

# Exploring the use of variable bandwidth kernel density estimators

Isaías H. Salgado-Ugarte  
F.E.S. Zaragoza U.N.A.M. Biología &  
Depto. de Biología, U.A.M. Iztapalapa  
isalgado@servidor.unam.mx

Marco A. Pérez-Hernández  
Depto. de Biología, U.A.M. Iztapalapa; México  
maph@xanum.uam.mx

**Abstract.** Variable bandwidth kernel density estimators increase the window width at low densities and decrease it where data concentrate. This represents an improvement over the fixed bandwidth kernel density estimators. In this article, we explore the use of one implementation of a variable kernel estimator in conjunction with several rules and procedures for bandwidth selection applied to several real datasets. The considered examples permit us to state that when working with tens or a few hundreds of data observations, least-squares cross-validation bandwidth rarely produces useful estimates; with thousands of observations, this problem can be surpassed. Optimal bandwidth and biased cross-validation (BCV), in general, oversmooth multimodal densities. The Sheather–Jones plug-in rule produced bandwidths that behave slightly better in this respect. The Silverman test is considered as a very sophisticated and safe procedure to estimate the number of modes in univariate distributions; however, similar results could be obtained with the Sheather–Jones rule, but at a much lower computational cost. As expected, the variable bandwidth kernel density estimates showed fewer modes than those chosen by the Silverman test, especially those distributions in which multimodality was caused by several noisy minor modes. More research on the subject is needed.

**Keywords:** st0036, kernel density estimation, bandwidth, cross validation, multimodality test