CONSTRUCTING, REFINING AND VALIDATING A TASK FOR DEVELOPING REASONING ON STABILIZED FREQUENCY DISTRIBUTIONS IN THE CONTEXT OF INFORMAL INFERENCES

Ana Serradó Bayés
La Salle-Buen Consejo, Spain
ana.serrado@gm.uca.es

In this paper we present the construction, refining and validation of a task to reason about sampling and stabilized relative frequencies distribution to be included in a possible learning trajectory in Secondary School that facilitates the link between Informal and Formal Inference. The task consists of an informal modeling process of the relative frequency of appearance of each vowel in strings of characters, with a pseudo-concrete model developed through a statistical process of investigation, statistical modeling and validation through the analysis of animations. Animations of the relative frequency distributions when varying the sample size were created using the dynamic open-source software, Geogebra, and were incorporated in an online quiz, constructed within the virtual learning environment, Moodle. The task, its learning goals and assessment activities were validated with n=49 Spanish students of Grade 9 in order to draw conclusions about the conjectured learning process.

INTRODUCTION

In the last few years there has been an increasing interest in research of how researchers, teachers and designers might plan tasks in order to truly engage students in knowledge construction. Acknowledging this interest, we present the construction, refining and validation of a task and its animations for reasoning about sampling and stabilized relative frequencies distribution to be included in a possible learning trajectory in Secondary School that facilitates the link between Informal and Formal Inference.

The Spanish Curriculum of Compulsory Secondary School indicates that students of grade 9 (ages 14 and 15) should use comparisons, formulate conjectures, reason about the accuracy and representativeness of a sample, and formalize the probability through the classical and frequentist approach of relative frequency estimations (Batanero et al., 2013). In the Spanish curricular design and its textbooks, this estimation is taken as the definition of the mathematical value, raising serious epistemological, ontogenic and didactic problems (Chaput et al., 2011; Serradó, 2005). With the aim of trying to overcome these obstacles, other countries have opted for a modelling perspective, where probability is a theoretical value of the degree of confidence that one can give to a random outcome obtained by an observation of the stabilized relative frequency when the same random experiment is repeated a larger number of times under the same conditions (Chaput et al., 2011).

This perspective concurs with the general process of contemporary statistical thinking, and contributes to the learning of a modelling process through the development of statistical investigations. The concepts of sampling, variation and distribution are key when engaging students in this modelling process (Wild, 2006). When modelling, students are asked to reason about distributions which implies stablishing relations between the data (data distribution), the population (distribution of probabilities) and the samples (sampling distribution). But, this kind of reasoning is introduced in the Spanish Curriculum three years later, in grade 11 (ages 17 and 18), in the context of Statistical Inference (Batanero et al., 2013). As a consequence, we are exploring here the possibility of introducing the stabilized relative frequencies distribution in the context of the Informal Inference Reasoning (IIR) through involving students in an informal modeling process. In this context, we make the following conjecture: *A task that engages students in the informal statistical modelling process of "growing samples" may help them to reason about sampling and stabilized relative frequency distribution in the context of IIR.*

SAMPLING AND RELATIVE FREQUENCY DISTRIBUTION FRAMEWORK

Bakker and Gravemeijer (2004) presented a theoretical framework that allows analyzing the relation between data and distribution in an informal way. They examined aspects as center,

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

spread, density (relative frequency,..), and skewness to structure the relationship between data (individual value), and distribution (conceptual entity). This structure can be read upward, from data to distribution, which is typical for novices in statistics; or downward, from distribution to data, in which case one uses probability distributions to model data.

Bakker and Gravemeijer’s (2004) theoretical framework provides a bridge between frequency, stabilized relative frequency and distribution. However, if we want to establish a bridge between exploratory data analysis in the context of IIR and statistical inference, then it is useful to consider the framework developed by Ben Zvi, Gil and Apel (2007). They analyse the cognitive aspect of distribution in the IIR context as: reasoning about variability, distributional reasoning, reasoning about signal and noise, contextual reasoning and graph comprehension. These two frameworks only provide part of the picture underlying the conceptual structure of distributions, but are not adequate when thinking about, exploring and describing distributions. Arnold and Pfannkuch (2012) propose the Distribution Description Framework (DDF) organized by: (1) overaching statistical concepts that underpin distribution, (2) characteristics of distribution, and (3) the specific features that are used when describing distributions. In these frameworks, there is no reference to the concepts underlying sampling reasoning: sampling size, random process, distribution, intuitive confidence interval, and relationship between sample and population (Dierdrop et al., 2012). We propose a theoretical framework of the cognitive aspects related to sampling: contextual knowledge (samples, population, sample size,…), distributional (error, reliability, law of large numbers), graph comprehension (smothing), variability (tendencies, intuitive confidence intervals).

The integration of the Distribution Description Framework (DDF) and cognitive aspects of the sample let us to propose the Stabilized Relative Frequency Distribution Description Framework (SRFD), in which we describe theoretically the overaching statistical concept and the features of the distribution: contextual knowledge (population, sample, sample size, variable, interpretation and explanation), distributional (shape, skewness, error, reliability, individual cases, law of large numbers), graph comprehension (decoding visual shape, unusual features, smothing, comparing samples), variability (spread, density, tendencies, intuitive confidence intervals) and signal and noise (center, modal clumps).

METHODOLOGY

To answer the main question of how students can develop a notion of stabilized relative frequency distribution, we carried out a design-based research study (e.g. Bakker & Gravemeijer, 2004). In line with the principles of Realistic Mathematics Education, we looked for ways to guide students in being active learners to abstract similarities. The teaching experiment involved two teachers: a main teacher, author of this paper, and a second one participating as assistant teacher, helping some students with learning difficulties. The 49 students composed two groups of *n=25* and *n=24* Grade 9 students (ages 14,15) from a Spanish Compulsory Secondary School in a low socio-economic coast city. Each classroom group was also subdivided to create hetereogenous groups for deliberative dialogue in a cooperative learning environment. Groups were coded as: Classroom A: Ai, from i=1 to 6; Classroom B: Bi, from i=1 to 5.

The students had not encountered statistics before, except for some descriptive statistics using center parameters, and graphs and sector diagrams. The teaching experiment lasted 11 sessions of 60 minutes each, in which research data was collected and coded using the SRFD description framework. The data collected included video recordings of general group work and small group work interactions with the teachers, student work, field notes, and individual interviews with a few students to clarify their responses, during the retrospective analysis of the data. In designing research, when analysing student work we usually use the final product, but in this paper we consider the need for obtaning information about the deliberative process. So, we included the task and animations in an on-line quiz constructed using the virtual learning environment Moodle. And, we asked students to send any partial animation used or idea developed. The retrospective analysis presented in this paper aims at providing an answer to the following question: *Do the animations (included in the task) help students to reason about sampling and stabilized relative frequencies?*

Task Design: Embodying the Conjecture

To assist students in exploring data and developing the concept of stabilized relative frequency distribution, we decided to use a four-stages task. The task was designed to generate activity which affords students the opportunity to encounter new statistical ideas and strategies, but also new modes of enquiry. Therefore, all the activity proposed engaged students in a deliberative dialogue, through involving them in answering some questions. The main problem posed to students was: “Can I guess which language my friend is speaking only by counting the vowels?”. This problem generated three questions to be answered by students, which guided their investigations and learning: (Q1) which vowels are not used when writing a mobile message? (Q2) Which problem can we plan to get to know what happens in Spanish? (Q3) Can we figure out which language a friend is speaking only by counting the vowels?

The first question (Q1) was introduced to contextualize the task in an environment where students have to think why certain vowels in Spanish are not used in SMS or Whatsapps. For example, the Spanish sentence “Mi casa es blanca” is written using sms language as “Mi cas s blnc”, omitting those vowels that intuitively appear more often in our language. This intuition leads us to think that students in this case may have an everyday sense of density, a key aspect for constructing the notion of distribution (Bakker & Gravemeijer, 2004). As it is stated in the conjecture, the second question (Q2) allows introducing an informal modeling process with three stages: a pseudo-concrete model developed through a statistical process of investigation of the vowels that appear in chains of characters, statistical modeling, and validation through the analysis of animations.

The learning objectives of these three stages of the task are: (*task stage 1*) formulate questions related to a context, devise a data collection plan coherent with the analysis and procedures planned and the hypotheses formulated, gain appreciation of the important role of the sample size in statistical sampling, analyze and interpret the collected data using Geogebra; (*task stage 2*) enhance understanding of the role of sample size on sample representativeness and variability, increase recognition of the potential for bias due to poor sampling design, enhance understanding of relative frequency and stabilized relative frequency distribution, reduce mistrust of single random sample, develop facility with Java Animations for analyzing data patterns, construct the Law of Large Numbers through Java Animations; and (*task stage 3*) compare distributions through Java animations, enhance recognition of the center and spread aspects of a relative frequency distribution, validate and interpret preliminary hypotheses using other samples, gain knowledge of bias error. The aim of *task stage 4* is to assess if students are able to transfer the informal modeling process developed to answer question (Q3) posed to them. Furthermore, it includes some questions to assess the purpose and utility expressed by the students about the task, in general, and the animations in particular, in the sense of Ainley, Pratt and Hansen (2004).

Geogebra Animations Construction

Although it has been largely researched how Tinkerplots or Fathom can help students to explore data analysis and construct the notion of distribution and sampling, Spain’s educational policy does not allow teachers to use any kind of commercial software at school. Thus, we decided to use the affordances of Geogebra, an open-source dynamic software, which has been enhanced with the incorporation of tools for descriptive statistical analysis. We selected samples of texts from 1 to 10000 characters, each character was numerically coded, we counted the relative frequencies of appearance of each vowel (coded number) in relation with the total number of characters, and plotted a bar diagram. In this proposal, we can find two main differences compared to the traditional simulations of the Law of large Numbers. First, animations were constructed using real samples of texts, not through the mathematical model previously constructed, giving them an upward vision of the structure from data to distribution (Bakker & Gravemeijer, 2004). Second, the distributions were not automatically animated. Students were able to animate them using a vertical scroll-bar that could be moved freely from sample size of characters 1 to 10000, or stop in any particular sample size, helping the analysis of the individual cases (Ben Zvi et al., 2007). The first animation (Figure 1, left) used only one sample of text and “growing samples” of characters from 1 to 10000. The second animation (Figure 1, right) used three samples of texts “and growing samples” of characters from 1 to 10000.

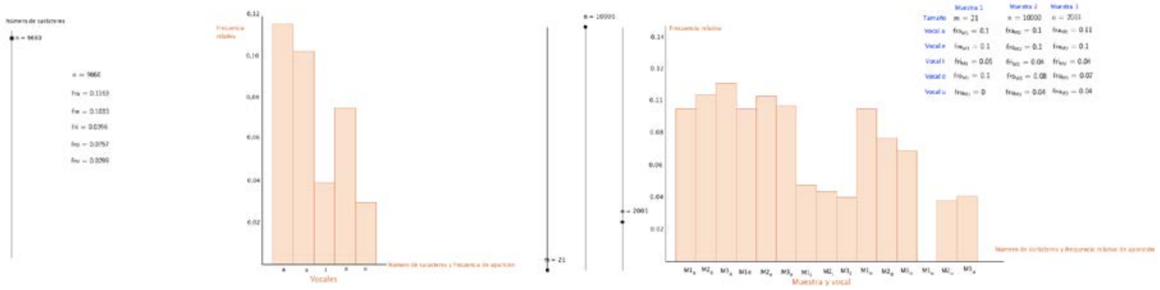

Figure 1. Left: Animation I, Analyzing tendencies. Right: Animation II, Comparing samples

In coherence with the framework introduced, the animations were created to help students decode visual shapes, analyze unusual features, infer smoothing, and compare graphs. Furthermore, students were asked to answer some questions, carefully designed to promote exploration, description and thinking about the animations used (Arnold & Pfannkuch, 2012). Students had to explore and describe *"how the distribution of frequencies varies in relation with the sample size of characters"*, interconnecting sampling, variation and distribution (Wild, 2006). They also had to investigate *"what happens when the sample size tends to 10000"*, considering the quotidian use of the word tend (Arnold & Pfannkuch, 2012). Students were also asked to improve their initial hypotheses. They were asked to answer reflective questions like: *"What can you conclude? Do you think that your conclusions would be different if you selected another text? How do you think you can validate your conclusion?"* All these questions were created before they were used in the classroom. We use them in the retrospective analysis to understand how the animations were used to explore and describe the distributions, and to validate their embodiment in the task.

RESULTS AND DISCUSSION

We structure the section in three episodes: the creation of the need for the animation, students' reasoning using the animations, and students' expressions about the purpose and utility of the animations.

The Need for the Animations

In Task Stage 1, with the aim of involving students in posing statistical questions, they were asked to answer the question *"Which problem can we plan to get to know what happens in Spanish?"* Group A posed the problem *"Is letter A the one that appears more often?"*, taking into consideration their contextual intuitions. In Group B, after a discussion of whether it was the letter a or e the one that appears more often, students posed the following problem "*Which is the letter with the most frequent appearance?"* When presenting the conclusions of their problems, none of the students in either group A or group B used a graph. Both classes argued that they had adequate information in a table, to summarize the frequency or relative frequency using different samples of characters. When the teacher asked: *"Which do you think is the most useful graph for analysing larger samples?*", all the students agreed that the most useful graph is a *"bar graph"* (Cristina, group A).

Reasoning with the Animation

In Task Stage 2, students used animation I to analyse variation in relation to the sample size, the distribution and the tendencies. We found differences between students of the two groups, A and B. Group A students A, when solving their problem (*"Is letter A the one that appears more often?"),* decoded the graph by looking only at what happens to letter A when increasing the sample size. And they concluded: *"When the sample size tends to 10000, the frequency of a is bigger than in the other graphs"* (Group A3); *"The samples are similar, but they are really different, giving us a different relative frequency, although similar"* (Group A3); *"The distribution of frequencies in relation with the sample size of characters has little variation because all the graphs are similar except the first of 1261"* (Group A6). The use of the word "similar" leads us to think that students develop some sense of stability. The animation has also helped them to look for unusual features and individual cases, possibly developing their informal distributional reasoning. However, Group B students' reasoning was different, maybe due to the fact that their problem was *"Which is the letter that appears more often?"* Students in Group B looked for differences between variables: *"The difference between A and E respect at the number of characters is that increases or decreases in relation with the other vowels"* (Group B2); *"When increasing the sample size to 10000, we discover that the number of the vowel A overcome while the rest of vowels decreases"* (Group B5). We agree with Pfannkuch et al. (2010), who concluded that different wording of questions promotes different ideas about distribution. The question posed by group A mainly promotes a distributional reasoning analysis of individual cases, while the question posed by group B seems to facilitate contextual knowledge reasoning relating the notions of sampling and variable.

Group A1 and B1 answered the question of tendencies by assigning the value of the relative frequency for a sample size 10000. For example: *"A is still the vowel used more often (0.1151), next is the vowel e (0.103), next is o (0.0765), then i (0.0402) and at the end u (0.0302)"* (Group B1). Thinking that assigning a value could occasion in the future an epistemological obstacle (Chaput et al., 2011), students were interviewed about their conclusion:

Silvia: When increasing the number of characters, the differences are small.
Teacher: Which differences?
Silvia: Between the value and the frequency. It is more precise.
Teacher: What is more precise?
Silvia: As we have more examples, the graph has more resolution.
Teacher: What does more resolution mean?
Silvia: It is more equal. There is less variation.
Teacher: What word of physics could we use to explain that there is less variation? (Silence) Can I use the word stable?
Silvia: Yes, I agree.

The use of the expressions "difference" or "more equal" let us to think that some students started to reason about range as a characteristic of the spread of a sampling distribution. And when the teacher asked the student to explain the word "precise", the argumentation of the student focused on the reliability of larger samples. But, it is when decoding the shapes of Animation II that the students reasoned about variability and sampling, incorporating in their language the concepts of stability or reliability: *"The result would be more reliable because it has more data"* (Group B3); *"When the sample size tends to 10000, the frequencies remain more stable in relation with the other samples"* (Group A1). On the last description, it is our hope as teachers that the students would shift the shape of the stabilized relative frequencies distribution instead of the distribution of size 10000.

Animation II also helped students to reason about the variability of samples of texts, examining aspects such as centre when the sample tends to 10000: *"In the third sample A is the smallest, in sample two it is in the middle, and in sample 3 it is the biggest"* (Group B1). We can assume here that the students are trying to describe some kind of an interval, an unknown concept for them. We interviewed them asking how they have come to this conclusion, and they explained: *"comparing the different vowels and samples, and having the same number of characters, and ordering from smaller to larger the number of repeated times in proportion with the number of characters of every sample"* (Anabel, B1). Again, the student presents a description of an interval where she considers the values of the confidence interval of a sample size of 10000. In this case, we agree with Bakker and Gravemeijer (2004), that comparing gives students the challenge of understanding other dimensions of a distribution such as centre or spread. In this case, when the comparison is made with a large sample of distributions, it gives the students the challenge of having an aggregate knowledge of sampling distributions, a key idea for probability understanding and hypothesis testing in an inferential context (Wild, 2006).

The Purpose and Utility of the Animations

We find three types of responses to the question: "*Why were animations useful for students?"* The first type is responses which point to the usefulness of animations in solving the initial problem posed to students. The second type is responses which stress the utility of animations in the construction of meanings: density as a property of a distribution, sample size, samples, and developing strategies as comparing graphs. Two students also inform us of the utility of the animation in improving the process of "*collecting data*" (Carlos group A) and "*in developing clearer hypotheses*" (Francisco group A). In the words of Ainley et al (2004), we can conclude that the teacher's purpose when planning the task coincides with the utility established by the students.

CONCLUSION

The four-stage task has addressed multiple and complex purposes as the development of the conceptual understanding of sampling and stabilized relative frequency distributions, procedural fluency in using the language of distributions, strategic competence of statistical problem solving and informal modeling, adaptive reasoning due to the context and the animations, and productive disposition to describe and think statistically. We conjecture that active engagement in the task of an informal statistical modelling process of growing samples helps students to reason about sampling and stabilized relative frequencies distribution and to construct the meaning of sampling distributions in the framework of Informal Inferential Reasoning (IIR). The work with animations is crucial when comparing distributions in order to relate the concepts of sampling and distribution. However, involving students only in this task is not sufficient in a learning trajectory that aims to connect informal and formal inference. We conjecture that this connection should be made by a series of successive tasks in which students engage in reasoning about sampling and distributions in statistical problem solving processes, then get involved in informal modeling processes of "growing samples", and finally model real data to construct the notions of data, probability, and sampling distribution.

REFERENCES

Ainley, J., Pratt, D., & Hansen, A. (2004). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23-38.

Arnold, P., & Pfannkuch, M. (2012). The Language of Shape. In D. Ben-Zvi, & K. Makar (Eds.), *Teaching and Learning of Statistics. Topic Study Group 12. 12th International Congress on Mathematical Education (ICME-12)*, (pp. 181-190). Coex, Seoul, Korea.

Bakker, A., & Gravemeijer, K. P. (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: Kluwer Academic Publishers.

Batanero, C., Ortiz, J., Roa, R., & Serrano, L. (2013). La Statistique dans le Curriculum en Espagne. *Statistique et Enseignement, 4*(1), 89-106.

Ben Zvi, D., Gil, E., & Apel, N. (2007, August). *What is hidden beyond the data? Helping young students to reason and argue about some wither universe*. Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), August 11-17. UK: University of Warwick.

Chaput, B., Girard, J. C., & Henry, M. (2011). Frequentist Approach: Modelling and Simulation in Statistics and Probability Teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School. Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study*. (pp. 85-96). Dordrecht, The Netherlands: Springer.

Dierdrop, A., Bakker, A., van Mannen, J., & Eijkelhof, H. (2012). Supporting students to develop concepts underlying sampling and to shuttle between contextual and statistical spheres. In D. Ben Zvi, & K. Makar (Eds.), *Teaching and Learning of Statistics. Topic Study Group 12. 12th International Congress on Mathematical Education (ICME-12)*, (pp. 249-258). COEX,Seoul, Korea.

Pfannkuch, M., Regan, M., Wild, C., & Horton, N. J. (2010). Telling Data Stories: Essential Dialogues for Comparative Reasoning. *Journal of Statistics Education, 18*(1), 1-38.

Serradó, A., Cardeñoso, J. M., & Azcárate, P. (2005). Obstacles in the learning of probabilistic knowledge: influence from the textbooks. *Statistics Education Research Journal, 4*(2), 59-81.

Wild, C. (2006). The Concept of Distribution. *Statistics Education Research Journal, 5*(2), 10-26.