# Selective inference for the problem of regions via multiscale bootstrap with applications to clustering and regression

**Yoshikazu Terada**

*Osaka University*
*E-mail: terada@sigmath.es.osaka-u.ac.jp*

**Abstract:** A general approach to selective inference is considered for hypothesis testing of the null hypothesis represented as an arbitrary shaped region in the parameter space of multivariate normal model. This approach is useful for hierarchical clustering where confidence levels of clusters are calculated only for those appeared in the dendrogram, thus subject to heavy selection bias. Our computation is based on a raw confidence measure, called bootstrap probability, which is easily obtained by counting how many times the same cluster appears in bootstrap replicates of the dendrogram. We adjust the bias of the bootstrap probability by utilizing the scaling-law in terms of geometric quantities of the region in the abstract parameter space, namely, signed distance and mean curvature. Although this idea has been used for non-selective inference of hierarchical clustering, its selective inference version has not been discussed in the literature. Our bias-corrected p-values are asymptotically second-order accurate in the large sample theory of smooth boundary surfaces of regions, and they are also justified for nonsmooth surfaces such as polyhedral cones. The p-values are asymptotically equivalent to those of the iterated bootstrap but with less computation. The proposed algorithm is applied to practical selective inference problems on hierarchical clustering and regression.