2 , we assume that it existed in t_2 even though V is not sampled in t_2 . This assumption is unsound only if taxon V is polyphyletic. As the synoptic data for taxa I and V are identical, we need a more detailed database to recognize cases like taxon V. Thus, $S_{t_2} = 5$ regardless of how many ‘range-through’ taxa (e.g. I and V) actually are sampled in t_2 , and $\mu_{\text{pt}} = 0.40$. (Per-taxon origination, λ_{pt} , simply replaces S_L with S_F , the taxa first appearing in t_2).

Per-capita extinction rate measures the survivorship of taxa present at the outset of an interval. This is:

$$\mu_{\text{pc}} = \ln \left(\frac{S_a}{S_{\text{az}}} \right) \quad (19.2)$$

where S_a is the number of taxa crossing the boundary from the prior interval and S_{az} is the number of the taxa also crossing through both the base and the top (i.e. the range-through taxa). Note that ‘per capita’ here refers not to individuals within a species, but to taxa as operational units. For t_2 in Fig. 19.1, S_a is 3 (taxa I, III, and V) and S_{az} is 2 (taxa I and V), giving $\mu_{\text{pc}} = 0.41$. The most likely probability of a taxon becoming extinct in an interval is now $1 - e^{-\mu_{\text{pc}}}$. However, this is identical to μ_{pc} if we restrict per-taxon rates to only boundary crossers.

Chronostratigraphic intervals vary considerably in temporal length (Gradstein et al. 2005). Accordingly, workers often divide rates by interval length to get rates per million years instead of rates per interval. This is appropriate if extinction is distributed throughout an interval. One expectation of continuous turnover rates is that longer intervals should have higher extinction rates. However, Foote (1994) showed that ‘raw’ extinction metrics are random with respect to the length of chronostratigraphic units, whereas per million year rates show a negative correlation. We will return to this issue below when we summarize tests of pulsed vs continuous rates.

The distribution of extinctions and originations plays a role in whether we should prefer per-taxon or per-capita extinction rates. If extinction and origination happen throughout an interval, then per-taxon rates overestimate immediate extinction risk as well as exaggerating S from any one slice of time. ‘Singletons,’ that is those taxa known from only one interval, such as taxon IV in Fig. 19.1, exacerbate this problem. Singletons often reflect differences in research effort and/or available fossils (Raup & Boyajian 1988) and excluding singletons greatly reduces the volatility of per-taxon rates (Alroy 1996). Although per-capita rates are excellent for describing continuous rates, they still can accurately reflected pulsed extinction.

19.2.2 Survivorship curves

Per-capita rates lead conceptually to survivorship curves. The per-capita rate for any one interval reflects the proportion of taxa expected to survive the entire interval into the next interval. Survivorship curves reflect the proportion of taxa that survive multiple successive intervals. Some survivorship analyses contrast rates among higher taxa in order to assess whether extinction rates in some groups differ markedly from those in other groups (e.g. Simpson 1944, 1953) or whether there are common patterns among higher taxa (Van Valen 1973, 1979; Raup 1991). Alternatively, cohort analyses contrast sets of taxa that originate in the same interval (i.e. Raup 1978; Foote 1988) to examine whether extinction rates change markedly over time. Given a per-taxon extinction rate, we expect the proportion of taxa that survive N intervals, $f(N)$, to be:

$$f(N) = \mu_{pc} = e^{-N\mu_{pc}} \quad (19.3)$$

Given $\mu_{pc} = -\ln(\frac{3}{2})$ from our example above, we expect 66.7% of the taxa to survive 1+ intervals, 44.4% to survive 2+ intervals, etc., and thus 33.3% of the taxa to have durations of 1 interval, 22.2% to have durations of 2 intervals, etc. Of course, in reality this will vary even if extinction rates are constant simply by chance. However, the slope of the logged frequency of taxa with durations of 1, 2, 3, etc. intervals plotted against duration approximates the average μ_{pc} .

Foote (1988) uses Monte Carlo simulations to demonstrate that survivorship slopes for cohorts of Cambrian trilobites are significantly steeper than those for cohorts of Ordovician trilobites. (Subsequent revisions to the timescale greatly reduced the average longevity of Cambrian trilobites, making the pattern more pronounced than Foote reports; Bowring et al. 1993). Raup (1991) also uses this approach to demonstrate that survivorship slopes for Phanerozoic cohorts fit the expectations of a general 'kill curve' of extinction rates.

19.2.3 The importance of sampling

For three reasons, equation 19.3 will not predict the distributions of stratigraphic ranges even if μ_{pc} is

constant over time and among taxa. First, if we measure stratigraphic ranges in discrete bins (as is almost always done), then $f(N=1) = e^{-\mu_{pc}}$ only if all taxa originate at the very base of intervals. If origination is continuous throughout intervals, then we expect (for example) half of the taxa lasting one half an interval to span from t_x to t_{x+1} . Thus, only the slope from taxa with ranges of 2+ intervals will reflect μ_{pc} under continuous diversification. Second, extinction rates do not make predictions about the ages of extant taxa: instead, the age distribution of contemporaneous taxa (i.e. a 'backwards survivorship' curve *sensu* Pease (1988) or a 'prenascence' curve *sensu* Foote (2001b)) reflects origination rates (Foote 2001b).

The third reason is the most critical: extinction rates make predictions about temporal durations, but we can observe only stratigraphic ranges. As preservation rates strongly affect stratigraphic ranges, 'survivorship' curves using fossil ranges actually reflect both preservation and extinction (Sepkoski 1975; Pease 1988; Foote & Raup 1996). Consider a simple example of two clades where origination and extinction occur at the beginning and end of each interval, and for which $\bar{\mu}_{pc} = 0.69$ (Fig. 19.2). For one clade, the probability of recov-

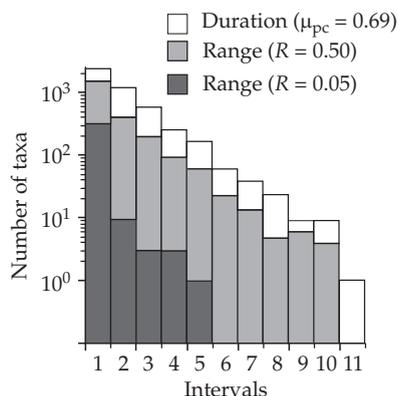


Figure 19.2 The relationship between durations (time from birth to death) and stratigraphic ranges (= number of chronostratigraphic units from first to last appearances). This simulation of 5000 taxa uses $\mu_{pc} = 0.69$, which predicts half of the taxa present at the outset of an interval to go extinct in that interval. Sampled stratigraphic ranges reflect sampling intensity (R) of 0.5 and 0.05 per interval. Following palaeontological studies, both durations and ranges are binned into discrete intervals.

ery, R , is 0.5 per interval; for the other, $R = 0.05$ per interval. These illustrate two patterns emphasized by Foote and Raup (1996). First, the slopes of the logged survivorship curves *after* the first interval parallel each other quite well. Second, the proportion of taxa sampled from 1 interval to those sampled over 2+ intervals increases as R decreases. At sufficiently low R , all taxa are known from single finds (and thus single intervals) only. From this, Foote and Raup derived preservation intensity as:

$$R_{FR} = \frac{S_{sr=2}^2}{S_{sr=1} \times S_{sr=3}} \quad (19.4)$$

where $S_{sr=x}$ is the number of taxa with stratigraphic ranges of x intervals. In our example here, $R_{FR} = 0.53$ when true $R = 0.50$ and $R_{FR} = 0.09$ when true $R = 0.05$.

The ability to contrast preservation rates over time or among taxa obviously interests palaeontologists. However, subsequent work embraced the fact that joint hypotheses of preservation and extinction make exact predictions about distributions of stratigraphic ranges. This led to a shift from contrasting slopes of survivorship curves to contrasting the likelihoods of hypothesized extinction and preservation rates (Solow & Smith 1997; Foote 1997). Consider a simple system where origination and extinction occur at the beginning and end of intervals. Extinction here is a per-taxon metric giving the proportion of taxa expected to become extinct at the end of each interval. The expected proportion of taxa with a stratigraphic range (sr) of 1 interval is:

$$f[sr = 1] = \frac{\sum_{d=sr}^w [P[\mu](1 - P[\mu])^{(d-1)} \times d(1 - R)^{(d-sr)} R]}{1 - \sum_{d=1}^w [P[\mu](1 - P[\mu])^{(d-1)} \times (1 - R)^d]} \quad (19.5a)$$

while the expected proportion of taxa known from $x = 2, \dots, w$ intervals (where w is the maximum possible) is:

$$f[sr] = \sum_{d=sr}^w \left[P[\mu](1 - P[\mu])^{(d-1)} \times (1 + d - sr)(1 - R)^{(d-sr)} \times \sum_{n=0}^{sr-2} \binom{sr-2}{n} (1 - R)^{(sr-2-n)} R^{(n+2)} \right] / \left[1 - \sum_{d=1}^w [P[\mu](1 - P[\mu])^{(d-1)} \times (1 - R)^d] \right] \quad (19.5b)$$

where d is the true duration and n is the number of intervals between the first and last interval in which a taxon might be found (Foote 1997). The three important recurring terms here are:

- $\mu(1 - \mu)^{d-1}$: the probability of avoiding extinction for $d - 1$ intervals and becoming extinct in interval d , where μ is equivalent to μ_{pc} for boundary crossers only;
- $(1 - R)^x$: the probability of being unsampled for x intervals;
- R^y : the probability of being sampled in y intervals.

The denominator in both equations estimates the proportion of taxa that are sampled as one minus the proportion that are unsampled, i.e. the probability of having duration d times the probability of going unsampled for d intervals summed over the entire range of possible durations (w). This is because we can test hypotheses based only on their predictions about *sampled* taxa. The numerator also sums over the range of possible durations. The second part of both numerators reflects the number of ways a stratigraphic range of sr fits within a duration of d , which is simply $1 + (d - sr)$. For example, there is only one way to fit $sr = 4$ within $d = 4(t_1, \dots, t_4)$ and two ways to fit $sr = 4$ within $d = 5(t_1, \dots, t_4 \text{ and } t_2, \dots, t_5)$, etc. The final term in equation 19.5a, $\sum_{n=0}^{sr-2} \binom{sr-2}{n} (1 - R)^{(sr-2-n)} R^{(n+2)}$, sums the probabilities given R of all of the different stratigraphic records that match a synoptic range of sr . All taxa with $sr = 4$ are found in at least two intervals (the first and last). However, there are four possible stratigraphic records consistent with that synoptic range (Fig. 19.3), therefore the likelihood function must sum the probabilities of all four possibilities given any R .

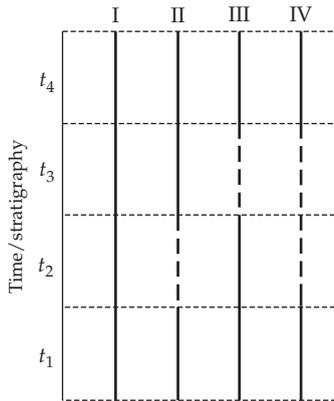


Figure 19.3 Four exact stratigraphic distributions resulting in a synoptic range of four intervals.

Foote (1997) presents more complex variations of equation 19.5 for continuous stratigraphic ranges. As most palaeontological studies use discrete bins, we will not cover this in detail. However, we will return to the effects of continuous μ and λ (origination) on the expectations for stratigraphic ranges below. At this point, equation 19.5 gives us an expected observed survivorship curve given any combination of μ and R . Multinomial probability then gives the probability of the data given that combination. However, this obviously is not wholly satisfactory for several reasons. In particular, this approach is useful for contrasting general differences in average extinction rates between taxa while controlling for the possibility of different average preservation rates. Most of the interesting palaeoecological questions concern differences in extinction rates in particular intervals. We also have ample evidence that sampling varies over time (e.g. Peters 2006; McGowan & Smith 2008), which might affect our estimates of extinction.

19.2.4 Relevant studies

Foote (2001c, 2003) modifies the approach outlined above to estimate interval-by-interval rates of μ , λ and R . One applies variations of equation 19.5 on 'birth' and 'death' cohorts, that is, how long do taxa first appearing in an interval last and how old are the taxa that last appear in an interval? In other

words, instead of analysing a vector of ages at last appearance, one now analyses a matrix X where each cell X_{ij} gives the number of taxa first appearing in t_i and last appearing in t_j . Now, one can apply separate μ_i and R_i (and λ_i) at each t_i on variations of equation 19.5. The expected number of taxa originating in t_i with $d = j$ now is affected by variation in μ from i to j , and the probability of any given $sr \leq d$ is affected by variation in R from i to j . Because the best R_i for any t_i must accommodate both the backwards survivorship curves as well as the forwards survivorship curve, λ_i is relevant for estimating μ_i . The algorithms to find the collection of parameters maximizing the probability of the observed matrix are computationally intensive. Foote's (2003) analysis of Sepkoski's (2002) compendium of marine genera shows, for example, that the major loss of taxa at the Big Five mass extinction events (Raup & Sepkoski 1982) and the high turnover of the Cambrian taxa (Sepkoski 1979) require high rates of extinction even in those cases where preservation rates were low.

Finally, estimates of extinction and preservation from equation 19.5 effectively assume that taxa extant are present throughout all intervals within their durations. This can be true only if both origination and extinction are pulsed, that is, concentrated in some particular interval. However, if they are continuous, then true durations will span fractions of the first and last intervals for most taxa. This has two effects (Foote 2005). One, if origination probabilities are uniform throughout intervals, then there is a 70% chance that a taxon with a true duration of 2.7 intervals exists in all or part of four intervals. Two, the probability of sampling a taxon in its first or last interval is not R , but $1 - (1 - R)^x$, where x is the fraction of the first/last interval in which the taxon occurs. Given this subtle difference, Foote (2005) found strong support for pulsed extinctions and origination in most intervals. One implication of this is that the tradition of standardizing extinction rates by millions of years often is highly misleading.

19.2.5 Occurrence-based diversity estimates

We require approaches such as outlined in the prior section when synoptic ranges are our only

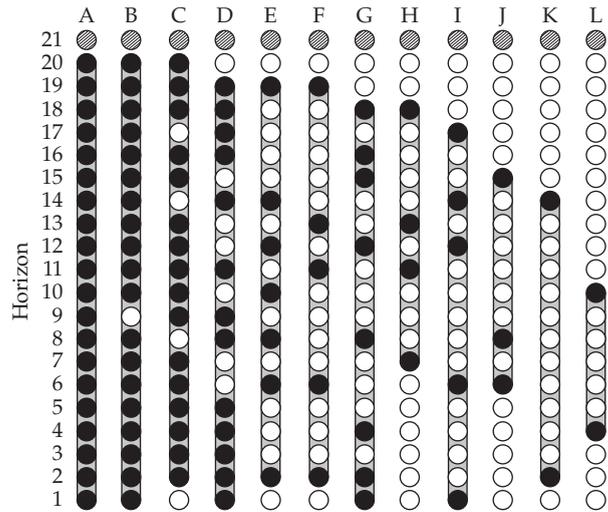


Figure 19.4 Hypothetical distribution of finds in a section, based on a single simulation. ‘Horizons’ represent fossiliferous beds. Black circles indicate that the species is sampled. White circles indicate that it is unsampled. Hatched circles represent the first fossiliferous horizon following the extinction event. Gray ellipses give the observed stratigraphic ranges. Although all 12 species are extant throughout the time represented, they appear to go extinct (and originate) gradually.

data. However, in many cases we have data concerning the distributions of finds within stratigraphic ranges, such as the Paleobiology Database (<http://pbdb.org>; see, for example, Alroy et al. (2001)). This provides us with an independent means of estimating R_i using range-through taxa for any t_i . In Fig. 19.4, four taxa are known to be extant during t_2 and t_3 because they were sampled before and after those intervals. This leads to a simple estimate (Connolly & Miller 2001b):

$$R_i = \frac{S_{i,obs}}{S_{i,az}} \quad (19.6)$$

where $S_{i,az}$ is the number of range-through taxa for t_i (see Fig. 19.1) and $S_{i,obs}$ is the subset of those taxa that we actually observe. If one uses empirical estimates of R_i instead of solving for R_i at the same time as μ_i (and λ_i), then it becomes much easier to solve for the diversity parameters.

Foote (2007a) uses the Paleobiology Database to measure R with range-through taxa and then assess the effect of preservation on recalibrated extinction rates. This suggested that extinction is *less* continuous through the Phanerozoic than previously thought, as immediate R_i explains the last appearances in many low extinction intervals. A corresponding result is that extinction rates of subsequent intervals increase, with the end result being that intervals with high extinction given ‘face value’

data have even higher extinction after optimizing for R . In particular, Foote’s analysis suggested that extinction rates in the intervals preceding the end-Permian, end-Triassic, end-Ordovician, and end-Devonian were lower than suggested by the ‘face value’ fossil record. Thus, accommodating preservation elevates rather than erases mass extinction episodes. Nevertheless, and as Foote emphasizes in that work, the ‘face value’ data leads to the same conclusions, for example estimates of discontinuity of extinction rates separating a few intervals of mass extinctions from other intervals remains (see Wang 2003).

Connolly and Miller (2001b) use the distribution of genus finds over time to estimate preservation, extinction, and origination likelihoods using a capture–mark–recapture (CMR) approach. This represents a modification of the CMR methods used to estimate the same parameters for Recent taxa given conservation data (Pradel 1996). One obvious advantage of this sort of approach is that one can estimate diversity parameters for separate environments or biogeographic units from the same stages (see, for example, Miller & Foote (1996) and Connolly & Miller (2001a)). Alroy (2000) uses likelihood estimates of these parameters to construct faunal ordinations that maximize the probability of observed species lists. The further development of such techniques in conjunction with the

expansion of large occurrence-based databases (e.g. the PBDB) represents an important future direction for palaeobiological studies (see Prospectus section below).

19.2.6 Gap analyses

The use of exact numbers of finds within stratigraphic ranges and the issue of continuous versus pulsed extinction rates points the way to the most exact methods for evaluating the timing of extinctions. This is perhaps best known in association with Alvarez et al.'s (1980) proposal that the end-Cretaceous extinction is a sudden event triggered by a major asteroid impact. However, the idea that extinction events are pulsed rather than gradual is not new (see, for example, Erwin (2006)). Some theory suggests that lower-level turnovers often are pulsed (Vrba 1985; Morris et al. 1995) and many of our concepts about stratigraphic zones tacitly assume pulsed turnovers (e.g. Shaw 1964). Foote's (2005) study argues against 'gradual' extinction mechanisms that distribute extinctions through entire stages. However, chronostratigraphic stages and substages typically are millions of years long whereas the 'gradual' extinction hypotheses often predict extinction over hundreds of thousands of years. Thus, although the 'pulsed' hypotheses of extinctions predict extinctions over tens of thousands of years or less, both 'pulsed' and 'gradual' might make the same predictions about synoptic data. We therefore must focus on the distributions of finds within chronostratigraphic units to find different predictions between gradual and pulsed extinction hypotheses.

Preservation again is very important when ascertaining the expectations of pulsed and gradual extinction hypotheses. The probabilistic nature of preservation and sampling means that we expect gradual disappearances of taxa from the fossil record even if all of the taxa share the same time of extinction (Signor & Lipps 1982). Simple simulations illustrate the Signor–Lipps effect (Fig. 19.4). Consider the first and last occurrences of 12 taxa over 21 fossiliferous beds (horizons), with the 12 taxa forming a Zipf abundance distribution ($\gamma = 2.0$) and having equal preservation probability per specimen. Although all 12 taxa are present throughout

the section, they appear to phase out gradually simply because of sampling failure: although the common taxa are sampled up to the top of the section, the less common taxa have highly truncated ranges. Solow (1993a) notes a similar problem for conservation biologists: it is easy to miss rare species for 5+ years by chance alone, so failing to sample one for a long period of time is weak evidence that it is extinct; conversely, it is difficult to miss a common species for any length of time, so failing to sample one is strong evidence that it is extinct.

One advantage that palaeontologists have that conservation biologists do not is that they can continue to sample all of the time planes. Ultimately, elevated sampling should reveal the rare taxa. However, this is too labour intensive and expensive to be done at many sites. Thus, we need solutions using the available data. This, in turn, leads to gap analyses, which attempt to use estimates of expected gaps between samples to evaluate both extinction and the Signor–Lipps effect.

Strauss and Sadler (1989; see also Marshall (1990)) use a Dirichlet distribution to assess possible gaps between last appearance and true extinction at any given confidence interval (CI) as:

$$g = sr \times \left([1 - C_1]^{-\frac{1}{H-1}} - 1 \right) \quad (19.7)$$

where g is the hypothesized gap, sr is the observed stratigraphic range, H is the number of horizons, and C_1 is the CI on one end of the stratigraphic range. Note that if we use continuous meters (e.g. Strauss & Sadler 1989) or time (e.g. Marshall 1990), then the stratigraphic range would be the last horizon minus the first horizon and single samples have no range; however, with binned horizons, the minimum range is 1. We use horizons here simply because the non-random distribution of horizons within units of strata or time require further modifications to these approaches (Marshall 1994, 1997; but see Prospectus section below). Thus, for taxon E, $g_{50\%}$ is 18 horizons $\times ([1 - 0.5]^{-\frac{1}{18-1}} - 1) = 2.2$ horizons beyond the last appearance. In other words, there is a 50% chance that the true extinction happened between the last appearance and this point and a 50% chance that the true extinction happened after this point.

Marshall (1995) proposes using 50% confidence intervals to test the null hypothesis of simultaneous extinction. Marshall uses the range that is above $g_{50\%}$ for half of the species as the optimal extinction zone. In Fig. 19.4, half of the taxa have $g_{50\%} \leq 2.2$ and half have $g_{50\%} \geq 3.1$. Thus, the 22nd horizon would be the optimum extinction horizon. As that follows the last appearance of all of the taxa, we cannot reject a single extinction. Using metres of sediment rather than fossiliferous horizons, Marshall (1995) shows that Cretaceous ammonoids not only fail to reject the simultaneous extinctions, but that the $g_{50\%}$ values do an excellent job of predicting the iridium layer terminating the Cretaceous. Conversely, Marshall and Ward (1996) demonstrate that the 50% line for Late Cretaceous inoceramid bivalves precedes the last appearance of nearly half the species. As many species clearly existed after the optimal extinction time for half the species, there almost certainly are multiple extinction events for inoceramids.

Marshall's basic approach lends itself well to likelihood tests of numbers of extinction events. Building on a model initially presented by Solow (1996), Wang and Everson (2007) consider the likelihood of the hypothesized extinctions, $\zeta_{1\dots S}$, of S taxa, noting that:

$$L[\zeta_{1\dots S} | sr_{1\dots S}, H_{1\dots S}] \propto \prod_{i=1}^S \left(\frac{sr_i}{\zeta_i - FA_i + 1} \right)^{H_i} \quad (19.8a)$$

where FA_i is the first appearance of taxon i , and therefore:

$$\zeta_i - FA_i + 1 = g_i + sr_i \quad (19.8b)$$

(i.e. the observed stratigraphic range, sr , plus the gap, g , implicit to the hypothesized extinction ζ). Again, the '+1' is necessary only when stratigraphic ranges are in discrete rather than continuous units. Note also that the likelihood is 0 rather than a negative number if $g < 0$ as this is an impossible hypothesis. For any taxon i , the likelihood decreases as g_i increases. The number of parameters is equal to the number of different extinction events. Thus, the simplest (minimum varying parameter) hypothesis has all taxa sharing the same extinction events and the most likely single-parameter (single-extinction) hypothesis sets all ζ_i to the latest last appearance

(e.g. horizon 20 in Fig. 19.4). Conversely, the most complex possible hypothesis posits a separate extinction for each taxon. The single most likely version of this basic model is one in which $g_i = 0$ for all taxa, and thus can have as many differing ζ values (and thus as many different parameters) as there are taxa. The support (log-likelihood rescaled to the maximum) for this hypothesis is 0. There are a large number of alternatives of intermediate complexity, for example one can examine a two-parameter hypothesis that posits two separate extinctions giving $\zeta_{A\dots F}$ at horizon 20 and $\zeta_{G\dots L}$ at horizon 19.

Log-likelihood ratio tests (e.g. Edwards 1992) now can assess whether the more complex hypothesis predicts the data significantly better than the simpler hypothesis. In Fig. 19.4, the support (i.e. the log-likelihood rescaled to the most likely hypothesis) for $\zeta_{A\dots L} = -7.34$. Whether we can reject this in favour of continuous extinction depends on how many parameters we ascribe to the best 'continuous' hypothesis. Wang and Everson (2007) use S on the grounds that there is one extinction for each of the S taxa. However, in this example there are only seven last appearance beds. This means that we need only seven extinction events to eliminate all 12 taxa, and that the degrees of freedom should be 6 ($7 - 1$) rather than 11. This is important because we would not reject the null hypothesis of a single extinction if we assume that the overall most likely hypothesis requires 12 parameters ($p = 0.20$) whereas we would if we assume that the overall most-likely hypothesis requires seven parameters ($p = 0.02$). However, a two-parameter hypothesis ($\zeta_{A\dots C} = 20$, $\zeta_{D\dots L} = 19$) represents a significant improvement over the one-parameter hypothesis (support = -4.45 ; $p = 0.02$) while withstanding rejection from the overall most likely hypothesis given the assumption of either 12 parameters ($p = 0.54$) or seven parameters ($p = 0.11$) or from the most-likely three-parameter hypothesis ($\zeta_{A\dots C} = 20$, $\zeta_{D\dots F} = 19$, $\zeta_{G\dots L} = 18$; support = -3.13 , $p = 0.10$). This clearly is erroneous given that all 12 taxa share the same time of extinction in the simulations, but restricting all 12 extinctions to the final two beds strongly contradicts the idea of a prolonged extinction event.

Information theory criteria give an alternative means for assessing the hypothesis. An appealing aspect is that these criteria can lead one to conclude that the simpler hypothesis is superior to the more complex hypothesis rather than simply unrejected (Burnham & Anderson 2002). Here, Akaike's information criteria (AIC; Akaike 1973), modified for sample size (Burnham & Anderson 2002), is:

$$\text{AICc} = -2 \ln L[\zeta_{A...L}] + 2K \left(\frac{S}{S - K - 1} \right)$$

where K is the number of parameters and the taxon richness, $S = 12$, is the number of data points. Because the $\text{AICc}[\zeta_{A...L} = 20] = 17.1$ whereas $\text{AICc}[7 \text{ separate } \zeta_i] = 42$, we now prefer the single extinction to the mostly-likely seven-extinction hypothesis. However, we still prefer the best two-parameter hypothesis ($\text{AICc}[\zeta_{A...C} = 20, \zeta_{D...L} = 19] = 14.2$) to the best single-extinction hypothesis. On one hand, one might fault the method for failing to corroborate the correct hypothesis that all of the extinctions happened after the last bed. However, it is reassuring that the favoured hypothesis is a near miss that concentrates all of the extinctions in the last two beds.

Huelsenbeck and Rannala (1997) estimate the probability of zero finds over some interval given preservation rate, R . Although this is intuitively reasonable, R itself is an unknown. R is not explicitly addressed in the Wang and Everson approach, but uncertainty about R should be a concern there too. An intuitive estimate of R is simply:

$$R = \frac{H}{sr} \quad (19.9a)$$

As Fig. 19.4 illustrates, this is biased towards overestimating preservation rates because the denominator under-estimates the true duration (Wagner 2000; Foote 2001a). Instead, an unbiased estimate is given by the finds *within* the known range, i.e.:

$$R = \frac{H - 2}{sr - 2} \quad (19.9b)$$

Even so, we should *not* estimate the likelihood of hypothesized extinction as:

$$L[\zeta|H, sr, R] = \left(1 - \frac{H - 2}{sr - 2} \right)^g$$

First, the uncertainty around the most likely R can be very large, especially for taxa sampled from a few beds over a short duration. Although the probability of the stratigraphic data is maximized with equation 19.7b, the joint probability of the observed data and the hypothesized gap (and thus the joint likelihood of ζ and R ; Wagner 2000) is maximized at:

$$R' = \frac{H - 1}{g + sr - 1} \quad (19.10)$$

giving the maximum:

$$\max L[\zeta, R|sr, H] \propto (1 - R)^g \times ([1 - R]^{sr-H} \times R^{H-1}) \quad (19.11a)$$

$$\propto (1 - R)^{g+sr-H} \times R^{H-1} \quad (19.11b)$$

where the first term in equation 19.9a is the probability of zero finds over the gap and the second term is the sufficient statistic from the binomial probability of $H - 1$ finds over the observed stratigraphic range after the first find. (The binomial combinatoric is constant for all hypotheses.) Note that we now subtract only one horizon as one end of the 'true' duration is implicit to the hypothesis and we truncate only the base.

We now can estimate the likelihood of one or more extinction events using the product of equation 19.11 for all species with $H \geq 2$. Returning to our example in Fig. 19.4, the log-likelihood of a single extinction at horizon 20 for the 10 species with $H_i \geq 2$ is -18.70 . The log-likelihood of each species going extinct at its last appearance increases only to -14.65 , which is insufficient to reject the single ζ hypothesis even when assuming only seven-parameters ($p = 0.23$) and which fares far worse given information theory ($\text{AICc}[\zeta_{A...L} = 20] = 39.8$ vs $\text{AICc}[\zeta_{A...L} = \text{LA}_{A...L}] = 71.3$). However, a hypothesis of extinctions at horizons 19 and 20 still is preferable ($\ln L = 16.25$) given either log-likelihood ratio tests ($p = 0.03$) or AIC ($\text{AICc}[\zeta_{A...C} = 20, \zeta_{D...L} = 19] = 37.8$).

The potential advantage of this approach over the approach of Solow/Wang and Everson is that it allows one to assess whether preservation rates change within a stratigraphic range. If so, then change in R can be explicitly accommodated (see Solow 1993b; Wagner 2000). For example, a sim-

ple test of linearly increasing/decreasing R over horizons would simply test the exact sequence of finds and misses with changing probabilities of finds and misses. Alternatively, one might posit a sudden shift in preservation rate. If the frequency of misses increases/decreases, then the gap between last appearance and extinction will increase/decrease. For example, taxon I in Fig. 19.4 shows an increasing recovery rate over time: it is found once over 10 horizons (2–11) then twice over the next five horizons (12–16). The most likely single R , $\frac{5-2}{17-2} = 0.2$ (equation 19.7b) has a support of $-7.51 (= \ln[0.2^3 \times 0.8^{12}])$. The best gradually shifting R , with $R_2 = 0.05 \rightarrow R_{16} = 0.34$, increases the support only to -7.14 . The best two-phase R , $R_{2...11} = 0.1$ and $R_{12...16} = 0.4$, increases the support only to -6.86 . In neither case would we reject the idea that sampling was homogeneous throughout taxon I's range and thus we have no particular reason to distrust the likelihood given sampling over those 20 horizons. However, Wagner (2000) presents empirical examples where this approach does reject homogeneous R and thus leads to different likelihoods on hypothesized extinctions.

19.3 Prospectus

The study of extinction can proceed in many directions, many of which workers have tentatively explored already. In particular, the use of databases such as the PBDB to ascertain more exact tests of extinction pacing and preservation rates within taxa represents the logical next steps to the methods described above. For example, using observed range-through taxa to estimate R provides an average sampling intensity within an interval. This obviously will not represent all genera. With higher taxa in particular, species richness within a genus or family will strongly affect preservation rate, as will other factors, such as the geographic distributions of the species within the taxa. Foote (2007b) shows that the occurrences of genera typically follow a bell curve over time, that is, occurrences are low both in the early and late intervals of a genus' range and typically peak in the middle. Because genera originate throughout time, this means that different genera will be at different points in their bell curves in any given interval. Moreover, the peaks of these

distributions (i.e. the maximum numbers of occurrences) differ among genera. These two factors combined means that it should be possible to derive a distribution of preservation intensities within each interval rather than a single R 'per genus'. What is most intriguing is that we might use the bell curves themselves in a Bayesian context to assess extinction risk (see, for example, Solow (1993b)). In other words, extinction risk should be relatively low at the peak of an occurrence distribution, when a genus typically is widespread and fairly speciose (see, for example, Foote et al. (2008)). Conversely, the extinction risk should be relatively high near the tail. Whether or not we wish to assume normal curves for occurrences over time requires further testing; however, data such as the PBDB make it possible to assess whether 'commonness' of genera is linked to extinction risk.

Part of the reason why we should expect common genera to be less extinction prone than uncommon genera is that we expect geographically wide-ranging genera to be more common in the fossil record than contemporaneous geographically narrow-ranging genera (e.g. Jablonski 2000). However, because sampling is not randomly distributed over geography, we cannot simply equate occurrences with geographic range. Instead, we need to actually examine occurrences in particular geographic units (see, for example, Foote et al. (2008)). Alroy (in Alroy et al. (2008)) presents methods for accommodating geographic dispersal in sampling standardized estimates of richness over time. We should adopt similar approaches for summarizing geographic distributions when examining the predictions of alternative extinction models.

Even occurrences within temporal and geographic units obscure important information about abundances within localities. In general, there is a good correlation between abundances and occurrences, in that taxa that are abundant anywhere within their geographic range tend to have many occurrences (Buzas et al. 1982; Alroy et al. 2008). However, in gap analyses of particular sections, variation in abundances of specimens has an important effect on the plausibility of missing an extant species (see, for example, Hurlbert (1971)). The simulations sample equal numbers of 'specimens' per bed. However, in reality, all beds are not equal: due

to access, taphonomy, weathering, etc., some beds in the same section will yield many more specimens than others. Weiss and Marshall (1999) explicitly accommodate this when examining extinction boundaries for late Cambrian trilobites. Similarly, McElwain et al. (2009) examine the end-Triassic extinction in plants not in loss of richness, but in shifts on most likely relative abundance distributions. A combination of the two approaches would assess the likelihoods of hypothesized abundances over time/strata, with extinction being an abundance of 0. The logic of shifting R over a section would apply here, except that we now would ask whether we could sample X specimens without finding a species given that it is known from n of N specimens prior to that.

Another promising avenue of research involves the use of ordination techniques to describe fossil turnover without assuming a particular chronostratigraphic scale (e.g. Alroy 1992, 1994, 2000; Sadler and Cooper 2003). Although presented as techniques for repeatable, high-resolution chronostratigraphy, these methods also describe fossil distributions in the same manner needed for confidence interval studies, but at the geographic and geologic scales of synoptic studies.

Finally, one might ask whether gap analyses of extinction events frame the basic hypotheses properly. As done above, the 'complexity' of the hypotheses is the number of extinction events. A single-extinction event thus is the null because it is only one event. Statistically, this is also appropriate because a single event necessarily is less likely than (as in Fig. 19.4) seven events, as it maximizes the minimum number of necessary gaps. However, this

is not quite the appropriate null hypothesis if we look at extinction *rates*. A single event requires two rates: a 'background' rate of no extinction and a single pulsed rate of (nearly) 100% extinction. Thus, we can ask, what is the probability of X extinctions from Y remaining taxa given a hypothesized extinction rate per bed? As Weiss and Marshall (1999) note, the extinction rate provides a prior probability of a lineage existing 1, 2, 3, etc., beds after its last find. This multiplied by likelihood of the gap gives a posterior probability of the taxon existing 1, 2, 3, etc., beds after its last finds. The sum of these posterior probabilities is what we would use as the count of the number of extinctions per bed. For example, in Fig. 19.4, we would have numerous extinctions in the last two beds given a hypothesis of uniform extinction rates and the sampling data. This would be strong evidence against a gradual hypothesis. However, if last appearances were distributed more evenly throughout the section, then it would be much more difficult to extend the durations to the top of the section as not only would they become increasingly improbable, but the plausibility of the taxon still being extant also would become increasingly improbable.

19.4 Key points

1. The fossil record provides data capable of testing a wide variety of extinction hypotheses.
2. This can be done at very broad temporal scales as well as over intervals closer to 'ecological time'.
3. Accommodating preservation is critical at all temporal scopes of palaeontological analyses of extinction.