Statistica Sinica **19** (2009), 1361-1365

Editorials

On the Marriage of Multiscale Analysis and Statistics

Not so long ago, I had the opportunity to attend a Public Lecture sponsored by the Clay Mathematics Insitute. Ingrid Daubechies spoke to a general audience about Wavelets. I wasn't the only 'professional statistician' in the audience; Harvard Statistics Chair Xiao-Li Meng sat nearby.

At this lecture, Professor Daubechies presented an overview of the many ways in which wavelets and multiscale ideas are finding applications in science and technology. As an historical aside, she took us back to the previous century − her famous paper on orthonormal wavelet bases is now more than 20 years old − and used last century's presentation tool (the overhead projector) to present a transparency she has been holding onto for some over 20 years. The transparency shows her first picture of what we today call the Daubechies D4 wavelet.

Daubechies' talk set me thinking about the ways the world has changed since the days of D4. We no longer use transparencies, and today we use fancier wavelets, without the visible 'fractal' texture that makes D4 so distinctive. Bell Labs, where Daubechies did her work, is no longer the technologically dominant force it once was, when it produced inventions, software, and ideas (transistor, unix, 'bit') that reshaped our life. But some new developments occurred, and some are welcome. The fact that Professor Meng and I were in attendance at Daubechies' lecture, along with other statisticians (we all arrived in an uncoordinated 'random' way), is the one I want to focus on here: a significant number of statisticians today are interested in and conversant with multiscale analysis, as this special issue demonstrates.

Twenty years ago, it would have been extremely unusual for statisticians to interact with harmonic analysts − I can think of only a few people who did so regularly. Moreover, it would have been extremely unusual for either species of mathematical scientist (i.e., mathematical analyst or mathematical statistician) to spend much time thinking about how to help engineers who gather and analyse signals. Today, there are many conferences of mathematical scientists in which signal-processing inspired research is presented and discussed. In fact, one such meeting, the Graybill Conference in June 2006, provided the spark for the present issue of *Statistica Sinica.*

This special issue points to a cross-fertilization of statistics by applied analysis. This interaction is young and we really don't know what effects it will have. I expect that the researchers whose work is presented here will, over time, produce global, socially-discernible benefits that we will all be proud of one day.

There's a high standard to live up to. But remember that an earlier generation of applied mathematicians focused on fluid dynamics-related issues, and an earlier generation of statisticians focused on agricultural research. The efforts of earlier generations really paid off: in the last 100 years airplanes have advanced tremendously far. Applied mathematicians can proudly point to high quality numerical airflow simulations validating better wing and engine designs. Agriculture has made similar dramatic advances. Statisticians can proudly point to efficient experimental designs and statistical inference tools leading to discovery of high-yield crops. These are major societal impacts.

There's really no telling what important applications will emerge from the marriage of applied analysis with statistics. I have worked for years combining the two in my own work, and have seen how applications came about. Things progress in unpredictable ways; the spread of ideas is often difficult to track. It seems that unexpected combinations of ideas have had the biggest impact. My own work features such an unexpected combination between ideas in applied analysis and ideas from statistics. I will first discuss the two sets of ideas or themes I am thinking about and then their combination.

Theme 1. Wavelet Denoising. When I first learned about orthonormal bases of wavelets in the late 1980s, I dove into harmonic analysis headfirst; I remember being utterly fascinated by all the great work in harmonic analysis in the '60s, '70s, and '80s, for example, papers of Jaak Peetre on the Besov-Triebel spaces and on ℓ_p spaces for $p < 1$, work of Raphy Coifman on atomic decomposition of H^1, and work of Ron Devore on approximations spaces A_p, $p < 1$. While such topics were not discussed in applied mathematics at the time, those abstract ideas inspired thoughts of applications. The key point was to recognize that, where these spaces offered something different from traditional spaces like L^2 Sobolev spaces, much of the math analysis was about showing that wavelet coefficients of objects in certain cases were somehow sparse $-$ had relatively few big coefficients.

I soon decided that there was something here for statisticians. At the time the literature of 'smoothing' considered estimation of functions in 'nice spaces' like L^2 Sobolev spaces. It was known that many different orthogonal systems could do an excellent job in such estimation problems. It became clear that wavelets didn't really make essential contributions in those settings; where they could really contribute would be apparent if more 'exotic' spaces were assumed. Iain Johnstone, Dominique Picard, Gerard Kerkyacharian and I worked out some implications of this view for removing noise from signals, estimating densities and spectra; see for example our Royal Statistical Society (RSS) read paper. This work was addressed explicitly to a statistical audience, proposing to perform statistical estimation of densities, signals, and images in Besov and Triebel classes. At the time, such assumptions in statistics papers were unusual; such spaces were known really only in analysis. Such assumptions have become more conventional. And much literature has sprung up using them.

However, although the formal assumptions (Besov and Triebel classes) are crucial to the scholarly angle, working out the consequence of the formal assumptions is not the real reason, in my view, for the growth of publication in this area. It is the larger insights and the algorithmic deliverables that have mattered more than the accuracy of the formal assumptions.

The larger insight: wavelets are useful because they sparsify objects being studied − taking completely dense signals and transforming them invertibly into sparse vectors with relatively few significant entries.

The algorithmic deliverable: many of these papers propose simple thresholding of noisy coefficient sequences. This is computationally practical and seems to be psychologically natural for many people working in signal processing, where numerous applications of wavelet thresholding have been published.

Theme 2. Recovery of Sparse Signals. In the early 1980s, I became fascinated by some empirical phenomena known in geophysical signal processing: one could take fewer than N measurements of an N element signal vector, and yet still exactly reconstruct the signal, provided the signal was sparse, the measurements were specially chosen linear combinations, and the reconstruction was by a specific kind of linear programming. I was inspired by a simple model: a sequence is sparse and its Fourier transform is undersampled.

I decided this was in some way connected with work on non-harmonic Fourier analysis and to analytic methods in number theory, and I dove into that literature. I immersed myself in the collected works of Beurling, Wiener, Littlewood and Polya. I eventually wrote papers in the late 1980s about this model with Phil Stark (of UC Berkeley) and Ben Logan (of Bell Labs). Although the model was not primarily statistical, my coauthors Iain Johnstone and NMR spectroscopists Jeff Hoch and Alan Stern (Rowland Institute) and I presented it to a statistical audience in an RSS read paper at about that time. Today there are many papers being published which consider extensions or outgrowths of models like this; the most well-known is the paper by Candes, Romberg and Tao (2006).

In those days, looking at sparse and/or random sets of Fourier coefficients of sparse and/or random objects was unusual, such situations were definitely a narrow specialty even in analysis, but they have since become more common. Sparsity of the object to be recovered now seems more natural, as spectroscopy and other technology have proliferated. Undersampling of the Fourier transform has also been discussed in spectroscopy at least since Jeff Hoch and Alan Stern's explicit proposal in 1994.

This literature has, in my view, not proliferated because of the strict applicability of the formal assumptions, but instead because of the larger insight and the algorithmic deliverable. The larger insight is that sparsity is a very valuable piece of side information, and can be exploited to allow undersampling. The algorithmic deliverable is the use of ℓ_1-penalized methods (later called Lasso) that allow reconstruction of sparse sequences from undersampled data. The computations seemed extravagant in the 1980s, where simple filtering ideas dominated signal processing. But now, heavy computations seem increasingly practical and natural in the signal processing community.

Cross-Fertilization. Our two themes each developed extensive literatures and many published applications. But the most memorable impact may not be the originally intended ones. Consider a combination of these themes, sometimes called Compressed Sensing. We represent an image or signal in a wavelet basis, where it is sparse; then we undersample some transform of the signal − for example, undersample the Fourier transform of the object − and reconstruct the wavelet coefficients an iterative application of wavelet thresholding. This approach, which is really a combination of the two themes just mentioned, is today stimulating lots of application work.

I recently attended a conference hosted by the International Society for Magnetic Resonance in Medicine at which novel schemes for Magnetic Resonance Imaging were presented and discussed. I witnessed a number of impressive demonstrations of speed-ups in MRI − factors of eight, for example; or even applications previously thought impossible, made possible by such speedups – 3D-spectroscopic imaging and dynamic cardiac imaging. The driving force allowing such speedups is very definitely the marriage of Themes 1 and 2 mentioned above, but these may be somewhat hidden in the actual implementation. For example, wavelets may not be explicitly in evidence; denoising of sparse sequences may also not be explicitly in evidence; but one can definitely show that such ideas, correctly combined, are the driving force in the applications being presented.

It is extremely hard to trace the complex chain of ideas and inspirations that lead from ideas in 'pure' mathematical statistics or harmonic analysis all the way to impacts in technological fields like signal processing and medical imaging. Ideas get into the intellectual atmosphere, jump disciplines in untraceable ways and reshape discussion far away from their original source.

In reading the papers in this special issue, please reflect on this. While multiscale analysis in statistics is an established area of its own, and is here accorded a special issue of *Statistica Sinica*, the most important or broadest impacts of the work you will read between these covers is likely to come in unexpected ways.

———— David Donoho

David Donoho has been Professor of Statistics at Standford University and, earlier, at the University of California, Berkeley. He received his AB in Statistics from Princeton (summa cum laude) where his undergraduate thesis Adviser was John Tukey and his PhD in Statistics from Harvard, where his PhD adviser was Peter Huber. He has also worked for Western Geophysical Company (Seismic Exploration) and Renaissance Technologies (Financial). He was co-founder of network management software company BigFix. He has published in Robust Statistics, Signal and Image Processing, Mathematical Statistics, Harmonic Analysis Scientific Computing, and High Dimensional Geometry. He has made fundamental contributions to theoretical and computational statistics, as well as to signal processing and harmonic analysis. His algorithms have contributed significantly to the understanding of the maximum entropy principle, of the structure of robust procedures, and of sparse data description. He is a member of the National Academy of Science and the American Academy of Arts and Sciences, is a MacArthur Fellow, and is a recipient of the COPSS Presidents' Award.