A Tribute to Professor C. R. Rao

Arijit Chaudhuri

While paying regards to a hugely phenomenal savant of eminence, a disclosure of a few features of the person submitting them is axiomatically inevitable. The article 'C. R. Rao: A life in statistics' by Prof. B L S Prakash Rao (2014) contained in this volume has admirably exposed adequately diverse aspects of the life and work of Prof. C. R. Rao. So, I must refrain from trying to add anything to such an exercise. But I like to reveal here a few personal matters only by way of paying regards to Prof. C. R. Rao.

My Ph.D. research supervisor was Prof. H. K. Nandi in Calcutta University (C.U) who was a classmate of Prof. C. R. Rao in the first batch of post-graduate students of statistics in C.U. I was never a student of Prof. C. R. Rao. I passed M.A in statistics in February, 1962 from C.U. But Prof. Nandi taught us his first published book 'Advanced Statistical methods in Biometric Research' (1952, John Wiley & Sons). We learnt Rao—Cramer inequality and Rao—Blackwell theorem from this book.

Treatment of vectors and matrices, linear models, limit theorems on probability, and distribution theories in this early book appeared to us formidably tough. Just before starting this write-up, I once again glanced through its pages and the contents looked no easier even now.

Once in 1977, I received Prof. Rao's first letter asking me to contribute a research article with a 4-page limit for Bulletin of International Statistical Institute (BISI) as he was the then President of ISI. To my utter delight in 1983, I received his second letter asking me to contribute one full-fledged article to the Handbook of Statistics Volume 6 he was to jointly edit with Prof. P. R. Krishnaiah for Elsevier Publishers of North Holland. This was eventually published in 1988 with my paper turning out to be the biggest in length. In 2015, when S. L. Warner, the father of randomized response (RR) techniques (RRT), was to attain 50 years of his invention, Prof. C. R. Rao in his email asked me to be his joint editor for the Handbook of Statistics

A. Chaudhuri (✉)
Indian Statistical Institute, Kolkata, India

A. Chaudhuri et al. (eds.), *A Tribute to the Legend of Professor C. R. Rao*,
Indian Statistical Institute Series,
https://doi.org/10.1007/978-981-33-6991-7_14

Volume 34 exclusively on RRTs. He immediately agreed to my suggestion to add Prof. T. C. Christofides of Cyprus University as the third joint editor, and the volume was published in April 2016. By that time, however, I developed a closeness to him in course of my repeated correspondences procuring his unmixed collaborative participation in his aiding our publication with authorship of three of the four books with his contributed articles therein brought out by ISIREA during 2009–2017. He not only wrote for us, but emailed me in appreciation of our publications which he claimed to have read cover to cover including the articles in Bengali and not to speak of those in English.

When we sought his blessings announcing our scheme to observe his centenary with a financial and moral backing from the ISI Director, he enthusiastically emailed me his elaborate CV plus a few photographs of himself.

Let me now recount a few of mine own academic contributions through which we may mean to pay our due respects to this man of outstanding stature with lifelong academic pursuits.

I shall narrate only a few simple findings of mine or jointly with those of my students:

(i) To guard against possibilities of committing silly mistakes, not uncommon in practice, it is good to know that for the existence of an unbiased estimator for a finite population total 'the inclusion–probability of every population unit' should be necessarily positive and it is 'sufficient' for it to be so; proof is non-trivial;

(ii) for the existence of an unbiased estimator of the variance of an unbiased estimator for a finite population total its 'positivity of inclusion-probability of every pair of distinct population units,' is 'necessary' as well as 'sufficient.' Why?
$V(t) = E(t^2) -$ (Sum of squared y's + Sum of cross-products of all y_i, y_j's for distinct pairs of units).
So, unbiased estimation of Sum of y_i, y_j's, in the cross-product herein positivity of π_{ij}'s is an 'NSC.' Hence the above claim.

(iii) With survey data in hand through 'Probability Proportional to Size Sampling Without Replacement' to unbiasedly estimate the gain in efficiency over a rival 'Simple Random Sampling Without Replacement' (SRSWOR) a simple trick is to use Desraj's (1956) unbiased estimator for the 'Population total of the squared variate-values.'

(iv) To rationally set the size of an SRSWOR to ensure

$$\text{Prob}[|t - Y| \leq fY] \geq 1 - a$$

with pre-assigned proper fractional positive values of f and a one may use Chebyshev's inequality

$$\text{Prob}[|t - Y| \leq K.\text{SE}(t) \text{ of } t] \geq 1 - \frac{1}{K^2}.$$

This equivalently needs tabulating N, f, α, CV$= 100\frac{\mathrm{SE}(t)}{Y}$ to lay down the values of n, the sample-size.

(v) It is a tough task to extend (iv) to cover Warner's RR alternative estimator e replacing t. To see this, note

$$V(e) = V(t) + \frac{N^2}{n}\frac{P(1-P)}{(2P-1)^2} = \mathrm{I} + \mathrm{II}, \text{ say}$$

This is so when we employ SRSWOR of size n,
$t = N$ Sample-mean of y-values,
$e = t$ (with y's replaced by RR-based unbiased estimates of y's)
$P =$ Proportion of cards marked 'Yes' $\left(0 < P \neq \frac{1}{2} < 1\right)$ about bearing a stigmatizing characteristic, say, 'Testing HIV Positive.'
It is verifiable that the Chebyshev's approach in (iv) above gives reasonable values of n versus N, f, α, CV to contact but the solution of n correspondingly for I + II turns out beyond limits of logic.

(vi) To cover the cases of unequal probability sampling is difficult but a few solutions we have reached.

(vii) To apply sophisticated method of sampling to audit Government financial transactions of huge dimensions, we found it convenient to apply 'Multi-stage Sampling,' choosing, e.g., district-wise PWD offices by Rao—Hartley—Cochran (RHC 1962) scheme in the first stage, with sanctioned budget amounts as size-measures and by SRSWOR each of the next four-stage units, namely the different office departments, department-wise ledger books, pages and the page-wise rows, and reading of the row-wise column entries. Unbiased estimation of totals and their variances is easy to come by [vide Chaudhuri, Arijit (2010)].

(viii) An Indian version of small area estimation (SAE) to apply to NSSO.
What is called small area estimation (SAE) or problem of developing small domain statistics arises when a suitable sample is taken from a population to estimate its total but in addition the same sample is used to estimate totals of the domains of the population of various sizes and unacceptably low efficiencies are encountered about estimates of several domain totals owing possibly to small numbers of sample units pertaining to several domains of interest.
In NSSO, rural surveys in India samples are taken in two stages from every district in each state or union territory (UT); the villages are the first stage and the households (HH) therefrom, if selected, constituting the second stage units. The survey data are used to derive 'District-level' estimates, and they are just added to provide state or UT-level estimates.
The sample-size for every district invariably turns out palpably small raising doubts about the efficiency level of state/UT-level estimates which may hopefully be improved by applying model-based SAE methods.
The Government of India in 2005 asked me to try and provide other alternative estimates for all the states/UT's, Chaudhuri (2005) provided his response

with estimates having reduced estimated coefficients of variation on applying generalized regression (Greg) method of estimation taking village 'Census-population values as exogenous regressors.' We also tried Empirical Bayes (EB) estimates which offered no tangible further improvements on the Greg estimates. The government appeared satisfied with a letter of thanks on receiving my reported alternatives.

(ix) Randomized response techniques since inception happened to be based mostly on exclusively simple random sampling with replacement (SRSWR) are seen in the early publications.
In order to convince NSSO of their applicability in Chaudhuri (2001), it showed extension of most of the known RRTs to be applicable in usual varying probability sampling situations.

(x) Observing wide discrepancies in reported amounts of rural loans 'incurred' from NSSO surveys and 'advanced' by banks from RBI accounts, I could document [vide Chaudhuri (2017)] an effort to reconcile them by dint of 'Network and Adaptive sampling'—a book-length publication being Chaudhuri, Arijit's (2015) text with this title.

References

Chaudhuri, A. (2001). Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. *Jour. Stat. Plan. Inf., 94,* 37–42.

Chaudhuri, A.: A report on completion of the first phase of an NSSO Project entitled "Application of Small Area Techniques for Estimation of District Level Parameters"—Phase I covering rural India for the 55th round NSSO Survey, MOSPI, GOI (2005)

Chaudhuri, A.:. Essentials of Survey Sampling, pp. 179–185. Prentice Hall of India (2005)

Chaudhuri, A.:. Network and Adaptive Sampling. CRC Press, Taylor & Francis. Boca Raton, Florida, USA (2015)

Chaudhuri, A. (2017). Unbiased estimation of total rural loans advanced and incurred in an Indian state aling with unbiased estimation of variance in estimation. *Cal. Stat. Assoc. Bull., 69*(1), 71–75.

Des, R. (1956). Some estimators in sampling with varying probabilities without replacement. *Jour. Amer. Stat. Assoc., 51,* 269–284.

Rao, BLSP.: CR Rao: A life in Statistics. In Current statistics series on "Living Legends in Indian Science" (2014)

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: Wiley.

Rao, J.N.K., Hartley, H.O., Cochran, W.G.: On a simple procedure of unequal probability sampling without replacement. J.R.S.S.B **24**, 482–491 (1962)