

Discussion of the Paper by Davies and Kovac

Rudolf Beran

University of California, Berkeley

“Today’s understanding of how well elementary techniques work owes much to both mathematical analysis and experimentation by computer.”—John Tukey (1977).

“The history of the arts and sciences could be written in terms of the continuing process by which new technologies create new environments for old technologies.”—H. Marshall McLuhan (1964).

Man makes tools whose use then reshapes his life. Technological advances in statistical computing and in empirical process theory have swept away statistics as a normative mathematical philosophy. Our subject is under reconstruction on a more scientific foundation of computational experiments linked to mathematical probes of statistical methods. Environmental stimuli, notably competition from data-analytic techniques that fall outside the statistical canon, favor the evolution of statistics towards empirically tested, falsifiable theory.

In presenting their innovative nonparametric regression estimators, Davies and Kovac carefully distinguish among data, statistical procedure, and probability model. Computational experiments have brought this fundamental distinction to the forefront of statistical thought. Tukey’s (1977) *Exploratory Data Analysis* made the point dramatically by not using probability models at all in the exposition.

1. Experimental and Theoretical Statistics. Davies and Kovac remark that “statistical procedures can be evaluated using real data sets and under the well controlled conditions of a stochastic model or test bed.” Their figures illustrate another possibility, trying a procedure on artificial data. We delve a little further into the matter.

Until recent decades, the primary tools available to a statistician were mathematics and logic. A powerful technology, measure-theoretic probability theory, directed statistics toward probability models for data, toward discussions of abstract principle, and toward test statistics or pivots whose distributions could be tabulated because they did not depend on nuisance parameters. Simple probability models can motivate promising classes of statistical procedures. The rules (1.4) and (1.6) that Davies and Kovac propose as part of their nonparametric regression technology are in this tradition. More complex probability models can be used in studying mathematically the performance of statistical procedures. The asymptotic theorems of Section 3 reflect this second tradition.

Advances in computing and graphical output have brought about rapid development of experimental statistics. One type of computational experiment elicits finer details of a procedure’s performance in repeated pseudo-random realizations from probability models. These details may supplement results from asymptotic theory, such as rates of convergence and asymptotic minimax bounds, whose implications for statistical practice are less than clear. Of course, pseudo-random realizations constitute a mathematical model for data that deserves study in its own right. Such artificial data mimics only some features of the

motivating probability model and has properties not envisaged by that probability model. This point tends to be minimized in the statistical literature.

Another type of experiment consciously investigates performance of procedures outside probability models. Case studies on actual data or on idealizations of actual data are important. So are comparative analyses with perturbations of real or artificial data. One might change signal-to-noise ratio or sample size in the Blocks example in Fig. 11. One might add a high frequency sine wave to that example, increasing the frequency of the sinusoid in order to discover at what point the regression estimators in this paper first lose sight of its regularity and at what point they fail completely to detect it. Such perturbation experiments provide interpretable ways of comparing competing nonparametric regression estimators. Though the human eye often fails to spot high-frequency patterns amidst noise, good regression techniques do better.

Assessing procedures under a probability or other mathematical model is a bold speculation that relies on two hopes: the ability of the model to mimic observational data and the ability of the mathematical analysis to address questions of data-analytic interest. Greater interplay between computational experiments and mathematical probes of procedures has replaced speculation about data analysis with the rudiments of scientific method. Nevertheless, reproducibility of published computational experiments still has low priority in core statistical journals. In an unpublished technical report, Buckheit and Donoho (1995) discussed what is required to make computational portions of statistical research reproducible.

2. The One-Way Layout in Nonparametric Regression. A regression model that motivates parts of the paper is

$$(1) \quad y(t_i) = f(t_i) + \epsilon(t_i), \quad 1 \leq i \leq n,$$

where the values $0 < t_1 < t_2 < \dots < t_n < 1$ are strictly ordered. Estimation of the function f on the basis of the observed $\{(y_i, t_i)\}$ is the task undertaken. In theoretical study of function estimators f_n under probability models, it is customary to compare f_n with f through quantities such as the supremum norm and to impose conditions on f such as differentiability (cf. Section 3 of the paper).

Nonparametric regression combines two distinguishable problems, each of which may be studied constructively on its own. The first problem is estimation of the values $\{f(t_i): 1 \leq i \leq n\}$. This amounts to estimation of the values $\{\mu_i\}$ in the one-way layout

$$(2) \quad y_i = \mu_i + \epsilon_i, \quad 1 \leq i \leq n,$$

where $y_i = y(t_i)$, $\mu_i = f(t_i)$, and $\epsilon_i = \epsilon(t_i)$. That the least squares estimator $y = \{y_i\}$ of $\mu = \{\mu_i\}$ is not a good answer to the problem was pointed out by Stein (1956). This naive estimator can have relatively high risk under a probability model.

Once we have devised a more efficient estimator of μ , the second problem is interpolation among its components so as to estimate the function f . This is essentially a problem in approximation theory and is considerably more sensitive to assumptions on the nature of f than is the estimation of the $\{f(t_i)\}$. Because the data will not tell us how many derivatives f has, we might settle, in the absence of strong prior information, for linear or

spline interpolation among the estimated components of μ . Instead, the Davies and Kovac estimators of f simultaneously determine the estimator of μ and the interpolation scheme by minimizing the number of local extrema in f_n , subject to achieving residuals at the $\{t_i\}$ that behave like white noise. Their idea is refreshingly novel. Comparing the performance of their estimators with linearly interpolated thresholding competitors when f is very wiggly seems to be a natural question.

To consider separately the estimation and interpolation aspects of nonparametric regression clarifies what we can achieve in each respect. In particular, handling the practically important case where the $\{t_i\}$ are not all distinct can begin with treating an unbalanced one-way layout.

3. Loss Estimates as Diagnostic Tool. A statistical practitioner needs credible indications of a procedure's success or failure in analyzing the data at hand. Ensemble results such as minimaxity of a procedure or asymptotic rate of convergence under a probability model do not diagnose adequacy of a procedure applied to specific data, though they may suggest instructive experiments with competing procedures in worst-case scenarios. The residency system for practical training of physicians arose in the second half of the nineteenth century, replacing a system whereby the training of medical practitioners was mostly theoretical. Surgeons and army doctors who worked with their hands long had lower social status, even as they pioneered important medical procedures. Improved computing environments now encourage developments in statistical diagnostics and statistical training that may parallel, at an abstract level, the evolution of modern scientific medicine.

Feedback about which nonparametric regression procedure to use in a particular data analysis can come from estimated performance summaries, from diagnostic plots, and from the substantive field in which the observations were obtained. A broadband diagnostic approach is surely more effective than any single tool. For example, estimated losses sharpen visual scrutiny of competing regression estimators and their residuals by prompting one to *see* why one estimated loss is larger than another. Estimated losses are especially helpful when the domain of the regression function is not one or two dimensional.

We discuss here technical aspects of estimating the quadratic loss

$$(3) \quad L_n(\hat{\mu}, \mu) = n^{-1}|\hat{\mu} - \mu|^2$$

for the estimator $\hat{\mu} = \hat{\mu}(y)$. Let $g(y) = \hat{\mu}(y) - y$. If the errors $\{\epsilon_i\}$ in the one-way layout are i.i.d. $N(0, \sigma^2)$ and g satisfies assumptions detailed in Stein (1981), then the risk of $\hat{\mu}$ is

$$(4) \quad R_n(\hat{\mu}, \mu, \sigma^2) = \sigma^2 + E[2\sigma^2 n^{-1} \sum_{i=1}^n \partial g_i(y)/\partial y_i + n^{-1}|g(y)|^2].$$

Let U be an orthogonal matrix whose columns form a basis for R^n and are increasingly wiggly as column index increases. For example, U might be the orthogonal polynomial basis or the discrete cosine basis for R^n . Let $z = U'y$. Some approaches to estimating σ^2 are typified by the formulae

$$(5) \quad \hat{\sigma}_D^2 = [2(n-1)]^{-1} \sum_{i=2}^n (y_i - y_{i-1})^2 \quad \text{and} \quad \hat{\sigma}_H^2 = (n-q)^{-1} \sum_{i=q+1}^n z_i^2$$

and robustifications like (1.7) in the paper. Consistency theorems for these estimators of σ^2 and their variants suggest diagnostics that guide their use.

Let

$$(6) \quad \hat{L}_n = \hat{\sigma}^2 + 2\hat{\sigma}^2 n^{-1} \sum_{i=1}^n \partial g_i(y) / \partial y_i + n^{-1} |g(y)|^2.$$

For consistent $\hat{\sigma}^2$ and for certain classes of estimators $\hat{\mu}$, the loss $L_n(\hat{\mu}, \mu)$ and the risk $R_n(\hat{\mu}, \mu, \sigma)$ converge together as n tends to infinity; and \hat{L}_n also converges, in an L_1 -norm sense, to the common asymptotic value of loss and risk. Further details are given in Beran (2000) and references cited there.

The proposal being made is to consider \hat{L}_n as a diagnostic tool for assessing nonparametric regression estimators. When $\hat{\mu}(y)$ lacks a tractable closed form, the partial derivatives needed in (6) may be approximated numerically. Let u_i denote the vector in R^n whose i -th component is 1 and whose other components are 0. Then, for small real values of δ ,

$$(7) \quad \partial g_i(y) / \partial y_i \approx \delta^{-1} [g_i(y + \delta u_i) - g_i(y)], \quad 1 \leq i \leq n.$$

Computing these difference quotients requires computing $\hat{\mu}(y) = y + g(y)$ and the n perturbed estimators $\{\hat{\mu}(y + \delta u_i) : 1 \leq i \leq n\}$.

Like other statistical procedures, diagnostic tools need experimental testing. In trials on pseudo-random artificial data, we may compare the actual loss of a regression estimator with loss estimates such as \hat{L}_n . In the author's experiments, approximate evaluations of \hat{L}_n on pseudo-Gaussian data have yielded respectable estimates of loss for the James-Stein estimator of μ and for more efficient linear shrinkage and soft-thresholding estimators. The findings suggest trial of \hat{L}_n as a performance diagnostic for the Davies and Kovac regression estimators and further theoretical investigation of loss estimators outside Gaussian models.

REFERENCES

- BERAN, R. (2000). REACT scatterplot smoothers: superefficiency through basis economy. *J. Amer. Statist. Assoc.* **95** 155–171.
- BUCKHEIT, J. B. and DONOHO, D. L. (1995). WaveLab and reproducible research. Technical Report 474, Dept. of Statistics, Stanford University, <www-stat.stanford.edu/~donoho/>.
- McCLUHAN, E. and ZINGRONE, F. (1995). *Essential McLuhan*. House of Anansi Press, Concord ON, Canada.
- STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Third Berkeley Symp. Math. Statist. Probab.* (J. Neyman, ed.) **I** 197–208. University of California Press.
- STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.
- TUKEY, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading MA.