David Donoho: Sparse Data, Beautiful Mine >>>

David Donoho

Excerpts of interview of David Donoho by Y.K. Leong

(Full interview at website: http://www.ims.nus.edu.sg/imprints/interview_donoho.htm)

David Donoho is world-renowned for many important contributions to statistics and its applications to image and signal processing, in particular to the retrieval of essential information from "sparse" data. He is reputed to be the most highly-cited mathematician for work done in the last decade (1994–2004) — a reflection of the impact of his work on engineering and the physical and medical sciences.

He has received numerous honors and awards, notably the Presidential Young Investigator Award and the Presidents' Award (of the Committee of Presidents of Statistical Societies). He is a member of the National Academy of Sciences, USA, and the American Academy of Arts and Sciences. He has been invited to give prestigious lectures of scientific bodies, such as the Wald Lecture and the Bernoulli Lecture, and at the International Congress of Mathematicians. He has served on the committees of professional scientific bodies and on the editorial boards of leading journals on probability, statistics and mathematics.

The Editor of *Imprints* interviewed him at the Department of Mathematics on 26 August 2004 when he was a guest of the Department of Mathematics and the Department of Statistics and Applied Probability from 11 August to 5 September 2004 and an invited speaker at the Institute's program on image and signal processing. The following is an enhanced and vetted account of excerpts of the interview. It reveals little-known facets of his early scientific apprenticeship in the primeval and almost unreal world of computer programming and data analysis of the seventies. He talks passionately about the trilogy of attraction and fascination with computing, statistics and mathematics and about the many statistical challenges and opportunities arising from the exponential growth in data collected in all branches of human knowledge.

(**Acknowledgement.** *Imprints* acknowledges the efforts of Ms Tombropoulos of Stanford University in the preparation of the final version of the interview.)

Imprints: How did you come to be interested in probability and statistics?

David Donoho: When I left for college, my father suggested I get a part-time job where I'd learn computer programming. The employment office at Princeton sent me to the Statistics Department. I went to work for a professor (Larry Mayer) doing statistical data analysis of household energy use; this taught me to use statistical computer programs. Through this I gradually became very interested in computers and also in data analysis. At the same time, I was taking mathematics courses. I saw that a career in statistics would let me do mathematics and use computers to analyze data. By the end of my first year I was hooked; I remember that I started pulling all-nighters hanging out at the computer center already by spring term.

I: Why would you need to stay up all night?

D: In those days, computing was much more difficult than today. It was a major effort to translate your program into a physical form (punch cards) acceptable to the computer. It was a long wait for the computer to process your work, and then often it would just spit out something like "IEHK6040 Job Control Language Operand Agreement Error". You had to be very persistent to get things done. Sometimes it would just take all night.

Continued on page 12

Continued from page 11

I: What was the attraction?

D: Today it's all the rage for young people to do "Extreme Sports"; triathlon, bungee, and so on. The point is the sheer exhilaration of taking on a daunting challenge and prevailing. Computing in those days was a kind of extreme intellectual sport. To get over all the physical and intellectual hurdles was really an achievement. By comparison, computing today is like jogging, or maybe just a brisk walk around the block.

I: What came next?

D: Computing improved very rapidly. The next year, the Statistics Department got a PDP-11 minicomputer — one of the first Unix computers outside of Bell Labs. Don McNeil and Peter Bloomfield gave me a job: to build up all the statistical software that was needed. I had to program linear regression and graphics, and I also had to provide graphical displays from exploratory data analysis. I had to learn C (a totally new language in those days) and even assembly language programming. I had to learn all the basic algorithms for statistics and numerical analysis, and implement and test things. At times I practically lived in the computer room almost 24 hours a day. I remember very clearly the room — named after Princeton statistician Sam Wilks. There was an oil painting of Wilks, the minicomputer, Wilks' personal books, the old-fashioned teletype terminals, the fancier pin writer terminals, the tape drives, the clack-clack, buzz-buzz sounds everything made. I remember all the strange things that came with staying up late at night working on the computer in this place. Sometimes the computer users held "afternoon tea" at 3 am in the Math lounge!

I: Tell us about the computing tools you were developing.

D: The computer package was called ISP (Interactive Statistical Package). In addition to regression and data manipulation, it had all the basic tools of Exploratory Data Analysis. John Tukey's book Exploratory Data Analysis was just being published during my Junior year in college. Later, the software was used at hundreds of universities, especially after people at UC Berkeley took it over, revamped it, renamed it as BLISS and worked on it as their daily bread. Gradually, this got displaced by the S and R languages.

I: What were your statistical interests?

D: In those days, robust statistics — being able to cope with small fractions of really bad data — was a big deal. For my Junior Paper and Senior Thesis, I immersed myself in Annals of Statistics papers on robust statistics. Many researchers were interested in knowing the "right" score function to use in a robust (M)-estimator. For example, my Senior Thesis adviser John Tukey had proposed the "biweight" score function which I had programmed into the ISP software. I studied the notions of minimax optimality due to Peter Huber; you play a game against nature where you pick the score function, and Nature decides how to contaminate the data with outliers. I formulated and solved the problem of unimodal contamination. Years later Jim Berger (now at Duke) came across an equivalent problem from a different viewpoint (Bayes Decision Theory) and published an equivalent solution.

I: I believe you had some industrial experience after college. Can you tell us something about that?

D: I lived at home in Houston and worked for the research labs of Western Geophysical on problems in signal processing for oil exploration. It was after the second oil shock of the 1970s; there was tremendous interest in the oil business in finding new oil and developing new imaging and signal processing methods. I was assigned to work on what seemed (for the time) massive imaging problems. They had to fit a linear model with thousands of unknowns and tens of thousands of observations. This was huge for 1978; they used the largest mainframe computers of the day, filling up rooms the size of basketball stadiums run by hundreds of people in white coats. Computer jobs had to run for weeks to produce a single image.

The key point was that the data were actually of very bad quality, with many outliers; since I knew robust estimation inside out, I showed the geophysicists how to do robust regression. They were very eager; it took only days for a senior researcher to take it on himself. The result, immediately, was a much clearer picture of the subsurface. Western right away wanted to make this into a product and send me to conferences to speak about it. They sent me to London for an extended stay and I made presentations to the chief geophysicist at ARAMCO and developed further ideas about image and signal processing for geophysical signals. My first scientific paper came out of that work; it studied the problem of "blind deconvolution" where a signal has been blurred, but you don't know in what way it has been blurred, and you want to sharpen it up. That's a problem of real interest today. I worked on that in early 1979.

I: Was that after your PhD?

D: No, just the Bachelor's degree. I wrote a paper on blind deconvolution which I finished in early 1980, and it turned out that three papers on this appeared about the same time — one in automatic control, one in astronomy and my paper. About ten years later, such techniques began to be used heavily in digital communications (e.g. mobile phones). A big review paper in Proc. IEEE referred to my paper prominently, in the second paragraph.

Continued on page 13

Continued from page 12

I: How did you feel about that?

D: I felt lucky to come to the problem as a statistician, which gave me a broader view. I noticed connections between deconvolution and robust statistics, and saw that the key point was small departures from the Gaussian distribution. In robust statistics you viewed departures from a Gaussian error term as annoyances you want to protect yourself from. In signal processing, it was exactly upside down: you could view departures from Gaussian in the signal term as gifts from heaven, allowing you to recover the signal, against all odds! To explain this, I used things I had learned from the literature of robustness in the Annals of Statistics. My paper could actually be read by people in digital communications ten years later with some profit; I felt that if I had been working in a more narrowly defined subject matter, my papers would have aged more quickly.

I: How did you get your PhD?

D: I went back to graduate school at Harvard, and after finishing my paper on blind deconvolution I worked on robust estimation with high-dimensional data. I became obsessed with the idea that robustness was hard in high dimensions. Ricardo Maronna had shown that existing robust methods (like affine-equivariant M-estimates) could break down under fairly light contamination; I showed new approaches would be coordinate-free and avoid breakdown even with 49 per cent of the data completely corrupt. Werner Stahel did related work in Zürich at the same time.

It was beginning to dawn on people in those days that we should all be thinking about high-dimensional data. In the mid 1970s, Jerry Friedman and John Tukey had made a movie of PRIM-9, a system for looking at 9-dimensional particle physics data using 3D computer graphics. I remember the thrill I got when I saw that movie as a Junior in college (the Statistics Department at Princeton showed it to students to attract enrollment in the statistics major). My advisor at Harvard, Peter Huber, had been bitten by the bug and had gotten a very fancy-for-the-day computer — the Evans and Sutherland Picture System 2 — which could be used to look at three-dimensional objects from different angles and was shared with the Chemistry Department at Harvard. It had been developed for architecture and was being used by chemists to look at molecules. We used it to look at high-dimensional statistical data and we made movies to display our results — point clouds of statistical data, spinning around so you could sense their 3D structure. I gave talks at conferences and the high point of my talk was to show a movie of computer results; for presentation techniques, that was about twenty years ahead of the curve. I got to be a movie producer and screen writer!

At the same time I had to write statistical software for the VAX minicomputer that hosted the picture system; I remember working late into the night surrounded by chemistry grad students and postdocs. It seems that chemists were just not as weirdly interesting as mathematicians late at night. A lot of other magical things were not so magical either. The air conditioning was always mighty cold. The operating system was not UNIX, etc.

I: How did your industrial experience affect your later work?

D: In a typical academic career you get an advisor, get introduced to a problem and a specific field, become an expert, develop linearly. My industrial experience added a whole set of other interests to my academic portfolio. I'll give you an example. One thing that I learned in industrial research is that sparsity of a signal is a very important element in doing any kind of analysis of that signal. If you look at seismic signals, they are, in some sense, sparse. The reflectivity is zero a lot of the time with non-zero values relatively rare. And this sparsity was a fundamental constraint. I saw that seismic researchers were using sparsity to do surprising things that didn't seem possible. They were solving problems with too many unknowns and too few equations — somehow using sparsity to do that. Empirically, they were successful, but linear algebra would say this is hopeless. Over the years, that paradox really stuck with me. I felt that science itself involves too many unknowns and not enough equations and that often scientists are solving those equations by adding sparsity as an extra element. This somehow rules out the need to consider all the variables at once. In the last twenty years, I returned again and again to this theme of solving under-determined systems that seem horribly posed, and yet are actually not if you think beyond linear algebra, and use sparsity.

As a result, I worked frequently in applied math and information theory in addition to statistics. I have two careers, and this goes back to having worked in oil exploration.

I: Can you give some examples?

D: In seismics they have band-limited signals — that do not have frequencies outside a certain range — but they want to recover wide-band signals with all the frequencies that were not originally observed. It sounds impossible, and you can cook up counterexamples where it really is hopeless. But for signals that are sparse (most samples are zero and a few are non-zero) people were having good empirical success in seismology, and I worked with Ben Logan to prove that the problem is solvable. If you exploit sparsity, even though you only have band-limited information, you can recover a broad-band signal. Later, I considered the problem of representing a Signal which was made up as a linear combination of elements coming from

Continued on page 14

Continued from page 13

more than one basis, say sinusoids and wavelets. It sounds impossible, since the underlying system of equations has n equations and $2n$ unknowns, so there can be no unique answer. With Xiaoming Huo, I showed that this problem could be uniquely solved if the signal was made up of any sufficiently sparse combination from the two bases, simply by singling out the linear combination having the smallest ℓ^1-norm.

I: Although we've talked in this interview about computing and robust statistics, we haven't talked much about theoretical statistics. Yet, you have worked in this area extensively. How did you get interested in that field?

D: My theoretical immersion started early, as my undergraduate thesis solved a problem in robust statistics basically calculus of variations. I published a few theoretical papers as a graduate student, and even when they concerned "practical topics" like signal processing, they were ultimately based on things I'd learned from the Annals of Statistics, the main Soviet probability journals, etc. Also, Persi Diaconis visited Harvard one year while I was there; he made it easy to believe that theory was where the fun was!

I was lucky enough to win a postdoctoral fellowship at MSRI, the mathematics institute in Berkeley. The other young visitors included Iain Johnstone, who had just joined the faculty at Stanford, and Lucien Birgé, now a professor at the University of Paris. Both were interested in decision theory — Iain, the exact finite sample "Charles Stein" kind, and Lucien, the asymptotic "Lucien LeCam" kind. I hadn't had much deep exposure to either, and my interest in such subjects really picked up. The long-run role of those interests in my career has been enormous.

My first academic job was in the Statistics Department at UC Berkeley. When I arrived there, I was equally interested in computing, data analysis and statistical theory. My career could in principle have gone in any one of several directions. I was immediately given the explicit advice "don't get swallowed up by the computer". A certain faculty member had been spending lots of time revamping the statistical software ISP that I had developed as an undergraduate. Some faculty told me directly that I probably would go back to my computing roots and get "swallowed up" in the same way. Another faculty member gave me the advice that if I wanted tenure, I should publish ten papers in the Annals of Statistics. So the message was clear: do theoretical statistics!

Peter Bickel and Lucien LeCam were very kind and patient in speaking to me about their own work and interests. These personal qualities supported me in doing what could have been very isolating work. Bickel and LeCam were also patient in listening to me explain the results of my own work, as were David Blackwell and Rudy Beran. There was a steady stream of visitors giving very interesting talks at Berkeley, with most talks emphasizing theory. It was said in those days that attending statistics seminars at Berkeley could be painful, because seminar speakers would often want to present the most challenging, abstract, and technical achievements of their life to date — meaning that some seminars would seem impenetrable. But I found them mind-expanding.

I: What's the attraction of theoretical statistics?

D: On the one hand, it's about exploring the boundary of what can be learned and what can never be learned from a given amount of measured data. On the other hand, it's about taking what scientists and engineers are inventing, and making loud claims about, and subjecting those claims to scrutiny. I sometimes feel that if we didn't have theoretical statistics, science would degenerate into a crass business of people claiming they can do the impossible from their datasets, without any fear of critical scrutiny. Finally, some ideas in theoretical statistics are just beautiful ideas, very intellectually rewarding, I think of Wald's decision theory itself, of Huber's minimax robustness theory, of Stein's insight on shrinkage in high dimensions, of LeCam's equivalence of experiments theory. You have to make a decision in Life about what ideas you want to spend your time with. These ideas wear well as constant companions.

I: According to the Institute for Scientific Information most-cited website, "incites.com/top/2004/third04-math.html" you are the most highly-cited mathematician for work in the period 1994–2004, with 23 highly-cited papers and well over 1500 citations to your work. Do you think that citation counts are important? How can statisticians increase their citation counts?

D: I'd like to emphasize that many of those papers are joint with my co-author, Iain Johnstone of Stanford. In fact he's number two in that list, close behind me. Statisticians do very well compared to mathematicians in citation counts. Among the top 10 most-cited mathematical scientists currently, all of them are statisticians. There's a clear reason: statisticians do things used by many people; in contrast, few people outside of mathematics can directly cite cutting-edge work in mathematics. Consider Wiles' proof of Fermat's Last Theorem. It's a brilliant achievement of the human mind but not directly useful outside of math. It gets a lot of popular attention, but not very many citations in the scientific literature. Statisticians explicitly design tools that are useful for scientists and engineers, everywhere, every day. So citation counts for statisticians follow from the nature of our discipline.

A very specific publishing discipline can enhance citation

Continued on page 15

Continued from page 14

counts: Reproducible Research. You use the Internet to publish the data and computer programs that generate your results. I learned this discipline from the seismologist Jon Claerbout. This increases your citation counts, for a very simple reason. When researchers developing new methods look for ways to show off their new methods they'll naturally want to make comparisons with previous approaches. By publishing your data and methods, you make it easy for later researchers to compare with you, and then they cite you.

The important thing: do the reproducible research; don't worry about citations. My website has a paper on reproducible research giving the philosophy in more detail.

I: You have written that statistics is an "invisible" profession. Could you elaborate on that?

D: Many people don't even recognise that statistics exists as a discipline in academia. They are surprised when they hear that one can be a "Professor of Statistics" . Statisticians, in general, don't do public relations. I think we're all too busy. There are not enough statisticians to go around. The world is flooded with data; scientists, engineers and doctors all wanting to analyze their data. Outside every statistician's office in the world, there is a line of people waiting to get in to get some help with their data. Since we are completely over-subscribed, no one is out there advertising the existence of our profession. It is a sort of secret.

I: How do you select the problems that you work on?

D: This is the problem of life, isn't it? Some problems are urgent because many people are interested in them; I like to do those once in a while because of the challenge. I often look at articles in Science and Nature. When people write articles that make a big splash, I try to understand what they did and I either criticize it or build on it. So that's one angle. Another angle is to study some fundamental area of mathematics where a breakthrough just occurred, and to trace out implications in the real world.

I: In your 2004 American Statistical Association President's Invited Address, you spoke about missed opportunities for statistics. Could you elaborate?

D: Many fun problems in computer science could be attacked by statisticians, but statisticians don't even know about these problems, partly because they are already "fully booked" . Today statisticians are immersed in genomics; but there are many, many other interesting problems that are equally urgent. Go to a conference like NIPS on neural information processing. There is work on analyzing catalogs of images and sounds, problems of all sorts in signal array processing that come up in electrical engineering. There are so many interesting datasets, so many interesting problems, so many great opportunities!

I: You have worked with wavelets. How is that related to statistics?

D: Wavelet theory is a fascinating branch of applied mathematics — harmonic analysis, numerical analysis, approximation theory all come together. Studying wavelet theory you learn about representing problems, about representing signals, about representing noise.

This background is useful in statistical theory. In non-parametric estimation, everything depends on your assumptions about some unknown regression function or unknown density function. Coming merely from a background in statistics, you don't have tools to think deeply about your assumptions and how they should be represented. By learning what wavelets are all about, you suddenly understand a lot of things that were mysterious in non-parametric estimation. A simple example: often nonlinear estimators dramatically outperform linear estimators in nonparametric estimation and regression, even in problems where everything seems linear and convex and banal. Once you understand wavelets it's very easy to understand this phenomenon and extend it in many directions.

There's a wide collection of signals and stochastic processes where modeling by wavelets is appropriate — any time you have impulsive events or long memory. Many non-Gaussian stochastic models are very important in applications — remember that the Gaussian is a myth. In certain application areas such as Internet traffic, if you come with only a Gaussian stochastic process background or only a Poisson process background, you just cannot analyze the data perceptively. So knowing about wavelets widens your scope quite a bit.

The wavelet transform broadens your mind in the following way. If all you know is the Fourier transform (which every statistician has to learn in the guise of the characteristic function) then you have in your mind only a very poor collection of transforms. As soon as you have wavelet transform, you suddenly realize that there are not just two — i.e., not only Fourier and Wavelets — there are many, many transforms. The right one can depend on the data you are studying.

Finally, for many kinds of signals, the wavelet representation is sparse. That gives an impetus to the statistician to study high-dimensional parameter vectors where the vector is sparse, with relatively few big entries. Iain Johnstone and I were very inspired by this viewpoint, and it has influenced all my later work.

I: What do you think will be the forces shaping the future

Continued on page 16

Continued from page 15

development in statistics?

D: Statistics is a data-driven discipline; each time someone invents a new kind of data there is always an infinite supply of new questions. Genomics is an example: microarrays came along and there were enough new questions to keep all statisticians busy. There are many new kinds of data. For example, we are now entering a world of ubiquitous sensors where there are sensors on your body, sensors in space, and everywhere sensors are talking to each other. Because of this sensor network, there will be many new questions. Another example: all sorts of data come out of analysis of blood chemistry. In proteomics, they subject blood samples to high-resolution mass spectrometry and get very finely-resolved spectra that reveal all the chemical constituents present in the blood. They hope to detect diseases early and forecast about your health. All the time we see new data sources creating enormous volumes of data with completely different structures from anything we have seen before. Basically, we need statisticians to cope with this onslaught of new data types. Each new one is going to cause a revolution in our field because you have so many new questions arising from each new data type.

I: You mentioned revolutions. Do you think there will be some conceptual revolution that will change the direction?

D: Over the last twenty years there was a shift away from an intellectual attitude, where you think very carefully before you do something, to a computational, experimental attitude where you quickly do something with the computer. At some point this will run its course and statisticians won't be able to really do much of value simply by running to the computer. Then there will be a whole bunch of new questions which arise out of dealing with these new data structures; they'll ask "what can we learn from graph theory?" or "what can we learn from theoretical computer science?" We'll go back to a much deeper level of thought in order to make the next step. I think that's coming soon.

I: In that case, do you think there is need to relook at the way statisticians are getting their undergraduate training in order to meet the challenges you just mentioned?

D: They should be good in mathematics and computers, and really care about analyzing data. In some parts of the world, statisticians are just trained at math and they aren't interested in science. In some other parts, they learn a lot about data but are not well-trained in math. In most parts of the world, they don't get enough computer background to really push the field. It's a three-legged stool — you need all three. That's really demanding for an undergraduate education, but I just don't see any other way.

On the one hand, things are much easier these days. We used to have to work really very hard to get the computer to do anything. For even the most routine analysis, I had to write a short computer program, get it into the computer, wait for the results, and if I made one tiny mistake, I had to start all over again. It's much easier these days. On the other hand, in economically advanced countries like Singapore, in Europe and the United States, kids have so many possible entertainments that few will choose to really use their minds. It is very unlikely that more than a small number are going to look at a field and say, "Oh, this is so inspiring, I want to know everything about it." Kids will pursue social life and many other diversions. Plus, they'll be "cool" and sophisticated and materialistic. Finally, in a comfortable society parents may be a little afraid if their kids are too intense about study and consider it unhealthy.

Every once in a while at Stanford, I see a kid with that "look" in the eyes. I know they still exist. We get some of them in the graduate program. I'm very fortunate to have had some great students who have gone on to become very distinguished scientists in their own right. I know that more great young minds are out there. That's for sure.