The Coordination of Distributional Thinking

Prodromou, Theodosia
Centre for New Technologies Research in Education, (CeNTRE), Institute of Education
University of Warwick, Coventry, CV4 7AL, United Kingdom
t.prodromou@warwick.ac.uk

1. An overview of the problem

My aim is to trace the thinking-in-change (Noss & Hoyles, 1996) during the co-ordination of two epistemologically distinct faces of distribution. By co-ordination here, I refer to the connection between a data-centric perspective on distribution, which identifies distribution as an aggregated set of actual outputs, and a modelling perspective on distribution, which views distribution as a set of possible outcomes and associated probabilities. The coordination requires that the learner connects in both directions the data that forms a distribution of results to make up the modelling distribution. The dual connection is, I believe, at the heart of informal inference.

2. Short review if the literature

At the heart of this study is the notion of statistical inference. Snedecor (1950) very clearly articulated that "A distinctive function of statistics is this: it enables the scientist to make a numerical evaluation of the uncertainty of his conclusion". Any kind of conclusions that are drawn before any numerical quantification of their uncertainty are subsumed under the umbrella of "informal inference". When a statistician analyses data for research purposes, the intention is to extrapolate the findings from a sample of individuals to the population of all similar individuals. Statistical inference (formal inference), therefore, is concerned with drawing conclusions, from observed data, about unobserved quantities: These conclusions may be distinguished into 1) conclusions on observable quantities, such as future observations of a process and 2) conclusions on quantities that are not directly observable, that is, parameters that governed the hypothetical process leading to the observed data (e.g. regression coefficients influencing the relation between independent and dependent variables; or, the probability of an event, which may be observed or not in a series of Bernoulli trials). Statistical conclusions about the unobserved data are made in terms of probability statements because there is no impediment in principle to fitting models with one parameter or many parameters and various probability specifications.

The above idea perhaps encapsulates what it is that expert statisticians do. But how do such powerful ideas emerge out of less sophisticated coherent views on inference?

Pfannkuch (2007) defines informal inference as the way "to describe the drawing of conclusions from data that is based mainly on looking at, comparing, and reasoning from distributions of data" (p. 1). She has perceived informal reasoning to be "interconnected to reasoning from distributions, reasoning with measures of centre, and sampling reasoning within an empirical enquiry cycle " (p. 1). Underlying this reasoning is the consideration of variation (Pfannkuch, 2005b).

Pfannkuch (2007) referred to a personal communication from Professor Dave Pratt (7 July 2005) in which he suggested that "Informal inference requires students to be aware of the game being played. He believes a part of students' difficulty in understanding the reasoning being used is that the game being played is not being made explicit to the students by the teacher. The students may believe that they are reasoning only about the data under consideration, which Pratt refers to as game *one*, whereas the teacher believes that the data are a sample from a population, which Pratt calls game *two*. It is the playing of game *two* that will lead students towards formal inferential reasoning" (p. 1).

Previous research suggests that early teaching of statistics should focus on informal methods of exploratory data analysis (Konold and Pollatsek, 2002). Pfannkuch, Rubin, and Yoon (2002) claim that those methods should be kept separate from probability, with only informal quantifications of variability to indicate the propensity or a spread of one sample distribution compared to another. However, Prodromou and Pratt (2006), taking into consideration that EDA is developing interesting pedagogic approaches towards informal inference, ask whether students can develop an appreciation of the robustness (in the sense of stability of conclusions for a variety of multiple differing analyses) of their inferences without constructing a *modelling* perspective alongside their *data-centric* perspective. According to their view this question is crucial to tax research in informal inference.

3. Research study

The whole study falls into the category of *design experiments* (Cobb et al, 2003). Design experiments aim to sensitise towards the complex *learning ecology* of the domain being investigated through iterative design. It also highlights the delicate process of *phenomenalising* (Pratt, 1998) a mathematical concept that can capture learners' needs by transforming powerful ideas into situated, meaningful and manipulable phenomena.

Influenced by the Constructionists' (Harel & Papert, 1991) accent on the affective, I have been concerned to place emphasis on developing activities for playful contexts in which students are likely to construct *purpose*, while at the same time coming to appreciate the *utility* (Ainley, Pratt & Hansen, 2006) of the modeling distribution and the data-centric distribution as central concepts.

According to these aims the Basketball microworld has been developed to investigate individual' ideas on several aspects of probability and modeling. The microworld has been substantially revised after each cycle of investigation and is now in its fourth and final version. Preliminary results of the fourth iteration form the basis of this paper.

This version of the microworld was used by eight pair of students (aged between 14 and 15 years old) in a UK secondary school. Each pair of students worked with the microworld for three sessions lasting about 270 minutes in total. Data on students' behaviour consists of their on-screen activity which was captured on video-tape and transcriptions of those sections to generate plain accounts of the sessions. Screenshots are incorporated as necessary to make sense of the transcriptions. Subsequently, the transcriptions are analyzed in attempts to produce extended narrative accounts for the students' actions and articulations.

In the third iteration, Prodromou and Pratt (2006) reported on how they designed a microworld that aspired to research thinking-in-change (Noss & Hoyles, 1996) about distribution. Their premise, in line with a constructivist approach and Pratt's (1998) prior work, was that thinking about distribution must develop from causal meanings already established. They conjectured that they would be able to build an environment that would enable the students to exploit their appreciation of the limited explanatory of causality to capture the essence of local variation and use this causality to construct new situated meanings for the distributions of throws and success rates. They demonstrated how the provided on-screen control mechanisms for average and spread could be deterministic or subject to stochastic error. The students used these controls to recognize that, while at the micro-level causality is shown to have limited explanatory power, at the global level, it can be harnessed to articulate the relationship between the parameters of the model (average and spread) and the emergence of distributional patterns.

In that study, Prodromou and Pratt suggested that the concept of distribution lies in co-ordinating the emergent *data-centric* and *modelling* perspectives for distribution. They regarded this paradox of seeing the limitations of causality at one level while recognizing its power at another level as the heart of coordinating the two perspectives on distribution. They asked whether and how do students co-ordinate the *data-centric* and modeling perspectives on distribution but they were not able at that stage of their research to elaborate on this aim to their complete satisfaction. Nevertheless, the results of the third iteration indicated support for their conjecture that it is possible to design an environment in which students' well-established causal meanings can be exploited to co-ordinate *data-centric* and modeling aspects of distribution. The students appreciated how not only themselves could be agents of variation, but also randomness, instantiated in the form of the quasi-concrete arrows, could create histograms in which variation is apparent. In this sense, randomness might become understood as reality once removed.

On this paper, I describe the insights gained from the participants' interactions with the two perspectives on distribution in the fourth cycle of investigations and trace their distributional thinking-in-change (Noss & Hoyles, 1996) in informal inference. The term "distributional thinking" is in fact the co-ordination of the two faces of distribution.

When students in this study pay attention to the emerging data, they consider the *data-centric* perspective on distribution that is the set of outcomes, so students play what Pratt calls game *one*. When they have access to both the *modelling* distribution and the *data-centric* distribution, they use information from a sample to make inference about the wider population in the background from which the data was sampled; this is what Pratt calls game *two*. Taking into account that formal inference adds an emphasis on substantiating students' conclusions by probability calculations, the students are encouraged to make a connection between the *data-centric* distribution and a probability model for the data.

The conjecture investigated in the present paper is that tools such as those found in the Basketball microworld support students to

1) make these two types of connections,
2) connect chance models and data by taking chance variation into consideration and distinguish between patterns that are consistent or inconsistent with the random location scenario,
3) reason intuitively about the most prominent types of statistical inference that are based on the sampling distributions of statistics, and
4) grasp the "subtle" ideas that lie behind the reasoning used in statistical inference.

4. Discussion

Previous stages of the study let the author develop and redesign a microworld, which can be used systematically to test out the conjecture of the fourth cycle of investigation referred to above. Prodromou (in press) reports on how students make an intuitive synthesis of the *modelling* and *data-centric* perspectives on distribution which can be schematised by the structural model as depicted in Fig.1. The model shows that the students can perceive of the *modelling* distribution (MD) as the intended outcome and the data distribution (DD) as the actual outcome, suggesting a connection being made, when the modeling distribution in some sense generates the data. The opposite connection from data to the probability distribution is made when students perceive of the *modelling* distribution (MD) as the target to which the data distribution (DD) is directed. These two types of connection appear to be central in informal inference. In fact, the bridging of data to the *modelling* perspective on distribution can act as a pathway from students' informal inferential statistical reasoning towards a more formal level, not only in terms of the discipline, but also in terms of students' cognition.

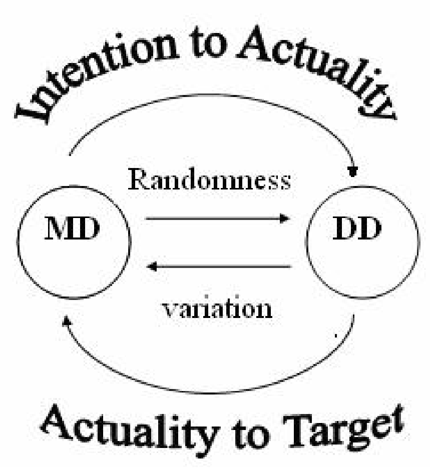

Figure 1: A tentative model for the connection of the data-centric and modeling perspectives on distribution.

REFERENCES

Ainley, J., Pratt, D. and Hansen, A. (2006). Connecting engagement and focus in pedagogic task design, *British Educational Research Journal*. 32(1), 21-36.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research, *Educational Researcher*, 32(1), 9–13.

Harel, I., & Papert, S. (1991). *Constructionism*. Norwood, New Jersey: Ablex.

Konnold, C., & Pollatsek, A. (2002). Data Analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*. 33(4), 259-289.

Noss, R. & Hoyles, C. (1996). *Windows on mathematical meanings: Learning cultures and computers*. London: Kluwer Academic Publishers.

Pfunnkuch, M. (2005b). Informal inferential reasoning: A case study. In K. Makar (Ed.), *Reasoning About Distribution: A Collection of Research Studies. The Fourth International Forum on Statistical Reasoning, Thinking and Literacy*, 2-7 July, 2005, Auckland, New Zealand [CD-ROM].

Pfunnkuch, M. (2007). Informal inferential reasoning. *Proceedings of the Seventh International Conference on Teaching Statistics,* 2007, Salvador. Bahia: Brazil.

Pfunnkuch, M., Rubick, A., & Yoon, C. (2002). Statistical thinking: An exploration into students' variation-type thinking. *New England Mathematics Journal,* 34(2), 82-98.

Pratt, D. (1998). *The construction of meanings IN and FOR a stochastic domain of abstraction*, Unpublished Doctoral Thesis, Institute of Education, University of London, May 1998.

Prodromou, T. (in press). Making Connections Between the Two Perspectives on Distribution. *Proceedings of the Fifth Conference Society for Research in Mathematics Education*, Larnaca: Cyprus.

Prodromou, T. & Pratt, D. (2006). The Role of Causality in the Co-ordination of the Two Perspectives on Distribution within a Virtual Simulation. Statistics Education Research Journal.

Snedecor, G. W. (1950). The statistical part of the scientific method. *Ann. NY. Acad. Sci.,* 52, 792-9.