1

Polling 101

Review

1. Why is the Agresti-Coull estimator not often used in polling analysis?
   Solution: *This question gets at the inherent tension between the value of a biased estimator and that of a precise estimator. The Agresti-Coull estimator is biased, but has greater precision than the usual sample proportion estimator. Thus, to a statistician, it is the better estimator.*

   However, it is rather difficult to make the case to use a biased estimator when an unbiased one is easily calculated. Furthermore, the sample proportion is easily understood, while the Agresti-Coull is not. Finally, unless the sample size is incredibly small (less than 10), the improvement on precision is rather minor. Thus, it is an improvement, but not much of one.

   For these reasons, the Agresti-Coull estimator is rarely used.

2. How would you convince a person that an estimator with a lower MSE is preferred to one that is unbiased?
   Solution: *Great question. I have yet to find a way of doing this successfully. The closest I've come is to emphasize the precision and the likelihood of being closer to the right answer (with the lower MSE estimator) rather than being right "on average." The "on average" requires performing the experiment many times in order to reap the benefit.*

3. What are the main differences between a confidence interval and a credible interval? When should you use each?
   Solution: *The main difference is that a credible interval is based on probabilities. That is, one can state "the probability of π being in the interval is 95%." For a confidence interval, such a statement is not true. All that can be said is "95% of the time, π is in the*

interval." We don't know if this is one of those 95% or one of the 5%.

I advocate using credible intervals at all times, because it provides more information. However, some hold that Bayesian analysis is inherently weak because it bases itself on an assumption about the distribution of π.

4. List several advantages and disadvantages to Bayesian analysis.
   Solution: *Bayesian analysis provides a way of incorporating prior information (or lack thereof) in the analysis. With that prior information, and the additional information from the collected data, one is able to calculate the distribution of the population parameter (or of its uncertainty).*

   The main drawback is that the results are based on the assumption that the prior distribution is correct. If the prior is wrong, then the posterior distribution is wrong.

   However, I do not see this as a weakness. Frequentist analysis also makes an assumption about the distribution of the parameter. It just is not as explicit about it.

5. Why is the conservative margin of error approximately $1/\sqrt{n}$?
   Solution: *The conservative margin of error is based on assuming the population parameter is $\pi = 0.500$. Using this as the estimate of π produces the widest confidence intervals.*

6. What does the thick line at the bottom of Figure 1.2, page 12, represent?
   Solution: *This line represents the 95% confidence interval. A total of 95% of the binomial distribution occurs within that interval (47 to 53%).*

7. Why is the beta distribution called the "conjugate prior" for the binomial distribution?
   Solution: *The beta distribution is the conjugate prior to the binomial distribution because if we make the prior distribution a beta, then the posterior distribution will also be a beta.*

8. What is "coverage," and how can a biased estimator have actual coverage closer to the claimed coverage than an unbiased estimator? Do all biased estimators have better coverage than unbiased estimators?

   Solution: *Coverage is the proportion of the time that the confidence interval will contain (or cover) the parameter (see Figure 1.6 for an illustration).*

   While a part of the confidence interval, the parameter estimator is not the whole story. The endpoints contain both the estimator and *the standard error (standard deviation of the sampling distribution). Thus, the unbiased estimator may have a standard error that is much greater than "it should be," which causes the coverage to be much higher than 95%.*

   No, not all biased estimators have better coverage. However, it does open a new field of study in trying to determine correct confidence intervals, even if the estimator is biased.

Conceptual Extensions

1. There were four counties (local authorities) in which the 2014 Scottish independence referendum received more than 50% of the votes cast (Figure 1.1). What do they have in common?
2. The turnout for the 2014 Scottish independence referendum was not 100%. Assuming that those who did not vote in each county were similar to those who did, would a higher turnout have helped independence or not?
3. The Survation poll estimated independence support to be 47%, which was 2.3% too high. Was this poll "too far off"? This question gets at your understanding of the confidence interval.
4. In the beta distribution, one can think of $a + b$ as the effective sample size, n. With this, explain why $\mathbb{E}\left[\, Y \,\right] = \frac{a}{a+b}$ makes sense. Also, compare the variance of a beta distribution with that of the sample proportion, $\frac{\pi(1-\pi)}{n}$.
5. Using words, answer these two questions:
   (a) How do we know that the binomial distribution is equivalent to the hypergeometric distribution when $n = 1$?
   (b) How do we know that the binomial distribution is equivalent to the hypergeometric distribution when $N = \infty$?
6. Using mathematical proofs, answer these two questions:

(a) How do we know that the binomial distribution is equivalent to the hypergeometric distribution when $n = 1$?

(b) How do we know that the binomial distribution is equivalent to the hypergeometric distribution when $N = \infty$?

Computational Extensions

1. Using Defn. 1.3, prove $\mathrm{MSE}\left[\ P\ \right] = \mathrm{bias}^2[P] + \ \mathbb{V}\left[\ P\ \right]$.
2. Repeat all of the Survation calculations for this chapter for a poll in which 256 out of 500 people supported Scottish independence. Do the conceptual conclusions significantly differ?
3. In several places, we focused exclusively on the kernel of the distribution and completely ignored the normalizing constants. Why can we do that? Are distributions *uniquely* determined by their kernel?