MSE, WLLN, and CLT

Nicholas Ray

2024-09-16

The following discussion puts concepts in terms of the sample mean as an estimator for the population mean. The below concepts apply to most estimators, however, and students should refer to the relevant citations to read more general statements of these concepts.

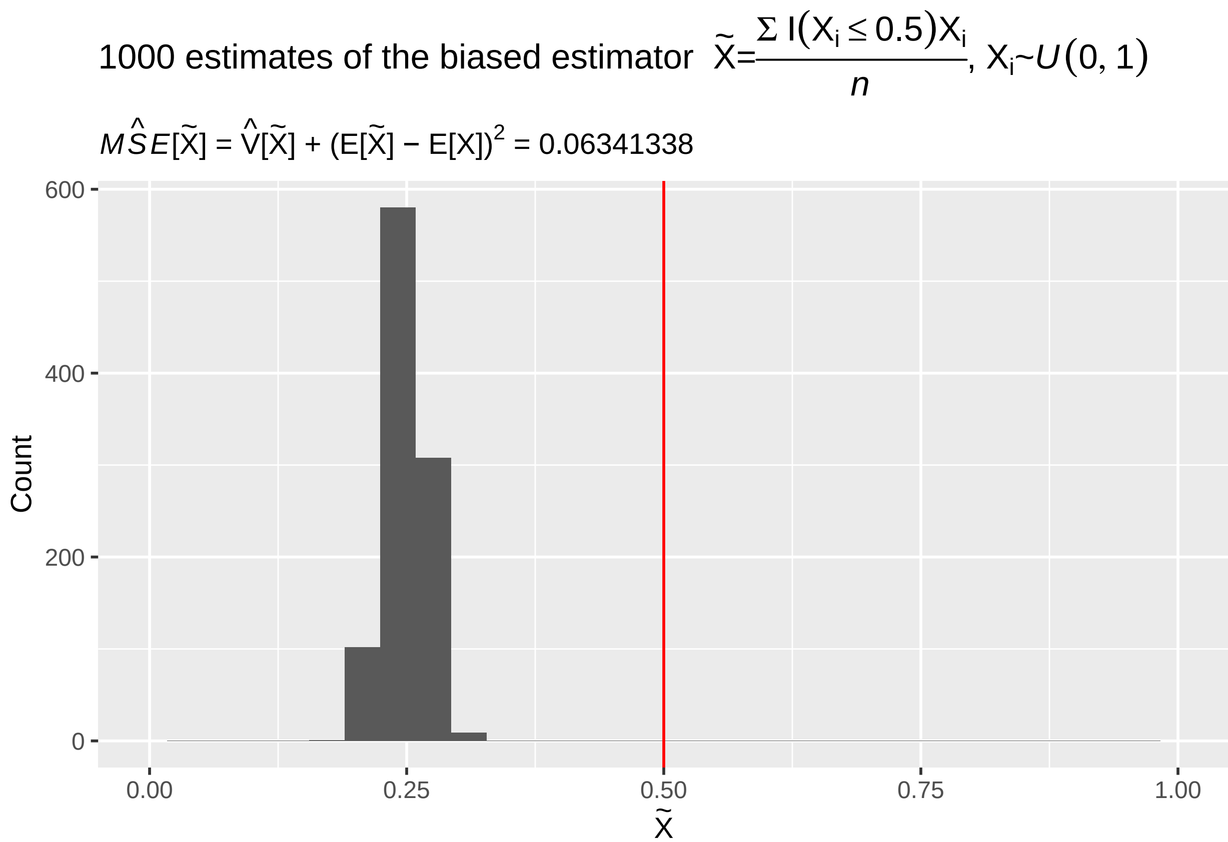
Bias, Variance, and MSE

Let's say we have an i.i.d sample (Aronow and Miller 2019, 92). We can estimate the mean of the population that produced this sample ($\mathbb{E}[X]$) with the sample mean ($\overline{X}$) (Aronow and Miller 2019, 97). We can also estimate the variance of the population ($V[X]$) by using the estimated (unbiased) variance of the sample $\left(\hat{V}[X] = \frac{n}{n-1}(\overline{X^2} - \overline{X}^2)\right)$ (Aronow and Miller 2019, 107). Further, we can estimate the variance of the sample mean (i.e, "sampling variance" of the "sampling distribution" $V[\overline{X}] = \frac{V[X]}{n}$) by using $\hat{V}[\overline{X}] = \frac{\hat{V}[X]}{n}$ (Aronow and Miller 2019, 114).

Mean squared error (MSE) combines information about bias ($\mathbb{E}[\overline{X}] - \mathbb{E}[X]$) and (sampling) variance ($V[\overline{X}]$). It is essentially the variance of an estimator (sampling variance) about the true population value. Therefore, it is defined as $\mathbb{E}\left[\left(\overline{X} - \mathbb{E}[X]\right)^2\right]$ but can also be written as $V[\overline{X}] + \left(\mathbb{E}[\overline{X}] - \mathbb{E}[X]\right)^2$ (Aronow and Miller 2019, 104). An estimator that has lower MSE is called more *efficient*. It is perhaps important to note that if you read different treatments of these subjects, a more efficient estimator usually refers to an estimator with lower variance while MSE is seen as combining information on bias and efficiency.

To "see" MSE, consider a biased estimator $\widetilde{X}$ for the population mean of the standard uniform distribution. Let $\widetilde{X} = \frac{1}{|X \leq 0.5|} \sum_{i=1}^{n} I(X_i \leq 0.5)$. The estimator $\widetilde{X}$ subsets the sample to observations below or equal to 0.5 and then estimates the sample mean. For a random variable X that follows a standard uniform, we (or the contributors to Wikipedia) know that $\mathbb{E}[X] = 0.5$ and $V[\overline{X}] = \frac{1}{12n}$. Unfortunately, our biased estimator $\widetilde{X}$ has $\mathbb{E}[\widetilde{X}] = 0.25$ and $V[\widetilde{X}] = \frac{1}{48n}$. Thus, for our estimator $MSE[\widetilde{X}] = V[\widetilde{X}] + (\mathbb{E}[\widetilde{X}] - \mathbb{E}[X])^2 = \frac{1}{48n} + (0.25 - 0.5)^2$. If we had a sample size of 100, $MSE[\widetilde{X}] = 0.0627$. The following function `mse()` simulates this by estimating the population mean with $\widetilde{X}$ 1000 times with each sample being of size 100. The $MSE[\widetilde{X}]$ is the sampling variance seen in the graph plus the square distance between the (expected) center of the sampling distribution and the true mean 0.5 (denoted by the red line).

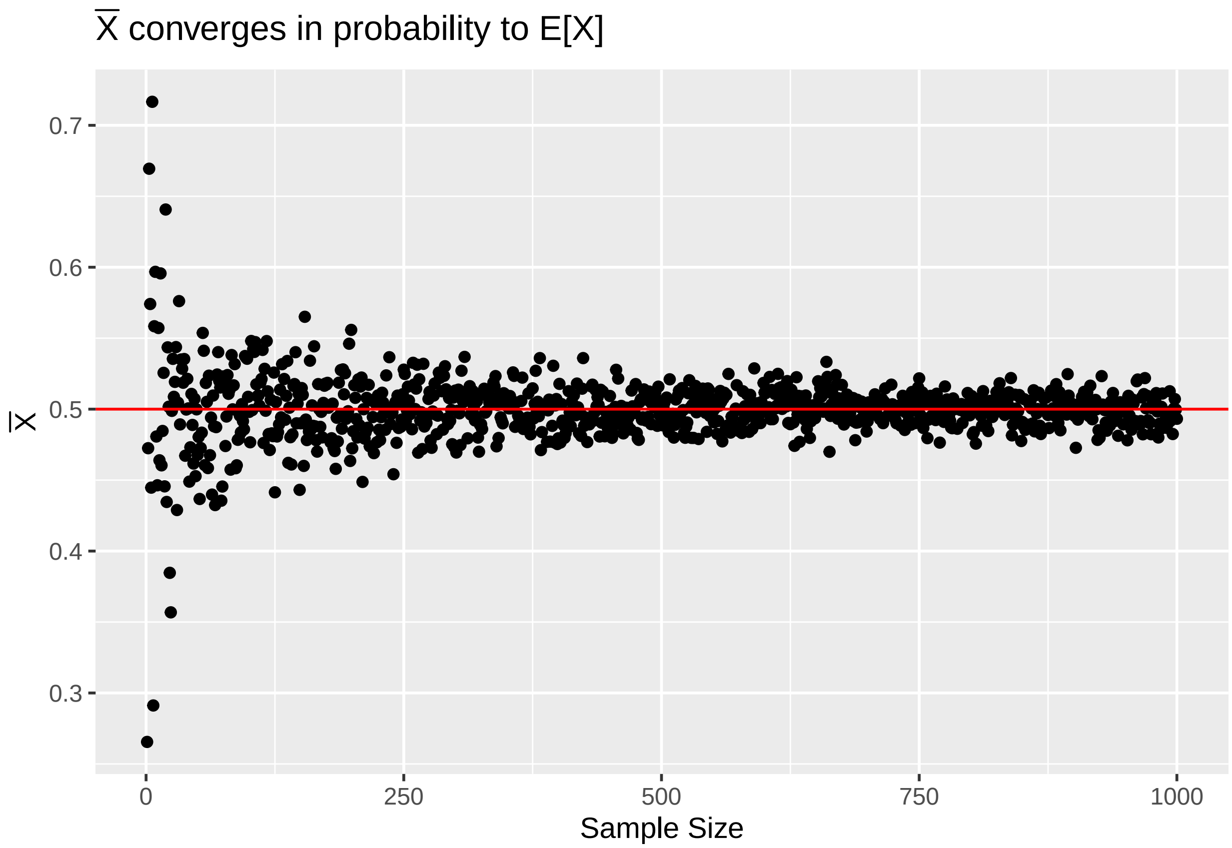
```
mse()
```



WLLN and CLT

The Weak Law of Large Numbers ($WLLN$) says that as (i.i.d) sample size goes to infinity, the probability that the distance between the sample mean and the true population average is greater than some arbitrarily small number goes to 0 (Aronow and Miller 2019, 99–100). In other words, the sample mean ($\overline{X}$) converges in probability to ($\xrightarrow{p}$) the population average ($\mathbb{E}[X]$). The WLLN underlies the concept of *consistency*, where saying "the sample mean is consistent for the population mean" (Aronow and Miller 2019, 105) means that it converges in probability to the population mean. This makes intuitive sense because as the data we have tends to infinity, the sample and its mean will increasingly resemble the "population" and its mean.

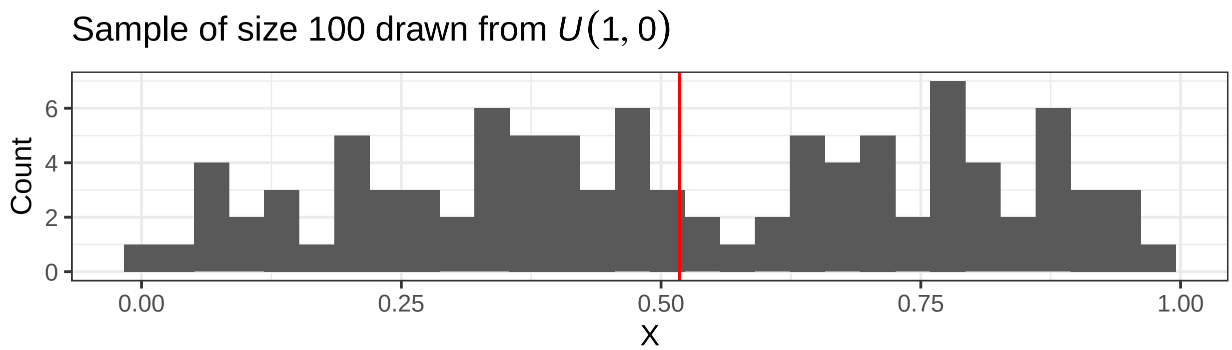
```
wlln(limit=1000)
```

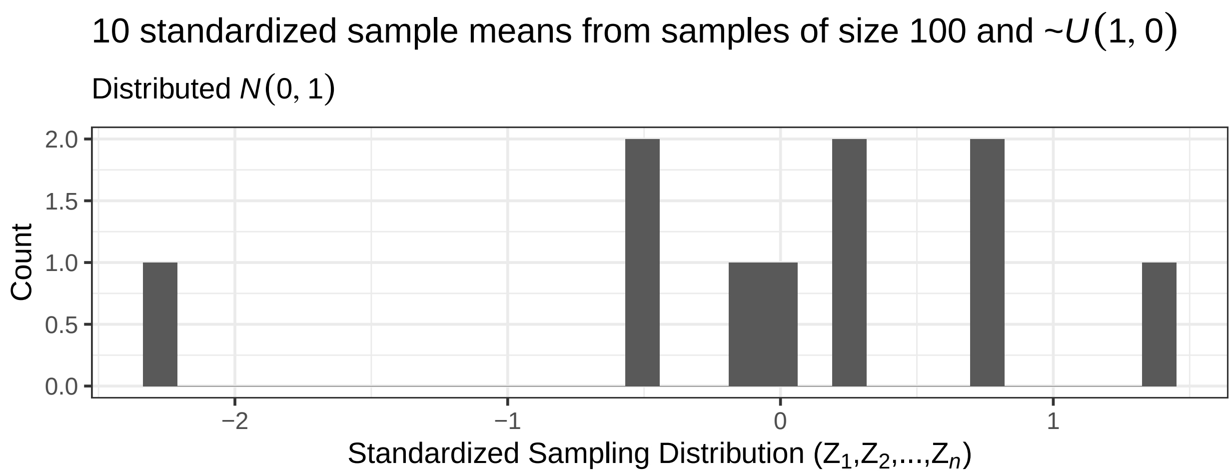


The Central Limit Theorem (CLT) states that an i.i.d. sampling distribution $(\overline{X}_1, \overline{X}_2, ..., \overline{X}_n)$ converges in distribution to ($\xrightarrow{d}$) a normal distribution $\big(N(\mathbb{E}[\overline{X}], V[\overline{X}])\big)$. In other words, the distribution of sample means will tend to be normally distributed as the number of sample means taken tends to infinity. If the sampling distribution is *standardized* such that $\overline{X}_1, \overline{X}_2, ..., \overline{X}_n = Z_1, Z_2, ..., Z_n$ where $Z_i = \frac{\overline{X}_i - \mathbb{E}[\overline{X}]}{\sigma[\overline{X}]}$, then the standardized sampling distribution converges in distribution to a standard normal distribution $\big(N(0,1)\big)$ (Aronow and Miller 2019, 109).

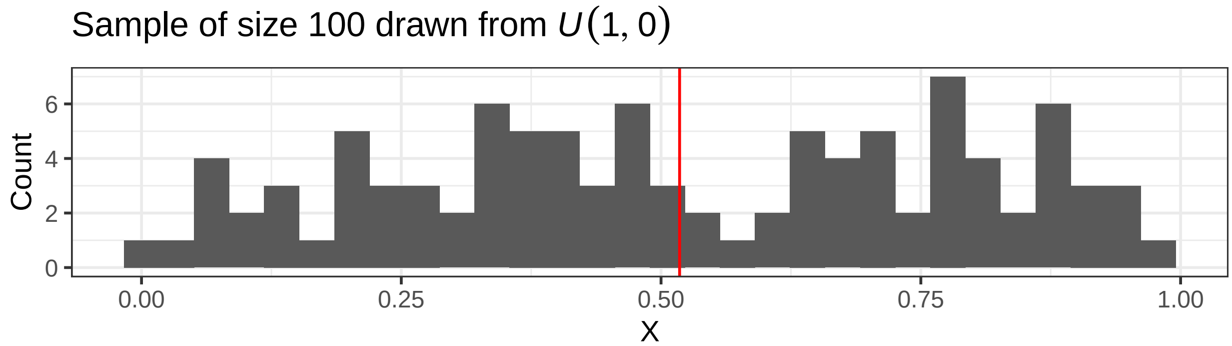
To see this, hold `sample_size` fixed and increase `n` (number of means/size of sampling distribution) in the function `CLT()` below. Toggle the argument `standardize` off by specifying `"no"`, and observe the differences between the standardized and unstandardized sampling distributions.

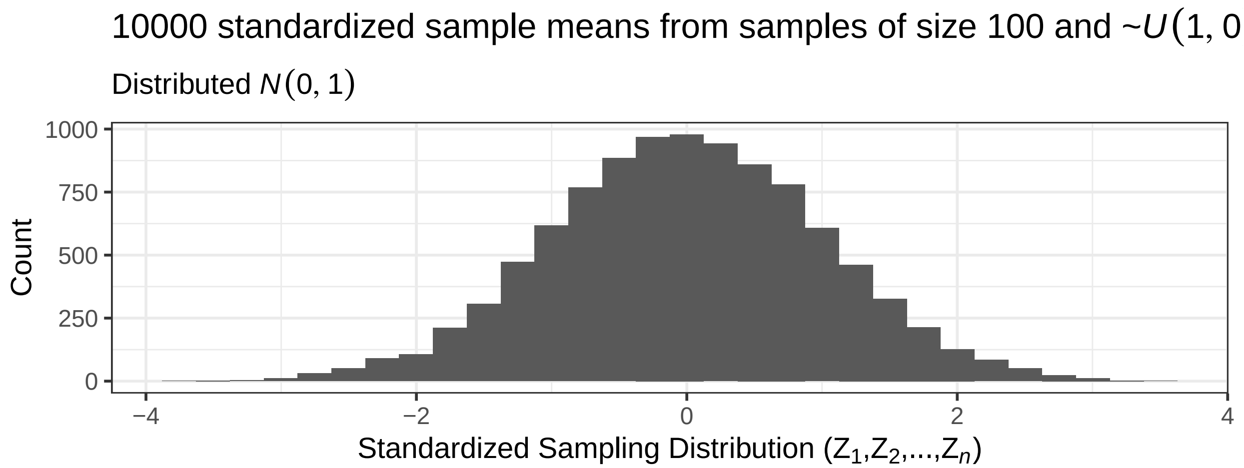
```
CLT(sample_size=100,n=10,standardize="yes")
```






```
CLT(100,10000,"yes")
```

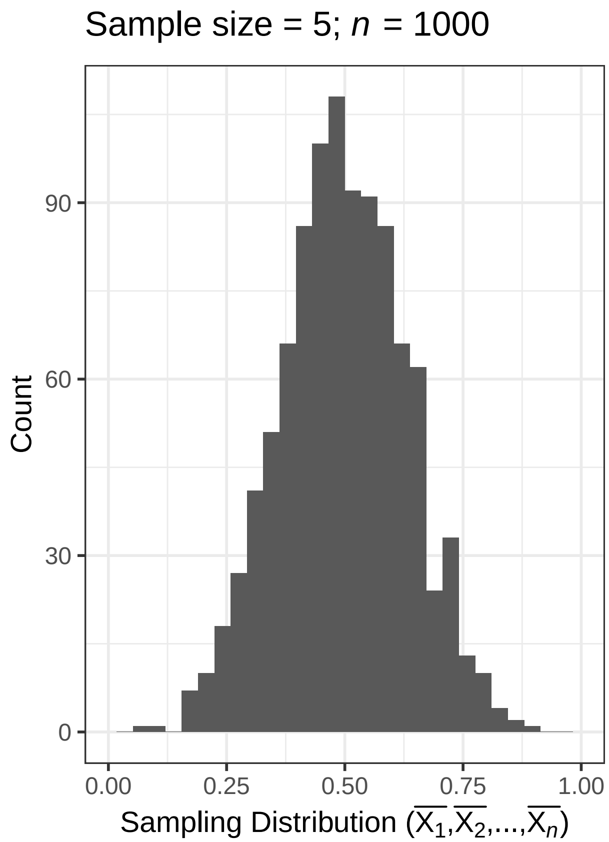




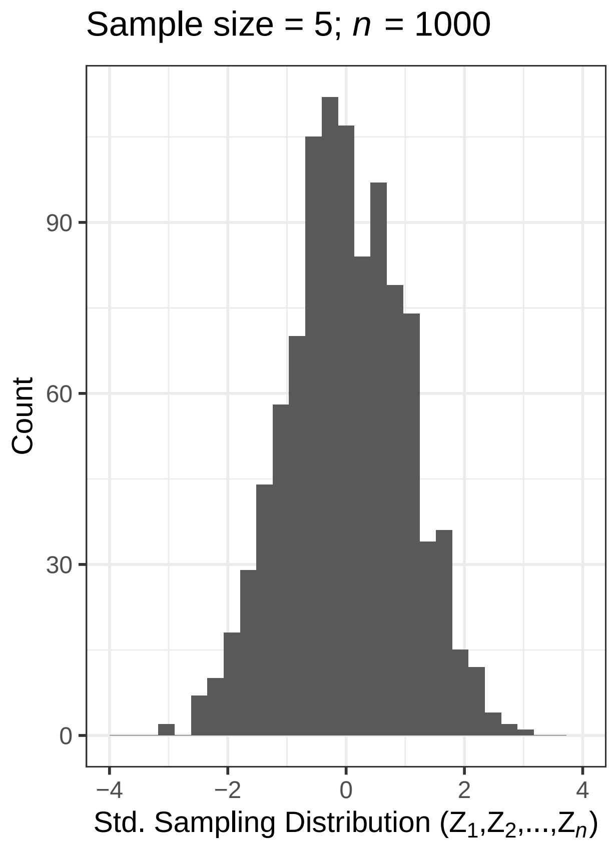
Note on Asymptotic Normality

Distinguishing between standardized and unstandardized sampling distributions is important because standardized or "rescaled" sampling distributions are asymptotically normally distributed while unstandardized sampling distributions may not be. With large sample sizes, unstandardized sampling distributions (for consistent estimators) will converge to a spike about $\mathbb{E}[X]$ due to the $WLLN$. However, standardized sampling distributions will still be (asymptotically) normally distributed. To see this, increase `sample_size` while holding `n` constant in the function `an()` below. Also note that Aronow and Miller (2019) discuss asymptotic normality in terms of "rescaling" an estimator and not necessarily standardization (Aronow and Miller 2019, 112–13).

```
an(sample_size=5,n=1000)
```

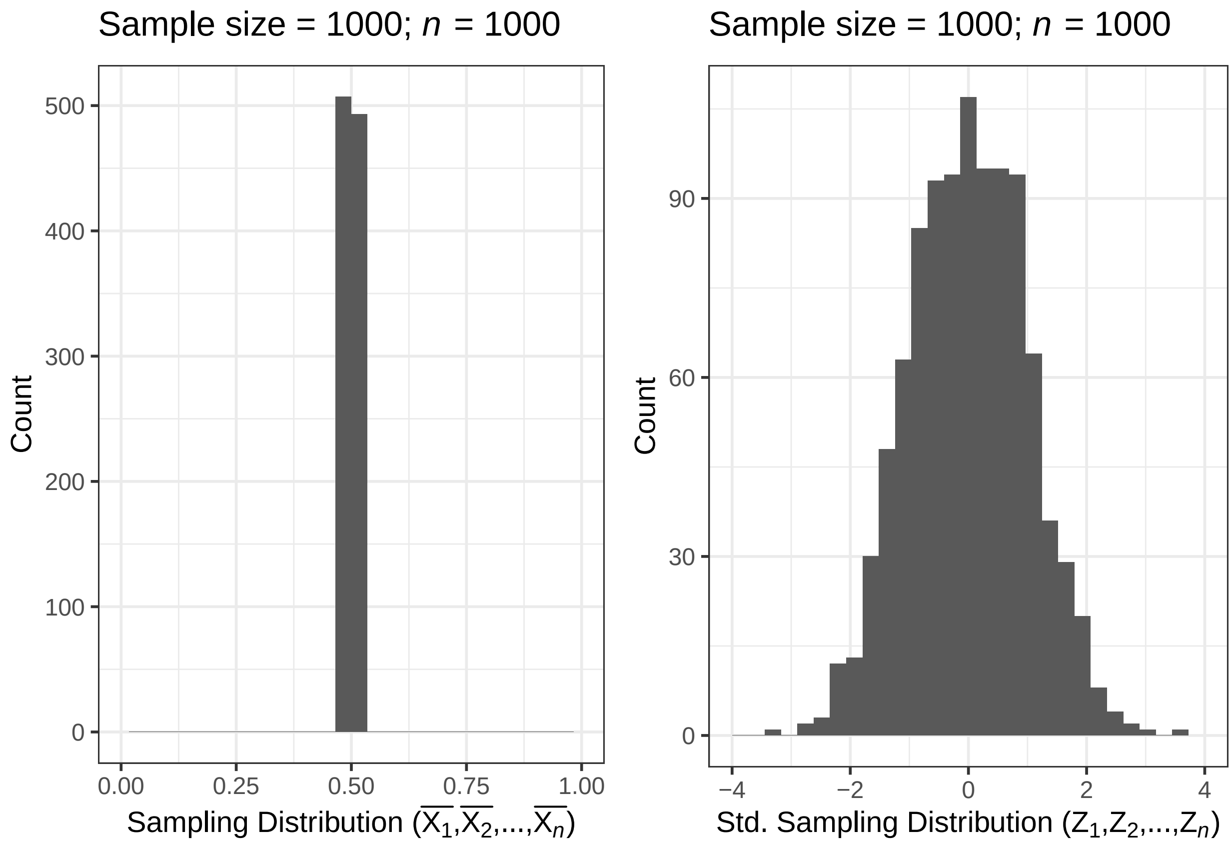


Sample size = 5; n = 1000
Count
90
60
30
0
0.00
0.25
0.50
0.75
1.00
Sampling Distribution (X̄1,X̄2,...,X̄n)



Sample size = 5; n = 1000
Count
90
60
30
0
−4
−2
0
2
4
Std. Sampling Distribution (Z1,Z2,...,Zn)


```
an(1000,1000)
```



Bibliography

Aronow, Peter B., and Benjamin T. Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge University Press.