Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 31–35
http://www.ssca.org.in/journal

CR Rao's Shadows on Our Academic Journey

Sumanta Basu[1] **and Jyotishka Datta**[2]
[1] *Cornell University, Ithaca, United States*
[2] *Virginia Polytechnic Institute and State University, Blacksburg, Virginia*

Received: 20 July 2024; Revised: 28 July 2024; Accepted: 31 July 2024

"*One could say if Europe is the mother of differential calculus based on deterministic analysis, India could be called the mother of statistics. When I think of modern statistics, Dr. C.R. Rao features on the top of the list. He once said that "statistics is the technology of finding the invisible and measuring the immeasurable.*"

- Abdul Kalam, Bharat Ratna (past president of India)

We, the authors of this short note, met Prof. C. R. Rao when we were students at the Indian Statistical Institute (ISI), Kolkata. Prof. Rao spent the morning of his visit at the hostel dining room, having breakfast with us, and sharing anecdotes and stories. Fifteen years have passed since that morning, and we have spent these years almost entirely in the United States, pursuing higher education and academic careers. Undoubtedly, our education and careers owe a great deal to the heritage of Statistics in India, and the growth of Statistics over the better part of the last century, both of which were influenced greatly by Prof. C. R. Rao. In what follows, we attempt to capture a few things we remember as lessons from his life and anecdotes that continue to guide and shape us today.

Integration of research and teaching

When Prof. Rao joined Indian Statistical Institute around 1942, he was one of the 15 or so technical workers at the institute, led by P. C. Mahalanobis, known as the 'Professor' at ISI, who did teaching and some research. There were not many textbooks on Statistics yet, and the teachers labored tirelessly turning original research papers into teaching materials, translating the state-of-the-art into classroom materials. During this time as a 'technical apprentice', Rao discovered some of the foundational results in classical statistics, that are still taught in any undergraduate statistics inference course anywhere in the world. One of these was the famous Cramér-Rao inequality that provides a lower bound on the variance of unbiased estimators, indicating the minimum possible variance that any unbiased estimator of a parameter can achieve. As Rao recounts[1], he was presenting a large sample result by Fisher regarding the lower limit of the error of an estimate, and a student in his class[2]

[1] The authors of the present article were fortunate to hear this story from Prof. Rao when he visited the Boys' hostel at ISI Kolkata.

[2] DasGupta (2024) identifies this student as V. M. Dandekar.

Corresponding Author: Jyotishka Datta
Email: jyotishka@vt.edu

asked if a similar result would hold true for small sample size, which is often the case in real application. Rao went home and worked out the solution the same night and answered the student next morning (Champkin, 2011). Due to wartime restrictions, it took two years for Rao's paper to finally appear in the *Bulletin of the Calcutta Mathematical Society* (Rao, 1945). In Sweden, Harold Cramér had derived an analogous result, and Neyman linked the two scientists' name. The beauty of the result is that it holds true for any data distribution under mild regularity conditions.

This story serves as a reminder of why cutting-edge research questions should be blended with teaching statistics, and why one should always encourage students to critically engage with the subject and ask good and possibly difficult questions.

Do not bury the lede

The 1945 paper Rao (1945) has another seminal result – Rao-Blackwellization – which provides a simple way to improve an estimator by conditioning on a sufficient statistics. This, too, was being discovered contemporaneously by David Blackwell in 1947, and the names were combined by Joseph Berkson. Interestingly, as Champkin (2011) points out, Rao did not mention this result in the Introduction, which probably contributed towards it being discovered later. Jokingly, Rao said "It was my first paper, and I was not aware that the introduction is generally written for the benefit of those who do not want to read the paper." This story has now become part of the folklore in our community, and it has a profound implication. For us, it is a reminder of one of the basic rules of journalism: "do not bury the lede", *i.e.*, writers should present the most important information at the beginning of an article or news story. However, we should note here that missed attribution is not an uncommon phenomenon in statistics or machine learning, and while Rao got his credit for the seminal results, Stiegler's law of eponymy (Stigler, 1980) is still commonplace (and perhaps worsening?).

Geometric intuition

It is worth noting that the seminal 1945 paper was written when Rao was only 25 years old, and yet to obtain a PhD. The paper not only introduced Cramér-Rao lower bound and Rao-Blackwellization, it 'introduced differential geometry to statistical inference' and opened the field of information geometry. While presenting a geometric interpretation of the parametric probability densities, Rao defined a 'population space' where Fisher information is used as a distance between densities and the invariant measure turns out to be the square root of the information matrix: an idea containing the essence of Jeffreys' prior. As Efron notes in 'C. R. Rao's Century' (Efron *et al.*, 2020), 'A notable characterization of Rao's work, and Fisher's too, is its reliance on geometric intuition, substituting what, for me, are vivid pictures in place of rote algebra and analysis.' Such 'geometric intuition' has probably been a distinguished characteristic of both the authors' education: the best parts of our theoretical or methodological pursuits were influenced by the geometric intuition about the low-dimensional structures in high-dimensional spaces.

LSI and Impact on ISI education

A major legacy of Prof. C. R. Rao is his iconic Wiley textbook 'Linear Statistical Inference and Its Applications' (LSI) (Rao, 1965). Its encyclopedic breadth aside, what makes this book special is that Rao managed to concisely present and contextualize all the abstract mathematical machinery required, not just to *learn*, but to also *develop*, statistical methods. If we view Statistics as a vehicle that researchers use to advance scientific knowledge, LSI can play the role of not only a driver's manual, but also a mechanic's manual. The book starts with vector spaces and covers linear algebra and probability before introducing statistical theory. Over the last 50 years since its publication, LSI has been used and is still used by statisticians worldwide. Don Rubin once said, "Bill suggested that I turn to Rao's famous textbook on linear models for its straightforward mathematical clarity, at least relative to some other "math-stat" texts that were in use at the time. Being an official dinosaur, I still use it as a "go to" resource" (Efron *et al.*, 2020). To quote Efron: "When the fat second edition of Rao's magisterial book on linear statistical inference arrived on my desk, it was a big event in the department, not just for me (The book is still in use, though it has gotten a little beat up)" (Efron *et al.*, 2020). One can get almost all the fundamental concepts in probability, linear and abstract algebra, distribution theory, linear models, the theory of least squares and analysis of variance, large sample techniques, and multivariate analysis between the two covers of LSI. In fact, in the preface of LSI, Rao states, "*the aim has been to provide in a single volume a full discussion of the wide range of statistical methods useful for consulting statisticians and, at same time, to present in a rigorous manner the mathematical and logical tools employed in deriving statistical procedures, with which a research worker should be familiar.*"

Personally, this book has served the role of a statistical dictionary throughout our academic journey. This comprehensive treatment of statistical methods, along with all the abstract tools needed to derive them, is also a signature style of our undergraduate and graduate (B.Stat. and M.Stat.) education [https://www.isical.ac.in/∼deanweb/brochure_bstat.pdf] at ISI, a learning experience that shaped both authors' scholarly outlooks. Indeed, the B.Stat. and M.Stat. degrees came out of a number of courses in statistics that were developed by C. R. Rao as the head of the Research and Training School at ISI. The three-year long B. Stat. and two-year long M. Stat. program prepared students in various aspects of statistics over the course of ten semesters. Every semester had five courses, and many of the earlier ones would give student rigorous exposure to skills that Rao thought Statisticians should need in their arsenal. Joining the B. Stat. program straight out of high school, we were introduced to three-semester long sequences of real analysis, probability, linear (including one course on abstract) algebra, computer programming (in lower level languages such as C or Fortran) data structures, two-semester long elective on a domain science of one's choice (economics, physics or biology), and only one sequence on Statistical Methods where key ideas will be introduced in intuitive albeit somewhat informal manner. Only after these introductory courses will come the more formal statistical topics in their full glory: linear models, parametric and nonparametric inference, stochastic processes, sample survey and design of experiments. By then, the students are well-trained to think through abstract concepts and recognize them in action in commonly used statistical methods. This integration of abstract and real enabled a generation of students to comfortably navigate between the two worlds.

Our professors, many of them leading researchers in their fields, would take advantage of this unique curriculum to creatively teach important concepts in a classroom that left long-lasting impressions on us. As a concrete example of this pedagogical style, we recall how we learned about linear regression in our first year B. Stat. Statistical Methods courses. We did not have any textbook. It was typical of our professor (Prof. Probal Chaudhuri) to come to class, pose a statistical question in simple terms, and encourage us to solve them using the tools we learned from our other courses such as analysis, algebra, and computer programming.

In one class, he drew a bivariate scatter plot of X and Y on the blackboard, and asked us to find the formulae for a "reasonable" straight line (*i.e.* two numbers, a slope β_0 and an intercept β_1) that passes through the plot. After a lively discussion in the classroom on how to even define "reasonable", the class settled on two loss functions: the squared error loss $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ and least absolute deviation (LAD) $\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|$, of two variables β_0 and β_1. We chose to focus with the first because it is differentiable. Students said they have only learned how to differentiate with respect to one variable in high school. So the professor asked: what if I tell you the value of β_0? Can you then find the best β_1 by taking a derivative? Alternately, if I tell you the value of β_1, can you find the best β_0? After some back-of-the-envelope calculations, the students came up with formulae that only involved some weighted means. Then the professor asked: if you keep computing β_0 and β_1 alternately by plugging in the most recent value of the other, would you eventually find their best values? The class was split: some of us thought it will surely work, while others were more skeptical. At this point, the professor reminded us that we don't need to wonder, we already knew enough programming to implement this strategy and see for ourselves. That programming exercise was our Statistics homework for the day. By the time we solved the homework problem and tried it on multiple synthetic data sets that we created ourselves, we had not only learned the concept of simple linear regression, but also a way to solve it using knowledge from our programming class.

We came back to the same problem later in the course, after learning about partial derivatives in other classes, and solved it analytically. This time we not only learned how the closed form solutions of β_0 and β_1 in simple linear regression looked and why the formulae made intuitive sense, we also recognized that even though this strategy is not applicable to the LAD problem, the alternating minimization algorithm introduced earlier is still a potential path to pursue. We revisited the linear regression problem a third time in our Statistics Methods course after we learned enough about vectors and matrices in our linear algebra class. This time, when solving the problem with our newly acquired skills, we recognized that the complicated formulae we derived earlier using multivariable calculus was a special case of a very simple-looking matrix-valued formulae: $\hat{\beta} = (X^\top X)^{-1} X^\top Y$, which even generalized to regression with more than two predictors. This experience helped us appreciate the power of reformulating a complex statistical problem in the language of matrix and vectors.

This pedagogical theme resonated throughout our entire 3-year undergraduate (B. Stat.) and 2 year Masters (M. Stat.) training in ISI. We received rigorous training in real and complex analysis, probability and measure theory, differential equations, and were able to see their application in designing statistical methods. The upshot of this learning style was that Statistics was never about formulae or recipe, it was the experience of solving a realistic problem by combining our intuition with some incredibly powerful yet seemingly disjoint abstract techniques. Later in our careers, we benefitted a lot from this outlook while doing both methodological and interdisciplinary research. Whenever we tried to adopt this learning style in our own undergraduate classes, we appreciated Rao's vision of presenting rigorous mathematical and logical tools in tandem with statistical methods. A solid foundation in the abstract tools can help students feel the joy of discovery when learning about statistical methods, and appreciate them through a developer's lens.

Importance of domain knowledge

Fifteen years ago, at the breakfast table in ISI boys' hostel, we asked Professor Rao for his advice to junior statisticians like us. He offered many valuable insights, but one in particular we remember clearly to this date. He advised us to study, along with statistics, another domain science rigorously. He stressed that it does not matter what the subject is: it could be physics, chemistry, biology or economics. But if we don't acquire expertise in another domain, he said, someone else will get the credit for our core innovation. Coming from a legend of mathematical statistics, this seemed quite unusual at the moment. We both took elective biology courses during our B. Stat. years, but never fully grasped their role in an otherwise quantitative curriculum. Years later, while doing our postdocs, each of us would spend a fair part of two years in molecular biology labs (SB in Brown/Celniker lab at LBNL, JD in Dave lab at Duke), learning from the domain experts in an immersive environment. Our postdoc advisors, Bin Yu and David Dunson, stressed the crucial importance of this immersive experience for carrying out good scientific work. The experience fundamentally changed the way we approach and conduct research, and also form new collaborations. In our academic journey through this era of data science and its widespread impact across disciplinary boundaries, Prof. Rao's words remain all the more relevant.

Prof. C. R. Rao's significant legacy can perhaps be best summarized by the popular Sanskrit phrase: *deepena prajjwalito deepah*, meaning '*from one lamp, another is lit*.' Prof. Rao's name will be remembered for a long time to come as one of the 'developers of statistics as an independent discipline,' through his many path-breaking contributions, his role in statistics education, and his influence on the numerous statisticians like us in the present and future generations.

References

Champkin, J. (2011). C. R. Rao. *Significance*, **8**, 175–178.

DasGupta, A. (2024). C. R. Rao: Paramount statistical scientist (1920 to 2023). *Proceedings of the National Academy of Sciences*, **121**, e2321318121.

Efron, B., Amari, S.-i., Rubin, D. B., Rao, A. S. R. S., and Cox, D. R. (2020). C. R. Rao's Century. *Significance*, **17**, 36–38.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, **37**, 81–91.

Rao, C. R. (1965). *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc., New York-London-Sydney.

Stigler, S. M. (1980). Stigler's law of eponymy. *Transactions of the New York Academy of Sciences*, **39**, 147–157.