

Nonparametric Density Estimation and Clustering with Application to Cosmology

Woncheol Jang
Carnegie Mellon University

Abstract

We present a clustering method based on nonparametric density estimation. We use Kernel smoothing and orthogonal series estimators to estimate the density f and then we extract the connected components of the level set using a modified Cuevas et al (2000) algorithm. We extend an idea due to Stein (1981) and Beran and Dümbgen (1998) to construct confidence sets for the level set $\{f > \delta_c\}$ using the asymptotic distribution of loss function. Specifically, we show the stochastic convergence of the pivot process, $B_n(\lambda_p) = \sqrt{n}(L_p(\lambda_p) - \hat{S}_p(\lambda_p))$ where $L_p(\lambda_p)$ and $S_p(\lambda_p)$ are the loss function and the estimated risk function with the smoothing parameter λ_p . Inverting the pivot provides a confidence set for the coefficient of the orthogonal series estimator and furthermore one can construct a confidence set for functionals of f . We consider applications in astronomy and other fields.

Acknowledgment

This is joint work with Larry Wasserman, Chris Genovese and Bob Nichol.

References

- [1] Beran, R. and Dümbgen. (1998). Modulation of Estimators and Confidence Sets. *Ann. Statist.*, **26**, 1826-1856.
- [2] Cuevas, A., Febrero, M. and Fraiman, R. (2000). Estimation the number of clusters. *The Canadian Journal of Statistics*, **28**, 367-382.
- [3] Jang, W. and Wasserman, L. (2003). Confidence Sets for Densities and Clusters. In preparation.
- [4] Stein, C (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135-1151.