SAGI
Statistics for the Australian Grains Industry
National | North | West | South

Ranked Set Sampling Symposium: Translating the theory to applications in agriculture and natural sciences

27 – 28, September, 2018
Waite Campus
University of Adelaide

THE UNIVERSITY of ADELAIDE

GRDC GRAINS RESEARCH & DEVELOPMENT CORPORATION

Government of South Australia
Primary Industries and Regions SA

SARDI SOUTH AUSTRALIAN RESEARCH AND DEVELOPMENT INSTITUTE

The University wishes to acknowledge the Kaurna people, the original custodians of the Adelaide Plains and the land on which the University of Adelaide's campuses at North Terrace, Waite, Thebarton and Roseworthy are built.

Sponsors

- Academic Consortium for 21st Century
- University of Adelaide and Waite Research Institute
- University of Canterbury
- Statistical Society of Australia, South Australia branch
- Grains Research and Development Corporation
- Crawford Fund
- Rho Environmentrics

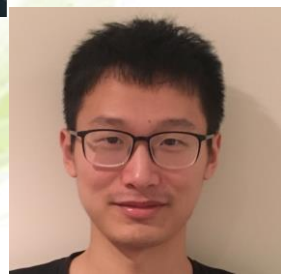
Organisers

- Jennifer Brown
- Olena Kravchuk
- Ray Correll



- Peter Kasprzak
- Wendy Li
- Enqi Wu
- Lisa Dansie
- And the wonderful team of the Biometry Hub!





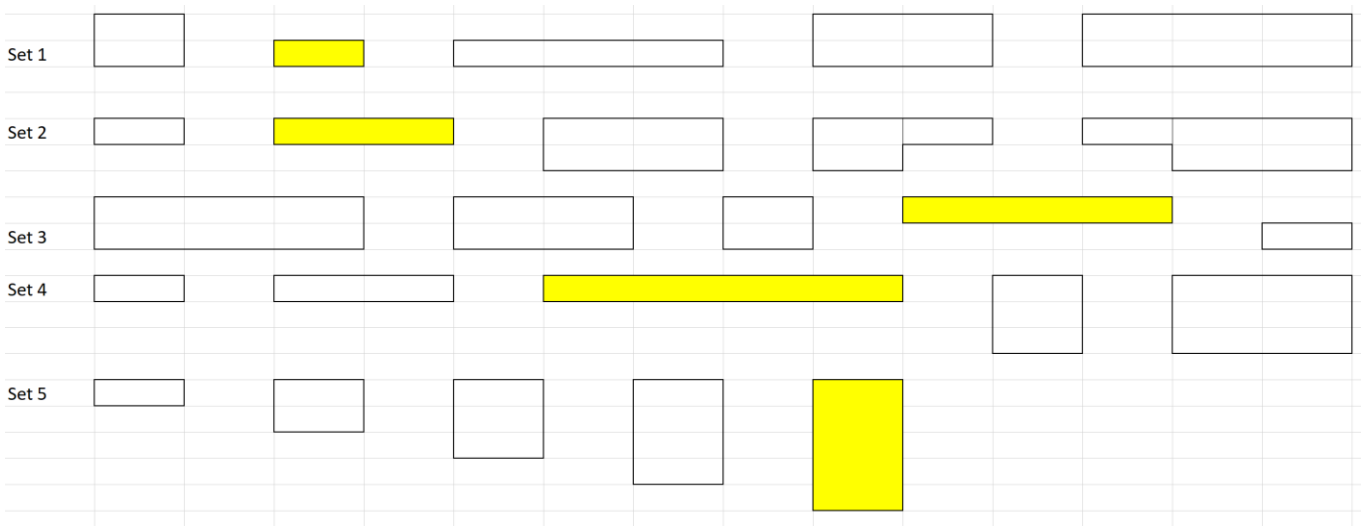

Housekeeping

- Presentations in the seminar room, Plant Genomics Centre (PG)
- Morning teas and lunches at the back of the seminar room, PG
- Round-table discussion on Thursday in the Aroma café on campus
- Symposium dinner at Edinburgh hotel
- Friday lunch tours: the Plant Accelerator or the Arboretum
- Friday afternoon tea at the back of the seminar room, PG
- Friday webinar, 7:30pm – 8:30pm in the Hub Central Seminar Room, North Terrace campus, University of Adelaide
- Internet access: eduroam or guest

Programme

- Methodological statistical talks
- Applied statistical talks
- Discussions with researchers
- Sampling day tour on Saturday in Adelaide Hills
- Social program and networking

A long time ago, in 1952…



The principle of the method can be illustrated by taking five sets of five random samples on the area to be sampled and ranking the samples in each set in order of magnitude with respect to some character. From the first set take, say, the highest ranking sample. From the second set take the second highest sample, and so on. The five samples so selected for cutting out of the 25 samples inspected would then include each rank position from the highest to the lowest.

This article first appeared in the *Australian Journal of Agricultural Research* (1952), 3, 385–390.

G.A. McIntyre in the 1950's – a rare photo from CSIRO archives

A method for unbiased selective sampling, using ranked sets

- Researcher can collect in an objective way a representative sample of field material
- also requiring less samples to be sent to laboratory in comparison to a simple random sample.

There are currently almost 1100 citations to the McIntyre's paper on Google Scholar, and about 550 on the Web of Science. The paper became classic, being cited more than 140 times only in the last two years.

The original publication was re-produced in the History Corner in *The American Statistician* in 2005 in recognition to its foundational impact to the area of Ranked Set Sampling (RSS).

Foundation of RSS

The original publication was re-produced in the History Corner in *The American Statistician* in 2005 in recognition to its foundational impact to the area of Ranked Set Sampling (RSS).

"Though there is no theoretical rigor, the work of McIntyre is pioneering and fundamental since the spores of many later developments of RSS are already contained in the paper."
(Chen et al, 2003, p. 7)

Talking directly to field researchers

- The paper contains a real-life motivation;
- Addresses practical issues and constraints;
- Suggests a possible resolution of the problem of subjective bias in field sampling;
- Demonstrates how that would work for a range of typical field scenarios.

McIntyre sparingly puts notes of caution throughout the text. He emphasises that researchers need to look at practicalities of their own projects before deciding on whether to accept the RSS in their practice.

McIntyre as an exceptional consultant

“George had an uncanny capacity for evaluation of each new problem presented to him. Indeed, many a good scientist has come away from a consultation a little wiser in their own subject, as well as with competent technical advice.” (panegyric by George’s colleagues from the Plant Industry Division in 1974)

Time of changes in sampling practices

- Random sampling in plant industries was a novelty. That approach was only recommended by the Forestry Authority in 1950.
- The contemporary sampling practice was dominated by either
  - systematic sampling or
  - subjective sampling of “representative” biological units.

The focus of McIntyre was on addressing the latter concern. He demonstrated a slight modification that could convert the inclination of biologists to search for representative samples in the field into a statistically valid and unbiased method.

Efficiency gain in pasture research

"The method will be of use where only a sample of the population is to be measured [O.K. e.g. not for map construction], where random samples can be readily ordered by visual inspection or other rough gauging method [O.K. at that time there was a strong trend to substitute human assessment with some objective measurements], and where the exact measurement of the sample is costly in time or effort."

Typical models for pasture yield in squares of various size

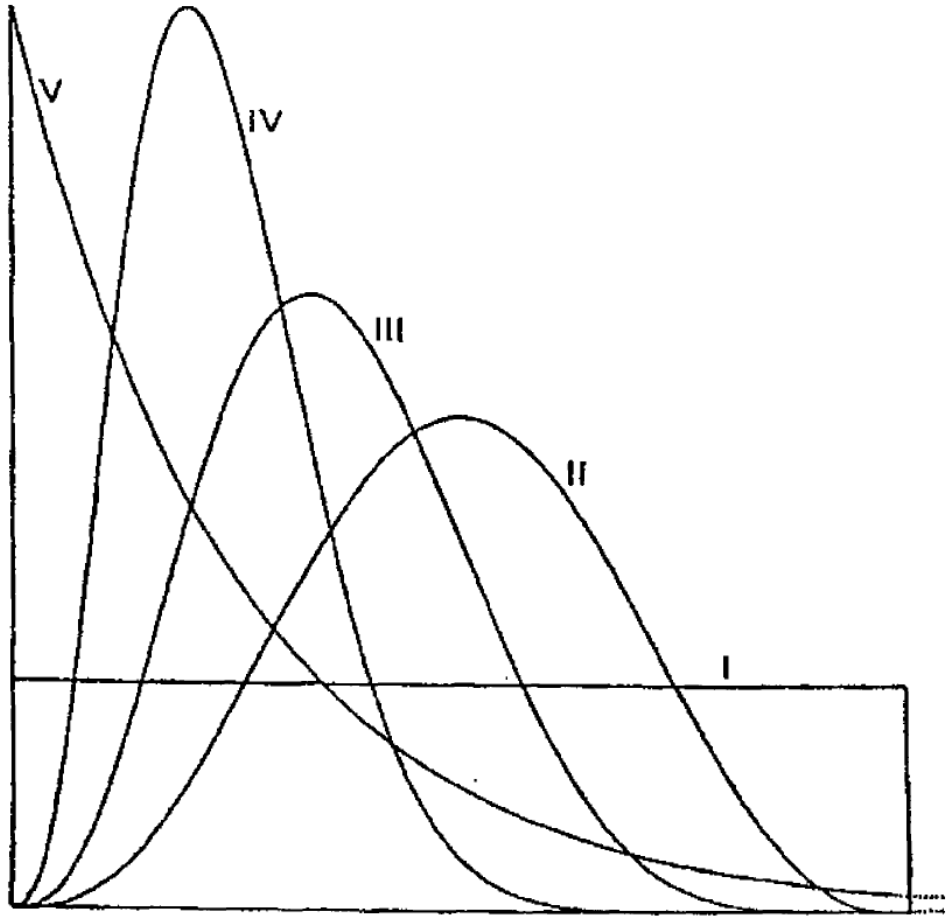


Figure 2. Arbitrary distribution selected to examine the efficiency of the ranked sample method.

McIntyre uses the beta distribution family, Beta(p,q), $q \geq p \geq 1$, to give examples of distributions coming close to what biologists would expect for the yield of pasture measured in quadrats of different size.

McIntyre examples include the uniform, symmetric and skewed Beta distributions; the exponential distribution is added as an extreme example.

Computational background for order statistics

Such a clever selection of distributions gives field researcher a visual guide to field distributions, as well as allows McIntyre to compute, ‘by integration’, the first two moments of order statistics for sample size, $2 \leq n \leq 5$.

McIntyre refers to ‘subpopulations of ranks’ rather than ‘order statistics’, but his rationale is obviously grounded on the order statistics computations.

This is all happening in the early 1950’s — his computational skills are both impressive and convincing to support that the method can be adopted in practice.

‘Ranked Set Sampling with Order Statistics Background’

Dell and Clutter (1972) communicated with McIntyre to check whether he expanded the work for a general mathematical context, and upon his negative answer presented their ‘Ranked Set Sampling with Order Statistics Background’ in *Biometrics*, somewhat reproducing McIntyre’s reasoning but with transparent and careful rigour.

Their calculations for his examples fully coincide with the original results in the McIntyre’s paper.

Why did not George get the idea to *Biometrics*?

It is not known why McIntyre decided not to put the method in a formal framework, as he did in *Biometrics* with his idea of multi-stage experiments (McIntyre, 1955).

This may be related to his cautious understanding that the efficiency of the method is sensitive to many factors of biophysical, human and cost nature.

The sheer volume of research in the last four decades produced in RSS under various generalisations provides a possible explanation why George did not embark on further theoretical developments of his idea in 1952.

Method adoption by the industry

The ranked set sampling method was gradually adopted by the Plant Industries, probably through personal meetings with McIntyre and annual CSIRO conferences in Australia.

In the 1950's, George was presenting it to his students in the Australian Forestry School as

"a method of restricted usefulness whereby one can use eye judgements to select a sample which is unbiased and whose mean has a lower standard error than a random sample." (McIntyre, Lectures in Statistical Methods in Forestry, 1957, p. 217)

Lecture notes by McIntyre

There is a method of restricted usefulness whereby one can use eye judgements to select a sample which is unbiased and whose mean has a lower standard error than a random sample. Suppose we were to locate 25 random sample plots on an aerial photograph within a height class and group these into five sets of five plots at random. If it is possible from inspection of the photograph to correctly order the plots within a set with respect to say gross bole volume and the best plot is taken in a set chosen at random, the second best in a further set and so on then the group of five plots so selected will be an unbiased sample with a lower standard error of the mean than a random sample of five plots. If the plots were grouped into sets at random and there is little or no error of ranking within sets then the efficiency of the sample so selected relative to a random sample of the same size will be slightly less than $\frac{n+1}{2}$, where n is the number of members in each set. For example with 5 plots per set the mean is almost three times as precise as random sampling. This is the maximum gain. With poor judgement in ranking the gain may be quite small and not worth the effort of selection.

Dissemination of the idea

The first citation to McIntyre (1952) appeared in 1954 in *Forestry Science* by Lawrence and Walker in their review paper on benefits from the random sampling practice to forestry inventory. Application of ranked set sampling was suggested there as an area for future research work.

11—FUTURE LINES OF RESEARCH

While the methods of assessment outlined above have become standardized, it is fully realized that there is considerable room for improvement in many directions. The two main points under consideration at the moment are the subdivision of the myrtle areas and a system of double sampling by ranked sets as described for agricultural work by McIntyre (1952). The sampling method, if successful, reduces very considerably the number of plot measurements in the field to obtain a given sampling error.

Tailoring the method to practical constraints

In 1965, Hall and Dell designed and conducted a study

"to determine whether the method could increase efficiency of sampling browse and herbage in shortleaf-loblolly pine-hardwood forests"

Their aim was to stimulate the evaluation of the method in other applications.

That call was followed by about a dozen more case-studies in range, pasture and forest management, and the method gradually grew to be included in textbooks in environmental research.

The *Environmental and Ecological Statistics* journal presented a special issue on RSS in 1995.

Ranking forage

FIGURE 1. *The ranked-set method seems efficient where extreme local variation is the main sampl ng problem. Here is a set of three quadrats for a forage inventory in a loblolly-shortleaf pine stand. The observer's task is to rank the quadrats visually as highest, intermediate, or lowest in weight of forage, and then select one for quantification.*

To rank samples in their correct order by inspection, it is necessary to have them fairly close to one another, that is, with five ranks the set of five quadrats which is being examined for the identification of quadrat of a particular rank will not be random samples from the whole plot but a cluster of samples within a restricted portion of it. If the local variability is small compared to the variation between different parts of the whole plot, then there will be a very serious loss of the advantage from the use of the method. Usually local variability is large and the sampling

RSS in modern practices

A typical statistical consultant would come across this sampling method only a few times in their careers of many years.

The absence of clear field protocols for RSS causes high variation in how the sampling method is implemented by field researchers, which affects the validity and the efficiency of the method.

It is as important today, as it was 66 years ago, for statisticians to engage with researchers in a constructive dialogue, of McIntyre's style, to enable best decisions about implementing RSS in the light of the current advances in agriculture and environment technology and research.

Acknowledgements

- Angela Mills, a librarian at the Waite library
- Many members of the anzstat mailing list
- Dr Warren Muller for his interest and personal reflections
- Dr Ray Correll and Prof Jennifer Brown for their encouragement with organising the Symposium
- Our invited speakers for communicating their ideas and suggestions
- Peter Kasprzak for his many questions and enthusiasm about the method

Special edition of Applications in ANZJS

Invitation to submit manuscript for a Special Application Issue of the Australian and New Zealand Journal of Statistics, “Ranked Set Sampling: adoption and impact”

ANZJS is inviting manuscripts for a special issue addressing adoption of ranked set sampling in the real-life sampling practice in wide biophysical context. The idea of RSS was suggested by an Australian statistician George A. McIntyre in application to plant industries in 1952. Since then, there have been a number of advances in the RSS methodology. Nevertheless, the adoption of the method by practitioners has been limited. This issue is celebrating the consulting wisdom of McIntyre by highlighting the drives and means for promoting the use of the methodology in the practice of statistics.

The editor of Applications Prof Triggs and guest editors Prof. Brown and Dr. Kravchuk are inviting applied and theoretical manuscripts that present topics in RSS immediately available for translating the methodology into research and statistics practice.

For consideration for inclusion, manuscripts must be submitted by October 1, 2019 as the usual ANZJS electronic submission, and will be peer reviewed. Authors must indicate in the cover letter to the Editor that submitting for the special issue.

Contacts: Prof. Jennifer Brown, jennifer.brown@canterbury.ac.nz or Dr. Olena Kravchuk, olena.kravchuk@adelaide.edu.au