

THE IMPACT OF I.J. GOOD ON DENSITY ESTIMATION*

David Scott
James Thompson

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 2.0[†]

Abstract

One of three appreciations of I.J. Good's work published in *The Good Book: Thirty Years of Comments, Conjectures and Conclusions* by I.J. Good. The book is available in print form from Rice University Press (<http://ricepress.rice.edu>).

(This module helps introduce **The Good Book: Thirty Years of Comments, Conjectures and Conclusions, by I.J. Good**. The book is available for purchase from the Rice University Press Store¹. You can also visit the Rice University Press web site².)

1 Density Estimation

It is a pleasure to review Jack Good's numerous contributions to the theory and practice of modern statistics. Here, we wish to remember his innovations in the field of nonparametric density estimation. Together with his student R. A. Gaskins, Jack invented penalized likelihood density estimation (Good and Gaskins, 1971). Given the computing resources available at that time, the implementation was truly revolutionary. A Fourier series approximation was introduced, not with just a few terms, but oftentimes thousands of terms. To address the issue of nonnegativity, the authors solved for the square root of the density. The penalty functions described were L_2 norms of the first and second derivatives of the density's square root.

The first author had the pleasure of attending a lecture by Jack at one of the early Southern Research Conference on Statistics meetings and returned to Rice University with a number of questions. For example, is the square root "trick" valid? Could a closed-form solution be found? Considering such questions led to collaborations with numerical analyst Richard Tapia and theses by Gilbert de Montricher and the second author. Gilbert was able to show that the first derivative penalty could be solved in closed form. (Klonias [1982] later provided a wider set of solutions.) But Gilbert also showed that the square root trick does not work in general in infinite-dimensional Hilbert spaces, such as those considered here. Scott (1976) examined a finite-dimensional approximation for which the square-root trick does apply. These research findings were collected in Tapia and Thompson (1978), one of the first surveys of nonparametric density estimation. In this and other venues, Jack's pioneering work led to a large body of research based on splines and other bases.

*Version 1.1: Sep 12, 2008 12:54 pm -0500

[†]<http://creativecommons.org/licenses/by/2.0/>

¹<http://my.qoop.com/store/3111075350609104/386563560345>

²<http://ricepress.rice.edu>

2 Nonparametric Density Research at Rice

Jack's inspiration came at a very fortuitous time for statisticians at Rice. NASA funding had switched from an emphasis on space exploration to that of agricultural intelligence gathering via remote sensing. (Thompson well remembers Jack in his IDA days walking around Princeton in a trenchcoat, affecting the pose of George Smiley. So Jack might appreciate what follows below.) The idea was to identify and exploit shortages in Soviet grain production.

The NASA prototype solution in 1970 used a huge and clunky multi-spectral scanner that recorded ground reflectivity in twelve channels. This involved flyovers in Kansas from large aircraft. Misclassification rates were running around 25% using the assumption the data were multivariate Gaussian. The solution (before we got into the problem) was to expand the hardware to an even larger twenty-four-channel device. NASA had not run into the heavy-tailed pathologies dealt with by the Princeton Robustness Project, but rather into the mixture of distributions problem which the Princetonians did not address. Of course, for the mixture problem under the Gaussian assumption, things get worse as the number of channels increases. Thompson was somewhat amazed to find during a drive around in the summer of 1971 that the LARYS group at Purdue and the Willow Run group at Michigan were also treating the data as though they were Gaussian.

At Rice we immediately discovered that we could gain dramatic improvements (misclassification rates of 5%) by dropping the number of channels to four and using template methods. We quickly moved to kernel approaches. Some years later, the NASA group, led by Richard Heydorn, took the solution for a four-channel scanner and easily packaged it into a satellite. They had arrived at a point where NASA forecasting of Soviet grain production was much better than that of the Soviet Ministry of Agriculture. The technology was exploited, according to the terms of the Jackson-Vanik Amendment, to trade the permission of the Soviets in their unusually bad years (there were no good years) to buy US grain at spot prices in exchange for very relaxed policies on the emigration of Soviet Jews to Israel. Thus, nonparametric density estimation has had indirect but dramatic effects on the demographics of Israel. Jackson-Vanik has permitted hundreds of thousands of Soviet Jews to resettle in Israel. Without the NASA results, Jackson-Vanik would have had no teeth.

We expanded our research to work on Defense modeling, ecology, and biomedical work in cardiology and oncology. Over the years, Rice has produced over a hundred papers, three books, and dozens of doctorates in nonparametric function estimation topics in a variety of application areas.

3 Bump Hunting

One of the most important applications of density estimation is the discovery and characterization of features. In Good and Gaskins (1980), the authors set forth an extremely influential application of their penalized likelihood estimator for assigning odds for the veracity of modes and bumps. Silverman (1981) introduced a bootstrap technique that examines the number of modes in a density all at once. However, Good and Gaskins examined individual modes (and more generally, bumps) one-at-a-time, which we believe is the more powerful approach. One would like to find the "closest" density without the mode or bump of interest. What the authors introduced was "bump surgery," where the raw data were gently massaged to reduce the size of the mode or bump, until it was just eliminated. Then a quantity rather like log-likelihood could be computed to quantify the odds on whether the feature is real, or just an artifact of the sample. This problem is very challenging, and many lines of research have ensued. But the paper was read at the Joint Statistical Meetings, accompanied by a lively set of discussions, and has been enormously influential in the field.

4 Appreciation

The second author had the pleasure of serving as Jack Good's chauffeur twice. At another SRCOS meeting in Arkansas, Jack flew into a nearby airport. The SRCOS meetings were a wonderful week-long affair that made research discussions informal and exciting. In 1993, Jack was invited to give a talk at the National Security Agency, where the the second author was spending a sabbatical. Jack agreed to visit if he did not

have to drive. A large collection of Jack's classified publications are available in the NSA library, typewritten with 1940s technology. Many of Jack's friends had long retired from NSA, but the excitement of problem-solving made for a memorable day.

Thus it is our great pleasure to share in the celebration of Jack's ninetieth birthday and to admire the depth and breadth of his work. Many happy returns.

5 References

Good, I. J. and Gaskins, R. A. (1971), "Nonparametric roughness penalties for probability densities," *Biometrika* , 58, 255-277.

Good, I. J. and Gaskins, R. A. (1980), "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data," *J. Amer Stat Assoc* , 75, 42-56.

Klonias, V. K. (1982), "Consistency of two nonparametric maximum penalized likelihood estimators of the probability density function," *Annals of Statistics* , 10, 811-824.

Scott, D.W. (1976), "Nonparametric Probability Density Estimation by Optimization Theoretic Techniques," unpublished doctoral dissertation, Rice University, Houston.

Silverman, B. W. (1981), "Using kernel density estimates to investigate multimodality," *J. Royal Statistical Society, Series B*, 43, 97-99.

Tapia, R. A. and Thompson, J.R. (1978), *Nonparametric probability density estimation* , Johns Hopkins University Press, Baltimore.