

## Robustness of coverage estimators for assessing the completeness of paleontological sampling

\*Takao Ubukata<sup>1</sup>

1.Division of Geology & Mineralogy, Department of Earth & Planetary Sciences, Kyoto University

The study on global diversity change requires an adequate method to correct diversity estimates that are readily affected by unevenness of sampling density. The diversity estimates depend on true diversity as well as sampling effort. Therefore, rarefaction to a uniform sample size causes overcorrection of diversity measures which results in an underestimate of diversity particularly when true diversity is high. Alroy (2010) introduced the shareholder quorum subsampling (SQS) method which allows to remove only the effect of sampling intensity without overcorrection. The SQS method needs to estimate the proportion of the total number of individuals belonging to an already observed species in the sample, or sample coverage. The sample coverage can be accurately estimated using Good-Turing frequency estimation when the number of occurrences for each species follows a binomial distribution and the sample size is large enough. In natural community, however, individuals are not evenly distributed among species and often a few species predominate in a community. In addition, a small sample size is common in paleontological studies, especially if we focus on regional or local diversity of a particular taxonomic group. Chao and Jost (2012) proposed coverage-based rarefaction and extrapolation methods using an improved version of the Good's estimator of coverage. The sample coverage can also be estimated using a rarefaction curve calculated from the sample of interest: the final slope of the rarefaction curve represents the probability of encountering an individual hitherto unseen in the sample which is equal to one minus the sample coverage. However, precision and accuracy of the coverage estimators have not yet been fully verified for samples with various sizes from populations with a variety of statistical distribution.

Here I assessed the robustness of the above estimators of sample coverage through a simulation study. 10,000 bootstrap samples with various sizes were taken from the populations that follow logarithmic normal distributions with various sets of parameters. The three kinds of coverage estimators as well as a true coverage were computed for each bootstrap sample. The random error of estimation of the sample coverage was assessed by calculating the standard deviation of true coverage at a fixed value of each coverage estimate. The result of the simulation shows that the precision of any of the estimators examined tends to decrease with decreasing the sample coverage. The random error of estimation is considerably great in the case with a low coverage even if the sample size is pretty large. It also increases fairly as the median of the population increases.

Keywords: paleobiodiversity, completeness of fossil record, rarefaction