Statistical Science
2005, Vol. 20, No. 2, 193–203
DOI 10.1214/088342304000000288

A Conversation with Larry Brown

Anirban DasGupta

Abstract. Lawrence D. Brown was born on December 16, 1940 in Los Angeles, California. He obtained his Ph.D. in mathematics from Cornell University in 1964. He has been on the faculty of the University of California, Berkeley, Cornell University, Rutgers University and, most recently, the Wharton School of the University of Pennsylvania, where he holds the Miers Busch Professorship of Statistics. Professor Brown was President of the Institute of Mathematical Statistics in 1992–1993, Coeditor of *The Annals of Statistics* for 1995–1997 and gave the prestigious Wald Memorial Lectures in 1985. In 1990, Professor Brown was elected to the U.S. National Academy of Sciences. In 1993, Purdue University awarded him an honorary D.Sc. degree in recognition of his distinguished achievements, and in 2002 he was named winner of the Wilks Memorial Award of the American Statistical Association.

Professor Brown is probably best known for his extensive work on the admissibility of estimators of one or more parameters. He has also published research on a broad variety of other topics including general decision theory, sequential analysis, properties of exponential families, foundations of statistical inference, conditional confidence, interval estimation and Edgeworth expansions, and bioequivalence. His current interests include functional nonparametrics, analysis of census data and queuing theory as involved in the analysis of call-center data.

Key words and phrases: Admissibility, ancillary, Bayes, bootstrap, census, conditional, consulting, decision theory, diffusion, exponential family, frequentist, infinitely divisible, information inequality, *M* estimate, minimaxity, nonparametric, prior, random walk, sequential.

This conversation originally took place in Professor Brown's office at the Wharton School on May 9, 2001 and was revised in consultation with him through October 2003.

CHILDHOOD AND SCHOOLING

DasGupta: Good afternoon, Larry. Many of us know at least some things about your research, but maybe not much about your childhood: where you grew up, your family, et cetera. Why don't you give us a glimpse into that? Did you grow up in Los Angeles?

Anirban DasGupta is Professor of Statistics, Department of Statistics, Purdue University, West Lafayette, Indiana 47907-2068, USA (e-mail: dasgupta@stat.purdue.edu).

Brown: Yes, West Los Angeles. That's where my family lived when I was born. That was the year before the U.S. entered the Second World War and then we moved to Beverly Hills shortly before I entered high school, when I was 12.

DasGupta: Did you go to a private school?

Brown: No. It wasn't common to go to private schools in Los Angeles in those days. I went to a public school; it was one of the best in the public school system.

DasGupta: Tell us about those days. Did you have many friends? Were you interested in any sports?

Brown: Actually, until high school age I was a pretty lonely kid. In elementary school I was not athletic. Later, my father used to take me running around the block in Beverly Hills to improve my athletic training, partly as a way to help me interact with other kids.

Gradually I did get involved in organized athletics. I played B-football and B-basketball in high school, and then when I went to college, I went to Cal Tech, I was involved in a number of athletic activities. I was a starter on the basketball team for three years in a row. There was a period of three or four weeks when I was among the leading scorers in the league, and of course then the other teams figured out they needed to keep watch on me. After that I didn't get to touch the ball much when we played the stronger teams.

DasGupta: Larry, please tell us a bit about your family. Were you the only child?

Brown: No. I have two younger brothers. One is a professor of comparative literature. The other is a partner in my mother's law firm in Los Angeles.

DasGupta: So your mother is an attorney?

Brown: Yes. She is a senior partner of a very successful, moderate sized firm in Los Angeles that specializes in entertainment law. (My mother died from cancer in September 2003 at the age of 87. She practiced law full time until about 9 months before that, and continued part time up to a few weeks before her death.)

DasGupta: Now, was your father an attorney too?

Brown: Yes. He was a founding partner of what is now the largest tax firm in Los Angeles, but he was an educator at heart, and after my mother began to earn enough money, he took an early retirement from his practice and went on to become a legal scholar and professor at USC. He always enjoyed teaching and research, and in addition to being my father he really is one of my intellectual heroes.

DasGupta: Larry, how did you get interested in mathematics? Did a teacher at school or someone in the family inspire you?

Brown: Well, actually my favorite teachers in high school were my English teachers, Mrs. Lehman and Miss Schmidt, but of course we also studied mathematics in high school and I enjoyed it and did well at it. That was before the days of advanced placement courses. We did no calculus in high school, but we did study a semester of solid geometry, stuff that is not done anymore. My paternal grandfather was an immigrant and very successful self-made man. He was interested in numbers and enjoyed telling us (repeatedly) all sorts of interesting properties of the number 9.

Much more important, as a child I spent a lot of time playing by myself. For many years my favorite game was a board game called "All Star Baseball." It had a spinner and a set of cards with a hole in the middle. The card is marked into segments of various sizes. You place the card into the spinner, spin the pointer and the segment it stops on tells what the batter does (strike out, fly out, home run, et cetera). These correspond to possible batting outcomes according to the records of various baseball players. If you play long enough and keep detailed records, as I did, you could see that the batting and slugging averages converge to those of the real-life players. It was a beautiful demonstration of the law of large numbers. You could say that was my first interest and exposure to statistics, or at least to data. I always liked data. The *Almanac* was one of my favorite books. On the other hand, there is a story my father always loved to tell. On parents' day my eighth grade math teacher told my father, "Your son is a nice kid, but one thing I can tell you for sure—he'll never be a mathematician."

COLLEGE AND UNIVERSITY LIFE

DasGupta: I understand you went to Cal Tech for college. Where else did you think of going?

Brown: My family, especially my mother and her mother, wanted me to go to the East; for her, that was the center of culture. I applied to Harvard, MIT and Princeton. I was accepted at all of those places. I actually didn't apply to Cal Tech 'til very late, almost near the deadline. There was a student visiting day. I went and was very impressed. I applied, and I really wasn't ready to go very far from home, so both intellectually and personally it was a good decision to choose Cal Tech.

DasGupta: What sort of things, mathematics and otherwise, did you do there?

Brown: Well, I majored in mathematics and physics. We used the calculus book by Thomas as a text at Cal Tech, with a supplement of hard problems later incorporated into Apostol (1961). Likewise in physics, the text was supplemented by a set of very challenging homework problems. They were written by a professor named Strong, so we referred to the homework problems as "Strong problems" (laughs). I remember going, among others, to Brad Efron who was one year ahead of me to get help in homework. I had a beautiful course in Mathematical Analysis taught by Professor Knowles, an applied mathematician, and what I learned there has been very useful to me.

DasGupta: What sort of math most appealed to you at Cal Tech?

Brown: I liked graph theory and combinatorics, analysis and statistics. I had a course in combinatorics from Marshall Hall, and it was through him that I got a summer job at RAND, although technically the job was with Bellman—the Bellman of dynamic programming. That summer I wrote my first paper. It solved a very specific problem using backward induction—what Bellman called dynamic programming. This later appeared as Brown (1965) after revisions that Jack Kiefer helped me with.

DasGupta: So how did you get interested in statistics?

Brown: Well, Cal Tech was in a quarter system, and the only statistics I had was a one-quarter course taught by an algebraist named Dilworth. There is a Dilworth theorem: It involves one of the equivalent forms of the axiom of choice. The subject almost immediately appealed to me. I suppose that my interest in statistics was rooted in a desire to use formal mathematics in a pragmatic way. So I could have gone into either statistics or something like computer science or applied math, but in those days you had to be almost superhuman to do computer science. You would write programs on cards, turn them in to be run and would get them back the next day, usually with an error message. If you wanted actual computer time, it had to be between 2:00 and 6:00 in the morning. The hours were worse than in the lab sciences and I just didn't get interested in it at all.

In the meantime, I started to do some calculations on what you would today call Bahadur efficiencies. I calculated exponential rates of convergence of the power to 1. I thought I was calculating asymptotic relative efficiencies, but actually I was not since I was working with fixed alternatives. This was my senior thesis.

DasGupta: And then you came to Cornell. You did move east, but why Cornell?

Brown: My parents convinced me I had to go east. Otherwise the most natural thing for me would have been to go to Berkeley or Stanford. Dilworth said that if you are going to go east, there is a very bright guy named Kiefer at Cornell. He said that actually there is also another very bright guy there: His name is Wolfowitz, but he is a kind of "curmudgeon," so you should stay away from him. You want to work with Kiefer. This was in 1961. Wolfowitz and I had very little interaction (partly because of divergent political opinions) until many years later.

DasGupta: What was the Ph.D. program like when you came? What courses did you take?

Brown: It was a Math Department. I took a year of algebra, a year of analysis, and a semester of topology and combinatorial topology. The analysis was measure theory and functional analysis. Then we had to pass an oral qualifying exam. I had an algebraist on my committee, and Spitzer and Kiefer. Though a probabilist, Spitzer qualified as the analyst on the committee, but on that exam he asked me a probability question about random walks. I tried and tried to answer it, and floundered and after one hour I was feeling totally insecure. Spitzer said, "So you don't know how to do this do you?" I said that I didn't. Spitzer replied, "Well, actually neither do I."

DasGupta: And you must have had some formal statistics courses also?

Brown: Some. Wolfowitz used to give a basic statistics course in those days, but I never took that course. I regret that. Kiefer told me that instead I should do a reading course using Lehmann's testing book (Lehmann, 1959), and I did that, which was a wonderful experience. Under Kiefer's guidance I worked out most of the problems and wrote out solutions for them.

Another regret in this regard is that I had almost no contact with the many other people that used to do statistics at Cornell outside of the Math Department. There were many excellent people of that sort then, just as there are today. I could have benefited from that.

DasGupta: Did you take any courses from Kiefer, Farrell or Spitzer?

Brown: I took optimal design from Kiefer, and Jack used to teach an undergraduate inference course and I sat in on this course. I don't think I took a course with Roger Farrell, but we did interact quite a lot. I also took a course on random walks from Spitzer and I had general probability from Harry Kesten. Many people warned me it was a dry course. I thought it was great. It was really precise and very organized.

DasGupta: And what about decision theory?

Brown: I had my major decision theory course from Peter Huber. Peter spent a year at Cornell in 1963–1964 and he gave a year long course on decision theory. Peter was a topologist by training and became a statistician. He had spent two years at Berkeley and had taken some courses from Le Cam. He had reworked much of Le Cam's notes and made them more accessible. I really enjoyed that course. Peter's course had a very major influence on me. It was also the basis for a set of lecture notes I later wrote on the topic. I am really glad I met Peter at Cornell. He was writing his robustness manuscript (Huber, 1964) at that time; the paper that showed that the minimax M function is quadratic within a bounded interval and linear outside. I think that paper is one of the best in all of statistics, even though there turned out to be some crucial limitations to the formal theory. It opened up a whole new way of thinking.

DasGupta: Tell us a little about your thesis work. Did you select the problem or did Kiefer do that for you?

Brown: He suggested the area, but asked me to select my problem. With me he acted in that perceptive but nondirective way throughout our entire relationship. He told me that Stein was doing some really interesting work on admissibility and I should take a look at that. Statistics was lovely in those days; I essentially had to read five papers to know all the necessary background. Three papers by Charles (Stein, 1955, 1959; James and Stein, 1961), a paper by Hodges and Lehmann (1951), and a paper by Blackwell (1951) on the translation parameter problem for discrete cases. I quickly realized that Blackwell's argument should work in general, in principle anyway, to prove admissibility in one dimension, and the Taylor series argument Charles gave for proving inadmissibility was the right method to give a general proof of inadmissibility in three or more dimensions. Then of course I had to work out the best regularity conditions to get to the most general class of loss functions and distributions. My overall goal was to show that what Stein did was not a particular feature of squared error loss or normal distributions, and that indeed there was a very general dichotomy, with something happening in one and two dimensions and the contrary in three or more dimensions.

DasGupta: This is the 1966 paper (Brown, 1966)? It unified the location parameter problem, all in one paper.

Brown: There was a reason I put all of it together in one paper. I was attempting to show the general structure of the dichotomy in all location parameter problems, but if I were writing it today I would break it into at least two papers so as to make it easier to follow.

YEARS AT LONDON AND BERKELEY

DasGupta: Where did you go after graduation?

Brown: Kiefer suggested that I should spend some time with David Cox in London and arranged for me to do so. At that time David was working with P. A. W. Lewis on their book (Cox and Lewis, 1966), but David and I talked often about many topics. I also attended a graduate course given in the evening on applied multivariate analysis. This was really the first "applied statistics" course I ever took, although applications were frequently mentioned in conversations with Kiefer, Cox and others. (As a sign of how times have changed, I note that several weeks of this course were spent describing how to efficiently organize calculations on a mechanical calculator and drilling the students to develop skill in this!) I had a friendly professional relationship then with David and have had ever since, but perhaps I didn't really get all I could have out of that year in London. I also met Dennis Lindley and Mervyn Stone there and enjoyed talking with them about the Bayesian implications of the James–Stein estimation results.

DasGupta: And then you came back to America?

Brown: Yes, to Berkeley, as a regular Assistant Professor. One day, when I was still at Cornell, Jack came and asked me if I would like to go to Berkeley. He said he had talked to them and they had a job for me. I said of course. So I never had to send applications or send my vita. In some ways life was much simpler in that era.

DasGupta: Who was the Chair then?

Brown: Scheffé. He was very kind as a Chair, and in his somewhat courtly European way, Neyman was as supportive and sweet as could be.

DasGupta: What about Lehmann and Peter Bickel?

Brown: I had a very nice relationship with Erich, though I didn't get to see him much. He had odd hours. He would come in at 3:00 in the morning and stay 'til 8:00, sometimes 9:00. I didn't see Peter that year because he was on leave. Actually I remember Peter from Cal Tech, where he was an undergraduate his (and my) freshman year, although he says he doesn't remember seeing me there. Among the senior statisticians there, I spent the most time talking with Lucien Le Cam, who was very open, and I found it often more helpful to talk with him than to try to read his papers.

THE VIETNAM WAR AND THE 1971 PAPER

DasGupta: So you spent only one year at Berkeley, is that right?

Brown: Yes. Those were the days of the Vietnam War and I got a draft notice. The office at Berkeley told me there was nothing to worry about. There was a national guideline according to which people in technical subjects such as mathematics, physics, engineering never get drafted, but my local draft board apparently thought statistics was not a technical subject and denied my appeal. It became clear that the obvious solution to avoid being drafted was to get a job in a math department. I called Jack and said I needed a job in a math department. He wrote back saying I was hired. So that was the reason I had to leave Berkeley. Otherwise, I was very happy there and I wanted to stay.

DasGupta: Larry, the 1971 paper (Brown, 1971) on admissibility and recurrence of diffusions came out after you returned to Cornell. Of all the influential papers you have written, that was perhaps the most influential. It made such an impact on the subsequent work on decision theory, for example, the early work of Jim Berger (1976a–c) and the more recent work of Joe Eaton (1992, 2001). Did your return to Cornell have anything to do with that paper?

Brown: Not very much. While at Berkeley (if not before) I had realized that the admissibility question could be thought of as a calculus of variations problem. I also had realized how this creates a connection with the issue of recurrence or transience of suitable diffusions. Much of the paper was thus laid out at that time. There was, however, a key technical gap relating to solutions of differential equations that needed to be filled in order to complete the proof at (nearly) the level of generality I was seeking. That gap was filled in a conversation with an applied mathematician, Jim Bramble. He was on the faculty at Cornell. After I described my difficulty in terms I thought he could relate to, he said he had an unpublished theorem in his file cabinet that he thought would be helpful. It was just what I needed. He hadn't published it because he didn't see a use for it!

DasGupta: In the 1971 paper, through the random walk connection, you also conjectured the result that no genuine Bayes minimax procedures can exist in less than five dimensions.

Brown: That's right. A few years later, Bill Strawderman gave some hierarchical priors in five or more dimensions that lead to minimax estimates (Strawderman, 1971). That was very nice. From my paper I knew the general shape of the prior that was needed, but I didn't realize at all that hierarchical priors would be the effective way to construct them. Opening this connection with the world of hierarchical priors has been an important innovation.

DasGupta: A few years ago, you wrote a review article on minimaxity (Brown, 1994). It gave a very thoughtful exposition on the place of minimaxity in statistical inference. Are you writing other similar expository articles?

Brown: Thank you for the compliment. More recently I wrote a review article for *JASA* (Brown, 2000) and a review entry for the *Encyclopedia of Social Sciences*, but I'd still recommend the minimaxity article as most reflective of my own point of view.

FIVE YEARS AT RUTGERS

DasGupta: Now, Larry, after the 1971 paper, your interests diversified and you wrote several papers on sequential analysis and on conditional inference. By that time you had moved to Rutgers.

Brown: At Cornell at that time I didn't have any connections outside the Math Department and so I wanted to be in a statistics department. I had already started correspondence with Arthur (Cohen). I was very happy to go to Rutgers.

DasGupta: Before going on I want to pull back a little and talk about your view of conditional confidence and the Bayes–frequentist controversy. Did your work on this begin with the 1967 paper (Brown, 1967)?

Brown: Well (laughs), at the time I would have described it as only a small result on the t test, but going back to it later did help me explore the area more fully.

DasGupta: Larry, would you like to briefly talk about the issues in conditional inference?

Brown: The issues are subtle. It does seem that if there is a clear choice of the ancillary statistic, then one should condition on it. Who doesn't? Take Cox's example. Suppose we flip a coin either twice or 100 times with a 50–50 probability. If we know the coin was actually flipped 100 times, our report on the quality of inference has to be conditional on that knowledge. The statement—say about the coverage of the confidence interval—has got to be conditionally correct. On the other hand, you could build your procedure by taking into account the possible values the ancillary can assume.

DasGupta: If I understand you correctly you seem to be making a distinction between the procedure [estimate or confidence interval] and its assessment. Doesn't Bayesian analysis do this automatically without having to make explicit distinctions?

Brown: It does in a way, but only if the prior is the correct one or at least is robust with respect to a suitable range of possibilities, and that's a big IF.

DasGupta: All right. Do you see any sort of real compromise at a methodological level, not just an academic level?

Brown: I think there is already a synthesis in many problems. Take hierarchical models as used with spatial statistics and elsewhere, for example. You could look at the procedures as coming from a big random effects model with parameters estimated by something like maximum likelihood or you could use a hierarchical Bayes model with a diffuse prior on the final hyperparameter. The models are philosophically different, but the resulting procedures are very similar. My feeling is that the point where I am a frequentist and somebody else is a Bayesian is that I think eventually you have to study the distribution of the proposed procedures. You don't necessarily have to look at a risk function coming from a single loss function, though that may be helpful, but you should have lingering doubts about recommending a procedure if you don't know enough about its distribution.

DasGupta: Hasn't the relationship between Bayesian and frequentist philosophies been a focus of your thought for many years?

Brown: Yes. Although it's a somewhat philosophical issue, it also has important practical consequences. Embedded in my 1971 paper is the fact that all sensible admissible procedures are generalized Bayes. Also, some generalized priors lead to admissible procedures and others don't, and one can tell from the prior (and the dimension of the problem) which is which. Thus, even in the late 1960s it seemed to me that a sensible way to generate procedures with desirable properties is to choose a prior among the class of those that lead to admissible procedures. Several considerations seemed relevant. For example, one could choose among this class of priors those which give heavier weight to parameters that seem a priori likely. Choosing among only those priors that lead to minimax estimators is another possible principle. This guarantees desirable robustness for the choice of prior. It turns out, however, that exact minimaxity is somewhat too restrictive for many practical purposes, but near minimaxity or ε minimaxity (for not too small an ε) works well. (The 1971 paper is, of course, only about the problem of estimating a normal mean, but the heuristic principle seems to carry over fairly well to a wide variety of problems.)

Jim [Berger] then carried these ideas well beyond this bare outline and modified them to create a theory of robust Bayesian procedures. So far as I understand Jim's perspective (and we have talked about it quite a bit over the years), he thinks of this as clearly Bayesian, whereas I think of it as using Bayesian methodology to generate satisfactory frequentist answers. Thus he gives a reality to the resulting posterior distributions, whereas my perspective is that they are only a step on the route to a procedure, and the litmus test of that procedure is how well it performs in a frequentist sense. On a practical level, the difference between these points of view often matters very little and may even be hard to discern, but there does seem to be a difference.

RETURNING TO CORNELL AND DEATH OF JACK KIEFER

DasGupta: In 1977 you returned to Cornell. Please tell us a little about that move.

Brown: That was basically for family reasons. Personally, I was very happy at Rutgers and had good friendships and collaborations. I enjoyed it there, but my family didn't want to live in New Jersey. Again Jack brought me to Cornell. You know, he had been my guardian angel all those years. I remember the important phone call; it was certainly a curious one. I called him and said that I had news for him. He responded by saying that he had news for me. I then reported my news first, and said that I wanted to leave Rutgers and would like to come back to Cornell, if it was at all possible. He then reported that he was planning to leave Cornell and go to Berkeley.

DasGupta: And then, most unfortunately, he passed away suddenly in 1982. How did that affect you?

Brown: It was a shock—very upsetting. In many ways, he was a father figure to me. Even after I returned to Cornell and he left for Berkeley, I saw him several times when he returned to visit his family, and I had chatted with him at the 1982 Purdue Symposium not long before his death.

DasGupta: You must have also felt that the profession lost a great scholar.

Brown: Oh absolutely; tremendous, tremendous range and depth of work. Many of his papers I hadn't read until after his death. Then I read them when I helped edit a volume of his collected works (Brown et al., 1985, 1986), and solicited discussions from many people. Here, in Brown et al. (1986) you see, is my commentary on Jack's proof of what is basically the Hunt–Stein theorem (Kiefer, 1957).

DasGupta: You mean the result that the best invariant procedure with respect to solvable groups has the minimax property?

Brown: That's right. Jack gave a very creative proof unlike anyone else's. The proof in my commentary was originally Lucien Le Cam's. I actually first saw it from Peter Huber. Later I came to know that Peter learned it from Le Cam.

THE EXPONENTIAL FAMILY MONOGRAPH

DasGupta: Larry, a few years after you returned to Cornell, you wrote a monograph on the exponential family. It is very well cited. Please tell us how it came about.

Brown: Actually, I hadn't originally planned on writing a monograph (Brown, 1986) on the exponential family. I had earlier written a set of notes on measure theoretic and topological aspects of decision theory that had some admissibility and minimaxity implications, but the treatment was much too formal to make clear the relevance of decision theory to practical statistics. So I thought that to make it somewhat accessible, I should add a chapter on applications and that one could do the applications in a unified way in the exponential family. So in that monograph, Chapter 4 has the decision theoretic applications. When I now teach a course on exponential families, I spend less time on those applications and concentrate more on maximum likelihood, the E–M algorithm, higher order asymptotics and Efron's curvature theory. Many such results can be more cleanly and directly proved in that setting, and if you want to go outside of the exponential family you need to add on special extra conditions.

INTEREST IN NONPARAMETRICS

DasGupta: Did your work on the information inequality (Brown and Gajek, 1990) precipitate your later interest in nonparametric function estimation? You gave a proof of the central limit theorem using the information inequality (Brown, 1982). You have been working on nonparametrics for more than ten years now.

Brown: Well, there is a connection. Actually my nonparametric work started with a paper with Roger Farrell (Brown and Farrell, 1990). Roger had done some very important work on density estimation. Parzen (1962) and Rosenblatt (1956) of course gave kernel estimates that do not have $\sqrt{n}$ rate of convergence, but Roger actually showed that that's the best you can do (Farrell, 1972). At the time I was very puzzled by that.

Roger's approach involved concentrating on a parametric subproblem. I knew how to obtain bounds on minimax risks in parametric problems by using the Cramér–Rao inequality. That was in my paper with Gajek (whom I've never met), but was really already implicit in work that Hodges and Lehmann did (Hodges and Lehmann, 1951). Once I saw this connection, my interest in doing something in nonparametrics developed. I later wrote another paper with Mark too (Brown and Low, 1996).

DasGupta: Please give us your thoughts on some of the developments in nonparametrics.

Brown: Well, the work by David Donoho and his coauthors (Donoho and Liu, 1991; Donoho, Liu and MacGibbon, 1990) was pivotal for me. Their work opened up a whole new horizon and showed you can say something about not just the optimal rates, but the constants. Of course, the first result with constants is Pinsker's and then Efromovich and Pinsker (1996), but that doesn't easily generalize until one sees how it fits into Donoho's framework.

DasGupta: Do you think that there are still unanswered important questions in the area of nonparametric function estimation?

Brown: Well, effective adaptivity and the use of real priors are some. (Thank you for the question, because it gives me a chance to mention Linda [my wife], since she has done some nice work related to construction of priors here [Zhao, 2000]. But the real reason for mentioning her is to thank her for all the brightness she has brought to my life.) Practical procedures in higher dimensions and correct, effective confidence statements are others. There are evolving connections to general pattern recognition and classification problems. We are learning a lot, but in many cases someone has to take the theoretical estimates, try them out and (perhaps) modify them to make them practical. The methodological part needs to make progress. The last decade was good for theoretical nonparametrics.

RETURN TO PARAMETRIC INFERENCE

DasGupta: But, Larry, nonparametrics is not the only thing you are doing these days. A particular series of recent papers of yours that, for selfish reasons, I too am fond of . . . .

Brown: I really like that sequence of papers with you and Tony Cai on the standard confidence interval for binomial proportions (Brown, Cai and DasGupta, 2001, 2002) and, of course, we've now extended the work to show the exact similarity of the phenomena in a certain subclass of the exponential family (Brown, Cai and DasGupta, 2003). The message is that the standard interval does not, at all, perform to the level that it is claimed or believed to, and the problem is not only a problem of small n or p near 0 or 1. It happens for rather large n and for p near $1/2$. Most people (including me) did not previously understand that. They did not understand the extent of the shortcoming in such a fundamental problem.

DasGupta: How do you feel about the apparently strong opposing views that only confidence intervals matter and that point estimates are irrelevant?

Brown: I am not sure that the views are strong as you suggest. Clearly confidence intervals are widely used. Something like a confidence interval seems obviously important to me. You have to give a measure of error.

DasGupta: Is it too formalistic to treat it as a decision problem? I mean in the point estimation domain most are happy with a squared error or absolute error loss, but in the set estimation problem, we cannot agree on a loss function.

Brown: That's right. Most statisticians will agree that coverage and a measure of size such as length have to be balanced, but there is no uniquely appealing way to average the two. This difficulty does not take away the fact that an error statement must accompany an estimate. Call it a confidence interval if you want.

APPLICATIONS, CONSULTING AND THE 2000 CENSUS

DasGupta: Larry, a lot of people don't know that since coming to Wharton, you have been doing a variety of applied work. In particular, please tell us a little about your involvement with the 2000 national census. How did that happen?

Brown: Perhaps David Freedman indirectly talked me into it. David had been writing a lot about census adjustments and we occasionally talked about the issues. I was always interested in data—looking at it and seeing what it says—but previously I had always been so busy with other types of activities that I had to tell people, either directly or sometimes by my level of inactivity (laughs), that maybe they should call up someone else. One day I got a phone call from the Chief of Staff of the Senate Internal Affairs subcommittee who wanted to know if I would be able to testify before the Senate about my view of the plans for the 2000 Census. It came at a good time. My term as Coeditor of *The Annals* was nearing its end and I was looking for new activities. The invitation came with enough lead time, so I read up all the material I could and testified before the Senate.

DasGupta: Is that project still going on?

Brown: A paper with David Freedman and others is out now (Brown et al., 1999). It gets cited a lot. One thing that really embarrasses me is that it gets cited as Brown et al., but David and the others deserve most of the credit for that publication. About the Census itself, I am on an NRC (National Research Council) oversight committee that is due to soon issue a final report on how well the census was done. This committee activity has involved a greater time investment than any other I have experienced. I also have a graduate student who has completed a thesis on some issues in the Census (Zhao, 2003). I think my interest in this and other data-driven projects is likely to continue.

DasGupta: Did you find these application oriented projects useful for finding good theoretical problems?

Brown: Definitely in some ways. I am especially involved now in projects that have evolved from work analyzing telephone call-center data. This has led me to think more deeply than I had before about both certain questions in nonparametric functional analysis and in prediction for longitudinal models. The questions of interest here are very explicitly derived from aspects of the data and its analysis. I think it's a nice story how this project arose and how work on some of the detailed questions has evolved. First, I got started working on this project when a former student, Avi Mandelbaum, visited Wharton and got me interested in trying to see how well actual call-center data matched predictions for it derived from queuing theory. One continuation of this project has been an attempt to build autoregressive infinitely divisible stochastic processes to help accurately and aesthetically model telephone call arrivals at the center. This work is continuing, but much of it already appears in a recent thesis (Li, 2003). It turns out that a key idea is due to P. A. W. Lewis. A surprise coming out of my past! Then, in attacking this problem we were able to use some techniques developed for a slightly different (but related) purpose in a paper I had written with Yossi Rinott in the 1980s (Brown and Rinott, 1988). Another surprise from the past! Finally, last year I was visiting at Duke and chatting at dinner with Robert Wolpert, a friend and collaborator who had been a student along with Jim Berger in my decision theory seminar at Cornell. Even after several glasses of wine (or, maybe because of them!), Robert was able to come up with a reinterpretation of the issues here that has led to a different and useful approach to the problem.

DasGupta: Are you involved in any other projects as part of the National Academy?

Brown: Yes, I am involved in several committee projects there. Perhaps I spend too much time on this, but I do enjoy it and seem to do reasonably well at it—at least much of the time—and it gives me a chance to provide some service to the profession and the country.

TEACHING AND GRADUATE PROGRAMS

DasGupta: Larry, obviously another integral part of our profession is teaching. What are some of your favorite topics that you like to teach?

Brown: Teaching statistics at any level is fun as long as the students are reasonably intelligent and conscientious. I suppose that guiding graduate students can also be referred to as teaching, but I usually think of it as a different type of activity. I find it extremely stimulating intellectually, in terms of formulating new ideas and approaches, and overall I find the interaction with students to be very rewarding on a personal level as well. I'm pleased to see how the careers of many of my former students have flourished and I continually hope that I have contributed in a positive way in this by helping them develop their skills and perspectives.

DasGupta: Let's talk a bit about the evolution of statistics graduate programs and the subject itself. The importance of mathematical statistics seems to be declining; in how we train graduate students, in journal publications, in awarding grants. Is mathematics becoming irrelevant in statistics? Please let us have your comments on that.

Brown: As regards the importance of mathematics, statistics remains and will remain a largely mathematical subject. That does not mean that the type of mathematics or the reasons for knowing and doing mathematics will remain the same. For example, think of the 1950s. We spent a lot of time worrying about measurability questions. You can now usually ignore them; that's because someone looked at them, and we now know that under widely satisfied conditions, you don't have to worry about them. So what type of mathematics is important is an evolutionary process. A lot of mathematics is still obviously going on in statistics: in imaging problems, climate modeling, MCMC schemes, meteorological sciences, genetics. Even though there is obviously less formal theorem-proving in statistics today than there was 40 years ago, the importance of mathematics persists. It's just a different sort of mathematics.

DasGupta: What about data? Should graduate students in statistics do statistical consulting as part of the curriculum?

Brown: I would not say that they must do consulting per se, but they have to see and analyze data. I cannot imagine what kind of a statistician one would be if he or she has no background with data or interest in it, and now there is no reason that students should not see and realistically analyze data. It is so easily accessible. When we were students, we would get toy data examples we could conveniently work with. We then had to try to imagine what the real thing would be like. Now it's practical for students to examine real data extensively.

DasGupta: What is the order in which students should see theory and data? Theory followed by data, the other way around or simultaneously?

Brown: I've struggled with that question, as have many of us. I hope there's a good answer to it, but I'm not sure what it is. I often feel that students need to see data and the corresponding theory simultaneously, but it seems hard to design a successful course that way.

THE IMPORTANT DEVELOPMENTS IN STATISTICS

DasGupta: Larry, in the last 25 years a number of fundamental problems in mathematics were solved: the four color problem, classification of simple groups, the Bieberbach conjecture, Fermat's last theorem. In statistics, there is probably no such thing as a universally agreed fundamental problem, but still what would you consider to be some of the most important and influential developments in statistics in the last 25 or 30 years?

Brown: The bootstrap (Efron, 1979) has had an obvious impact and is clearly useful in many types of contexts. M estimates (Huber, 1964) were an important development, less for what they were designed to be, as robust estimates, but certainly for organizing asymptotic theory and for showing that you could think of robustness in a mathematical way. MCMC sampling (Gelfand and Smith, 1990) grew out of a Bayesian motivation, but has been useful in other ways as well. It remains to be seen how satisfactorily some of the issues of convergence and convenience can be settled. Shrinkage estimation has led to many useful developments outside of the narrow focus of finding dominating estimates. Stein's (1981) unbiased estimate of risk has had a lasting impact on statistics. The area of nonparametric function estimation (nonparametric regression and density estimation) has grown from a baby to a healthy adolescent in the last 15 years and there have been a number of serious developments in how we analyze data, such as in classification problems (e.g., Breiman, 1998). Advances are almost inevitable as long as you have creative thinkers and important problems as we still do these days.

FUTURE OF STATISTICS

DasGupta: In closing, how do you feel about the future of statistics? Will it continue in the foreseeable future to be useful to the human enterprise?

Brown: Oh yes! Statistics has been and will remain useful, if anything, in more contexts than ever before. I do see a red flag on the horizon within the discipline. There seems to be a danger of fragmentation.

Branches of statistics, like biostatistics, could become essentially independent subjects without a link through the fundamental core to other fragments of the field. It has happened in other sciences, physics and chemistry. Fortunately in those subjects, over time, there have also been interesting recombinations, like biochemistry and biophysics, for example. I hope that statistics becomes useful in more and more areas with enough commonality that we still exist as a discipline with a unifying core.

DasGupta: Thank you, Larry, for giving us this chance to have a conversation with you. I wish you a very long, healthy and productive life. Thank you also for the scientific inspiration I have always received from you. I wish you the best.

Brown: Well, thank you for all the effort you made and for your kind words. I am very flattered.

REFERENCES

APOSTOL, T. M. (1961). *Calculus*. Blaisdell, New York.

BERGER, J. (1976a). Tail minimaxity in location vector problems and its applications. *Ann. Statist.* **4** 33–50.

BERGER, J. (1976b). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. *Ann. Statist.* **4** 302–333.

BERGER, J. (1976c). Admissibility results for generalized Bayes estimators of coordinates of a location vector. *Ann. Statist.* **4** 334–356.

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Statist.* **22** 393–399.

BREIMAN, L. (1998). Arcing classifiers (with discussion). *Ann. Statist.* **26** 801–849.

BROWN, L. (1965). Optimal policies for a sequential decision process. *J. Soc. Indust. Appl. Math.* **13** 37–46.

BROWN, L. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

BROWN, L. (1967). The conditional level of Student's t test. *Ann. Math. Statist.* **38** 1068–1071.

BROWN, L. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

BROWN, L. (1982). A proof of the central limit theorem motivated by the Cramér–Rao inequality. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 141–148. North-Holland, Amsterdam.

BROWN, L. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. IMS, Hayward, CA.

BROWN, L. (1994). Minimaxity, more or less. In *Statistical Decision Theory and Related Topics V* (J. Berger and S. S. Gupta, eds.) 1–18. Springer, New York.

BROWN, L. (2000). An essay on statistical decision theory. *J. Amer. Statist. Assoc.* **95** 1277–1281.

BROWN, L., CAI, T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion (with discussion). *Statist. Sci.* **16** 101–133.

BROWN, L., CAI, T. and DASGUPTA, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30** 160–201.

BROWN, L., CAI, T. and DASGUPTA, A. (2003). Interval estimation in exponential families. *Statist. Sinica* **13** 19–49.

BROWN, L. ET AL. (1999). Statistical controversies in Census 2000. *Jurimetrics Journal* **39** 347–375.

BROWN, L. and FARRELL, R. (1990). A lower bound for the risk in estimating the value of a probability density. *J. Amer. Statist. Assoc.* **85** 1147–1153.

BROWN, L. and GAJEK, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** 1578–1594.

BROWN, L. and LOW, M. (1996). A constrained risk inequality with applications to nonparametric function estimation. *Ann. Statist.* **24** 2524–2535.

BROWN, L., OLKIN, I., SACKS, J. and WYNN, H., EDS. (1985). *Jack Carl Kiefer Collected Papers* **1**–**3**. Springer, New York.

BROWN, L., OLKIN, I., SACKS, J. and WYNN, H., EDS. (1986). *Jack Carl Kiefer Collected Papers* **Suppl**. Springer, New York.

BROWN, L. and RINOTT, Y. (1988). Inequalities for multivariate infinitely divisible processes. *Ann. Probab.* **16** 642–657.

COX, D. and LEWIS, P. A. W. (1966). *Statistical Analysis of Series of Events*. Methuen, London.

DONOHO, D. and LIU, R. (1991). Geometrizing rates of convergence, II, III. *Ann. Statist.* **19** 633–667, 668–701.

DONOHO, D., LIU, R. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437.

EATON, M. L. (1992). A statistical diptych: Admissible inferences—recurrence of symmetric Markov chains. *Ann. Statist.* **20** 1147–1179.

EATON, M. L. (2001). Markov chain conditions for admissibility in estimation problems with quadratic loss. In *State of the Art in Probability and Statistics. Festschrift for Willem R. van Zwet* (M. de Gunst, C. Klaassen and A. van der Vaart, eds.) 223–243. IMS, Beachwood, OH.

EFROMOVICH, S. and PINSKER, M. (1996). Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statist. Sinica* **6** 925–942.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

FARRELL, R. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

GELFAND, A. and SMITH, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

HODGES, J. L., JR. and LEHMANN, E. L. (1951). Some applications of the Cramér–Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 13–22. Univ. California Press, Berkeley.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

KIEFER, J. (1957). Invariance, minimax sequential estimation and continuous times processes. *Ann. Math. Statist.* **28** 573–601.

LEHMANN, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.

LI, X. (2003). Infinitely divisible time series models. Ph.D. dissertation, Univ. Pennsylvania, Philadelphia.

PARZEN, E. (1962). On the estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

ROSENBLATT, M. (1956). Some remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

STEIN, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Math. Statist.* **30** 970–979.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.

ZHAO, L. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.* **28** 532–552.

ZHAO, Z. (2003). Analysis of dual system estimation in the 2000 Decennial Census. Ph.D. dissertation, Univ. Pennsylvania, Philadelphia.