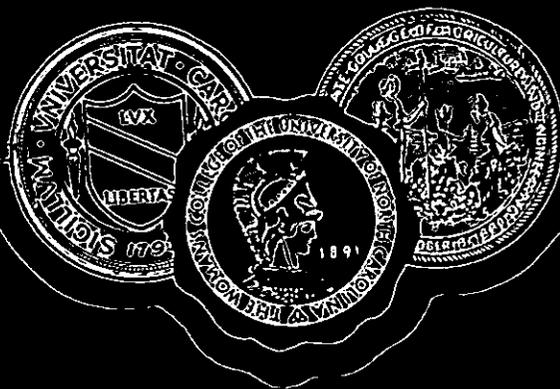# THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA

THE NONEXISTENCE OF MOMENTS OF SOME KERNEL REGRESSION ESTIMATORS

by

Wolfgang Härdle
Universität Heidelberg, Sonderforschungsbereich 123

and

James Stephen Marron
University of North Carolina, Chapel Hill

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

THE NONEXISTENCE OF MOMENTS OF SOME KERNEL REGRESSION ESTIMATORS

by

Wolfgang Härdle[1]

Universität Heidelberg, Sonderforschungsbereich 123

and

James Stephen Marron[2]

University of North Carolina, Chapel Hill

ABSTRACT

In the setting of nonparametric density estimation, it is seen that the moments of kernel based estimators (with high order kernels) may not exist. Thus the popular error criterion of mean square error may be useless in this setting.

KEY WORDS AND PHRASES: Nonparametric regression, kernel estimation, nonexistence of moments.

Please send all correspondence to:

James Stephen Marron
Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27514

The (stochastic design) regression estimation problem may be defined as follows. Let $f_{X,Y}(x,y)$ be a joint probability density, and let $f_X(x)$ denote the marginal density of X. Define the regression function,

$$r(x) = E[Y|X=x] = \int y \, f_{X,Y}(x,y) f_X(x)^{-1} dx \; .$$

The object is to estimate the function $r(x)$ using a iid sample,

$$(X_1,Y_1),\ldots,(X_n,Y_n), \text{ from } f_{X,Y}(x,y).$$

Nadaraya (1964) and Watson (1964) have proposed "kernel estimators" of $r(x)$. These are defined as follows. Given a "kernel function", $K(x)$, and a "band-width", h, let

$$\hat{r}(x) = \frac{\sum\limits_{i=1}^{n} K(\frac{x-X_i}{h}) Y_i}{\sum\limits_{i=1}^{n} K(\frac{x-X_i}{h})} \; .$$

A discussion of this estimator and some related estimators may be found in the survey by Collomb (1981).

The most common means of assessing the accuracy of statistical regression estimators is the Mean Square Error, given by

$$MSE = E[\hat{r}(x)-r(x)]^2 \; .$$

This paper demonstrates that, under commonly occurring circumstances, MSE is a poor error criterion, because the moments of $\hat{r}(x)$, and hence MSE may fail to exist.

To help put these results in proper perspective , using the notation

$$Z_i = K(\frac{x-X_i}{h}) \; ,$$

note that

$$E[\hat{r}(x)|X_1,\ldots,X_n] = \frac{\sum_{i=1}^{n} Z_i r(X_i)}{\sum_{i=1}^{n} Z_i} \ . \tag{1}$$

Thus, if the Z_i are nonnegative, $r(X_i)$ is bounded, and 0/0 is appropriately defined, then $E\hat{r}(x)$ is easily seen to exist. However it is well known, see for example Parzen (1962) or Gasser and Müller (1979) that the rate of convergence for kernel estimators in either regression or density estimation can be greatly improved by allowing K to take on negative values. The rest of this paper is devoted to showing how this practice can easily cause nonexistence of $E\hat{r}(x)$.

Problems arise when the denominator in (1) is very close to 0 but the numerator is not. For example, consider the case $n = 2$. Routine computations show that, if Z_1 and Z_2 are absolutely continuous with respect to Lebesgue measure, if there is a point z_0 such that the densities of Z_1 and Z_2 are bounded above 0 on neighborhoods of both z_0 and $-z_0$, and if $r(x)-r(-x)$ is nonzero on some neighborhood of z_0, then $E\hat{r}(x)$ fails to exist.

Similar examples may be easily constructed where Z_1 and Z_2 are not absolutely continuous (for example when K is compactly supported), but have an absolutely continuous component which satisfies the above conditions. In the case where Z_1 and Z_2 are discrete (corresponding to K a step function) counter-examples of the above type arise much less naturally, since it is required that for some $c > 0$, K takes on both the values c and $-c$.

For $n > 2$, analogous (but more complicated) examples can be constructed. It should be noted that for reasonable choices of K (ie: more "positive" than "negative") the probability that the denominator of (1) is close to 0 will decrease as n increases. Thus the difficulties discussed in this paper tend to disappear in the limit. However, for each n, MSE may still be undefined and so will not be a reasonable error criterion.

In the case of K a step function (considered by Serfling (1980)), note that for $n > 2$, counterexamples arise much more easily. Indeed, if K takes on the values $c_1, \ldots, c_k$, then little more is required than $\sum_{j=1}^{k} n_j c_j = 0$ where $n_1, \ldots, n_k$ are nonnegative integers whose sum is n.

Having seen that MSE can be a treacherous error criterion, one might look for substitutes. A first choice would probably be some sort of truncated MSE. Other approaches may be found in Härdle and Marron (1983) and Marron and Härdle (1983).

REFERENCES

COLLOMB, G. (1981), Estimation nonparamétrique de la régression: revue bibliographique, <u>Int. Statist. Rev.</u> 49, 75-93.

GASSER, T. and MÜLLER,H.-G. (1979), Kernel estimation of regression functions, in: T. Gasser and M. Rosenblatt, eds. <u>Smoothing Techniques for Curve Estimation</u> (Lecture Notes in Mathematics 757, Springer Verlag) pp. 23-68.

HÄRDLE, W. and MARRON, J.S. (1983), Optimal bandwidth selection in nonparametric regression function estimation. North Carolina Institute of Statistics Mimeo Series #1530.

MARRON, J.S. and HÄRDLE, W. (1983), Random approximations to an error criterion of nonparametric statistics. North Carolina Institute of Statistics Mimeo Series # 1538.

NADARAYA, E.A. (1964), On estimating regression, <u>Theory Prob. Applic.</u> 9, 141-142.

PARZEN, E. (1962), On estimation of a probability density and mode, <u>Ann. Math Statist.</u> 35, 1065-1076.

SERFLING, R.J. (1980), Properties and applications of metrics of nonparametric density estimators, in: <u>Coll. Math. Soc. János Bolyai,</u> 32 Nonparametric Statistical Inference, Budapest.

WATSON, G.S. (1964), Smooth regression analysis, <u>Sankhyā</u> A, 26 359-372.