A Conversation with S. R. S. Varadhan

RAJENDRA BHATIA

S. R. S. Varadhan was awarded the Abel Prize for the year 2007. I met him on 14 and 15 May—one week before the prize ceremony in Oslo—in his office at the Courant Institute to interview him for the Mathematical Intelligencer. *My qualifications to interview him were that he and I are Ph.D.'s from the same institute, my Varadhan number is 2, and his was the first research talk that I attended as a graduate student. My major disqualification was that I know little of probability, and I felt like someone destitute of geometry daring to enter Plato's Academy.*

Though we had planned to talk for two or three hours, our conversation was spread over nearly eight hours. What follows is the record of this with very minor editing. To help the reader I have added a few "box items" that explain some of the mathematical ideas alluded to in the conversation.

Professor Varadhan, before coming here this morning I was in a Manhattan building whose designers seem to believe that the gods look upon the number 13 with an unfavourable eye, and they can be hoodwinked if the 13th floor is labelled as 12A. The Courant Institute building not only has 13 floors, your office here is 1313.

Well, the two thirteens cancel each other.

Excellent. I am further encouraged that I saw no sign prohibiting those ignorant of Probability from entering the Academy. So we can begin right away.

Early Years

Our readers would like to know the mysteries of your name.

In South India a child is given three names. My name is Srinivasa Varadhan. To this is prefixed my father's name Ranga Iyengar, and the name of our village Sathamangalam. So my full name is Sathamangalam Ranga Iyengar Srinivasa Varadhan.

What part of this is abbreviated to Raghu, the name your friends use?

The child is given another short name by which the family calls him. Raghu is not any part of my long name.

And you were born in Madras, in 1940. Your father was a high-school teacher. Did he teach mathematics?

He taught science and English. He had gotten a degree in physics, after which he had done teachers' training.

And your mother, Janaki?

My mother didn't go to school after the age of 8, as in those days it was not the custom to send young girls to school. But she was a versatile woman. She learnt to read very well, was knowledgeable and smart. For example, she taught me how to play chess. I could play chess even before I went to school.

Was the school in your village?

No, we had some land in the village but did not live there. My grandfather died when my father was 18. My father became the head of the family with two younger brothers one of whom was one year old, and he had to look for a job.

Did he teach in Madras?

He was in the District School System in Chengalpat district that surrounds the city of Madras on three sides. When I was born he was in Ponneri, a village 20 miles north of Madras. He moved from one place to another and I changed school thrice. I skipped some grades and was in elementary school for only two or three years. I spent three years in the high-school in Ponneri.

Do you remember some of your teachers?

Yes, I remember my high-school teachers very well. My father was the science teacher. I remember my maths teacher who was very good. His name was Swaminatha Iyer. He used to call some students to his home on the weekends and gave them problems to work on. His idea of mathematics was solving puzzles as a game. He gave us problems in geometry.

I remember that about my father too. He was a school teacher in Punjab. He would also teach on holidays and the parents of Sikh boys had to beg him to give at least one day off for the boys to wash and dry their long hair.

(Laughs) Yes, teachers those days thought it was their mission to educate. They enjoyed it. They were not very well paid but they carried a lot of respect. Now things have changed.

Did you have any special talent for mathematics in high-school?

In most exams I got everything right. I usually got 100 out of 100.

Was this so in other subjects as well?

In other subjects I was reasonably good but I had problems with languages. I was not very enthusiastic about writing essays.

What languages did you study?

English and Tamil; a little bit of Hindi but not too much.

Were you told about Ramanujan in school?

No. I learnt about him only in college.

Interesting, because in a high-school over a thousand miles away from Madras I had a teacher who worshipped Ramanujan and told us a few stories about him, including the one about the taxi number 1729.

Where did you go after high-school?

In those days one went to an Intermediate College. So, I went to Madras Christian College in Tambaram, and then to the Presidency College for a bachelor's degree.

At the Presidency College you studied for an honours degree in statistics. Why did you choose that over mathematics?

My school teacher Swaminatha Iyer told me that statistics was an important subject, and that Statistics Honours was the most difficult course to get into. In the entire state of Madras there were only 14 seats for the course. Statistics seemed to offer a possible profession in industry. My teacher had aroused my curiosity about it. So I did not apply for admission in mathematics, but in statistics, physics and chemistry.

Did you get admission in these other subjects also?

I think I did in physics but not in chemistry. I had applied for physics in the Madras Christian College, Tambaram, and for chemistry in Loyola College. You know admissions are a nerve-racking process. They do not put up all the lists at the same time. They want you to join the course immediately, and take away all your certificates and then you cannot switch your course. The Presidency College is different, being a government college. They put up all the lists on one day. My name was there in the statistics list.

You mean it is somewhat of a coincidence that you joined Statistics. If the other colleges had put up their lists earlier, you might have chosen another subject.

Yes.

Did you read any special books on mathematics in College?

I never learnt anything more than what was taught. But I found that I was not really challenged. I could understand whatever was taught. I did not have to work for the examinations, I could just walk in without any preparation and take the exams.

The newspapers in India have been writing that in the honours examination you scored the highest marks in the history of Madras University.

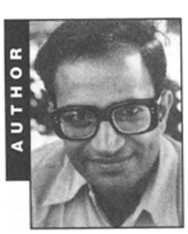
I think I scored 1258 out of 1400. The earlier highest score had perhaps been 1237, and one year after I passed out this course was stopped. So there was not any chance for any one to do better than me.

V. S. Varadarajan was also in the same college. Did you know him there?

He was three years ahead of me. I met him for the first time in Calcutta.

I was struck by the fact that the two persons from India who won the physics Nobel Prize—C. V. Raman and S. Chandrasekhar—and now the one to win the Abel Prize, all studied at the same undergraduate college. Was there anything special in the Presidency College?

I think at one time the Presidency Colleges in Madras, Calcutta and Bombay were the only colleges offering advanced courses. So, it is not surprising that the earlier Nobel Prize winners studied there. If you wanted to learn science, these might have been the only colleges. They were showpieces of that time. In my time the Presidency College was the only college in Madras that offered honours programs in all science subjects, and these were very good.



RAJENDRA BHATIA suggests that his exposure in the course of interviewing Professor Varadhan has been quite sufficient and a biographical note about the interviewer would be overdoing it. He quotes the character Insarov in *On the Eve* by Ivan Turgenev: "We are speaking of other people: why bring in yourself?"

Indian Statistical Institute Delhi
New Delhi, 110016
India
e-mail: rbh@isid.ac.in

Figure 1. The Guru and his disciples: A. N. Kolmogorov, dressed in a dhoti and kurta in Calcutta 1962. Standing behind him are L to R, K. R. Parthasarathy, B. P. Adhikary, S. R. S. Varadhan, J. Sethuraman, C. R. Rao, and P. K. Pathak.

Indian Statistical Institute

Now that you had chosen Statistics it was but natural that on graduating in 1959 you came to the Indian Statistical Institute (ISI) in Calcutta. Was the Institute well-known in Madras? In Delhi we had not heard about it.

We knew about it because C. R. Rao's book (*Advanced Statistical Methods in Biometric Research*) was one of the books we used. There were not too many books available at that time. Feller's book had just come out. Before that there was a book by Uspensky. These were the only books on Probability. In Statistics there was Yule and Kendall which is unreadable. C. R. Rao's was a good book.

Did you join the Ph.D. program?

Yes. My goal was to do a Ph.D. in Statistical Quality Control and work for the Industry. I did not know much mathematics at that time except some classical analysis. Then I ran into [K. R.] Parthasarathy, Ranga Rao and Varadarajan who started telling me that mathematics was much more interesting (Laughs) . . . and slowly I learnt more things.

What are your memories of the Institute? Do you recall anything about [P. C.] Mahalanobis?

Yes, Mahalanobis would come and say he would like to give lectures to us.

Were they good?

No! (Laughs). . . . He wanted to teach mathematics but somehow he also made it clear that he did not think much of mathematics. It is difficult to explain . . . C. R. Rao was, of course, always there. He was very helpful to students. But he didn't give us any courses. There were lots of visitors. For example, [R. A.] Fisher used to come often. But his lectures on Fiducial Inference were ununderstandable. (Laughs)

Did R. R. Bahadur teach you?

Yes, in my first year two courses were organised. One on Measure Theory by Bahadur and the other on Topology by Varadarajan. I went through these courses but did not understand why one was doing these things. I was not enthused by what I was learning and by January was feeling dissatisfied. By then Parathasarathy, Ranga Rao and I decided to start working on some problem in probability theory. In order to do the problem we had to learn some mathematics—and that is how I learnt and found that the things I had studied were useful.

So your getting into probability or mathematics, was because of the influence of your fellow students.

Yes, it was because of Parthasarathy and Ranga Rao. We studied a lot of things. I was interested in Markov processes, stochastic processes, etc. We used to run our own seminar at 7:30 AM. J. Sethuraman also joined us.

What did you study at this time?

We went through Prohorov's work on limit theorems and weak convergence, Dynkin's work on Markov processes; mostly the work of the Russian school. At that time they were the most active in probability.

Were their papers easily available?

Yes, some of them had been translated into English, and we had a biochemist Ratan Lal Brahmachary who was also an expert in languages. He translated Russian papers for us. We also learnt some languages from him. I learnt enough Russian and German to read mathematics papers.

What books did you read?

We read Kolmogorov's book on limit theorems. Dynkin's book on Markov processes had not yet come out. We read his papers, some in English translation published by SIAM, some in Russian.

Was mathematics encouraged in the Institute, or just tolerated?

It was encouraged. C. R. Rao definitely knew what we were doing and encouraged us to do it. There was never any pressure to do anything else. Mahalanobis was too busy in other things. But he also knew what we were doing.

How did the idea of doing probability theory on groups arise?

Before I came to the Institute, Ranga Rao and Varadarajan had studied group theory. So Ranga Rao knew a fair amount of groups. When we read Gnedenko and Kolmogorov's book on limit theorems it was clear that though they do everything on the real line there is no problem extending the results to finite-dimensional vector spaces. So there were two directions to go: infinite dimensions or groups. The main tools used by Gnedenko and Kolmogorov were characteristic functions. I did not know it at that time, but Ranga Rao knew that for locally compact groups characteristic functions worked well, though they did not work so well for infinite dimensional spaces. So our first idea was to try it for locally compact groups. Then I did some work for Hilbert spaces.

Your first paper is joint work with Parthasarathy and Ranga Rao. The main result is that in the space of probability measures on a complete separable metric abelian group indecomposable measures form a dense G_δ set. Why was this surprising?

At that time we were learning about Banach spaces, Baire category, etc. To show that a distribution on the real line is indecomposable was hard. You can easily construct discrete indecomposable distributions. The question (raised by H. Cramér) was whether there exist continuous indecomposable distributions. We proved that continuous distributions and indecomposable distributions both are dense G_δ sets. So their intersection is non-empty, in fact very large.

I read a comment (by Varadarajan) that this work was sent to S. Bochner and he was very surprised by it.

No . . . , I don't think so. Certain things are appearing in print [after the Abel Prize] about which I do not seem to know.

After this you studied infinitely divisible distributions on groups.

We studied limit theorems on groups. The first paper was just really an exercise in soft functional analysis. The second problem was much harder. In proving limit theorems you have to centre your distributions by removing their means before adding them. The mean is an expectation of something. In the group context this is clear for some groups and not for others. To figure this out for general groups we had to use a fair amount of structure theory. The main problem was defining the logarithm of a character in a consistent way.

Your Ph.D. thesis was about the central limit theorem for random variables with values in a Hilbert space.

Yes, then we thought of extending our ideas to Hilbert spaces, and there characteristic functions are not sufficient. You need to control some other things.

Is that the Lévy concentration function?

Yes.

Was that the first work on infinite-dimensional analysis of this kind? Had the Russian probabilists done similar things?

They had tried but not succeeded.

So this is the first work on measure theory without local compactness.

Yes.

What happened after this?

The work on Hilbert space suggests similar problems for Banach spaces. Here it is much harder and depends on the geometry of the Banach space. There has been a lot of work relating the validity of limit theorems of probability to the geometry of the Banach space.

Was Kolmogorov your thesis examiner?

Yes, one of the three.

Some newspapers have written that C. R. Rao wanted to impress Kolmogorov with his prize student and brought him to your Ph.D. oral exam without telling you who he was.

(Laughs) Yes, the story is pure nonsense. We knew Kol-

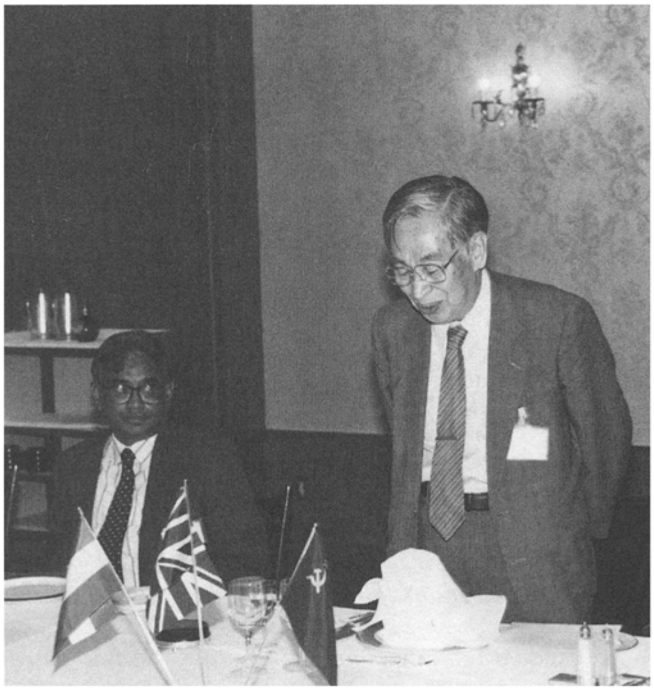


Figure 2. Varadhan and Kiyosi Itô at the Tanigushi Symposium in 1990.

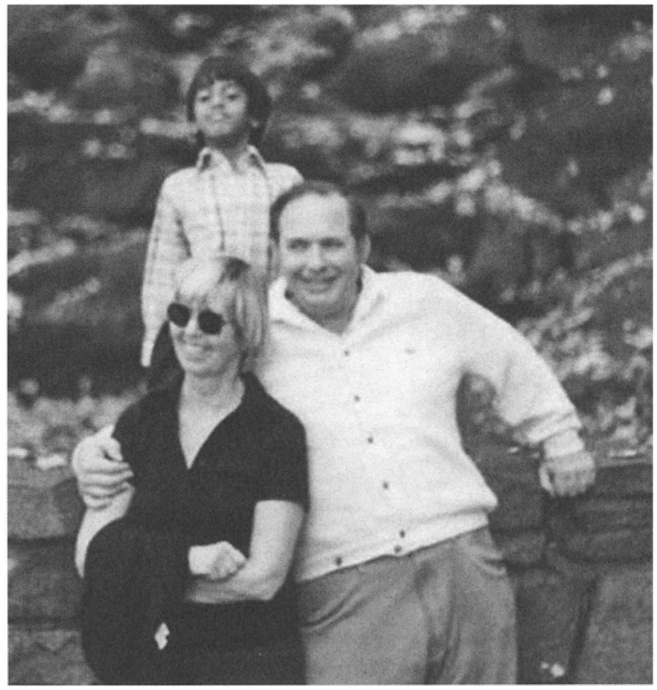


Figure 3. A rare photograph of Monroe Donsker with his wife and Varadhan's son Ashok.

mogorov was going to visit and were prepared for it. He attended my talk on my work and I knew he was going to be one of my thesis examiners. My talk was supposed to be for one hour but I dragged it on for an hour and a half and the audience got restless. Then Kolmogorov got up to make some comments and some people who had been restless left the room. He got very angry, threw the chalk on the floor, and marched out. And I was worried that this would be the end of my thesis. (Laughs) So we all went after him and apologised. He said he was not angry with us but with people who had left and wanted to tell them that when someone like Kolmogorov makes a remark, they should wait and listen.

Do you remember any of his lectures?

Sure, I attended all of them. In one of them he talked about testing for randomness and what is meant by a random sequence. If you do too many tests, then nothing will be random. If you do too few, you can include many systematic objects. He introduced the idea of tests whose algorithmic complexity was limited and if you did all these your sequence would still be random. He insisted on giving his first lecture in Russian and Parthasarathy was the translator.

I learnt that Kolmogorov travelled by train to other places in India. Did you accompany him?

Yes, Parthasarathy and I, and perhaps some others, travelled with him. We went to Waltair, Madras, and then to Mahabalipuram where Parthasarathy fell from one of the temple sculptures and fractured his leg. Then he did not travel further and I accompanied Kolmogorov to Bangalore and finally to Cochin, from where he caught a ship to return to Russia.

Varadarajan was not in Calcutta all this time. He returned in 1962 and pulled you towards complex semisimple Lie groups.

Yes, he returned during my last year at ISI. He had met Harish-Chandra and wanted to work in that area.

This was a different area, and considered sort of difficult. Was it difficult for you?

Not really. We were just learning, it was hard learning because it was different.

Very few people, even among those working on the topic, understood Harish-Chandra's work at that time. What is the wall you had to climb to enter into it?

I wouldn't say we understood all of it. We just made a beginning. Varadarajan, of course, knew a lot more and guided me. We had a specific goal, a specific problem. When you have a specific problem you learn what you need and expand your knowledge base. I find that more attractive than saying I want to learn this subject or that and face the whole thing at once.

Was this work completely different from what you had been doing with groups?

It was completely different. So far we had been working on abelian groups and not on Lie groups.

Was there a feeling that Lie groups and not probability was real *mathematics?*

No, I don't think so. Varadarajan was interested in mathematical physics, and he thought Lie groups were important there.

In the preface to his book on Lie groups he says his first introduction to serious mathematics was from the works of

Harish-Chandra. That would suggest that what he had been doing earlier was not serious mathematics.

Perhaps what he meant by serious mathematics is difficult mathematics. I think probability came easy to him. On the other hand, Harish-Chandra's work is certainly hard because it requires synthesizing many things. In probability theory, especially limit theorems, if you know some amount of functional analysis and have some intuition, you can get away with it.

So, he thought it was much more difficult.

It was much more inaccessible. One gets much more pleasure out of going to a place that is inaccessible.

And you never had that feeling.

No, for me I was quite happy doing whatever I had been doing.

Is there any other work from ISI at that time that influenced your later work? For example, the paper by Bahadur and Ranga Rao related to large deviations?

Yes, very much so. Cramér had a way of computing large deviations for sums of independent random variables and it led to certain expansions. Bahadur and Ranga Rao worked out the expansions. So I knew at that time about the Cramér transform and how large deviation probabilities are controlled by that.

Would it be correct to say that at ISI you got the best possible exposure to weak convergence and to limit theorems? Varadarajan was one of the early pioneers in weak convergence.

Prohorov's paper came in 1956 and he studied weak convergence in metric spaces. Varadarajan knew that and took it further to all topological spaces. Ranga Rao in his Ph.D. thesis used weak convergence ideas to prove difficult theorems in infinite-dimensional spaces, such as an ergodic theorem for random variables with values in a Banach space. That was very important for me, as I saw how weak convergence can be used as a tool, and I have used that idea often.

In the preface to his book Probability Measures on Metric Spaces, *Parthasarathy talks of the "Indian school of probabilists". Did such a thing ever exist?*

Ranga Rao, Parthasarathy, Varadarajan, and I worked on a certain *aspect* of probability—limit theorems—where we did create a movement in the sense that our work has influenced others, and we brought in new ideas and techniques.

The "school" lasted very briefly. What makes a school?

The school does not exist but the ideas exist. (Laughs).

With hindsight, do you still *consider this work to be important?*

I think it is important. It has influenced others, and I have used ideas from that work again and again in other contexts.

Later generations in the Institute look at that period with a sense of reverence and of longing. The burst of creativity in Calcutta in the 1950's and 60's was perhaps like a comet that will not return for a long time. For the Tata Institute also that seems to have been the golden period.

One must remember that at that time if anyone wanted to do research in mathematics in India, there were only two places, the TIFR or the ISI. If you went to any university, you would be attached to exactly one professor and do exactly what he did. There was no school there. But now that has changed. There are lots of places in India where a student can go. ISI is not the only place, and even ISI has other centres now.

Courant Institute

You came to the Courant Institute in 1963 at the age of 23. How did you choose this place?

When I learnt about Markov processes, I learnt they had links with partial-differential equations. Varadarajan had been here as a post-doctoral fellow in 1961–62. When he returned to India he told me that if I wanted to learn about PDE, then this was the best place for me.

The reason for his recommending you this place was its strong tradition in differential equations, not probability. There were some probabilists here like [H. P.] McKean and [Monroe] Donsker.

McKean wasn't here. Donsker had come just the year before.

And in PDE, Courant, Friedrichs, Fritz John, Nirenberg, and Lax were all here.

Yes, Moser and Paul Garabedian too. Almost everybody (important) in PDE was here.

Stroock, in one of his write-ups on you, says that few other probabilists knew statistics at that time. That was one of your advantages.

I think that is an exaggeration. In the United States probabilists came either from the mathematics or the statistics departments. Those who came from the statistics department surely knew statistics. Stroock himself had a mathematics background.

How about the converse? Did statisticians know probability well at that time?

I think they knew *some* probability. You cannot do statistics without knowing probability. Those who worked on mathematical statistics definitely knew enough probability to be proving limit theorems. That is what mathematical statistics at that time was.

What was the status of probability theory itself in 1960, within mathematics. For example, I have here with me an obituary of [J. L.] Doob by Burkholder and Protter. They say that before Doob's book "probability had previously suffered a cloud of suspicion among mathematicians, who were not sure what the subject really was: was it statistics? a special use of measure theory? physics?"

Doob's book was the first one to put probability in a mathematical context. If you read the book, it is clear that what he is doing is mathematics; everything is proved.On the other hand people at that time were also influenced by Feller who came from a different background—he was a classical analyst. I don't think he cared much about Doob's book. I think there was some friction there.

Did Feller think the book was too mathematical?

I think it was too . . . theoretical. It is not so much the mathematics. It is totally devoid of any intuition, it is very formal. For that reason Feller did not like the book. Doob's book is difficult to learn from. For certain topics like martingales it was perhaps the ideal book. I was interested in Markov processes, and Dynkin's books were the first ones that treated the subject in the way it is done today.

I return to my question about the status of probability in 1960. Was it indeed under a cloud of suspicion and mathematicians did not know where to place it?

It is hard for me to say. . . . I think there were some like Mark Kac who knew what exactly it could do or not do. He used it very effectively to study problems in physics. Donsker knew it was a branch of mathematics and he was interested in using it to solve problems of interest in analysis. And then there was [G. A.] Hunt who did excellent things in probability and potential theory.

I think again it was Doob who made the connection between probability and potential following the work of Kakutani.

Yes, Doob made the initial connections but the decisive work was done by Hunt.

Courant was 75 when you came here. Do you have any memories of him?

I met him two or three times at social dinners. I had no scientific interaction with him. He had retired and came to his office on some days.

Let me ask you a few questions about his spirit and his influence on the thinking here.

In her famous biography of Courant, Constance Reid says he resisted the trend towards "generality and abstraction" and tried to "shield" his students from it. She cites Friedrichs as saying Courant was "a mathematician who hates logic, who abhors abstractions, who is suspicious of 'truth', if it is just bare truth." Later in the book she says Courant told her he did not hate *logic, he was* repelled *by it. At the same time he regarded himself as the "intellectual son" of Hilbert. Now Hilbert certainly solved several concrete problems. But he had a major role in promoting abstraction in mathematics, and also worked in logic itself.*

When Reid pointed this out to Courant he replied, "Hilbert didn't live to see this overemphasis on abstraction and the self-emulation and self-adulation that some of these abstractionists show." This quote in the book is followed by one from Friedrichs: "We at NYU recognised rather tardily *the achievements of the leading members of 'Bourbaki'. We really objected only to the trivialities of those people whom Stoker calls 'les petits Bourbaki'."*

(Laughs) I think there is a difference in the point of view. I think abstraction is good—to some extent. It tells you why certain things are valid, the reason behind it; it helps you put things in context. On the other hand the tradition in the Institute has always been that you start with a concrete problem and bring the tools needed to solve it, and as you proceed do not create tools irrespective of their use. That is where the difference comes, with people who are so interested in the formalism that they lose track of what it is good for.

When you began your career, the Bourbaki style was on the rise. Did that affect your work?

Not here!

Is there a clear line between "too abstract" and "too concrete"? Let me again quote from Reid's book. Lax is cited there as saying that there was "provincialism at NYU which was somewhat Göttingen-like." He quotes Friedrichs to say that von Neumann's operator theory was considered too abstract there. I find this surprising. First, I thought Göttingen was very broad, and second, if we apply the same yardstick what do we say of Hilbert's work on the theory of invariants? Gordan had dismissed this work of Hilbert as "too abstract" and called it "theology", not mathematics. Now we find Friedrichs saying von Neumann was considered too abstract. Is there some clear line here, or everyone feels comfortable with one's own idea of abstraction?

I do not attach much importance to these things. I think abstract methods are useful and one uses whatever tool is available. My philosophy always has been to start with concrete problems and bring the tools that are needed. And then you try to see if you can solve a whole class of problems that way. That is what gives you the ability to generalise.

I will persist with this question a little more. Lax says that what they felt in Göttingen about von Neumann's theory of operators, here at this Institute they felt the same way about Schwartz's theory of distributions. He says it is one of those theories which has no depth in it, but is extremely useful. He goes on to say they resisted it because it was different from the Hilbert space approach that Friedrichs had pioneered. Later both he and Friedrichs changed their minds because they found distributions useful in one of their problems. One of Courant's last scientific projects was to write an appendix on distributions for Volume III of Courant-Hilbert that he was planning. Is there a lesson here?

The lesson is precisely that if you do not see any use for something, then it is abstract. Once you find a use for it, then it becomes concrete.

What Lax calls the "provincialism" at NYU, did it exist at other places? Most of the elite departments in the US those days hardly had anyone in probability or combinatorics.

Yes, there has been a certain kind of snobbery. If one does algebraic geometry or algebraic topology, one believes that is the golden truth of mathematics. If you had to actually make an estimate of some kind, that is not high mathematics. (Laughs)

Has this changed in the last few years?

I think fashions change. Certain subjects like number theory have always been important and appeal to a lot of people. Some other subjects that had been peripheral become mainstream as the range of their applications grows.

In this shift towards probability and combinatorics has computer science played a major role?

Computer science has raised several problems for these subjects. There are whole classes of problems that cannot be solved in polynomial time in general, but for which algorithms have been found that solve a typical problem in short time. What is 'typical' is clearly a probabilistic concept. That is one way in which probabilty is useful in computer sciences. Indirectly many of the problems of computer science are combinatorial in nature, and probability is one way of doing combinatorics.

I come back to my question about admiration for and resentment against Bourbaki. Do you think this had unhealthy consequences? Or, is it that mathematics is large enough to accommodate this?

I think we have large enough room for different people to do different things. Even in France, those brought up on the Bourbaki tradition, if they need to learn other things, they will do it. People want to solve the problems they are working on, and they find the tools that will help them. Sometimes you have no idea where the tools come from. Ramanujan's conjectures were solved eventually by Deligne using the étale cohomology developed by Grothendieck.

Did you ever feel, as some others say they have felt, that some branches of mathematics have been declared to be prestigeous and very good work in others is ignored?

I never felt so. At ISI there was no such thing. At the Courant Institute there was no snobbery.

Except that there was no need to do distributions!

No, I don't think so. I will put it this way. Distributions are useful because they deal with objects that are hard to define otherwise. But, more or less, the same thing can be achieved in a Hilbert space context. It is true that duality in the context of topological vector spaces is much broader but a major part of it can be achieved by working in Hilbert spaces. A problem does not come with a space. You choose the space because it is convenient to use some analytical methods there. Some people find Hilbert spaces more convenient than (general) topological vector spaces. That is what Friedrichs did initially. When you come to a problem where one space does not work you go to another one.

Now let me ask a question to which I know your answer. But I will ask it and then put it in context. Were you disappointed that you did not get the Fields Medal?

No.

What I really mean to ask you is whether you did not get the Fields Medal because at that time probability theory was not considered to be the kind of mathematics for which Fields Medals are given.

I can't say. (Laughs) It is true that, historically, Fields Medals have gone much more to areas like algebraic geometry and number theory. Analysis, even analysis, has not had as many. It is only this time that probability has got its first Fields Medal. Sure, I would have been happier if one had been given to a probabilist earlier. But after all, (at most) four medals are given every four years. Many people who deserve these awards do not get them.

Let us come back to 1963. Did you start getting involved in PDE soon after coming here?

I was still continuing my work in probability, and whatever PDE I needed I learnt as I went along. And here you do not even have to make an effort to learn PDE, you just have to breathe it.

Stroock says that the very first problem you solved after coming here was done simultaneously by the great probabilist Kiyosi Itô, and you did not publish your work. What was the problem?

Varadhan's Lemma

There is a simple lemma due to Laplace that is useful in evaluating limits of integrals: For every continuous function h on [0,1]

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^1 e^{-nh(x)}\, dx = -\inf h(x).$$

(The common fact $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$ can be used to get a one-line proof of this lemma:

$$\lim_{n\to\infty} \log \|e^{-h}\|_n = \log \|e^{-h}\|_\infty = \log \sup e^{-h(x)} = -\inf h(x).)$$

Now suppose we are given a family of probability measures and are asked to evaluate the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \int_0^1 e^{-nh(x)}\, d\mu_n(x).$$

In his 1966 paper Varadhan argues that if we have

$$d\mu_n(x) \sim e^{-nI(x)}dx,$$

then by Laplace's lemma this limit would be

$$-\inf\,[h(x) + I(x)].$$

The function $I(x)$ is now called the *rate function*. It is defined for spaces much more general than the unit interval [0,1].

Let X be any complete separable metric space (Polish space). A rate function I is a lower semicontinuous function from X into $[0, \infty]$ such that for every $\ell < \infty$ the level-set $\{x : I(x) \le \ell\}$ is compact. A family $\{\mu_n\}$ of probability measures on X is said to satisfy the *large-deviation principle* (LDP) with the rate function I if

(i) for every open set U

$$\underline{\lim}\ \frac{1}{n} \log \mu_n(U) \ge -\inf_U I(x),$$

(ii) for every closed set F

$$\overline{\lim}\ \frac{1}{n} \log \mu_n(F) \le -\inf_F I(x).$$

Varadhan's Lemma says that if $\{\mu_n\}$ satisfy the LDP, then for every bounded continuous function h on X

$$\lim_{n\to\infty} \frac{1}{n} \log \int_X e^{-nh(x)}\, d\mu_n(x) = -\inf\,[h(x) + I(x)].$$

There is an amazing variety of situations where the LDP holds. Finding the rate function is a complex art that Varadhan has developed over the years.

It was a question about giving a more precise meaning to Feynman integrals. In the Schrödinger equation there is a differential part (the Laplacian) and there is a potential part (some function). The measure you want to construct depends on the Laplacian. Without the i this will be the Brownian motion. The presence of i makes it the Feynman integral, and not so well-defined. If you take the Fourier transform of the Schrödinger equation, then the potential part (multiplication) becomes a convolution operator and plays the role of the differential operator, and the Laplacian becomes multiplication by x^2. The idea was that now you base your measure on (the Fourier transform of) the potential part. That is not as bad as the Feynman integral; it may even be a legitimate integral for some nice potentials—like functions with compact support, some functions with rapid decrease, or the function x.

Was this the first work you did after coming to the Courant Institute?

Well, Donsker asked a very special question. There are several approximations that work for the Wiener integral. Do the same approximations work for the Feynman integral? If you take the Fourier transform, then they do work because the measure based on the potential is nicer. That was the context of my work.

Your first paper at the Courant Institute appeared in 1966 and has the title "Asymptotic probabilities and differential equations". Can you describe what it does?

When I came here in 1963 Donsker had a student by the name Schilder. He was interested in the solution of certain equations whose analysis required Laplace type of asymptotics on the Wiener space. You have the Wiener measure and the Brownian motion that has very small variance, and you are interested in computing the expectation of some thing like exp $(\varepsilon^{-1}f)$. So you have something with very large oscillations and you are computing its expectation with respect to something with very small variance. If you discretize time, then you get Gaussian densities instead of the Wiener measure and this becomes standard Laplace asymptotics. So you do it for finite dimensions and interchange limits, and that is what Schilder had done. Having been brought up in the tradition of weak convergence it was natural for me to think of splitting the problem in two parts. One was to abstract how the measures behave asymptotically and then have a theorem linking the behaviour of the integrals to that of measures. That is not a hard theorem to prove, once you realize that is what you want to do. Then if you know a little bit of functional analysis, that Riemann integrals are limits of sums, and how to control errors you can work out the details. It was clear that if you have probabilities that decay exponentially and functions that grow exponentially you can do it by formulating a variational problem that can be solved.

Is this paper the foundation for your later work with Donsker?

Yes. This paper has two parts. First I prove the theorem I just mentioned and then apply it to a specific problem. Schilder had studied the case of Wiener measure with a small variance. I do it for all processes with independent increments.

Your address on this paper is given as the Courant Institute and ISI. Were you still associated with the ISI?

I was on leave from the ISI for three years and resigned later.

You have stayed at this Institute since 1963. What has been the major attraction? Is it New York? the Institute? something else?

I like New York. After living in Calcutta I got used to living in big cities. In 1964 I got married and my wife was a student here. So when the opportunity came to join the faculty here, I did so. By that time I had got used to the place, and I liked it and stayed. It is a good place and has been good for me. It is always exciting and interesting with lots of people coming here all the time.

The Martingale Problem

Most of your work has been in collaboration. You began by collaborating with a small group at ISI. Then in 1968 appears your first paper with Stroock. Were you working by yourself between 1963 and 1967? You have single-author papers in these years, which is unusual for you.

I was a post-doctoral fellow working mainly by myself. But I had lots of conversations with Donsker.

How did your collaboration with Stroock begin?

He was a graduate student at Rockefeller and we met at joint seminars. In 1965–66 I wrote a paper on diffusions in small time intervals and he was interested in that. He came here as a post-doc and joined the faculty after that. He was here for about six years from 1966 to 1972. We talked often and formulated a plan of action, a series of things we would like to accomplish together.

I have never met him but from his writings I get the impression that he will like it if I say that your coming together was a stroock of good luck.

(Laughs.)

Your work with Stroock seems to have flowed like the Ganga. In three years between 1969 and 1972 you published more than 300 pages of research in a series of papers in the Communications on Pure and Applied Mathematics. *Can we convey a flavour of this work to the lay mathematician?*

Let us understand clearly what you want and what you are given. In the diffusion problem certain physical quantities are given. These are certain diffusion coefficients $\{a_{ij}(x)\}$ which form a positive-definite matrix $A(x)$ for each x in $\mathbb{R}^d$, and you are given a first-order drift, i.e., a vector field $\{b_j(x)\}$. We want to associate with them a stochastic process, i.e., a measure P on the space Ω consisting of continuous functions $x(t)$ from $[0,\infty)$ into $\mathbb{R}^d$ such that $x(0) = x_0$ almost surely.

When we started our work there were two ways of doing it. One is the PDE method in which you write down the second-order PDE

$$\frac{\partial u}{\partial t} = \frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_j b_j(x)\frac{\partial u}{\partial x_j}.$$

This equation has a fundamental solution $p(t, x, y)$. You use this as the transition probability to construct a Markov process P, and the measure coming out of this process is

The Martingale Problem

For the discussion that follows it might be helpful to remind the reader about a few facts about diffusion processes.

Let us begin with the prototype Brownian motion (the Wiener process) in $\mathbb{R}$. It is a process with stationary independent increments that are normally distributed. The transition probability (the probability of a particle starting at x being found at y after time t) has normal density $p(t, x, y)$ with mean x and variance at, where a is a positive constant. This is related to fundamental solutions of the heat equation as follows. For every rapidly decreasing function φ,

$$u(t, x) = \int_{-\infty}^{\infty} \varphi(y)\, p(t, x, y)dy$$

satisfies the heat equation (or the diffusion equation)

$$\frac{\partial u}{\partial t} = \frac{1}{2}\, a\, \frac{\partial^2 u}{\partial x^2},$$

and further, $\lim_{\substack{t\to 0\\ y\to x}} u(t, y) = \varphi(x)$.

More generally, one may study a problem where the constant a is replaced by a function $a(x)$, and the particle is subjected to a *drift* $b(x)$. (For example, the Ornstein-Uhlenbeck process is one in which $b(x) = -\rho x$, an elastic force pulling the Brownian particle towards the origin.) Then we have the equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\, a(x)\, \frac{\partial^2 u}{\partial x^2} + b(x)\, \frac{\partial u}{\partial x}.$$

In higher dimensions $a(x)$ is replaced by a covariance matrix $[a_{ij}(x)]$ whose entries are the *diffusion coefficients,* and $b(x)$ is now a vector.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{X_t\}_{t\geq 0}$ be a family of random variables with finite expectations. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be an increasing family of sub-σ-algebras of $\mathcal{F}$. If each X_t is measurable with respect to $\mathcal{F}_t$, and the conditional expectation $E(X_t|\mathcal{F}_s) = X_s$ for all $s \leq t$, then we say $\{X_t\}_{t\geq 0}$ is a *martingale.* (A common choice for $\mathcal{F}_t$ is the σ-algebra generated by the family $\{X_s : 0 \leq s \leq t\}$.)

The Brownian motion $\{B(t)\}_{t\geq 0}$ in $\mathbb{R}^d$ is a martingale. The connection goes further. Let φ be a C^2-function from $\mathbb{R}^d$ into $\mathbb{R}$. Then

$$X_\varphi(t) = \varphi\,(B(t)) - \int_0^t \frac{1}{2}\, \Delta\varphi\,(B(s))\, ds$$

is a martingale. It was shown by Itô and Lévy that this property characterizes the Brownian motion—any stochastic process for which $X_\varphi(t)$ defined as above is a martingale for every φ must be the Wiener process.

The *Martingale Problem* posed by Stroock and Varadhan is the following question. Let

$$\mathcal{A} = \sum_{i,j} a_{ij}(x)\, \frac{\partial^2}{\partial x_i \partial x_j} + \sum_j b_j(x)\, \frac{\partial}{\partial x_j}$$

be a second-order differential operator on $\mathbb{R}^d$. Can one associate with $\mathcal{A}$ a diffusion process with paths $x(t)$ such that

$$X_\varphi(t) = \varphi(x(t)) - \int_0^t (\mathcal{A}\varphi)\,(x(s))\, ds$$

is a martingale? (If $\mathcal{A} = \frac{1}{2}\,\Delta$ such a process exists and is the Wiener process.)

your answer. All this requires some regularity conditions on the coefficients. In the other method, due to Itô, you write down a stochastic differential equation (SDE) involving the Brownian motion $\beta(\cdot)$. Let σ be the square root of A. The associated SDE is

$$dx(t) = \sigma(x(t))d\beta(t) + b(x(t))dt; \qquad x(0) = x_0.$$

This equation has a unique solution under certain conditions. This gives a map Φ_{x_0} from Ω into itself and the image of the Wiener measure under this map is the diffusion we want. The conditions under which the two methods work overlap, but neither contains the other. The PDE method does not work very well if the coefficients are degenerate (the lowest eigenvalue of $[a_{ij}(x)]$ comes close to zero); the Itô method does not work if the coefficients are not Lipschitz. When they fail it is not clear whether it is the method or the problem that fails.

We wanted to establish a direct link between P and the coefficients without any PDE or SDE coming in. This is what we formulated as the Martingale Problem: Can you find a measure P on Ω such that

$$X_\varphi(t) = \varphi(x(t)) - \varphi(x_0) - \int_0^t (\mathcal{A}\varphi)(x(s))ds$$

is a martingale with respect to $(\Omega, \mathcal{F}_t, P)$, where $\mathcal{F}_t$ is the σ-field generated by $\{x(s) : 0 \leq s \leq t\}$ and

$$\mathcal{A} = \frac{1}{2} \sum_{i,j} a_{ij}\, \frac{\partial^2}{\partial x_i \partial x_j} + \sum_j b_j\, \frac{\partial}{\partial x_j}.$$

In this general formulation $\mathcal{A}$ can be replaced by any operator. This method works always when the other two do, and in many other cases. (Just as integration works in more cases than differentiation.)

I believe after the completion of your work the field of PDE started borrowing more from probability theory, while the opposite had been happening before.

No, we too use a lot of differential equations; we do not avoid them.

Between distribution solutions of differential equations and viscosity solutions, that came later, is there another layer of solutions that one may call probability solutions?

Yes, . . . , there is something to that. If you take expectations with respect to the probability measure that you have constructed, then you get solutions to certain differential equations. Usually they will be distributions but the conditions for the existence of a generalized solution may not be fulfilled. So you can call these a new class of generalized solutions, and they can be defined through martingales.

So, are we saying that for a certain class of equations there are no distribution solutions but there are solutions in this new probability sense?

It is difficult to say what exactly is a distribution solution. It is perfectly clear what a *classical* solution is. Then everyone can create one's own class—nothing special about the Schwartz class—in which a unique solution exists, as long as it reduces to the classical solution when that exists.

Talking of classical solutions, what is the first instance of a major problem in PDE being solved by probabilistic methods? Is it Kakutani's paper in which he solved the Dirichlet problem using Brownian motion?

Sure, that is the first connection involving probability, harmonic functions, and the Dirichlet problem.

What is it that Wiener did not know to make this connection? The relation between Brownian motion and the Laplace operator was obvious to everyone. Is it because things like the strong Markov property were not known at that time?

Also, Wiener was much more of an analyst. I don't think he thought as much of Brownian paths as of the Wiener measure. Unless you think of the paths wandering around and hitting boundaries you will not get the physical intuition needed to solve some of the problems.

Who were the other players in the development of this connection between probability and PDE?

Kac, for example, with the Feynman-Kac formula, surely knew the connections.

As you were working on this, who were the other people doing similar things?

In Japan: Ikeda, Watanabe, Fukushima, and many students of Itô. The brilliant Russian probabilist Girsanov. He died very young in a skiing accident. He had tremendous intuition. Another very good analyst and probabilist Nikolai Krylov, now in Minnesota. Then there were Ventcel, Freidlin, and a whole group of people coming from the Russian school. In the United States McKean who collaborated with Itô, and several people working in martingales: Burkholder, Gundy, Silverstein; and the French have their school too.

I am curious why hyperbolic equations are excluded from probability methods.

Except one or two cases. There are some examples in the work of Reuben Hersh. But they are rare. If you want to apply probability, there has to be a maximum principle, and not all equations have that. The maximum principle forces the order to be two, and the coefficients to be positive-definite.

Large Deviations

Your papers with Stroock seem to stop in 1974—I guess that is because he left New York—and there begins a series of papers with Donsker. How did that work start?

I was on sabbatical leave in 1972–73 and on my return Donsker asked me a question about the Feynman-Kac formula which expresses the solution of certain PDE in terms of a function-space integral. Asymptotically, this integral grows like the first eigenvalue of the Schrödinger operator, and this can be seen from the usual spectral theory. Donsker asked whether the variational formulas arising in large deviations and Laplace asymptotics and the classical Rayleigh-Ritz formula for the first eigenvalue have some connection through the Feynman-Kac representation. I thought about this and it turned out to be the case. This led to several questions like whether there are Sanov-type theorems for Markov chains and then for Markov processes; and if we did the associated variational analysis for the Brownian motion, would we recover the classical Rayleigh-Ritz formula. It took us about two years 1973–75 to solve this problem. The German mathematician Jürgen Gärtner did very similar work from a little different perspective.

What are your recollections about Donsker?

He had a large collection of problems, many of them a little off-beat. He had the idea that function-space integrals could be used to solve many problems in analysis, and in this he was often right. We worked together a lot for about ten years till he died, rather young, of cancer.

It is mentioned in Courant's biography that Donsker was his confidant when he worried about the direction the Institute was taking.

There was a special relationship between the two. I think a part of the reason was that most of the others at the Institute were *too* close to Courant—they were his graduate students or sons-in-law. (Laughs) Donsker was an outsider and Courant respected the perspective of some one like him. But in the end Courant did what he wanted to do in any case.

Almost all the reports say that the large-deviation principle starts with Cramér.

The idea comes from the Scandinavian actuarial scientist Esscher. He studied the following problem. An insurance company has several clients and each year they make claims which can be thought of as random variables. The company sets aside certain reserves for meeting the claims. What is the probability that the sum of the claims exceeds the reserve set aside? You can use the central limit theorem and estimate this from the tail of the normal distribution. He found that is not quite accurate. To find a better estimate he introduced what is called tilting the measure (Esscher tilting). The value that you want not to be exceeded is not the mean, it is something far out in the tail. You have to change the measure so that this value becomes the mean and again you can use the central limit theorem. This is the basic idea which was generalized by Cramér. Now the method is called the Cramér transform.

Is Sanov's work the first one where entropy occurs in large-deviation estimates?

It is quite natural for entropy to enter here. Sanov's and Cramér's theorems are equivalent. One can be derived from the other by taking limits or by discretizing.

The Shannon interpretation of entropy is that it is a measure of information. Is it that a rare event gives you more information than a common event and that is how entropy and large deviations are related?

. . . The occurence of a rare event gives you more information, but that may not be the information you were looking for. (Laughs)

What happens in large deviations is something like in statistical mechanics. You want to calculate the probability of an event. That event is a combination of various micro events and you are adding their probabilities. It is often possible to split these micro events into various classes and in each of these the probability is roughly the same. It is very small, exponentially small in some parameter. So each individual event has probability = exponential of $-n$ times something. That something is called the "energy" in physics. But then the number of micro events making an event could be large—it could be the exponential of n times something. That something is the "entropy". So the energy and entropy are fighting each other and the result gives you the correct probability. That is the picture in statistical mechanics. So, for me entropy is just a combinatorial counting. Of course you can say that if I pick a needle from a hay stack, then it gives me more information than picking a needle from a pin cushion. But then entropy is the size of the hay stack.

In the notes in their book Deutschel and Stroock say that Sanov's elegant result was at first so surprising that several authors expressed doubts about its veracity. Why was that so?

I do not know! . . . It is something like Bochner having been surprised [by our first theorem]. (Laughs)

One comment I heard about your work was that before you most people were concerned only with the sample mean, whereas you have studied many other kinds of objects and their large deviations.

Let me put it this way. Large deviations is a probability estimate. In probability theory there is only one way to estimate probabilities, and that is by Chebyshev's inequality.

Coin Tossing and Large Deviations

The popular description of the theory of large deviations is that it studies probabilities of rare events. Some simple examples may convey an idea of this. If you toss the mythical fair coin a hundred times, then the probability of getting 60 or more heads is less than .14. If you toss it a thousand times, then the probability of getting 600 or more heads reduces very drastically; it is less than 2×10^{-9}. How does one estimate such probabilities?

Let us enlarge the scope of our discussion to include unfair coins. Suppose the probability for a head is p, and let S_n be the number of heads in n tosses. Then by the weak law of large numbers (which just makes formal our intuitive idea of probability) for every $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = 0.$$

The elementary, but fundamental, inequality of Chebyshev gives an estimate of the rate of decay in this limit:

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

It was pointed out by Bernstein that for large n this upper bound can be greatly improved:

$$P\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq e^{-nh_+(\varepsilon)},$$

where for $0 < \varepsilon < 1 - p$,

$$h_+(\varepsilon) = (p+\varepsilon)\log\frac{p+\varepsilon}{p} + (1-p-\varepsilon)\log\frac{1-p-\varepsilon}{1-p}.$$

As $\varepsilon \to 0$, $h_+(\varepsilon)$ is approximately $\varepsilon^2/2p(1-p)$. For a fair coin $p = 1/2$ and this is $2\varepsilon^2$. So,

$$P\left(\frac{S_n}{n} \geq \frac{1}{2} + \varepsilon\right) \leq e^{-2n\varepsilon^2}$$

(In our example at the begining we had $\varepsilon = .1$, and for n we chose $n = 100$ and 1000.)

Bernstein's inequality is an example of a large-deviation estimate. It is optimal in the sense that

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq p + \varepsilon\right) = -h_+(\varepsilon).$$

The function h_+ is the *rate function* for this problem. The expression defining it shows that it is an entropy-like quantity.

Let us now go to a slightly more complicated situation. Let μ be a probability distribution on $\mathbb{R}$ and let $X_1, X_2, \ldots$ be independent identically distributed random variables with common distribution μ. The sample mean is the random variable

$$\overline{X}_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega),$$

and by the strong law of large numbers, as $n \to \infty$ this converges almost surely to the mean $m = EX_1$. In other words, for every $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|\overline{X}_n - m| \geq \varepsilon\right) = 0.$$

Finer information about the rate of decay to 0 is provided by Cramér's theorem. Let

$$k(t) = \log E(e^{tX_1}) \qquad \text{(the cumulant function)}$$

and

$$I(x) = \sup_t\,(tx - k(t)).$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(|\overline{X}_n - m| \geq \varepsilon\right) = -\inf_{|x-m|\geq\varepsilon} I(x).$$

In other words, as n goes to ∞, $P(|\overline{X}_n - m| \geq \varepsilon)$ goes to 0 like e^{-nc}, where $c = \inf\{I(x) : |x - m| \geq \varepsilon\}$.

The functions I and k are convex conjugates of each other, and I is the Fenchel-Legendre transform of k. Convex analysis is one of the several tools used in Varadhan's work on large deviations.

Sanov's Lemma

Let μ be a probability measure on a finite set $A = \{1,2, \ldots, m\}$ and let $X_1, X_2, \ldots$ be A-valued i.i.d random variables distributed according to μ. Each X_i is a map from a probability space $(\Omega, \mathcal{S}, P)$ into A such that $P(X_i = k) = \mu(k)$. The "empirical distribution" associated with this sequence of random variables is defined as

$$\mu_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\omega), \qquad \omega \in \Omega.$$

For each ω and n, this is a probability measure on A. (If among the values $X_1(\omega), \ldots, X_n(\omega)$ the value k is assumed r times, then $\mu_n(\omega)(k) = r/n$.) The Glivenko-Cantelli lemma says that the sequence $\mu_n(\omega)$ converges to μ for almost all ω.

Let $\mathcal{M}$ be the collection of all probability measures on A and let U be a neighbourhood of μ in $\mathcal{M}$. Since μ_n converges to μ, the probability $P(\omega : \mu_n(\omega) \notin U)$ goes to 0 as $n \to \infty$. Finer information about the rate of decay is given by Sanov's Lemma.

For every ν in $\mathcal{M}$, the relative entropy of ν with respect to μ is defined as

$$H(\nu | \mu) = \begin{cases} \sum_{i=1}^{m} \nu(i) \log \frac{\nu(i)}{\mu(i)} & \text{if } \nu \ll \mu \\ \infty & \text{otherwise.} \end{cases}$$

Let $c = \inf_{\nu \notin U} H(\nu | \mu)$. Then

$$\lim_{n \to \infty} \frac{1}{n} \log P(\omega : \mu_n(\omega) \notin U) = -c.$$

For this problem $I(\nu) = H(\nu | \mu)$ is the rate function, and our probability goes to 0 at the same rate as e^{-cn}.

In this case we have studied limits of measures instead of numbers. This is what Varadhan calls the LDP at the *second* level. At the third, and the highest, there are LDP's at the stochastic-process level.

The usual Chebyshev inequality applies to second moments, Cramér's applies to exponential moments. You compute the expectation of some large function and then use a Chebyshev-type inequality to control the probability. That is you control the integral, and the probability of the set where the integrand is big cannot be very large. So people have concentrated on the expectation of the object that you want to estimate. That stems from the generating-function point of view. My attitude has been slightly different. I would start from the Esscher idea of tilting the measure. His exponential tilting is just one way of tilting. It works for independent random variables. If you have some process with some kind of a model and you are interested in some tail event, then you change the model so that this event is not in the tail but near the middle. The new model has a Radon-Nikodym derivative with respect to the original model, and you can use a Jensen inequality to obtain a lower bound. This may be very small. Then you try to optimise with respect to the choice of models. If you do this properly, then the lower bound will also be the upper bound.

What kind of optimisation theory is used here?

It depends on the problem. For example for diffusion in small time for Brownian motion on a Riemannian manifold it is the geodesic problem. If you want to get Cramér's theorem by Sanov-type methods, the ideas are similar to those in equilibrium statistical mechanics. The Lagrange-multiplier method is an analogue of the Esscher tilt. If you want a Sanov-type theorem not for i.i.d. random variables but for Markov chains, then a Feynman-Kac-like term is the Esscher tilt. It is the same idea in different shapes.

It is said that whereas in the classical limit theorems the nature of individual events is immaterial, in the large-deviation theory you do have to look at individual events.

I guess what people mean to say is that in the large-deviation theory you solve an optimisation problem. So events near the optimal solution have to be examined more carefully.

You and Donsker have a series of papers on the Wiener sausage (a tubular neighbourhood of the Brownian motion) where you have asymptotic estimates of its volume. What is the problem in physics that motivated this study?

The Laplace operator on $\mathbb{R}^d$ has a continuous spectrum. If you restrict to a box of size N it has a discrete spectrum. As you let N go to infinity and count the number of eigenvalues in some range and normalize it properly this goes to a limit, called the density of states. If you add a potential, you get another density of states. A special class of potentials of interest is where you choose random points in $\mathbb{R}^d$ according to a Poisson distribution, put balls of small radius around them where the potential is infinite. There are two parameters now, one is the density (of the Poisson distribution of traps) and the other is the size of the traps. Now you want to compute the density of states. This is done better if you go to the Laplace transform. Then it becomes a trace calculation, and by the Feynman-Kac formula this can be done in terms of the Brownian motion. Entering the infinite trap means the process gets killed. So we are looking at a Brownian motion that must avoid all these traps—which are distributed at random. This problem was posed by Mark Kac.

You are looking at the behaviour of density of states at low levels of energy. That is the same as the behaviour of the Laplace transform for large t. So you want to know what is the probability that the Brownian motion avoids these traps for a very long time. The conjecture made by the Russian physicist Lifschitz was that this probability decays like $\exp(-c\, t^{d/(d+2)})$.

It is easy to calculate the probability of having a big sphere with no traps in it. Then you calculate the probability that a Brownian motion that is in this sphere stays there for ever, or at least up to time T. This can also be easily calculated and turns out to be like $e^{-\lambda T}$, where λ is the first eigenvalue of the Laplacian on this sphere with the Dirichlet boundary condition. You multiply the two probabilities to get the answer.

Are Monte Carlo methods relevant to this kind of problem?

No, they are not useful here. As this theory tells you, if a Brownian path has found a safe territory without traps, then it must try to stay there. A typical path will not do that. So the contribution comes from paths that are not typical, and Monte Carlo methods can not simulate such paths.

In another series of papers you and Donsker study the polaron problem. Can we describe this to our readers, even if loosely?

Our work started with a question posed to us by E. Lieb. It comes from a problem in quantum statistical mechanics and the work of Feynman. In the usual Feynman-Kac formula you have to calculate the expectation of integrals like

$$\exp\left[-\int_0^t V(x_s)ds\right].$$

In the polaron problem you have a quantity depending on a double integral:

$$A(t,\alpha) = E\left\{\exp\left[\alpha\int_0^t\int_0^t e^{-|\sigma-s|}V(x_\sigma - x_s)ds\, d\sigma\right]\right\},$$

and you have to evaluate

$$G(\alpha) = \lim_{t\to\infty}\frac{1}{t}\log A(t,\alpha).$$

This is complicated, and there was a conjecture about the value of the limit of $G(\alpha)/\alpha^2$ as $\alpha \to \infty$.

The potential $V(x_\sigma - x_s)$ is something like $1/\|x_\sigma - x_s\|$. So the major contribution to the integral comes from those Brownian paths that tend to stay near themselves. Typical paths do not behave like that. So the asymptotics require a large-deviation result quite like the one in our work on the Wiener sausage.

Here is where large deviations at various levels come in. At the first level you have a sequence of random variables and you look at their means. At the next level you look at $\delta_{X_1} + \cdots + \delta_{X_n}$ and think of it as a random measure. That is like in Sanov's theorem. I will call that level two. From the higher level you can project down to a lower level using a contraction principle. It is like computing the marginals of a bivariate distribution. At level three you think of the measures as a stream and in addition to $\delta_{X_1}, \delta_{X_2}, \ldots$, look at $\delta_{(X_1,X_2)}, \delta_{(X_2,X_3)}, \ldots$, and then at $\delta_{(X_1,X_2,X_3)}, \delta_{(X_2,X_3,X_4)}, \ldots$, and so on with tuples of length k. Now you can first let n, and then k, go to ∞. This is *process level* large deviation. Here I am considering the following question: I draw a sample from one stochastic process and ask what is the probability that it looks like a sample drawn from a different process? This is what I call a level-three large-deviation problem. The rate function for this can be computed and turns out to be the Kolmogorov-Sinai entropy. This is used in the solution of the polaron problem.

Is this the highest level, or can you go beyond?

Level three is the highest because the output and the input are in the same class—both are stochastic processes. (At the first level, for example, the input is random variables and you consider quantities like their means.)

The interesting thing is that at level three the rate function is universal. You take any two processes P and Q and the rate function is the Kolmogorov-Sinai entropy between them. So at level three there is a universal formula. They are different at the lower level because the contraction principles are different.

Let me turn to lighter things now. In 1980 I attended a talk by Mark Kac. He began by saying that Gel'fand, who was three months older than him, advised him that as you grow old you should talk of other people's work and not your own. And he said he couldn't do better than talking of the Donsker-Varadhan asymptotics. Kac was 66 at that time, your present age. If you were to follow that advice, whose work will you talk about?

In probability theory the most exciting work in the last ten years has been on SLE (stochastic Loewner equations). This is mostly due to Wendelin Werner, Greg Lawler, and Oded Schramm.

Kac says a large part of his scientific effort was devoted to understanding the meaning of statistical independence. For Courant, a very large part of the work is around the Dirichlet principle. Is there one major theme underlying your work?

It is hard for me to say something in those words. I can talk of my *attitude* to probability. I don't like it if I have

The Feynman-Kac Integral

The Lagrangian of a classical mechanical system $L(x, \dot{x}) = \dot{x}^2/2 - V(x)$ has its quantum mechanical counterpart $\frac{1}{2}\frac{\partial^2}{\partial x^2} - V$. The wave function $\psi(t, x)$ is a solution of the Schrödinger equation

$$\frac{1}{i}\frac{\partial\psi}{\partial t} = \frac{1}{2}\frac{\partial^2\psi}{\partial x^2} - V(x)\psi,$$
$$\psi(0,x) = \varphi(x).$$

Feynman's solution to this is in the form of a curious function-space integral

$$\psi(t,x) = \int_{\Gamma_x}\exp\left\{i\int_0^t\left[\frac{(\dot{x}_\tau)^2}{2} - V(x_\tau)\right]d\tau\right\}\varphi(x_\tau)\prod_\tau dx_\tau,$$

where Γ_x is the space of all paths

$$\{x_\tau : 0 < \tau \le t,\ x_0 = x\}$$

and $\Pi_\tau\, dx_\tau$ is a "uniform measure" on $\mathbb{R}^{(0,t]}$. For a mathematician such a measure does not exist.

Kac observed that if the i in Schrödinger's equation is taken out, one gets the heat equation. A solution similar to Feynman's now reduces to a legitimate Wiener integral

$$\psi(t,x) = \int_{\Gamma_x}\exp\left\{-\int_0^t V(x_\tau)d\tau\right\}\varphi(x_\tau)dW_x.$$

These function-space integrals occur very often in the work of Donsker and Varadhan.

The Wiener Sausage

This may give the best example of the reach, the power, and the depth of the work of Donsker and Varadhan.

Let β be a Brownian path in $\mathbb{R}^d$. The Wiener sausage is an ε-tube around the trajectory of β till time t; i.e.,

$$S_t^{(\varepsilon)}(\beta) = \{x \in \mathbb{R}^d : |x - \beta(s)| < \varepsilon \text{ for some } s \text{ in } [0, t]\}.$$

Let V be the volume (Lebesgue measure) in $\mathbb{R}^d$, and W the Wiener measure on the space X of continuous paths from $[0,\infty)$ into $\mathbb{R}^d$. For a fixed positive number c, let

$$\mathcal{A}_t^{(\varepsilon)} = \int_X \exp\,[-c\; V(S_t^{(\varepsilon)}(\beta))]\;W(d\beta).$$

Varadhan explains in this conversation why physicists are interested in studying the behaviour, as $t \to \infty$, of $\mathcal{A}_t^{(\varepsilon)}$. Donsker and Varadhan proved the marvelous formula

$$\lim_{t\to\infty} \frac{1}{t^{d/(d+2)}} \log \mathcal{A}_t^{(\varepsilon)} = -k(c),$$

where

$$k(c) = \frac{d+2}{d}\left(\frac{2\lambda}{d}\right)^{d/(d+2)} c^{2/(d+2)},$$

and λ is the first eigenvalue of the Laplace operator in the unit ball of $L_2(\mathbb{R}^d)$.

to do lots of calculations without knowing what the answer might turn out to be. I like it when my intuition tells me what the answer should be and I work to translate that into rigorous mathematics.

Where are these problems coming from?

Usually from physics. Physicists have some intuitive feeling for the answer and mathematics is needed to develop that.

Do you talk often to physicists?

Yes, For the last few years I have been working on hydrodynamic scaling. I often talk to J. Lebowitz at Rutgers, and to others.

What is your work connecting large deviations to statistical mechanics, thermodynamics and fluid flow?

This may loosely be described as non-equilibrium statistical mechanics. The ideas go back far; for example in the derivation of Euler's equations of fluid dynamics from classical Hamiltonian systems for particles.

You ignore individual particles and look at macroscopic variables like pressure, density, fluid velocity. These are quantities that are locally conserved and vary slowly, and others that wiggle very fast but reach some equilibrium. There are ergodic measures indexed by values for different conserved quantities. These are local equilibria or Gibbs states. If your system is not in equilibrium, it is still locally in equilibrium. So certain parameters that were constants earlier are now functions of space and time. You want to write some differential equations that describe how these evolve in time. Those are the Euler equations.

How do large deviations enter the picture?

In the classical model there is no noise. I can't touch things that have no noise. (Laughs) Think of a model in which after a collision who gets to leave with what momentum is random.

What are the applications of these ideas in areas other than physics. I believe there are some in queuing networks. How about economics?

There are some applications. I do know some people working in mathematical finance—I don't know whether that is real finance. (Laughs) But it is conceivable that these ideas are used. These days you can write "options" on anything. Let us make up a problem. I write an option that if a certain stock rises to $1000, then I will pay you the average closing price for the last 90 days. So what I pay does not depend just on the current value but also on the past history how it got there. But the past history counts only if it reached this high value. Therefore you want to know first the probability that the stock will reach this high value, and then if it did so what is the most likely path through which it will reach this value. For this you have to solve a large-deviation problem.

There is a joke that two economists who got the Nobel Prize for their work on stock markets lost their money in the stock market.

(Laughs) I have no idea. But whatever they lost they made up in consultancy.

It is remarkable that the equations of Brownian motion were first discovered by Bachelier in connection with the stock market, and only later by Einstein and Smoluchowski. Are there many instances of this kind where social sciences have a lead over physical sciences?

I think many statistical concepts, now used in biology, were first discovered in the context of social sciences.

Another major collaborator of yours has been G. Papanicolaou. Whereas your work with Stroock and Donsker was concentrated over a few years to the exclusion of other things, here it is spread over several years. Is this the beginning of your interest in hydrodynamic limits? Can you summarize it briefly?

George and I were at a conference in Luminy in Marseille. We always went for a walk after lunch. Luminy is on top of a hill and there is a steep walk down to the sea. We walked down and up for exercise after lunch, and discussed mathematics. George explained to me this problem about interacting Brownian motions. You have a large number of Brownian particles that come together and are repelled from each other. The density of paths satisfies a nonlinear diffusion equation. You have to compute some scaling limits for this system. I was intrigued by the problem as it looked like a limit theorem, and I always thought I should be able to prove a limit theorem, especially if everyone believed it was true. But when I looked at it closely there was a serious problem—of the kind I mentioned before. How to prove certain quantities are in local equilibrium. Pretty soon I found a way of doing it. In some sense large deviations played a role. If you are in equilibrium you can compute probabilities. If you have a small probability with respect to one measure and if there is another measure absolutely continuous with respect to this, then the probability in this measure is small if you have control over the Radon-Nikodym derivative. This derivative is given by the relative entropy. In statistical mechanics the relative entropy of nonequilibrium with respect to the equilibrium is of the order of the volume. So events that had probabilities that were super-exponentially small in the equilibrium case still have small probabilities in the nonequilibrium case. My idea was that to control something in the nonequilibrium case you control it very well in the equilibrium. At this time Josef Fritz gave a seminar at the Institute where he was looking at a different problem on lattice models. Some ideas from there could handle what we had been unable to do in our problem. That is the history of my first entry into this field.

You have worked with several collaborators. Do you have any advice on collaborations?

I think you should talk to a lot of people. A part of the fun of doing mathematics is that you can talk about it. Talking to others is also a good source of generating problems. If you work on your own, no matter how good you are, your problems will get stale.

Have you thought of problems coming from areas other than physics?

Problems that come from physics are better structured. In statistical mechanics one knows the laws of particle dynamics and can go from the micro level to the macro level where observations are made. There is a similar situation in economics where you want to make the transition from individual behaviour to what happens to the economy. This is fraught with difficulties. Whereas we know which particles are interacting, we do not know how persons interact and with whom. The challenge is to make a reasonable model. Physics is full of models.

How about biology, does it have good mathematical models?

It seems to me that at this moment most of their problems are statistical in nature, . . . like data mining.

On Probabilists

I would like to talk a little about some major figures in probability theory in the 20th century, and get a feel of the recent history as you see it.

Does the modern theory of probability begin with Kolmogorov, as is the general view?

Kolmogorov was really responsible for making it a legitimate branch of mathematics. Before that it was always suspect as mathematics, something that was intuitively clear but was definitely not mathematics. The person who contributed the most to probablistic ideas of the time was Paul Lévy, but he was considered an engineer by the French.

What was Wiener's role? In 1923 Wiener wrote a paper Differential space *and several years before Kolmogorov he introduced a measure on a function space.*

Wiener measure was just one particular measure. Kolmogorov advanced the view that, very generally, the models in probability or statistics have legitimate measures behind them. After that it became easier to make new models. Kolmogorov must have known for several years what is in that book and decided to write it at some point. There is perhaps nothing there that he discovered just before he wrote it.

In the preface to Itô's selected papers you and Stroock say Wiener (along with Paley) was the Riemann of stochastic integration.

Yes, though he did it more by duality and completion arguments. The Wiener integral is very special, but it must have been the motivation for Itô's more general theory.

Do you think Cramér's Mathematical Methods of Statistics (1945) *did for statistics what Kolmogorov's book had done for probability?*

It is quite unreadable! When I was a student there were not any statistics books that were readable. The best I found were some lecture notes on statistical inference by Lehmann from Berkeley. Statistics has two aspects to it. One is computing sampling distributions of various objects, and this is

The Ubiquitous Brownian Motion

A gambler betting over the outcome of tossing a coin wins one rupee for every head and loses one for every tail. Let $N(n)$ be the number of times that he is a net gainer in the first n tosses.

For $a > 0$ what is

$$\lim_{n\to\infty} P\left(\frac{N(n)}{n} < a\right)?$$

The answer is that this limit is equal to the Wiener measure of the set of Brownian paths (in the plane) that spend less time than a in the upper half-plane.

This very special example is included in a very general "Invariance Principle" proved by Donsker in his Ph.D. dissertation.

Let $(\Omega, \mathcal{F}, P)$ be any probability space and $X_1, X_2, \ldots$ i.i.d. random variables on it with mean 0 and variance 1. For each $n \geq 1$ associate with every point ω in Ω an element γ_n of $C[0,1]$ as follows. Let

$$S_n(\omega) = X_1(\omega) + \cdots + X_n(\omega).$$

Let $\gamma_n(0) = 0$; for $k = 1,2, \ldots, n$, let $\gamma_n(k/n) = S_k(\omega)/\sqrt{n}$, and define $\gamma_n(t)$ for other values of t as a piecewise linear extension of this.

This defines a map φ_n from Ω into $C[0,1]$ given by $\varphi_n(\omega) = \gamma_n$. Let $\mu_n = P \circ \varphi_n^{-1}$ be the measure induced on $C[0,1]$ by φ_n.

The Donsker Invariance Principle says that the sequence μ_n converges weakly to the Wiener measure on $C[0,1]$.

just an exercise in multiple integrals. This is really more analysis than statistics. The other aspect, real statistics, is inference. Very few books did that. At that time many expositions came from Berkeley.

We already talked of Paul Lévy. Though you said he was thought of as an engineer, he was proving theorems to the effect that the set where the typical Brownian path intersects the real axis is homeomorphic to the Cantor set. How is this kind of thing useful in probability?

There is an interesting way of looking at Brownian motion. The zero-set of a Brownian motion is a Cantor set. So there is a measure that lives on the Cantor set, and parametrised by that measure it becomes an interval. This gives a map from the Cantor set into an interval; under this map several open sets are closed up. You can try to "open" them up again. This means there are randomly distributed points on the interval where you "open up" things using Brownian paths that wandered into the upper-half or the lower-half plane. These are "Brownian loops" or "excursions". Is it possible to reconstruct the Brownian motion by starting with the Lebesgue measure on the interval and opening random intervals with excursions? This "excursion theory" is a beautiful description of the Brownian motion. Lévy saw all this, and later Itô perfected it.

Among other important names there are Khinchine, Feller, . . .

Khinchine did probability, number theory, and several other things. I think he would have thought of probability as an exercise in analysis.

Did Feller like analysis? He seems to have been critical of Doob for making probability too abstract.

Feller's work is all analytical. For example, the law of the iterated logarithm is hard analysis, as is his description of one-dimensional diffusion. It is not that he did not like analysis; he thought Doob's book was too technical. In his own book he cites Doob's book among "books of historical interest". That made Doob very angry.

How do you assess Doob's Stochastic Processes (1953)?

My view of Doob's book is that it is very uneven. Some parts like martingales are very original. But if you look at a book you should compare the number of pages with what is proved in those pages. Doob's book is large, over 600 pages, but does not prove that much.

In one interview he says he intended to minimise the use of measure theory because probabilists thought it was killing their subject. But then he found the "circumlocutions" became so great that he had to rewrite the whole book. Is it that even that late probabilists did not want measure theory to intrude into their subject?

I don't think so. But you see, in probability what do you do with measure theory? The only thing you need is the dominated convergence theorem, what else? It is always in the background. But to say that you were avoiding measure theory in an advanced book sounds strange.

No, Doob says he tried to avoid it because probabilists thought it was killing their subject.

That is only because they allowed it to. Let us take Doob's own book, for example. One of the concepts in the study of stochastic processes is the notion of separability. That is where measure theory really intrudes. The problem is that sets depending on more than a countable number of operations (like those involving a supremum) are not measurable. If you change a random process on a set of measure zero nobody will notice it. But if for each t you change it on a set of measure zero, then as a function you change it on the union of these sets which is no longer of measure zero. So one has to be careful in choosing certain sets and functions from an equivalence class. Of course if you don't know measure theory this does not bother you. (Laughs) But soon you notice you can choose versions that are reasonable, and then you don't have to worry about the matter. So you should know "separability" can be a problem, learn to avoid it, and then avoid it forever. Doob, on the other hand, makes a whole *theory* out of it. That is because you let the measure theory intimidate you.

At another place Doob attributes the popularity of martingales to the catchy name. Do you believe that the name "martingale" made the theory popular?

I don't think so. Maybe when Doob started the theory, no one cared. But then it turned out to be a very useful concept. Today even people on Wall Street know of martingales. (Laughs)

Let us come to Itô now. In your preface to his Selecta *you and Stroock say that if Wiener was the Riemann of stochastic integration, then Itô was its Lebesgue. Is that an accurate analogy? I thought Wiener's integral is very special, while Itô's is much more general.*

. . . I am sure that was written by Stroock; it is not my style. If you read Lévy's work you will get some idea of what a diffusion should be like. It is locally like a Brownian motion, but the mean and the variance depend on where you are. It is clear from Wiener's integral that you are changing the variances by a scale factor, but the factor depends on time and not on space. If you want it to depend on both time and space, you get an equation, and that is a stochastic differential equation. This is what Itô must have seen; and he made precise the ideas of Wiener and the intuition of Lévy, by defining this equation.

Do you have any special memories of Mark Kac?

Oh, he was a lot of fun. He would often call us up and invite us to come to Rockefeller University where he would talk of many problems. He had a tremendous collection of problems.

Is there anyone else you would like to mention?

Dynkin made big contributions. He started out in Lie groups and came to probability a little late, around 1960. Then he founded a major school on Markov processes. I learnt a lot from his work, from his books, papers, and expository articles. Around 1960 he wrote a beautiful paper in *Uspehi* on problems of Markov processes and analysis, that I remember very well.

What, in your view, is the most striking application of probability in an area far away from it?

Although I am not quite familiar with it, it is used in law some times. (Laughs)

I meant an application in mathematics, but in an area not traditionally associated with probability. For example, Bismut's purely probabilistic proof of the Atiyah-Singer index theorem. Is that unexpected?

Well, McKean had already done some work on it with Singer and it was clear that probability or, at least, the fundamental solution of the heat kernel plays a role. Since the Laplacian operator is involved, the role of probability is not far fetched.

Is there an area where you would not expect probability to enter, but it does in a major way?

. . . Number Theory. For example the work of Furstenberg. In PDE probability now plays a major role but that is not unexpected. If you use martingales, the maximum principle just reduces to saying that the expectation of a nonnegative function is nonnegative.

The Prize

Did you anticipate your being chosen for the Abel Prize?

No, not at all.

I have read that potential Nobel Prize winners are usually tense in October and jump every time their phone rings at 6 AM. I also heard a talk by a winner who told us that when they phone you about the Nobel Prize they have with them someone whom you know so that you are sure no one is pulling your leg. How was it for you?

They called me at 6:10 A.M., gave me the news and said I should not tell anyone till 7 when they would announce it at a news conference in Oslo. They told me there would be a live interview on the Norwegian radio.

Were you allowed to tell your wife?

I told my wife.

My question was whether you were allowed to. The Fields Medal winners have to be told in advance because the awards are announced just before they are actually given, and they are told they can tell their spouses but no one else. That must be difficult for them!

But this was only between 6:30 and 7. You cannot call too many people anyway. I did not tell anyone except my wife till 7.

There has been some discussion about the purpose of such prizes—beyond honouring an individual. Lennart Carlson said they draw public attention to the subject. When I entered college, physics was the most prestigious subject. Then Hargobind Khorana got a Nobel Prize, and for a few years many top students in India wanted to study biochemistry. I see little chance of mathematics displacing management even after your Abel Prize.

I think it does make the subject more visible, and may attract a few individuals who otherwise had not thought about it.

At 67 you are the baby among the Abel Prize winners. Itô got the first Gauss Prize last year when he was about 90. Is it good to have age limits for such prizes?

I don't think it is important. Now you have several prizes of high level. There are the Wolf Prize, the Crawfoord Prize, the Kyoto Prize, the King Faisal Prize, . . . And although they don't say it, they rarely go to the same individual.

Still most of the other prizes have not caught the public imagination in the same way as the Nobel Prize.

The Nobel Prize is a century old and has got etched into people's consciousness.

One purpose the prizes could be made to serve is that a serious attempt is made to explain the winner's work to people.

It is hard to explain what a mathematician has done, compared to a new cure for cancer or diabetes.

But we don't even explain it to mathematicians. At the ICM's there are talks on the work of the Fields Medalists.

Some Thoughts on Prizes

The Nobel is awesome to most of us in the field, probably because of the luster of the recipients, starting with Roentgen (1901). The Prize gives a colleague who wins it a certain aura. Even when your best friend, one with whom you have peed together in the woods, wins the Prize it somehow changes him in your eyes.

I had known that at various times I had been nominated. . . .

As the years passed, October was always a nervous month, and when the Nobel names were announced, I would often be called by one or another of my loving offspring with a "How come . . . ?" In fact, there are many physicists—who will not get the Prize but whose accomplishments are equivalent to those of the people who have been recognized. Why? I don't know. It's partly luck, circumstances, the will of Allah.

When the announcement finally came, in the form of a 6 A.M. phone call on October 10, 1988, it released a hidden store of uncontrolled mirth. My wife, Ellen, and I, after very respectfully acknowledging the news, laughed hysterically until the phone started ringing and our lives started changing.

—Leon Lederman, in *"The God Particle"*

I think it's a good thing that Fields Medals are not like the Nobel Prizes. The Nobel prizes distort science very badly, especially physics.

The difference between someone getting a prize and not getting one is a toss-up—it is a very artificial distinction. Yet, if you get the Nobel Prize and I don't, then you get twice the salary and your university builds you a big lab; I think that is very unfortunate.

But in mathematics the Fields Medals don't have any effect at all, so they don't have a negative effect.

I found out that in a few countries the Medals have a lot of prestige—for example, Japan. Getting a Fields Medal in Japan is like getting a Nobel Prize. So when I go to Japan and am introduced, I feel like a Nobel Prize winner. But in this country, nobody notices at all.

—Michael Atiyah
The Intelligencer, 1984

On Bach

The Stroock-Varadhan book proceeds on its inexorable way like a massive Bach fugue.

—David Williams

(Book Review in *Bull. Amer. Math. Soc.*)

There's nothing to it. You just have to press the right keys at the right time with the right force, and the organ will make the most beautiful music all by itself.

Johann Sebastian Bach

Some are very good and others do not convey much even to a competent mathematician from a neighbouring field.

To explain something very clearly and very well takes a lot of effort, thought and time. It is not an easy job.

You have been an editor, for many years, of Communications on Pure and Applied Mathematics, *and of the Grundlehren series. Do you make any effort to make your authors write better?*

In a journal it is difficult to do so. But for books in the Grundlehren series we are very meticulous. We try to have books from which people can learn.

I began our conversation with India and would like to end with it. You left India at the age of 23. Do you think you could have done something more for mathematics in India?

. . . Perhaps I could have. But these things are complicated. Since my family and my work are here, I could at best make short visits and give some lectures. Some students could then keep in contact, or come here. Some of that was done in the 1970's when we had more scholarships. But then our funds for these things were reduced.

I have a very specific question here. If you see a person like S. S. Chern, he played an enormous role in grooming mathematicians of Chinese origin, even before China opened up. Perhaps Harish-Chandra could have played a similar role for Indians but he didn't. It could be that the two personalities were different. Several mathematicians of Chinese origin became outstanding differential geometers. Nothing like that happened to Indians in the fields of representation theory or probability. Is this something you could have done, or would like to do in the future?

I don't know. I think the Indian psyche is different from the Chinese. The Chinese like the role of an emperor much more and Chern enjoyed that role. Indians seem to be much more individualistic, and even within India I do not see anyone with that much influence.

What are your other interests?

I like sports. I play either tennis or squash for one hour every day. I listen to music, though I do not have special knowledge of it. I like Karnatak music. I like to watch movies, I see a lot of English as well as Tamil movies.

Are these the masala *movies in Tamil?*

Yes, a lot of them. These days you have DVD players and you can fast-forward whenever you want to. I also read Tamil books, both new and old.

Nobel Prize winners are often asked a silly question: what will you do with the money?

I haven't made detailed plans but I have a rough idea. I would like to put some of it for public good. My parents' last residence was in Madras (Chennai) and they were associated with a school. There is also a hospital there which is doing good work. I would like to help such ventures. Perhaps I will use one third of the prize money for that. Then, of course, I have to pay taxes—nearly one third of it. The remaining one third I will keep for my own use.

Thank you very much for giving me so much of your time, and best wishes for the Award Ceremony next week.