STAT 522 Project
Introduction of C.R.Rao

Fengli Zou

Tianmeng Wang

April 2021

Contents

1 Introduction

In this project report, we introduced C.R.Rao and some of his major contributions in statistics and many other fields. In section 2, we briefly present a life story of Rao, including his background and his academic path. In section 3, we talked about his contributions in multivariate analysis, such as MANOVA test and the best-known approximation for Wilks lambda, which are related to our course 522. In section 4, we introduced three of his contributions related to statistical inference and linear models–Cramer-Rao lower bound, Rao-Blackwell theorem, and C.R.Rao's score test. In section 5, we discussed orthogonal array, another essential contributions in experimental design. Later in section 6, we introduced generalized inverse of matrix, which is widely used in matrix theory and linear inference of statistics. Additionally, in section 7, some real and potential applications based on Rao's contributions was further discussed.

2 Life Story of Rao : The great man comes from the small

C.R.Rao, who we've known worldwide, is short for Calyampudi Radhakrishna Rao, born on 10th September, 1920 in Huvanna Hadagali, Karnataka, India. He came from a large family with six brothers and four sisters, among which he is the eighth child. C.R.Rao was named Radhakrishna because there is a tradition at that time to name the eighth child after God Krishna. Maybe he is deemed to become a great person after all. He completed his M.A. degree in mathematics at the Andhra University in Waltair, Andhra Pradesh and also got his M.A. degree in statistics from Calcutta University in Kolkata, West Bengal in 1943. He was among the first few people in the world to hold a Master's degree in Statistics. However, he never cease to pursue his academic career. During the period working as a researcher in Calcutta, Rao was invited to work on a project at Cambridge University, UK, furthermore, due to his great contributions in this project, he earned his Ph.D. in 1948 from Cambridge University with the help of R.A. Fisher, the father of modern statistics, who is his thesis advisor. Rao hold a prestigious position in statistical field in India working as a director of ISI and professor, after retiring, he was invited to work as a professor in University of Pittsburgh and Eberly Professor of Statistics and Director of the Center for Multivariate Analysis at Pennsylvania State University in USA. He devotes his life to his academic career, making contributions that could benefits various of fields, even currently, he is in his early 80s, and he is still keep a regular routine working and publishing certain academic papers. And it is fortunate for us to immerse ourselves into something we love, and we would spend our life time to pursue this goal.

3 Contributions to Multivariate Analysis (Related to 522)

3.1 MANOVA Test Statistic for Subsets of p Variates

One part of Rao's research field of multivariate analysis is MANOVA test. MANOVA is a generalized form of univariate analysis of variance (ANOVA), it uses the covariance between outcome variables in testing the statistical significance of the mean differences. Based on Rao [1], several Multivariate analysis of variance (MANOVA) problems are addressed. And later in his book [2] where "Hotelling T^2 showed no significance between the two populations with two variables, while two sample t-tests based on each of the variables were highly significant". In multivariate analysis it is considered as the first example of "curse of dimensionality", which is also named as "Rao's paradox" . In general, "the curse of dimensionality" phrase was introduced by Bellman for describing the problem caused by the exponential increase in volume associated with adding extra dimensions to the Euclidean space. More specifically when it comes to one of Rao's finding, if we are dealing with the analysis of p variables, we might want to test the significance of different subset of p variables. And in Rao's paper, he showed how to test the subset of p variables is still significant or not. Here is a simple conclusion of the test statistic proposed by Rao.

It is assumed that all of the $n(= n_1+\cdots+n_q)$ observations are normally distributed with the common covariance matrix $\mathbf{\Sigma}$. Let $\mathbf{Y}_{i1}, \cdots, \mathbf{Y}_{in_i}$ be samples from the ith treatment group $N_p(\mu^{(i)}, \mathbf{\Sigma})$. For testing the equality of the mean vectors, we have null hypothesis, $H_0 : \mu^{(1)} = \cdots = \mu^{(q)}$, let $\mathbf{B}$ and $\mathbf{W}$ be the matrices of sums of squares and products.

$$\mathbf{B} = \sum_{i=1}^{q} n_i \left(\overline{\boldsymbol{Y}}_{i.} - \overline{\boldsymbol{Y}}_{..}\right)\left(\overline{\boldsymbol{Y}}_{i} - \overline{\boldsymbol{Y}}_{.}\right)', \quad \mathbf{W} = \sum_{i=1}^{q}\sum_{j=1}^{n_i} \left(\boldsymbol{Y}_{ij} - \overline{\boldsymbol{Y}}_{i}\right)\left(\boldsymbol{Y}_{ij} - \overline{\boldsymbol{Y}}_{i}\right)'$$

Letting $\mathbf{T} = \mathbf{B} + \mathbf{W}$, an LR test for H_0 is based on $\Lambda = |\mathbf{W}|/|\mathbf{T}|$, which is the Wilks Lambda. Now we proposed a test for whether $\mathbf{Y}_2 = (Y_{k+1}, \cdots, Y_p)'$ brings out further differences in q populations when the differences due to $\mathbf{Y}_1 = (Y_1, \cdots, Y_k)'$ are removed. Let us consider this problem in a multivariate one-way MANOVA model. Therefore, we could redefine our hypothesis as

$$\boldsymbol{\mu}_{2\cdot 1}^{(1)} = \cdots = \boldsymbol{\mu}_{2\cdot 1}^{(q)}$$

where $\boldsymbol{\mu}_{2\cdot 1}^{(i)} = \boldsymbol{\mu}_{2}^{(i)} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\boldsymbol{\mu}_{1}^{(i)}, i = 1, \ldots, q$. Here, $\boldsymbol{\mu}_{(i)}$ and $\mathbf{\Sigma}$ have been decomposed as in previously. Let us decompose B and W as

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}$$

where $\mathbf{W}_{22\cdot 1} = \mathbf{W}_{21} - \mathbf{W}_{21}\mathbf{W}_{11}^{-1}\mathbf{W}_{21}$ and $\mathbf{T}_{22\cdot 1} = \mathbf{T}_{21} - \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{T}_{21}$. The null distribution is $\Lambda_{p-k}(q-1, n-q-k)$.

3.2 The Best-known Approximation for Wilks's Lambda

Rao has developed different types of tests based on real data, substantially these tests are likelihood ratio tests. Normally, when we conduct likelihood ratio test, we would consider the test statistic Λ, and find its null distribution. Based on the previous set up in section 3.1, there are two different types of approximation of Wilks Lambda, One approximation is attributed to M.S. Bartlett and he estimates Wilks' lambda to be an approximate chi-squared distribution with large sample size.

Another approximation attributes to Rao, he considers the distribution of a Lambda statistic to be a highly accurate F approximation for a transformed version of $\Lambda_p(q, n-q)$, which specified as below:

Based on the section 3.1, Rao [3] proposed a better F approximation of the distribution of another function of $\Lambda = \Lambda_p(q, n-q)$. The approximation is to consider

$$\frac{1-\Lambda^{1/s}}{\Lambda^{1/s}} \cdot \frac{ms+2\lambda}{pq}$$

as an F approximation with pq and $ms+2\lambda$ degrees of freedom, where

$$\lambda = -\frac{1}{4}pq + \frac{1}{2}, \quad s = \left(\frac{p^2q^2-4}{p^2+q^2-5}\right)^{1/2}$$

For $p = 1$ or 2 (or $q = 1$ or 2) the F-distribution is exactly as given. If $ms+2\lambda$ is not an integer, interpolation between two integer values can be used.

4 Contributions to Statistical Inference

Rao has made formidable contributions to advance knowledge in statistical inference. His book Linear Statistical Inference and Its Applications [4] has been cited over 17,000 times. Here are several well-known theorem or methods discovered by him.

4.1 Estimation Theory

Cramér-Rao bound, also known as Cramér-Rao lower bound, expresses a lower bound on the variance of unbiased estimators of a fixed and unknown parameter, stating that the variance of any such estimator is at least as high as the inverse of the Fisher information. Therefore we would take this specific parameter estimate as efficient.

Suppose $g(\theta)$ is a function of parameter for a distribution with some probability density function $f(x;\theta)$, and $T(X)$ is an unbiased estimator of $g(\theta)$, then under some conditions, the variance of $T(X)$ is bounded by

$$\mathrm{var}(T) \geq \frac{[g'(\theta)]^2}{I(\theta)}$$

where the Fisher's information $I(\theta)$ is defined as

$$I(\theta) = n\mathrm{E}_\theta \left[\left(\frac{\partial \ell(X;\theta)}{\partial \theta} \right)^2 \right]$$

and $\ell(X;\theta)$ is the log likelihood function. Furthermore, if $\ell(X;\theta)$ is twice differentiable, then Fisher information can be written as

$$I(\theta) = -n\mathrm{E}_\theta \left[\frac{\partial^2 \ell(X;\theta)}{\partial \theta^2} \right]$$

In the multivariate case, define $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_d) \in \mathbb{R}^D$ with probability density function $f(x;\boldsymbol{\theta})$, and let $\mathbf{T}(X)$ be an estimator function of $q(\boldsymbol{\theta})$. Then the Cramér-Rao bound is defined as

$$\mathrm{Var}_\theta \, \boldsymbol{T}(X) \geq (\nabla q(\boldsymbol{\theta}))(I(\boldsymbol{\theta}))^{-1}(\nabla q(\boldsymbol{\theta}))^t$$

Where $\nabla q(\boldsymbol{\theta})$ is the Jacobian matrix and $I(\boldsymbol{\theta}) = E\left[(\nabla L(\boldsymbol{\theta} \mid X))^t \nabla L(\boldsymbol{\theta} \mid X)\right]$ [4].

Another important contributions in estimation theory is Rao and David Blackwell found Rao-Blackwell theorem founded by Rao [5] and David Blackwell [6]. Rao-Blackwell theorem states that let $U = U(X)$ be a sufficient statistic for θ and let

$T(X)$ be any kind of estimator of parameter θ. Then $\tilde{T} = E(T \mid U)$ is an improved estimator of T, or we can say $\operatorname{Var}_\theta \tilde{T} \leq \operatorname{Var}_\theta T$.

The Rao-Blackwell theorem improves efficiency by taking conditional expectation with respect to a sufficient estimator. The theorem has many applications, including estimation of prediction error and producing estimates from sample survey data. More specifically, observations in adaptive sampling are found in sequence, each new observation depends on one or more characteristics from prior observations, and the improved estimators can also be found by taking an average of estimators over every possible order.

4.2 C.R.Rao's Score Test

Rao's Score test is an alternative to Likelihood Ratio and Wald tests. All these tests are equivalent to the first-order of asymptotics, but differ in the second-order properties to some extent. The main advantage of the score test over the Wald test and likelihood-ratio test is that the score test only requires the computation of the restricted estimator. This makes testing feasible when the unconstrained maximum likelihood estimate is a boundary point in the parameter space. Additionally, the discovering history for score test is quite interesting. During Rao's time as a research scholar at ISI, a scientist named S.J. Poti asked him for help in testing a one-sided alternative hypothesis ($\theta > \theta_0$). Rao finds it very interesting and quite new in this area, then he set to work to develop a test that was most powerful for alternatives close to θ_0. Later he and Poti published a paper "On Locally Most Powerful Tests When the Alternatives Are One Sided" [7]. After moving to Cambridge, Rao continued working on this problem and making breakthrough. Nowadays it is known that many of the most-used significance tests for categorical data analysis are score tests by improving or transforming Rao's score test. which include McNemar's test, the Mantel-Haenszel test, the Cochran-Armitage test, and Cochran's test. In long history of developing and improving, score test has had a profound effect on econometric inference and is now the most common method of testing used by econometricians. Below are some simple comparison based on null distribution of three tests.

We want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. For Wald Test, under H_0 we have

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, I^{-1}\left(\theta_0\right)\right)$$

where $I(\theta_0)$ is the Fisher's Information Matrix.

The LR test is based on

$$\Delta_n = I\left(\widehat{\theta}_n\right) - I\left(\theta_0\right) = \log\left(\frac{\sup_{\theta\in\Theta} L(\theta \mid \mathbf{x})}{L\left(\theta_0 \mid \mathbf{x}\right)}\right) \geq 0.$$

And by Slutsky's Theorem, under H_0 we have null distribution

$$2\Delta_n \xrightarrow{\mathrm{d}} \chi_1^2$$

Lastly, for Rao's score test, under H_0 we have

$$\frac{l'\left(\theta_0\right)}{\sqrt{nI\left(\theta_0\right)}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $l'(\theta) = \frac{\partial \log L(\theta|\mathbf{x})}{\partial\theta}$. In practice, for complex models, one would use

$$R_n = \frac{l'\left(\theta_0\right)}{\sqrt{n\widehat{I}\left(\theta_0\right)}}$$

where

$$\widehat{I}\left(\theta_0\right) = -\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \log f\left(x_i \mid \theta\right)}{\partial\theta^2}\right|_{\theta_0} \xrightarrow{P} I\left(\theta_0\right)$$

5 Orthogonal Array

Another great contributions made by C.R.Rao is Orthogonal array, which is widely used in factorial design and combinatorial design. A standard notation of an orthogonal array is (N, n, s, d). Here N stands for number of rows, and in combinatorics the rows represent different tests to be run; n stands for number of columns, and in combinatorics each column represents a different parameter; s stands for number of possible variable values, and in combinatorics represents different levels to be tested; d stands for the number of variables in each N*d subarray, or strength in combinatorics [8]. For example, a (8,4,2,3) orthogonal array may written in the form

0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

This array contains 8 rows, 4 columns, and for each cell the possible value is 0 or 1. And if randomly choose 3 columns, the combination can represent 8 different values.

Furthermore, orthogonal array have many applications in statistical design of experiments, coding theory, cryptography and various types of software testing. One of the application is orthogonal array testing. Orthogonal array testing is a software testing technique that uses orthogonal arrays to create test cases, and it is particularly effective in finding errors associated with faulty logic within computer software systems.

6 G-Inverse

In 1962, C.R.Rao introduced the concept of generalized inverse (g-inverse) of a matrix. A generalized inverse (or g-inverse) of a matrix A of order $m \times n$ is a matrix of order $n \times m$, denoted by A^{-}, such that for any y for which $Ax = y$ is consistent, $x = A^{-}y$ is a solution. An equivalent definition of g-inverse is an order $n \times m$ matrix A^{-} such that $AA^{-}A = A$.

Furthermore, Rao introduced the relationship between g-inverse, projection, and best linear unbiased estimator (BLUE) in 1972 [9]. Consider the general Gauss-Markov model

$$(\boldsymbol{Y}, \boldsymbol{X\beta}, \boldsymbol{\sigma^2 V}), \quad E(\boldsymbol{Y}) = \boldsymbol{X\beta}, \quad Var(\boldsymbol{Y}) = \boldsymbol{\sigma^2 V}$$

The BLUE of an estimable parametric function $\boldsymbol{c'\beta}$ is $\boldsymbol{c'}\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is defined as

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T\boldsymbol{Y}$$

Also Rao introduced MINQUE estimate of $f\sigma^2, f = R(\mathbf{V} : \mathbf{X}) - R(\mathbf{X})$, can be written in form of

$$\mathbf{Y}'\mathbf{T}^{-}\left(\mathbf{I} - \mathbf{P}_{\mathbf{XT}^{-}}\right)\mathbf{Y}$$

Where R stands for rank, $(\mathbf{A} : \mathbf{B})$ stands for matrix obtained by adjoining the columns of $\mathbf{B}$ to those of $\mathbf{A}$, and $\mathbf{T} = \mathbf{V} + \mathbf{XUX}'$ for any symmetric $\mathbf{U}$ such that $\mathbf{T}$ is n.n.d. and $R(\mathbf{T}) = R(\mathbf{V} : \mathbf{X})$ [10].

When calculating OLS estimator, or $\widehat{\boldsymbol{\beta}}$, ideally we want that $\boldsymbol{X}^T\boldsymbol{X}$ is nonsingular, so finding inverse of matrix would be much easier, but in practice singular matrix appears more often. So finding OLS estimator and BLUE can be fulfilled by using generalized inverse (and more often the Moore-Penrose g inverse) in practice.

G-inverse and its application is essentially in Multivariate Statistical Analysis. In multivariate multiple regression, each component $\boldsymbol{\beta}_{(i)}$ has OLS estimator

$$\widehat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{Z}^T\boldsymbol{Z})^{-}\boldsymbol{Z}^T\boldsymbol{Y}_{(i)}$$

and $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_{(1)}, \cdots, \widehat{\boldsymbol{\beta}}_{(m)}]$ can be written as

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}^T\boldsymbol{Z})^{-}\boldsymbol{Z}^T\boldsymbol{Y}$$

This estimator is widely used since it has many nice properties: unbiasness, minimized $\mathrm{tr}((\boldsymbol{Y}-\boldsymbol{ZB})'(\boldsymbol{Y}-\boldsymbol{ZB}))$, and minimized $|(\boldsymbol{Y}-\boldsymbol{ZB})'(\boldsymbol{Y}-\boldsymbol{ZB})|$. It is still one of the most widely used estimator in linear regression.

7 Real and Potential Applications Based on Rao's Contributions Connected to 522

Fields like Economics, Genetics, Anthropology, Geology, Demography, Biometry, and Medicine. C. R. Rao establish his prestigious reputation by applying his discovering and theorem to all kinds of field, which also make him one of the world leaders in statistical science over the last six decades. His research, scholarship, and professional services have had a great influence on statistical theory and applications. More specifically, besides what we've introduced above, Rao's findings that has been used frequently in academic world are beyond counting, such as Cramer-Rao inequality, Isher-Rao and Rao's Theorems on second order efficiency of an estimator, Rao metric and distance, others like Canonical Variate analysis and for mathematics, G-inverse of matrices appear in all standard books on statistics. His name also bear in some other technical terms such as are Rao's F and U tests in multivariate analysis, Rao's Quadratic Entropy [11], Cross Entropy and Rao-Rubin, Lau-Rao, Lau-Rao-Shanbhag and Kagan- Linnik-Rao theorems on characterization of probability distributions.

"Cramer-Rao Bound and Rao-Blackwellization are the most frequently quoted key words in statistical and engineering literature." Quantum Cramer- Rao Bound [12] have been created in Quantum Physics, which is one of the special use of Cramer-Rao Bound under the technical term. Rao-Blackwellization has found applications in adaptive sampling, particle filtering in high-dimensional state spaces, dynamic Bayesian networks etc. As a matter of fact that these newly developed theory along with Rao's fundamental creation advance some strategic application, such as signal detection, tracking of non-friendly planes and recognition of objects by shape.

8 Summary

We've introduced Rao's contributions in multivariate analysis, estimation theory, orthogonal array and G-inverse, among which the MANOVA test in multivariate analysis and F. approximation of Wilks lambda are connected with current course 522, we could do one-way or two-way MANOVA test to check the significance of our treatments, and use Rao's F approximation when dealing with the test statistic for likelihood ratio test, which is the Wilks lambda. For estimation theory part, we could use Rao's finding to conduct statistical inference for our parameter estimation in order to get a more informative estimator. Orthogonal array and G-inverse are mathematical based, and widely used in statistical inference and multivariate analysis, especially g-inverse, it provides us with different solutions of our lease squared estimator for parameter by using one formula when our design matrix is not of full-rank.

C.R.Rao is one of the most contributed statistician. He is currently professor emeritus at Pennsylvania State University and Research Professor at the University at Buffalo. Rao has been honoured by numerous colloquia, honorary degrees and festschrifts Rao received 38 honorary doctorates from universities in 19 countries spanning six continents. His special awards include the National Medal of Science, USA; India Science Award; and the Guy Medal in Gold, United Kingdom. The American Statistical Association has described him as "a living legend whose work has influenced not just statistics, but has had far reaching implications for fields as varied as economics, genetics, anthropology, geology, national planning, demography, biometry, and medicine." He is one of 77 contemporary scientists in all fields listed in Mariana Cook's book, Faces of Science, and among a galaxy of 57 famous scientists of the 16th through the 20th centuries listed in the "Chronology of Probabilists and Statisticians" compiled by M. Leung at The University of Texas, El Paso. He produced 14 books, 475 research papers, 51 PhD students, and 42 edited volumes of the Handbook of Statistics. We would like to appreciate for Rao's contributions in statistics, economics, biometry, and many other areas.

References

[1] Rao, C. Radhakrishna (1948). *Tests of significance in multivariate analysis.* Biometrika, 35, 58-79.

[2] Rao, C. Radhakrishna (1952). *Advanced Statistical Methods in Biometric Research.* John Wiley & Sons, New York.

[3] Rao, C. Radhakrishna (1951). *An asymptotic expansion of the distribution of Wilk's criterion.* Bulletin of the International Statistical Institute, 33, Part 2, 177–180.

[4] Rao, C. Radhakrishna (1973). *Linear Statistical Inference and Its Applications(2nd ed.).* John Wiley & Sons, New York.

[5] Rao, C. Radhakrishna (1945). *Information and accuracy attainable in the estimation of statistical parameters.* Bulletin of the Calcutta Mathematical Society. 37 (3): 81–91.

[6] David Blackwell (1947). *Conditional Expectation and Unbiased Sequential Estimation.* The Annals of Mathematical Statistics, 18(1) 105-110 March, 1947.

[7] Rao, C., & Poti, S. (1946). *On Locally Most Powerful Tests When Alternatives Are One Sided.* Sankhyā: The Indian Journal of Statistics (1933-1960), 7(4), 439-439.

[8] Rao, C. Radhakrishna (1947). *Factorial Experiments Derivable from Combinatorial Arrangements of Arrays.* Supplement to the Journal of the Royal Statistical Society, vol. 9, no. 1, pp. 128–139.

[9] Rao, C. Radhakrishna (1962). *A Note on a Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 24, no. 1, pp. 152–158.

[10] Rao, C. Radhakrishna (1974). *Projectors, Generalized Inverses and the BLUE's.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 36, no. 3, pp. 442–448.

[11] Rao, C.R. (2010). *Quadratic entropy and analysis of diversity.* Sankhya 72, 70–80.

[12] Braunstein, Samuel L.; Caves, Carlton M.; Milburn, G.J. (1996). *Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance.* Annals of Physics. 247 (1): 135–173.