THE ET INTERVIEW:
PROFESSOR JOSEPH B. KADANE

Interviewed by Ngai Hang Chan
Chinese University of Hong Kong

Joseph B. (Jay) Kadane.

Joseph B. Kadane, usually known as Jay among his friends and colleagues, is a well-known figure among statisticians and econometricians. He has made substantial contributions to the fields of Bayesian statistics, econometrics, and many applied areas. He is well known for his work in applying statistics to how people make decisions and draw conclusions from data.

Kadane was born in 1941 in Freeport on Long Island. He finished his bachelor's degree in mathematics at Harvard in 1962 and his Ph.D. degree in statistics at Stanford in 1966 under the supervision of Herman Chernoff. He joined Carnegie Mellon in 1971, serving as the head of the statistics department from 1972 to 1981. He currently holds the Leonard J. Savage Professorship of Statistics and Social Sciences. Jay was named as the University Professor at Carnegie Mellon in 2000.

Among his many contributions and honors, Jay is an elected fellow of the American Association for the Advancement of Science, the Institute of Mathematical Statistics, and the American Statistical Association. Jay was named

Jay has been actively involved in leadership activities with many prestigious statistical societies and groups, including the International Statistical Institute, the American Statistical Association, the Committee of Presidents of Statistical Societies, the Institute of Mathematical Statistics, the Institute of Statisticians, and the International Society of Bayesian Analysis. Jay was the editor of *JASA* during 1983–1985.

Jay said that he learns the most by choosing problems in disparate fields. He has worked on problems in econometrics, political science, archaeology, law, psychophysics, environment, and medicine and computer science, among other topics. He is interested in many aspects of Bayesian theory. It is fair to say that Jay is a truly cross-disciplinary statistician.

Jay has been a consultant to several business and government organizations and has also done pro bono work for public defenders on death penalty cases and racial profiling. His testimony was instrumental in underscoring a general pattern of racial profiling by police patrolling the New Jersey turnpike. The results of his analysis are reflected in several research papers and in news coverage of the issue.

Jay enjoys traveling, and he is a familiar face and a regular participant at conferences in statistics and econometrics. The following interview was conducted in May 2000 at the DeGroot Memorial Library at Carnegie Mellon.

1. BACKGROUND AND EARLY STAGE

Perhaps you may want to start by telling us something about your background and the early stages of your life.

Well, I grew up on Long Island, in Freeport, which is on the south shore near Jones Beach. My early interests were bird-watching, chess, and reading. I did a lot of reading and also some stamp collecting.

Besides chess, were there any other indications about your interest in analytic subjects?

I liked math. I was also a baseball fan, in particular, a Brooklyn Dodgers baseball fan, back when there was such a thing as the Brooklyn Dodgers. And of course in baseball there are many averages, batting averages, fielding averages, and so on, that are very important, and I think that might have been part of my early interests in statistics.

You mentioned that you read a lot; any particular books that strike you?

Gee, it's hard to remember now. But I just remember doing a lot of reading and enjoying it.
You started out in mathematics at Harvard in the early 1960's. Was there any distinction between pure mathematics and applied mathematics at that time?

Yes, I started out as a math major in Harvard, but there was no distinction between pure and applied. There were some of each. On the applied side, there were some computing courses. On the theoretical side, I took calculus, real analysis, the standard stream. Since I had quite a lot of math in prep school, that set me up for the math at Harvard.

Were there any statistics in the math program at that moment? Were they taught within mathematics or statistics?

Well, I did have a full year of statistics there. The first semester was taught by Art Dempster and the second semester by Fred Mosteller. I believe there was a separate statistics department, but I don't remember whether it was taught under a math number or a statistics number.

How do you describe the atmosphere at Harvard during those years, and what made you choose to become a math major?

The atmosphere was quite puzzling for an undergraduate in that it was clear that there were a lot of things going on in mathematics, but in a sense it wasn't clear what they were. The communication between the faculty and the undergraduates was not very strong. So it was puzzling to me. I should also say about my time at Harvard that I probably spent more time and more effort on the social sciences than I did on mathematics. Just because I was a math major doesn't mean I was necessarily spending the majority of my time doing math. Since I didn't know what I wanted to be interested in, I figured that mathematics was the most difficult thing around so I thought I would major in that and that would help me in whatever I decided to do, and that's pretty much turned out to be true. I still don't know what finally I'm interested in, but mathematics turns out to be useful in all of the things that I'm interested in, so that's fine.

You mentioned that you were interested in social sciences at an early stage in Harvard. How did that manifest in your interest in statistics?

Well, I was, at the time, very much involved in political matters at Harvard, especially about arms control, and took a number of courses involving defense policy and arms control at the time. I think my interest in statistics came out of having been a math major and having been interested in political matters and thinking of statistics as being sort of halfway in between. So that is one of the attractions of statistics for me. I was sent to Howard Raiffa at the business school in Harvard for advice for what I should do about graduate school given my interests. He told me that he thought the two best places for me might be either Harvard or Stanford, depending on what a particular professor did. Kenneth Arrow was thinking about moving from Stanford to Harvard at that point,
but he did not. So Raiffa’s recommendation was that I should go to Stanford, which is what I did. Arrow did move to Harvard later though.

Were you interested in economics when you went to Stanford?

Yes, I took quite a number of economics classes at Stanford. Some of the most influential people on me at that point in economics were Arrow and Uzawa, Hirofumi Uzawa.

Did you start your work at the Center for Naval Analysis (CNA) before going to Stanford?

Yes, I had a summer job with them in 1962, which was right after I graduated from Harvard. I worked with them for the next summer as well. I had taken a number of courses with Thomas Schelling at Harvard, and he recommended me to the Center for Naval Analysis.

2. POST-PH.D. YEARS

We will come back to discuss something more about your thesis later. After spending 4 years at Stanford, you had already been working on an on-and-off basis for the Center for Naval Analysis as a consultant. What made you start an academic career at Yale?

I worked for the center not just through the summers but during the semester as well. After finishing my Ph.D., I was appointed both to the statistics department and the Cowles Foundation at Yale. Given my interest in social sciences, I was excited about this opportunity, and I decided to start my career at Yale.

Could you tell us something about the Cowles Foundation then?

Well, it was a remarkable group of people there, I suppose there still is. Koopmans was the head of Cowles at that time. The person that I probably worked most with was Mark Nerlove. David Grether was also working with Mark at the time. Jim Tobin was there, among the younger people there in my group, Marty Wiesmann, Al Kleverick, and Joe Stiglitz. Also, Herb Scarf was a senior faculty member there, and Shubik in game theory was there. In fact he and I shared an office for a while. It was a very lively group. There were probably about 20 people in Cowles. Statistics was much smaller at that time, and it is still small nowadays. Jimmy Savage and Frank Anscombe were the leading figures in statistics at that time. It was a very rich environment. I was there for only 2 years but learned a great deal from being there.

Were there any people having a joint appointment like you?

I was the intersection at the time. There were no formal connections between the Cowles and statistics except my joint appointment. Fortunately, they were
just a few houses apart on Hillhouse Avenue, so it was not difficult to separate my time between the two.

After that, you went back to the Center for Naval Analysis. Would you tell us something more about the nature of the center and what kind of work you got involved in?

CNA is funded by the Department of Defense (DoD), but it is not a government agency. It was part of the University of Rochester, which received a contract from the DoD. Although it received money from the navy, it was somewhat independent of the navy as well. It's called a federal-contract research center. Some 70% of the budget for CNA, at the time, and I believe still is, was dedicated to projects sponsored by one or another office in the navy. Some 20–25% of the budget was for studies which CNA thought were important to the navy, even if there was no navy sponsor. And this enabled CNA to do things which, in fact, the navy was interested in having done but which it did not wish to officially sponsor. For example, at the time there was a draft. There was talk about the possibility of ending the draft. CNA sponsored on its own, as part of its 20–25% budget, to study the question of whether there would be enough volunteers for the navy if the draft were ended. And it was able to do this before it became an official doctrine, to look into the possibility of ending the draft. Later on, there was a presidential commission set up to examine the all-volunteer armed force as a concept, and because CNA was the only group that had been doing work on the manpower response to the end of the draft, my colleagues at CNA became the staff for the president's commission on the all-volunteer armed force.

There were several different divisions in CNA, ranging from defense related matters such as missile defense studies to operations research of the navy and to whatever was bothering the admiral in the morning. There were also people in CNA doing long-range strategic studies about the politics of the use of armed force in general. Of course the Vietnam War was going on at the same time.

How big was the statistics group in CNA?

Actually, there was no separate statistics unit within CNA at that time. Because I had been active politically and in particular was very concerned about the war in Vietnam, my deal in going to CNA was that I would not be assigned to a single project but rather that I would float among projects, and if I didn't like the politics of a project I could have nothing to do with it. So, I was really as a methodologist to poke into whatever I thought was interesting. Another thing they wanted me to do was publishing papers; they thought it would help with recruiting and make it a more interesting place to work. I also offered some courses in statistics to people on the staff. Although it was government related, my job there had an academic flavor.
Besides you, were there any other statisticians?

Yes there were a number of other statisticians around. Probably the one I worked most heavily with was Warren Rogers, who had gotten a Ph.D. from Stanford a couple years behind me and who was a naval officer but was assigned to CNA. There were a quarter to a third of the staff who were navy people in uniform. Warren and I worked together on a number of things.

The 1960's to 1970's was an exciting but intriguing period of time in U.S. history. You mentioned earlier that you were politically active, but on the other hand, you carried a top secret security clearance to work for CNA. How did that happen?

The politics of the time was much more interwoven and confusing than your question suggests. While I was at Stanford, as we've said, I was a consultant for CNA, the whole time. I was carrying a top secret security clearance the whole time. I was also a student activist. I was Stanford coordinator for the free-speech movement. I was Speaker of the student legislature. And I had a very large hand in running the student government. Many of my friends were very heavily involved in the antiwar movement. I was not so heavily involved, but I was certainly involved in confronting the Stanford administration about many things, and so in the Stanford political context I would be regarded, probably, by the Stanford administration as a dangerous radical student. Yet, on the other hand, I was involved in this defense work on the other side. That was just part of life at the time. One of my very close friends who was editor of the *Stanford Daily* for a while, with whom I planned a great many political things, joined the army as a Green Beret and became a prisoner of war in North Vietnam for 5½ years.

Needless to say I didn't participate in the decision to give me a top secret security clearance, so I have no idea what those discussions were about. Although CNA was 100% sponsored by the government, the deal between the navy and CNA was that studies done by the center were to be the opinion of CNA. Reports from CNA could be distributed within the Department of Defense without any constraint. The navy had the right to put a letter on top of the report which would give the navy's opinion of the matter dealt with within the report. For example, in terms of the issues that were hot at the time, CNA could write a report saying that aircraft carriers are very vulnerable, and CNA suggests that the navy stop building them. The navy, which was very attached to its aircraft carriers, could write a letter on top of that saying they think this report is wrong because of a number of reasons. But the report and the letter would go forward to the secretary of defense. So it was set up so that there would be a great deal of intellectual honesty in what happened. These were not studies where the navy was directing the outcome of the study. The navy would know about the study, would know about what it had to say, would know about
the reasons, but could agree or disagree with them as it chose. And, the navy supported having a group with that kind of independence because they knew that in the long run, they did better, even if sometimes they were hearing things that they didn’t want to hear. The center was a very interesting place for me to work, and I learned a great deal there. When I was at Yale, really intellectually the dominant person there for me was Jimmy Savage. When I was confused about something about statistics, I could talk to Jimmy, and Jimmy would straighten me out. What was different at CNA is that at CNA I was like Jimmy, and all sorts of people were turning to me for that kind of advice. And that made me think much harder about the principles and about what was fundamentally important and what wasn’t in statistics than I had before. And so I think it shaped my future development in statistics quite a bit.

Was that the time period you started thinking seriously about a Bayesian paradigm? Were you involved in Bayesian statistics when you were at Stanford?

No, at Stanford I was purely a classical statistician, although Herman Chernoff was sort of open to Bayesian ideas. He had written that book on decision theory with Lincoln Moses. But Herman was open to Bayesian ideas. I learned a great deal from Lincoln Moses at Stanford, who was quite classically minded, but more empirically minded, more open to doing applied work than some of the others at Stanford. But yes, it was Jimmy who introduced me to Bayesian ideas most seriously. He and I taught out of Dubins and Savage (1965). I learned a lot from doing that with him. But it wasn’t really until I was at CNA that I started to become much more seriously Bayesian.

Why was that so?

Well, because I had to think through for myself what I really believed. And I had some early experiences that showed me that the methods I had been brought up with were not all that useful to me. When I was going to Yale the book that I most wanted to teach out of and which I thought of as the fundamental place to begin in statistical theory was Lehmann (1959). I worked with a sociologist, Gordon Lewis, who is now at Carnegie Mellon University (CMU) and with Gerry Ramage, now at AT&T. We were looking at participation rates in small groups. And there was a theory about participation rates in small groups, and we tested that theory and found that the theory was being rejected at the .05 level, and the .01, level and in fact the $10^{-6}$ level. And then I had to think to myself, would I be more impressed with how false this theory was if it were rejected at the $10^{-13}$ level? Answer, well no, not really. We found a way to plot the data, and it turned out that really the center of the data was very close to the theory. But we had thousands of observations. Every utterance of someone in one of these small groups was an independent observation in our theory. So,
when you hold up the square root of \( n \) magnifying glass, it was many standard deviations from the theory, but it was still quite a good theory, as things go in sociology. So here was a method that was really in my way and not helping me, when I had a great deal of data. Well, that already made me wonder about it, because you would think that a statistical procedure should work best when you have a lot of data.

At CNA I had another experience where a study had been done on a new machine that had been worked over very carefully in the laboratory and then taken to the field. And some field testing had been done on it. An analyst had written up the results, which said it doesn’t work significantly differently in the field than it does in the laboratory. And this went all the way up the chain of command, and just before it left CNA one of the senior scientists at CNA said, "I have a funny feeling that there’s a problem here. I want a statistician to look at this." So one of us and ultimately several of us looked at it, and what had happened is that there were five observations. They cost about a million dollars apiece to collect. So that’s the reason why there weren’t a huge number of observations. The machine was working roughly 75% as well in the field as it was in the laboratory. But with only five observations, it was not significantly different at the traditional .05 level. So the analyst was completely correct but entirely wrong.

What I learned from this is here’s a technique that I thought was the place to start in statistical theory, and it obscures the truth when the sample size is too small and it obscures the truth when the sample size is too large. And we certainly have better methods for examining sample size than all this testing theory. So then what good is it? A question I still haven’t managed to find an answer to. That was the beginning of seeing that in terms of the applied problems that I wanted to do, the usual classical methods simply were not helpful to me. And for that reason, I had to change.

3. PITTSBURGH YEARS

Perhaps we can move on to something which is more familiar to both of us. You came here in 1971 and then served as the head of the statistics department for the next 9 years. You have been here for almost three decades now, which is a very long period of time. Although you and I unequivocally agree that Carnegie Mellon is an exciting place, are there any particular reasons that attracted you to stay here for such a long period of time?

Well, I think that certainly one reason is that this department, as it has developed, reflects my values and what I think is important. And so the reason for doing all that work as head of the department was to make it a good place in that it works well internally and it presents what I think is a good policy toward the university as a whole. The way we do our courses, the joint work that
Jay (with student at left) and Morris DeGroot (partly obscured at right), Statistics Lounge, Carnegie Mellon University, 1984.

we do with people in other departments, and so on. So having built all of that, why move?

Maybe you could tell us something about the kind of difficulties you ran into when you first came. Did you have a model for the department in mind?

There were only two tenured professors in the department in 1971, DeGroot and myself. DeGroot was on leave my first year here in 1971. During that year his wife became very seriously ill with multiple sclerosis, and it was just clear that he couldn't be effective as a department head given his home situation, so I had to be department head. I was just 6 years out of graduate school, and I was supposed to run the show. In order to accomplish that task, I had a great deal of growing up to do in a very short period of time. Although there was a department of statistics at that time, we did not belong to any college. I was reporting jointly to the dean of the business school and the dean of the engineering school, and our computer money came through the science school, so we were doing business all over the place, and I had a lot of learning to do about how the university worked and how to get things done.

We had about six faculty in 1971. At the time, most of the statistics departments that I knew about were very theoretically oriented. Because of my interest in econometrics and social sciences, I wanted to have a department which was much more engaged in applied work. It was a matter of having a balance between theory and application. One of the early decisions that we made was not to have a consulting center, and the reason for that was that consulting cen-
ters were and maybe still are a way for people to have a patina of statistics in their work without it being a fundamental part of what they’re doing. So, they would come in and say, “How do I do a t-test on this data,” and only reluctantly tell you where the data came from. That’s not what I wanted. What I wanted was that statisticians should be partners in the whole enterprise from the beginning, so what we established was we conduct joint research with you, but we’re not consulting. We set out to do that and at the same time to absorb the statistics courses that were being taught in various different departments in the university. What we found was that conducting joint research in this way helped us a great deal in overcoming the diplomatic obstacles to absorbing the courses. Once we had been doing joint research with them, it would be much harder for others to make arguments like: statisticians don’t understand what we do, so statisticians won’t be able to teach courses for our students. So these went together and have gone together well at Carnegie Mellon ever since.

Another important early move we made was the agreement that we would judge applied work solely by its contribution to the applied area. Thus a junior statistician joining an applied project would not feel pressure from the department to use fancy new statistical tools but rather would have the same goal as his or her scientific partners. We also agreed that if the applied area was one we didn’t know about, we would get outside advice from others who did. Thus we assured our junior colleagues that applied work would be taken seriously.

4. RESEARCH

Let us begin with the question: what fascinates you most in econometrics?

Well, there are certainly many interesting things going on in econometrics. I’m still very interested in the relationship between game theory and Bayesian decision making, which lies at the heart of a whole bunch of econometrics. It seems to me that the classical game theory that comes really from the zero-sum two-person game recommends things which I would still not recommend in that I think that they are much too conservative. I would prefer to ask myself what do I think the other player is likely to do? What are my probabilities on what the other player will do, and then given that, my expected utility maximizing action is not a difficult calculation. It seems to me that the mini-max solution is unstable in that I can always think of a worse consequence than the consequences I’ve been thinking about so far. And then the mini-max idea would have me put all of my attention against mitigating this even worse consequence that I just thought of. I don’t understand why people really believe in the mini-max solution to games: Why do they get out of bed in the morning? The worst things that can happen if they get out of bed are surely worse than the worst things that can happen if they stay in bed. So, I don’t think there’s much to the whole foundation of the mini-max solution and Nash equilibrium. I think they are very shaky. And much of modern econometrics—
modern econometric thinking—is based on those ideas. And yet I find people
very resistant to thinking hard about the foundations of what they’re saying.
Also in econometrics, as it is presently engaged in, the model that econo-
metricians have of the rational man is basically a Bayesian. In each econo-
metric model, when you model individual behavior that’s generally how it’s done.
On the other hand, most econometric analyses are classical, not Bayesian. So
why is it that the rational man is Bayesian but none of the econometricians are?
I don’t understand this.

I guess I don’t know enough economics to answer your question, but
I’m sure we’ll be hearing responses from our colleagues in economics.
Let me move on to a different subject. You have been involved in devel-
oping data mining activities here at CMU. What is your opinion about
data mining, and how do you see its role both in statistics and in econo-
metrics, in particular?

I think there are both good things and bad things about data mining. Certainly
data mining is paying attention to what one can learn from large databases. It is
an attractive subject for statisticians and econometricians because perhaps many
things can be learned from those databases. On the other hand, some of those
databases raise important questions having to do with the origin of the data-
base. For example, suppose one has the records of one or many hospitals and
one has patients with a particular diagnosis, some of whom are treated one way
and some of whom are treated another. It is a tempting thing to think that one
would, by analyzing the results of how those patients fared under their treat-
ments, be able to say something about the effectiveness of the treatments. We
know however, from bitter experience in medicine, that this is not the case.
There may be important differences in determining why some patients got the
treatment they got which may not show up in the database. And that difference
in turn may be correlated with the response they have to treatment. One of the
things that we know is that sicker patients don’t do as well as less sick patients
do, regardless of what treatment you give them. So if the sicker patients are
being given treatment A and the less sick patients are given treatment B, then
A is going to look worse than B. By just looking at the data, one would not be
able to tell whether that was the case or not. Certainly, there are things to be
learned from data mining, but there are also inferential traps that need to be
thought hard about in order to “mine” the data well. We need to free the data
from hidden biases.

Maybe we can talk a little bit about your personal research. At the
beginning of your career, Bayesian statistics wasn’t considered to be in
the mainstream of statistics. Could you tell us something about that en-
vironment in those days?

Well, one of the most useful institutions for me was the NBER seminar on
Bayesian inference in econometrics, which got going I would guess around 1969
or 1970. I was involved with it when I was at CNA before I came to Carnegie
Mellon. It met every 6 months. There were usually 30 people from out of town and 20 people in whatever place we were meeting. I found that for quite some years the papers that I was writing were being written in response to questions or conversations or papers that had come up in the previous meeting. And for many of us that was true, so it became sort of an intellectual center of our activities. And it was extremely helpful particularly in the early days in helping people to understand what the difference was about Bayesian statistics as opposed to classical statistics and what the strengths of it were, and so on. So that was a very important influence on my development.

Is that still going on nowadays?

It is, although I think it’s a bit dormant now. It doesn’t have the same regularity and the same centrality that it had when there were only 20 or 30 of us, which is really what there was of Bayesian statistics. Of course, that is no longer true now; things have changed enormously.

Incidentally, a similar NBER time series seminar in econometrics and statistics has also been developed for the last 20 years which I find to be very useful, at least personally. How do you see Bayesian statistics’ relevance to statistics as of today?

Certainly Bayesian statistics is doing very well. In fact, its main problems now are the problems of success. Particularly with the development of Markov chain Monte Carlo (MCMC) methods everybody wants to use it; one can get an array of models with MCMC that one really can’t get in any other way. Consequently, all sorts of people who before had no use for Bayesian methods at all suddenly now are very interested in at least running chains. The problem with it is that these same people don’t really understand what Bayesian methods are about. They sometimes take the attitude that they’ll assume whatever it is one has to assume in order to get to run a chain without really understanding what the implications are. But as I said, that’s a problem of success, and it will just take a while for people to understand the philosophy behind the methods they are using.

Are you seeing the same phenomenon happen in economics?

I don’t know enough of what’s happening in econometrics. I haven’t seen it as much there though, but I certainly see it in statistics.

One of your first pieces of work out of graduate school is on the so-called small-sigma asymptotics, which had been capturing a considerable amount of attention in econometrics. Even as of today, papers are being published relating to that particular area. Could you tell us something about the background of the problem, what led you to think about it, and getting the solution?

Well, that was my thesis. I remember that the first thesis topic that I tried to work on was about the two-armed bandit. Chernoff had done a lot of work on
the one-armed bandit, and his idea was that I should extend this to multiarmed bandits. The difficulty was that Chernoff could write down the differential equations that controlled the bandit problem and he would sort of know that those were the right differential equations and later justify them. There was a series of four papers that he wrote on the bandit problem, and the equations were in the first paper, and then there was a justification in the second paper, and then there was further justification in the fourth paper. So I would ask him, "Where's the proof?" and he would say, "Well, it's not quite like that; it's sort of like you know it has to work this way." I just somehow couldn't think the way he thought about that problem. And so, fairly far into working on a thesis, we came to the conclusion, he and I, that the two-armed bandit and I were just not made for each other and I should work on something else. I had several other papers that I had written at that point, and he thought that I needed something to go with it. In some of the earlier work that I had done and had shown him, he had sort of casually remarked that, you know, asymptotics doesn't have to be large-sample asymptotics, you can do asymptotics with respect to any parameter. And just sort of left it lying there. But I was thinking about econometrics and thinking about the fact that ordinary least-squares was asymptotically biased, where two-stage least-squares and limited information maximum likelihood were not. Consequently, ordinary least-squares was dominated by these other two methods in large samples. But somehow that didn't seem right, because for lots of problems ordinary least-squares seemed just as good, and so, somehow the asymptotics weren't being as informative as they might. Then I thought that perhaps letting the variance of the residuals get small was a better ideal case than letting the sample size get large for comparing econometric methods. Asymptotics are always an ideal case. So this is simply asymptotics as the model gets very good, rather than as the sample size gets large. That was the start of my dissertation, and then I found this terrific paper by Nagar, who had done the large-sample case, and I essentially adapted his methods for the small variance case. In the space of about 4 months that was my thesis, and it was alone my thesis. The other papers were published but just didn't appear as part of my thesis. With that kind of asymptotics, I showed that all three of the estimators can be compared to one another. They're all the same order of magnitude as sigma gets smaller.

I published a few papers after my thesis on it. The main one came out in *Econometrica*, and then there were a couple in *JASA* on test statistics, but as I was writing those papers I was also becoming Bayesian. I became less interested in all these classical calculations and less convinced that it was the right thing to do. I had also started looking at time series problems from that point of view, autoregressive moving average models. The difficulty is that the likelihood was the same, and therefore any Bayesian calculation would be the same. But from a sampling point of view they're not the same. And that puzzled me for a while, and as I chewed on it I decided that the Bayesian thinking was more appealing, and I moved my research in the Bayesian direction and didn't pursue small-sigma asymptotics anymore.

Your joint research in Laplace approximations with Rob Kass and Luke Tierney has made important contributions to Bayesian methodology and has connected in many important ways in econometric work. Could you comment for us on how you started out with this idea and came up with the approximation?

Well, it’s a funny story. I was working on an applied problem having to do with missing data in the National Crime Survey, a household survey about whether people had been the victim of a crime. And the idea that I was interested in there is that nonresponse in that survey could be an indicator that the household had been the victim of a crime. They would be less likely to open their door to an enumerator if they’d been the victim of a crime than if they had not. So nonresponse was informative. And so I had a model and was estimating that model. I was doing it in the Bayesian way, but it was, shall we say, inelegant! Luke was in the department at the time, and I described this problem to him. Luke said something like, “Oh, for pity’s sake,” and over the weekend he did an approximation. Numbers that were taking me an hour apiece to get, he was getting in 19 seconds. His approximation was the same to two to three significant figures, and I was pretty impressed. At that time, Rob Kass was also interested in integral approximations, and so we started to look into how Luke did this work. As we looked into it, I said that there was something very general there, and out of that came the work on Laplace approximation.

How do you see it goes from here? Do you think the problem is pretty much resolved?

No, I still think there are interesting things to do with it. Since that time, one thing that I did do is some work with Partha Bagchi, who was a postdoc here at the time, and he and I looked at circular distributions. The interesting aspect is
that once you know how to do it for circles, you know how to do it for spheres of any dimension; once you know how to do it for cylinders, you know how to do it for tori, and so on. It's a different kind of Laplace approximation because what plays the role of the normal distribution is the Fisher-von Mises distribution. I have also talked with Sid Srinivasan about doing the same thing for covariance matrices. This time we have another kind of topology; we're dealing in a space of positive definite matrices rather than the whole real line, or circles, or whatever. I still think that there are good things to be done on these other kinds of spaces.

As you alluded to earlier, the idea of MCMC has been getting a lot of successes recently. It seems to leave the impression that MCMC has taken over the Laplace approximation. Could you offer some comments on this issue?

I think the Laplace approximation and MCMC are useful for different things. The MCMC idea is a very powerful one. But it is likely to be used especially for analysis of a particular data set. On the other hand, when one comes to experimental design, say, to assess the expected value of a design, one has to draw from the prior and then draw from the likelihood, and then approximate the analysis that one will do on that data set and see what its value is at the end, and then draw another. Of course, one has to do that very quickly. In this case, one doesn't want to run a chain for each of the samples one might draw.

In this situation, I think the Laplace approximation is going to work very well. It may not be as accurate as a full chain would be, but it is very good as a fast approximation. It can also be used effectively as a method for getting good starting values for a chain. There are lots of ways in which these ideas can be useful to each other in different contexts. There is room in the toolbox for both of them, and there are different occasions on which each of those tools would be useful.

One of the main themes throughout your research across disciplines is the idea of elicitation. Could you elaborate a little bit more about this idea?

Sure, I think elicitation is a critical issue for subjective Bayesians. It is the issue we cannot avoid, namely, if you're going to use a subjective likelihood and a subjective prior, which subjective one would you use? How are you going to get it written down? And how are you going to find out what others think about a process? If you are dealing with experts in a field, how are you going to find out what they think? And in particular how are you going to deal with them when they don't understand statistical models very well? They don't understand the parameterizations that we use. In the context of a linear regression, asking them what is their covariance between \( \beta_2 \) and \( \beta_3 \) is hopeless. It's hopeless, in fact, for most statisticians and econometricians as well. As a little challenge, try to write down a non-diagonal covariance matrix that you think is
your covariance matrix for any three or four variables you care to name and then check to see whether what you’ve written down is positive definite or not. It’s very hard to make it positive definite. So, we don’t do these things intuitively very well. The question is, can we find methods that will ask people a series of questions that they find natural, from which we can deduce what their prior is. That’s what elicitation is all about. I did some early work in that area, a five author paper that appeared in JASA in 1980, and I’m still interested in it. Lara Wolfson, one of my former students, and I did a review paper on elicitation for JRSS a couple of years ago, and she’s been pursuing elicitation since then.

A related question along this line is that you have also done quite a bit of work on the foundation of statistics with Mark Schervish and Teddy Seidenfeld, the “three musketeers” at Carnegie Mellon. Would you say something about how you see the relevance and importance of the foundation of statistics?

Oh, we’ve been working together I think about 20 years, and we’ve certainly learned a lot together, and it’s been wonderful fun. The most recent published work has been what we’ve been calling reasoning to a foregone conclusion. There’s a paper in JASA with that title. And what that’s about is that with finitely additive models you can be in a position where you have a prior today, you’ll see data tonight, you already know what your posterior will be tomorrow, and it is different from your prior today. This observation leads to two questions. First, which is your prior, what you think today or what you know you’re going to think tomorrow? Second, it leads to decision problems in which you would pay not to see data! This is contrary to a theorem which is valid only in the case of countable additivity. And so it sharpens the issue of whether finite additivity is worth going into. Since then, we’ve been trying to quantify incoherence. There are classical statistical procedures which we know are incoherent, but how badly incoherent are they? That is, can we distinguish some that are real catastrophes from others that are perhaps not so bad? So we’re working on ways of quantifying how incoherent, incoherent procedures are.

I think it is important for there to be some work going on in foundations. The foundations of statistics are not as secure as one might think. There are implications of the assumptions that it’s easy to gloss over. Consequently, I find it useful in my own thinking to be doing some work in the foundations because it helps me think through the implications of the methods that I’m using when I do applied work.

It seems to me what you have done is mainly on the foundations of Bayesian statistics. What would you consider to be important for the non-Bayesian world as far as the foundation of statistics is concerned?

There are no foundations there, and therefore it is very hard to identify foundational issues. Basically non-Bayesian statistics is sort of a grab bag of meth-
ods and no ordering of principles. When should you, say, use a unbiased estimator, when should you use an invariant estimator, when should you use something that's least-squares, and so on, and so on. Well, there are no principles for these things. Just different ideas that people have had, and they will say things like "Well, I would use them only when they are sensible" without saying anything about a theory of sensibility. So you don't know what they really mean. And consequently it's not really a theory; it's basically a mess. And so there's a sense in which Bayesian ideas are the only theory of statistics that we have where people are willing to say, "These are my principles, and I would seek to apply them in every instance." So without such a statement it's very hard to have anything to really work with.

I agree with you that the Bayesian paradigm seems to be more axiomatic, while classical statistics seems to be lacking this type of articulation.

It's not just articulation. The reason that classical statistics hasn't done that isn't a lack of articulation on the part of classical statisticians; it's that their principles contradict each other and do not stand up to the kind of scrutiny that Bayesian principles do. For example, there's a whole wing of classical statistics that wants to do what they call conditional inference. Do they condition on part or all of the data? If they condition on all of the data, then they're doing likelihood and probably Bayesian inference. If they want to condition on part of the data, which part of the data do they wish to condition on? Well, I asked one practitioner of conditional inference, and what she tells me is that "We've been studying that for 40 years, and we still don't know." How then can one examine the foundations of conditional inference? It is in this context that I say that there are no foundations for classical statistics.

5. CONSULTING, LEGAL AND CENSUS CASES

That's very interesting. I'm sure we'll be hearing many interesting comments from our colleagues on this issue. Let me change the subject a little bit. In studying your CV, I think everyone agrees that you are truly cross-disciplinary. You conduct research in areas in economics, sociology, political science, medical science, physical science, computer science, environmental science, law, archaeology, and the list goes on and on. You have contributed to almost every discipline that requires statistical expertise. You have already discussed your view on statistical consulting. Could you also tell us something about your experiences as an expert witness in legal cases? Is there a single case you would like to share with us?

I do work as an expert witness, I enjoy it, but it is very demanding work in that I have to think very carefully about what I'm saying and how to say it, and so on, in that context. It is very different from the usual scientific discourse in that
the attorneys for the other side are seriously out to get you as an expert witness, where that is not usually true in academics. In the law, they're out to embarrass you as much as they can. I find it useful to publish my cases because I find it constrains the competitive impulse that I might have to "win the case," as opposed to trying to represent the truth as best as I can. So there's always a double loyalty that an expert witness has. Putting myself in a position where I know that whatever I say in court I will be explaining to all of my colleagues in print helps me to restrain that competitive impulse and I think makes me a better and more persuasive expert witness as a result. I've enjoyed that a lot. I think the single case that I'm most pleased about is the turnpike case concerning the southern end of the New Jersey Turnpike, the so-called Soto case. The state police in New Jersey were accused of discriminating against blacks in who they were stopping and who they were arresting. Norma Terrin and I worked on that together; I gave the testimony. The attorney general's office in New Jersey was on the other side. I was called by the public defender's office. We found that, despite a large amount of missing race data, the evidence was overwhelming that blacks were being stopped at rates several times those of whites, but that virtually everyone speeds on the turnpike. The attorney general representatives who were there had nothing but scorn for the analyses that we were doing. They lost the case; they appealed, saying it was a terrible insult to the state police. Just before the appeals court was to hear the case the attorney general and the governor of New Jersey had to withdraw the appeal because they had done their own study and found out that we were right. So this was a
very pleasing outcome, and Norma and I have published two papers about the analyses that we did in that case.

You have also been involved heavily in the census undercount problem. To some extent that also involves a lot of legal issues there. Would you care to say something about it?

Sure. I was an expert witness in the 1980 case; that is, the case having to do with the 1980 census, which I think, if I remember correctly, got to court in about 1983 or 1984. It was ultimately decided in 1986. New York City and State were suing the Census Bureau and the secretary of commerce about this. I was involved in a less time consuming way as an adviser to the expert panel that was set up having to do with the 1990 census on similar matters. Now I am watching developments in the census of 2000.

In the 1980 case, basically the position of the Census Bureau at the time and of the department of commerce was that adjusting for the undercount could not be done. Since I had done it in research with Gene Erickson, basically the import of my testimony was “What do you mean it can’t be done? I just did it; here it is. You may not like how I did it, but it certainly can be done.” The only way I know of to address an argument that says something can’t be done is to do it. The census is doing it both ways in 2000. They’re doing both an unadjusted and an adjusted census, and then every state is going to have to argue out for itself which it should use, so we can anticipate many lawsuits coming from the census of 2000.

I recall that the other side also had statisticians as expert witnesses who argued against adjustment. Could you comment a little on this issue?

I don’t think the argument now was whether it can be done but more about whether it should be done. There were many discussions about the quality of the adjustment against the quality of the raw census, and so on. The way the adjustment is done is by taking a random sample and using that as the basis to figure out how much adjustment should be made in what places and then how much, what direction? So you would think that sampling would be generally agreed as a reasonable thing to do and that this is not terribly difficult. I interpret the issue now as being far more a partisan political issue than I do a technical one. Certainly there are statisticians on the other side who would disagree with that.

6. TEACHING, STUDENTS, AND SUPERVISION

You have directed a large number of students, and a number of them are now having very successful careers. If you were to offer advice to potential thesis supervisors, what would you like to tell them?

I think the single most important thing is to learn your student. That is, to listen to your student, understand your student, understand what they’re inter-
ested in. Understand what their particular bent or talent is. Understand what kinds of issues are most likely to get in the way of this student being successful in terms of what they want to do and then address those issues. And those might be technical, and they might not be technical. And my way of doing it is, I will tackle whatever the most important issues are.

Some people feel that any good statistics, be it methodological or theoretical, has to be subjected to the test of the data. Do you feel the same way? If you were to have a student who worked out a theoretical result, do you feel it would be kind of incomplete until the student applies the theory to a data set and sees how it works?

No, I don’t think that. Because my observation is that problems have solutions, but solutions don’t necessarily have problems. If I have developed a technique, I may not have a particular application for it; that’s fine. On the other hand, if I start with the data and work on the data, then presumably at the end I will have something to say about that data set. I may not have anything new theoretically to say. If a student takes a theoretical problem and works on it, then no, I don’t think necessarily that it needs to be a contribution applied to data in order to be a satisfactory dissertation. It’s interesting if it can be; it’s a plus if it can be applied. But as a requirement, no.

But what about statistics? Do you consider a real contribution to statistics has to be subjected to the test of data?

Well, statistics is such a broad field. There’s so much going on in it, it seems to me that is too narrow a perspective on it. We virtually wiped out the whole Annals of Statistics by saying that. One can think of important ideas in statistics that did not immediately have an application and yet were certainly useful things to have published. One can also think of many things for which there was a section in the paper saying, “Well, here’s some data, and here I did it,” which would have died right there because nobody thought it was a useful idea. So I don’t see it either as necessary or as sufficient to make it an interesting piece of work. But it is certainly a good thought. But I don’t—no, I don’t think a necessary one.

Statistics has been changing rapidly during the last decade. These changes bring in new research as well as new pedagogical issues about the statistics curriculum at all levels. For example, many statistics departments are cutting back the theoretical requirements of the students, measure theoretic probability and statistics, for example. On the other hand, our colleagues in econometrics and finance are sending more and more of their students to take courses in advanced probability and stochastic calculus. What is your opinion about these changes?

I think that over the years I have become an old curmudgeon in these things. I’m on the more conservative side of the debate about the curriculum in statistics. I still think that it’s important for students to be exposed to measure theo-
retic probability. As for econometrics and finance, I think that those tools, especially the Itô's calculus, have become very useful and important tools and students would naturally want to learn them. For statisticians who are not working in that area maybe going that far in measure theoretic probability and so on isn't as important. It all depends on the kinds of problems you want to solve and the tools that you need to solve those problems.

Do you think that statisticians are shortchanging their students in reducing the theoretical training?

Well, I think those are very useful kinds of training to have because it opens a literature that otherwise is going to be very difficult for people to learn on their own. And that really, I think, is the importance of it. There are other things that we teach that perhaps aren't as critical in the sense that people can pick them up on their own later, if they need to. But I don't know very many people who teach themselves that kind of mathematics.

Also, I think that part of this new trend is a response to the wonderful things we can do computationally that we couldn't before. And to some extent we are attracting some students who would otherwise possibly be in computer science, instead of statistics. Those students tend to be less mathematically oriented than the ones that we are used to, who came mainly through math departments and math undergraduate degrees. As I said before there's room for lots of different kinds of statisticians, and that's one of the reasons why you have to listen to your student. Certainly I have had students who came through a computer science kind of training and whose impulses are more like those of a computer scientist than those of a mathematician, and such students are making very valuable contributions to statistics now. So I have no complaint with that or any desire to say that isn't as good or important a student or a kind of training, but I do believe that measure theory and measure theoretic probability are very useful things for a student to study.

7. GENERAL

Throughout your career, you have been appointed to the editorial boards of many premier journals. At the same time, you have served on many professional committees, both on the national and the international levels. In addition, you have also engaged in all kinds of consulting activities. The fact of answering all the e-mails and inquiries from these activities seems to be a daunting task to many people. Besides having a very effective secretary like Heidi, could you tell us your secret on how do you manage that?

Heidi\(^1\) is certainly an important part of it. Another part of it is, try to be organized. I set priorities for myself and try to do the most pressing, most important things first. And I keep a to-do list, and I look at it and I check things off as I do them, and I add new things to it as I think of things I should do. And so on.
I remember you had a very good idea about refereeing: could you share it?

Sure. What I try to do with refereeing is that, when things are sent to me to referee, I either referee or don’t referee in a week. It either gets to the top of my list and I do it, because it’s sufficiently interesting to me and I have the time to do it, or if I don’t I send it back to the editor saying, “I don’t have the time to do this.” And that way I do not keep a long guilt list of refereeing that’s getting older and older that I haven’t done. Being a former editor of JASA, I found that the referees that were problematic for me were the people who kept a paper for a long time and didn’t do it. But if I heard back quickly from a referee saying, “Sorry, I can’t do it,” that was just fine. I think I do maybe two-thirds of the papers that are sent to me. I think that the length of time that it takes to handle papers in statistics in general is too long, and I think the reason for that is that people put off doing the work. The amount of work is the same whenever they do it, and therefore I think that if more people had a policy similar to mine, we would hear much sooner about our papers than we now do. As a result, everyone would be much happier.

Although nothing is typical here, could you tell us what would be a typical working day schedule for you when you are at Carnegie Mellon?

Recently this past semester much of my time has been absorbed by being chair of the faculty and chair of the faculty senate. And consequently going to many meetings and doing a lot of problem solving for the university in general. In addition I was teaching two courses this spring, one master’s level course for our students and one physics graduate course jointly with a professor in physics. These together with research meetings were the main things I was doing this spring.

Do you have any particular paper that you’ve written that you think of as a favorite?

Oh, I guess if I had to choose one I guess it would be the elicitation paper because I think it is the most important issue given my Bayesian convictions. That’s certainly the problem which I think needs more attention in research.

It’s certainly very enlightening to discuss these various issues with you. On behalf of ET, may I thank you for your time and giving us this opportunity to conduct this interview.

You are most welcome.

NOTE

1. Ms. Heidi Seeherich has been the secretary for Jay during the last 15 years. Her assistance in this interview is gratefully acknowledged.
REFERENCES


PUBLICATIONS

RESEARCH ARTICLES

1966

1967

1968

1969

1970

1971
1972


1974


25. A family of distributions with nearly constant ratios of expected values of order statistics. In *Proceedings of the All-India Seminar on Demography and Statistics*, pp. 70-75.

1975


1976


1977


1978


1979


1980


1981


1982

658 ET INTERVIEW


1983


1984

63. After Hovey: A note on taking account of the automatic death penalty jurors. Law and Human Behavior 8, 115–120.

1985


1986


1987


1988


1989


1990


**1991**


**1992**

1993

130. Several Bayesians: A review (with discussion). *TEST* 2 (1–2), 1–32.

1994


1995


1996


1997


1998


1999


2000


Forthcoming


BOOKS

1984


1986


1996

185. Editor and author of seven chapters. *Bayesian Methods and Ethics in a Clinical Trial Design*. New York: Wiley. (Author of the following chapters: 1. Introduction; 4. Statistical issues in the analysis of data gathered in the new designs (with T. Seidenfeld); 5. Introduction to the Verapamil/Nitroprusside study; 11. Operational history and procedural feasibility; 12. Results of the clinical trial I (with N. Sedransk); 13. Verapamil versus Nitroprusside: Results of the clinical trial II (with E.S. Heitmiller, N. Sedransk, & T.J. Blanck); 19. Epilogue.)

1999


**PUBLISHED COMMENTS**

1979


1980


1981


1982


1984


1985


1986

1987


1988


1990


1991


1992


1993


1994


1996

209. Response to Keith O’Rourke. *Controlled Clinical Trials* 17, 352.


1997


1998


1999


FULL BOOK REVIEWS

1973


1975


1976


1978

1984


1990


1991


**SHORT BOOK REVIEWS**

1971


1972

