Lab 10 – MATH 240 – Computational Statistics

Reagan Sernick

`rsernick@colgate.edu`

04/09/25

Keywords: Binomial Distribution: a probability distribution that summarizes the likelihood of a given number of successes in a fixed number of independent trials, each with the same probability of success. ; **Margin of Error (MOE):** range of values above and below a sample statistic that is likely to contain the true population parameter, given a certain confidence level.

1 Introduction

This report explores the relationship between binomial distributions and margin of error in the context of political polling data. Using simulations and resampling methods based on a Gallup poll, we examine how changes in sample size and population proportion affect the variability of survey estimates. By modeling the sampling distribution with binomial assumptions, we highlight how statistical tools can quantify uncertainty and improve confidence in public opinion measurements.

2 Methods

The data we use in this lab came from a Gallup polls political poll of 1004 randomly sampled Americans. Gallup reported that 39% of respondents were satisfied with the position of the United States in the world today, 59% were not, and 2% had no opinion. They also reported a 4% margin of error and claimed that if the sample size were doubled the MOE would be halved.

2.1 Simulation Assuming the Value of p

If we assume the value of p to be $\hat{p}$ (0.39) we can use `rbinom()` to simulate a sampling distribution at any given sample size.

2.1.1 Sample size of 1004

For this task we ran a simulation of 10,000 samples of size 1004, to see what a sampling distribution might look like if the population parameter p was 0.39.
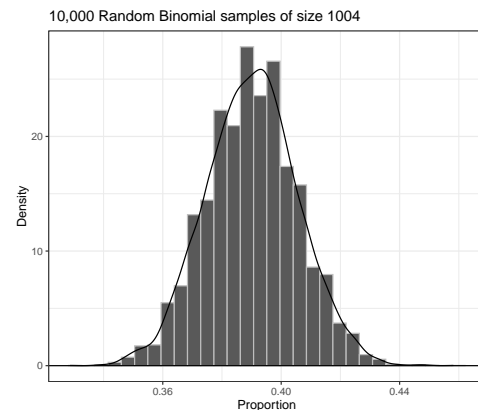

Figure 1: Sampling distribution of sample size 1004

2.1.2 Sample size of 2008

Much like the last task, we ran a simulation of 10,000 samples assuming p to be 0.39, but doubled the sample size to be 2008.
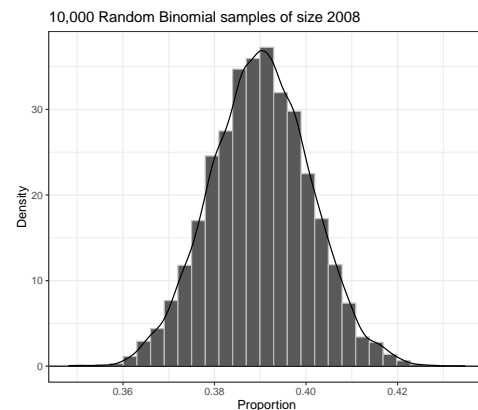

Figure 2: Sampling distribution of sample size 2008

This graph shows significantly less variation than the previous one, and appears to be more normally distributed as well.

2.2 Simulation Without Knowing the Value of p

While `rbinom()` can help create a realistic sampling distribution based on the parameter p, it is not often that we know the value of the population parameter. In this case it would

make more sense to use resampling. However, in order to do resampling, some conditions must be met:

Approximately Gaussian Sample size (1004) > 30 so Central Limit Theorem (CLT) applies

Representative The researchers performed this survey carefully to make sure that the sample is representative

Knowing the conditions were met I was able to resample the data using `sample()` on sample data I created from the given proportions.
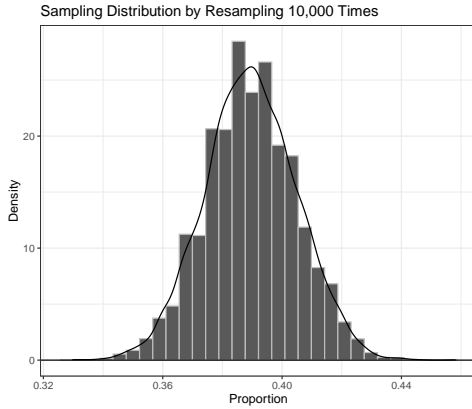

Figure 3: Resample from original sample of size 1004

2.3 How do n and p Affect Margin of Error?

In order to see how both sample size and the probability parameter affect MOE, we ran a simulation across n in {100, 110, 120, ..., 3000} and p in {0.01, 0.02, ..., 0.99}. This simulation was the same as the first 2, 10,000 samples using `rbinom()`, but n and p were varied. For this complicated simulation we generated $2,900 * 99 * 10,000 = 2,871,000,000$ random responses. This method was used once for our MOE estimate (half the range of the middle 95%) and once for Wilson's MOE:

$$\frac{z_{1-\alpha/2}\sqrt{n\hat{p}(1-\hat{p}) + \frac{z_{1-\alpha/2}^2}{4}}}{n + z_{1-\alpha/2}^2}$$

3 Results

3.1 Does Resampling Work?

The theory behind resampling is that if the conditions are met, as n (number of resamples) gets larger the resampled sampling distribution tends towards a sampling distribution of n samples the same size. The Gallup poll data meets the conditions for resampling so our resampled distribution should be close to a random sampling distribution. This is exactly what we see in Figure 1 and Figure 3, the distribution of the `rbinom()` data and the resample are almost identical.

	Sample	MOE
1	rbinom size 1004	0.02988048
2	rbinom size 2008	0.02116534
3	resample	0.02988048

Table 1: Estimated MOEs for simulations

Furthermore, as seen in Table 1 above, the margin of error of the first simulation and the resampling distribution are identical to 8 decimal places.

3.2 Relationship between n, p and MOE

Both margin of error raster plots seen in Figure 4 reveal interesting details about MOE as a function of n and p. The first thing I noticed is that at the extremes of p (near 1 and 0), and any value of n, MOE is low. Another detail is that the largest MOE is at p = 0.50 for every value of n. This relationship can be generalized to as n increases MOE increases, and as p gets further from 0.50 in any direction MOE increases.

Additionally, Figure 4 reveals the accuracy of our MOE estimate compared to Wilson's MOE. The MOE estimate plot is noisy, meaning it is slightly imprecise, but the Wilson's MOE is significantly less noisy, meaning it is much more precise.
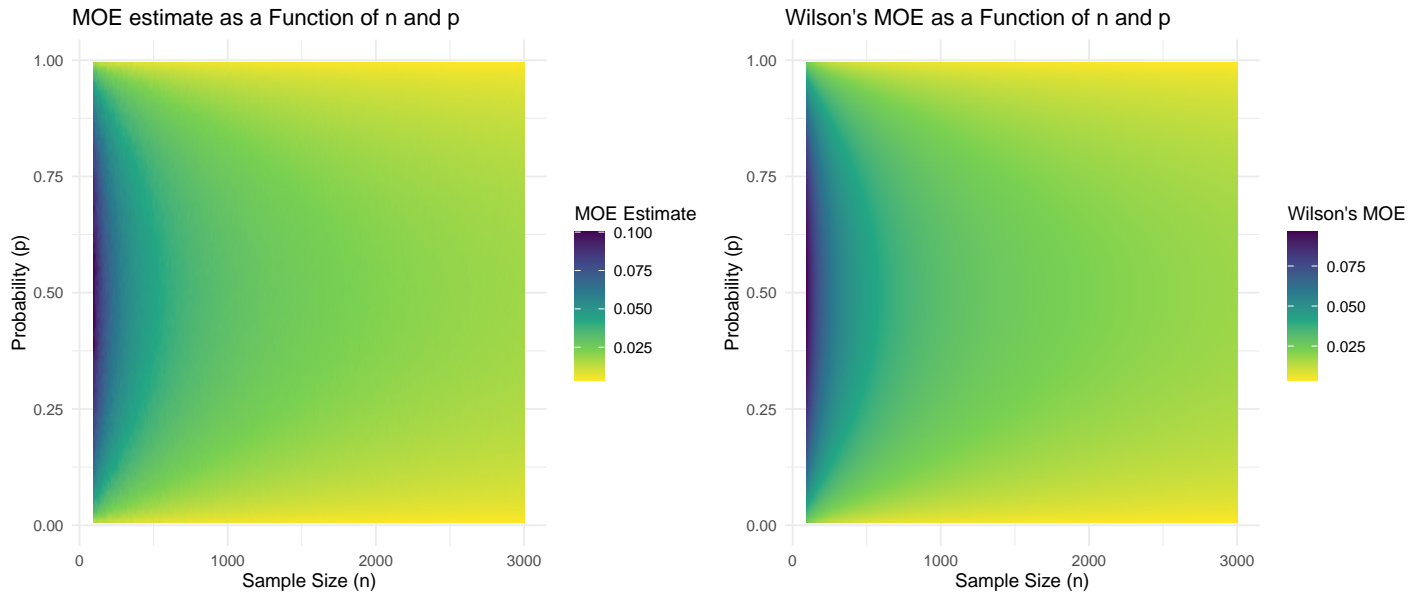
4 Appendix

Figure 4: Combined raster plots of MOE estimates and Wilson's MOE (*I reversed the colors because it makes more sense to me this way*)