

Using Stata and Python for nonparametric smoothing estimation

Salgado-Ugarte, I.H.*, V.M. Saito-Quezada, M.M. Salgado-Saito y N.I. Plascencia-Díaz

The choice of the bidwidth/bandwidth (smoothing parameter) is one of the most relevant issues of density estimation. For univariate data, Cross validation (CV) is a well-known procedure for automatic choosing the smoothing parameter. Least squares CV minimizes an estimation of the Integrated Squared Error by means of the “leave one out estimate”, a computationally intensive method ideal for WARPing implementation. On the other hand, by directly estimating the asymptotic mean integrated squared error, it is obtained a biased estimation called Biased Cross Validation (BCV). With asymmetrical or heavy tailed distribution BCV tends to oversmooth while L2CV despite having larger dispersion, in average, produces better estimations. In regression, there are several estimators of discrepancy between the estimation and the true curve. At this respect, from several methods, the leave-one-out Cross Validation stands out. In this contribution we present ado.file-Python script combinations to calculate univariate kernel density estimators and kernel regression (Nayarada-Watson). Besides we include programs to calculate L2VC and BCV for univariate bandwidth choosing, and a program which employs combination of several penalizing functions and cross-validation algorithms to select bandwidth in kernel regression based on the algorithms from Scott (1992;2015), Härdle (1991), Salgado-Ugarte(1995; 2002) and Salgado-Ugarte & Saito-Quezada (2020).

* Presenter