

Lab 10 – MATH 240 – Computational Statistics

Michael Boateng
Math Department
mboateng@colgate.edu

Abstract

According to Gallup polls, a doubling in the sample size for polls from 1000 to 2000, results in decreasing the margin of error by two percentage points. In this lab, we test this inference by first conducting a basic simulation of random binomial distribution, calculating its margin of error, and then doing same for double the sample size. Then, using data from the Gallup survey we perform resampling to approximate \hat{p} . Finally, we calculate the margin of error for multiple sample sizes (n) and probabilities (p). We find that the margin of error depends not only on sample size but also p .

Keywords: resampling; simulation; distribution; point estimate

1 Introduction

Gallup polls in *"How Are Polls Conducted?"*, claimed that with an initial sample size of carefully selected adults of 1000, the margin of error of the results obtained from such a sample is likely to be within $\pm 4\%$. They then followed that doubling the sample size leads to a two percentage point decrease in the margin of error. We aim to test how true this assertion from Gallup polls is.

As an example data and frame of reference, we'll use the February 3-16, 2025 poll of 1004, adults living in all 50 U.S states and D.C.. In this poll 39% of respondents were satisfied with the position of the U.S in the world, 59% were dissatisfied (and 2% of no opinion).

We begin our test by making simulations of the Gallup survey and then making estimates of the margin of error from these simulations. We then calculate the true margin of error on multiple samples with different sizes and sampling distributions to effectively know how well sample size affect margin of error values.

2 Basic Simulation

Given that the Gallup poll reported 39% satisfaction, we take this to be our probability of success and conduct a simulation study. Using `rbinom()`, we generate 10k polls of the same sample size ($n = 1004$). We see the the distribution of this sample in figure 1

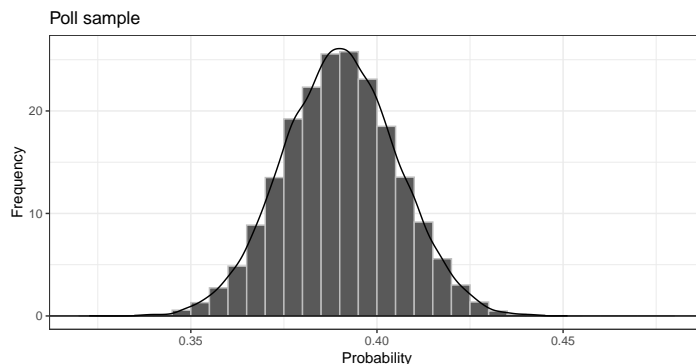


Figure 1: Simulation of Gallup Poll with $n = 1004$

From this sample, we estimate the margin of error by halving the range of the middle 95% of the data. We find the value to be 3.03%, which is less than the estimated 4% from Gallup. Next we, double our sample size to obtain the distribution in figure 2. The margin of error for this sample was 2.1%, which is close to Gallup's reported value of 2%. In both figure 1 and 2, we see the the data are centered around are true probability value and are also normally distributed.

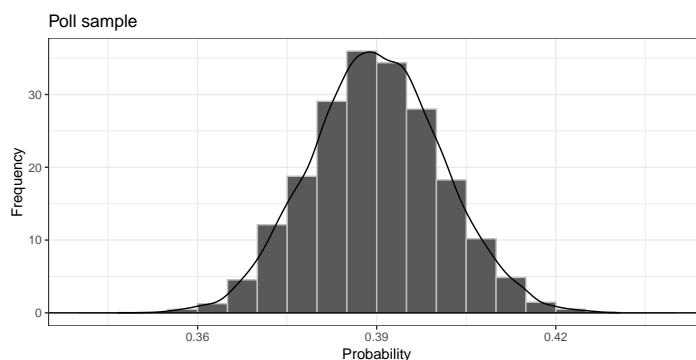


Figure 2: Simulation of Gallup Poll with $n = 2008$

3 Resampling

During our simulation, we created samples of data under the assumption that ($p = 0.39$). Another way of obtaining the value of p , is through resampling. In resampling, we make repeated samples of an original sample and calculate our statistic of interest with each iteration, and then after we make summaries on the particular statistic.

We performed resampling on the Gallup data (with our statistic of interest being the mean (i.e percent "satisfied")),

and created a graphical summary in figure 3 below. We see that our sampling distribution is normally distributed, and also heavily centered around our true probability value (0.39). We calculated the margin of error on the resampling data to be 3.07%, which is less than Gallup's report of 4%.

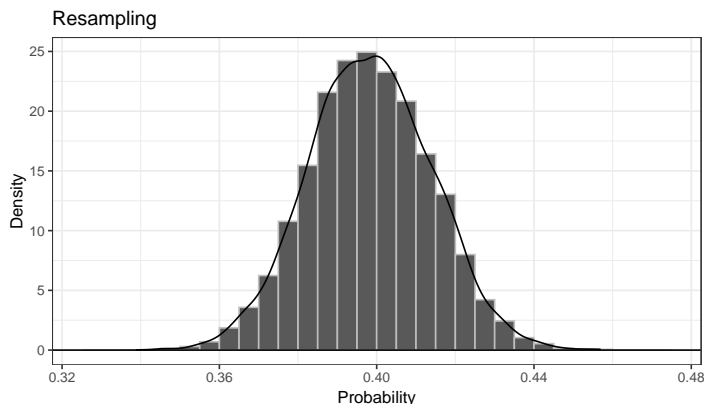


Figure 3: Resampling Distribution

4 Simulation over n and p

We now consider the effect of not just sample size but also the probability of success on margin of error values.

For n in $\{100, 110, 120, \dots, 3000\}$ and p in $\{0.01, 0.02, \dots, 0.99\}$, we perform 10k simulations and each time calculating and storing the estimated margin of error. The simulations are summarized in figure 4 below.

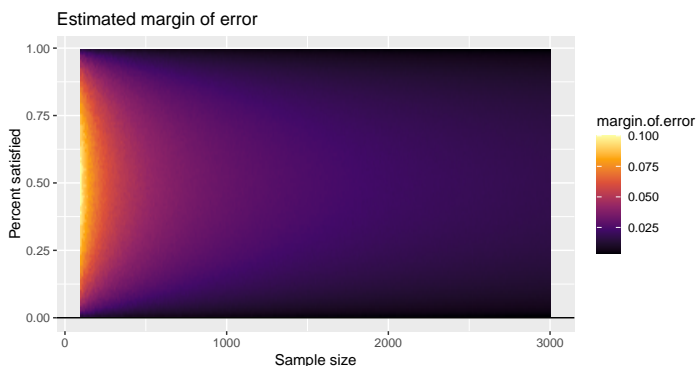


Figure 4: Margin of error as a function of n and p

4.1 Actual Margin of Error Calculation

Previously, we had estimated the margin of error by halving the range of the middle 95%. However using Wilson's estimate we can find the actual margin of error of the data.

We computer the margin of error using the Wilson formula for n in $\{100, 110, 120, \dots, 2000\}$ and p in $\{0.01, 0.02, \dots, 0.99\}$. Results summarized in figure 5.

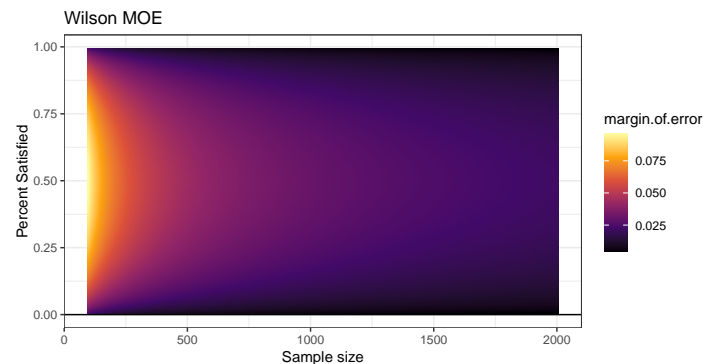


Figure 5: Margin of error as a function of n and p (wilson's estimate)

5 Discussion

While Gallup's report of how a double in sample size decreases the margin pf error is true to some extent, it does not reflect how the value of p can also influence the change in margin of error. Regardless of sample size, we see in figure 4 and 5, is smallest as p takes a value close to the extremes (0 or 1). However, looking at the same graph we see that margin of error, is greater for smaller for larger sample sizes that smaller one for non-extreme values of p .

This suggests that the actual story is not as simple Gallup made it out to be. For extreme population proportion values, doubling our sample data will not yield such a change in margin of error.