BOOTSTRAP CONFIDENCE BANDS

Wolfgang HÄRDLE
C.O.R.E.
Voie du Roman Pays 34, B-1348 Louvain-la-Neuve

Michael Nussbaum
Karl-Weierstrass-Institut für Mathematik
Mohrenstr. 39
D-1086 Berlin

Abstract

Bootstrap confidence bands are constructed for nonparametric regression. Resampling is based on a suitably estimated residual distribution. The procedure is called the *Wild Bootstrap*. The method is to construct first a fine grid of error bars with simultaneous coverage probability. Second the end-points of these error bars are joined via polygon pieces or parabolae using assumptions on the local curvature of the regression curve.

1. Motivation

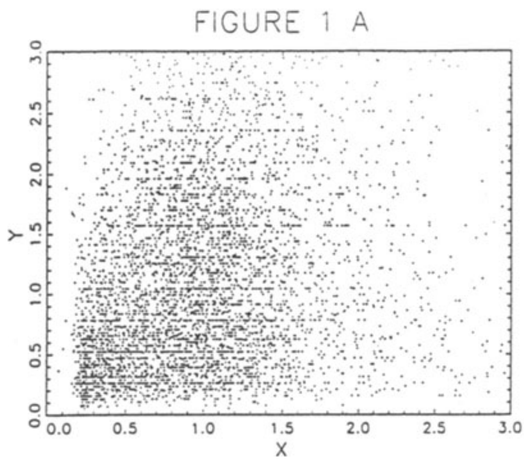
Nonparametric regression smoothing is a flexible method for estimation of mean curves. Since this technique makes no structural assumptions on the underlying curve, it is very important to have a device for understanding when observed features are significant. An often asked question in this context is whether or not an observed peak or valley is actually a feature of the underlying regression function or is only an artifact of the observational noise. For such issues confidence bands should be used.

This paper proposes and analyzes a method of obtaining confidence bands based on simultaneous error bars at a grid of points. The method is simple to implement and relies on local smoothness of the regression curve. The construction is based on a residual resampling technique which models the conditional error distribution and also takes the bias properly into account.

For an understanding of these ideas, consider Figure 1. Figure 1a shows a scatter plot of the expenditure for potatoes as a function of income for the year 1973, from the Family Expenditure Survey (1968-1983). Figure 1b shows a nonparametric regression estimate which was obtained by smoothing the point cloud, using the kernel algorithm described in Section 2. As a means of understanding the variability in the kernel smooth, Figure 1b also shows error bars, constructed by the Wild Bootstrap method proposed in Härdle and Marron (1990). These bars are estimated

K.-H. Jöckel et al. (eds.), *Bootstrapping and Related Techniques*

simultaneous 80 % confidence intervals. Note that the error bars are longer on the right hand side, which reflects the fact that there are fewer observations there, and hence more uncertainty in the curve estimate. The error bars are also asymmetric in particular at points with high curvature which reflects the correct centering of the bars by a bias term. We propose a method of joining these error bars in order to obtain a bootstrap confidence band.



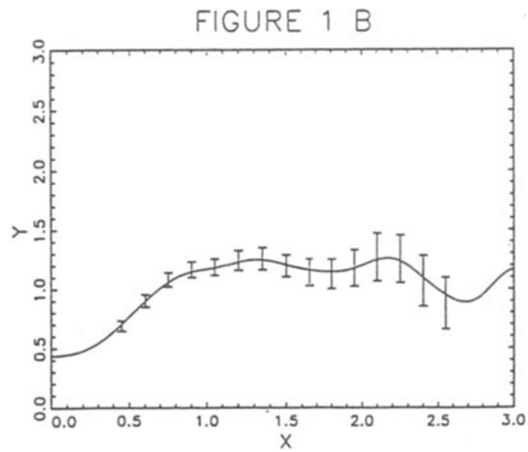


Figure 1 a,b. Expenditure for potato (Y) vs. income (X) (a) Scatter Plot (b) Regression kernel smooth (quartic kernel with band with h=0.3) and errors bars.

Clearly there is a need for confidence bands in all applications of nonparametric regression. Hall and Titterington (1986) constructed a confidence band for calibration of radio carbon dating assuming Normal errors. Knafl, Sacks and Ylvisaker (1984) derived uniform variability bands under the same error structure.

Our approach is based on resampling from estimated residuals. This form of bootstrapping preserves the error structure in the data and guarantees that the bootstrap observations have errors with mean zero. There are two main advantages to this approach. First it correctly accounts for the bias and hence does not require additional estimation of bias or the use of a sub-optimal (under smoothed) curve estimator. Second, no assumption of homoscedasticity is required, the method automatically adapts to different residual variances at different locations.

In Section 2 we give a technical introduction into simultaneous error bars constructed via the Wild Bootstrap. In Section 3 we consider bootstrap confidence bands. Proofs are given in the forthcoming paper by Härdle and Marron (1990).

2. Simultaneous error bars via the Wild Bootstrap

Stochastic design nonparametric regression is based on iid. observations $\{(X_i, Y_i)\}_{i=1}^n \in \mathbb{R}^{d+1}$ and the goal is to estimate $m(x) = E(Y|X = x) : \mathbb{R}^d \to \mathbb{R}$. The form of the kernel regression