Some Class-Participation Demonstrations for Decision Theory and Bayesian Statistics

Andrew GELMAN

We present several classroom demonstrations that have sparked student involvement in our undergraduate course in decision theory and Bayesian statistics. Some of the demonstrations involve student participation, while others are essentially lectures with extra class discussion.

KEY WORDS: Calibration; Expected value; Instruction; Probability; Utility.

1. INTRODUCTION

We have taught several times an undergraduate or M.A.-level elective course in decision theory and Bayesian statistics. The course requires one term of probability as a prerequisite and typically attracts about 15 students. In addition to the usual lectures, homework, and problem solving, we have found it useful to conduct frequent classroom demonstrations. Some of these are essentially lectures, but with extensive class participation, whereas others involve actions or calculations by the students. This article outlines some of our more effective demonstrations. Our contribution here is in the tricks used to involve students (see Cobb [1992] and Gnanadeskikan et al. [1997] for some discussion of the benefits of class participation in statistics classes); the ideas behind the demonstrations are well known, and we refer instructors and students to textbooks in applied decision analysis and Bayesian statistics (e.g., Watson and Buede 1987; Smith 1988; Clemen 1996; Berry 1995; Gelman, Carlin, Stern, and Rubin 1995; and Carlin and Louis 1986) for further references.

Table 1 lists the demonstrations, the concepts they are intended to convey, and the additional materials they require. The demonstrations in Sections 2–6 may be of general interest to teachers of probability, whereas those in Sections 7–8 relate more specifically to Bayesian inference.

2. HOW MANY QUARTERS? INTRODUCTION TO THE PRINCIPLES OF DECISION ANALYSIS

This demonstration, which we do on the first day of class, needs to be preceded by the in-class demonstration of subjective probability intervals (see Sec. 4).

The current demonstration begins by showing a glass jar full of quarters to the class and letting the students pass it around and examine it. The jar has previously been filled to a specified level, and the instructor does not know how many quarters are in the jar—the answer has been written on a sheet of paper and is in a sealed envelope, which the instructor places on a table. The students are asked how many quarters they think are in the jar. A few students will state guesses, and we then encourage them to explain and discuss them (for example, "So, Ned, given that Louise guessed that the jar has 200 quarters, do you still want to guess 100?"). A student is then asked for a 50% subjective probability interval, so that the probability is 25% that the true value is below the low point of the interval and 25% that it is above the high point. The instructor and the class then prod the student (for example, "Your interval is $[125, 150]$. If I offered you an even-money bet, where you would win if the true number of quarters is between 125 and 150, and I would win if it is below 125 *or* above 150, would you take this bet?"). Once the student has settled on an interval, we ask if any students disagree with the interval. Someone will answer and give their interval, and the class is led in more discussion until they are brought to agreement on an interval that seems about right for everyone (for example, $[120, 200]$).

We then sketch a normal density on the blackboard and ask what the mean and standard deviation should be for the specified interval to contain 50% of the probability. The computation is easily done, using the fact that $[\mu - \frac{2}{3}\sigma, \mu + \frac{2}{3}\sigma]$ is an approximate 50% interval for the $\mathrm{N}(\mu, \sigma^2)$ distribution (the students are already familiar with the normal distribution from their probability prerequisite). The horizontal axis of the density is then labeled appropriately, and the students are asked if this seems to represent their uncertainty ("You are 90% sure the number of quarters is less than X?", and so on). We ask if their uncertainty can be expressed exactly by a normal distribution. Some students will realize the answer is no, because the true number of quarters must be (a) positive, and (b) an integer. We discuss how, with a distribution with mean 160 and standard deviation 60, for example, zero is far in the tail of the distribution, and the discreteness is a minor issue, so it can be reasonable to characterize the students' uncertainty by a normal distribution. At the end of the demonstration, we return to this issue and discuss why the results are basically valid for any unimodal distribution.

We then state the puzzle: you (the class) will be given *a single guess* as to the number of quarters in the jar. We will then open the envelope to reveal the true value. If the guess turns out to be correct, the money will be given to the class (yes, we would really do this) and split equally among the students. If the guess is incorrect, the students get nothing. What should you guess? In answering, assume that the

Andrew Gelman is Associate Professor, Department of Statistics, Columbia University, New York, NY 10027 (E-mail: gelman@stat.columbia.edu). We thank Tom Little, Deborah Nolan, Phillip Price, Seth Roberts, Andrea Siegel, Kimmy Szeto, the editor, and two referees. We also thank the National Science Foundation for grants DMS-9404305, SBR-9708424, and Young Investigator Award DMS-9457824.

Table 1. Concepts that are Intended to be Conveyed and Additional Materials Required to Conduct the Demonstrations

Demonstration	Concepts covered	Additional materials required
How many quarters?	expected values, optimization, subjective probability	Jar filled with money
Where are the cancers?	adjustment of data, shrinkage, sampling variability	Handouts of Figures 1 and 2
Subjective probability intervals	calibration, overconfidence	Handouts of Figure 3; list of uncertain quantities and their true values
Utility of money	coherence, risk-aversion, cognitive illusions	none
What is the value of a life?	utilities, calibration of low probabilities	none
Drawing parameters out of a hat	Bayesian inference (normal model), coverage of posterior intervals	Hat filled with draws from a normal distribution
Where are the cancers, a simulation	Bayesian inference (Poisson model), prior distributions, shrinkage	List of counties with populations; envelope filled with draws from a gamma distribution

normal distribution sketched on the board represents your true state of uncertainty about the number of quarters; thus, you believe that 160 (say) is the most likely value, 159 and 161 are the next most likely, there is only a 5% chance that it will be more than two standard deviations away, etc. Recall also that the jar was filled to a specified level—the number was not picked in advance. So you need not worry about psychological issues like, "He wouldn't have picked 150, because it's a round number." It's just a problem of geometry—how many quarters are in the jar—and the distribution on the blackboard represents your uncertainty.

After a short pause, a student speaks up and says they should guess 160, the mode of the distribution. Does everyone agree? Yes, everyone in the class agrees, although some are wary, suspecting a trick. We state that, in fact, the "obvious" answer is wrong—and there is no trick! Why is this? A pause to let the students think.

We pull a quarter out of our pocket and shake it between our two hands, then hold out both fists. One fist contains the quarter. A student is asked to pick a hand, and then we say, "This hand contains 0 or 1 quarters. There's a 50% chance that this hand holds a quarter, and if you guess right, you will get all the quarters in the hand. Should you guess 0 or 1?" The students start to realize—if you guess 0 and you're right, you don't win anything anyway, so you might as well guess 1. "Suppose there's only a 10% chance that there's a quarter—what should you guess?" You should still guess 1—it can't hurt.

Now to a more complicated problem. Suppose the jar contains either 100 or 200 quarters, and you think the two possibilities are equally likely. Should you guess 100 or 200? What if there is a 51% chance of 100 quarters and a 49% chance of 200? Which should you guess? At this point, some students will choose 100 and some will choose 200. Which choice is better? We bring two students to the blackboard—suppose Ned would guess 100 and Louise would guess 200—and play the guessing game repeatedly. At each play, we choose a random number from 00 to 99 by rolling dice (see Appendix); if the outcome is in the range $[00, 48]$, we give Louise \$200, and if it is in the range $[49, 99]$, we give Ned \$100 (Monopoly money in both cases). After playing ten or so times, it becomes clear that Louise is doing better than Ned. We derive this on the blackboard by showing Ned's and Louise's expected value per play and recalling the law of large numbers.

So Louise's strategy is better. But, a student asks, we are only playing the quarters game once, not playing repeatedly, so why are expected values and the law of large numbers relevant? Well, life is full of uncertainties—in a given week, you may buy insurance, bet on a football game, make a guess on an exam question, and so forth. As you add up the uncertainties in the events, the law of large numbers comes into play, and the expected value determines your long-run gain (just like Ned and Louise). As long as no single decision or small set of decisions are dominant (this condition can be made more precise in a more theoretical course on probability), you can go with the expected value. (We are ignoring nonmonetary gains such as the thrill of getting the correct guess.)

Now back to the quarters. Should you guess 160, or something higher, or something lower? Yes, something higher. Let's work it out mathematically. Your goal is to maximize your expected return. Let x be your guess; then your expected gain is just x times the probability that the number of quarters is x. For our distribution with mean μ and standard deviation σ, that probability is approximately the normal density at x, so the expected gain is approximately $x\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$. To find the maximum of the expected gain, differentiate with respect to x and set the derivative to zero; after some cancellation this yields a quadratic equation, with solution $x = \frac{1}{2}\left(\mu \pm \sqrt{\mu^2 + 4\sigma^2}\right)$. The answer cannot be negative, so the $\pm$ must be $+$. We plug in μ and σ to compute an answer, rounding to the nearest integer (for example, in the above example, with $\mu = 160$ and $\sigma = 60$, the optimal guess is $x = 180$. Are the students happy with this guess? It is useful to tell the story of the motorist who is stranded at night and is looking for his keys, not by his car (the most probable location), but near the street lamp (less probable, but more likely that he will find his keys, if that is where they are). We then open the envelope and find the true answer, paying out in the unlikely case that the guess is exactly correct.

One nice thing about this demonstration is that the answer can be computed exactly, but it requires some nontrivial analysis (differentiation and the quadratic formula). This is probably the first problem they have ever seen in which the exact formula for the normal density is useful. We conclude by noting that their expected gain (and thus the instructor's expected monetary loss from doing the demonstration) equals the value of x times the probability that x is the correct guess. Computing this for the chosen x and the approximate normal distribution shows the instructor's expected loss to be reassuringly small (if $\mu = 160$ and $\sigma = 60$, an expected loss of 28.3 cents).

3. WHERE ARE THE CANCERS? INTRODUCTION TO BAYESIAN STATISTICS

We do this demonstration also on the first day of class (or on the second day if we run out of time). We begin by passing out copies of Figure 1, reproduced from Manton et al. (1989); this is a map of the United States, shading the counties with the highest rates of kidney cancer from 1970–1979. We ask the students what they notice about the map; one of them points out the most obvious pattern, which is that many of the counties in the Great Plains states but relatively few near the coasts are shaded. Why is this? A student asks whether these are the counties with more old people. That could be the answer, but it is not—in fact, these rates are age-adjusted. We point out that many people retire to Arizona and Florida but they do not have many shaded counties on this map of age-adjusted cancer death rates. Any other ideas? A student notes that most of the shaded counties are in rural areas; perhaps the health care there is worse than in major cities. Maybe so, but what if we told you that if we were to highlight the counties with the *lowest* age-adjusted kidney cancer death rates, we would still be mostly highlighting rural areas?

At this point, the students are usually stumped. To help them, we consider a county with 1,000 people: if it has even one kidney cancer death in the 1970s, its rate will be one per thousand, which is among the highest in the nation. Of course, if it has no kidney cancer deaths, its rate will be lowest in the nation (tied with all the other counties with zero deaths). The observed rates for smaller counties are much more variable, and hence they are much more likely to be shaded, even if the true probability of cancer in these counties is nothing special. A small county has an observed rate of one per thousand, this is probably random fluctuation, but if a large county such as Los Angeles has a very high rate, it is probably a real phenomenon. We now hand out copies of Figure 2 (also from Manton et al. 1989), a map that shades the counties with the highest *adjusted* kidney cancer death rates, where the adjusted rate for each county is a weighted average of (a) the observed rate in the county and (b) the national average rate. In this weighted average, the relative weight attached to the observed rate is approximately proportional to the population of the county, so that in counties with extremely small population the rate is shrunk virtually all the way to the national average, in counties with moderate population the rate is shrunk part way toward the national average, and in very large counties the adjusted rate is essentially equivalent to the observed rate. Figure 2 looks much different from Figure 1, with much more of the shading appearing in populous counties. Most of the extreme rates in Figure 1 occurred in low-population counties, and they got shrunk so much toward the national average that they were no longer extreme in Figure 2. (In fact, the shading in Figure 2 overemphasizes the high-population counties—see Gelman and Price [in press] for a discussion of problems of mapping adjusted rates—but this is not an issue we discuss in our introductory class.)

This example is a nice introduction to applied statistics because it shows a case where statistical adjustment is clearly appropriate and is, in fact, a standard tool in epidemiology (see Clayton and Bernardinelli 1992), but we can explain it without worrying about prior distributions or probability theory. We return to this example later in the class to illustrate Bayesian inference (see Sec. 8.).

4. SUBJECTIVE PROBABILITY INTERVALS AND CALIBRATION

A well-known, and very useful, demonstration involves the calibration of probability intervals. We start by asking a student to give his or her guess at some uncertain quantity (for example, the number of master's degrees conferred in the United States in 1990). The student is then asked to give a 90% probability interval (more precisely, the 5% and 95% quantiles) for the uncertain quantity, at which point we ask the other students to comment. If a student in the class would place more or less than 90% probability on the stated interval, we point out the opportunity for a bet that both parties should accept. We then give each student a minute to write down his or her 90% interval, then we write the intervals on the board. It is typically the case that the intervals are relatively short with little overlap, so that it is in fact *impossible* for 90% of them to contain the true value, whatever it happens to be. Thus, substantial differences of opinion have been revealed. The students are now given the opportunity to adjust their intervals, and then the true value is revealed. It becomes clear from the discussion that most students, when told the guesses of the others, tend to widen their uncertainty intervals—that is, they recognize their original intervals to be overconfident. The students may be reassured to see the well-known example displayed Figure 3, reproduced from Hynes and Vanmarcke (1977), in which a set of internationally known geotechnical engineers show overconfidence (in retrospect) in their probabilistic forecasts of the failure height of an embankment.

Later in the term, we follow up with a written exercise in which each student is given a list of several unknown quantities and asked to write down 50% and 90% intervals for

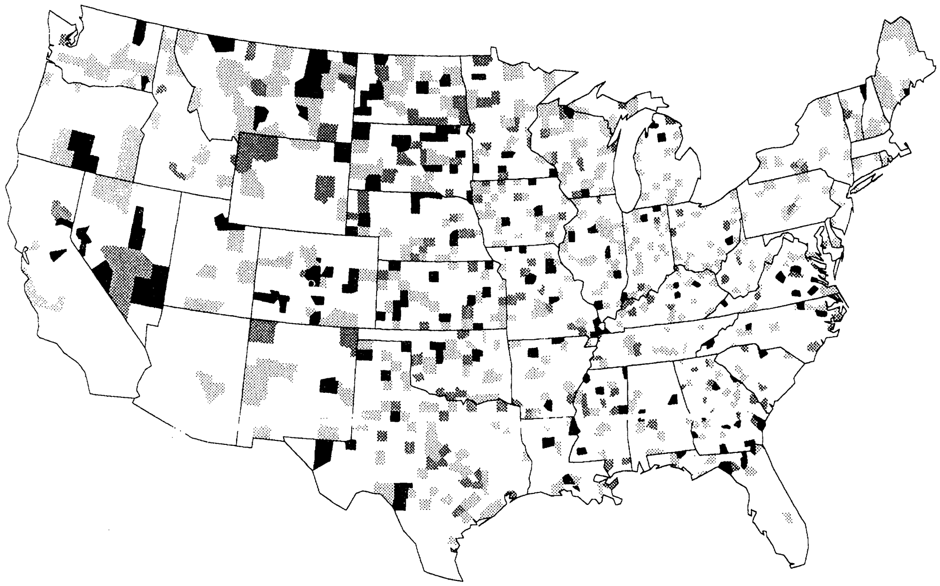

Figure 1. Counties of the U.S., with Three Levels of Shading Indicating the Counties with the Highest 25%, 10%, and 5% Age-Standardized Death Rates for Cancer of Kidney/Ureter for U.S. White Males, 1970—1979. From Manton et al. (1989).

each. They will still tend to be overconfident (even though we warned them about this in the class discussion), in that, for most students, fewer than 50% of their 50% intervals and fewer than 90% of their 90% intervals will contain the true value (see Alpert and Raiffa [1984] for an involved discussion of this phenomenon). To keep students interested, we include questions about themselves (for example, the average weight of the students in the class) and topics in education (for example, the number of master's degrees), as well as quantities about which the students know very little (for example, the total number of eggs produced in the U.S. in 1965, a question from Alpert and Raiffa [1984]). We perform the discussion before the written exercise because we have found that many students do not understand the concept of a probabilistic forecast until we have discussed it in class together.

5. UTILITY OF MONEY AND RISK-AVERSION

To introduce the concept of utility, we ask each student to write on a sheet of paper the probability p_1 for which they are indifferent between (a) a certain gain of \$1, and (b) a gain of \$1,000,000 with probability p_1 or \$0 with probability $(1 - p_1)$. (For brevity, we write this as $\$1 \equiv p_1\$1{,}000{,}000 + (1-p_1)\$0$. We use $\equiv$ rather than $=$ because utility for money is not necessarily linear.) The students are then asked to write down, in sequence, the probabilities p_2, p_3, p_4, for which $\$1 \equiv p_2\$10 + (1-p_2)\$0$; $\$10 \equiv p_3\$100+$

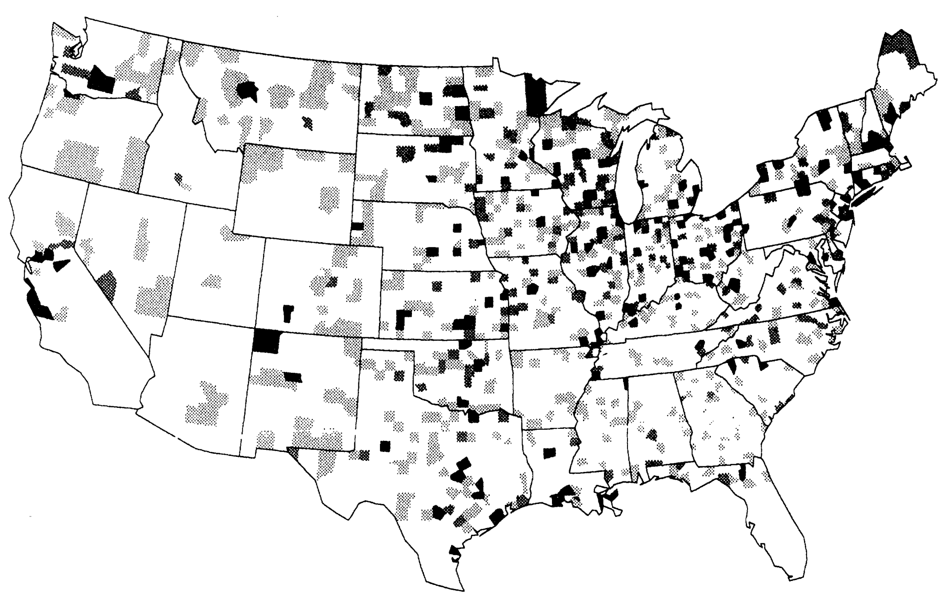

Figure 2. Counties of the U.S., with Three Levels of Shading Indicating the Counties with the Highest 25%, 10%, and 5% Adjusted *Age-Standardized Death Rates for Cancer of Kidney/Ureter for U.S. White Males, 1970—1979. From Manton et al. (1989).*

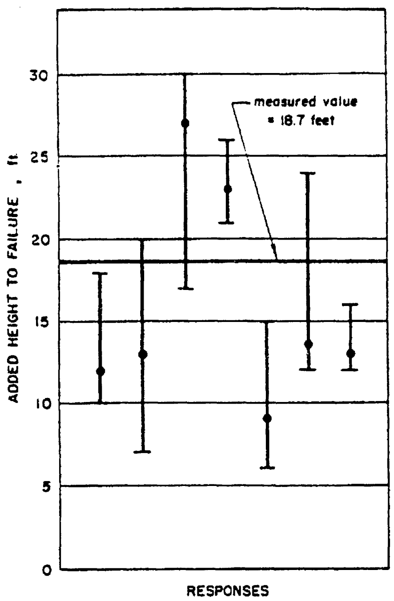

Figure 3. Experts' Predictions and 50% Predictive Intervals of the Height at Which an Embankment Would Fail, Along with the True Value. Note that none of the predictions included the true value. From Hynes and Vanmarcke (1977).

$(1-p_3)\$1$; $\$100 \equiv p_4\$1000 + (1-p_4)\$10$; and $\$1000 \equiv p_5\$1{,}000{,}000 + (1-p_5)\$100$. One of the students is then brought to the blackboard to give his or her answers to the questions. The probabilities are checked for coherence (that is, the existence of a consistent set of utilities and preferences), as follows. First, the answers to the questions involving p_2 and p_3 are combined to yield a comparison between \$1, \$100, and \$0. For example, suppose $p_2 = .1$ and $p_3 = .15$ (when working through the example in class, using the student's actual numbers is clearer than working with the algebra of p_1, p_2, p_3, and so on). Then,

$$\begin{aligned} \$1 &\equiv .1\,\$10 + .9\,\$0 \\ &\equiv .1(.15\,\$100 + .85\,\$1) + .9\,\$0 \\ &\equiv .015\,\$100 + .085\,\$1 + .9\,\$0 \\ &\equiv .085\,\$1 + .915\left(\frac{.015}{.915}\,\$100 + \frac{.9}{.915}\,\$0\right). \qquad (1) \end{aligned}$$

This holds if and only if $\left(\frac{.015}{.915}\,\$100 + \frac{.9}{.915}\,\$0\right) \equiv \$1$. (Note: in class we work this and subsequent utility computations out using decision trees, as illustrated in Figure 4, rather than equations.) Given this student's answers to the questions, we have deduced that $\$1 \equiv .0164\,\$100 + .9836\,\$0$. We then repeat this procedure, using the student's value of p_4, to determine the utility of \$1 relative to \$1,000 and \$0, and then once again, using p_5, to determine the utility of \$1 relative to \$1,000,000 and \$0. Finally, this derived value is compared to the student's original value of p_1. These will disagree, meaning that the student's preferences are incoherent. The students in the class then discuss with the student at the blackboard how to change $p_1, \ldots, p_5$ to give coherent and reasonable answers. It may be necessary to remind the students that coherence does *not* require the utility for money to be linear. The student at the blackboard then is asked to sketch his or her utility function for money, as implied by the equivalence statements above.

A related demonstration goes as follows. A person is somewhat risk-averse and is indifferent between (a) a certain gain of \$10 and (b) a 55% chance of \$20 and a 45% chance of \$0. Similarly, he or she is indifferent between (a) a certain gain of \$20 and (b) a 55% chance of \$30 and a 45% chance of \$10; and, in general, indifferent between (a) a certain gain of $\$x$ and (b) a 55% chance of $\$(x+10)$ and a 45% chance of $\$(x-10)$; for $x = 30, 40, 50, \ldots$.

Is this reasonable? The students assent.

[Conversely, this can be done as more of a "set-up" by picking a student and asking for what value p is he or she neutral between the alternatives (a) \$10 and (b) a p probability of \$20 and a $(1-p)$ probability of \$0; then repeating this question for $\$20 \equiv p\$30 + (1-p)\$10$, $\$30 \equiv p\$40 + (1-p)\$20$; etc. The student will probably give a value of p that is near or greater than .55.]

Then answer the following question: For what dollar value y is this person indifferent between (a) a certain gain of $\$y$, and (b) a 50% chance of \$1 billion and a 50% chance of \$0? The answer, surprisingly, is that $\$y$ is between \$30 and \$40, as can be derived easily by mathematical induction. For example, using utility notation, the given indifferences can be written as $U(\$x) = .55U(\$(x+10)) + .45U(\$(x-10))$ for each x, and thus $U(\$(x+10)) - U(\$x) = \frac{.45}{.55}\,(U(\$x) - U(\$(x-10)))$. Setting $U(\$0) = 0$ and $U(\$10) = 1$ (the location and scale of the utility function can be set arbitrarily) and evaluating

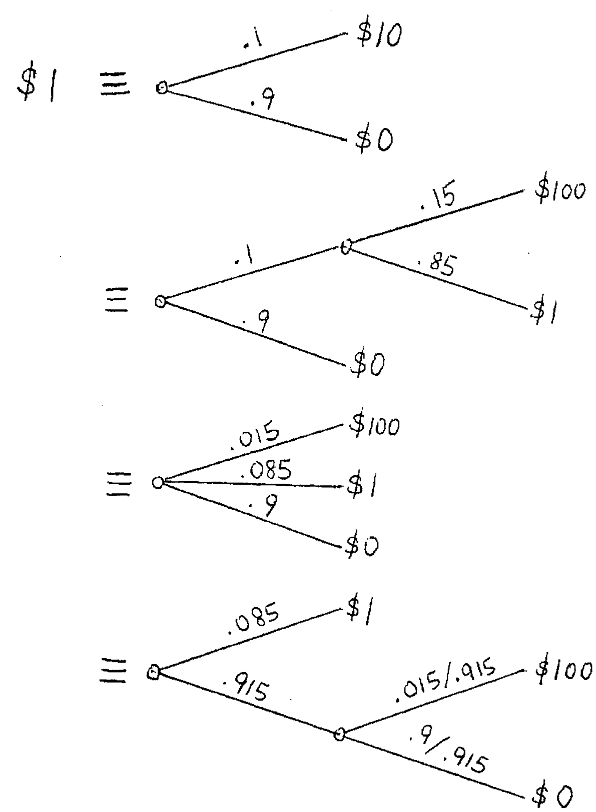

Figure 4. Illustration of a Derivation of Relative Utilities Using Decision Trees, Replicating the Steps of Equation (1).

the expressions in order yields $U(\$20) = 1 + \frac{.45}{.55} = 1.818$, $U(\$30) = 1 + \frac{.45}{.55} + \left(\frac{.45}{.55}\right)^2 = 2.487, \ldots, U(\$1 \text{ billion}) = 1 + \frac{.45}{.55} + \left(\frac{.45}{.55}\right)^2 + \left(\frac{.45}{.55}\right)^3 + \cdots + \left(\frac{.45}{.55}\right)^{99{,}999{,}999} \approx \frac{.55}{.55-.45} = 5.5$. Since $U(\$30) < .5U(\$1 \text{ billion}) + .5U(\$0)$, $U(\$40) > .5U(\$1 \text{ billion}) + .5U(\$0)$, and utility is an increasing function of money, $\$y$ must be between \$30 and \$40.

The student believes each step of the argument but is unhappy with the conclusion. Where is the mistake? It is that fearing uncertainty is not necessarily the same as "risk aversion" in utility theory: the latter can be expressed as a concave utility function for money, whereas the former implies behavior that is not consistent with any utility function (see Kahneman and Tversky 1979). This is a good time to discuss cognitive illusions, many of which have been demonstrated in the context of monetary gains and losses. (See Kahneman, Slovic, and Tversky [1984] and Thaler [1992] for much more on this topic, including many experiments that can be performed in the classroom.) Is decision theory descriptive? Is it normatively appropriate?

6. WHAT IS THE VALUE OF A LIFE?

We begin by asking the students what is the dollar value of their lives—how much money would they accept in exchange for being killed? They generally answer that they would not be killed for any amount of money. Now flip it around: suppose you have the choice of (a) your current situation, or (b) a probability p of dying and a probability $(1-p)$ of gaining \$1. For what value of p are you indifferent between (a) and (b)? Many students will answer that there is no value of p; they always prefer (a). What about $p = 10^{-12}$? If they still prefer (a), let them consider the following example.

To get a more precise value for p, it may be useful to consider a gain of \$1,000 instead of \$1 in the above decision. To see that \$1,000 is worth a nonnegligible fraction of a life, consider that people will not necessarily spend that much for air bags for their cars. Suppose a car will last for 10 years; the probability of dying in a car crash in that time is of the order of $10 \cdot 40{,}000/250{,}000{,}000$ (the number of car crash deaths in ten years divided by the U.S. population), and if an air bag has a 50% chance of saving your life in such a crash, this gives a probability of about 8×10^{-4} that the bag sill save your life. Once you have modified this calculation to your satisfaction (e.g., if you do not drive drunk, the probability of a crash should be adjusted downward) and determined how much you would pay for an air bag, you can put money and your life on a common utility scale. At this point, you can work your way down to the value of \$1 (as in the demonstration in Sec. 5). This can all be done using a student volunteer at the blackboard and the other students making comments and checking for coherence.

The student discussions can be enlightening. For example, one student, Julie, was highly risk-averse: when given the choice between (a) the current situation, and (b) a .00001 probability of dying and a .99999 of gaining \$10,000, she preferred (a). Another student in the class pointed out that .00001 is approximately the probability of dying in a car crash in any given three-week period. After correcting for the fact that Julie does not drive drunk, and that she drives less than the average American, perhaps this is her probability of dying in a car crash, with herself as a driver, in the next six months. By driving, she is accepting this risk; is the convenience of being able to drive for six months worth \$10,000 to her? This demonstration is especially interesting to students because it shows that they really do put money and lives on a common scale, whether they like it or not. There is a vast literature on the practical, political, and moral issues involved in equating dollars and lives; see, for example, Rhoads (1980) and Dorman (1996).

7. DRAWING PARAMETERS OUT OF A HAT

When introducing Bayesian statistics, a vivid way to illustrate population (or prior) and data distributions is by physical sampling. The demonstration goes as follows. Two students out of the class are picked to be "statisticians" and are taken out of the room. Each of the remaining students then draws a slip of paper out of a hat, which, before the lecture, the instructor filled with random samples from a $\mathrm{N}(100, 15^2)$ distribution. This slip of paper represents θ_j, the "true IQ" of student j. (IQ test scores are scaled so that their distribution is approximately normal.) Each of these students rolls a die several times to create a random $\mathrm{N}(0, 10^2)$ random variable (see Appendix) to represent "measurement error," and then adds it to his or her true IQ to obtain a "measured IQ," y_j.

The two "statisticians" are then brought back into the room. They are told the population distribution of "true IQ's," the distribution of measurement error, and the "measured IQ's" y_j, and are asked to estimate the "true IQ" θ_j for each student in the class and to supply 90% posterior intervals. The length and coverage of these intervals are compared to the classical 90% intervals obtained from the "measurements" alone. We find that both sorts of intervals have the correct coverage properties (on average), but the Bayesian intervals are shorter, which makes sense because the Bayesian intervals make use of the known population distribution.

This example can be stretched out further by discussions of the prior distribution, the likelihood, and so forth. The "IQ" context is a nice hook to get students involved.

8. WHERE ARE THE CANCERS? A SIMULATION

Near the end of the course, we work through the basics of Bayesian inference, including results for normal, binomial, and Poisson models. Where possible, we use prior distributions that correspond to actual populations, thus treating all Bayesian models as implicitly hierarchical (that is, with a prior distribution that represents the distribution of an actual population of parameters). As an example, we adapt the "drawing parameters out of a hat" demonstration to the kidney cancer mortality rates in U.S. counties (see Sec. 3). To do this requires first setting up a probability model for the parameters θ_j (the underlying ten-year kidney cancer death rates in U.S. counties j) and the data y_j (the observed number of deaths out of a population n_j in each county j). We assume a Poisson distribution for each y_j with mean $n_j\theta_j$ (ignoring the age-adjustment) and a conjugate gamma population distribution for the θ_j's, with hyperparameters set by matching moments (itself an interesting discussion topic), which comes out to $\text{Gamma}(20, 20/(4.65 \times 10^{-5}))$ (see Manton et al. 1989 and Gelman and Price in press). A student in the class is asked to compute the mean and standard deviation of this gamma distribution and to sketch its density function on the blackboard.

The demonstration now begins. Two students out of the class are picked to be "public health officials" and are taken out of the room. Each remaining student in the class is assigned a county j (identified by its name and population), taken from a list chosen to include a wide range of populations, ranging from about a thousand to over seven million (Los Angeles), and also, to keep the students' interest, counties that are well known (e.g., New York), or with amusing names (e.g., Jim Hogg County, TX). Each student then selects a true (underlying) kidney cancer rate θ_j for his or her county by drawing from an envelope that I had previously stuffed with random simulations from the $\text{Gamma}(20, 20/(4.65\times 10^{-5}))$ distribution. The physical sampling brings an immediacy to the meaning of the prior distribution in a hierarchical model.

Each student then multiplies the county population n_j by the underlying rate θ_j to get an expected number of kidney cancer deaths in a ten-year period. The student then draws a random number, y_j, from the Poisson distribution with this mean (see Appendix).

All this is written on the blackboard. Then we erase the true rates and the expected rates, leaving only the county names, populations, "observed" deaths y_j, and "observed" rates y_j/n_j. The "public health officials" are then brought back into the room and asked to estimate the ranking of the counties in order of the true death rates θ_j. The highest and lowest observed rates, of course, will tend to be in low-population counties.

The class is then led through the Bayesian analysis, which yields the posterior mean and standard deviation of θ_j, conditional on y_j, for each county in the table. We conclude the demonstration by writing the true values of θ_j back on the blackboard, checking the confidence interval coverage, and comparing the underlying, observed, and posterior mean death rates.

9. CONCLUSION

In-class demonstrations serve several purposes, including (1) focusing student attention on difficult conceptual issues that are hard to learn in a lecture or by solving homework problems (for example, the principle of expected gain in the quarters example; determining the value of a life); (2) alerting students to their cognitive illusions and that they are shared with others (for example, the uncalibrated subjective probability intervals and the incoherent utilities for money); (3) bringing personal issues into the class, thus allowing each student to make a personal contribution to the discussion (for example, different areas knowledge in the subjective probability intervals and different preferences regarding the value of a life); (4) dramatizing counterintuitive results which a student might not realize as counterintuitive if he or she were not forced to guess out loud (the quarters and the cancer maps); and (5) demonstrating the multiple levels of uncertainty in a Bayesian analysis, as well as the coverage property of posterior intervals (drawing parameters out of a hat and the cancer rate simulations). In addition, eliciting discussion in these demonstrations has been useful in introducing the students to the instructor and each other and has led to a high level of student participation.

APPENDIX: RANDOM NUMBERS VIA DICE

At the beginning of the term, each student in the class is given a 20-sided die on which each of the digits from 0 to 9 are written twice. (These dice can be bought in a game store for about 40 cents each.) Rolling the die once gives a random digit. We ask the students how to create a random variable with an approximate $\text{N}(\mu, \sigma^2)$ distribution, using five rolls of the die. After some discussion, they can derive that the sum of five independent random digits has mean 22.5 and standard deviation $\sqrt{41.25} = 6.42$, and we inform them that the distribution is close to normal. (This can be demonstrated by asking each student in the class to roll five dice, and then displaying a histogram of the students' totals.) Thus, the sum of five random digits, minus 22.5, times $\sigma/6.42$, plus μ, has the desired distribution. (The students are required to bring calculators to class.)

A more difficult problem is creating a random sample from a $\text{Poisson}(\lambda)$ distribution. We break this into two tasks: large and small λ. For large λ, we can use the normal approximation, drawing from the $\text{N}(\lambda, \lambda)$ distribution and rounding to the nearest integer. We ask the class: how big must λ be for this to work? Well, at the very least, we do not want to be drawing negative numbers, which suggests that the mean of the distribution should be at least two standard deviations away from zero. Thus, $\lambda > 2\sqrt{\lambda}$, so $\lambda > 4$. For smaller λ, we can compute the distribution function directly, using the formula for the Poisson density function. We can round the density function to two decimal places and then simulate using two random digits obtained by rolling a die twice.

We tell the students that, in practice, more efficient and exact simulation approaches exist; the methods here are useful for developing students' intuitions about distributions, means, and variances.

[Received April 1996. Revised November 1996.]

REFERENCES

Alpert, M., and Raiffa, H. (1984), "A Progress Report on the Training of Probability Assessors," in *Judgment Under Uncertainty: Heuristics and Biases*, eds. Kahneman, D., Slovic, P., and Tversky, A., New York: Cambridge University Press, pp. 294–305.

Berry, D. (1995), *Statistics: A Bayesian Perspective*, Belmont, CA: Duxbury Press.

Carlin, B., and Louis, T. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall.

Clayton, D., and Bernardinelli, L. (1992), "Bayesian Methods for Mapping Disease Risk," in *Geographical and Environmental Epidemiology:*

Methods for Small-Area Studies, eds. P. Elliott, J. Cusick, D. English, and R. Stern, Oxford: Oxford University Press, pp. 205–220.

Clemen, R. T. (1996), *Making Hard Decisions* (2nd ed.), Belmont, CA: Duxbury Press.

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L.A. Steen, Mathematical Association of America, pp. 3–34.

Dorman, P. (1996), *Markets and Mortality: Economics, Dangerous Work, and Value of Human Life*, Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Gelman A., and Price, P. N. (in press), "All Maps of Parameter Estimates are Misleading," *Statistics in Medicine*.

Gnanadesikan, M., Scheaffer, R.L., Watkins, A.E., and Witmer, J.A. (1997), "An Activity-Based Statistics Course," *Journal of Statistics Education*, 5.

Hynes, M. E., and Vanmarcke, E. (1977), "Reliability of Embankment Performance Predictions," *Mechanics in Engineering*, University of Waterloo Press, pp. 367–384.

Kahneman, D., Slovic, P., and Tversky, A. (1984), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Kahneman, D., and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–291.

Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P., and Pellom, A. C. (1989), "Empirical Bayes Procedures for Stabilizing Maps of U.S. Cancer Mortality Rates," *Journal of the American Statistical Association*, 84, 637–650.

Rhoads, S.E. (1980), *Valuing Life: Public Policy Dilemmas*, Boulder, CO: Westview Press.

Smith, J. Q. (1988), *Decision Analysis: A Bayesian Approach*, London: Chapman & Hall.

Thaler, R. H. (1992), *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, New York: Free Press.

Watson, S. R., and Buede, D. M. (1987), *Decision Synthesis: the Principles and Practice of Decision Analysis*, New York: Cambridge University Press.