The bootstrap achieves that goal very simply. Since it provides us with all these "incarnations" of $\widehat{\theta}$, i.e. $\widetilde{\theta}_1, ..., \widetilde{\theta}_k$, we can in principle estimate any apsect of the distribution of $\widehat{\theta}$ that we wish.

For example, suppose we are in a situation covered by the by the delta method, so that $\widehat{\theta}$ is approximately normally distributed. Instead of calculating messy derivatives, we could get a standard error for $\widehat{\theta}$ as the sample standard deviation of $\widetilde{\theta}_1, ..., \widetilde{\theta}_k$.

Efron's percentile method is more general, though unfortunately not so intuitive.

8.8 Bayesian Methods

Everyone is entitled to his own opinion, but not his own facts—Daniel Patrick Moynihan, senator from New York, 1976-2000

Whiskey's for drinkin' and water's for fightin' over—Mark Twain, on California water jurisdiction battles

Black cat, white cat, it doesn't matter as long as it catches mice—Deng Xiaoping, when asked about his plans to give private industry a greater role in China's economy

The most controversial topic in statistics by far is that of **Bayesian** methods. In fact, it is so controversial that a strident Bayesian colleague of mine even took issue with my calling it "controversial"!

The name stems from Bayes' Rule (Section 2.6),

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{not } A)P(B|\text{not } A)} \tag{8.126}$$

No one questions the validity of Bayes' Rule, and thus there is no controversy regarding statistical procedures that make use of probability calculations based on that rule. But the key word is *probability*. As long as the various terms in (8.126) are real probabilities, there is no controversy. But instead, the debate stems from the cases in which Bayesians replace some of the probabilities in the theorem with "feelings," i.e. NON-probabilities, arising from what they call **subjective prior distributions**. The key word is then *subjective*. Our section here will concern the controversy over the use of subjective priors.

Say we wish to estimate a population mean. Here the Bayesian analyst, before even collecting data, says, "Well, I think the population mean could be 1.2, with probability, oh, let's say 0.28, but on the other hand, it might also be 0.88, with probability, well, I'll put it at 0.49..." etc. This is the analyst's subjective prior distribution for the population mean. The analyst does this before even collecting any data. Note carefully that he is NOT claiming these are real probabilities; he's just trying to quantify his hunches. The analyst then collects the data, and uses some mathematical procedure that combines these "feelings" with the actual data, and which then outputs an estimate of the population mean or other quantity of interest.

The Bayesians justify this by saying one should use all available information, even if it is just a hunch. "The analyst is typically an expert in the field under study. You wouldn't want to throw away his/her expertise, would you?"

The non-Bayesians, known as **frequentists**, on the other hand dismiss this as unscientific and lacking in impartiality. "In research on a controversial health issue, say, you wouldn't want the researcher to incorporate his/her personal political biases into the number crunching, would you?"

8.8.1 How It Works

To introduce the idea, consider again the example of estimating p, the probability of heads for a certain penny. Suppose we were to say—before tossing the penny even once—"I think p could be any number, but more likely near 0.5, something like a normal distribution with mean 0.5 and standard deviation, oh, let's say 0.1." The prior distribution is then $N(0.5, 0.1^2)$. But again, note that the Bayesians do not consider it to be a distribution in the sense of probability. We are just using our "gut feeling" here, our "hunch." This is an absolutely central point.

So, Bayesians would not regard p as random here. They would simply be using the normal "distribution" for p to describe a degree of belief, rather than a probability distribution. (I will continue to use quotation marks below for this reason.)

Nevertheless, in terms of the mathematics involved, it's as if the Bayesians are treating p as random, with p's distribution being whatever the analyst specifies as the prior. Under this "random p" assumption, the Maximum Likelihood Estimate (MLE), for instance, would change. Our data here is X, the number of heads we get from n tosses of the penny. In contrast to the frequentist approach, in which the likelihood would be

$$L = \left(\begin{array}{c} n \\ X \end{array} \right) p^X (1-p)^{n-X} \tag{8.127}$$

it now becomes

$$L = \frac{1}{\sqrt{2\pi}\, 0.1} \, \exp -0.5[(p-0.5)/0.1]^2 \left(\begin{array}{c} n \\ X \end{array} \right) p^X (1-p)^{n-X} \tag{8.128}$$

This is basically P(A and B) = P(A) P(B|A), though using a density rather than probability mass functions. We would then find the value of p which maximizes L, and take that as our estimate.

Note how this procedure achieves a kind of balance between what our hunch says and what our data say. In (8.128), suppose the mean of p is 0.5 but n = 20 and X = 12 Then the frequentist estimator would be X/n = 0.6, while the Bayes estimator would be about 0.56. (Computation not shown here.) So our Bayesian approach "pulled" our estimate away from the frequentist estimate, toward our hunch that p is at or very near 0.5. This pulling effect would be stronger for smaller n or for a smaller standard deviation of the prior "distribution."

A Bayesian would use Bayes' Rule to compute the "distribution" of p given X, called the **posterior distribution**. The analog of (8.126) would be (8.128) divided by the integral of (8.128) as p ranges from 0 to 1, with the resulting quotient then being treated as a density. The MLE would then be the **mode**, i.e. the point of maximal density of the posterior distribution.

But we could use any measure of central tendency, and in fact typically the mean is used, rather than the mode. In other words:

> To estimate a population value θ, the Bayesian constructs a prior "distribution" for θ (again, the quotation marks indicate that it is just a quantified gut feeling, rather than a real probability distribution). Then she uses the prior together with the actual observed data to construct the posterior distribution. Finally, she takes her estimate $\hat{\theta}$ to be the mean of the posterior distribution.

8.8.2 Extent of Usage of Subjective Priors

Though some academics are staunch, often militantly proselytizing Bayesians, only a small minority of statisticians in practice use the Bayesian approach. It is not mainstream.

One way to see that Bayesian methodology is not mainstream is through the R programming language. For example, by my rough count in March 2010 of CRAN, the R repository, shows that only about 1% of R packages involve Bayesian techniques. There is actually a book on the topic, *Bayesian Computation with R*, by Jim Albert, Springer, 2007, and among those who use Bayesian techniques, many use R for that purpose. However, almost all general-purpose books on R do not cover Bayesian methodology at all.

Significantly, even among Bayesian academics, many use non-Bayesian (called **frequentist**) methods when they work on real, practical problems. Choose an academic statistician at random, and you'll likely find on the Web that he/she does not use Bayesian methods when working on real applications.

8.8.3 Arguments Against Use of Subjective Priors

As noted, most professional statisticians, including me, are frequentists. What are the arguments made in this regard?

First, the following must be noted carefully:

> Ultimately, the use of any statistical analysis is to make a decision about something. This

could be a very formal decision, such as occurs when the Food and Drug Administration (FDA) decides whether to approve a new drug, or it could be informal, for instance when an ordinary citizen reads a newspaper article reporting on a study analyzing data on traffic accidents, and she decides what to conclude from the study.

Frequentists believe that there is nothing wrong using one's gut feelings to make a final decision, but it should not be part of the mathematical analysis of the data. One's hunches can play a role in deciding the "preponderance of evidence," as discussed in Section 7.4.4, but that should be kept separate from our data analysis.

If for example the FDA's data shows the new drug to be effective, but at the same time the FDA scientists still have their doubts, they may decide to delay approval of the drug pending further study. So they can certainly act on their hunch, or on non-data information they have concerning the drug. But the FDA, as a public agency, has a responsibility to the citizenry to state what the data say, i.e. to report the frequentist estimate, rather than merely reporting a number—the Bayesian estimate—that mixes fact and hunch.

Thus, the Bayesian rallying cry, "It would be wrong to ignore any information we possess to supplement our data, even if that information is just a hunch," is presenting us with a false choice. The frequentists have never advocated ignoring hunches.

The Bayesians say that in some cases, a Bayesian estimator may, for instance, produce smaller mean squared estimation error (recall Section 8.2.3) than its frequentist counterpart, even if the prior distribution was just in our imaginations. But again, this argument is incorrectly implicitly presuming that frequentists ignore their hunches, which is not the case.

Moreover, in most applications of statistics, there is a need for impartial estimates. As noted above, even if the FDA acts on a hunch to delay approval of a drug in spite of favorable data, the FDA owes the public (and the pharmaceutal firm) an impartial report of what the data say. Bayesian estimation is by definition not impartial. One Bayesian statistician friend put it very well, saying "I believe my own subjective priors, but I don't believe those of other people." His statement should be considered by any potential user or consumer of Bayesian statistics.

Furthermore, in practice we are typically interested in inference, i.e. confidence intervals and significance tests, rather than just point estimation. We are sampling from populations, and want to be able to legitimately make inferences about those populations. For instance, though one can derive a Bayesian 95% confidence interval for p for our coin, it really has very little meaning, and again is certainly not impartial.

Consider the following scenario. Steven is running for president. Leo, his campaign manager, has commissioned Lynn to conduct a poll to assess Steven's current support among the voters. Lynn takes her poll, and finds that 57% of those polled support Steven. But her own gut feeling as an expert in politics, is that Steven's support is only 48%. She then combines these two numbers in some Bayesian fashion, and comes up with 50.2% as her estimate of Steven's support.

She then reports to Steven that she estimates Steven's support to be 50.2%. Leo asks Lynn how she arrived at that number, and she explains that she combined her prior distribution with the data. **But Leo then says, "Lynn, I really respect your political expertise, but I'd like you to tell me separately—what did the data say, and what is your own gut feeling? Lynn then tells Leo the two numbers, 57% and 48%, separately, and Leo finds both of them useful, in different senses.**

8.8.3.1 What Would You Do?

In evaluating the frequentist/Bayesian debate, you might wish to ask yourself what you would do in the following situations:

- As a personal investor, you've developed a statistical model for the day-to-day price variation of Google stock prices, and will use it to decide whether to buy the stock today. You wish to predict the price of the stock tomorrow, based on its price the last few days. Here are your choices:

  - As a frequentist, you could use a classical mathematical model, say regression analysis (Chapter 10), say fitting a linear or polynomial model. You could use the data to estimate the parameters of the model. This would give you a predicted price for tomorrow. Note that you can still choose to ignore that predicted price in the end, based on a hunch, but you've kept that hunch separate from your data analysis.
  - As a Bayesian, you might use the say linear or polynomial regression model, but you would specify a subjective prior distribution for the parameters. Your predicted price would then be affected by that subjective prior.

  So, what would you deem wise here—a frequentist or Bayesian approach?

- We are in a presidential election, complete with opinion polls as to who is currently winning. As an involved citizen, would you rather that the pollsters simply report the data as is, with their reported margin of error being computed from the traditional frequentist methods we've seen so far, or would you prefer that they factor in their own feelings via subjective priors?

  So, what would you deem wise here—a frequentist or Bayesian approach?

- You are a physician reading a medical journal article about the effectiveness of a certain drug for alleviating high blood pressure. Would you rather that the authors of the article simply report a straightforward analysis of the data, or would you prefer that the author incorporate a subjective prior distribution into his/her mathematical model?

  So, what would you deem wise here—a frequentist or Bayesian approach?

Exercises