Econ 325
Simulation: the Law of Large Numbers, the Central Limit Theorem, and the Consistency of Estimator [1]
By Hiro Kasahara

Let p be the fraction of voters who vote for Trump in Ohio. In 2016, Donald Trump received 52.1 percent of votes in Ohio so that $p = 0.521$. Suppose that a survey was conducted just before the presidential election by asking n eligible voters who are randomly sampled and who honestly answered the survey question. The i-th voter's preference was coded as $X_i = 1$ if voting for Trump and $X_i = 0$ if voting for someone else. Therefore, the data set $\{X_1, X_2, ..., X_n\}$ is a random sample, where each observation is a Bernoulli random variable, i.e.,

$$X_i = \begin{cases} 0 & \text{with prob. } 1 - p = 0.479 \\ 1 & \text{with prob. } p = 0.521 \end{cases} \tag{1}$$

Given the data set, we compute the sample fraction as

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The sample fraction $\hat{p}$ is a random variable. By simulation, we examine how the distribution of $\hat{p}$ changes as n increases.

In computer, we may generate a data set of n observations by drawing a random variable X_i for $i = 1, ..., n$ from the Bernoulli distribution with $p = 0.521$. Then, we may compute the sample fraction by taking the average of n observations. We simulate a large number of data sets (10000000 data sets), where each data set contains n observations. For each of 10000000 data sets, we compute the sample fraction $\hat{p}$. Then, we plot the histogram for 10000000 realized values of $\hat{p}$.

Figure 1 presents the histograms of $\hat{p}$ when n = 1, 2, 5, 10, 50, 100, 1000, 10000, and 1000000, where each figure is generated from 10000000 data sets. For example, when $n = 1$, each of 10000000 data sets contains only 1 observation so that $\hat{p} = X_1$ for each data set. Therefore, when $n = 1$, $\hat{p}$ takes the value equal to 1 for about $10000000 \times 0.521 = 5210000$ data sets while $\hat{p}$ takes the value equal to 0 for about $10000000 \times 0.479 = 4790000$ data sets.

When $n = 2$, each data set contains two values $\{X_1, X_2\}$, where $\hat{p} = 1$ if $X_1 = X_2 = 1$, $\hat{p} = 1/2$ if $(X_1, X_2) = (0, 1)$ or $(1, 0)$, and $\hat{p} = 0$ if $(X_1, X_2) = (0, 0)$. When $n = 2$, out of 10000000 data sets, $\hat{p}$ takes the value equal to 1 for about $10000000 \times (0.521)^2 \approx 2714410$ data sets, $\hat{p}$ takes the value equal to 1/2 for about $10000000 \times 2 \times 0.521 \times 0.479 \approx 4991180$ data sets, and $\hat{p}$ takes the value equal to 0 for about $10000000 \times (0.479)^2 \approx 2294410$ data sets. As n increases, the number of possible values $\hat{p}$ can take increases.

In Figure 1, as n increases from $n = 1$ to $n = 1000000$, *the distribution of* $\hat{p}$ *shrinks toward a point* $p = 0.521$. This is a consequence of **the Law of Large Numbers**. When n is very large, all realized values of $\hat{p}$ fall in an arbitrary close neighborhood of $p = 0.521$ with probability one. In such a case, we say that $\hat{p}$ **converges in probability** to p and write as

$$\hat{p} \xrightarrow{p} p.$$

Also, $\hat{p}$ is said to be a **consistent estimator** of p when $\hat{p}$ converges in probability to p.

The shape of histogram for $\hat{p}$ approaches the normal density function in Figure 1, as n increases from $n = 1$ to $n = 1000000$. To see this more clearly, consider a standardized random variable by subtracting the mean of $\hat{p}$ and then dividing by the standard deviation of $\hat{p}$, where $E[\hat{p}] = p$ and $Var(\hat{p}) = \frac{p(1-p)}{n}$. By **the Central Limit Theorem**, a standardized random variable $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ **converges in distribution** to a standard normal distribution, i.e.,

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1).$$

To see this, Figure 2 plots the histogram of the standardized random variable $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ when $n = 1$, 2, 5, 10, 50, 100, 1000, 10000, and 1000000. These figures in Figure 2 are generated similarly to those in Figure 1 except that $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is used in place of $\hat{p}$. *The distribution of* $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ *approaches the standard density function* as n increases from $n = 1$ to $n = 1000000$.

Figure 1: Frequency distribution of $\hat{p}$ when $n = 1, 2, 5, 10, 50, 100, 1000, 10000$, and 1000000 across 10000000 simulated data sets

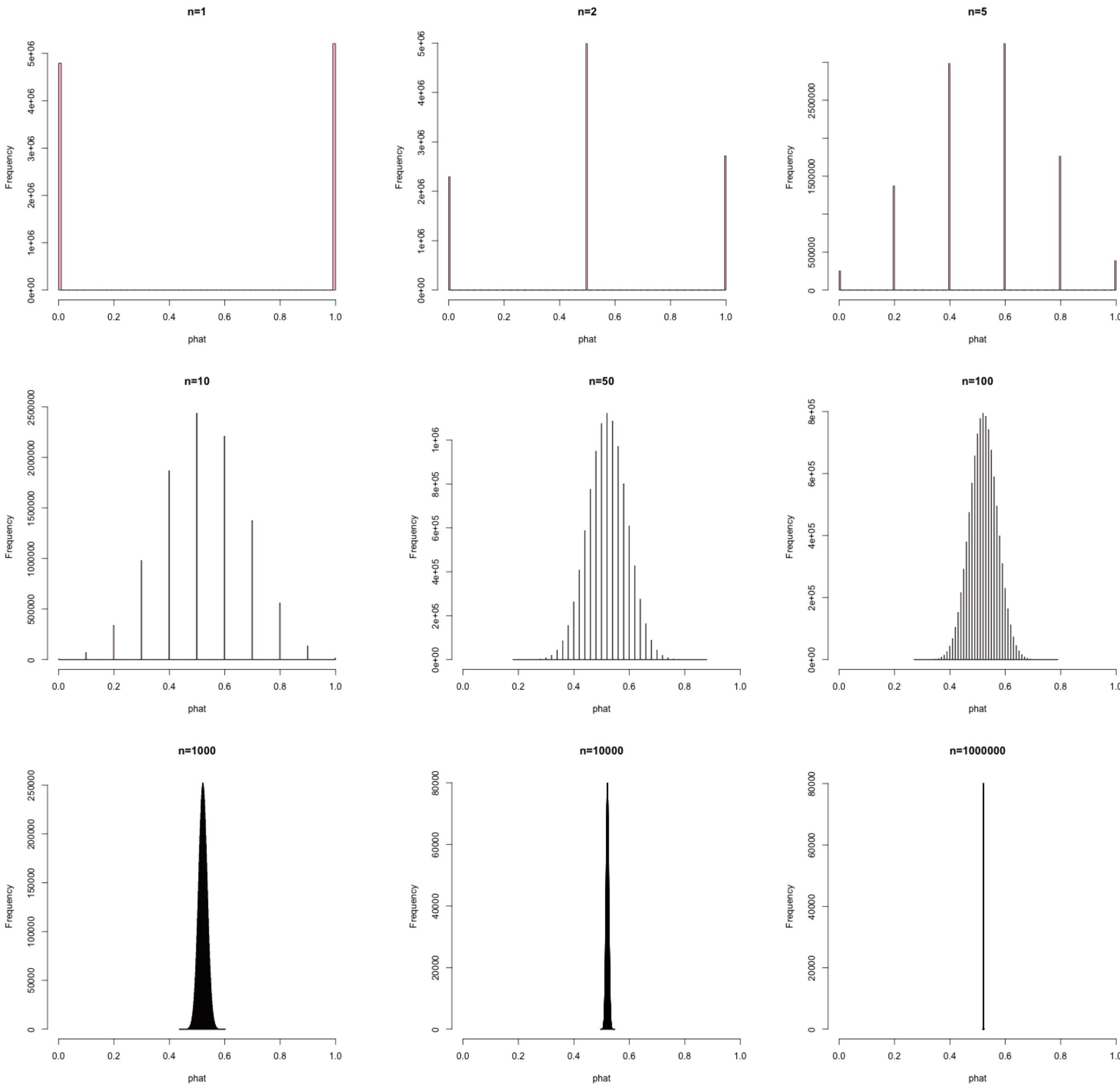

Figure 2: Frequency distribution of $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ when $n = 1$, 2, 5, 10, 50, 100, 1000, 10000, and 1000000 across 10000000 simulated data sets