ENABLING STUDENT INVESTIGATION OF ROBUST STATISTICS AND COST-BASED STATISTICAL DECISION-MAKING: LAYING A FOUNDATION FOR STATISTICAL SOPHISTICATION

Hunter Ellinger, Univ. of Texas at Austin; Mary Parker, Austin Community College
Hunter Ellinger, 1622 Waterston, Austin TX 78703 (hunter.ellinger.org, ellinger@io.com)

Key Words: Education, Decision-Making, Cost

Abstract

We report on our initial experience in a project to develop and test web-based programs designed to enable students to investigate the behavior of various measures of central tendency (e.g., mean, median, trimmed means) and dispersion (e.g., variance, interquartile distance, full width at half-maximum) for samples from arbitrary populations, and to learn about constructing cost-minimizing decision methods using statistical concepts.

Analytical, numerical, and simulation approaches are planned to illustrate such questions as the efficiency and robustness of estimators and rates of convergence to normal. Because users will be able to enter arbitrary distributions and adjust the parameters controlling the statistics (e.g., the trimming fraction for a trimmed mean), they will be able to investigate a very wide range of questions.

Cost-based decision making is examined in a context in which statistical principles can be applied to situations that resemble realistic decisions. The cost of various types of error (and the cost of measurements) can be deduced from the problem description (or supplied from a real problem); the software will both illustrate and solve the resulting minimization problems.

Project Goals

The software reported on in this article is being developed because the authors feel that computer technology provides many unrealized opportunities to expand students' mathematical and statistical sophistication, in addition to the expanded numerical-calculation power that is more widely appreciated and used.

The basic goal is to *provide helpful windows into mature statistical and mathematical thinking*. While the materials may be useful in teaching the standard content of some statistics courses, they have not been developed with that in mind. They are intended primarily as enrichment materials, for both students and teachers, that will not require an accompanying course or text to provide their value.

A related goal is to help people connect statistical, mathematical, and operational viewpoints. The variety of perspectives and goals that has at times fragmented the discipline of statistics also provides opportunities to show how particular issues or problems can have solutions that are interesting (and finally accessible, due to computers) in several different ways.

A final general goal is to stimulate broad participation in the development, criticism, use, and free distribution of software developed along these lines by ourselves or others. We plan to make full use of the internet, with most projects implemented as web sites.

Project Approach

The fundamental tactic of the project is to take full advantage of computers to provide transitions both from introductory concepts to the powerful abstractions of mature mathematical thinking and from informal interactions with statistical processes to the productive insights and methods of professional practice and theory. While the biggest challenges are the educational ones of motivation, clarity, and accessibility, we also see it as vital that the software be scrupulously authentic (if still somewhat simplified) from a mathematical and professional level.

While presentation and manipulation of numbers is a pervasive part of statistics, graphical techniques have proven of greater value for the presentation of data and its derived statistics. In addition to the rich legacy of standard statistical graphics, we have made use of a variety of techniques mode possible by the advent of full-color computer monitors and efficient computation: multidimensional graphs, animations of mathematical transformations, and coordinated diagrams that give multiple simultaneous perspectives.

We have also tried to use the intermediate complexity of the material to construct connections between separated professional domains, showing how an issue such as making an optimal decision about a measurement process connects to more advanced statistical, mathematical, and operational methods and issues.

A final element of the approach we espouse is the visible provision of messages as many levels. Students should be shown not just "facts" (usually specific techniques and results) but ideas, approaches, and styles. This layering of several levels and directions of abstraction, which historically has been fostered by direct contact with skilled and learned practitioners, is an essential element in the development of sophistication and maturity in a field. If computer-based systems are going to provide serious educational support, they need to take full advantage of their opportunities to promote awareness of such distinctions.

Areas of Application I: the Central Limit Theorem

While the Central Limit Theorem holds a prominent place in statistics courses, most students learn it rather than understand it. We think that this is largely because it contains too many ideas to be fully explored in the packed introductory course. But it is an excellent focus for enrichment materials, due to its broad applicability and the possibility of separate investigation of its various elements.

Unpacking the ideas: In its assertions about the behavior of the distributions of statistics with increases in sample size, the theorem deals with three distinct ideas. The first (which might be called "centering") is that the expected mean of the samples is that of the population. The second ("narrowing") is that the variation of the central measure around its expected value decreases in a predictable way as the size of the sample increases. The third idea ("normalizing") states that the *shape* of the distribution of the central measure approaches a standard form unrelated to the population form (except for a shift to match the population mean and a single scale factor based on the square root of the population variance).

Expanding the horizons: While students generally get the main point (that the statistics of means of even moderate-sized samples can safely be treated as if they were normal distributions with the population mean and variance), it is rare for them to look at just how these ideas fit together. Would the theorem hold for other central measures than the mean? Other dispersion measures than variance? How could you measure how different the expected distributions are from the corresponding normal? Such questions, while often analytically difficult or intractable, can be investigated numerically even by beginning students by using a simple set of computer-simulation tools.

Learning to roll the dice: Our programs, which have been prototyped in a leading computer-algebra system but are suitable for conversion to Java applets, permit students to investigate the properties of a variety of central measures (e.g., median, trimmed mean, midrange) as well as a wide range of populations. They permit the students to pose and answer interesting questions like the ones above. Extensive simulations based on random sampling provide both convincing answers and the experience of research, as well as extensive exposure to probabilistic situations.

. *Presentations to separate and reunify the ideas:* Simulation results are not the only approach to take to these ideas. Once the idea has been established (usually by simulation, but alternatively by calculus for some audiences) that each sampling distribution has a "smooth" asymptotic form, the familiar properties of such graphs can be used to illustrate some other connections. We have used such "theory curves" both to give students a clearer picture of the overall effect of the convergence-to-normality process and to separate the narrowing and normalizing aspects.

Figure 1 below shows an example. The panel on the left shows a population (solid line) and the asymptotic distributions for the means for samples of size 2, 4, and 8 (short, medium, and long dashes, respectively). This shows the predicted combination of narrowing and normalizing effects of increased sample size. The omission of the noise from the randomization in a simulation makes the comparisons more striking.

But the real improvement in comparison is shown on the right. Here the same curves are graphed, except that each of the sample curves has been widened by a scale equal to the square root of the sample size, thus offsetting the narrowing effect to make exact shape comparison possible. While the graph on the left only showed that the curve became "bell-shaped", this graph on the right indicates much more strongly that a uniform shape is being converged to, after the predictable scale factor is taken into account. The effect is even stronger if the results from populations with different shapes (but the same mean and variance) are overlaid using colored graphs (not shown).

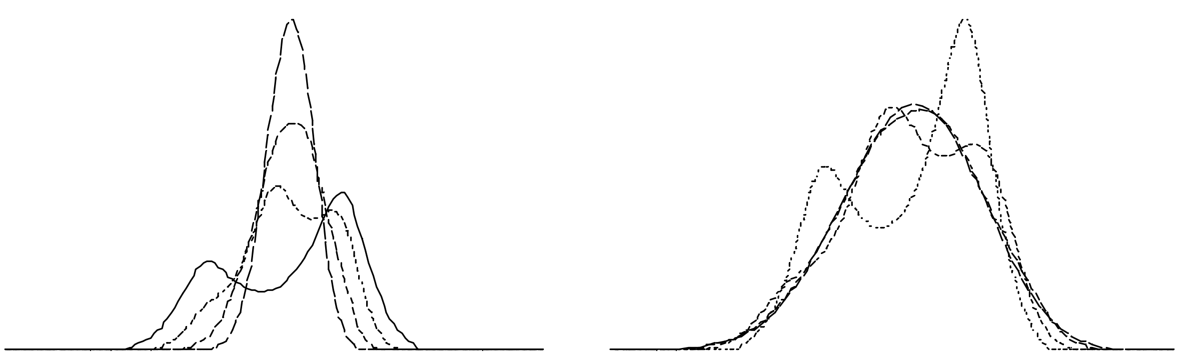
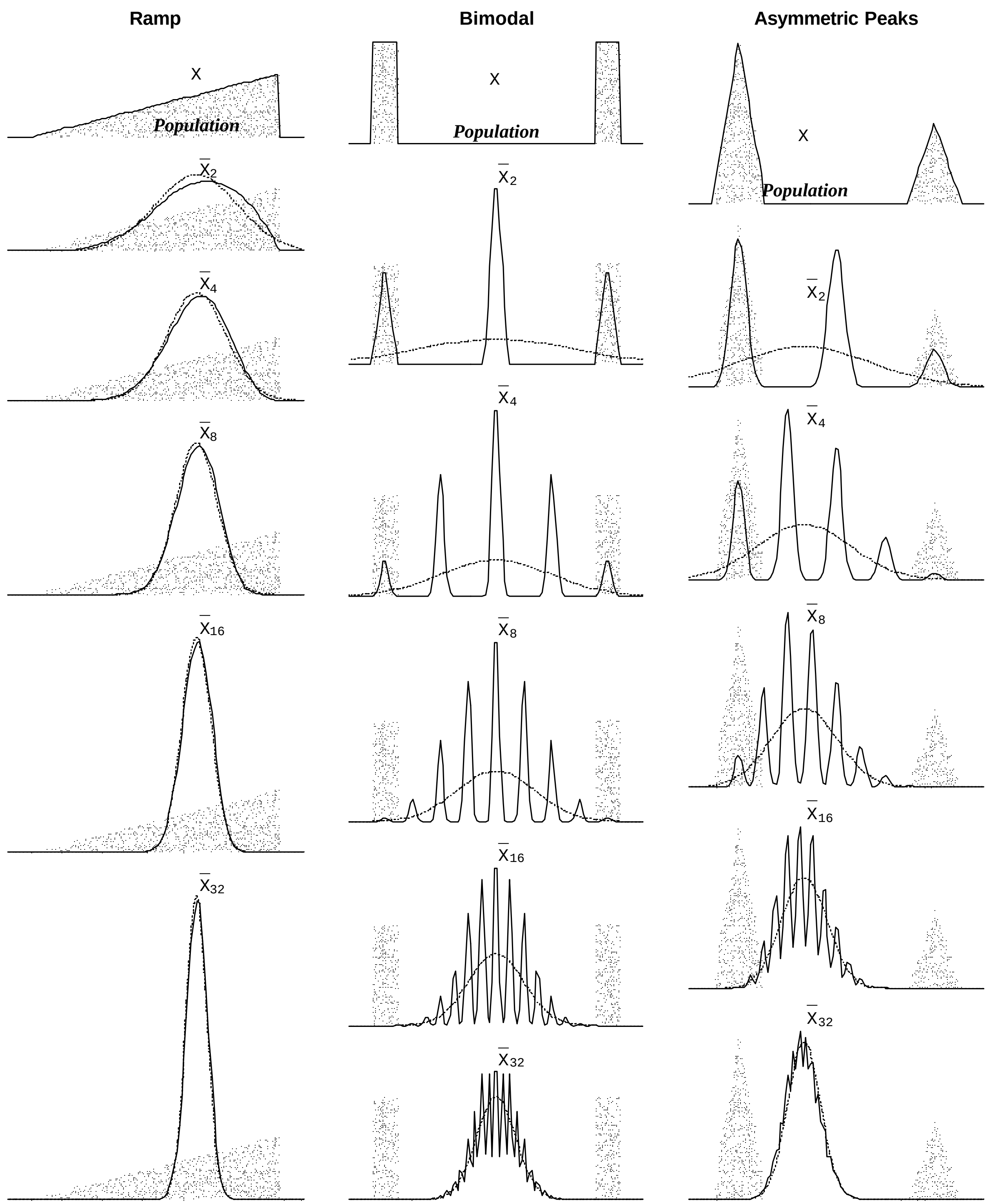


Figure 1. Left: an population (solid line) and the expected distribution of the means of samples of size 2, 4, and 8 (short, medium, and long dashes, respectively). Right: the same set of curves (here the population is the dotted line), with each curve expanded horizontally by the square root of the corresponding sample size. This balancing of the narrowing effect permits clear illustration that the predicted convergence to a standard shape takes place.

Distributions of means for samples of increasing size from various distributions



Areas of Application II: Cost & Statistical Decisions

The second area we have chosen for development is a sector of inferential statistics, *making optimal decisions with uncertain information*, that we feel has been underutilized but has great potential to enliven an area that many students still find difficult and boring.

While calculation of the probability that the deviation from expectation shown by a particular set of data was a result of chance will always be a central part of statistics education, few people are interested in error analysis *per se*, in the absence of connections to goals they might plausibly pursue. The main interest people have is in producing results of maximum benefit in their areas of activity; the detection and analysis of mistakes is relevant to that task only to the extent that these mistakes entail significant avoidable costs.

We are developing modules that use the general interest in net-benefit maximization to motivate a variety of inferential-statistics investigations. The approach is to direct attention beyond the simple likelihood of mistakes to their *consequences* (it is of course common for different types of mistake to have very different costs). The idea is also introduced that data has a cost, and that one must decide not only what threshold to use to make decisions, but how much data it is worthwhile to acquire.

Properly developed, we feel that this area could make a major contribution to statistics education, and we are preparing a complete set of modules to address it. We give below one of the more complex examples we have used, which includes all the features mentioned. In the software, the images shown are part of an animation sequence.

The situation:

A sealed machine is built that weighs exactly 1000 pounds when correctly assembled. However, some units have omitted a necessary piece that weighs exactly one pound.

A mechanical scale with $\sigma = 10$ pounds for individual measurements is the only device available for determining whether to reject a particular unit.

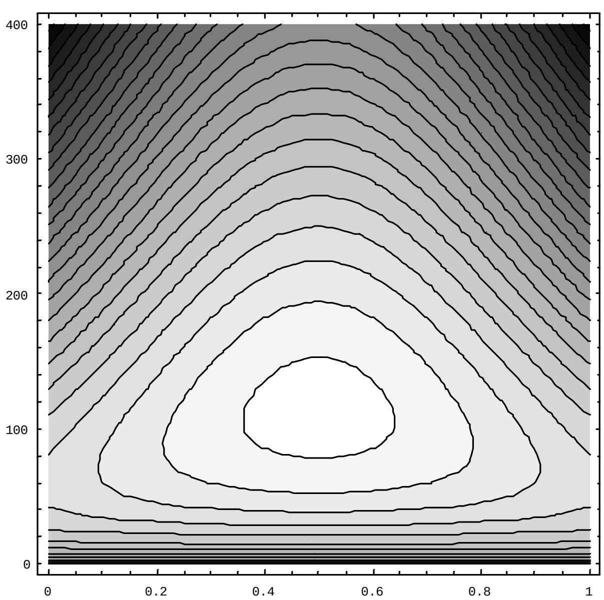
Measurements with the scale cost $1 each. Mistakes in sorting (either direction) cost a penalty of $1000 each.

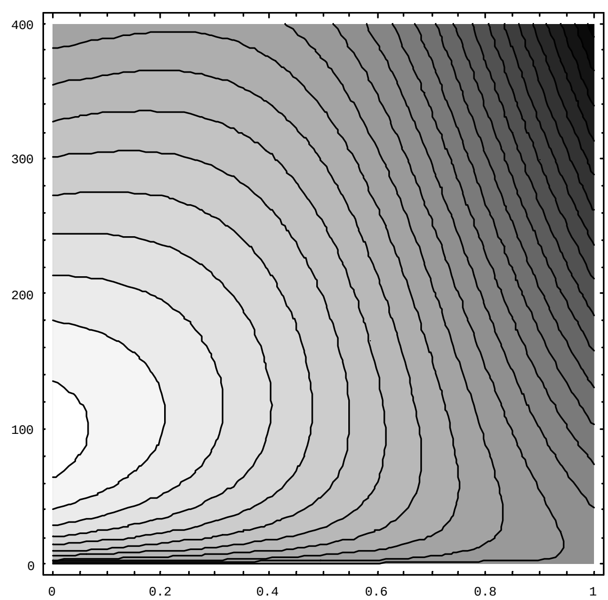
The person controlling the sorting makes two specifications:

p – the partition weight (above→accept, below→reject) – between 999 and 1000

n – how many measurements to average for each decision

What *p* and *n* setting should be used? It depends on another: *the expected relative frequency of omissions*





The two contour graphs above map the average cost per item of the sorting process as a function of partition weight (horizontal) and number of measurements (vertical), given the particular costs listed above. Because both measurements and errors have costs, a maximum occurs at intermediate values (at a partition weight of 1000.5 and approximately 113 measurements in this case if widgets and gizmos are equally frequent—see the left graph below).

The optimal strategy for setting the controllable parameters changes if gizmos and widgets are not equally likely. If widgets are 9 times more likely than gizmos, the same other conditions displace the maximum toward a partition weight much closer to 1000 and to somewhat fewer measurements (about 75).