TEACHING DISCRETE DISTRIBUTIONS USING CONTINGENT TEACHING WITH CLICKERS

Wayne Stewart and *Sepideh Stewart*
University of Oklahoma, USA
wayne.s.stewart@ou.edu

In most undergraduate statistics courses students are introduced to discrete random variables which are later accompanied by a list of probability functions and examples of their applications. The students are required to determine from a story problem what discrete probability distribution is appropriate. This process involves appreciating the defining characteristics of the distributions and distilling from the problem those characteristics that enable the learner to decide on the correct distribution. This initial classification which is an important first step in solving the problem must be accompanied by more guidance on how to proceed. This study will describe a teaching routine which uses contingent teaching in conjunction with clicker provided feedback from students to enhance students' categorization strategies.

INTRODUCTION

Discrete distributions are important in many statistical applications, and form an integral part of undergraduate and graduate statistics courses. They are useful in their own right but also have great pedagogical use in introducing students to the idea of a formulaic probability function with known moments and distributional characteristics that do not require integration to acquire useful information like probability and population moments. These distributions are often introduced before probability density functions, after discrete random variables have been defined. In our experience of teaching discrete distributions, students often find distinguishing between the distributions difficult and consequently are unable to select the correct distribution to solve a particular problem. Wroughton and Cole (2013) in examining three of the distributions (Negative Binomial, Binomial and Hypergeometric) found that students have difficulties in distinguishing between them. They noted that: "students had a tendency to believe the distribution would be Binomial more than any other of these three distributions" (p. 5). To resolve this matter Wroughton and Cole used card activities to teach the differences between the distributions.

In this article we will discuss an application of clickers to help students to recognize a distribution in a given statistics problem. We will employ a model built on Schoenfeld's (2002) decision making teaching routine to better target areas of weakness in students' understanding and address them dynamically.

THE DISTRIBUTIONS

The following is a list of seven distributions commonly encountered in text books: *Bernoulli; Binomial; Multinomial; Negative Binomial; Geometric; Hypergeometric* and *Poisson.*

These distributions can be seen in some cases as connected and advancing, for example the Binomial is the repetition of n independent Bernoulli experiments, the multinomial is the extension of the Binomial outcome categories from two to k, with probability of outcome i, p_i.

DEVELOPING A CLASSIFICATION ALGORITHM

The distinguishing characteristics of the distributions can be clarified by finding the parts of the descriptions that are the same and then noting differences. In this approach a particular problem can be classified. The binomial, negative binomial and geometric distributions are all repeated independent Bernoulli experiments (see figure 1). The geometric is a special case of the negative binomial where the random variable is the number of trials till the first success. The three distributions differ in that the negative binomial and geometric do not have the number of trials fixed as is the case of the binomial, the random variable in the case of the negative binomial is the number of trials till the rth success.

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

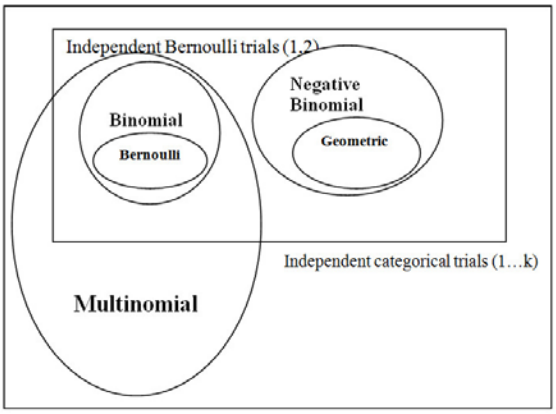

Figure 1. Venn diagram showing commonalities and differences in some of the distributions.

The multinomial is related to the binomial in that each trial can result in one of k categories (see figure 2) rather than two in the case of the binomial, the trials are not repeated independent Bernoulli but repeated independent categorical distributions, the Bernoulli then is seen as a special case of the categorical distribution. This distribution along with all the Bernoulli based distributions utilize at least one probability of success (k of them in a Multinomial).

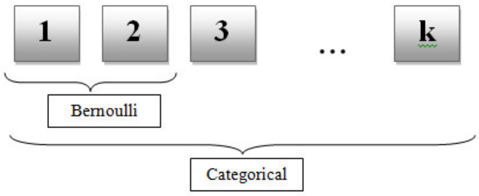

Figure 2. The Bernoulli is a special case of the categorical distribution.

The hyper-geometric distribution can be viewed as a series of non-identical, non-independent success/fail trials, but the probability of a success is not constant due to non-replacement with a relatively large sample size (n/N>0.05). This non-independence is a clarifying difference. There is the possibility that students could confuse this type of problem with a Binomial by not observing the dependence issue and the fact that there is no constant p (probability of success) which is related to the dependent samples. This would be observed however, once the formula for a hyper-geometric was viewed.

The Poisson distribution has an average rate of events, this rate is over a unit of measurement (say time). The rate and no p are distinguishing features which can be used to aid categorization.

THE CATEGORIZATION ALGORITHM

The following is an algorithm that a learner may invoke while attempting to categorize a problem into one of the seven discrete distributions (see table 1). Although, there are other ways of expressing the commonalities and differences, the best choice would be the one that most easily and consistently secures the correct categorization. The algorithm below is one of many contending algorithms:

Are there independent categorical trials?
- a. Yes
  - i. Is the number of trials fixed?
    - 1. Yes
      - a. Are there more than two categories?
        - i. Yes – Multinomial
        - ii. No
          - 1. Are there 2 or more trials
            - a. Yes – Binomial
            - b. No -- Bernoulli
    - 2. No
      - a. Trials till first success?
        - i. Yes – Geometric
        - ii. No -- Negative Binomial
- b. No
  - i. Is there a constant rate?
    - 1. Yes – Poisson
    - 2. No – Hyper-geometric

Table 1. The categorization algorithm.

Some Examples of the Proposed Algorithm

We will use two examples from Mendenhall and Sincich to test and demonstrate the algorithm. Example 4.15 page 148 concerns an experiment conducted to select the most suitable catalyst for making ethylene-diamine (EDA). Three catalysts were randomly selected from a group of ten catalysts, four having high acidity and six having low acidity. The problem asks the student to calculate two probabilities:

- The probability that no highly acidic catalyst is selected.
- The probability that one highly acidic catalyst is selected.

The distribution appropriate to the problem must be chosen from among the seven. Upon invoking the algorithm we ask "Are there independent categorical trials?" The answer is "no". This highlights the fact that as the catalysts are drawn (assuming one at a time) they are not replaced so that the second draw depends on the first and the third draw depends on the previous draws. Independence is the issue here, the decision for determining if independence is sufficiently violated is calculated by the ratio $n/N=3/10=0.3>0.05$(cut-off), therefore dependence must be assumed. We then ask the question "Is there a constant rate?" to which we answer "no" since no rate is in the problem and therefore select Hyper-geometric as the distribution appropriate for this problem.

Example 4.9 page 133 introduces students to a question relating to high neutral currents in computer systems and their potential problem. 10% of US surveyed sites had high neutral to full-load current ratios. If a random sample of five computer power systems was selected from *the large number* of sites in the country what is the probability that:

- Exactly three will have a high neutral to full-current load ratio?
- At least three?
- Fewer than three?

Here again the student must first find the appropriate distribution. "Are there independent categorical trials?", the answer is "yes", the meaning of categorical is that there is a fixed probability for each category but since the sample is taken without replacement this is not strictly true. However since a sample of size 5 is taken from a large number of computer power systems we would assume that the probabilities are practically the same for the five selected sites. This is again emphasizing the nature of the categorical trials as used in the algorithm. We then ask "Is the number of trials fixed?" the answer is "yes" since there are n=5 trials. "Are there more than two categories?" the answer is "no" there are only two categories, high or low neutral currents. Finally we ask "Are there 2 or more trials", the answer is "yes" since n=5 and we therefore select the Binomial as the correct distribution for this problem.

A CONTINGENT MODEL TO TEACH CATEGORIZATION OF DISCRETE DISTRIBUTION

The proposed Contingent Teaching for Discrete Distributions (CTDD) model uses the Stewart and Stewart (2013, see figure 3) Contingent Teaching (CT) model with clickers which was based on Schoenfeld's teaching routine (2002), created through observing good teaching practices over time by frequently taking the pulse of the class and making appropriate in-the-moment decisions. Schoenfeld's routine is well practiced, easily accessed and is relatively undemanding for the teacher, thus it was used as a foundation to build consequent teaching models. Stewart and Stewart's (2013) CT model was created as a general structure for any particular lesson, however, it is versatile and can be used as part of any contingent teaching model when further explanation is required.

Components of the CT Model

Research shows that although the instructors appreciate the feedback that they receive from the students (Abrahamson, 2006) the challenge would be whether the instructor can react to these responses on the spot, according to students' needs. This is 'contingent teaching' (Draper & Brown, 2004) and requires the interactivity of lecturers as well as students. In this setting "lecturers must develop their plans beyond the factory machine stage of executing a rigid, pre-planned sequence regardless of circumstances" (Draper & Brown, p. 91) and have more strategies on hand, based on students' responses. In their belief no class of students is ever the same.

One important aspect of contingent teaching would be that the instructor is constantly confronted with decision making. As Schoenfeld (2011, p. 36) points out "the quality of people's decision making in problem solving, teaching, and most everything we do affects how successfully people attain the goals they set for themselves". Although decision making is complex, nevertheless this comes naturally from one's Resources, Orientations and Goals (Schoenfeld, 2011). By resources he focuses mainly on knowledge, which he defines "as the information that he or she has potentially available to bring to bear in order to solve problems, achieve goals, or perform other such tasks" (Schoenfeld, 2011, p. 25). Goals are defined simply as what the individual wants to achieve. The term orientations refer to a group of terms such as "dispositions, beliefs, values, tastes, and preferences" (Schoenfeld, 2011, p. 29).

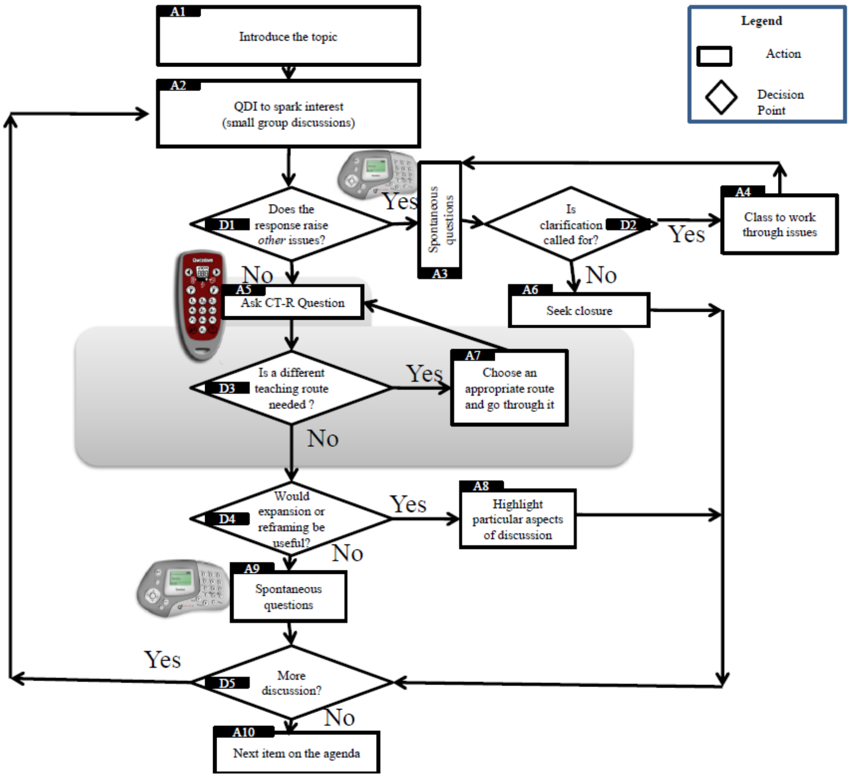

Figure 3. A contingent teaching (CT) model with clickers (Stewart & Stewart, 2013).

CONTINGENT TEACHING FOR DISCRETE DISTRIBUTIONS (CTDD)

The basic idea behind the CT model is to ask questions that relate to students understanding and contingently through clickers address their problems, misunderstandings and lack of knowledge dynamically through prepared routes of teaching. This teaching model can be greatly facilitated through powerpoint since hyperlinking is relatively easy to perform and is usually easy to integrate with clicker software (see HITT and Qwizdom for example).

In the case of teaching students how to categorize a problem in terms of the appropriate distribution, we assume that each distribution has been separately developed and taught with examples. CTDD is an application of the CT model to the classification problem, and proceeds after a probability question has been asked usually in the form of a word descriptive, the first question in the algorithm is asked and a clicker poll is taken. The problem we are considering is the second of the two discussed above and is included in the first shown in figure 4 below.

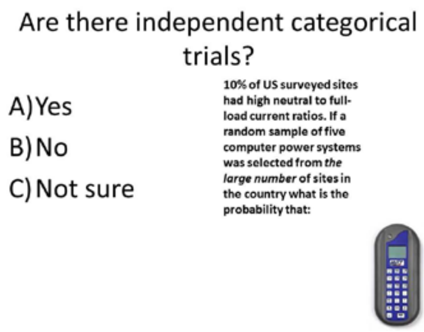

Figure 4: Slide of the first question in the algorithm.

The clicker response may reveal that a large number of students are unsure or have incorrectly analysed the problem. If that is the case then explanation of what defines independent categorical trials is in order. This can be facilitated in powerpoint by using hyperlinks to a slide sequence, see figure 5. Once this has been covered the main sequence can be returned to. The questions that follow can be in a similar fashion be linked and appropriately addressed. If a large enough percentage of the students correctly respond at the first question then the next question can be posed. There is no limit to how carefully and in depth the explanatory portions may go as the CT model demonstrates.

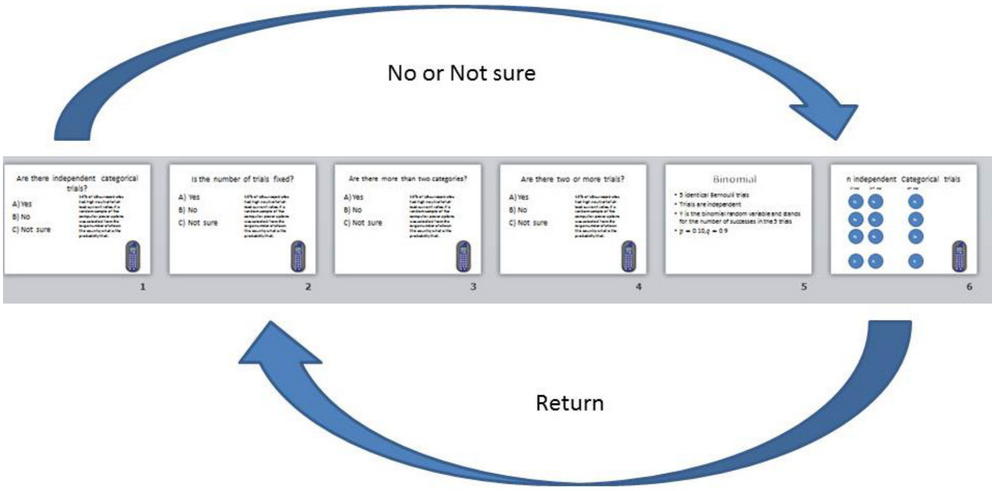

Figure 5: A portion of the slide sequence in the CTDD model.

THE WAY FORWARD

While many studies suggest that clickers are useful, more in depth research on how to effectively incorporate them in teaching is vital and there is a need to move the research towards more pedagogical aspects. In this paper we have proposed a model of teaching for recognizing various discrete distributions. We expect appropriate use of the contingent teaching model together with the teacher's preparation in advance to be of great help in making more accurate decisions and ultimately help improve the teaching of discrete distributions. The authors are planning to use this model in the first author's *Applied Statistical Methods* class in the Spring 2014, and based on the data collected refine the model.

REFERENCES

Abrahamson, L. (2006). A brief history of networked classrooms: Effects, cases, pedagogy and implications. In D. A. Banks (Ed.), *Audience response systems in higher education: Applications and Cases* (pp. 1-25). Hershey, PA: Information Science Publishing.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, *20*(2), 81- 94.

Mendenhall, W., & Sincich, T. (2007). *Statistics for engineering and the sciences* (5th ed.). New Jersey: Pearson Prentice Hall.

Schoenfeld, A. H. (2002). A highly interactive discourse structure. In J. Brophy (Ed.), *Social constructivist teaching: Its affordances and constraints* (pp. 131-170). New York: Elsevier.

Schoenfeld, A. H. (2011). *How we think*: *A theory of goal-oriented decision making and its educational applications*. New York: Routledge.

Stewart, S., & Stewart, W. (2013). Taking clickers to the next level: A contingent teaching model. *International Journal of Mathematical Education in Science and Technology*, *44*(8), 1093-1106.

Wroughton, J., & Cole, T. (2013). Distinguishing between binomial, hypergeometric and negative binomial distributions. *Journal of Statistics Education*, *21*(1), 1-15.