

## SIMULATION AND STATISTICAL EDUCATION

Jorge Luis Romeu

Department of Mathematics  
SUNY-Cortland  
Cortland, New York 13045, U.S.A.

### ABSTRACT

Discrete event simulation has nurtured from statistical analysis for many years. The converse is not true. However, recent advances in computer technology and software development has made it possible to have PC's running specialized simulation languages, readily available. This paper discusses how discrete event simulation, developed via specialized simulation languages (e.g., GPSS) can become a useful class resource to teach and motivate statistics students. In addition, simulation helps to present, more effectively, interdisciplinary case studies, to increase group learning and to relieve students and instructors from statistical drudgery. Examples of such GPSS simulation are developed.

### 1 INTRODUCTION

For some time now, many statisticians have been trying to change the way statistics is taught in college. The 1992 ASA Winter Conference, in Louisville, KY, was dedicated to Teaching. The Statistical Education Section and some hard copy and electronic journals are dedicated to improving the teaching of statistics (e.g., the February 1995 issue of *The American Statistician*). And most agree that it is time for a change in the way statistics should be taught, especially to non majors, applied statistics students and practicing professionals.

For example, Hogg (1985, 1991) describes the workshops in statistical education and discusses the need for clearly specifying the **objectives** that statistics service courses should pursue, according to the audience and the level of the students to which it is intended for.

Bisgaard (1991) describes his engineering statistics course, where he uses practical experimentation in the classroom to motivate the students and to introduce statistical concepts. Bisgaard states that such live

experiments constitute the core of his entire course, since he uses them to introduce basic concepts in experimental design, data collection problems, parameter estimation and statistical comparisons of the results.

With respect to teaching applied statisticians, Kettenring argues that industry needs problem solvers who are able team players and can communicate effectively their findings (Kettenring 1995). He states that in industry, problems are interdisciplinary (as in most real life environments). Hence, that bridges should be built between disciplines in college. Bickel also argues in favor of an interdisciplinary approach to teaching statistics (Bickel 1995) and lists some topics and types of talents that statisticians should have.

In the same vein, Ross expresses the requirements to succeed as a government statistician (Ross 1995). Again he demands for team players due to the interdisciplinary problems encountered. The entrepreneurial element, that is the drive to search within one's own organization, for potential statistical problems that would not otherwise be proposed to the government statistician by the (non statistician) client, is another positive characteristic. Bailar states that Academe fails to involve students because we teach heavily abstract Statistics topics, unrelated to the student taking a service course (Bailar 1995). He argues that there must be cuts in the present curriculum; that priorities must be established and that specific topics must be selected. Also that the range of Statistics electives should be widened even at the expense of currently required material, perhaps far too theoretical.

Finally, Lehoczky discusses ways of modernizing the doctoral curriculum (Lehoczky 1995). He describes the Carnegie Mellon Ph.D. program (master's as well as upper level courses) and argues that it is evenly distributed between theory and the practice of Statistics. In addition, the program does train the doctoral student in teaching as well as in written and

oral communication and in cross disciplinary activity.

All of the above shows that there is a serious and longstanding concern among the statistical community about the teaching of our subject. And that a serious quest for new methods of teaching it in a more practical and interdisciplinary manner is justified.

This paper presents one such method: teaching Statistics through simulation modelling and output analysis. Through this approach, we believe applied and service students can be better reached and motivated. And that the student's perception that Statistics is a useless, dull course requirement can be changed to that of a useful and interesting topic, worth dedicating time to. We must remember that we are competing for the students' attention and time among several other subjects and extra curricular activities.

In addition, simulation allows the application of several new educational approaches such as *workshops and cooperative learning*. Traditional lecturing lends poorly to developing these new techniques. Finally, via simulation we can *teach more statistical methods to undergraduate science and engineering students*. Many undergraduate curriculums today teach at most two statistics courses. And most undergraduates never pursue a masters degree, even when they deal with data on a daily basis, in research, development and production environments.

The rest of this paper expands on these problems. Section 2 discusses several uses of simulation in the teaching of Statistics and the advantages of using simulation packages. Section 3 deals with several types of teaching approaches using simulation and discusses their advantages and disadvantages. Section 4 briefly presents Monte Carlo methods as a teaching tool. Finally, in Section 5, we summarize our work.

## 2 SIMULATION IN TEACHING

There are several ways of teaching Statistics: from the traditional lecturing approach to that of developing in-class experiments, as in Bisgaard (1991). The first, essentially uses book examples to illustrate the theory, often remote from the students' main interests and lacking in data collection and manipulation. The second, experimental approach, is complicated in nature and even risky (e.g., the ladder experiment may be prone to accidents). Or at least, time consuming (inside and out of the classroom) so that only a very limited number of real life experiments can be carried out during the semester.

Simulation modelling and analysis, on the other hand, is an intermediate solution, half-way between the above two. Simulation still retains the flavor of

the uncertainty in the experimental outcomes, as occurs in real-life. This uncertainty is produced by the simulation model (for it represents a real situation). On the other hand simulation does not require that the student spends hours in or outside of classroom, gathering data. For the student can do this automatically, during the simulation run. Students are no longer subjected to the inherent physical risks of an experimentation process. Nor they incur in the time and money expenses that some experiments necessarily bring about.

Throughout this paper, we will refer to simulation as the art of modelling in a computer, via a specialized language (e.g., GPSS), a real system whose operation through time we study by running the computer model. This definition contrasts with the concept of (static, Monte Carlo) simulation sometimes used when teaching elementary service courses. There, a set of data following a pre-specified distribution (say the normal), is generated via some statistical package (e.g., Minitab). Then one performs some simple statistical procedures on them (say, a test of hypothesis).

Such elementary uses of static (Monte Carlo) simulation are adequate for introductory courses but will not be discussed further in this paper. We, in turn, will concentrate in the pedagogical uses of discrete event simulation models in intermediate and advanced service and/or applied statistics courses.

In the past, simulation was seldom used in teaching Statistics. At least two good reasons can be argued for this. One is that simulation models were written in an (HOL) programming language (e.g., Fortran). These were time consuming and difficult to implement. The second reason is that specialized simulation languages (e.g., GPSS) were available only in main frame computers making them (i) quite expensive and (ii) constrained to a small group of specialized users.

With the advent of the PC and the subsequent transfer to it, of many of these specialized simulation languages, the above two disadvantages have all but disappeared. First, it is now possible to write a complicated and challenging simulation model with relative ease. And they are still easier to run by the students. Then, there are several new simulation books that include an inexpensive student version, for example, Karian and Dudewicz (1991), Thesen and Travis (1992), Schreiber (1991). Such versions are, for all practical classroom purposes, quite adequate.

The approach presented in this paper was developed from several years of teaching applied Statistics, then simulation and finally Statistics via simulation. And more recently, from our experiences in

teaching short courses in the uses of simulation in the intermediate and advanced statistics courses, to Statistics Faculty of several Mexican and Spanish universities. And from helping them implement these techniques in their classrooms. In general, Statistics Faculty have been very positive to it, and so have their students.

### 3 SIMULATION APPROACHES

We propose here three different approaches to teaching Statistics and/or stochastic OR courses via simulation with a specialized (e.g., GPSS) language. And these depend on the level of expertise the instructor wants to have with the language. They are: (i) simulation as an independent, statistically oriented course; (ii) simulation as a companion course lab and (iii) simulation as integral part of the statistics course.

As an independent course (as it is most often taught today) simulation is an excellent complementary course for statistics and OR students. In them, students have an opportunity to apply, in an interdisciplinary environment (models may differ widely) multiple data analyses techniques.

Majors in Industrial engineering, OR, systems analysis, computer science and business form the bulk of its student body. However, we should encourage (even require) applied or service course statistics students with at least two semesters (including regression and ANOVA) to take simulation as an applications course.

Requirements to teach simulation include, in addition to statistics, knowledge of a specialized simulation package (e.g., GPSS) and of the simulation techniques themselves. For, simulation is an art that includes programming, modelling and statistical analysis. Hence, its great potential for students in statistical data analysis.

As a Lab course, companion to the statistics course, simulation is excellent and the requirements for the instructor are much smaller. One does need a minimum level of understanding of the simulation activity and language. But (introductory level of) languages such as GPSS are easy to learn. And the examples in the introductory books are easily understood and can be quickly modified to provide useful applications of statistical methods.

The Lab would follow the theory class, providing an inexpensive alternative to physical experimentation. One would explain the physical meaning, needs and objectives of the simulation model in question. Then, the student would be provided with the simulation program and an instruction sheet on how to run it. And of course, an individual seed. This is the key

issue.

With individual seeds we guarantee individual and even possibly conflicting results. And such capability provides many advantages. First, students can now work in *teams* and cooperate. Different seeds generate different random samples. With this, students can engage in **cooperative learning** while remaining **individually accountable** for their results.

In addition, if we have a class with enough students (say 20 or more) and we are, say, testing at level  $\alpha = 0.05$ , it is likely that one or more of the students will obtain, by chance, contradictory results. We can make excellent use of such contradictory results to elaborate on the practical implications of testing and of the size of the test and of the sample.

We can also control the model variables and the model itself. In real life, one seldom knows the exact model, nor all the model assumptions hold. One usually is not even sure which is the best approximation to the *real* model. In simulation, since we are *building* the model, it is totally under our control.

This capability also allows us to *violate* model assumptions and have the students assess its impact in the experimental results. This way we show (i) practical importance of checking model assumptions and also (ii) which assumptions are more important and which can be relaxed and to what extent.

Another advantage of the realistic character of the examples presented, via simulation, is the training into potentially inter-disciplinary problems and team work. Also, a final report and presentation can be required from each student or team of students (if using cooperative learning techniques). In this way, the communication skills, oral and written, so important for our students, are developed.

Labs can take place once a week, for two or three hours. Students understand the simulation model and learn how to operate it. Objectives of the statistical analyses are explained and the task is rehearsed. The students, then and on their own, design and develop the experiments and obtain the results.

The third possibility is that of using simulation as an intrinsic part of the Statistics course. This approach has yet smaller requirements on the instructor. One only needs to operate the simulation programs and understand how to change the control values of the variables under analysis.

There can be one instructor who can act as a *focal* point. This instructor develops and passes on to the other instructors the required simulation programs for their use. The focal point does need an intermediate level of programming in such simulation languages. However, the other instructors only need to have an elementary programming knowledge.

Under this setting, the course instructor would use, as examples, the simulations in class. And out of class, as homework and/or project materials. Course work with a mid complexity simulation model of a network of small dams is shown next.

#### 4 CLASS EXAMPLE

A system of small, interconnected dams is modelled as follows: Each dam has, as design parameters, maximum and minimum capacity levels, a daily (deterministic or stochastic) water delivery quota and three sources of water input. These inputs can be: from rain, from the water table and from another dam. Associated with each source of water there is a cost. In addition, there is a stochastic rain pattern. And the system has to deliver the required quantities of water to the agricultural region on a daily basis.

This optimization model has several discrete (dams, maximum and minimum capacity, water table) and continuous (costs, rain distribution) variables. It also has several response variables (Total/partial operation costs, number of water transfers from dam to dam, water stock-outs).

If different seeds are assigned to the students, one can easily form cooperative learning groups (teams) and still have students working independently. Also, different sets of independent variables and different responses may be assigned to different members of a team. Then, different seeds need only to be given to the different teams. This insures *individual accountability*.

Many statistical topics can be reviewed with this simulation model. For example, one can use different number of dams, with different (max/min) capacities (replenishing levels). One can also have different costs associated with inter-dam water transfers. And one can have different rain distributions (dry, rainy season). This provides multiple factors for implementing a full or fractional factorial design. One can also design as many replications (or batches) as one needs and implement (i) ANOVA or ANCOVA, (ii) multiple regression, (iii) goodness-of-fit, (iv) residual analysis, (v) methods of variable selection and (vi) response surface methodology, just to mention a few.

In addition, if one makes the simulation batches or replications short enough, one will violate the assumptions of normality and independence of the observations. One can then discuss the importance of residual analysis and of data transformations. Discrete variables can be used in ANOVAs and continuous ones in regression. Combinations of discrete and continuous variables allow the implementation of ANCOVAs. There is much flexibility and realistic flavor

in these different combinations, allowing the use of many interesting models and statistical procedures.

Finally, since we are able to computer generate and store all these statistical results, regardless of size, we can have students easily use statistical packages in our course. Under such setting, students' numerical results and data will be (i) meaningful for them (not retrieved from a package library) and (ii) they won't have to enter them by hand, as one would have to with those obtained from a physical experiment.

#### 5 MONTE CARLO SIMULATIONS

Monte Carlo is also a very useful tool for teaching Statistics. Hence, we would not like to conclude without dedicating a few paragraphs to its many potential educational uses.

Assume one gives the students of an intermediate or advanced Statistics course a final project to compare the power of two or more hypotheses tests. For example, students are requested to compare several multivariate normality goodness-of-fit tests (Romeu, 1994a). And one gives them the assignment to do so under a subset of non normal alternatives, say under kurtic or skewed alternatives.

Students can then review several important problems. First, the impact of the dimensionality of the problem. Then, the importance of defining the specific type of alternative one is covering against. Also, the effect of the sample sizes and the inter-correlation between the variable components.

Then, students can get into different methods of random variate generation and compare them (by testing them for fit). This also helps uncover the nature and relevance of the concepts of skewness and kurtosis. And of how to measure and interpret these concepts in a data set or in the shape of a density function. Many Statistics students have a good handle of the theory but are unable to use theoretical results in practical situations. But in real-life data analysis, one needs to see through the data during the EDA phase of a study.

Finally, Monte Carlo simulation gives students another opportunity to use statistical packages and another excuse to browse through statistical papers, becoming more familiar with statistical journals.

#### 6 CONCLUSIONS

During many years, simulation modelling has nurtured from Statistics. Simulation output analysis is another good example of applied Statistics. On the other hand, statisticians have seldom used simulation in their teaching. For, writing simulations in a HOL

programming language was complicated and because simulation packages were difficult to get and operate.

With the advent of the PC, these two problems have been solved. Today one has ready access to good simulation packages such as GPSS, easy to learn and to use by students and faculty.

Simulation, in turn, provides an intermediate approach to pure lecturing or to physical experimentation. Computer or simulation modelling is easier and faster to implement and to learn and more efficient to control in the classroom. It provides more flexibility for the Statistics instructor, by allowing a small simulation model to serve as a teaching tool for several statistical methods, with the adequate change of a few parameters and instructions.

This author has experimented with this approach, with excellent results, in several institutions in the U.S. and abroad (Romeu, 1986 and 1994b). This author has also conducted short training workshops in teaching statistics with simulation in GPSS, during a recent Fulbright Lecturing Award, in several universities in Mexico and also in Spain. Student evaluations and responses to such approach have consistently been very positive.

## REFERENCES

- Bailar, John C. 1995. A Larger Perspective. *The American Statistician* 49:10-11.
- Bickel, Peter J. 1995. What Academia Needs. *The American Statistician* 49:5-6.
- Bisgaard, S. 1991. Teaching Statistics to Engineers. *The American Statistician* 49:274-283.
- Karian, V. and E. Dudewicz. 1991. *Modern Statistical, System, and GPSS Simulation: The First Course*. New York: Computer Science Press.
- Hogg, R. et al. 1985. Statistical Education for Engineers: An Initial Task Force Report. *The American Statistician* 39:168-175.
- Hogg, R. 1991. Statistical Education: Improvements are Badly Needed. *The American Statistician* 45:342-343.
- Kettenring, Jon. 1995. What Industry Needs. *The American Statistician* 49:2-4.
- Lehoczky, John. 1995. Modernizing Statistics Ph.D. Programs. *The American Statistician* 49:12-17.
- Romeu, J. L. 1986. Teaching Engineering Statistics with Simulation: A Classroom Experience. *The Statistician (RSS Series D)* 35:441-448.
- Romeu, J. L. 1994a. Validation of Multivariate Monte Carlo Studies. In *Proceedings of the IMSIBAC-4*, ed. M. L. Puri and J. P. Vilaplana. VSP Int. Science Publ. Netherlands.
- Romeu, J. L. 1994b. Simulacion, Pedagogia y Estadistica. *Actas del Primer Seminario de la Enseñanza de las Matematicas*. ITAM. Mexico.
- Ross, P. 1995. What Government Needs. *The American Statistician* 49:7-9.
- Schreiber, T. 1991. *An Introduction To Simulation Using GPSS/H*. New York: Wiley.
- Snee, R. et al. 1980. Preparing Statisticians for Careers in Industry. *The American Statistician* 34:65-75.
- Thesen A. and L. Travis. 1992. *Simulation for Decision Making*. St. Paul: West.

## AUTHOR BIOGRAPHY

**JORGE LUIS ROMEU** is an Associate Professor in the Department of Mathematics at the State College at Cortland, New York. He received a Licenciado en Matematicas degree from the University of Havana, in 1973, and M.S., M.Ph. and Ph.D. degrees from Syracuse University in 1982, 1987 and 1990, respectively. His research interests are in simulation and statistical modelling and analysis. Dr. Romeu is a Chartered Statistician Fellow of the RSS, a member of ASA and INFORMS and a recipient of a 1994 Fulbright Senior Lecturership Award to teach Simulation at ITAM, in Mexico City.