

CONSULTANT'S FORUM

Multiple Comparison of Entropies with Application to Dinosaur Biodiversity

Kathleen S. Fritsch

Mathematics and Computer Science, The University of Tennessee at Martin,
Martin, Tennessee 38238, U.S.A.

and

Jason C. Hsu

Department of Statistics, The Ohio State University, Columbus, Ohio 43210-1247, U.S.A.
email: jch@stat.ohio-state.edu

SUMMARY. Did the biodiversity of dinosaurs decline, or did it remain more or less constant before their mass extinction 65 million years ago? Sheehan et al. (1991, *Science*, 835–839) reported that the biodiversity of families of dinosaur species remained more or less constant preceding their extinction, suggesting extinction due to a cataclysmic event such as an asteroid strike. But that claim was based on the incorrect interpretation that a large p value associated with a test of null hypothesis of equality supports that null hypothesis. To assess whether there is a basis for such a claim, we formulate the problem as one of *practical equivalence*, in analogy to *bioequivalence*. We then develop reliable practical equivalence confidence intervals for differences of entropies by applying the bootstrap-t technique to a nearly pivotal quantity. Confidence intervals for changes in the biodiversity of dinosaurs are then computed, allowing the reader to assess whether there is evidence of near constancy of dinosaur biodiversity before extinction.

KEY WORDS: Biodiversity; Entropy; Equivalence; Multiple comparisons.

1. Dinosaur Biodiversity and Dinosaur Extinction

The reign of the dinosaurs lasted till the end of the Cretaceous period 65 million years ago. Why did the dinosaurs become extinct? Some theories state that the extinction was due to a cataclysmic event such as a meteor strike, whereas other theories state the extinction was due to causes that worked slowly over time such as climatic changes (cf., Bakker, 1986; Norman, 1991; Lambert, 1993). To statistically answer the question whether the extinction was sudden or gradual, Sheehan et al. (1991) reported on an effort to sample the relative abundances of different types of dinosaurs from three time intervals in the late Cretaceous period, immediately preceding dinosaurs' extinction. If the community of dinosaurs was declining in "biodiversity" over time, this would indicate that they reached an evolutionary dead end and might have had difficulty adjusting to changing environments, giving credence to the gradual extinction theory. If, however, "biodiversity" remained relatively constant in the waning periods of the Cretaceous, this would give credence to the theory that the extinction occurred suddenly.

Suppose that the maximum number of families of species

in a community is S . Let p_1, \dots, p_S be the proportions of individuals from each family. Then, to measure biodiversity, one index, that Sheehan et al. (1991) used was entropy, usually denoted by H and called the Shannon Index in ecology, $H = -\sum_{m=1}^S p_m \log p_m$. For a detailed discussion of the rationale of using entropy to measure biodiversity, see Pielou (1977, Section 19.2). Note that H is maximized when all p_m are equal and minimized when the community comprises only one family.

Sheehan et al. (1991) tested the hypothesis

$$H_0^- : H_1 = H_2 = H_3, \quad (1)$$

where H_i is the entropy of the i th time period, with 1 = upper (most recent) interval, 2 = middle interval, and 3 = lower (oldest) interval. Having accepted the null hypothesis (1), they "reject(ed) the hypothesis that the dinosaurian part of the ecosystem was deteriorating during the latest Cretaceous" (Sheehan et al., 1991, p. 838). Obviously, they had the common misconception that the acceptance of a null hypothesis is a support for that hypothesis, ignoring the possibility that the acceptance may be because of insufficient data.

To determine whether the conclusion of Sheehan et al. is warranted, we develop confidence intervals for the differences of entropies $H_i - H_j$, $i \neq j$, so that if entropies within δ (>0) of each other can be considered practically equivalent, then these confidence intervals can be used to test the null hypothesis

$$H_0 : |H_i - H_j| > \delta \quad \text{for some } i \neq j \quad (2)$$

against the alternative

$$H_a : |H_i - H_j| \leq \delta \quad \text{for all } i \neq j. \quad (3)$$

Readers with their own δ can then decide for themselves whether data were sufficient to support the alternative hypothesis (3) of practically equivalent biodiversity.

Another misconception in ecological literature (e.g., Magurran, 1988, p. 76) is that having replications of sample diversity measures makes statistical methods based on i.i.d. normal errors (e.g., analysis of variance) applicable. This is false because, as an estimate of the population entropy, the variance of the sample entropy depends on the unknown p_1, \dots, p_S . Our development of confidence intervals for $H_i - H_j$ in Section 4 is based on the bootstrap-t method, incorporating expressions for the biases and standard deviations to make the quantity being bootstrapped as pivotal as possible.

2. The Data

The dinosaur bones collected for the study of Sheehan et al. (1991) were gathered from sites in North Dakota and Montana, that are part of the Hell Creek formation. The entire Hell Creek formation was deposited over about 2.2 million years near the end of the Cretaceous period. The formation was divided into three equal stratigraphic intervals, with each third representing approximately 730,000 years.

Individual species of dinosaurs are relatively difficult to distinguish from their bones, but the family groups are distinct enough to enable the bones to be more easily classified at this level. Thus, the researchers classified the bones according to their family group. In all, eight families were identified. Table 1 lists the eight dinosaur families, and Table 2 lists the number of individual dinosaurs identified per family from North Dakota and Montana combined.

3. Distribution of Sample Entropy

For convenience, we shall often say “family” when referring to “families of species.” Let $\mathbf{p} = (p_1, \dots, p_S)$ be the vector of proportions of S families in a community. Let $\mathbf{X} = (X_1, \dots, X_S)$ be the vector of individuals counted in each

Table 1
Dinosaur families identified

Family
Ceratopsidae
Hadrosauridae
Hypsilophodontidae
Pachycephalosauridae
Tyrannosauridae
Ornithomimidae
Saurornithoididae
Dromaeosauridae

Table 2

*Number of individual dinosaurs per family.
The order in which individual families are listed follows the order of families in Table 1.*

Interval	Counts
Upper	(50, 29, 3, 0, 3, 4, 1, 0)
Middle	(53, 51, 2, 0, 3, 8, 6, 0)
Lower	(19, 7, 1, 0, 2, 0, 3, 0)

family, and let N be the total sample size. One reasonable way to estimate H from a sample is to plug in the maximum likelihood estimate for \mathbf{p} , $\hat{\mathbf{p}} = \mathbf{X}/N$, yielding

$$\hat{H} = - \sum_{m=1}^S \hat{p}_m \log(\hat{p}_m).$$

By the delta method, as $N \rightarrow \infty$,

$$\sqrt{N}(\hat{H} - H) \xrightarrow{d} N(0, \log(\mathbf{p})' [\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}'] \log(\mathbf{p})).$$

However, the estimator \hat{H} is biased (Basharin, 1959):

$$E(\hat{H}) = H - (S-1)/2N + O(1/N^2).$$

Because $-(S-1)/2N$ does not depend on any unknown parameters, it can be used as a bias-correction term. In this and subsequent development, we gauge the accuracy of point and interval estimates using simulated data similar to the observed dinosaur data generated from geometric distributions. Details of this model and the rationale for using it are given in the Appendix. Our simulation showed that $-(S-1)/2N$ captures most of the bias.

4. Tests Based on Equivalence Confidence Intervals

The null hypothesis (2) being tested is the union of the hypotheses $H_0^{ij} : |H_i - H_j| > \delta$, whereas the alternative (3), which represents the desired inference, is the intersection of the alternatives $H_a^{ij} : |H_i - H_j| \leq \delta$. It follows from Berger (1982) that when the hypothesis-alternative pair is in this form, rejecting (2) when each of H_0^{ij} is rejected at level- α results in a level- α test for (2).

Let U_{ij} be an upper and L_{ij} be a lower $1 - \alpha$ confidence bound on $H_i - H_j$. Berger and Hsu (1996, Section 5.1) showed that $[L_{ij}^-, U_{ij}^+] = [\min(L_{ij}, 0), \max(U_{ij}, 0)]$ is a $100(1 - \alpha)\%$ confidence interval for $H_i - H_j$ and that rejecting H_0^{ij} when $[L_{ij}^-, U_{ij}^+] \subseteq [-\delta, \delta]$ is a level- α test for H_0^{ij} . Thus, it remains to develop reliable confidence intervals for differences in entropies. We first investigate confidence intervals for a single entropy, then confidence intervals for differences of entropies.

4.1 Confidence Bounds for a Single Entropy

For any entropy H , let L_c and U_c be the lower and upper $100(1 - \alpha)\%$ confidence bounds on H with bias correction, written as:

$$L_c = \hat{H} + (S-1)/2N - q_{1-\alpha}^c \hat{\sigma} / \sqrt{N},$$

$$U_c = \hat{H} + (S-1)/2N - q_{\alpha}^c \hat{\sigma} / \sqrt{N},$$

where

$$\hat{\sigma}^2 = \log(\hat{\mathbf{p}})' [\text{diag}(\hat{\mathbf{p}}) - \hat{\mathbf{p}}\hat{\mathbf{p}}'] \log(\hat{\mathbf{p}}).$$

(Note that $\hat{\sigma}^2 = 0$ when $\hat{p}_1 = \dots = \hat{p}_S$. A provision was made to use a resampling-based estimate for $\hat{\sigma}$ should this happen; however, because of the skewness of the distribution of the dinosaur family, this unlikely event never occurred in our subsequent computations.)

We need the quantiles q_α^c and $q_{1-\alpha}^c$ such that:

$$P\left(q_\alpha^c \leq \frac{\hat{H} - H + (S-1)/2N}{\hat{\sigma}/\sqrt{N}}\right) = 1 - \alpha,$$

$$P\left(\frac{\hat{H} - H + (S-1)/2N}{\hat{\sigma}/\sqrt{N}} \leq q_{1-\alpha}^c\right) = 1 - \alpha.$$

The exact values of these quantiles depend on \mathbf{p} , which is unknown. In the following sections, we investigate the accuracy of confidence bounds that result from various approximations to these quantiles.

Normal and t confidence bounds. Using first-order asymptotics, the standard normal confidence bounds without bias correction would refer $(\hat{H} - H)/(\hat{\sigma}/N^{1/2})$ to the standard normal distribution, whereas the standard normal confidence bounds with bias correction refers

$$\frac{\hat{H} - H + (S-1)/2N}{\hat{\sigma}/\sqrt{N}} \quad (4)$$

to the standard normal distribution, setting $q_\alpha^c = z_\alpha$ and $q_{1-\alpha}^c = z_{1-\alpha}$, where z_γ is the γ quantile of the standard normal distribution.

Our simulation showed that the distribution of the bias-corrected standardized sample entropy (4), although more closely centered around 0 than the non-bias-corrected standardized sample entropy, does not agree with the standard normal distribution in the tails very well. We were also unable to find a t distribution with some degrees of freedom that offers a better approximation. In the subsequent sections, we explore the feasibility of using bootstrap techniques to better approximate the distribution of (4).

Bootstrap- t confidence bounds. Our application of the bootstrap- t technique in Efron and Tibshirani (1993) refers (4) to the distribution of

$$\frac{\hat{H}^* - \hat{H} + (S-1)/2N}{\hat{\sigma}_*/\sqrt{N}}, \quad (5)$$

where \hat{H}^* is the entropy computed from a typical bootstrap sample generated from the observed data, and $\hat{\sigma}_*$ is its corresponding standard deviation estimate. Having generated B bootstrap samples, the bootstrap- t lower and upper $100(1-\alpha)\%$ confidence bounds L_b and U_b , written as $L_b = \hat{H} + (S-1)/2N - q_{1-\alpha}^b \hat{\sigma}/N^{1/2}$ and $U_b = \hat{H} + (S-1)/2N - q_\alpha^b \hat{\sigma}/N^{1/2}$, respectively, use quantile estimates q_α^b and $q_{1-\alpha}^b$, which are the $[B\alpha] + 1$ and $[B(1-\alpha)] + 1$ order statistics of the B bootstrap quantities (5), where $[\cdot]$ is the greatest integer function. If the standardized sample entropy (4) is nearly pivotal, then the distribution of (5) will be close to the distribution of (4), and the bootstrap- t technique will work well.

Percentile and BC_a confidence bounds. Suppose that $\hat{H}^{*1}, \dots, \hat{H}^{*B}$ are B bootstrap estimates of H . Although the

bootstrap- t technique uses order statistics of the standardized statistic (5) as quantiles of the statistic (4), the percentile and bias-corrected and accelerated (BC_a) techniques of Efron and Tibshirani (1993) use select order statistics of \hat{H}^{*i} itself as confidence bounds for H .

The percentile method uses the $[B\alpha]$ and $[B(1-\alpha)]$ order statistics of the B ordered values of \hat{H}^{*i} as the confidence bounds. We also computed bias-corrected percentile bounds from the order statistics of $\hat{H}^{*i} + (S-1)/2N$.

The BC_a method incorporates "automatic" bias correction. Let $\hat{H}^{*(\gamma)}$ be the 100γ percentile of $\hat{H}^{*1}, \dots, \hat{H}^{*B}$. Then we define

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{H}^{*i} < \hat{H}\}}{B}\right),$$

where Φ is the standard normal distribution function, and

$$\hat{a} = \frac{\sum_{i=1}^N (\hat{H}^{(\cdot)} - \hat{H}^{(i)})^3}{6 \left\{ \sum_{i=1}^N (\hat{H}^{(\cdot)} - \hat{H}^{(i)})^2 \right\}^{1.5}},$$

where $\hat{H}^{(i)}$ is the sample entropy computed without the i th observation and $\hat{H}^{(\cdot)}$ is the average of \hat{H} ; the lower and the upper $100(1-\alpha)\%$ confidence bounds for H given by the BC_a technique are $\hat{H}^{*(\alpha_1)}$ and $\hat{H}^{*(\alpha_2)}$, respectively, where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_\alpha}{1 - \hat{a}(\hat{z}_0 + z_\alpha)}\right),$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha})}\right).$$

Comparison of confidence bounds for a single entropy. We carried out a simulation study to compare the accuracy of the confidence bounds. Data were generated from a multinomial (50, \mathbf{p}) distribution, where \mathbf{p} is defined by various geometric models with $S = 8$. Table 3 lists the percentage of confidence bounds that fail to bound the true entropy value. "Below" represents the estimated probability that the upper confidence bound is smaller than the true value, $P(U < H)$, whereas "Above" represents the estimated probability that the lower confidence bound is higher than the true value, $P(L > H)$. The target values for "Below" and "Above" are α .

The noncoverage probabilities for the standard normal bounds were estimated based on 10,000 sets of simulated bounds, whereas the noncoverage probabilities, for all other confidence bounds, were estimated based on 3000 sets of simulated bounds. Each bootstrap bound was computed using $B = 7500$ bootstrap samples.

The percentile and BC_a confidence intervals appear to be the most liberal. Bias correction improves the percentile method, making its total coverage probability $1 - (\text{Below} + \text{Above})$ comparable to that of the standard normal method. However, the total coverage probability of the bootstrap- t interval comes closest to the desired $1 - 2\alpha$ when compared with all other methods. However, the bootstrap- t interval has some imbalance in that its lower bound appears to be conservative, whereas its upper bound appears to be liberal.

Table 3
Noncoverage probabilities of confidence bounds for H

α (%)	Method	$k = 0.8$		$k = 0.6$		$k = 0.4$	
		Below (%)	Above (%)	Below (%)	Above	Below (%)	Above
5	Percentile	18.2	1.2	24.2	0.6%	35.9	0.2%
	Bias-corrected percentile	9.1	6.2	10.8	3.5%	16.9	1.3%
	Bootstrap-t	10.1	3.2	8.0	3.5%	8.5	4.6%
	BC _a	12.6	5.2	11.7	5.6%	14.2	7.0%
	Standard normal	5.9	10.0	7.1	8.7%	8.0	7.5%
2.5	Percentile	13.7	0.5	16.8	0.2%	27.1	1/3000
	Bias-corrected percentile	5.7	2.9	6.9	1.3%	12.0	0.5%
	Bootstrap-t	5.2	1.6	4.0	1.3%	5.6	2.1%
	BC _a	11.4	3.0	8.4	2.2%	11.2	3.9%
	Standard normal	4.3	6.6	4.2	5.2%	4.9	4.2%
1	Percentile	9.3	0.2	11.3	2/3000	19.0	0/3000
	Bias-corrected percentile	3.3	0.9	4.7	0.4%	7.8	0.1%
	Bootstrap-t	0.8	0.3	2.5	0.7%	2.5	0.7%
	BC _a	7.1	1.3	5.3	1.6%	8.6	2.1%
	Standard normal	2.1	3.5	2.5	2.5%	2.7	2.6%
Target H		0.6255		1.1162		1.5958	

4.2 Confidence Bounds for the Difference of Two Entropies

For each i , $i = 1, 2, 3$, let

H_i = entropy of the i th interval,

N_i = number of observation from the i th interval,

\hat{H}_i = sample entropy estimate of H_i ,

B_i = first-order bias of $\hat{H}_i = (S - 1)/(2N_i)$,

$\hat{\sigma}_i$ = sample estimate of the standard deviation of \hat{H}_i .

Recall that, to assess practical equivalence of entropies, we need individual $100(1 - \alpha)\%$ lower and upper confidence bounds L_{ij} and U_{ij} on $H_i - H_j$ to test $H_0^{ij} : |H_i - H_j| > \delta$ versus $H_a^{ij} : |H_i - H_j| \leq \delta$. Writing the bounds as:

$$L_{ij} = \hat{H}_i - \hat{H}_j - (B_i - B_j) - q_{ij,1-\alpha} \sqrt{\hat{\sigma}_i^2/N_i + \hat{\sigma}_j^2/N_j},$$

$$U_{ij} = \hat{H}_i - \hat{H}_j - (B_i - B_j) - q_{ij,\alpha} \sqrt{\hat{\sigma}_i^2/N_i + \hat{\sigma}_j^2/N_j},$$

we need the quantiles $q_{ij,\alpha}$ and $q_{ij,1-\alpha}$ such that:

$$P \left(q_{ij,\alpha} \leq \frac{\hat{H}_i - \hat{H}_j - (H_i - H_j) - (B_i - B_j)}{\sqrt{\hat{\sigma}_i^2/N_i + \hat{\sigma}_j^2/N_j}} \right) = 1 - \alpha,$$

$$P \left(\frac{\hat{H}_i - \hat{H}_j - (H_i - H_j) - (B_i - B_j)}{\sqrt{\hat{\sigma}_i^2/N_i + \hat{\sigma}_j^2/N_j}} \leq q_{ij,1-\alpha} \right) = 1 - \alpha.$$

To obtain these quantiles, the standard normal technique refers the distribution of

$$\frac{\hat{H}_i - \hat{H}_j - (H_i - H_j) - (B_i - B_j)}{\sqrt{\hat{\sigma}_i^2/N_i + \hat{\sigma}_j^2/N_j}}$$

to the standard normal distribution, whereas the bootstrap-t technique refers it to the distribution of

$$\frac{\hat{H}_i^* - \hat{H}_j^* - (\hat{H}_i - \hat{H}_j) - (B_i - B_j)}{\sqrt{\hat{\sigma}_{*i}^2/N_i + \hat{\sigma}_{*j}^2/N_j}},$$

where \hat{H}_i^* is the sample entropy computed from a typical bootstrap sample generated from the i th interval data, and $\hat{\sigma}_{*i}$ is its corresponding standard deviation estimate. The bias-corrected percentile method uses the $\lfloor B\alpha \rfloor$ and $\lfloor B(1 - \alpha) \rfloor$ order statistics of the B ordered values of $\hat{H}_i^* + (S - 1)/2N_i - (\hat{H}_j^* + (S - 1)/2N_j)$ as the confidence bounds. Because of the inferior performance of the BC_a interval in the single-sample case, it was not pursued in the two-sample case.

Table 4 lists estimated noncoverage probabilities of confidence bounds for $H_i - H_j$. Noncoverage probability estimates for the bootstrap bounds are based on 3000 simulated data sets, whereas noncoverage probability estimates for standard normal bounds are based on 10,000 simulated data sets. Each bootstrap bound was computed using $B = 7500$ bootstrap samples, and each data set is simulated from a multinomial $(50, \mathbf{p})$ distribution, where \mathbf{p} is defined by one of three geometric models with $S = 8$. Again, "Below" is the estimated probability that the upper confidence bound is lower than the true value, $P(U_{ij} < H_i - H_j)$, whereas "Above" is the estimated probability that the lower confidence bound is higher than the true value, $P(L_{ij} > H_i - H_j)$. The target values for "Below" and "Above" are α .

Although all the confidence bounds appear to be somewhat liberal, the bootstrap-t bounds are more accurate in the sense that their estimated coverage probabilities come closest to $1 - \alpha$ for both the upper and lower bounds. When U_{ij} and L_{ij} do not have coverage probabilities exactly equal to $1 - \alpha$, Berger and Hsu (1996) have shown that the size of the test that rejects H_0^{ij} , when $[L_{ij}^-, U_{ij}^+] \subseteq [-\delta, \delta]$, is

Table 4
Noncoverage probabilities of confidence bounds for $H_1 - H_2$

α (%)	Method	k_1, k_2	0.6, 0.6	0.6, 0.4	0.6, 0.8
		$H_i - H_j$	0	-0.4795	0.4907
5	Percentile	Below	5.4%	3.7%	8.2%
		Above	6.5%	8.8%	3.8%
	Bootstrap-t	Below	6.2%	6.9%	6.2%
		Above	6.3%	6.3%	6.1%
	Standard normal	Below	7.4%	6.6%	8.9%
		Above	7.6%	8.2%	5.9%
2.5	Percentile	Below	3.0%	2.1%	5.2%
		Above	3.6%	5.0%	2.3%
	Bootstrap-t	Below	3.6%	4.1%	3.5%
		Above	4.0%	3.3%	3.3%
	Standard normal	Below	4.6%	4.1%	5.5%
		Above	4.3%	5.0%	3.3%
1	Percentile	Below	1.5%	1.0%	2.6%
		Above	1.5%	2.4%	1.2%
	Bootstrap-t	Below	1.3%	1.9%	1.5%
		Above	1.6%	1.7%	1.0%
	Standard normal	Below	2.4%	2.2%	2.6%
		Above	2.2%	2.4%	1.7%

$$\max \left\{ \sup_{H_i - H_j > \delta} P(U_{ij} < H_i - H_j), \sup_{H_i - H_j < -\delta} P(L_{ij} > H_i - H_j) \right\}. \tag{6}$$

Therefore, for the two cases of $H_i - H_j \neq 0$ simulated corresponding to $\delta = 0.48$ and 0.49 , approximately, bootstrap-t bounds appear to be the most accurate in assessing practical equivalence of entropies as they have the smallest max{Below, Above} for all α values.

5. Dinosaur Biodiversity Data Reanalyzed

Table 5 gives the 95% bootstrap-t and standard normal practical equivalence confidence intervals $[L_{ij}^-, U_{ij}^+]$ for the differences in entropy for all possible pairs of intervals for the dinosaur data. Note that the standard normal and bootstrap-t intervals involving only the upper and middle intervals are similar, but the bootstrap-t intervals involving the lower interval are quite a bit wider than the standard normal counterparts. This is because of the relatively large sample sizes for the upper and middle intervals (90 and 123, respectively) and the relatively small sample size for the lower interval (32).

Table 5
Ninety-five percent practical equivalence confidence intervals for differences in entropy

Difference	Bootstrap-t	Standard normal
$H_M - H_U$	$[-0.12, 0.31]$	$[-0.11, 0.29]$
$H_L - H_M$	$[-0.35, 0.29]$	$[-0.27, 0.12]$
$H_L - H_U$	$[-0.27, 0.41]$	$[-0.17, 0.23]$

Given the results in Section 4.2, practical equivalence inference should be based on bootstrap-t confidence intervals. Thus, to reject the null hypothesis $H_0 : |H_i - H_j| > \delta$ for some $i \neq j$ in favor of the practical equivalence alternative $H_a : |H_i - H_j| \leq \delta$ for all $i \neq j$, the quantity δ defining practical equivalence has to be at least 0.41.

Currently, there is no suggested δ in the literature on paleontology. To determine such a δ , one possibility is to consider an analogous community, one that has declined in biodiversity (because of pollution, for example). If a positive lower confidence bound on the decrease in biodiversity can be calculated, using data before and after the community has encountered difficulty, this lower confidence bound may serve as a delta defining practical equivalence in biodiversity. Another possibility is to examine multinomial probability histograms that correspond to the maximum observed confidence limit for differences in entropy and contemplate their practical equivalence.

ACKNOWLEDGEMENTS

We thank Dr Peter Sheehan and Prof David Fastovsky for use of the data in Table 2, as well as for discussions on measures and practical equivalence of biodiversity. We thank Rachel Thiet for references on biodiversity measures in ecology and the associate editor and the referee for their suggestions.

RÉSUMÉ

Qu'en était-il de la biodiversité des dinosaures, préalablement à leur extinction totale, il y a de cela 65 millions d'années? Cette biodiversité avait-elle décliné, ou était-elle restée plus ou moins constante? Selon Sheehan et al (1991), la biodiversi-

té des espèces de dinosaures aurait été stable dans la période précédant leur extinction (ce qui suggérerait que cette extinction serait due, selon eux, à un événement cataclysmique—par exemple une collision de notre planète avec un astéroïde). Mais cette affirmation repose sur une erreur d'interprétation statistique: sous couvert d'avoir observé une p -value importante à l'occasion d'un test destiné à mettre en évidence une différence, l'hypothèse nulle de l'égalité des valeurs testées se verrait prouvée. Pour essayer d'évaluer sérieusement si l'on peut étayer ou non la thèse des auteurs, nous proposons, par analogie avec la bioéquivalence, de formuler le problème comme un problème d'équivalence pragmatique. Dans ce cadre, nous construisons des intervalles de confiance de différences d'entropie, à l'aide de bootstraps réalisés autour d'une quantité quasi-pivotale; ces intervalles de confiance, qui portent sur les changements, dans le temps, de la biodiversité des dinosaures, permettent en fin de compte au lecteur d'évaluer la solidité ou non de la thèse en présence, à savoir que la biodiversité des dinosaures, avant leur extinction, aurait été à peu près stable.

REFERENCES

- Bakker, R. T. (1986). *The Dinosaur Heresies*. New York: William Morrow.
- Basharin, G. P. (1959). On a statistical estimate for the entropy of a sequence of independent random variables. *Theory of Probability and Its Applications* 4, 333–336.
- Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* 24, 295–300.
- Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests, and equivalence confidence sets. *Statistical Science* 11, 283–315.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.
- Lambert, D. (1993). *The Ultimate Dinosaur Book*. London: Dorling Kindersley.
- Magurran, A. E. (1988). *Ecological Diversity and Its Measurements*. Princeton, New Jersey: Princeton University Press.
- Norman, D. (1991). *Dinosaur!* New York: Prentice Hall.
- Pielou, E. C. (1977). *Mathematical Ecology*. New York: Wiley.
- Sheehan, P. M., Fastovsky, D. E., Hoffmann, R. G., Berghaus, C. B., and Gabriel, D. L. (1991). Sudden extinction of the dinosaurs: Latest Cretaceous, Upper Great Plains, U.S.A. *Science* 254, 835–839.

Received February 1998. Revised January 1999.
Accepted January 1999.

APPENDIX

The Geometric Model

Among ecological models for species distribution, the geometric model seems to be most compatible with the observed dinosaur data. If p_i is the proportion of individuals in the i th most abundant family, then the model is

$$p_m = \frac{k(1-k)^{m-1}}{1-(1-k)^S}, \quad m = 1, \dots, S \quad (0 < k < 1).$$

Figure 1 shows the sample proportions from the dinosaur data for the three stratigraphic intervals, as well as the geometric probabilities with $k = 0.6$. Because, at least for the upper and lower intervals, the sample proportions follow a shape that is similar to probabilities from a geometric model, we generated data from a range of geometric models ($k = 0.4, 0.6$, and 0.8) to assess the accuracy of point and interval estimators.

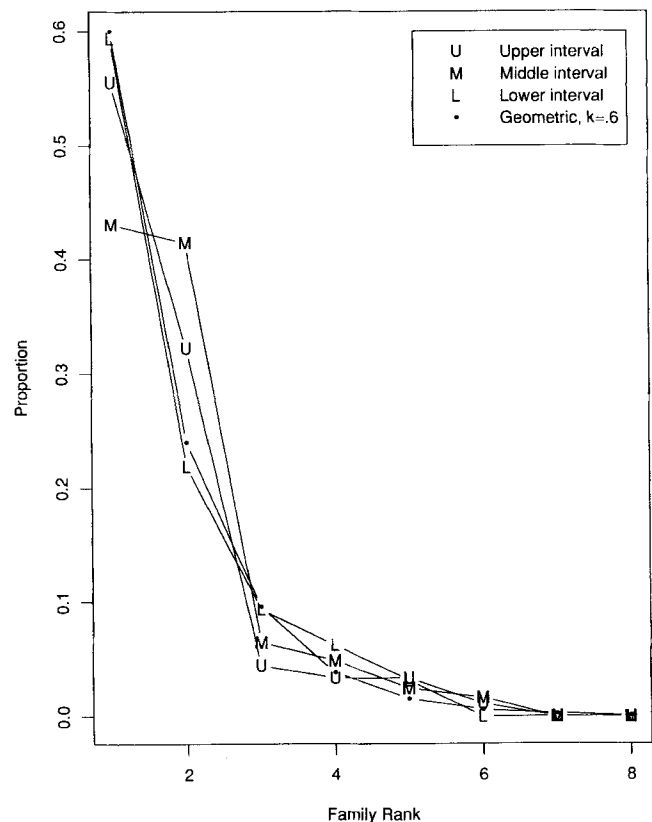


Figure 1. Sample proportions for the three intervals and geometric $k = 0.6$ probabilities.