A VISUAL APPROACH IN THE TEACHING OF STATISTICS AND PROBABILITY

Jorge Dominguez-Dominguez[1] and J. A. Dominguez-Lopez[2]
[1]CIMAT Mathematics Research Center, Mexico
[2]Conteck, Mexico
jorge@cimat.mx

From experience, we know there are several difficulties to convey different concepts in statistics and probability as well as being assimilated by the learner. Looking to contribute in the solution of this problem, a technology development, named CalEst, has been accomplished. It aims to provide a set of tools with an educational approach to assist the teaching/learning process. This development generates information in a visually-attractive manner, improving enormously the understanding of the concepts and motivating the learning of statistics and probability. Probability notions play a crucial role in the analysis and interpretation of statistical data. Accordingly, this project incorporates several animations to illustrate and experiment several probability concepts. Furthermore, CalEst assists in an animated approach to calculate probabilities; hence illustrating the concepts of density and cumulative distribution functions for a diverse number of distributions. Other concepts are also included, such as the hypothesis test using a biometric system as example.

INTRODUCTION

Experience shows that there are several difficulties to communicate and understand different concepts in the teaching/learning process of statistics and probability. To help solve this problem, a technology development, which is named CalEst, has been accomplished. This development focuses in two components: educational and operational. In the former, it has been developed a series of teaching ideas and proposals in order to facilitate the presentation and understanding of concepts in the two areas. The latter aims to assess the understanding level of statistical concepts. This operational component consists in performing straightforwardly the description and analysis of data generated in a given study. Several of the operating results can be viewed using the educational materials.

The application of these two components will allow the teacher to work with the topics proposed in the curriculum of statistics and probability for high school level and above. On the other hand, the usefulness of this material will give the student the opportunity to study the examples and find solutions to the exercises posed in several statistical texts. It has also been developed complementary materials within the educational module. In addition, the methodology of the information processing comes from both real problems and data generated by the user.

Students need a good knowledge and skill in several statistical and probability concepts such as probability calculus, probability distributions, hypothesis tests and statistical modeling. Bearing in mind this situation, it has been developed a new and attractive educational material with computer animation effects and simulations. This material is intended to contribute in improving the teaching-learning in higher education. The development of this visual approach is based on the idea of learning by playing and do-it-yourself. Notions of probability play a crucial role in the analysis and interpretation of statistical data. Accordingly, this project incorporates a series of animations that considers the classic games of probability such as coin toss, dice, coin-die, roulette, draw of marbles and mazes. These activities are designed to present a clear perception of the probability density and the cumulative distribution functions. These animations motivate students to increase their interest in this area and seek new ideas by interacting with the material.

The integration of this visually-aid technology development with the strategies to generate data has created a dynamic which provides a greater ease in the understanding of concepts and motivate learning these fields. With the purpose of illustrating this situation, in this paper, three modules are presented in detail. They have been developed in the CalEst project to motivate and illustrate various concepts of statistics and probability.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]

EDUCATIONAL MATERIAL

The discipline of statistics is the process of discovering more about the real world by gathering, analyzing and interpreting data. Subsequently, statistical studies are designed as a search procedure. Firstly, a problem is posed which lead to a series of questions. Then, these questions are answered and explained with an appropriate collection and analysis of data. It is a common practice to provide data for students to do calculations but frequently no interpretations of the results are made. Moreover, the study of statistics is based on concepts of probability theory. And there is a certain difficulty in teaching and learning notions about these topics.

Sample Educational Material 1: Calculation of Probabilities

Within the educational component, there are ten different activities for calculating the probability of events as well as for applying the rules for calculating the probability of compound events. In this section, the roulette activity is presented. It has two roulettes and the user can select in how many groups each roulette is divided. Figure 1 shows the case of both roulettes with four groups. This creates different scenarios using the colors of the roulette or the numbers and letters. For these scenarios, different sample spaces are generated to calculate probabilities. For instance:

1. What is the probability of both roulettes having the same color? From the sample space, we have that the same color case happens 4 times of 16 cases, then $P(E : same\ color) = 4/16 = 0.25$
2. What is the probability that at least one roulette is blue? Again, using the sample space, we notice there are 7 cases in which at least one roulette is blue, hence $P(E : at\ least\ one\ is\ blue) = 7/16 = 0.4375$
3. Considering the number and letters of the first and second roulettes, respectively. We define the events *E={1,2,3,4} and F={A,B,C,D}.* What is the probability of getting an even number and a consonant? For the first roulette, Red (R) and Yellow (Y) are considered as even numbers while for the second roulette, Red (R), Green (G) and Yellow (Y) are defined as the consonants. Hence, we have *H={(R,G), (Y,Y), (R,Y), (Y,R), (R,R), (Y,G)}* equivalently the event *H* is described by *H={(2,C), (4,D), (2,D), (4,B), (2,B), (4,C)}*. Thus, the probability is 0.375.

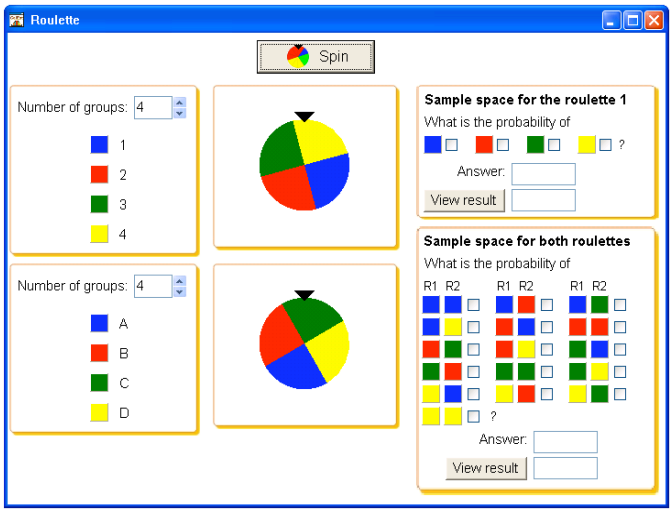

Figure 1. Two roulettes divided into four groups

Sample Educational Material 2: Probability Distribution

An example will illustrate the simplicity to calculate probabilities of a density function using the developed materials. This example considers a study of the brain. It is desired to know the response time of a person to a visual stimulus. It is assumed that the variable has a normal distribution with mean $\mu = 130$ and standard deviation $\sigma = 38.6$. Furthermore, we want to know which percentage of people takes less than 30 seconds to respond or estimate the percentage of persons that respond between 70 and 180 seconds. This situation is considered in terms of

probabilities; $P\ (30 < x)$ for the former case and $P\ (70 \leq x \leq 80)$ for the latter. This raising generates a number of questions such as understanding the concepts of probability, random variable, density function and probability distribution.

When probability distributions are addressed in basic courses of statistics and probability, it is assumed the variable follows a normal distribution. Commonly, the first thing that it is explained to the student before making any calculation is the process of standardization. This consumes time and distracts the students' attention from the main objective. It is suggested that students should focus on understanding what a distribution is, the differences among the several distributions and their applications. This tool has a high potential to be used in a diversity of applications. Nevertheless, we will limit our discussion only to the visual approach used to present the concepts. Accordingly, for the problem described above, we obtain its visual result for the calculation of the probabilities. Figure 2 illustrates the calculation of the probability described by the density and cumulative plots of the normal distribution.

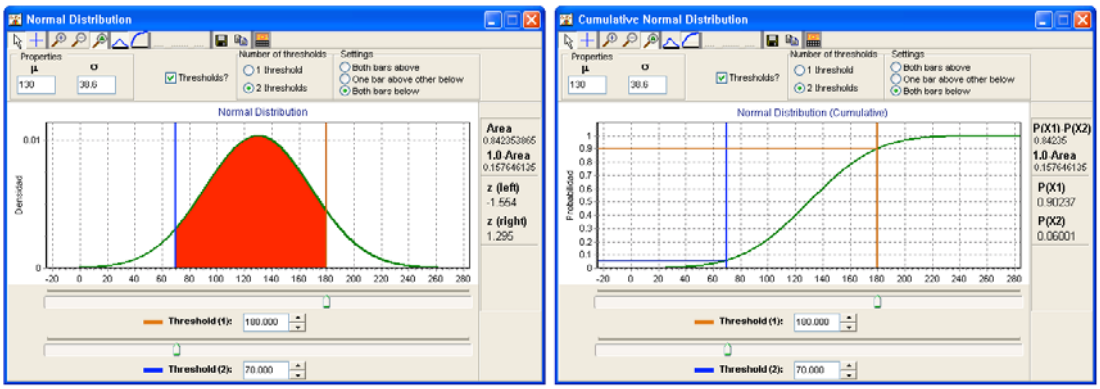

Figure 2. Visual representation of the calculation of probabilities.

It is possible to calculate different probabilities just by moving the track bars (i.e., thresholds), changing the parameters values or by moving the mouse over. Notice that for the normal distribution, the probabilities are obtained without the need of standardization. We only need to declare the distribution parameters (i.e., mean and standard deviation). We can compare against the standardized normal distribution by using $\mu = 0$ *and* $\sigma = 1$. In addition to the graphic approach, there is a calculator which the user can use to obtain any probability (to the left or right, between two points) using distribution parameters and the value(s) of the random variable. Also, it is possible to use it for the inverse case. Given the probability, we can obtain the random variable.

The technological development proposed by CalEst also uses animation to directly calculating distribution probabilities, illustrating both the density function and probability distribution concepts. This approach is applied to a wide range of distributions.

Sample Educational Material 3: Hypothesis Test

Following the example described above, we want to verify the hypothesis that the persons respond fast. In statistical terms, the respond is fast when the person reaction takes less than 130 seconds. Hence, we have the formal statistical approach:

$$H_0 : \mu = 130$$

$$H_1 : \mu < 130$$

The verification of a hypothesis test is based on concepts of probability theory. Thus, it is necessary to understand the meaning of the probability of rejecting the hypothesis H_0, when it was true and the probability of fail to reject it when it was false. These concepts are visually illustrated using a plot with the ability to simulate different scenarios using the CalEst. Figure 3 shows the case when the alternative hypothesis is H_1: $\mu = 120$. We can explore different scenarios and evaluate the probabilities of Type I and Type II errors. This tool has been used successfully in illustrating the concepts of hypothesis test for several problems. The calculator of distributions is

also included to assist in the different hypothesis test calculations. Moreover, the concept of hypothesis test is also illustrated using an animation/simulation of a practical system, i.e., a biometric security system. The system has to reject impostors and accept authorized users.

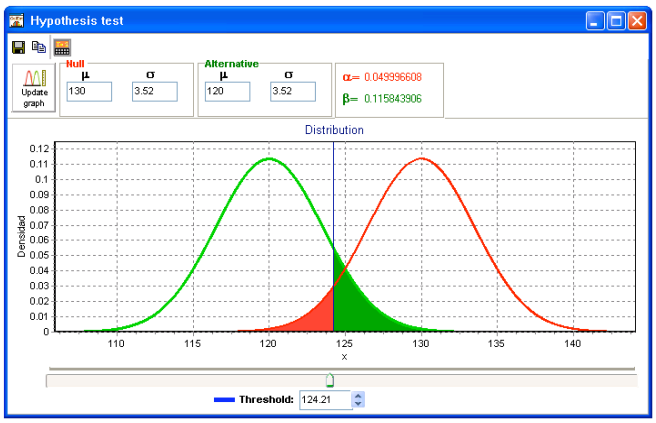

Figure 3. Animation to carry out a hypothesis test

OPERATIONAL MATERIAL

Finally, once the data have been obtained, the CalEst has a diverse variety of tools for calculation, analysis, data manipulation and graphics. These functions are illustrated with a tutorial. The modules included in this component cover the curriculum in high school and undergraduate courses. Some of the modules are: measures of central tendency; dispersion and position; histogram; frequency polygon; empirical distribution; pie chart; dot plot; stem-and-leaf plot; box plot; scatter plot; probability paper; hypothesis testing for 1 or 2 means, proportions and for the variance; nonparametric confidence intervals; analysis of variance; regression; goodness-of-fit test; etc. Other functionality is its capability to exchange information with other software tools. For instance, data can be imported and/or exported to/from other packages, a graphic (i.e., image) can be place in the clipboard or saved, and the results printed or exported to a file.

DISCUSSION

The described technology development is presented as a package which uses computer-aid animations and simulations. This development is a new, visually-appealing tool that benefits in the learning, teaching and application of concepts of statistics and probability. On one hand, it gives elements to the teachers to explain different topics in a more enjoyable manner and straightforward to understand. It also provides to the teacher the option of going deeper into the subjects. Its visual environment and animations not only allows but also encourages the student to explore and learn by doing-it-yourself using the auxiliary practice material of the package. The material helps students to understand clearly the concepts, motivate themselves to learn more and explore for themselves. CalEst serves as supporting material for a better understanding of some concepts as well as a tool to solve problems from different statistics books. The animations and simulations can also be used to introduce elementary and secondary students into data processing.

A key part of this development is given in the computer knowledge and skills applied to explain and develop concepts and results in the teaching of statistics and probability. This project has been developed with advanced programming techniques which integrate a collection of resources (e.g., animations, graphics, and simulations) to aid in the teaching/learning of statistics and probability concepts. CalEst runs on Windows XP, Vista and 7 operating systems.

REFERENCES

Chase W., & Bown F. (2000). *General Statistics*. New York: John Wiley & Sons, Inc.

Domínguez-Domínguez J., & Domínguez-López J. A. (2006). *Estadística y Probabilidad: El mundo de los datos y el azar*. Mexico DF, Mexico: Oxford University Press.

Johnson R. A., & Tsu K.,W. (1998). *Statistical Reasoning and Methods*. New York: John Wiley & Sons, Inc.

Martz, P. (2006) *OpenGL Distilled*. New York: Addison-Wesley.
Pelosi M. K., & Sandifer T. M. (2003). *Elementaty Statistics*. New York : John Wiley & Sons, Inc.
Press, W., Teukolsky, S., Vetterling, W., & Flannery B. (2002) *Numerical Recipes in C++* (2nd edition). New York, NY: Cambridge University Press.
Triola, M. (2008). *Essentials of Statistics* (3rd edition). Boston, MA: Pearson Education, Inc.