Every Lesson in the 'Computer Lab': Teaching Introductory Probability and Statistics with Personal Laptops in the Classroom

Richard C. Bell, Jr., Department of Mathematical Sciences, United States Military Academy, West Point, NY 10996

Andrew G. Glen, Department of Mathematical Sciences, United States Military Academy, West Point, NY 10996

KEYWORDS: laptop, computer, statistical software, EDF.

Abstract:

What if every day of a probability and statistics class was a computer laboratory day? How would this change your method of teaching an introductory course in probability and statistics? At the United States Military Academy at West Point wireless laptop computers have become a permanent part of the classroom and have changed our approach to teaching. Over the last three years, we have made a concerted effort to find the improvements technology has to offer and to steer clear of the pitfalls technology can bring to the classroom. Our method of teaching a calculus-based probability and statistics course has evolved into data-oriented approach to understanding distributions. We present some methods that we have developed that use spreadsheets and computer algebra systems to create an environment that helps students understand the foundations of probability theory and statistical inference.

1. INTRODUCTION.

The introduction of wireless laptop computers into the classroom at the United States Military Academy has made a profound impact on the way we teach introductory probability and statistics. Simply bringing technology in the classroom is no guarantee of an improved learning environment. The complexity of the issue begins with a large number of software choices, ranging from the friendly (Minitab), to the free (R), to the fiendish (SAS, no offense). Next there is the choice of which text, as they all highlight technological issues differently. Also compounding the problem are the difficult issues of the faculty accepting the vision of technology in the classroom (welcome to the math wars) as well as the faculty development necessary to make the effort viable. We have in the end adopted a time honored principle in our approach to incorporating technology into our classroom: we keep it simple. We will present our use of basic spreadsheet technology to introduce the notion of ‘distribution’ in a modeling context. We then will share our efforts at exploring fundamental probability theory by having students create their own statistical software from first principles using a computer algebra system. We purposefully do not use existing software packages, as we believe their use pre-supposes some statistics education. Instead we find greater learning value in having cadets create their own spreadsheets, templates, files, and software that teach probability and statistics notions. Then when they use a commercial package later on, they will have some understanding of its underlying principles.

All students are issued laptop computers with a pre-loaded software package upon their arrival at West Point. They are required to bring them to mathematics classes every day for their two years of core classes. Every cadet regardless of academic major, approximately 1100 students per year, must take our probability and statistics course in the fourth semester. Thus we see the gamut of abilities from students confident in their mathematics talents to students insecure in their technical skills. Prior to our course, all students have successfully completed the first three semesters of core mathematics at West Point: Mathematical Modeling, Differential Calculus (single and multi-variable), and Integral Calculus (single and multi-variable). Thus, not only does each student bring a calculus-based, modeling background into our classroom, but they also come with experience with Mathematica and Excel. Additionally, our small class sizes (no more than 18 students per class) facilitate excellent student-teacher interaction and more flexibility for instructors to assist students in troubleshooting computer problems that may arise during class. Typically, most mathematics classes consist of about 20 minutes of lecture and 35 minutes of exercises and problem solving. Under these conditions, in which every day can be treated as a laboratory day by the instructor, our classrooms are the ideal setting for exploring the appropriate way to integrate computer technology within a probability and statistics education. The two major innovations we will discuss here are 1) a data-driven modeling introduction of random variables and 2) the creation of basic, first-principled software to understand the essential properties of distributions.

2. A DATA ORIENTED APPROACH: RANDOM VARIABLES AS MODELS.

Rather than beginning our program of instruction with the usual definition of a random variable followed by the introduction of discrete random variables, we begin our semester with a sample of *n* data points. Our intent is to graphically model data in order to foster a better understanding of what a “distribution” function actually represents without immediately overwhelming our students with the myriad notations and symbols associated with probability theory. After calculating and discussing the importance of basic sample statistics (mean, median, variance, and standard deviation), we teach an algorithm using a computer spreadsheet to create an empirical distribution function (EDF) of the data. We define the EDF as

$F_n(x) = \frac{\#\,observations \le x}{n}$ and explain to our students that the EDF represents the accumulation of the proportion of the sample data and is thus a non-decreasing, step-function that ranges from 0 to 1 and has a domain of the entire number line with steps of height 1 / *n* at the sample points. We demonstrate to our students the usefulness of the EDF by using it to approximate various probabilities and percentiles about the population from which we drew the sample data. We ensure that every student can master the algorithm in class so that they can produce their own EDFs in out of class exercises.

After noting that the EDF is merely a graphical representation of the cumulative distribution of our sample, we next turn our attention to modeling the EDF with a smooth continuous function. Although the continuous functions we choose to model with are the exponential and Weibull cumulative distribution functions (CDFs), we only identify them as models of the EDF. In fact, at this point in our course, we have not yet even introduced terms such as the CDF and probability density function (PDF). We simply characterize these models as potentially useful functions that are widely used to model real-world distribution accumulation. We are able to save the new terminology and acronyms for future lessons once our students are firmly grounded in the notion of building such accumulation models.

An example is offered. A data set of 251 car weights for all 2002 model cars is provided at Oswego University's econometrics website (link provided in References). An EDF of the data is illustrated in Figure 1.

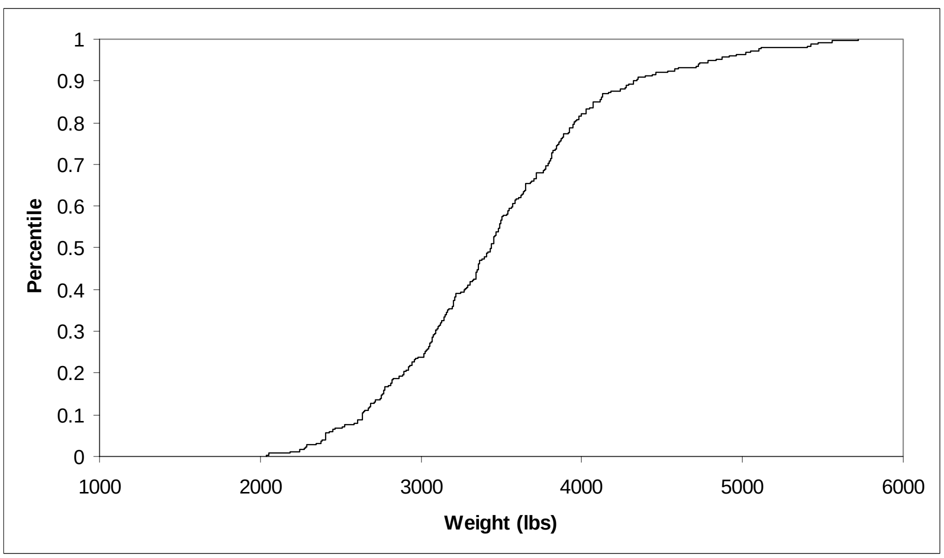


Figure 1: EDF of Car Weight Data

The first function we attempt to fit to the EDF is the model $F(x) = 1 - e^{-ax} \quad x, a \geq 0$. We find the parameter a that ensures the best fit by minimizing the sum of squared error (SSE) between $F(x)$ and the EDF. We use Excel as the technology of choice for modeling EDFs by setting up separate columns for the sample, the EDF, the model, the error, the squared error, and a SSE cell. A separate cell contains the parameter a. Initially we graph the EDF and the model on the same set of coordinates. As we vary *a*, we see the model move in relation to the EDF, sometimes improving, sometimes not. Eventually we show the use of the "Solver" tool in Excel which will vary the parameter until a minimum SSE is attained. An example of a few rows of such modeling can be seen in Figure 2.

weight	percentile	model	error	error^2	SSE
2035	0	0.346716	0.346716	0.120212	25.033
2035	0.003984	0.346716	0.342732	0.117465	
2055	0.003984	0.349443	0.345459	0.119342	
2055	0.007968	0.349443	0.341475	0.116605	
2183	0.007968	0.366633	0.358665	0.128641	
2183	0.011952	0.366633	0.354681	0.125799	
2242	0.011952	0.374403	0.362451	0.131371	

Parameter	
a	0.000209

Figure 2: Excel Screen Capture showing the Modeling of the EDF with *F*(*x*)

As the graph in Figure 3 indicates, $F(x)$ is an inadequate model of the EDF, clearly stemming from using a one parameter model when one is insufficient. SSE is minimized when a = 0.000209, but the convex nature of the function produces a poor fit. This introduces the learning point for students that there is a trade off between number of parameters and complexity of models.

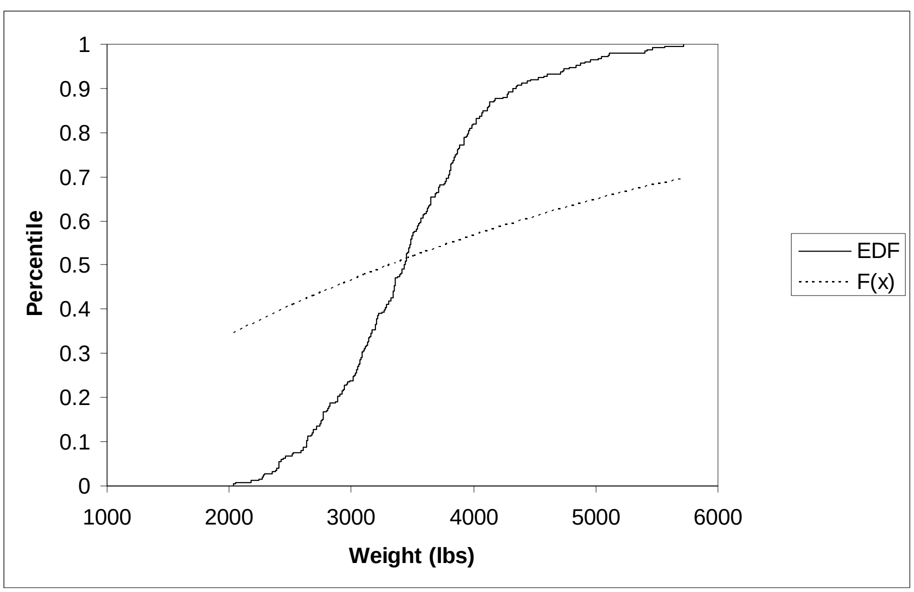


Figure 3: Modeling the EDF With *F*(*x*)

Once our students become comfortable with using $F(x)$, we then ask them to fit the more robust, two-parameter model, $G(x) = 1 - e^{-(ax)^b} \quad x, a, b \geq 0$, to the same EDF. When they find the parameters a and b that minimize the SSE between $G(x)$ and the EDF, they discover that $G(x)$ fits at least as well as $F(x)$ because when $b=1$, $G(x)$ is equivalent to $F(x)$. They further realize that the presence of the parameter b in the new model $G(x)$ greatly increases the flexibility (and hence the usefulness) of the fitted model. The two parameter model $G(x) = 1 - e^{-(ax)^b} \quad x, a, b \geq 0$, indeed, does model the classic S-shaped accumulation well as we can see in Figure 4. Using this model, a minimum SSE of 0.2197 (much lower than the previous SSE of 25.033) is achieved when a = 0.000274 and b = 6.145.

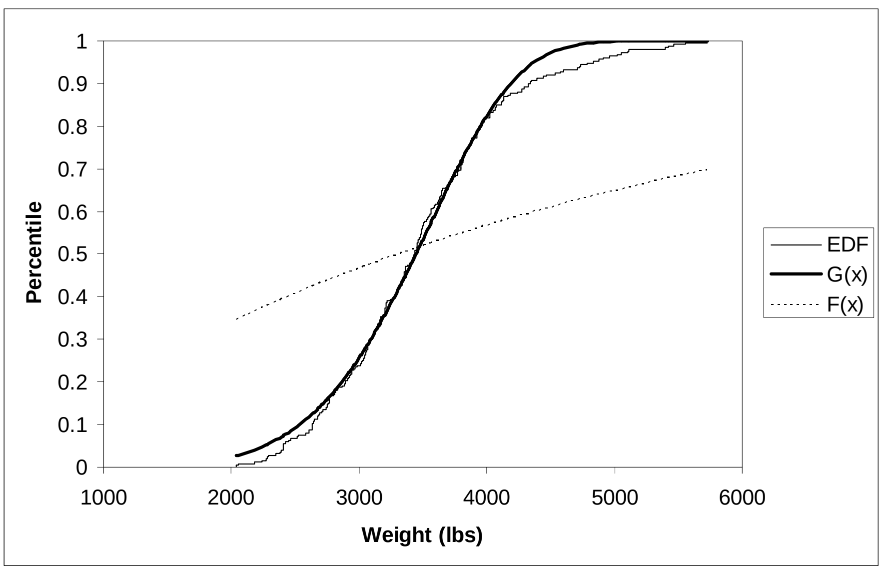


Figure 4: Modeling the EDF With *F*(*x*) and *G*(*x*)

After modeling EDFs, we next move on to exploring the derivative of the model. In this example, we choose to explore the derivative of $G(x)$ since it was the best model for this particular problem. Before we even attempt to compute the derivative of $G(x)$, we already recognize two properties that it must possess. First of all, since the EDF is a non-decreasing function and the best fit model $G(x)$ must likewise be non-decreasing, we know that the derivative $g(x) = G'(x)$ must always be a positive function. Furthermore, because the range of the EDF and therefore the model $G(x)$ is from 0 to 1, this tells us that the total area under the curve of the derivative $g(x)$ must be equal to exactly one. At this point in our course, we also begin to introduce some of the more conventional probability and statistics terminology by identifying the model $G(x)$ as the cumulative distribution function and its derivative $g(x)$ as the probability density function. We also reveal for the first time that the function $G(x)$ we have been studying for several lessons is a well-known continuous distribution called the Weibull CDF and the one parameter model is the exponential CDF. This marks the point at which we begin to transition from our data-oriented approach to our textbook's more common PDF-based approach to understanding continuous distributions.

The general form of the computed derivative of $G(x)$, or the PDF, is as follows: $g(x) = abe^{-(ax)^b}(ax)^{b-1}$. For this particular problem, substituting in values for the parameters a and b yields $g(x) = (0.000274)(6.145)e^{-(0.000274x)^{6.145}}(0.000274x)^{5.145}$. The graph of $g(x)$ is shown in Figure 5.

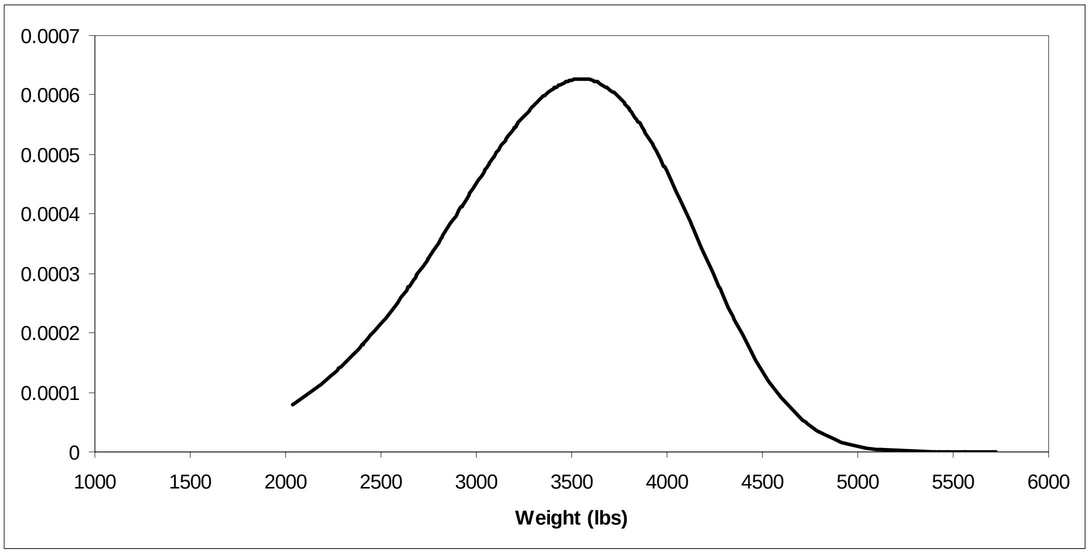


Figure 5: The Probability Density Function, *g*(*x*)

As expected, the graph verifies that $g(x)$ is always positive and a quick check will reveal that the total area under the curve of $g(x)$, represented by $\int_0^\infty g(x)dx$, is equal to one. Therefore, $g(x)$ is a legitimate PDF which we will use to illustrate a number of useful properties of continuous distributions.

3. NINE PROPERTIES OF CONTINUOUS DISTRIBUTIONS.

After modeling EDFs to help give students an understanding of a distribution function, our next task is to establish a framework to study some continuous and discrete probability distributions. In order to do so, we use the computer algebra system Mathematica and a methodology based on nine basic properties of continuous distributions. The purpose of learning probability theory with these nine properties is twofold. First, one sees that all distributions have these common elements that are derived from the same principles. Second, we develop the nine properties in such a fashion that the code is very portable. In fact, by changing only a few lines in the preamble of the ‘notebook’, different distribution ‘notebooks’ are created with their own nine properties. We call each of these computer files a ‘notebook’ as that is the naming convention Mathematica uses for its files. The nine properties themselves are very straightforward: verifying and plotting the PDF, deriving and plotting the CDF, calculating probabilities with PDFs and CDFs, calculating mean, variance and standard deviation, and calculating percentiles.

As an example of how we create these nine-property ‘notebooks’ we will use the model $G(x)$ from above to work on the Weibull distribution. Perhaps it would also be appropriate to name them ‘distribution templates.’ For the Weibull distribution, we begin by entering the low and high support, the parameter values and the PDF into a notebook.

```
lowsupport = 0;
highsupport = Infinity;
a = 0.000274;
b = 6.145;
g[x_] = a * b * e^-(a*x)^b * (a * x)^(b-1)
```

$$7.91588\times10^{-22}\, e^{-1.28818\times10^{-22}\, x^{6.145}}\, x^{5.145}$$

Figure 6: Defining the PDF for car weights

Now we begin building the nine properties into the notebook.

<u>Properties 1 and 2:</u> Verify that a function $g(x)$ is a legitimate PDF.

$$g(x) \geq 0 \qquad \forall x \tag{1}$$

$$\int_{-\infty}^{\infty} g(x)dx = 1 \tag{2}$$

As we enter these properties into Mathematica, we do so in such a way that our electronic worksheet can be reused in the future for other distributions without completely rewriting every line of code. For example, in Property 2 below, we use the already defined terms *lowsupport* and *highsupport* for our limits of integration rather than 0 and ∞. While the actual values of these variables will be different for every subsequent distribution, we will not have to change the references to the variables *lowsupport* and *highsupport*.

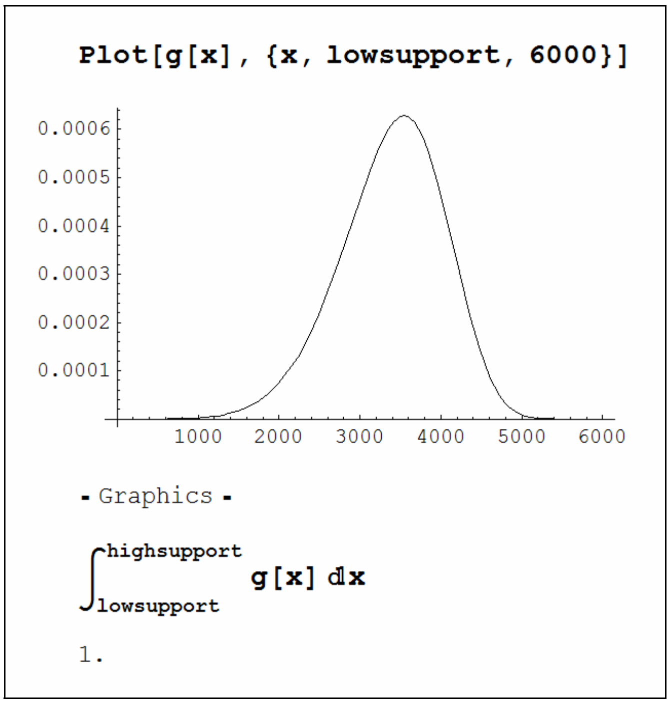


Figure 7: Verifying the legitimacy of our PDF for car weights

We proceed to cover the remaining seven properties, entering each of them into our electronic notebook as we discuss, analyze, and show examples of them in class.

<u>Property 3</u>: The probability assigned to any particular value of a continuous random variable is zero (this property does not require entry into Mathematica). The primary use of the property is to distinguish between discrete and continuous random variables later in the course.

$$P(X = c) = 0 \tag{3}$$

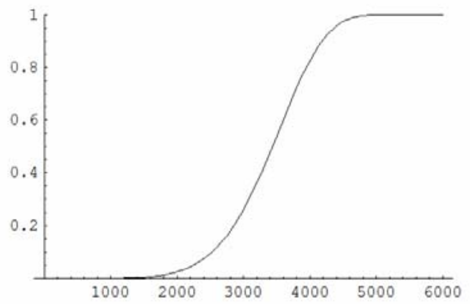
<u>Property 4</u>: Create the CDF using a continuous random variable's PDF.

$$G(x) = P(X \le x) = \int_{-\infty}^{x} g(y)dy \tag{4}$$

```
G[x_] = ∫_lowsupport^x g[y] dy
```

$1. - 1.\, e^{-1.28818\times10^{-22}\, x^{6.145}}$

```
Plot[G[x], {x, lowsupport, 6000}, PlotRange → {0, 1}]
```



- Graphics -

Figure 8: Creating the CDF for a car's weight

<u>Properties 5 and 6</u>: Calculate probabilities associated with continuous random variables using either the PDF or the CDF.

$$P(a \le X \le b) = \int_{a}^{b} g(x)dx \tag{5}$$

$$P(a \le X \le b) = G(b) - G(a) \tag{6}$$

```
a = 2000;
b = 3000;
PDFProb = ∫_a^b g[x] dx

0.234547

a = 2000;
b = 3000;
CDFProb = G[b] - G[a]

0.234547
```

Figure 9: Calculating the probability that a car weighs between 2000 and 3000 lbs

Properties 7 and 8: Calculate the expected value and variance of a continuous random variable.

$$\mu_x = E(X) = \int_{-\infty}^{\infty} x * g(x)dx \tag{7}$$

$$\sigma_x^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu_x)^2 * g(x)dx \tag{8}$$

$$\text{ExpValue} = \int_{\text{lowsupport}}^{\text{highsupport}} x * g[x]\, dx$$

```
3390.3
```

$$\text{Var} = \int_{\text{lowsupport}}^{\text{highsupport}} (x - \text{ExpValue})^2 * g[x]\, dx$$

```
412990.
```

Figure 10: Determining the mean and variance of a car's weight

Property 9: Find the (100*p*)th percentile of a continuous random variable.

$$\text{Set } p = G(x^*) \text{ and solve for } x^* \tag{9}$$

```
p = .90;
Solve[G[xstar] == p, xstar]
{{xstar → 4180.17}}
```

Figure 11: Finding the 90th percentile of a car's weight

Clearly, the strength of this approach lies not in the nine properties themselves, since they are simply a restatement of the material found in any text, but rather in the method in which we implement these nine properties. For instance, once our students have successfully completed their first notebook for the Weibull distribution, they can very quickly and easily modify this notebook for use with other continuous distributions such as the exponential, uniform, gamma, normal distributions, t, and chi-squared distributions. They return to the beginning of the notebook, enter the new PDF, add its associated parameters (if any), and input the high and low support values. Once that task is complete, a student can execute any or all of the nine properties at their discretion without changing the code for the nine properties at all. This approach makes precise measures of the once formidable distributions such as the gamma, chi-squared, normal, and t distributions attainable. Furthermore, as each of these distributions is introduced, we are able to discuss certain features of each, such as what the distribution best models, any special features of the moments, or perhaps the memoryless property of the exponential. Students are no longer constrained by charts and tables in the back of their textbooks, nor are they limited by pre-established degrees of freedom or book-imposed limits on significant digits when solving problems. For example, consider the normal distribution. With our approach, conducting a z-transformation and using the standard normal tables in order to calculate probabilities is no longer as necessary. Probabilities can be calculated from a normal distribution with any parameters.

Another benefit to teaching probability principles with these nine properties is evident in our third block of instruction, statistical inference. We give students the PDFs of the *t* and chi-square distribution and have them produce the nine-property notebooks for each distribution. When it is necessary to calculate a critical value for a confidence interval, the student uses property number nine to calculate the appropriate percentile from the appropriate distribution, *e.g.*, the *t* distribution with 31 degrees of freedom or the chi-square distribution with 44 degrees of freedom. When it is necessary to calculate a p-value for an hypothesis test, the student uses property 5 from the *t* distribution, calculating a probability with the PDF. All the statistical measures they need for this block are found from first principles with “software” that they created, as opposed to using some software package and/or set of tables that may or may not be intuitive to the novice statistician. They never use a table or chart and they are not required to interpolate or reduce degrees of freedom. We opine that the connection to first principles make the understanding of these statistical inference procedures more attainable the first time students encounter them. Our vision is not that statistical software will be eclipsed by these first-principle notebooks. In fact, we would predict (and prefer) that students taking a follow-on elective would move on to a professional statistical package and leave these notebooks behind. But in creating these notebooks, those students who rely on statistics but who are not as mathematically inclined as most statistics majors are will have a better understanding of what the software is doing by having created their own software at one time.

4. CONCLUSION.

This paper reports our experience of what has been done for the last three years at an institution that has been working diligently to find the appropriate use of technology in the classroom. We appeal, in effect, to the art of statistics instruction, not the science. Furthermore, this paper does not completely describe the entire course, nor does it describe all the ways that we use technology in the classroom. For instance our project work with Monte Carlo simulation and our statistical subjects have their own ‘technology’ implications, but we leave them out here, highlighting only what we view as the fundamental changes to probability and statistics education that the ubiquitous presence of laptops in the classroom can provide. Also, we have not addressed how assessment changes with technology in the classroom. (We allow open book, open notes, and open computers without access to wireless connections during our tests). These conditions make test-writing a challenge for our instructors. We also must point out that much of our course content is not in any single text, thus we must create our own supplemental readings (available on course web pages). Finally we have not addressed the faculty development issues that have arisen with our implementation. All the subjects are worthy of their own conversations. We have, however, found that this data-driven approach to understanding continuous distributions is highly effective for our students. As the presence of laptops becomes the norm, and we believe it will, all institutions will have to struggle to some degree with finding the most appropriate way they will use technology in the classroom to help learning. We intend for this summary of our experiences to be another contribution to the discussion among statistics professors on how best to incorporate technology in the classroom.

5. REFERENCES.

Link to Oswego University's econometrics data accessed on May 6, 2006.
http://www.oswego.edu/~kane/econometrics/data.htm