

## Teaching Confidence Intervals with Java Applets

### KEYWORDS:

Teaching;  
Central Limit Theorem;  
Confidence intervals;  
Java applets.

*Andrew Bertie and Paddy Farrington*

Open University, England.

e-mail: a.bertie@open.ac.uk

c.p.farrington@open.ac.uk

### Summary

Confidence intervals are pedagogically important but often misinterpreted. This article describes Java applets designed to help students understand two interpretations of confidence intervals.

### ◆ INTRODUCTION ◆

Confidence interval estimation is an established component of the conventional statistics curriculum, and some have argued that it is a key one. Tversky and Kahneman (1971) suggest that a confidence interval provides a useful index of sampling variability and is less prone to what they call the 'belief in the law of small numbers' – particularly compared to hypothesis tests. Wonnacott (1987) gives further reasons for preferring confidence intervals to hypothesis tests. Thomason et al. (1994) describe confidence intervals as the link between sampling variability and hypothesis testing. Mills (2002) suggests that students' understanding of confidence intervals can provide important long-term benefits.

In the Neyman–Pearson framework, confidence intervals and hypothesis tests are just two sides of the same coin, so pedagogically it makes sense to teach confidence intervals first. The correct interpretation of a confidence interval is, however, notoriously difficult for students to grasp (Hawkins et al. 1992, p. 83).

Robison-Cox (1999) focuses on students' difficulties in understanding confidence as an abstract process, and describes physical experiments to overcome this. Others have used computers to similar ends: statistical software is used to generate many confidence intervals and show that the proportion of intervals that include the true parameter corresponds to the confidence coefficient. This may be described

as the *repeated experiments* interpretation of confidence intervals. Specially developed simulation programs are likely to be more successful than a commercial statistical package for this, since they can be designed to display the intervals graphically, provide greater interactivity and instant visual feedback – making it easier for the student to experiment with different confidence levels, sample sizes, distributions and parameters. A number of such programs were developed in the 1990s – usually for the Windows/PC platform and written in programming languages such as C++.

More recently the World Wide Web has provided several Java applets for demonstrating the repeated experiments interpretation that can be run by anyone with access to the Internet; a typical example is described by West and Ogden (1998).

Simulation experiments such as these aid understanding of the reasoning behind the construction of a confidence interval, but do not help with the interpretation of an interval calculated from a single sample. The repeated experiments interpretation relates only to the properties of random intervals, and provides no guidance to students on what the intervals they have calculated actually mean. Consequently, students are tempted to apply probability statements that are valid for random intervals to the intervals they have calculated, for which these statements are not valid.

There is, however, an alternative to the repeated experiments interpretation of confidence intervals,

which is less commonly taught but may be more meaningful to students as it relates to the particular interval they obtain from a single sample. This interpretation defines a confidence interval as a *plausible range* of values of the unknown parameter. This range includes all values of the parameter not rejected in a fixed-level test.

This interpretation is equivalent to the repeated experiments interpretation in large samples, and forms the basis of some commonly used algorithms for calculating ‘exact’ confidence intervals in small-sample settings. The plausible range interpretation also corresponds to how confidence intervals are used in practice, and is easily conveyed graphically by sliding the sampling distribution along the parameter axis. Therefore, from both the pedagogical and the practical points of view it is beneficial to teach both the repeated experiments and the plausible range interpretations of confidence intervals, and to emphasize the links between them. But the teaching of the plausible range interpretation has not, until now, been well supported by software or applets.

*M248 Analysing Data* is a new introductory statistics course presented by the Open University from 2003. As the course title suggests, the emphasis is on learning statistics through analysing data rather than mathematics. MINITAB will be provided for data analysis, but it was decided to develop interactive Java applets to support the teaching of more difficult statistical concepts. Three of these applets supporting the teaching of confidence intervals are described here.

It was not absolutely necessary to make this software Web based, but statistical programming in Java has a number of advantages (see Murdoch 1999; Lang 2000). A program written in Java can be deployed as a Web-based applet or as a stand-alone application with little modification – the latter requiring no modifications to run on different computers and operating systems. The option of providing the software as stand-alone applications avoids the costs of accessing the Internet incurred by students off campus – a problem highlighted by Hunt and Tyrrell (2000).

---

## ◆ REPEATED EXPERIMENTS ◆

---

The *Repeated Experiments* applet demonstrates the repeated experiments interpretation of confidence

intervals for six standard distributions and an arbitrary distribution called ‘Your distribution’ that the student defines by manipulating its pdf curve. In all cases the approximate large-sample  $100(1 - \alpha)\%$  confidence interval for the population mean is calculated using  $\bar{x} \pm zs/\sqrt{n}$ , where  $\bar{x}$  is the sample mean,  $s$  is the sample standard deviation and  $z$  is the  $100(1 - \frac{1}{2}\alpha)\%$  percentile of the standard Normal distribution.

There is a tabbed panel for each of the distributions. Figure 1 shows the applet with the **Your distribution** tab selected. Each panel has the same layout. The population is represented in the top left of each panel – by a plot of the distribution and fields for its parameters and the corresponding mean  $\mu$  and standard deviation  $\sigma$  of the population. In the bottom-left area of each panel are controls for the simulated sampling of confidence intervals, consisting of fields for specifying the confidence level, sample size and number of samples, a check box for switching to single-step mode, and buttons for starting the simulation, sorting the confidence intervals and resetting the applet.

When the **Take samples** button is clicked, the specified number of samples of the specified size is randomly generated from the population. For each sample, the large-sample confidence interval for the mean of the population is calculated and plotted as a horizontal line in the scrollable panel on the right. Where an interval fails to include the population mean, the colour of the line is red; otherwise it is black. The intervals can be sorted with respect to their means (midpoints) by clicking on the **Sort** button.

Summary results about the simulation appear below the plot: the number and percentage of intervals that include  $\mu$ . The individual simulated lower and upper confidence limits are also displayed. In single-step mode these limits are displayed as each sample is generated; otherwise the limits of each interval are displayed as the mouse cursor passes over the interval’s line in the plot.

The applet can also be used to explore the accuracy of the approximate confidence interval, and the Central Limit Theorem that it depends on. A further panel behind the plot of intervals, labelled **Sampling distributions**, displays the superimposed pdfs of the exact and approximate (large-sample) distributions of the sample mean, and a histogram of sample standard deviations.

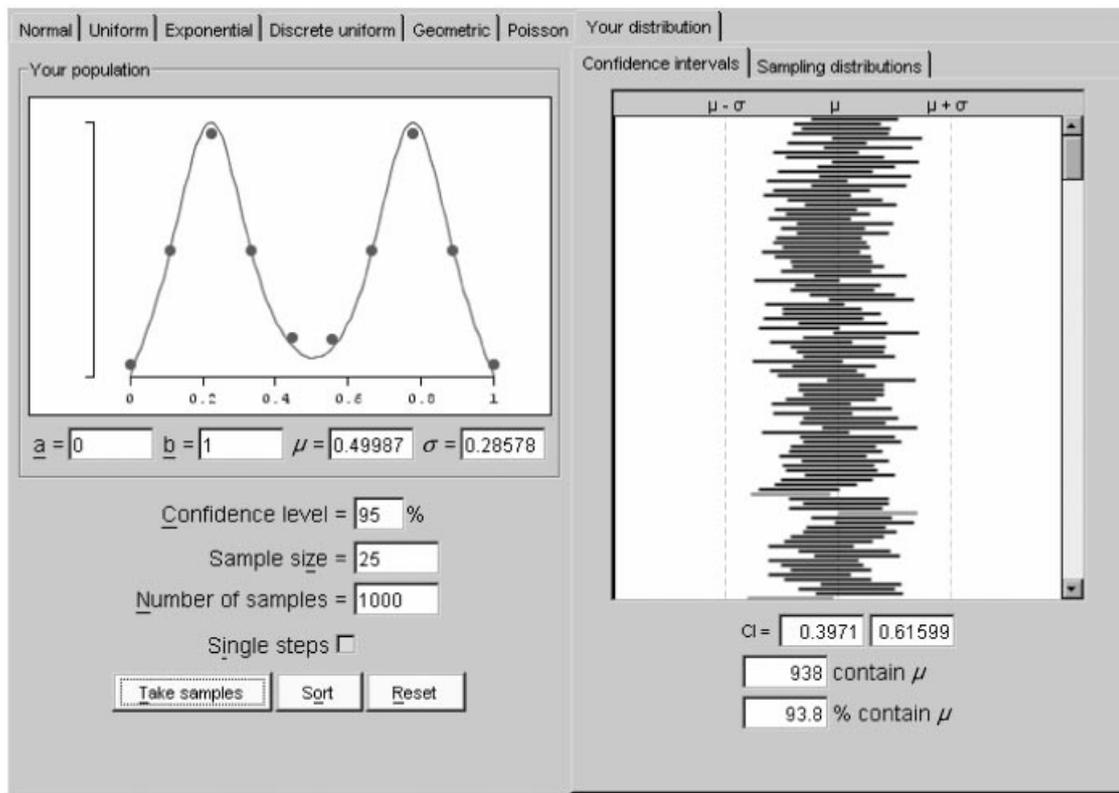


Fig 1. The *Repeated Experiments* applet

Although in some respects the applet resembles some applets on the Web, we believe our version is an improvement in combining some of their better features (e.g. sorting the intervals as in the Rossman/Chance applet at <http://statweb.calpoly.edu/chance/applets/confsim.html>), and introducing innovations such as 'Your distribution' and the sampling distribution plots.

### ◆ PLAUSIBLE RANGES ◆

The *Plausible Ranges* applet demonstrates the interpretation of large-sample confidence intervals as a range of plausible values of the unknown parameter. The layout is similar to the *Repeated Experiments* applet and deals with the same seven population distributions. Figure 2 shows the applet with the **Normal** tab selected.

When the **Take sample** button is clicked, a sample of the specified size is randomly generated from the population. The sample mean, sample standard deviation and large-sample confidence interval for the mean of the population are calculated and displayed in boxes below the buttons. The values of the sample are plotted as a rug on the plot of the population distribution.

On the right-hand side of the screen, two tabbed panels labelled **Plausible  $\mu$**  and **Plausible range** are designed to help the student understand the plausible range interpretation of the confidence interval obtained from the sample. Both these panels display the pdf of the approximate (large-sample) distribution of the sample mean for the given population and sample size,  $n$ . The sample mean is indicated on these plots by a vertical dashed line. In the **Plausible range** panel (shown selected in figure 2), the student explores the plausibility of values of  $\mu$ , and their corresponding tail areas, by adjusting the slider or typing values of  $\mu$  into the field. In the **Plausible  $\mu$**  panel, students choose a value for  $\mu$  that they think is plausible and choose which tail area to display. The percentage value of this area is displayed and the student is asked to decide whether the chosen value of  $\mu$  is plausible at the specified confidence level.

The *Exact CI for a Proportion* applet (figure 3) demonstrates the plausible range interpretation of exact confidence intervals for a proportion.

A random sample is generated from the Bernoulli population and the number of successes, number of failures, estimated value of  $p$ , approximate (based on the Normal approximation to the

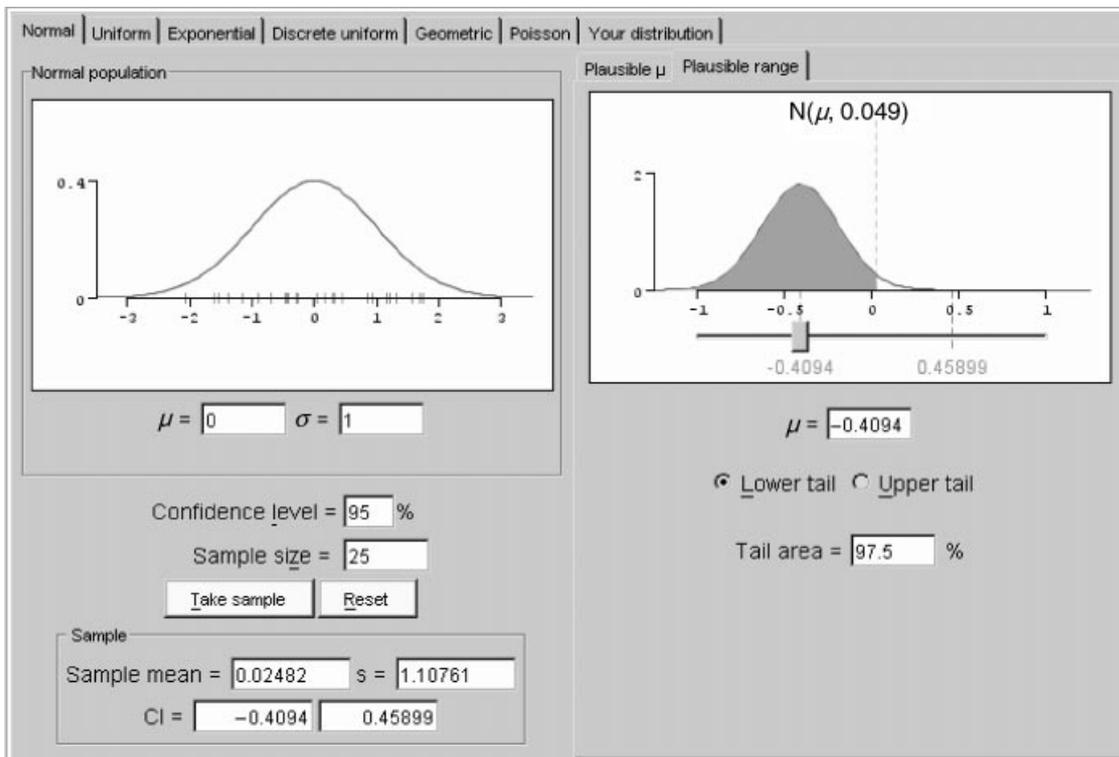


Fig 2. The *Plausible Ranges* applet

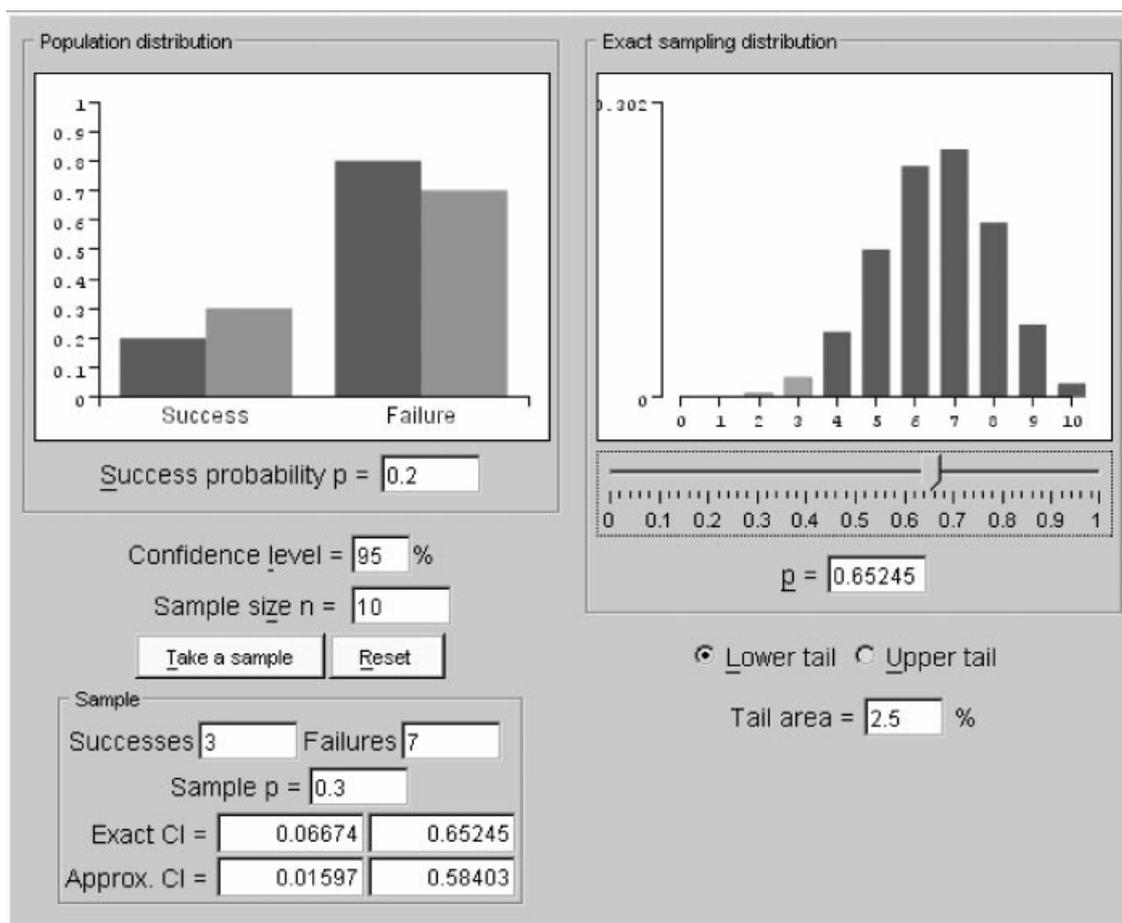


Fig 3. The *Exact CI for a Proportion* applet

binomial distribution) and ‘exact’ (Clopper–Pearson, as calculated by MINITAB) confidence intervals for  $p$  are calculated and displayed. The sample proportions of successes and failures are plotted on the population plot. On the right-hand side of the applet the student explores the plausibility of values of  $p$ , and their corresponding tail areas, by adjusting the slider or typing values of  $p$  into the field. This applet differs from the *Plausible Ranges* applet in that the approximate and exact intervals can be compared, and the special issues associated with confidence intervals for a proportion can be explored – such as what happens when the sample has zero or  $n$  successes.

---

◆ CONCLUSION ◆

---

Confidence intervals are taught before hypothesis testing in *M248*, using the duality between confidence intervals and two-sided tests of significance referred to above. A further applet on hypothesis testing has a similar structure to the *Repeated Experiments* and *Plausible Ranges* applets.

Rossman et al. (2000) have suggested presenting students with alternative formulae for constructing approximate and exact confidence intervals for a proportion, and asking them to compare their performance through simulations. The choice of confidence interval for a proportion is still a matter of debate (e.g. Agresti and Coull 1998), and such simulations, they suggest, would have pedagogic value in demonstrating the dynamic nature of statistics and involving the students in evaluating recent research. This might be an interesting development of the *Exact CI for a Proportion* applet.

Open University students will use the stand-alone application versions of these applets with printed computing activities, but basic instructions and suggested activities can be found in the Web pages containing the applets at <http://www2.open.ac.uk/CES/projects/SUStats/SUStatsApplets.html>. The authors welcome comments on the applets from teachers and students.

## References

- Agresti, A. and Coull, B. (1998). Approximate is better than ‘exact’ for interval estimation of binomial proportions. *The American Statistician*, **52**, 119–26.
- Hawkins, A., Jolliffe, F. and Glickman, L. (1992). *Teaching Statistical Concepts*. London: Longman.
- Hunt, N. and Tyrrell, S. (2000). Learning statistics on the Web – DISCUSS. *Teaching Statistics*, **22**(3), 85–90.
- Lang, D.T. (2000). The Omegahat environment: new possibilities for statistical computing. *Journal of Computational and Graphical Statistics*, **9**, 423–51.
- Mills, J.D. (2002). Using computer simulation methods to teach statistics: a review of the literature. *Journal of Statistics Education*, **10**(1).
- Murdoch, D.J. (1999). What’s Java? *Chance*, **12**, 44–7.
- Robison-Cox, J.F. (1999). Having a ball with confidence intervals. *Teaching Statistics*, **21**(3), 81–3.
- Rossman, A., Chance, B. and Ballman, K. (2000). A data-orientated, active learning, post-calculus introduction to statistical concepts, methods and theory. To appear in *Proceedings of the Section of Statistical Education, Joint Statistical Meetings*, August (<http://rossmanchance.com/scmt/JSMpaper00.htm>).
- Thomason, N., Cumming, G. and Zangari, M. (1994). Understanding central concepts of statistics and experimental design in the social sciences. In: K. Beattie, C. McNaught and S. Wills (eds), *Interactive Multimedia in University Education: Designing for Change in Teaching and Learning*. Amsterdam: Elsevier Science (North-Holland).
- Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, **76**(2), 105–10.
- West, R.W. and Ogden, R.T. (1998). Interactive demonstrations for statistics education on the World Wide Web. *Journal of Statistics Education*, **6**(3).
- Wonnacott, T. (1987). Confidence intervals or hypothesis tests? *Journal of Applied Statistics*, **14**(3), 195–201.