Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 13–17
http://www.ssca.org.in/journal

The Importance of C. R. Rao to the Graduate Student

Ezra Becker and Dibyen Majumdar
Department of Mathematics, Statistics, and Computer Science
University of Illinois Chicago

Received: 06 June 2024; Revised: 08 June 2024; Accepted: 09 June 2024

Abstract

This article discusses many of the contributions of C. R. Rao to the education of the statistics graduate student. We submit that Rao's foundational results are not only critical elements of the graduate statistician's analytical toolkit, but also shape the graduate statistician's philosophical approach to statistical research and real-world problem solving.

Key words: Rao-Cramer lower bound; Rao-Blackwell theorem; Linear statistical inference; Graduate student education; C. R. Rao.

AMS Subject Classifications: 01A70, 62F25, 62J05, 97D20

1. Introduction

One might approach the task of honoring the memory of C. R. Rao by selecting one of his myriad contributions to modern statistics and expanding upon why it is so important. We imagine many contributors to this edition of Statistics and Applications might do just that, demonstrating the breathtaking scope and diversity of Rao's foundational results with subtle and nuanced derivations that are sure to be appreciated by the seasoned professors and other established scholars in this journal's readership. However, we believe that one does not need a doctorate and a strong publication history to appreciate what Rao did - and continues to do - for statistics. Indeed, our focus in this article is the immense importance of Rao to beginning graduate students, those just starting out on their journey toward a master's degree or doctorate.

2. In the classroom

STAT 401 is the introductory master's-level probability course at the University of Illinois Chicago, and is taken by most of the first-semester graduate students in the statistics program. We would guess that it is more-or-less similar in content and approach to the beginning master's level probability course at universities across the globe. The course includes an introduction to five major theorems/concepts: the Kolmogorov probability axioms, the Rao-Cramér lower bound, the Central Limit Theorem, Sufficiency and Completeness of

Corresponding Author: Ezra Becker
Email: ebecke1@uic.edu

sample statistics (including Basu's Theorem), and the Rao-Blackwell Theorem. Of course there are several other ideas introduced, such as various discrete and continuous probability distributions, parameter estimation, confidence intervals, and hypothesis testing. But of the five foundational concepts of the course, Rao developed two of them. That fact alone should convince any skeptic how important Rao is to the budding statistician. A closer look at these two concepts yields even more insight into why they are so foundational for beginning graduate students.

Consider the Rao-Cramér lower bound. This elegant relation provides a lower bound for the variance of any unbiased estimator $\widehat{\theta}$ of a parameter of interest θ to the Fisher information for that parameter $\mathcal{I}(\theta)$. In the one-dimensional case, it may be expressed as

$$Var(\widehat{\theta}) \geq \mathcal{I}^{-1}(\theta). \tag{1}$$

In the multidimensional case, where one considers a vector of parameters $\boldsymbol{\theta}$ and a vector of estimable functions of those parameters $\boldsymbol{f(\theta)}$, where the covariance matrix of the estimator $\widehat{\boldsymbol{f(\theta)}}$ is denoted by $\boldsymbol{\Sigma}(\widehat{\boldsymbol{f(\theta)}})$ and the Fisher information matrix is denoted by $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$, it can be shown via a first-order Taylor expansion that

$$\boldsymbol{\Sigma}(\widehat{\boldsymbol{f(\theta)}}) \geq \left(\frac{\partial \boldsymbol{f(\theta)}}{\partial \boldsymbol{\theta}^T}\right) \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{f(\theta)}}{\partial \boldsymbol{\theta}^T}\right)^T, \tag{2}$$

which approaches equality asymptotically for any maximum likelihood estimator $\widehat{\boldsymbol{f(\theta)}}$ under certain regularity conditions. Thus one may loosely say that "maximizing" the information matrix for $\boldsymbol{\theta}$ is equivalent to "minimizing" the variance of $\widehat{\boldsymbol{f(\theta)}}$. In other words, this lower bound says that one's ability to minimize the uncertainty of a parameter estimate is limited in a simple and direct way by the information in the data available regarding that parameter. And note, this lower bound does not come into consideration directly when one is learning about confidence intervals, or hypothesis testing. Why, then, is it so critically important to the beginning graduate student? The answer is that it teaches the beginning graduate student a critical fact about both the theory and the practice of statistics, namely that we are limited by our data in meaningful ways. But it also teaches the student that important and often fruitful avenues of inquiry might be found through attempts to maximize how that available information is utilized.

The Rao-Blackwell theorem provides an elegant application of that concept. The theorem states that an unbiased parameter estimator conditioned on a sufficient statistic for the parameter under consideration will have less uncertainty than an unconditional estimator. And while sufficiency may not be an entirely glamorous concept in this age of computational methods and big data, the lesson of seeking sound theoretical justification for a proposed "better" estimator or classification method certainly endures. In this sense, Rao's work provides not just a technical foundation for how beginning graduate students engage in the practice of statistics, but also a philosophical starting point for what can be achieved with statistics.

Let us next touch upon the course on linear statistical inference that we assume every graduate statistics program offers. At UIC, this course is listed as STAT 521 at the Ph.D. level. Rao's classic *Linear Statistical Inference and its Applications* (Rao, 1965), and in particular the 1973 edition, is considered by many to be the bible of linear regression. It is the textbook that many programs use to teach the course, either as the primary text or as an important reference text. In this book Rao builds a coherent foundation for how to think about and approach statistical modeling. But what's also important is that this course - and by extension, Rao's framework for statistical inference - is the anchor point for a whole field of statistical modeling; for example, it is usually (although not always) the case that nonlinear regression is introduced as an extension of linear regression concepts. And, nonparametric models are often not introduced until linear statistical inference has been taught, again so that nonparametric approaches may be explored in contrast to parametric linear methods. In short, Rao's presentation of linear statistical inference is often the foundation for how statistics students think about and approach modeling; how they seek to describe, explain, and anticipate the world around them in mathematical terms; and how they make sense of complex dynamics.

3. Rao's students and their descendants

There is another measure available that speaks to Rao's influence on the emerging statistics graduate student. The Mathematics Genealogy Project ("MGP"), established by Harry Coonce and currently coordinated by the Department of Mathematics (1996), aspires to inventory as many mathematics Ph.D. holders worldwide (or those with equivalent degrees), and as far back historically, as possible. As of May 15, 2024, the MGP held data on 308,994 mathematicians, with the subset identified as statisticians containing 16,596 degree holders. The MGP lists Rao as having 52 direct Ph.D. students, with total descendants - *i.e.*, students of his direct students, and their students, *etc.*–of 822. That is, in a sense Rao is the direct patriarch of almost 5% of all the Ph.D. statisticians listed in the Project. And while the MGP is not comprehensive in its coverage, its coordinators are confident that it represents a comfortable majority of Ph.D. holders worldwide, certainly for degrees granted more than five years ago. The point is that Rao had a direct influence on the development of many, many graduate students, and a one- or two-step removed influence on many more.

It is also worth pointing out that many of Rao's students were themselves hugely influential in the development of diverse fields of statistics, such as Debabrata Basu (author of the famous Basu's Theorem referenced above), T. E. S. Raghavan (game theory), S. R. S. Varadhan (probability theory and large deviations), and others. Thus not only did Rao help forge the statistical perspective of many students directly, in particular he did so for a set of extraordinary students who in turn helped shape statistics – and how graduate students learn and approach the field – in important ways.

4. A walk down memory lane

The Indian Statistical Institute (ISI) was established by P. C. Mahalanobis. C. R. Rao joined the Institute, and in 1972 succeeded Mahalanobis as Director. Under Mahalanobis and Rao's leadership, ISI became one of the earliest, and most eminent, statistics institutes in the world, achieving great advances in both theory and applications. The Institute created an atmosphere of excellence in research, where the mind could roam and attain great heights. There was no boundary – many areas of research, even those not really a part of statistics, were encouraged to blossom. For example, as mentioned above T. E. S. Raghavan got his Ph.D. under the supervision of C. R. Rao in game theory! Graduate students were inspired to learn and discover, but the unique atmosphere created by Rao put the onus on the student to advance. There were faculty and postdocs who helped excite and advance minds, but learning and problem solving was entirely the responsibility of the student. In the early days, there were essentially no classes for Ph.D. students.

When Dibyen Majumdar (DM) joined ISI Kolkata as a Ph.D. student, he got an opportunity to meet Sujit Kumar Mitra (SKM), a statistician of eminence and a strong collaborator of C. R. Rao. DM wanted to work on linear models. He was told by SKM to read several chapters of Rao (1965) *Linear Statistical Inference and Its Applications*, read all chapters of Rao and Mitra (1971) *Generalized Inverse of Matrices and its Applications*, solve all the exercise problems in those texts and related problems, and to move to ISI New Delhi (Rao's and SKM's campus at the time) only if he succeeded in this learning project. There were no classes for DM, but ISI Kolkata had a galaxy of star researchers who helped DM enormously, even though they were not solving the problems themselves. One person who helped DM was the postdoctoral fellow P. Bhimasankaram, a former student of SKM who had great knowledge of g-inverses.

Here is one example of an exercise that DM had to solve at the start of his graduate student life at ISI Kolkata, for which Bhimasankaram helped him derive the proof.

Exercise 1: Let A $(n \times p)$ and B $(m \times p)$ be two (real) matrices such that

$$\mathcal{M}(A') \cap \mathcal{M}(B') = \{0\}\,. \tag{3}$$

Then

$$\begin{aligned} A'A(A'A + B'B)^- A'A &= A'A & (4)\\ B'B(A'A + B'B)^- B'B &= B'B & (5)\\ A'A(A'A + B'B)^- B'B &= 0. & (6) \end{aligned}$$

Proof: Since

$$\mathcal{M}(A'A) \subset \mathcal{M}(A'A + B'B), \tag{7}$$

$$\begin{aligned} A'A(A'A + B'B)^-(A'A + B'B) &= A'A,\\ \text{i.e., } A'A(A'A + B'B)^- A'A - A'A &= A'A(A'A + B'B)^- B'B. \quad (8) \end{aligned}$$

It follows from (7) and $\mathcal{M}(B'B) \subset \mathcal{M}(A'A + B'B)$ that we can assume with no loss of generality that $(A'A + B'B)^-$ is symmetric. Taking transposes in (8) we get

$$A'A(A'A + B'B)^- A'A - A'A = B'B(A'A + B'B)^- A'A.$$

In this expression,

$$\begin{aligned} \mathcal{M}(A'A(A'A + B'B)^- A'A - A'A) &\subset \mathcal{M}(A'A) = \mathcal{M}(A'),\\ \mathcal{M}(B'B(A'A + B'B)^- A'A) &\subset \mathcal{M}(B'B) = \mathcal{M}(B'). \end{aligned}$$

Hence condition (3) implies

$$A'A(A'A + B'B)^- A'A - A'A = B'B(A'A + B'B)^- A'A = 0.$$

This establishes (4) and (6). Clearly, (5) will have a similar derivation. □

These results have two lessons for a graduate student. The first, demonstrated by equations (4) and (5), is the richness of the family of generalized inverses that was introduced by Rao (1967), going far beyond the Moore-Penrose inverse. In other words, $A'A$ has a wide variety of generalized inverses with potentially desirable properties. The second, demonstrated by equation (3), starts with the fact that $\mathcal{M}(A')$ and $\mathcal{M}(B')$ are disjoint but not necessarily orthogonal under the inner product $< x, y > = y'x$. However, it is possible to find a positive definite matrix, say Q, that is a generalized inverse of $(A'A + B'B)$; $Q = (A'A+B'B)^-$. Then $\mathcal{M}(A')$ and $\mathcal{M}(B')$ are orthogonal under the inner product $< x, y > = y'Qx$, *i.e.*, $AQB' = 0$.

5. Conclusion

Many beginning statistics graduate students have trouble wrapping their arms around Rao's results the first time they are exposed to them (or even the second or third time!). And, not every beginning graduate student has the wherewithal to understand the more nuanced ramifications of Rao's results, let alone the ability to prove those results. Deeper insights and the ability to "connect the dots" of Rao's results usually come with time and experience. But what all aspiring graduate statisticians have, whether they realize it at the time or not, is the influence of Rao's work laying the foundation for how they perceive the nature and function of statistics. It is in this regard that a hugely important contribution of Rao endures, and will continue to do so as long as graduate students study classical statistics.

Acknowledgements

We are grateful to the editors and the Journal for the opportunity to contribute these thoughts on the impact of C. R. Rao.

References

Department of Mathematics, N. D. S. U. (1996). Mathematics genealogy project. `https://genealogy.math.ndsu.nodak.edu/index.php`. Accessed May 15, 2024.

Rao, C. R. (1965). *Linear Statistical Inference and its Applications.* John Wiley & Sons.

Rao, C. R. (1967). Calculus of generalized inverses of matrices Part I. General theory. *Sankhya-Series A*, **29**, 317–342.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* John Wiley and Sons, New York.