Statistical Simulations in the Web

Juha Puranen
University of Helsinki, Department of Mathematics and Statistics
Gustaf Hällströmin katu 2b, P.O. Box 68
00014 University of Helsinki, Finland
Juha.puranen@helsinki.fi

1. Introduction

Use of simulations is an essential part of teaching statistics. On low level courses the students make the simulations in the class by themselves, but on the high level cources it is more common to use computers. At the moment there are a lot of Web pages for statistical simulations. The quality varies, but many high quality pages are available.

In order to improve my teaching I have both constructed a variety of statistical simulations and presented these on my course pages. For simulation I have mainly used HTML, JavaScript, DHTML and Flash. At the moment it seems to me that Flash gives more degrees of freedom than other methods I used.

In the following I will outline the essential points of my experiences using statistical simulations on the Web and as important part of my teaching.

2. Tools and examples

The simplest way to make simulations in the Web is to use JavaScript. JavaScript is a programming language where it is possible to generate random variables, simulate random processes and show the result on the Web page.

JavaScript also has advantages in presenting ready-made pictures which can be used to visualize statistical simulations that have been made in advance. This is always a safe choice, because then one can be sure that the text and the simulation are well integrated. The teacher knows exactly, what the student will see on the screen. (Showing real life simulation can be risky).

Figure 1 is the last picture from a series of 19 pictures, shown one by one, to visualize how one can generate samples from the normal distribution, plot them on the probability paper and in this way gets a 95 % confidence interval for the empirical cumulative density function.

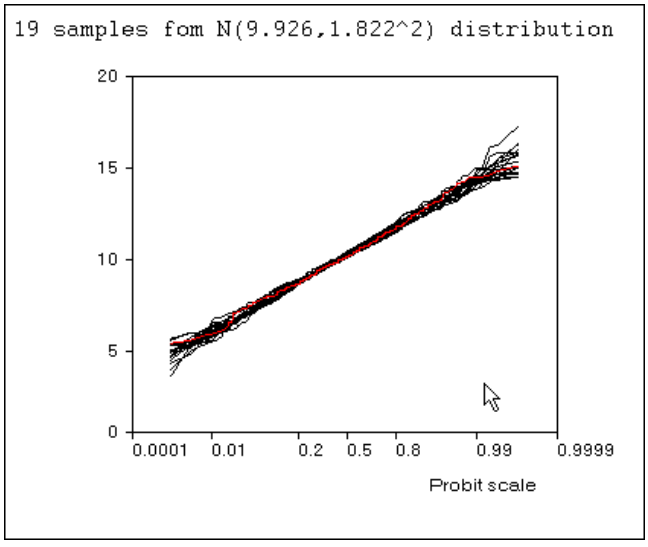

Figure 1. Probability paper test.

Just showing the result on the screen is not interesting, the students wants some Aaction@. One way to get more life to the simulations, is to use DHTML. In DHTML it is possible to move elements on the screen.

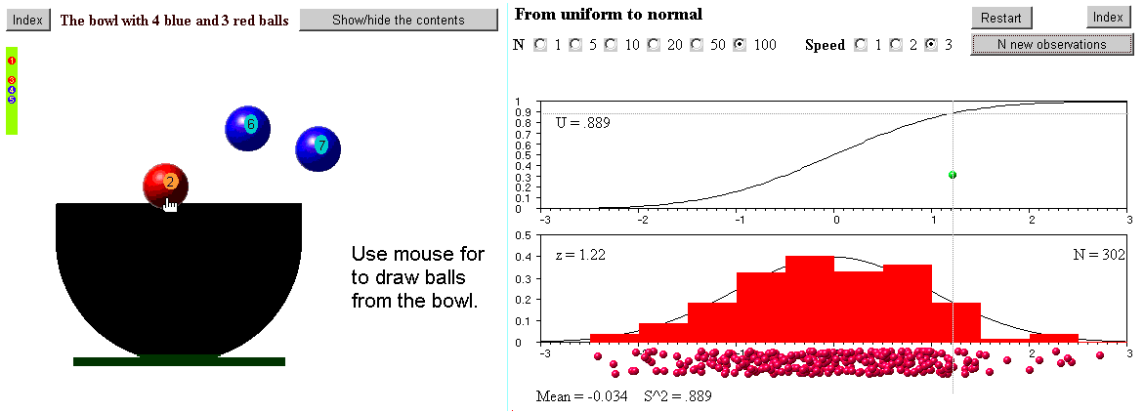

Figure 2. DHTM demos

On the left of figure 2 is an urn simulator, where it is possible to draw balls from the bowl and show the difference between "with replacement" and "without replacement". On the right there is an example for to show how normal distributed random variables can be generated using uniformly distributed random variables and the cumulative normal distribution function.

Unfortunately there is no standard for DHTML, so the urn simulator works only in Netscape 4.7 or 4.8. It is a pity, because DHTML was quite a flexible tool for making educational material.

At the moment I am rewriting my old simulation again, this time using Macromedia Flash. With Flash one can create dynamic graphics, interactive simulations, multilingual examples and parameter controlled applications.

As an example here is the output of one Flash demo with 2 different parameter-files.

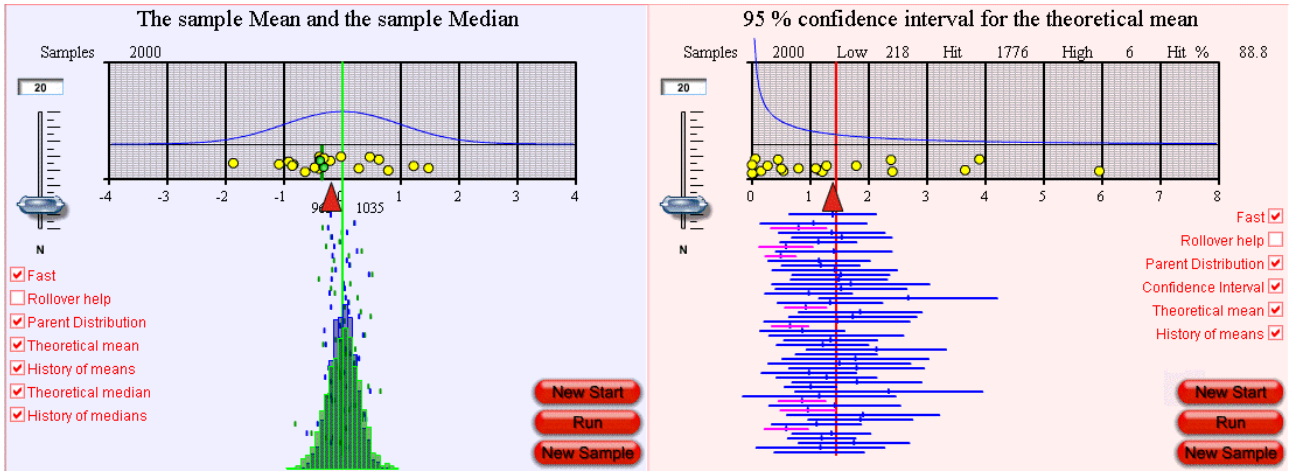

Figure 3. Flash demos

On the left of figure 3 are the distributions of the sample mean and the sample median from a symmetric normal distributed parent population. The sample mean is a slightly better estimate for the location parameter than the sample median. On the right is the t-confidence interval for the theoretical mean when the parent population is extreme skew to the right. When the sample population is skewed and the sample size is small the assumptions for the t-confidence interval are not valid and the true confidence level is .88 instead of .95.

Although the statistical characters in these demos are different, both are almost identical at the programming level.

In the class it is good if the simulation program has many characters. The teacher can take one character at a time and use it in his/her teaching. The situation is different on the Web where the simulation program can teach only one character at a time - everything else is disturbing. In order to maintain these demos it is much easier if we can put everything in the same package and then by the means of a parameter file define those characteristics used at any time.

In Flash it is relatively easy to add content sensitive help to every element in the applet and in this way compose interactive educational material.

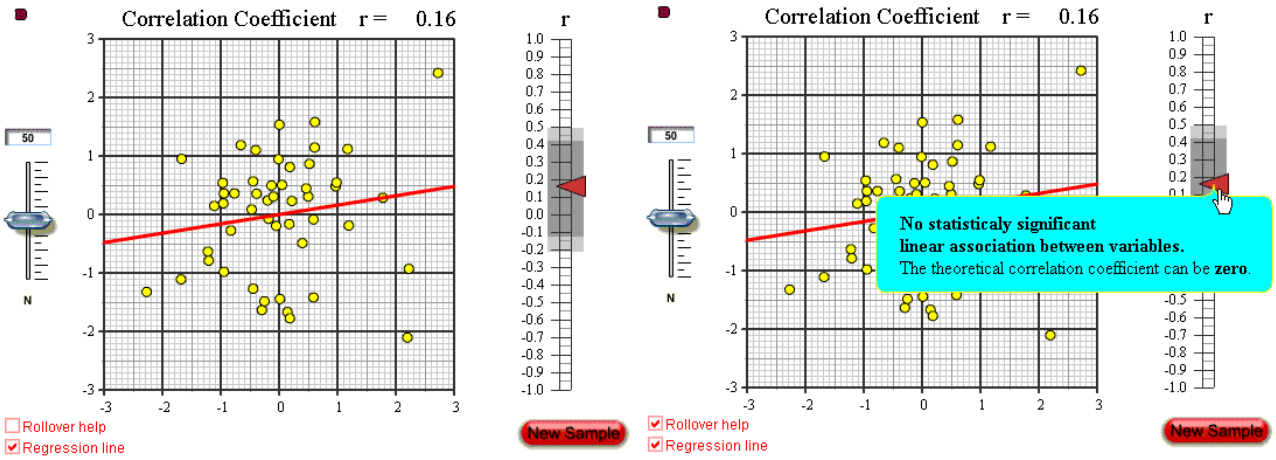

Figure 4. Context sensitive help

In the figure 4 there are two screen shots from a Flash version of my old correlation demo. On the left there are the 95% and 99 % confidence intervals for the theoretical correlation coefficient. On the right the same situation, when the box "Rollover help" is checked and the student points the arrowhead at the correlation scale.

3. Problems

Using Web-based simulations in the teaching is not without problems. Here is a short list of problems, I have met.

Lack of help.

There are many excellent statistical demos in the net. Unfortunately they are usually for classroom use only.

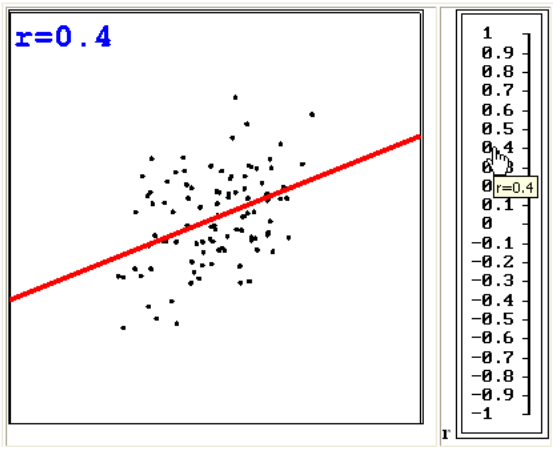

Figure 5 is perhaps the most popular of my demos. When one moves the mouse pointer on the scale, the form of the scatter plot changes correspondingly. This demo works well in the class, where the teacher is explaining, what happens and what the students should look at. But when the student sits alone in a computer class and tries to study correlation, he can be totally lost if this the only demo he/she uses.

Most of the statistical demos in the Web are like this. No theory, no context sensitive help.

Figure 5. Scatter plot and the correlation coefficient

Screen resolution. Usually the teacher plans the demo so that it looks good on his/hers computer but other teachers use other resolution and then he/she can only see only a part of the demo or everything is too small for classroom use. (It is the same with pictures and formulas.) I have solved this problem by using two versions of the same demo -one small one with background theory - and one saleable for classroom use.

The way of perceiving things. Students are different and so are the teachers. A simulation that is good for one is not good for another. Usually there are many versions of the same simulations available on the Web, so it is possible to take one into the class, and link to other versions on the course page.

Slow line. At home I have only a 56 Kb modem connection (usually slower), and downloading a demo over the net can take an eternity. Java applets which load several libraries first are especially slow to download.

Blug-in's. The browser needs a plug-in for Java, Flash and many other demos. If the computer one is using does not have this plug-in installed, then a good demo is useless. Nowadays when the people are afraid of computer viruses, they are unwilling to download plug-ins that they do not understand.

Technical development. Technical development is also a problem. Every day there are better and better methods for to present statistical simulations in the net (class). Old methods are not supported anymore. Now there are demos, for which I have done several versions (Basic, C, HTML, DHTML, and Flash). What next?

4. Conclusions

Although I sometimes suspect that the students are just looking enchanted at the moving elements on the screen without hearing or thinking anything, I am sure that simulations are effective way to teach statistical phenomena. This is because when I stop the simulation, they listen me again and this time with more attention.

REFERENCES

Blejec, A. (2003), Teaching statistics by using simulations on the internet. IASE satellite Conference on Statistics Education and The internet, Berlin.
http://www.ph-ludwigsburg.de/iase/proceedings/index.html

Puranen, J. (1984). DISTEP a Dynamic Interactive Statistical Teaching Package, Compstat 1984 Proceedings in Computational Statistics p 105-110

Puranen, J. (1997-2004) Some Experimental WWW pages for to teach Statistics
http://noppa5.pc.helsinki.fi/koe/index.html

RÉSUMÉ

Dans cet papier, je vais présenter les points centrales de mes expériences d'utiliser les simulations statistiques dans l'internet.