# The Maximal Smoothing Principle in Density Estimation

GEORGE R. TERRELL*

We propose a widely applicable method for choosing the smoothing parameters for nonparametric density estimators. It has come to be realized in recent years (e.g., see Hall and Marron 1987; Scott and Terrell 1987) that cross-validation methods for finding reasonable smoothing parameters from raw data are of very limited practical value. Their sampling variability is simply too large. The alternative discussed here, the maximal smoothing principle, suggests that we consider using the most smoothing that is consistent with the estimated scale of our data. This greatly generalizes and exploits a phenomenon noted in Terrell and Scott (1985), that measures of scale tend to place upper bounds on the smoothing parameters that minimize asymptotic mean integrated squared error of density estimates such as histograms and frequency polygons. The method avoids the extreme sampling variability of cross-validation by using ordinary scale estimators such as the standard deviation and interquartile range, which have order $n^{-1}$ variability; cross-validated parameters have orders of variability such as $n^{-1/5}$. The disadvantage is that maximal smoothing parameters are conservative, rather than asymptotically optimal. Because they tend to lose information, they should be used in conjunction with other data displays that retain more of the features of the original sample. On the other hand, such conservative methods are widely valued by statisticians because they discourage naive overinterpretation of one's data. Maximal smoothing parameters are here derived for histograms and kernel methods, using not only the standard deviation but several more resistant methods of scale estimation. The method is then applied to density estimation on the half-line, on finite intervals, and in several variables.

KEY WORDS: Histograms; Kernel density estimates; Maximum entropy principle; Multivariate density estimates.

## 1. INTRODUCTION

Nonparametric density estimates attempt to reconstruct the probability density from which a random sample has come, using the sample values and as few assumptions as possible about the density. These methods are smoothing operations on the sample distribution. When we choose a density estimate, therefore, we must decide how much smoothing is appropriate; the larger the sample size the less smoothing required. A large part of the literature on density estimation is concerned with the issue of how to choose the degree of smoothness of the estimate. We propose as one possible guideline the *maximal smoothing principle:* Choose the largest degree of smoothing compatible with the estimated scale of the density.

This generalizes the method of Terrell and Scott (1985). The argument for the principle will proceed as follows: We will briefly discuss some earlier methods for choosing the degree of smoothing. The next section will develop

individualistic approach does not lead to replicable results; nor does it lead reliably to sensible estimates from a novice.

There is an extensive literature on the process of automatically choosing the best smoothing parameters; see Silverman (1986) for an extended discussion and bibliography. The most important include assuming a parametric family, which has the disadvantage of often being somewhat arbitrary, and cross-validation techniques such as least-squares cross-validation (Bowman 1984; Rudemo 1982) and biased cross-validation (Scott and Terrell 1987). Cross-validation techniques seem to be subject to enormous sampling variation (see Hall and Marron 1987; Scott and Terrell 1987).

To choose the best degree of smoothing, we must have a criterion for optimality of density estimates. Purely for reasons of tractability in the arguments that follow, our goal will be to make the expected $L^2$ metric $E((\int \hat{f}(y) - f(y))^2)$, or mean integrated squared error (MISE), as