REASONING ABOUT DISTRIBUTION: A COMPLEX PROCESS

GUEST EDITORS:

MAXINE PFANNKUCH
The University of Auckland, New Zealand
m.pfannkuch@auckland.ac.nz

CHRIS READING
SiMERR National Centre, University of New England, Australia
creading@une.edu.au

1. THE NATURE OF DISTRIBUTIONS

We are very pleased to introduce this special issue of the *Statistics Education Research Journal* (*SERJ*) on *Reasoning about Distribution*, which presents research at the forefront of building conceptual foundations for statistics education. According to Moore (1990, p. 136) statistical thinking is an "independent and fundamental intellectual method that deserves attention in the school curriculum." Equally he could have stated that statistical thinking deserves attention by research. He also hoped that "in the future pupils will bring away from their schooling a structure of thought that whispers, 'Variation matters … Why not draw a graph?'" (Moore, 1991, p. 426). With considerable foresight Moore not only encapsulated the building blocks for statistical thinking but also two deep research questions with which statistics education researchers are currently grappling: How do students actually reason about variability and distribution? How do these two types of reasoning develop?

Variation is at the heart of statistical thinking but the reasoning about variation is enabled through diagrams or displays that "represent intuitively the original reality via an intervening conceptual structure" (Fischbein, 1987, p. 165), such as graphs or frequency distributions of data. The conceptualization of variation "through a lens, which is 'distribution'" (Wild, 2005) was originally fostered by Quetelet in the 1840s (Porter, 1986). Connecting variation in nature to distribution structures was a major conceptual obstacle in the history of statistics. It was not until the end of the 19th Century that the astronomers' error curve was re-conceptualized as a distribution governing variation in social data. According to Bakker and Gravemeijer (2004) distribution is the conceptual entity for thinking about variability in data. Therefore a discussion about the nature of distributions involves both conceptual and operational aspects to be considered. A conceptual perspective focuses on clarifying what notions underpin distributions and why these notions are important whereas an operational perspective focuses on how a specific set of data is captured, displayed and manipulated by distributions. Reasoning about distributions involves interpreting a complex structure that not only includes reasoning about features such as centre, spread, density, skewness, and outliers but also involves other ideas such as sampling, population, causality and chance. These other ideas lead towards connecting empirical data with probabilistic notions, which in turn develop cognizance of empirical and theoretical distributions. In fact Bakker and Gravemeijer (2004), in the context of data analysis, believe that focusing on distribution might bring

Statistics Education Research Journal, 5(2), 4-9, http://www.stat.auckland.ac.nz/serj

more coherence to the statistics curriculum. Similarly, Scheaffer, Watkins, and Landwehr (1998, p. 17) considered that "the unifying thread throughout the probability curriculum should be the idea of distribution."

It would seem that distribution provides a strong connection between statistics and probability, a connection that is currently lacking in curricula and teaching. Distribution is a key concept in statistics yet statisticians and educators may not be aware of how difficult it is for students to develop a deep conceptual and operational understanding of distributional structures. When students are given tasks involving comparing distributions or making inferences, they often fail to utilize relevant information contained in the underlying distributions. Curricular materials often focus on construction and identification of distributions, but not on the meaning and interpretation of these distributions or on how to manipulate them to derive further information from the data. Different distributions of the same data require students not only to understand how their structures are connected but also how these different distributions may unlock different parts of the story of the same dataset.

Thus, distributions are conceptual organizing structures or mental devices that allow for a statistical intellectual method to develop. These structures are complex and subtle and require a long enculturation into understanding them. Many questions arise about conceptual, pedagogical, and research-related aspects of reasoning about distributions.

Some questions that need to be addressed by research are:

- What does distribution mean to students?
- What are the simplest forms and representations of distributions that children can understand?
- When and how do children begin to develop the idea of distribution?
- How does reasoning about distribution develop from the simplest aspects or forms of distribution to the more complex ones?
- What type of understanding of distribution is sufficient for a statistically literate person?
- What instructional tasks and technological tools can promote the understanding of distribution?
- What are the common misconceptions involved in reasoning about distribution?
- What are the difficulties that students encounter when working with, analyzing and interpreting distributions?
- How does an understanding of distribution connect and affect understanding of other statistical concepts and how does it relate to other kinds of statistical reasoning (e.g., reasoning about variation, covariation, inference)?
- What methods can be used to assess understanding of distribution?
- What are useful methodologies for studying (researching) the understanding of distribution?

2. ABOUT THIS SPECIAL ISSUE

Since reasoning about distribution is a complex and challenging research topic, this special issue presents a series of papers which address some of the questions posed above. This special issue arose from the fourth international research forum on Statistical Reasoning, Thinking and Literacy (SRTL-4) and from a subsequent call from SERJ for other researchers to submit papers on this topic. After considering "reasoning about variability" in the third forum (SRTL-3), the fourth forum (SRTL-4) held in July 2005 at The University of Auckland, New Zealand built on the core idea of variation by focusing

on “reasoning about distribution.” These SRTL forums bring together a small number of researchers whose work is focused on a particular area and presented in extended sessions that permit lengthy discussions among the participants. In addition, many researchers present primary data in the form of video clips and transcripts of students or teachers in the process of reasoning as well as discussing and explaining their actions; this allows for intensive review and discussion of findings and research methods by all participants. At SRTL-4 twenty researchers in statistics education, from six countries, discussed eight studies that examined different aspects of reasoning about distribution (Makar, 2005). After five days of presentations and discussion the participants believed that they were only in the initial stages of understanding reasoning about distribution but felt that as a community they were getting closer to important breakthroughs. The papers in this special issue represent some of the many efforts now underway to deepen our knowledge and respond to some of the challenging research questions listed earlier.

The first paper in this Special Issue, by *Wild*, a well-known statistician, is based on his opening address at SRTL-4 and is designed to delve deeply into issues of distributional reasoning and its purpose from a statistician’s perspective. He presents distribution as a lens through which variation is viewed and then discusses the conundrums of connecting empirical distributions to theoretical distributions, the position of the sampling distribution, and why all distributions are conditional. The paper by *Pfannkuch* proposes a model for reasoning from the comparison of box plots based on one secondary teacher’s articulation of these comparisons whilst teaching. This model is intended as a guide for teacher reasoning and to inform the design of teaching sequences. The paper by *Reading and Reid* describes levels of reasoning about distribution based on the SOLO taxonomy that could be used to assess students’ development of such reasoning ability and to structure learning sequences. This hierarchy emerged from the re-analysis of tertiary students’ responses that had shown various levels of reasoning about variation. The paper by *Prodromou and Pratt* reports on a virtual simulation designed to allow students to use causality to articulate features of distribution. This latest iteration of software under development acts as a “window on thinking-in-change” by allowing the students to explore the relationship between causality and variation. The paper by *Leavy* reports on the developing understanding of distribution as elementary pre-service teachers compared distributions of data that were created during practical investigations. The use of the experimental context in this study was found to support the construction of a distributional perspective.

3. EMERGING KEY THEMES

There are four themes common across these papers that have important implications for future statistics education research. These relate to research purpose, educational context, methodology and the importance of variation. First, education research is evolving to have a more cognitive focus. The purposes of the various research studies undertaken were to either describe the reasoning about distributions (Pfannkuch; Reading & Reid) or investigate ways to assist students to develop such reasoning (Prodromou & Pratt; Leavy). In unpacking the concept of distribution, Pfannkuch’s key elements of reasoning help to position “distribution” within the wider “inference” context, while Reading and Reid’s “understanding” and “using” cycles provide a cognitive developmental framework for assessing the concept. In assisting development of the concept, Prodromou and Pratt are improving a microworld to assist students in co-ordinating different perspectives of distribution, while ways of building on existing notions, as identified by Leavy, provide a foundation for creating richer learning

environments. Future research must continue to address both the assessment of reasoning about distribution, and ways of supporting the development of this reasoning.

Second, the educational context in which research is positioned is becoming increasingly important for generating meaningful qualitative data. In all four studies the concept of distribution was investigated by the researchers during learning activities and involved comparison of datasets. Increasingly research data are being collected during actual teaching/learning episodes, rather than with participants who have been withdrawn from their classes or given artificial tasks as part of a research project. This is reflected in all but one of the studies, and even then Prodromou and Pratt worked with students who were involved in a learning situation, although it was outside class time. The comparison of datasets was either explicit, as the actual task given, or implicit, as a necessary action to achieve a more general task. These studies have demonstrated that rich environments are available for collecting qualitative research data on reasoning about distribution, when learners are allowed to explore their own meanings for distribution and the necessary related reasoning process.

Third, the analysis of qualitative data is proving to be a rich source of information for investigating reasoning about concepts. The methodologies employed in all four studies reflect this recent research trend. In each case the researcher(s) analysed qualitative data based on episodes or responses that were produced during learning activities, from either the teacher's (Pfannkuch; Leavy) or the student's (Reading & Reid; Prodromou & Pratt) perspective. This often necessitates smaller sample sizes to achieve the depth of analysis desired, with the implication that findings are more in-depth but sometimes more exploratory in nature. Frameworks provided by Pfannkuch and by Reading and Reid are valuable stepping-stones to more detailed assessment of students' reasoning. The "thinking-in-change" investigated by Prodromou and Pratt provides a particularly interesting approach to the analysis of student thinking in action and could profitably be pursued by future researchers.

Finally, variation is a recurring concept in each paper. The underlying importance of variation is demonstrated in its role in the various descriptions of models, frameworks and understandings of distribution, and in supporting key decisions when adjusting distributions. Variation was acknowledged as one of the key elements in being able to reason about distributions (Pfannkuch) and, as such, it was used as an initial variable for identifying better quality student responses before searching for what constituted weaker and stronger reasoning about distribution (Reading and Reid). Increase in awareness of variation (Leary) and co-ordination of two different perspectives of variation (Prodromou & Pratt) were both found to be important in supporting the development of the concept of distribution. Thus all four studies reinforce the now generally accepted linking of the concepts of variation and distribution. Future studies of either concept should not preclude the other.

The juxtaposition of these four studies also raises a question about the connection between teaching methods and students' reasoning about distribution. Starting with students' tendency to think deterministically, Prodromou and Pratt develop a novel way of building up students' concepts of distribution. Their teaching strategy raises questions about the other research. For example, if new teaching methods different from the current practice are used by teachers and researchers: Will Reading and Reid's hierarchical model change? Will Leavy's students' reasoning show the same misconceptions? Will Pfannkuch's teacher have the same problems with her reasoning? Conversely: Will Prodromou and Pratt's method give rise to new student misconceptions? To improve teaching, future research needs to approach the problematic issue of reasoning about distribution from many different perspectives.

4. CLOSING THOUGHTS

Together these studies have provided an insight into research methodology for investigating the reasoning process, as well as detailed knowledge and frameworks on which to base investigations into the concept of distribution. The focus thus far on qualitative studies to inform exploratory research in the area of reasoning about distribution has been enlightening, but given the limitations of qualitative research, researchers now need to develop quantitative studies to substantiate the wide range of findings being espoused. At the same time it is worth recognizing that there is a noticeable trend to investigate the cognition of teachers and those training to become teachers, as well as their students. These papers suggest that such research would be profitable for the development of statistics education. Wild's paper on the concept of distribution also points to many research avenues that need to be explored and thought about by researchers.

Importantly, researchers need to expand on this useful research work on reasoning about distribution. We hope to see further papers on reasoning about distribution and related issues such as variation appearing in future issues of *SERJ*. We challenge researchers to determine when the first notions of distribution begin to develop for students and how they extend into an understanding of more complex forms. Integral to this is the need to determine how the understanding of distribution connects to and affects understanding of other statistical concepts and related statistical reasoning. In particular, statistics educators are interested in knowing about common misconceptions held, and difficulties encountered, by students when reasoning about distribution and which instructional tasks and technological tools promote a better understanding of distribution. In particular, there is a lack of research with post-secondary and college level learners, who encounter distribution and variation in a more formal context of learning about statistics, that needs to be addressed. Underlying all this work, researchers should continually strive to identify useful methodologies for studying student understanding of distribution. By responding to these challenges the statistics education research community will enrich the available knowledge relating to reasoning about distribution and thus assist statistics educators to improve the quality of learning about fundamental statistical concepts.

We appreciate the opportunity to collate and devote a set of research papers to reasoning about distribution. Especially, we value the contribution of the coeditor, Iddo Gal (University of Haifa, Israel), who co-ordinated this special issue and offered many suggestions to improve the quality of the papers. Special thanks also go to all SRTL-4 participants who contributed to the research forum discussions of earlier versions of some of the papers and to those researchers who contributed as reviewers of all papers. Readers are now invited to comment or make suggestions by contacting the authors. Finally, all researchers are invited to consider contributing to the forthcoming SRTL-5 (see "Forthcoming Conferences" in this issue), to be held in 2007 in England, which will be devoted to *Reasoning about Statistical Inference: Innovative Ways of Connecting Chance and Data.*

REFERENCES

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield, (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Fischbein, E. (1987). *Intuition in Science and Mathematics*. Dordrecht, The Netherlands: Reidel.

Makar, K. (Ed.). (2005). *Reasoning about distribution: A collection of research studies. The Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

Moore, D. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, D.C., USA: National Academy Press.

Moore, D. (1991). Statistics for all: Why? What and how? In D. Vere-Jones (Ed.) *Proceedings of the Third International Conference on Teaching Statistics (ICOTS-3)*, Dunedin, August 1990, Vol. 1 (pp. 423-428). Voorburg, The Netherlands: International Statistical Institute.

Porter, T. M. (1986). *The rise of statistical thinking 1820-1900*. Princeton, NJ: Princeton University Press.

Scheaffer, R., Watkins, A., & Landwehr, J. (1998). What every high-school graduate should know about statistics. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in Grades K-12* (pp. 3-31). Mahwah, NJ: Lawrence Erlbaum Associates.

Wild, C. J. (2005). The concept of distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. The Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.

MAXINE PFANNKUCH
Department of Statistics
The University of Auckland
Private Bag 92019
Auckland, New Zealand

CHRIS READING
National Centre of Science, Information and Communication Technology and Mathematics for Rural and Regional Australia
Education Building
University of New England
Armidale, NSW 2351, Australia