A Conversation with Larry Brown

by

Anirban DasGupta
Purdue University

Technical Report #01-09

(Invited Article for Statistical Science)

Department of Statistics
Purdue University
West Lafayette, IN USA

May 2001

★

A Conversation with Larry Brown

by

Anirban DasGupta
Purdue University

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 1963, and has been in the faculty of University of California, Berkeley, Cornell University, Rutgers University, and most recently the Wharton School at the University of Pennsylvania, where he holds the Meier-Busch Professorship of Statistics. Professor Brown is widely regarded as one of the most eminent statisticians in the world, and is especially known for his wide and deep contributions in decision theory and mathematical statistics. He was the President of the Institute of Mathematical Statistics for 1992-93, the Editor of the Annals of Statistics for 1995-98, and gave the prestigious Wald lecture in 1985. In 1990, Professor Brown was elected to the US National Academy of Science, and is the Secretary of its Applied Mathematics section. He is serving on a National Research Council Panel on the quality of the 2000 national census and is a consultant for a national survey on telephone calls. His current research interests include functional nonparametrics, Edgeworth expansions, foundations of statistical inference, bioequivalence, and analysis of financila data. This conversation took place at Professor Brown's office at the Wharton school on May 9, 2001.

A Conversation with Larry Brown

by

Anirban DasGupta

Purdue University

Childhood and schooling

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

Brown: Yes, west Los Angeles. That's where my family lived when I was born. That was the year before the US 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 I was sort of a lonely kid. I was, if anything, anti-athletics at the beginning. I remember my father used to take me running around the blocks in Beverly Hills to improve my athletic skills, and to help me meet other kids. Gradually, I did get involved in 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 the leading scorer in the league. And of course then the other teams figured out how to keep the leading scorer in check and I don't think I touched the ball much when we played the stronger teams.

DasGupta: Do you still keep track of sports?

Brown: I keep track, but the real active interest in sports pretty much ended after my college days.

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.

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 make enough money, he took an early retirement from his practice and went on to become a professor at USC. He always enjoyed teaching and research, and 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, my best teachers in school were my English teachers, Mrs. Lehman and Miss Schmidt. But we did a lot of mathematics at school, and I did okay. That was before the days of advanced placement courses. We did no calculus in high school, but a semester of solid geometry, stuff that is not done anymore. In the family, I had a grandfather who was interested in math. I think he knew about most interesting properties of the number 9. Take a two digit number, rearrange the digits, subtract the smaller from the larger, and then the difference is divisible by 9. He inspired me quite a bit. And as I said, I was a lonely child. I spent a lot of time with myself. I remember one particular board game. It had a spinner and a card with a hole in the middle. You take the card and mark it into segments. These correspond to possible batting outcomes according to the records of various baseball players. If you play long enough, as I did, and keep detailed records, 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 had an interest in data. There is a story my father always loved to tell. On parent's 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 till very late, almost near the deadline. There was a student visiting day; I went and was very impressed. I applied. And really I wasn't ready to go very far from home; so it was a good decision to choose Cal Tech. I remember one story about MIT. I had an interview for admission there. The person conducting the interview said that I would certainly get in but MIT is sort of an engineering place and he thought I was too smart and had a scientist's mind and should go somewhere else. Anyway, it was nice being close to home. I would find an excuse to occasionally drop by for dinner or something and it worked out well.

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

Brown: Well, we certainly did mathematics and physics. I didn't do calculus in high school. We used Thomas as the text at Cal Tech, with a supplement of hard problems from Apostol (Apostol(1961)) - he was at Cal Tech. Likewise in Physics the text was supplemented by a set of very challenging homework problems. The professor was someone by the name Strong - so we referred to the homework problems as "strong problems "(laughs). I remember going to Brad Efron who was one year ahead of me to get help in homework. He was one of these upper class undergraduates who had an open door policy; anyone could go to them and ask for help.

DasGupta: I see; so you knew Brad Efron at Cal Tech? Did you interact with him in any other way?

Brown: Brad was very prominent and active in many ways. He was president of the student body and was many other things. I don't recall if we had a class together, but probably we did.

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

Brown: I liked graph theory and combinatorics. I had a course in algebra and one in combinatorics from Peter 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 that solved a very specific problem using backward induction-what Bellman called dynamic programming.

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 10 week course taught by an algebraist named Dilworth. There is a Dilworth theorem - it involves one of the equivalent forms of the axiom of choice. My interest in statistics grew out of my desire to use formal mathematics in a pragmatic way. So I could go to 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. It was worse than the lab sciences. 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. I was working with fixed alternatives. Dilworth read the work. He was an algebraist, so he probably didn't go through all the details. All he did was to write 'excellent work' on it. I think parts of what I did were not correct, but I was glad to do something original.

DasGupta: And this was at Cal Tech, right?

Brown: Yes, this was my senior thesis.

DasGupta: That's very interesting. You were able to create important problems on your own at that age with so little formal training in statistics.

Brown: Years later, I saw someone publish those results of mine. And he did all of it correctly.

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

Brown: My parents convinced me I had to go east. 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 at Cornell by the name Kiefer. 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 another oral qualifying. I had an algebraist on my committee, and Spitzer and Kiefer. Spitzer qualified as the analyst on the committee, but he asked me mostly probability questions.

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

Brown: Oh yes. 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 could do a reading course on Lehmann's testing book (Lehmann (1959)). And I did that which was a wonderful experience. I worked out most of the problems and wrote out solutions for them.

DasGupta: It's highly impressive that you were working out most of Lehmann's problems as a graduate student.

Brown: But I couldn't do all of them. So Kiefer used to meet me about an hour, sometimes more, every week. He was really excellent that way. One thing I do regret is that I had almost no contact with all the 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 benefitted from that.

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

Brown: Roger Farrell, I don't think so. I used to talk to him, but I didn't have a formal class from him. I had optimal design from Kiefer. And Jack used to teach an undergraduate inference course and I was his TA in that course for one semester. I sat in his course that semester. When Jack died, I really wanted his lecture notes for that course converted to a book. But I felt I really didn't have the time at the time to do so. I had to impose on Gary Lorden who did a very nice job of writing those notes into a book.

DasGupta: This is the Springer text (Kiefer (1987)) with a lot of rare and absolutely charming examples?

Brown: Yes; it is also the kind of book that would appeal to you.

DasGupta: It does; I love that book. What about Spitzer? Did you have a course from him?

Brown: Yes, on random walks. I had my 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.

I remember two stories about Spitzer. On my first day at Cornell, I went to the math department to check out who are the faculty and what courses are being offered. I was reading the bulletin board and someone came and stood next to me and asked if it was my first day. I said it was and he said it was also his first day at Cornell. I later learned that was Spitzer. He essentially replaced Marc Kac. Then on my qualifier, Spitzer asked me a random walk question. I tried and tried, and floundered, and after one hour I was feeling totally insecure. And Spitzer said "so you don't know how to do this, do you?" I said that I didn't. And Spitzer replied, "well, actually, neither do I."

DasGupta: And what about decision theory?

Brown: I had my major decision theory course from Peter Huber. Peter spent one year at Cornell, 1963-64 I think. 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, and later when I wrote my notes, Peter's course had a very major influence on me. I am really glad I met Peter at Cornell. He was writing his robustness manuscript (Huber (1964)) at that time. You know, 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 most influential and one of the best in all of statistics. It really opened up a whole new way of thinking, although there are some shortcomings.

DasGupta: How did you reach the decision that you were going to work with Kiefer?

Brown: Well, Dilworth had told me to do so. And it was clear to me that that's what I should do.

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

Brown: He suggested the area, but asked me to select my problem. He acted that way throughout my entire career. Jack wanted me to be my own man but I talked with him a lot. You know, the most conspicuous example of this is my 1978 paper on conditional inference. In the 1978 (Brown (1978)) paper, I had said that you have to go beyond what Jack did, but it was entirely influenced by him and his previous work. I suggested to Jack that we should write it up together. He said ' no, you write it on your own.' I am pretty sure that if it was anyone else, he would have agreed to make it a joint paper.

Anyway, coming back to my thesis, Kiefer 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 three papers to know all the necessary background. Two papers of Charles (Stein (1955), James and Stein (1961)) and a paper of Blackwell (Blackwell (1951)) in the translation parameter problem for discrete cases. I quickly realized that Blackwell's argument should work in general, in principle anyway. And the Taylor series argument Charles gave for proving inadmissibility was also the right one to generalize for the other part of the work. And then of course I had to work out the best regularity conditions to get the most general loss functions and the distributions. So I could show that what Stein did was not a particular feature of squared error loss or normal distributions and that indeed there was this general dichotomy there, with something happening in one and two dimensions and the contrary in three or more dimensions.

DasGupta: This is the 1966 paper? All of us regarded that paper (Brown (1966)) as a monumental one.

Brown: Well, monumental it was in length (laughs). There was a reason I put all of it together in one paper. In an overarching way, I was attempting to show the general structure of the dichotomy in all location parameter problems. But if I was writing it today, I would break it into more than one paper.

DasGupta: Well, I know that special corollaries of that paper of yours got published as separate papers because many people didn't know you had already solved their problems as particular cases in the '66 paper.

Brown: Yes, some of that happened.

DasGupta: But, Larry, then the 1964 paper on sufficiency (Brown (1964)) was not a part of your thesis?

Brown: No, although it could have been. I spoke to Bahadur about it when I spent a summer at Stanford. He was generous and nice to me; I was only a graduate student at that time.

There is a footnote in Lehmann's book (Lehmann (1959)) about the family of distributions being an Exponential family if a one dimensional sufficient statistic exists for all sample sizes. Dynkin wrote a key paper on that (Dynkin (1961)). The original paper was in Russian; I read only the English translation. Dynkin had a result that showed Lehmann's result assuming that the statistic was almost everywhere differentiable. But actually the proof had a technical error. It was just a technical error. You could take the CDF of the Cantor function and produce from there a counterexample. So Dynkin needed to assume differentiability everywhere, and then the proof would have worked.

DasGupta: You corrected that error?

Brown: Well, yes, that part of the 1964 paper was fine. But Dynkin had also assumed differentiability conditions on the densities and I wanted to see if Lehmann's assertion was right with conditions on just the statistic. So I worked on it, very hard actually, and wrote out what I thought was a proof. Ten years later, Pfanzagl gave a counterexample. My proof was off-base, and I engaged in wishful thinking, I would say. Later, I corrected the errors. But it took odd regularity conditions and I took many more pages. I wasn't happy with the conditions. Fortunately, Christian Hipp, a student of Pfanzagl could obtain a good result under a condition that was only slightly more stringent than necessary (Hipp (1974)). But his condition was clean. I consider that to be the correct resolution of that problem and the end of the story. I am glad it was settled. So it is true that under a condition, which is like absolute continuity, if a one dimensional sufficient statistic exists for all sample sizes, then the family of densities would have to come from the Exponential family.

Years at London and Berkeley

DasGupta: Where did you go after graduation?

Brown: Kiefer suggested that I should spend some time with David Cox. So I went to Birkbeck college in London as a postdoc for one year. David told me that he was working with P.A.W. Lewis on their book (Cox and Lewis (1966)). He and I talked and I attended many seminars in London. I met Dennis Lindley and Mervyn Stone and spent much time talking with them. I had a friendly professional relationship with David that year and have had ever since then. But perhaps I didn't really get all I could have out of that year at London.

DasGupta: And then you came back to America?

Brown: Yes, 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 somebody and they have 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: Scheffe. He was very kind as a Chair. And Neyman was as sweet as could be. Roger Purves was my office mate.

DasGupta: You wrote a very technical paper with him (Brown and Purves (1973)) on measurable selection of extremas?

Brown: Yes; it took years to come out. I am not the quickest person to write things up, and Roger was even less so (laughs).

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 till 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.

DasGupta: What did you teach there?

Brown: Well, I taught a classic multivariate course. I had a similar course from Narayan Giri at Cornell. Steve Stigler and Hira Koul were in that class. It would be regarded as a classic sort of course today. Quite a bit of distribution theory. I think I also taught a graduate decision theory class which was a bit below the level of my decision theory notes. And I taught a course in applied design theory.

The Vietnam War and Returning to Cornell

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

Brown: Yes. That was the days of the Vietnam war. I was reaching 26, and 26 was the magic age. If you didn't get drafted by 26, you were exempt. Anyway, I got the 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. They asked me to just appeal. I appealed, and it got denied by the local board. They apparently thought statistics was not a technical subject. The office at Berkeley said there was nothing to worry about. They would appeal to the state board. The state board denied the appeal. They upheld the local board's determination that statistics didn't fall under the President's guideline. It was upsetting; I didn't know what to do. My father got me an attorney in Oakland. I went to see my attorney and the attorney heard my story and said, "let me guess, you are either from west Los Angeles, or from Portland, Oregon. Those are the only two draft boards that could do something so stupid." It became clear that the obvious solution 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 numerous great papers you have written, that was one of the most influential. Did your return to Cornell had anything to do with that paper?

Brown: No, not really. I already knew that the admissibility question could be thought of as a calculus of variation problem. The basic ideas of the '71 paper were already there before I had to leave Berkeley.

DasGupta: I see. That paper is regarded as a masterpiece. What would you say was the best thing about that paper? Would you say that the best thing was that the paper gave a characterization of admissible procedures, that it was if and only if?

Brown: Looking back, I can tell you that making it if and only if was almost an obsession. You could say that the if and only if nature was the most striking. Actually one part still has a small qualifier.

DasGupta: You need a condition.

Brown: Right. But in retrospect, if I had left a little bigger gap between the if and the only if part, it would have greatly reduced the mathematical difficulty of the proofs and would have made the paper much more readable.

DasGupta: In that sense, the '71 paper carries some similarity to the '66 paper. Many of us wrote papers that proved special cases of the '71 paper. But Larry, the '71 paper had a lot more that was striking about it. It was this connection you revealed between admissibility and existence of recurrent random walks. When you discovered that connection between two completely unrelated problems, what was your state of mind?

Brown: I was very excited. But it was a natural analogy. I would tell you I was always hoping that there would be some such connection. I knew from my course from Spitzer at Cornell about the role of the dimension in the recurrence of the random walk; they were random walks, not diffusions. And I knew from my physics classes that you need a different theory of electrostatic potential in two and three dimensions. And that is a calculus of variation problem. So as soon as I realized that the admissibility question was a calculus of variation question as well, it seemed natural that there should be a connection. I am glad that everything came together.

DasGupta: Yes, and people who read the paper were excited and glad about the connection. Now the paper also had the Bayes risk and Bayes estimate representations that we later called the Brown identities.

Brown: Yes, as part of the proofs. Another thing that was buried in the proof was the Stein identity (Stein (1981)). But at that time, I didn't even notice that it was there and so I consequently didn't develop it. But of course, in terms of influence, that is the most influential discovery in that area of research.

DasGupta: Certainly so; and in fact, it had its influence in other areas as well, like regularization.

Brown: Oh yes; and later, in nonparametrics and wavelets also.

DasGupta: In the '71 paper, through the random walk connection, you also conjectured the result that no genuine Bayes minimax procedures can exist in less than 5 dimensions.

Brown: That's right; a few years later, Bill Strawderman gave some hierarchical priors in 5 or more dimensions that lead to minimax estimates (Strawderman (1971)). That was very nice; I didn't know, at all, that hierarchical priors would be the effective way to do it, although I knew about the general shape of the prior that was needed.

DasGupta: A few years ago, Joe Eaton gave another recurrence characterization, but through a markov chain on the parameter space (Eaton (1992)). It seems to match your 1979 (Brown (1979)) heuristics paper more closely.

Brown: Yes, and there are still challenging unanswered questions. I'm glad that someone of Joe's talent is still interested in pursuing this.

Frequentist vs. conditional inference

DasGupta: Now, Larry, after the 1971 paper, your interests diversified and you wrote a couple of classics on conditional inference. By that time, you had moved to Rutgers. Did you actually quit from Cornell?

Brown: Yes, I did. I very much appreciated Jack being in the faculty at Cornell. Jack had a good professional relationship with Federer in the Biometry department there although they were mostly interested in different sorts of design questions. I didn't have interest then in working in design. I have taught some optimal design, and it was a lovely contribution, and is somewhat under-appreciated now. Jack also had a close on (and then) off relationship with Bechhofer in the Operations Research department at Cornell. But I didn't have any connections outside the math department. 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: And there you started writing a series of papers on sequential testing, and on conditional inference.

Brown: Yes, the 1978 paper was mostly written while I was at Rutgers.

DasGupta: Actually, I want to pull back a little and talk about that little gem of yours, the 1967 paper.

Brown: Well (laughs); at that time, I would have described it as a small result on the t test. But it led me later to explore the topic in more detail.

DasGupta: In the 1967 paper (Brown (1967)), you showed that if one uses the t interval after having rejected the hypothesis, then the coverage is wrong. There is a relevant subset.

Brown: Yes, and the relevant subset is a very important one. I got into the area only because of previous literature on the problem. There were papers by Buehler, Feddersen, and some others (Buehler and Feddersen (1963), Buehler (1959)).

DasGupta: Larry, can you please help us understand the issues? For example, can you tell us a little about the relevance and importance of conditional statements and conditional inference?

Brown: Well, the issues are subtle. If there is a clear choice of the ancillary statistic, then one should condition on it. For example, the sample size. Who doesn't ? Take Cox's example. If we flip a coin either twice or 100 times, with a 50-50 probability, and if we know that the coin was flipped 100 times, we should condition on that. The report on the quality of inference has to be conditional. But you could build your procedure by looking at the entire risk function. For example, if we know the distribution of the ancillary, we can build a compound procedure by taking into account all the possible values that the ancillary can assume. But the statement, say the confidence statement, has got to be conditionally correct. But as you know, there are problems in which there are plenty of ancillaries and you don't know which one to condition on, and there are others where the ancillaries are just odd and no one wants to look at them. I don't think the area has reached any sort of a resolution and I can't say that the debate will go away.

DasGupta: Let me see if I understood you correctly. So for example if one has invariance with respect to a transitive group, and we know the distribution of the maximal invariant, then we can build our procedure by considering the whole risk function. You seem to be making a distinction between selection and assessment.

Brown: Exactly.

DasGupta: Ok, that part is clear now. But there seem to be some paradoxes in the conceptual framework of the whole context. Here is an example that has troubled me personally. In the normal case, the two sided interval does not have any relevant subsets, but the one sided interval does. So that says the one sided case is problematic. But now consider the results about irreconcilability of the P-value with the minimum posterior probability of the hypothesis over appropriate prior classes (Berger and Sellke (1987), Casella and Berger (1987)). Now the one sided case isn't problematic at all, and the results say that if there is a problem it's in the two sided case. There is no mathematical contradiction; both theorems are correct. But still it does cause a confusion in the conceptual understanding of the whole thing. Are we suffering from an incorrect or incomplete formulation?

Brown: Now you are asking good questions (laughs); tough ones. Well, I am tempted to give two disparate answers. One is what you said. They are different formulations and we have to take each one separately. There need not be a connection. If you ask about the incompleteness of the formulation, yes, there is. The standard two action formulation in the two sided case is a bit naive. In practice, if you reject, you want to know if it is larger or smaller. It's two one sided actions. I should think more about this; but perhaps the two sided case has a formulation problem.

DasGupta: All right. But even if we agree, in principle, that reports should be conditional, is there or will there be any impact in terms of how people behave, in their day to day uses of statistics? Do you see any sort of a real compromise at a methodological level, not just an academic level? Neyman discarded conditional confidence procedures because they are not unconditionally optimal even if they are unconditionally valid.

Brown: Well, I think there is already a synthesis at least in some problems. Take the spatial statistics models, for example. You could look at some of their procedures as coming from a big random effects model with parameters estimated by maximum likelihood, or from a hierarchical Bayes model with a diffuse prior on a final hyperparameter. The procedures have the correct interpretation according to each model. But it is basically the same procedure. So the debate about which is the correct model is actually philosophical. I would say that the synthesis is at the methodological level. The remaining debate is not so practically relevant. I think there will be a gradual compromise in some problems, but not in the entire spectrum of statistics. But I guess 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 procedure. You don't necessarily have to look at a risk function coming from a single loss function, but you will have lingering doubts about recommending a procedure if you don't know something about its distribution; the appropriate criterion would be a risk function in some problems.

DasGupta: That's interesting; Herman Rubin has been saying that too for many years. What about Jim Berger?

Brown: Well, Jim does risk calculations. He considers them useful in assessing robustness, that sort of thing (Berger (1986)). I think that's right. How do you assess robustness? There are other ways, but risks are useful there.

DasGupta: What about formal risk calculations to build procedures that beat a standard procedure everywhere? That enterprise generally leads to some form of shrinkage estimation. Is that a curiosity or is there more to it?

Brown: It cannot be just a curiosity partly because of how prevalent it is. We do shrinkage in problems involving multinomial data estimation all the time. Think of adding half a success in order to avoid 0 and 1. But it is shrinkage. Or think of ridge regression to avoid multicollinearity. That is shrinkage. Or for an example closer to our own joint interests think of the Wilson interval for a binomial proportion. This involves shrinkage to 1/2. So the shrinkage idea has much to offer. You don't have to do it for the formality of making something always less than or equal to zero. But you can and do benefit from it in other ways. The calculation obviously helps one in understanding the route to building desirable procedures. How do we form procedures that perform better in the area of maximum interest? The shrinkage literature has helped in that. This goes beyond the narrow shrinkage issue. People would sometimes say that a formal calculation isn't of any use; it is just mathematics. But formal calculations often ultimately lead to understanding. I couldn't predict 25 years ago that the unbiased estimate of risk would be useful in the many sorts of ways that we see now.

DasGupta: You mean in cross validation and regularization?

Brown: That's right.

Five years at Rutgers

DasGupta: Larry, the 1978 paper was written when you were at Rutgers. How much time did you spend there?

Brown: Five years; 1972 to 1977.

DasGupta: And during that time you wrote a series of papers on testing problems, complete class theorems and sequential analysis. Tell us a little about that.

Brown: I developed an independent interest in sequential problems at that time. My main interest was in drawing the analogy between sequential and fixed sample problems. The SPRT is a beautiful thing. In fixed sample problems, the next thing is a UMPU test, or something like that. That was a goal for the sequential problem. Of course there is no UMPU procedure, but we got some good results.

DasGupta: These are for the Exponential family?

Brown: Ah, yes, most needed an Exponential family. One particular result we were able to prove says that for any admissible test, the continuation region has a maximum length. So parabolic boundaries give you inadmissible tests. So, you know, these tests that people call repeated significance tests, are inadmissible (Brown, Cohen and Strawderman (1979), Berk, Brown, and Cohen (1981)). Almost all of that was collaborative work. Arthur Cohen was a collaborator in most of them. And Harold Sackrowitz, Bill Strawderman, Bob Berk also in various combinations.

DasGupta: Larry, how do you feel about the declining popularity of sequential procedures?

Brown: Well, yes, a part of the subject is kind of lost now. The idea was to review the experiment each time for a long time. Now in practice you usually don't want to do that. You can review the experiment every once in a while, may be six or ten potential stopping points. Thus, you do batch sequential trials. One area where the sequential idea is still heavily used is clinical trials. And change point problems. David Siegmund and Michael Woodroofe have contributed extensively on sequential problems.

DasGupta: They were able to give much more accurate approximations (Siegmund (1975), Pollak and Siegmund (1975), Woodroofe (1970, 1976, 1982)) to the error probabilities of the SPRT than Wald (Wald (1947)) did. What do you think of that?

Brown: Error probabilities and the average sample number. That was very very nice and they got much further than anyone else. But it is also true that with a computer, you can figure most things out more accurately than you could with the most sophisticated and best available asymptotic theory.

Returning to Cornell and Death of Jack Kiefer

DasGupta: After that period, you returned to Cornell. Please tell us a little about that move.

Brown: That was basically for family reasons. I was very happy at Rutgers. Many good friendships. I enjoyed it there. But my family didn't want to live in New Jersey. And again Jack brought me to Cornell. You know, he had been my guardian angel all those years. I remember the important phone call; you could call it a curious one. I called him and said that I have news for him. He responded by saying that he had news for me. I went first and said that I wanted to leave from Rutgers and come back to Cornell, if it was at all possible. And he said he was leaving Cornell and going to Berkeley. Jack was disenchanted with Cornell for a number of reasons. There was no statistics department. At one time we very much wanted to hire Peter Huber. It was vetoed by the topologists. Jack was very upset about that. And then of course he was depressed about my departure to Rutgers also. Anyway, he said he would do what he can to get me back to Cornell. Now you know, it is very very difficult to come back to a place that you have once left. But he did it. For him, the move to Berkeley was right. There were many more students, faculty, and a lot more interesting seminars. He was really very happy there.

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

Brown: It was a shock and very upsetting. In many ways, he was a father figure to me. It was shortly after the Purdue symposium, actually. I saw him there. And he used to come back to Ithaca still. It was very upsetting.

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

Brown: Oh absolutely; tremendous, tremendous range and depth of work. I hadn't read until his death many of his papers. After he passed away, I edited a volume on Jack's work (Brown et al. (1985)), and solicited discussions from many people. I was in Israel when the final publication was coming out, and the publisher dropped the ball. They forgot to put in many of the discussions. Afterwards, a supplement (Brown et al. (1986)) containing those discussions came out. I think most people don't know about the supplement. I have a copy. Here, you see, is my commentary on Jack's proof of what is basically the Hunt-Stein theorem (Kiefer (1963)).

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 LeCam's. I actually first saw it from Peter Huber. Later I came to know that Peter learned it from Le Cam. You see, Le Cam also didn't publish a lot of his results.

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 this material on measure theoretic and topological aspects of decision theory, the set of risks, that had some admissibility and minimaxity implications. But almost nobody understood the relevance of that to practical decision theory. It was much too formal. So I thought that to make it somewhat accessible, I should add on a chapter on applications, and you could do the applications in a unified way in the Exponential family. That's the context. So in that monograph, I think it came out in 1986, chapter 4 has the decision theoretic applications. But when I now teach a course out of that, I spend less time on those applications, and concentrate more on maximum likelihood, higher order asymptotics, and Efron's curvature theory. But it is sort of nice that the standard Exponential family material is in there in a unified way. If you want to go outside of the Exponential family, you need to add on special extra conditions.

DasGupta: And at around that time, you started experimenting with the information inequality and got a number of very interesting things out of it. Tell us a little about those works.

Brown: Well, the information formula for the Bayes risk was already in the 1971 paper. So it was a natural idea to want to study what could you say in general about Bayes risks by using the information function. You wouldn't expect an exact formula. The next useful thing is a bound (Brown and Gajek (1990), Brown and Low (1991)).

DasGupta: Did you get influenced by the Russian literature? Borovkov and others (Borovkov and Sakhanenko (1981)) were working on similar lines.

Brown: Actually I was completely ignorant of the Russians' interest in that problem. The Russians still hadn't started to come out of Russia then, and I have to confess I wasn't keeping track of the statistics that the Russians were doing. It was difficult to understand their writings, in a way. They were very good in talking to each other about their results. But the actual writing didn't come through well, especially in the translations.

DasGupta: You also gave an unusual proof of the central limit theorem (Brown (1982)) by using the information inequality.

Brown: Yes, but I consider that to be a side development. And I had unnecessary regularity conditions. Andrew Barron removed those extra regularity conditions and had an improved theorem (Barron (1986)).

DasGupta: Recently, we (Brown et al. (2000)) obtained some of these Bayes risk bounds by using the heat equation, a totally different approach. Something mysterious is going on.

Brown: Yes; I want to look at this more carefully. I haven't had enough time to study the connection. It is certainly interesting.

Interest in Nonparametrics

DasGupta: Did these papers precipitate your later interest in nonparametric function estimation ? You have been working on nonparametrics for more than 10 years now.

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

Now you can make it a parametric problem. I knew how to obtain minimax risks in parametric problems by using the Cramer-Rao inequality. It was really already implicit in work that Hodges and Lehmann did (Hodges and Lehmann (1951)). Once I saw this connection, the interest in nonparametrics developed.

DasGupta: You have had many students in the area of nonparametrics now. Did you do more work with Roger Farrell on this?

Brown: No, not with Roger. Roger doesn't mind listening. If I write something on the board, he would sit and listen and only say "hmm; let me think about it." A week later I would find a note from him in my mailbox. You too do that to me sometimes; you write me letters. But yes, most of my students in the last years have worked on nonparametrics.

DasGupta: What of this work is the most important?

Brown: Well, the most important work is by David Donoho and his coauthors (Donoho and Liu (1991), Donoho, Liu and MacGibbon (1990)). Really it is their work that opened up a whole new horizon; that 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 Pinsker and Efromovich (Pinsker and Efromovich (1996)). But that doesn't generalize. The techniques are really powerful. I remember reviewing a paper of Lucien Birge on estimating a monotone density. I first rejected it because he had a rate of convergence result for his procedure, but didn't say anything about the constant. I pushed him to have a result on the constant. Pushing is easy, you see (laughs). So anyway, Birge worked very hard, carefully, with the error terms and did return with a bound on the constant. I think he produced a bound within a factor of about 10 (Birge (1989)). Donoho, Liu, and MacGibbon showed that you can have access to the precise constant. I told David that was the best theorem of decision theory in the last decade. Mark Low and I (Brown and Low (1991, 1996)) were able to show that in some special cases it is possible to improve their constant, which was roughly 1.25.

DasGupta: Are you working on adaptive minimax estimates over large classes by using priors?

Brown: Well, that's really Linda Zhao.

DasGupta: And these are real priors; they don't depend on n and other weird things?

Brown: Yes, they are real priors. It should be useful to know in what types of classes you can do that.

DasGupta: So would it be fair to say that your prior work on decision theory inspired your recent interest in nonparametrics?

Brown: Yes, sure. But it was a confluence of considerations. That I could use the information bounds, and Roger's result that you cannot have $\sqrt{n}$ rates, and the desire to say something concrete about the constants. And of course I thought these were good problems. It was a combination of all of that.

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

Brown: Well adaptivity by using real priors is one. Consistency is another question. Persi Diaconis and David Freedman have done truly fundamental work on that (Diaconis and Freedman (1986, 1998)). We are learning a lot. But someone has to take these estimates and try them out. 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: Well, yes, but that was your lovely gesture to invite me to join in these works. I really like this sequence of papers with you and Tony Cai on the standard confidence interval for binomial proportions (Brown, Cai and DasGupta (2001a, b)). 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 (2000)). 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 large n, even very large, and for $p = 1/2$. Most people (including me) did not understand that. They did not understand the extent of the shortcoming in such a fundamental problem. And we show that the score interval is a very major improvement. In fact, there is a pedigree there, including the Jeffreys interval and the likelihood ratio interval that do equally well, and these two intervals are virtually identical in almost every respect. The challenging next question is to uncover an underlying deep reason for that. Then one could try to attack other problems once you understand a deep reason. It doesn't have to be certain Exponential family distributions or a lattice problem.

DasGupta: Are you happy that the asymptotic expansions in the Annals paper turned out to be so accurate at sample sizes as small as 20? I mean a lot of people treat asymptotic theory with skepticism and suspicion.

Brown: Certainly it was satisfying. It was fun to carry out the expansions. But we show that there is more. We actually learn many things about the problem from the expansions. Yes, I think, the Annals paper nicely supplements the Statistical Science paper. Others have shown me some other problems in which second order asymptotics can accurately explain moderate sample behavior. I think higher order asymptotics can be quite useful at least in some problems. Jayanta Ghosh (Ghosh (1994)) has written a very compehensive monograph on that.

DasGupta: Yes, I have much enjoyed this monograph. He is my advisor too. Are you working on any similar problems now ? How about the case of logistic regression or the curved Exponential family ?

Brown: A student of mine looked at the logistic regression case. But we didn't find any consistent pattern of results. Of course, that does not mean that there are none. It could be worth a more careful examination. It is an important problem.

DasGupta: How do you feel about the apparently strong opposing views that only confidence intervals matter and that confidence statements 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 in the sense we agree that coverage and length have to be balanced. But there is no uniquely appealing way to average the two. If you do, then you can't give assurance of any coverage. People don't like that. So then you come to assuring a specific coverage or something close to it. But then there won't be an agreement on what is close. But all of that 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 the 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: One reason was that David Freedman talked me into it. David had been writing a lot about census adjustments. One day he asked me if I would like to get involved. I was always interested in data, looking at them and seeing what they say. 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 may be they should call up someone else. David asked me at a good time and so I said yes. And another reason is that 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 the 2000 census. They gave me enough lead time. So I read up all the material and testified before the Senate.

DasGupta: Well, so you must have been on C-Span?

Brown: Actually I don't know. But the hearing itself was poorly attended. Senator Thompson was the chair of the committee. He was there. He asked a lot of very intelligent questions.

DasGupta: Is that project still going on?

Brown: The 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 a NRC oversight committee that is supposed to issue a final report on how well the census was done. But it takes time to decide the quality of your data in this kind of a problem. I think it will take another year, may be a bit more.

DasGupta: Are you also involved in a national telephone survey?

Brown: I am involved with a data set on telephone answering at various centers nationally. It's a lot of data and I find it interesting because it needs to be carefully modelled and analyzed. It doesn't have the kind of social significance as the census, but it is nevertheless interesting.

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

Brown: In some specific ways, yes. There are some capture-recapture schemes that are very specific to the census, much more complicated than the text book capture-recapture methods. There are confidence statements that are asymptotically approximately correct in the text book situation. It would be theoretically interesting to find out what sort of performance they provide in the schemes that are practically relevant, the ones that are used in that problem. Tony, you and I have sort of started to look at that. In the telephone survey, a student and Linda Zhao have found some innovative methods to use functional nonparametrics.

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

Brown: The National Academy oversees various panels of the NRC that study all kinds of scientific issues. The Academy and the NRC are loosely associated. The census subcommittee is one such panel. I am now on another committee that advises all the Federal statistical agencies, the Census bureau, the bureau of labor statistics, and so on. I have learned a lot about how the government works from the perspective of a statistician. It gives me a chance to provide some service to the profession. It is the same sort of reason that you try to serve on journal editorial boards as much as you can. You want to serve your subject.

DasGupta: Are you the present secretary of the Applied Mathematics and Statistics section of the National Academy?

Brown: Yes, but that's purely a function of who gets elected to the Academy.

DasGupta: It seems you are enjoying your varieties of involvements in applications and projects of social importance.

Brown: Yes, I am. The questions are quite interesting.

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: I certainly like to teach things in which I can use some mathematics, but I like teaching almost anything related to statistics. Something I am doing at Wharton that I didn't do before is to teach large service courses. I have done it a number of times now, and I have enjoyed it. Teaching anything is fun as long as the students are good.

DasGupta: Yes; are you seeing a change in student attitude about learning? Is there an expectation that things would come free, such as a good grade? Does grade inflation worry you?

Brown: I can speak about the students at Wharton. They work hard. Now if you talk about learning for its own sake, these students don't necessarily accept that as a model. They study and learn with money and professional success as their ultimate motivation. It is not really inappropriate. Learning for learning's sake is certainly one model, but does not have to be the only one. Clearly that would not be realistic. As regards grade inflation, at Wharton we are careful about that and have been able to cap grades at a modest level of inflation. But at other places, the story could be different.

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?

Brown: Well, about statistics graduate programs, there is something I am concerned about. US universities have always had many international students and in statistics many of the best graduates, historically, are international students. But we are getting less and less of US students. Actually it is not just in statistics, but in sciences as a whole, with the possible exception of biology. We can try to do some things, although only in a limited way, to get more good US students. The students are very much influenced by external factors, like the economy, in their decisions, and many are leaving after a college degree. It is of concern.

As regards importance of mathematics, statistics remains and will remain a largely mathematical subject. That does not mean that the type of mathematics or the reason for knowing and doing mathematics will remain the same. For example, think of the fifties. 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 frequently 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 in all of these things, you do need to know a lot of mathematics, just to model well, to understand what is going on, to study your procedures. Measure theory may not be the right tool for these things, but perhaps partial differential equations are. So even though there is obviously less of formal theorem proving in statistics today than there was 40 years ago, the importance of mathematics will not go away. You would always need mathematics, but for changing purposes. A broad training in mathematics will remain useful.

DasGupta: I am glad you said that. Larry, give me your ideal graduate program in statistics. What would you have them learn?

Brown: It is pretty clear what they should learn. But it is related with the questions of budgeting time, and not giving students a very false impression. They need to learn the core theory, at the level of Bickel and Doksum. You know that a new edition (Bickel and Doksum (2001)) just came out and I greatly enjoyed just teaching out of it. They also need to learn a subset of the material in the two books of Lehmann (Lehmann (1986), Lehmann and Casella (1998)). Look, these are fundamental things of statistics. They should learn Bayesian statistics at a moderately theoretical level, and although Bickel and Doksum has some asymptotics, they should have a better grounding in asymptotics. The reason that they need to learn about say the Neyman-Pearson theory, or unbiasedness, or Bayesian theory, is not that they are supposed to strictly adhere to one foundation or one principle, but because surrounding this core, an awful lot of useful other tools have developed. Asymptotic theory has produced many obviously useful methods. And they must learn some statistical computing. There was none of that when we were students but I wish I had learned more of that since then.

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

Brown: I would not say that they have to do consulting per se, but they have to see and analyze data. I cannot imagine what kind of a statistician would one be if he or she has had no background with data. And now there is no reason that students should not see and analyze data. It is so easily accessible. When we were students, we would get toy data examples. A few instructors and text book writers could make a toy example say something worthwhile. But only those that were imaginative and knowledgeable so they knew what would be worthwhile to pull out of a toy data set. Now you can attempt to examine data extensively.

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

Brown: I myself have struggled with that question. I hope there is a good answer to that question, but I am not sure I know it. I used to feel that they have to see data with the corresponding theory simultaneously.

DasGupta: It is difficult to design the courses that way.

Brown: Exactly. So I am just not sure that there is a good practical answer to that question.

DasGupta: And what about the background in probability?

Brown: If a student can manage it, then a measure theoretic background at the level of Billingsley (Billingsley (1995)). It is useful, still, to have that background in certain contexts. But it is possible to learn effectively and apply a lot of useful probability theory without learning the formal measure theory. At the very minimum one should learn distribution theory and limit theorems, and then a good bit of probabilistic modeling.

DasGupta: Can you please expand on that? Precisely what do you mean by probabilistic modeling?

Brown: Markov chains, Poisson processes and Brownian motion, some martingales and stochastic differential equations. For many students and many types of statistical academics, such a background could make more sense than having measure theory. There is the issue of budgeting of time.

The Important Developments in Statistics

DasGupta: Larry, in the last 25 years a number of fundamental problems in Mathematics got solved : the four color problem, classification of simple groups, the Biebarbach 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 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. But the limitations of the theory are anathema to the very concept of robustness;

that you could only do the calculations under quite specific restrictions. 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 convergence questions can be settled. Shrinkage estimation has led to many useful developments outside of the narrow focus of finding dominating estimates; Stein's unbiased estimate of risk (Stein (1981)) has had a lasting impact on statistics. And there have been a number of serious developments in how we analyze data, such as in classification problems (Breiman et al. (1984)). That's almost inevitable as long as you have creative thinkers because of the much better computing facilities that we have today.

DasGupta: What about robust Bayes procedures?

Brown: At a conceptual level, it continues to be an appealing theme. But again, as of now one can do the calculations only in special problems with special restrictions.

DasGupta: Is there a need of an analog to the M estimate in Bayesian robustness considerations?

Brown: Well, at a purely practical level, it has become common to use a diffuse prior at the top level of a hierarchical model. I guess everyone believes that is robust. But I haven't seen solid evidence of that. I would be interested in seeing a proof of that. It would be comforting.

Mathematics, Humans, Beauty, Truth and the World

DasGupta: Larry, I cannot resist the temptation of going with you into Philosophy at the end. You are not just a statistician; you are a mathematician too. How do you feel about mathematics? I mean what is it?

Brown: Personally, I have found mathematics to be useful and enjoyable because you could take a pragmatic problem in the real world and you can use mathematics to understand what is really going on. I think mathematical calculations are valuable for that reason.

DasGupta: Yes; but why can mathematics explain what is going on? Notable physicist Eugene Wigner says (Wigner (1990)): 'The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift that we neither understand nor deserve.' Are we to accept mathematics as something that is germane to the universe or did we make it up? If we made it up, we have the absurd situation that real things that are obvious for everyone to see are explained by something unreal that the human imagination simply made up. But if mathematics is germane, how did it get put there? This is no longer in the domain of just theology; scientists are asking these questions too (Davies (1984, 1992)).

Brown: Anirban, I wish I knew the answer! Physical laws as we have written them down may be an approximation to something very complex. Who knows if there are really permanent laws. Let's see what Russell says about this, for example. In his Lowell lectures (Russell (1914)), Russell asks what reason can be given for believing that causal laws will hold in future, or that they have held in unobserved portions of the past? Or, to ask a more modest question, have we any reason to believe that the law of gravitation will continue to hold in the future?; and he concludes '(there is) absolutely no logical ground for the belief that we can continue to expect the continuation of experienced uniformities.' Or, to take another popular example, you could not find one person bet a dollar against any amount of money that the sun would not rise tomorrow. We act like we just know that the sun will rise tomorrow. But don't we know things that have been observed? So we believe, rather than knowing it, that the sun will rise tomorrow, and we seem to assign 100% probability to it. What is the logical foundation for assigning 100% probability to it? Is it that if an event has happened everytime in a very long sequence of Bernoulli trials, then it is logical to assign it 100% probability at the next trial? It doesn't seem so. For argument's sake, take the day one day before the sun stops to rise; after all, the sun will die one day, after 4 or 5 billion years. Don't you get a contradiction with the predictive probability reasoning for the day just before doomsday? So you see, it is difficult to think about laws, or to say precisely what they mean. They have fantastic predictive accuracy. How and where they're from seems to be beyond the limits of science now. We can ask these questions, and maybe they are helpful in understanding the origin and the nature of the universe. But I am not sure we can unravel these mysteries. Many have thought about the need for the existence of a creator to explain consciousness and the universe. Now, it seems clear that people will have their personal beliefs on such an issue that really lies outside the limits of science. Science works on the basis of observation and inference and extrapolation from your observations. Even when we have a so called law, it is not provable. It is only consistent with our experience. So the ultimate questions on which we will probably never have observations may well be beyond human knowledge, although by asking these questions, the human beings are collectively raised to a great intellectual depth which has an aesthetic value.

DasGupta: Mathematics is also often linked to beauty, and not just to explaining a truth. James Jeans said that 'the Great Architect seems to be a mathematician'. Richard Feynman (Feynman (1965)) says: 'To those who do not know mathematics, it is difficult to get across a real feeling as to the beauty, the deepest beauty of nature.' Hardy, Poincare, Herman Weyl, and Nobel Laureate physicist S. Chandrasekhar have said that a mathematician, like a musician, has a responsibility to do beautiful things. Although beauty is purely in the mind, would you agree that beautiful mathematics is worthwhile regardless of any considerations of use? For example, take the stunning formula (Polya and Szego (1998))

$$\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}} \quad = \quad 1+\frac{1}{1+\frac{1}{1+\cdots}} \quad ;$$

do you think whoever proved it wasted his time?

Brown: Not all of mathematics can be driven by a desire to create a beautiful thing. Much of mathematics has to be related to a real question, an application, or a potential for an application in a real question. On the other hand, some mathematics can be a search for beautiful formulas, or abstract reasoning, for aesthetic reasons. But I personally hope that more mathematicians will do mathematics with a better appreciation of practical context. This is not a black and white issue and I don't think there can be a black and white answer.

DasGupta: How do you feel about the future of statistics? Would it continue in the foreseeable future to be useful to the human enterprise?

Brown: Oh yes; statistics will remain useful, if anything, in more human enterprises than ever before. I do see a red flag in the horizon. I see a danger of fragmentation. Branches of statistics, like biostatistics, could become essentially independent subjects without a link to the fundamental core. It has happened in other sciences, physics and chemistry. But in those subjects, over time, there have been interesting recombinations, like biochemistry 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 core.

DasGupta: Larry, let me draw your attention to a practical problem. It is becoming difficult to figure out what on a specific topic is known. How are we going to preserve and remember our knowledge?

Brown: Very interesting problem. They can put some of it on microfilm and tapes. But which ones? How many copies? Humans will have much thinking and decision making to do in probably 50 to 100 years.

DasGupta: And, Larry, how do you feel about the future of Homo Sapiens? Are you optimistic? Let me ask you a speculative question - do you think Homo Sapiens will survive as a species for another 1,000 years?

Brown: I'm sure we all hope the answer is "yes, and many more." We probably have the means to protect ourselves substantially from a small asteroid. Total self-destruction by nuclear war, fortunately, appears less likely than it was twenty years ago. Barring a sudden emergence of a deadly virus that is passed on easily, or unpredicted disastrous effects of things like possible global warming or genetic engineering, I am very hopeful we can last and reach much greater heights well into the future.

DasGupta: Technologically, yes. But how about human morality? We were told, as children, to be 'compassionate and gracious, slow to anger, and abundant in kindness and truth.' Are we seeing less of all of that?

Brown: Perhaps an advance in technology makes us more aggressive and our life more materialistic. There is less time for doing good. But, of course, yes, a degradation of human values is a degradation of humanity. After all, values make humans so different from animals. It cannot be DNA, because then we would be almost the same as the chimpanzees.

DasGupta: Larry, what about selfishness? Can an altruist culture survive?

Brown: A fully altruist society may not be functional. Take the example of people waiting for an elevator. Usually there is no such thing as a queue. If everyone was an altruist, the elevator would leave before anyone got in. So we cannot have a world of only the positives. You could probably make the same sort of argument about good and evil. There is a place for some evil in this world. Due to our evil actions, we cause pain and suffering to other humans, a destructive act. But to heal that pain and suffering for family and friends, we show love, loyalty and compassion. It seems like the evil act had a positive redeeming feature! It's very difficult to figure these things out.

DasGupta: Thank you Larry. You have had a remarkably distinguished career and I am sure you would continue to make more fundamental contributions. Thank you for giving us this chance to have a conversation with you. I wish you the best.

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

DasGupta: It is my pleasure. Good night Larry.

Bibliography

[1] Apostol, T.M. (1961). Calculus, Blaisdell Publ. Co., New York.

[2] Barron, A.(1986). Entropy and the Central Limit Theorem, Ann. Prob., 14, 1, 336-342.

[3] Berger, J.(1986). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

[4] Berger, J. and Sellke, T.M.(1987). Testing a point null hypotheses : irreconcilability of P-values and evidence (with comments and rejoinder by authors), Jour. Amer. Stat. Assoc., 82, 397, 112-139.

[5] Berk, R., Brown, L. and Cohen, A.(1981). Properties of Bayes sequential tests, Ann. Stat.,

9, 3, 678-682.

[6] Bickel, P.J. and Doksum, K.(2000). Mathematical Statistics, Prentice Hall, New York.

[7] Billingsley, P. (1995). Probability and Measure, 3rd Ed., John Wiley, New York.

[8] Birge, L.(1989). The Grenander estimator: a nonasymptotic approach, Ann. Stat., 17, 4, 1532-1549.

[9] Blackwell, D.(1951). On the translation parameter problem for discrete variables, Ann. Math. Stat., 22, 393-399.

[10] Borovkov, A.A. and Sakhanenko, A.I.(1981). Estimates for average quadratic risk (in Russian), Prob. and math. Stat., 2, 185-195.

[11] Breiman, L. et al.(1984). Classification and Regression Trees, Wadsworth, Stat. and Prob. Series, Belmont, CA.

[12] Brown, L.(1964). Sufficient statistics in the case of independent random variables, Ann. Math. Stat., 35, 1456-1474.

[13] Brown, L.(1966). On the admissibility of invariant estimators of one or more location parameters, Ann.Math.Stat., 37, 1087-1136.

[14] Brown, L.(1967). The conditional level of Student's t test, Ann. Math. Stat., 38, 1068-1071.

[15] Brown, L.(1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems, Ann. Math. Stat., 42, 855-903.

[16] Brown, L. and Purves, R.(1973). Measurable selections of extrema, Ann. Stat., 1, 902-912.

[17] Brown, L.(1978). A contribution to Kiefer's theory of conditional confidence procedures, Ann. Stat., 6, 1, 59-71.

[18] Brown, L.(1979). A heuristic method for determining admissibility of estimators - with applications, Ann. Stat., 7, 5, 960-994.

[19] Brown, L., Cohen, A. and Strawderman, W.E.(1980). Complete classes of sequential tests of hypotheses, Ann. Stat., 8, 2, 377-398.

[20] Brown, L.(1982). A proof of the central limit theorem motivated by the Cramer-Rao inequality, Stat. and Prob.: Essays in Honor of C.R. Rao, 141-148, North Holland, Amsterdam.

[21] Brown, L. et al.(1985). Collected Papers of Jack Carl Kiefer, I, II, and III, Springer-Verlag, New York.

[22] Brown, L. et al.(1986). Collected Papers of Jack Carl Kiefer, Supplementary Volume, Springer-Verlag, New York.

[23] Brown, L.(1986). Fundamentals of statistical Exponential families with applications in statistical decision theory, IMS Lecture Notes, Ser. 9, IMS, Hayward, CA.

[24] Brown, L. and Gajek, L.(1990). Information inequalities for the Bayes risk, Ann. Stat., 18, 4, 1578-1594.

[25] Brown, L. and Farrell, R.(1990). A lower bound for the risk in estimating the value of a probability density, Jour. Amer. Stat., Assoc., 85, 412, 1147-1153.

[26] Brown, L. and Low, M.(1991). Information inequality bounds on the minimax risk, Ann. Stat., 19, 1, 329-337.

[27] Brown, L. and Low, M.(1996). A constrained risk inequality with applications to nonparametric function estimation, Ann. Stat., 24, 6, 2524-2535.

[28] Brown, L. et al.(1999). Statistical controversies in Census 2000, Jurimetrics Jour, 39, 347-375.

[29] Brown, L., DasGupta, A., Haff, L.R. and Strawderman, W.E.(2000). Partial differential equations and expectation identities, with applications, Purdue Univ. Technical Report.

[30] Brown, L., Cai, T. and DasGupta, A.(2000). Confidence intervals in discrete Exponential families (Submitted).

[31] Brown, L., Cai, T. and DasGupta, A.(2001a). Interval estimation of a binomial proportion (with discussions), Stat. Sc. (To appear).

[32] Brown, L., Cai, T. and DasGupta, A.(2001b). Confidence intervals for a binomial proportion and asymptotic expansions, Ann. Stat. (To appear).

[33] Buehler, R.J.(1959). Some validity criteria for statistical inferences, Ann. Math. Stat., 30, 845-863.

[34] Buehler, R.J. and Feddersen, A.P. (1963). A note on a conditional property of Student's t, Ann. Math. Stat., 34, 1098-1100.

[35] Casella, G. and Berger, R.(1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem (with comments and rejoinder by authors), Jour. Amer. Stat. Assoc., 82, 397, 106-111.

[36] Cox, D. and Lewis, P.A.W.(1966). Statistical Analysis of Series of Events, Methuen, London.

[37] Davies, Paul(1992). The Mind of God, Simon and Schuster, New York.

[38] Davies, Paul(1984). God and the New Physics, Simon and Schuster, New York.

[39] Diaconis, P. and Freedman, D.(1986). On inconsistent Bayes estimates of location, Ann. Stat., 14, 1, 68-87.

[40] Diaconis, P. and Freedman, D.(1998). Consistency of Bayes estimates for nonparametric regression; normal theory, Bernoulli, 4, 411-444.

[41] Donoho, D. and Liu, R.(1991). Geometrizing rates of convergence, Ann. Stat., 19, 2, 633-667, 668-701.

[42] Donoho, D., Liu, R. and MacGibbon, B.(1990). Minimax risk over hyperrectangles, and implications, Ann. Stat., 18, 3, 1416-1437.

[43] Dynkin, E.(1961). Necessary and sufficient statistics for a family of probability distributions, Selected Translations in Math. Stat. and Probability, 1, 17-40, IMS and AMS, Providence, Rhode Island.

[44] Eaton, M.L.(1992). A statistical diptych : admissible inferences and recurrence of symmetric markov chains, Ann. Stat., 20, 3, 1147-1179.

[45] Efron, B.(1979). Bootstrap methods: another look at the jackknife, Ann. Stat., 7, 1, 1-26.

[46] Farrell, R.(1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, Ann. Math. Stat., 43, 170-180.

[47] Feynman, R.(1965). The Character of Physical Law, MIT Press, Cambridge, Massachusetts.

[48] Gelfand, A. and Smith, A.(1990). Sampling based approaches to calculating marginal densities, Jour. Amer. Stat. Assoc., 85, 410, 398-409.

[49] Ghosh, J.K.(1994). Higher Order Asymptotics, CBMS-NSF Regional Conference Series, IMS, Hayward, CA.

[50] Hipp, C.(1974). Sufficient statistics and Exponential families, Ann. Stat., 2, 1283-1292.

[51] Hodges, J. and Lehmann, E.L.(1951). Some applications of the Cramer-Rao inequality, Proc. 2nd Berk. Symp. in Math. Stat. and Prob., 13-22, Univ. Calif. Press, Los Angeles.

[52] Huber, P.J.(1964). Robust estimation of a location parameter, Ann. Math. Stat., 35, 73-101.

[53] James, W. and Stein, C.(1961). Estimation with quadratic loss, Proc. 4th Berk. Symp. Math. Stat. and Prob., 1, 361-379, Univ. Calif. Press, Berkeley.

[54] Kiefer, J. and Sacks, J.(1963). Asymptotically optimum sequential inference and design, Ann. Math. Stat., 34, 705-750.

[55] Kiefer, J.(1987). Introduction to Statistical Inference, Springer-Verlag, New York.

[56] Lehmann, E.L.(1959). Testing Statistical Hypothesis, 1st Ed., John Wiley, New York.

[57] Lehmann, E.L.(1986). Testing Statistical Hypothesis, 2nd Ed., John Wiley, New York.

[58] Lehmann, E.L. and Casella, G.(1998). Theory of Point Estimation, Springer, New York.

[59] Parzen, E.(1962). On the estimation of a probability density function and mode, Ann. Math. Stat., 33, 1065-1076.

[60] Pinsker, M. and Efromovich, S.(1996). Sharp optimal and adaptive estimation for heteroscedastic nonparametric regression, Stat. Sinica, 6, 4, 925-942.

[61] Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests, Ann. Stat., 6, 1267-1282.

[62] Polya, G. and Szego, G.(1998). Problems and Theorems in Analysis, I, Springer- Verlag, New York.

[63] Rosenblatt, M.(1956). Some remarks on nonparametric estimates of a density function, Ann. Math. Stat., 27, 832-837.

[64] Russell, B.(1914). Our Knowledge of the External World, Routledge, London.

[65] Siegmund, D.(1975). Error probabilities and the average sample number of the sequential probability ratio test, JRSSB, 3, 394-401.

[66] Stein, C.(1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proc. 3rd Berk. Symp. Math. Stat. and Prob., 1, 197-206, Univ. Calif. Press, Los Angeles.

[67] Stein, C.(1981). Estimation of the mean of a multivariate normal distribution, Ann. Stat., 9, 6, 1135-1151.

[68] Strawderman, W.E.(1971). Proper Bayes minimax estimators of the multivariate normal mean, Ann. Math. Stat., 42, 1, 385-388.

[69] Wald, A.(1947). Sequential Analysis, John Wiley, New York.

[70] Wigner, E.(1990). The unreasonable effectiveness of mathematics, Communications in Pure and Applied Mathematics, 13, 1-14.

[71] Woodroofe, M.(1970). On the first time $S_n > cn$, Ann. Math. Stat., 41, 2179-2183.

[72] Woodroofe, M.(1976). A renewal theorem for curved boundaries and moments of first passage times, Ann. Prob., 1, 67-80.

[73] Woodroofe, M.(1982). Nonlinear renewal theory in sequential analysis, CBMS-NSF Regional Conference Series in Appl. Math., SIAM, Philadelphia.