

# Adjusting the bias of bootstrap probability with “negative” sample size and its applications to clustering and multiple comparisons

Hidetoshi Shimodaira

*Kyoto University / RIKEN AIP*

*E-mail: shimo@i.kyoto-u.ac.jp*

**Abstract:** Bootstrap resampling has been widely used for calculating confidence level of discrete decisions. The bootstrap probability of a specified hypothesis is calculated as the frequency of observing the same hypothesis in bootstrapped datasets. Although the bootstrap probability can be interpreted as a Bayesian posterior probability, it is biased as a frequentist p-value caused by the “curvature” of the boundary surface of the hypothesis in the parameter space (Efron and Tibshirani 1998). This bias is successfully adjusted via the multiscale bootstrap (Shimodaira 2004) by utilizing the scaling-law of the bootstrap probability in “m-out-of-n” bootstrap. It is very surprising that the unbiased p-value is obtained by extrapolating the normalized bootstrap probability to  $m = -n$ . By now, the method has been widely used in clustering (pvclust) and phylogenetics (CONSEL). We review the method, and also its extension to selective inference (Terada and Shimodaira 2017).