Teaching statistics by using simulations on the Internet

Andrej Blejec
National Institute of Biology, Vecna pot 111
Ljubljana, Slovenia, andrej.blejec@uni-lj.si

Simple probabilistic simulations, such as flipping a coin or rolling a dice, are used for illustration of statistical concepts and stochastic properties of modeled processes. With availability of personal computers in past decade, computer supported statistical simulations were introduced to statistics teaching. Such simulations - usually combined with graphic presentation of results and simulation itself - can be used for more obvious demonstration of statistical phenomena. With the advent of the Internet, new dimensions were introduced to teaching. The simulation software is easily available for those that are connected to the Internet, regardless of physical distance. Internet also brought cross-platform compatibility of teaching material, which was sometimes limited to one or another type of operating system. But there are some cultural and language barriers, usually combined with low availability of computers and Internet in some countries.

1. Introduction

There are certain fundamental statistical concepts that are difficult for students to comprehend at an intuitive level. Teachers of statistics are continuously exploring new ideas and teaching practices to overcome such difficulties. One useful and widely used approach is demonstration of statistical concepts with simulations and simulated data. With the advent of personal computers computer simulations, combined with dynamical and interactive graphics became available and several authors prepared excellent demonstrations of statistical phenomena. At least two technological barriers prevented wide use of such systems: the incompatibilities of different computer platforms and the difficulty of software distribution. In the past decade, Internet evolved into main communication system for academic world. There are a number of World Wide Web (WWW) locations that provide various forms of information for statistics education in electronic form. Using the latest WWW technology, educators can now easily access and use interactive demonstrations in their teaching, regardless of particular computer platform. The importance of access to high-quality education and training materials whenever and wherever they are required are leading to standardization of specifications and guidelines such as the Sharable Content Object Reference Model (SCORM, http://www.adlnet.org). Such standardizations are necessary for compatibility of future e-learning environments and need cooperation of academia, governments, and industry (for example IMS Global Learning Consortium Inc., http://www.imsglobal.org).

In this paper is to point out the potential of computer simulations for teaching of statistical concepts and to overview some interesting teaching materials with simulations available on the Internet. Finally, some indications about the efforts for standardization of E-learning material is given.

2. Teaching statistics with computer simulations and simulated data

Teaching and learning of statistics has gained recognition in many disciplines over the past two decades. Statistics is integral part of the post secondary curriculum. In almost every discipline, the ability to understand, interpret and critically evaluate research findings is becoming essential skill (Gal, 2002; Giesbrecht, 1996). Microcomputers have an important role in modern statistics teaching (Garfield and Burril, 1996). They allow students to accomplish computational tasks more quickly, freeing them to focus more on statistics concepts. The computer is not only an efficient computational tool but can also help to clarify specific concepts

by providing settings in which one can apply statistics techniques (Ben-Zvi, 2000; Biehler, 1997). On the other hand, even if students use software packages to apply techniques and master the mechanics of data analysis, abstract statistical concepts may still be difficult to comprehend (Marasinghe et al., 1996).

Using modern computer technology, standard data analysis assignments can be supplemented with additional statistical experience through the use of computer simulation methods and systems. Such simulations can help teach statistics in general and particularly difficult or abstract concepts and theorems. Using simulations, combined with computer visualizations, such concepts and theorems can be effectively demonstrated even to students with limited mathematical abilities or interest, and can sometimes serve as a substitute for rigorous mathematical proof. Computer simulation systems allow students to experiment with random samples from a population with known statistical properties (with known parameters and quantiles) which enable them to get insight into specific abstract concepts (for example sampling distribution and central limit theorem) or to compare computed results with the expected (theoretical) ones (Blejec, 2002). Computer simulations can be used for graphical demonstration of concepts, either in large lectures or by students with some guidance from the instructor.

The concept of computer simulation in statistics has a vague meaning (Mills, 2002), but in essence it involves the use of random number generators to generate samples from desirable distribution of virtual population. Different programming environments are used for stand-alone simulation activities: from traditional statistical packages (such as MINITAB and SAS) to powerful computing environments (for example Lisp-Stat, S-Plus, R) to popular and widespread spreadsheets (mostly Excel). In the literature review Mills (2002) lists popular simulation topics: central limit theorem, sampling distribution, confidence intervals, different distributions, regression analysis, hypothesis testing, and survey sampling.

The central limit theorem is usually demonstrated by repeated selection of samples with selected size from different distributions (for example uniform, bimodal, exponential, normal) and plotting the distribution of means. Similar technique is used for demonstration of sampling distributions of other statistics such as variance, median, and proportion. Sampling distribution demonstration is sometimes combined with display of confidence intervals (Blejec, 2002), which are usually shown as a result of repeated sampling. Each sample gives a confidence interval that is plotted along the line representing true value of estimated parameter. Meaning of the confidence level as the percentage of intervals containing true value can be explained in obvious way. Some other concepts, such as bias of variance estimator and asymmetric confidence interval for variance are easily shown. For demonstration of linear regression properties, data based on linear model $Y = \beta_0 + \beta_1 X + \varepsilon$ are simulated. For each sample, estimated regression line can be compared with the model line and the relation between total, explained, and residual variation can be shown. Hypothesis testing simulations show the connection of sample size and significance level with power of the test and Type II error. Survey sampling, as described by Mills (2002), enables students to simulate the composition of the population and try to get the correct sampling scheme in order to get the best estimates for simulated structure. Maybe some simulations dealing with sampling in natural and environmental sciences would be appreciated.

3. Statistical simulations on the Internet

Since most educators have access to WWW, statistics simulations and lot of other statistics teaching and data analysis resources have provided an exciting new medium for teaching and learning (Mills, 2002). Internet proved to be an extremely efficient information delivery method, bridging physical gaps and barriers between users around the world. Many statistical simulation systems are easily accessible and can be downloaded for educational use. The list of statistical resources at Statistical Consulting Service of Yorku University[1] is just one of available and rich

[1] http://www.math.yorku.ca/SCS/WebResources.html

collections of WWW statistical links. Another influential WWW novelty is Java, which seems to be a powerful programming language, suitable for preparation of visual and interactive simulations on the WWW. Many mentioned simulations were rewritten into Java and are even more easily used without the need to download some special programming environment or software. Many simulations are available as separate Java applets[2]. They are included in interactive web textbooks, as found in, for example, "Rice Virtual Lab in Statistics"[3] with HyperStat[4] by D.M. Lane, "Concepts and Applications of Inferential Statistics"[5] by R. Lowry, and Computer Assisted Statistics Teaching[6] (Stirling, 2002). Many excellent simulations can be found in demonstration systems, such as WebStat[7] (West and Ogden, 1998) or Statistical Java[8] by S. Dorai-Raj, C. Anderson-Cook and T. Robinson.

4. E-learning

In the past decade Internet evolved into the main communication media. Along with all other e-something, idea of e-learning also evolved and is developing very fast. Many universities and institutions around the world are offering distant learning courses and trainings, and field of statistics is not an exception. The wealth of statistical resources on the web is enormous and is growing very fast. At the same time, many off-line systems for electronic textbooks evolved into the web supported course management systems, for example WebCT[9] which is at the moment providing over 30 statistics e-packs, mostly for business and elementary statistics. Such systems take communications advantages of the Internet in providing teaching material according with the paradigm of anytime and anywhere lifelong learning. Students have possibility to learn at their own pace and at the times that suit them best. Internet based e-learning systems enable tracking of students' activities, assessment and grading, student discussion forums and personal communications via e-mail. Widespread use of web based learning systems led to efforts to set up standards and recommendations for development of learningware in order to establish interoperability of systems and reusability of available information. Such efforts of many organizations and initiatives (like Advanced Distributed Learning Initiative (ADL) in USA and ARIADNE in EU) are coordinated by organizations such as IMS Global Learning Consortium.

While parts of the world are keeping up with all described, there are parts of the world that are trying to catch up with the changes in the information society. Even in European Union, not to speak about some transition countries in Eastern Europe, the availability of Internet at home, the speed of connection and computer usage proficiency is much lover than in US (SIBIS, 2003). The situation in some other parts of the world, in most of Asia, Africa and central and south Americas the situation is even worse.

5. Conclusions

The use of computer simulations in statistics teaching proved to be a useful mean for explanation of difficult statistical concepts. Internet is, on one hand, an efficient information delivery method for rapid exchange of information including simulation software. On the other hand, Internet provided recommendations and de facto standards for information interchange. With Java enabled browser teachers and students can use excellent pieces of software for demonstration of statistics, regardless of computer platform that was used by developer or end user. The setting of recommendations and standards for development of web based cooperative learning systems is under way and will enable wide spread availability and use of distant

[2] For example Statlets (http://www.statlets.com) or http://www.stat.duke.edu/sites/java.html
[3] http://www.ruf.rice.edu/~lane/rvls.html
[4] http://davidmlane.com/hyperstat/index.html
[5] http://faculty.vassar.edu/lowry/VassarStats.html and http://faculty.vassar.edu/lowry/webtext.html
[6] http://cast.massey.ac.nz
[7] http://www.stat.sc.edu/~west/webstat/
[8] http://www.stat.vt.edu/~sundar/java/applets/
[9] http://www.webct.com

learning possibilities. On the other hand, guidelines for the fair use of electronic materials in a variety of nonprofit educational contexts are already in use[10].

The paradigm of availability of learning materials "whenever and wherever" needed has limited applicability: there are many countries in the world, where information infrastructure is not tuned to wide use of the Internet, even if it is present in most of the universities around the world. Majority of teaching materials on the web is written in English and cannot be directly incorporated into teaching process.

But the universal language of graphics (and some mathematical symbols), which is mostly used for the presentation of statistical simulation results, makes simulations, either run or just available on the web, very useful and attractive teaching tool.

Acknowledgment

The author is grateful to an anonymous referee for valuable comments.

REFERENCES

Ben-Zvi, C. (2000). Towards understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, 2(1&2), 127-155.

Blejec, A. (2002), Teaching statistical concepts with simulated data, in: Phillips, Brian (ed.). ICOTS 6 : 7-12 July 2002 Cape Town, South Africa. Cape Town: IASE: SASA.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167-190.

Gal, I. (2002). Adult statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.

Galmacci, G. (2001). The impact of the Internet on researchers' training. In C. Batanero (Ed.). Training Researchers in the Use of Statistics (pp. 159-169). Granada: ISI & IASE

Giesbrecht, N. (1996), "Strategies for Developing and Delivering Effective Introductory-Level Statistics and Methodology Courses," ERIC Document Reproduction Service, No. 393-668, Alberta, BC.

IASE Round Table Conference (1996): Garfield y G. Burrill (Eds.), Research on the role of technology in teaching and learning statistics. Voorburg, The Netherlands: International Statistical Institute.

Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342-351.

Mills, J. D. (2002), Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education,* 10(1).

SIBIS Pocket Book 2002/03 (2003), Measuring the Information Society in the EU, the EU Accession Countries, Switzerland and the US, IST-2000-26276. Statistical Indicators Benchmarking the Information Society (SIBIS) project and European Communities http://www.sibis-eu.org.

Stirling, D. (2002), Interactive Content in Web Pages Teaching Statistics, , in: Phillips, Brian (ed.). *ICOTS 6 : 7-12 July 2002 Cape Town, South Africa*. Cape Town: IASE: SASA.

West, R. W., and Ogden, R. T. (1998), Interactive demonstrations for Statistics Education on the World Wide Web, *Journal of Statistics Education,* 6(3).

[10] http://www.utsystem.edu/ogc/intellectualproperty/confu2.htm