Mathematical Conversations

James V. Zidek: Bridges Bayesians Build >>>



James V. Zidek

Interview of James V. Zidek by Y.K. Leong

James V. Zidek is world-renowned for his research on Bayesian decision analysis, monitoring network design and spatial prediction.

He received his education from University of Alberta and Stanford University. Since 1967, except for a few short stints elsewhere, he has established a distinguished career in teaching and research at the University of British Columbia in Canada. He has also been actively involved in consultancy work in public health, engineering and industry; in particular, he did pioneering statistical work on long span bridges. An emeritus professor since 2005, he continues to apply his expertise and professional experience in addressing statistical problems arising in environmetrics, a multi-disciplinary discipline that has recently emerged to deal with environmental problems like pollution and climate change.

He has been invited to give lectures at conferences and workshops throughout the world. He has actively served on numerous committees of professional bodies, societies and international scientific meetings and on editorial boards of leading statistical journals like *Annals of Statistics, Canadian Journal of Statistics, Environmetrics and Journal of American Statistical Association*. He is a Fellow of the Institute of Mathematical Statistics, American Statistical Association and Royal Society of Canada, and an elected member of the International Statistical Institute. He received the Eugene Lukacs Symposium Distinguished Service Award, Izaak Walton Research Prize, Gold Medal of the Statistical Society of Canada, and Distinguished Achievement Medal (Environmental Statistics Section of the American Statistical Association). He was a President of the Statistical Society of Canada. In addition to an impressive record of research papers, he has also produced a long line of masters and doctoral students.

Zidek's association with National University of Singapore dates back to 1995, when he was invited to the Department of Mathematics for a short period. He came back as a member of the organizing committee and speaker at the IMS program (6 – 28 January 2008) on *Data-driven and Physically-based Models for Characterization of Processes in Hydrology, Hydraulics, Oceanography and Climate Change,* jointly organized with Pacific Institute for Mathematical Science, University of British Columbia. He was interviewed by Y.K. Leong on behalf of *Imprints* on 24 January 2008. The following is an edited transcript of the interview in which he exuded tremendous energy and passion as he talked about his views and experiences in both theoretical and applied aspects of statistics.

Imprints: Your undergraduate degree was in mathematics and yet you went on to do your graduate studies in statistics, ending up at Stanford University for your PhD. What shaped the choices you made, what benefits do you see retrospectively in choosing Stanford for your degree?

Jim Zidek: I did enjoy my mathematics undergraduate training. In fact, I was generally quite good, not brilliant, but quite good in doing math. However, I found the statistics courses rather challenging, I think my love at that time was number theory. When it was time to choose my graduate program, I decided that I would be into statistics. That was at the University of Alberta. My Masters program had an immensely stimulating man by the name of John McGregor as my thesis supervisor. What made him particularly good was the fact that he let his students do pretty much what they liked under some general supervision. In my Masters thesis I worked in mathematical learning theory – construction of models which try to predict how things like rats would learn

Continued on page 12

Continued from page 10

In this program, particular emphasis is placed on stochastic processes as models in ecology, and especially on the part which both population genetics and network structure play in their behaviour. The program aims to bring together people actively involved in different aspects of mathematical biology, to exchange ideas and to further promote the development of the field. Since many of the models of interest have characteristics similar to those encountered in other disciplines, such as discrete mathematics, statistical physics and computer science, members of these communities are warmly encouraged to participate.

Coding, Cryptology and Combinatoric Designs (15 May – 11 June 2011)

... Jointly organized with Nanyang Technological University

Computational Prospects of Infinity II, incorporating a Summer School and a Workshop on Infinity and Truth (15 June – 5 August 2011)

... Jointly organized with Department of Mathematics, National University of Singapore, funded by the John Templeton Foundation

when they are in those maze experiments. The high point came one day when I discovered a result about something or other, which I presented to John, and he seemed immensely positive about it. He was so excited that he actually went out of his office to tell some of his colleagues about this result. That was the first time that I began to appreciate the joy of discovery. I was turned on to research at that point and decided to go on to Stanford.

I was fortunate to get in there with its high reputation in the field of statistics. By then I knew I wanted to go further on. What led me to Stanford in the first place? You ask about the benefits of Stanford. The obvious one is the great intellectual benefit that I derived from the faculty. At that time there were very few departments of statistics in the world. I picked that one in part because it was one of the few and one of the more recognized. One of the benefits I began to appreciate long after I graduated was the valuable experience I gained from meeting people from all over the world and many of them have remained friends ever since. Many students and even faculty overlook the benefit of getting to know people when the opportunity arises.

I: Was there any person at Stanford who influenced you greatly?

Z: Yes, Charles Stein, my supervisor at Stanford had a profound influence on my thinking. He is a man who has very high standards for himself and for me, but he did not impose on me in an autocratic fashion, again leaving me to my own devices. I think we all owe a great debt of gratitude to our teachers. They have given us a great gift which we sometimes overlook. I single him [Charles Stein] out though there were also other people – many of the professors there and other visitors – who influenced me greatly.

I: What were you working on at Stanford for your PhD?

Z: I worked in the field of statistical decision theory which was a subject that Charles Stein was involved in. That subject enjoyed a tremendous amount of interest in the statistical world in that period. It was stimulated by a guy named Abraham Wald at Columbia University. In his relatively short life that unfortunately ended in an airplane crash in 1951, he introduced two subjects: sequential analysis and statistical decision theory. He had emulated in statistical decision theory what Kolmogorov had done in probability theory. Kolmogorov recognized that people had been trying to define probability for a hundred years or so. He had the brilliant idea of adopting the more mathematical approach of axiomatizing probability theory. He set down fundamental axioms saying that although we may not know what probability is, or how to define it in some operational sense, let us say instead that any quantity that satisfies these simple rules will be deemed a probability or probability distribution. That frees the mathematician to go on to develop probability theory without having to figure out what it means. The subject has become a tremendously important part of mathematics. Wald's idea was to build statistics on axiomatic foundations. His decision theory was an attempt to account for uncertainty in decision making and make it a rigorous discipline.

I: Was Wald's axiomatization of statistical decision theory done after Pearson-Neyman's work?

Z: That's a really good question. It sort of generalizes it. Some credit should certainly go to Neyman and Pearson who realized that the testing theory lacked one or two important elements and added those in, and Wald then further generalized it. That came in the late 1930 or 1940s. Some 30 years later, the subject became a premier subdiscipline of statistics in North America and attracted a lot of great minds, like Charles Stein.

I: After Stanford, you went immediately to University of British Columbia and you essentially stayed there for your scientific career until you retired as Professor Emeritus in 2005. Are there any particular reasons for being so loyal to one university?

Z: That's a very interesting question. The answer is that I was never loyal to the University. In fact, I think one attaches oneself to one's department or college. So any loyalty I had was really to the department I was in. My first years in that university were spent in the mathematics department. It was a department of some size, about 70 members. While I was very well treated in that department, I found it to be an increasingly poor environment for nurturing statistics. When I was asked by a senior administrator (I guess it was 2 years after I arrived) what might be the ideal model I told him that a separate department of statistics would be a good idea. That wasn't achieved, but we did propose and get around 1971 instead as a compromise an institute of applied mathematics and statistics. At the graduate level anyway, even then at that time beginning around1970, statistics in North America had emerged from its mathematical shell and had become more focused on science. It was moving towards statistical science. I think we all sensed that, so the Institute of Applied Mathematics and Statistics was actually meant in some way to synthesize these two subjects and bring them closer. It gave us some freedom in organizing graduate programs at the institute. I realized later though the decision to set up an institute solved some of our problems, but not all of them. In fact, our effort to get a department of statistics was held back. The administrators had argued, "Listen, the institute may not be a perfect answer but it's a lot easier than setting up a department. You've already got it, so let's leave it at that." It was not until 1983 that we got a separate department of statistics at the university. By then, I had realized that things weren't going so well for statistics in the department of mathematics even though they were fine for me personally. So I decided to leave. I did, in fact, join the University of Washington's Department of Statistics. It turned out to be a wonderful department and still is. Although that move was intended to be permanent, for various reasons that we need not go into here, I eventually left and accepted the University of British Columbia's invitation to me to come back to form a new department. The department was, in fact, formed in my absence in 1983. I was lured back to become its official founding member in 1984 but my colleagues really deserve tremendous credit for pushing it through and getting it up and running before my return. Although I regretted leaving the University of Washington in many ways, I don't think with hindsight it was a mistake to go back [to British Columbia]. At the end of the day, I have certainly enjoyed my life there since.

I: You prefer to live in Canada than in the States?

Z: Not really, I enjoyed both countries very much although the cultures are different – the way the funding agencies are set up for example. Also because we inherited our system from England to some extent, our academic programs are shaped a little bit by our history.

I: But now the Canadian is more geared towards the US system.

Z: Yes, there's a lot of evolution around the world now, [with the US system] gradually emerging as the most common choice. But at the graduate level, I think we remain different in terms of less course work and more emphasis on the thesis ... as they do in England and Europe. In that sense, we remain separate. At the undergraduate level, we [Canada and US] are very similar, I think.

I: Your early work in statistics was quite theoretical dealing with more fundamental conceptual issues than applications to concrete problems. What led you into the area of applications?

Z: When I first graduated, I thought I was really smart. I thought I really understood statistics. When I wrote all those abstract symbols on the blackboard or in my notes, I understood perfectly what they meant within the mathematical context. But it was consulting that revealed my overly optimistic self-evaluation and started me on the road to getting interested in applications. I would meet with the engineers and talk to them about things like, for example, the technical idea of independence of two random variables. They wouldn't understand what I meant, so I would start to explain it to them only to realize that when I started to translate this mathematical language into something that they could understand, I was having difficulties. I realized for the first time that what I called "independence" is actually conditional independence. This kind of interaction began to make me realize that I didn't really understand what I once thought I did. I learnt a lot about my subject from trying to explain to them. They helped me in that way and, of course, I eventually helped them. I found it interesting to work on the design of long span highway bridges. Consulting did spur me on. I think statistical consulting is a very sensible or at least complementary alternative to research, as a way of keeping up with the discipline and being motivated to study more. I certainly would encourage all statisticians to get involved in some such activities. It is even better when you get to do research work with people in some other field.

I: Did any of your consultation work give rise to some theoretical problems?

Z: Yes, indeed, that's the whole thing, isn't it? Unless you are working according to a very strict time table, which you sometimes have to do in consulting work, if you have the time, there is always an unsolved theoretical problem of interest hidden in the consulting problem. Indeed, textbook solutions almost never work. They have to be adapted in some way to fit the particular problem you are addressing. That is even more true today because generally academics and even non-academics have become more sophisticated

about statistical methods. So when they finally come to talk to the statistical consultant the problem will likely be quite sophisticated and that means there is likely to be some interesting theoretical alternatives. In my case, the bridge design problem got me into extreme value theory. The work with my colleagues developed an alternative approach to extreme value theory. It was a useful contribution in some ways.

I: On that particular problem, did the engineers seek you?

Z: Yes, they knew they needed a statistician. Unlike structural bridges where you have to build to carry the maximum imaginable load such as an army tank or truck, in the case of the long span bridge, spanning a thousand meters say, now you have to go to the maximum statistical load of traffic since you can't realistically picture the bridge to be completely covered with say army tanks or heavily loaded trucks. That would be unimaginable as a statistical event. They understood that and that's why they came to me. I must say that they already knew a fair bit about statistics when they came to see me. That enabled a fairly rich communication on the problem and I benefited from it on what was to be my first big consultation project though I had quite a few smaller ones. This was done over one to two years, and I had 2 or 3 publications coming out of it. Also, there was a bridge design code, the first one ever published for the design of long span bridges. It was eventually endorsed by the American Society of Civil Engineers and at that time anyway, it became the code that could be used by engineers designing long span bridges – so my colleagues told me. But I'm not an engineer and I've not followed the history of what happened since then.

I: You once described yourself as a Bayesian. Could you explain to a non-specialist what that means?

Z: I think all of us are Bayesians in the sense that when we get up in the morning and go through a day, we have to make a lot of judgments based on our experience and knowledge gained in the past and our anticipation of things to come. What distinguishes the Bayesian theory is that it tries to formalize what you and I actually do to make decisions, namely use our background. The formalization equates uncertainty and probability. Now that I think of it, we all do that anyway. When we are say betting on a soccer match, we arrive at a bet where I offer you 5 to 3 odds on team A. That is a quantification of my belief about the outcome of the game. I think that was actually the famous de Finetti's idea when he proposed probability as a measure of uncertainty. He realized that in ordinary conversation we are always making predictions about things being likely or probable, rain being less likely tomorrow, etc. Since probability is already a natural language in ordinary human conversation formalizing its role in inference seems appealing. Coming back to Kolmogorov, this formalization would be a way of defining probability to fit into the Kolmogorov framework and thus make the probabilities in this Bayesian framework bona fide probabilities. This is what really distinguished Bayesian theory. When Kolmogorov developed his axioms, the relative frequency definition of probability is the one involved. You think of the probability of heads being one half for a toss of a coin as meaning the fraction you will get

by tossing the coin over and over again, and calculating the fraction of those tosses when it turns up heads. But, of course, that theory falls on hard ground when you think about the probability of a highway bridge falling down in any one year as being 1 in 100 by design. You cannot imagine building that bridge over and over again repeatedly and running each replicate for a year to find out what fraction of them actually fell in that year. So the Bayesian theory came into its own because it gave an effective alternative to the relative frequency theory. As a result, it has become immensely fashionable these days in a whole variety of disciplines.

I: Who was Bayes?

Z: [Thomas] Bayes was actually a minister in the Presbyterian Church. Although he was a theologian, he actually wrote and published one paper on something like number theory. But he never published his famous thesis on probability. It was only published after he died, in 1764, I think, through Richard Price, a friend of his who had gone through his papers on his desk and found the famous thesis. He [Price] presented this paper to the Royal Society in England after Bayes had died. It pretty much lay dormant. The famous Bayes Theorem appeared in that paper in a certain way and he is therefore accorded the honor of being its discoverer. However, Laplace in France came to the same approach independently and some years later, Laplace developed a theory of probability based on applying probability to uncertainty. But it was only in the last half of the 20th century that the theory began to bloom and nowadays it's quite standard.

I: It wouldn't sound the same to call one a Laplacian.

Z: No, that's right [*laughs*]. Bayesian and Laplacian mean quite different things.

I: Mathematicians often see the conclusions of statistical investigations as non-rigorous and even subjective. Does statistics have rational foundations that validate the methods of inference that have been developed and used?

Z: The Wald framework was an attempt to rationalize decision making. The Bayesian framework is based on axioms. In fact, there is a complete Bayesian decision theory based on axioms. This framework is taken in conjunction with the axioms developed by economists to imply the existence of something called the "utility function", a measure of gain when an action is taken. So the Bayesian framework incorporates probability, which is axiomatic, and utility, which is also axiomatic. The rational theory of Bayesian statistics is based on axioms of both theories. Wald's theory was deficient in that his so-called "loss function" was not itself predicated on axiomatic foundations. Its role seems to be analogous to that of 'point' or 'line' in geometry as a basic building block on which to create an axiomatic theory. But the meaning of the loss function proved more elusive. In reality, the business of statistics doesn't derive from any axiomatic foundations and is inductive rather than deductive. One of the great things about the subject is the great freedom in exploring data and knowledge discovery.

Yet there is the deductive side, which is one of the hallmarks of the subject. It does stand on rational principles. There are the algorithms and you have to know how they work. There are the various performance criteria like unbiasedness or asymptotic efficiency. These are things meant to justify the methods even though you know the samples are never going to be infinite. Nowadays what has become a fairly standard alternative to having a broader based performance theory is computer simulation – you try a whole variety of artificial situations to see whether you get the right answers. That is not the same as theorems, of course, and we can't ignore theorems. Simulation may be reassuring but it is not quite the same as the truth expressed in a theorem.

I: Are there cases where for a given problem different statistical methods actually give different results?

Z: Oh yes, indeed. It is important not to apply statistics mechanically. One has to develop some understanding of the problem and apply the methods in an intelligent way. If you do get different answers, you face a real challenge and you have to go after the data to find out why.

I: In that case what is true becomes subjective.

Z: Yes ... it depends on what is meant by "true". The result of it is that the state or validity of belief depends on the evaluation of the data and the degree to which one or another of these analyses will have contributed to the change in opinion or belief. The outcome is not the same as the outcome of a theorem where by a matter of definition you have the notion of 'truth'.

I: Thirty years ago, the common perception of statistics is that it is about finding averages, standard deviations, confidence intervals and other statistical quantities. How much has this perception changed since then?

Z: Tremendously. I think the biggest thing to have changed statistics has been the recognition that it has something to contribute to science. I mentioned earlier that statistical science started to emerge as an important discipline in the latter part of the last century. It made a shift in favor not so much of applications but collaborative inquiry in other disciplines. Statistics took on a much different nature. It became a sort of detective job to look at data often in conjunction with scientists from other disciplines to try to divine some new knowledge from that data. That in turn has led to a lot of theoretical challenges for statistics. It's now in a very healthy state with questions coming from other disciplines. At the same time, I think that the core must be preserved. I fear sometimes that we are awash in applications and that people who do really work on the hardcore of mathematical statistical theory may be losing out in postgraduate programs, research programs and so on. I fear that this may lead to a loss of our identity. Statistics students know there are all kinds of options in areas like biology or biostatistics where they tend to gravitate rather than work on subjects that require a lot of hard mathematical background. But I must emphasize that this kind of work can require a lot of difficult, sometimes even mathematical, thinking.

I: Some people seem to think that statistics does not involve too much deep theoretical thinking when one is applying it.

Z: Yes, but even in applied statistics, it can involve an awful lot of thought and understanding as I learned when I first started out in consulting. On the other hand, you are absolutely right – applied statistics can offer opportunities for purely routine analysis.

I: Your recent research interest is in environmetrics. Is that a new discipline? Could you tell us something about it?

Z: Thank you for that question. It's actually quite new and it began, I think, with a group at Stanford University in the seventies organized under the auspices of an organization called SIMS (Society of Institutes of Mathematical Scientists). It was set up to try and find important societal problems that could be addressed by mathematicians and statisticians. So a group was created at Stanford to look at air pollution problems. It was under the direction of Paul Switzer. They did a lot of very good things, both theoretical as well practical, studying air pollution. The great thing about that group is that a great many people, academics and non-academics, students, faculty got involved in seminars and projects in learning about this world of environmental statistics. That was how the subject got started. Air pollution was quite a problem in California at that time. The name itself may have come from the President of SIMS, Don Thomson, or it may have come from Abdel El-Shaarawi who is at this workshop for a couple of weeks. In any case, it was born as a discipline in the latter part of the 20th century. On the other hand, there wasn't really a lot of interest in the subject in mainstream statistics until about 10 years or so ago. Then it started to really take off. It's now a flourishing discipline with lots of sessions at conferences and so on.

I: Is it very multi-disciplinary?

Z: Yes, it's inherently multi-disciplinary. At conferences and workshops, you will find people from statistics and non-statistics coming together and talking about this kind of questions – research scientists, meteorologists, even civil engineers.

I: Recently there has been much concern about global climate change. Moreover, the Intergovernmental Panel on Climate Change said it is very likely due to anthropogenic sources rather than natural sources. What is your position on this matter as a statistical scientist?

Z: Wow, I should start out by saying that I don't have the expertise of the panel that won the Nobel Prize for their work. But from the statistical perspective, I think what is interesting is the great uncertainty that abounds in that field. A lot of discussion at our workshop has been around the question of which model to use, for example, how you plug-in the uncertainty about these models, which kind of scenarios to use, and so on. There is a healthy recognition that there is a lot of uncertainty about this whole question of climate change. In particular, there is a lot of uncertainty about how much is exactly due to anthropogenic causes, how much is due to natural process. I know that the

International Panel on Climate Change has come down saying it is very likely that climate change is, to a substantial extent, due to anthropogenic causes, but trying to figure out how much seems quite a challenge. Of course, statistics is always about analyzing uncertainty and quantifying it and so on. It's an important opportunity for statisticians to get involved in what is arguably the most important issue of our age. We must do that and we must get involved in this kind of questions.

I: Were there any statisticians on the Panel itself?

Z: Hardly any, Peter Guttorp being the only one I know. But I was involved in the early 1990s, thanks to the International Statistical Institute, in trying to get ourselves as statisticians on that Panel, and we did not succeed. I don't know why. At the same time I do know that these scientists do know a lot of statistics, so I'm not saying their work is flawed. On the other hand, there's a lot of discussion recently about something called a "hockey stick", with a blade that rises steeply from the handle and tells us that the climate changed, tentatively anyway, a lot over the last century. There has been a lot of controversy about that stick among non-statisticians, as to what that really represents and there is an argument that it is flawed. That analysis anyway might have benefitted from some input by statisticians. How such an expensive project was launched and collected so much data without having statisticians on board is a mystery.

I: Do you have any reservations about the general findings?

Z: No, it only exemplifies that all these things are uncertain and that things could be a lot worse than you would have it ... One other thing is the "Prudence Principle" says that if you don't know what's going on, you had better be conservative. Even though we are uncertain about what has happened, I think it's appropriate to take some action to reduce our impact on the environment, just in case the worst case scenario might in fact obtain.

I: Do you think that statistics should be made compulsory in the undergraduate science curriculum or even in high school?

Z: I certainly do. In my university, a great many students do and they take it over a period of 4 years – 3500 to 4000 students in any one year. They are not all taught by statisticians. High school is a bit trickier because I know some examples where teachers, who don't have the resources in terms of projects or interesting demonstration examples, tend to rely on using these methods that you described in one of your earlier questions on standard deviations, confidence intervals and that that kind of mechanical exercises. I've seen some of that in my own experience. In that case, it might do more harm to do statistics in high school because it might just turn students off and make them not do it in the university. .. About 100 years ago, H.G. Wells said, "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." I believe that the time has come and everybody ought to have some knowledge of statistics. In the modern world, we are inundated with data. It used to be that there

Publications >>>

Continued from page 15

was not enough data, but nowadays, there is far too much data. I think the average citizen has to cope with figures and information and make important decisions about his or her life, the government and so on. I think that knowledge of statistics will certainly be needed.

I: Statistics is often perceived to have wide applicability to other disciplines and hence have higher market value in terms of career opportunities. What advice would you give to undergraduate or graduate students who are motivated to specialize in statistics?

Z: I guess these are valid reasons. I think the subject has its own beauty and worth studying for its own sake, but I am amazed to have found a demand for statisticians over the entire span of my career. Except for a brief period following the Tiananmen Square episode, statistics graduates, particularly at the Masters and doctoral degree level had no difficulty finding work. Over the last 30 years or so, there has been a tremendous demand and the trend seems be growing. The specific advice I would give to someone interested in a non-academic career in statistics would be to attain the Masters level because at that level you learn statistics to some depth where you can apply it on a wide range of problems. Undergraduates sometimes get jobs in that field but I think those jobs tend to be less than interesting. They



Order direct from publisher at http://www.worldscibooks.com/physics/7797.html

do not really open up a wide range of interesting problems. It's not the money issue. The job satisfaction is much greater based on a Masters degree than an undergraduate degree.

1



Institute for Mathematical Sciences National University of Singapore

3 Prince George's Park Singapore 118402

Phone: **+65 6516-1897** Fax: **+65 6873-8292**

Email: ims@nus.edu.sg

Website: http://www.ims.nus.edu.sg

Editors: TAN Ser Peow mattansp@nus.edu.sg

> Denny LEUNG matlhh@nus.edu.sg

Drafts & Resources: Claire TAN Web: Stephen AUYONG Printer: World Scientific Publishing Pte Ltd For calls from outside Singapore, prefix **65** to local eight-digit telephone numbers. For email, add the following domain name: **userid@nus.edu.sg**

IMS Staff

Louis CHEN	Director	6516-1900	imsdir
TAN Ser Peow	Deputy Director	6516-1898	mattansp
Emily CHAN	Assistant Manager (Finance)	6516-1893	imscec
Claire TAN	Executive (Human Resource)	6516-1892	imstlf
Stephen AUYONG	IT Manager	6516-1895	imssass
Jolyn WONG	Laboratory Technologist	6516-1890	imswwy
Agnes WU	Management Assistant Officer (Secretary)	6516-1897	imswua
Nurleen BINTE MOHAMED	Management Assistant Officer (Housing)	6516-1891	imsnm
Rajeswri	Operations Associate		imsrs