Econometrics 1
Lecture 5: Large sample theory and statistical inference

ENSAE 2014/2015

Michael Visser (CREST-ENSAE)

Motivation

Until now we have studied what are called the finite sample (or small sample or exact) properties of OLS estimators:

- For instance, the unbiasedness of the OLS estimator is a finite sample property because it holds for any sample size n.
- Similarly, the normal sampling distribution of the OLS estimator is the exact sampling distribution which holds for any n.

The methods of statistical inference of last week's lecture (tests of hypotheses, confidence intervals) are based on the finite sample properties of OLS estimators.

In today's lecture we study the asymptotic (or large sample) properties of estimators (OLS estimators in particular). These are properties when the sample size tends to infinity. It is important to investigate large sample properties of estimators because:

Motivation (cntd)

- It is not always possible to find unbiased estimators. In those cases, we may settle for estimators that are consistent, meaning that, as $n \to \infty$, the distribution of the estimator collapses to the true parameter value. An example is the Instrumental Variables estimator studied later in the course (it is biased but consistent).
- Similarly, it is sometimes not possible to derive the finite sampling distribution of an estimator. We may then attempt to find the asymptotic sampling distribution, i.e., the sampling distribution of the estimator as $n \to \infty$.
- Even if we know the finite sample properties of OLS estimators, it may be of interest to know their asymptotic properties as these can be obtained under weaker assumptions (relatively to MLR.1-6).

Today's lecture also presents methods of statistical inference based on asymptotic properties of estimators: large sample statistical inference.

Consistency

Loosely speaking, an estimator of a parameter θ is consistent if its sampling distribution becomes more and more tightly distributed around the parameter when the sample size grows: as the size tends to infinity the distribution collapses to the single point θ. Formally:

Consistency

Let W_n be an estimator of θ based on a sample $Y_1, Y_2, \ldots, Y_n$. Then W_n is a consistent estimator of θ if $\forall \epsilon > 0$

$$\Pr(|W_n - \theta| > \epsilon) \to 0 \text{ as } n \to \infty$$

we write: $\text{plim } W_n = \theta$ or $W_n \xrightarrow{p} \theta$

Example: let $Y_1, Y_2, ..., Y_n$ be i.i.d. random variables with mean μ. Then

$$W_n \equiv \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y}_n$$

is a consistent estimator of μ: version of the law of large numbers.

Example: coin toss

Consider a coin toss:

$$Y_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

(and note that $E[Y] = p < \infty$)

Consider a sample of coin tosses $Y_1, \dots, Y_n$, then $W_n \equiv \overline{Y}_n$ is a consistent estimator of p. This is illustrated in the following simulation study.

Example: coin toss

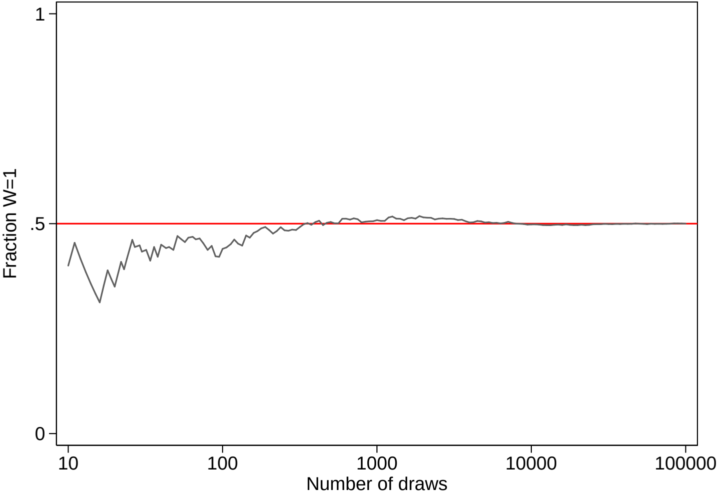
Generate a sequence of fair coin tosses:

```
. set obs 10
. gen toss = uniform()<0.5
. gen stoss = sum(toss)
. gen n = _n
. gen mtoss = stoss/n
. clist

         toss      stoss          n      mtoss
  1.        1          1          1          1
  2.        1          2          2          1
  3.        1          3          3          1
  4.        1          4          4          1
  5.        0          4          5         .8
  6.        1          5          6   .8333333
  7.        1          6          7   .8571429
  8.        0          6          8        .75
  9.        0          6          9   .6666667
 10.        1          7         10         .7
```

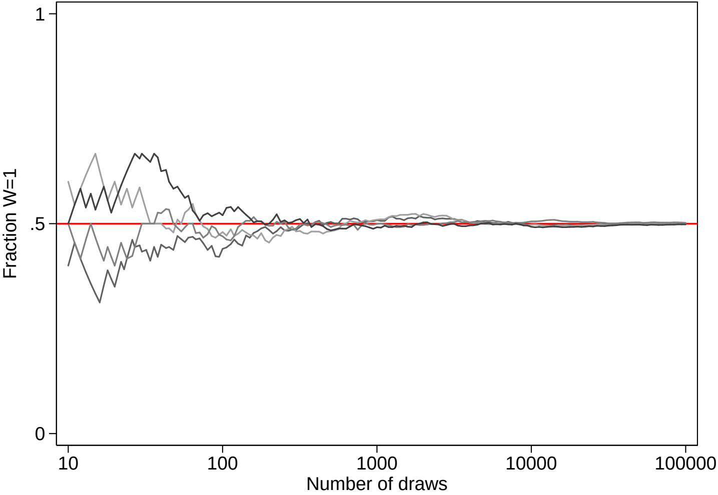
Example: coin toss

One sequence



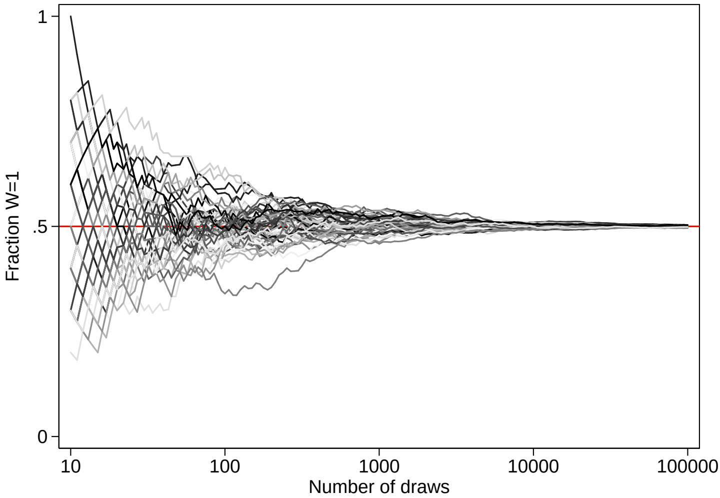
Example: coin toss

4 sequences



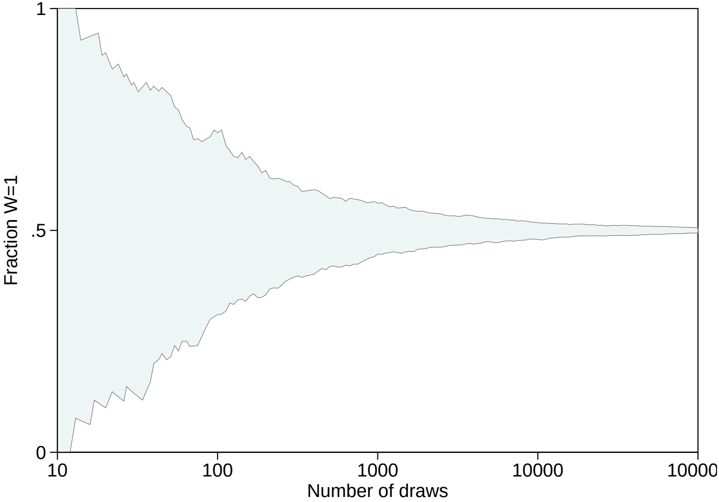
Example: coin toss

50 sequences



Example: coin toss

Many many (10000) sequences



Consistency of OLS estimator

Consider the MLR model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \tag{1}$$

Consistency of OLS estimator in model (1)

Under assumptions MLR.1-4, the OLS estimator $\hat{\beta}_j$ is consistent for β_j, $\forall j = 1, ., k$

Remark that the set of assumptions guaranteeing unbiasedness of the OLS estimator is the same as the set of assumptions implying consistency.

We prove this in the case $k = 1$ (SLR model). Focusing on $\hat{\beta}_1$:

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)^2}.$$

Using properties of probability limits, and repeated applications of the law of large numbers, we have

Consistency of OLS estimator (cntd)

$$\begin{aligned} plim(\hat{\beta}_1) &= \beta_1 + \frac{plim\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)u_i\right)}{plim\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)^2\right)} \\ &= \beta_1 + \frac{Cov(x_1,u)}{Var(x_1)} \end{aligned}$$

where the 1st equality follows from the property $plim(\frac{v_n}{w_n}) = plim(v_n)/plim(w_n)$. We have $Cov(x_1, u) = E(x_1 u) - E(x_1)E(u)$. MLR.3 implies $E(u) = 0$, and $E(x_1 u) = E(x_1 E(u|x_1)) = 0$. So $Cov(x_1, u) = 0$ and thus $plim(\hat{\beta}_1) = \beta_1$.

The arguments of the proof indicate that OLS estimators are consistent under MLR.1,2, and 4, and a weaker version of MLR.3, namely:

- MLR.3': Zero mean and zero covariance: $E(u) = 0$ and $Cov(x_j, u) = 0$, for $j = 1, ..., k$.

MLR.3' is weaker than MLR.3 because the latter implies the former, but not vice versa.

Inconsistency of OLS estimator

Suppose that the true population model is the following

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

and MLR.1-4 hold $\Rightarrow$ OLS estimators are consistent.

If we estimate instead $y = \beta_0 + \beta_1 x_1 + v$, (with $v = \beta_2 x_2 + u$), we will get an inconsistent estimate for β_1 if $\beta_2 \neq 0$ and x_1 and x_2 are correlated:

$$\text{plim } \hat{\beta}_1 - \beta_1 = cov(x_1, v)/var(x_1) = \beta_2 cov(x_1, x_2)/var(x_1)$$

- Similar to the omitted variable bias expression from Lecture 3
- We can think about the inconsistency as about the bias
- Again, in the general model with k regressors, if one x is correlated with u, then all OLS estimators will in general be inconsistent

Asymptotic sampling distribution of OLS estimator

Next we study the asymptotic sampling distribution of the OLS estimator. This is the sampling distribution of the estimator as the sample size tends to infinity.

Below we show that the OLS estimator has an asymptotic standard normal distribution.

Definition: a random variable Z_n has an asymptotic standard normal distribution if for each number z:

$$Pr(Z_n \leq z) \to \Phi(z) \text{ as } n \to \infty$$

where $\Phi(\cdot)$ is the standard normal distribution function. This is often written as: $Z_n \overset{a}{\sim} N(0, 1)$.

Example: suppose $Y_1, Y_2, ..., Y_n$ are i.i.d. random variables with mean μ and variance ρ^2. Then

$$Z_n \equiv \frac{\overline{Y}_n - \mu}{\rho/\sqrt{n}}$$

has an asymptotic standard normal distribution. This is a version of the central limit theorem.

Asymptotic normality of OLS estimator

Asymptotic Normality of OLS estimator in model (1)

Under assumptions MLR.1-5,

1. For $j \geq 1$ we have $\sqrt{n}(\hat{\beta}_j - \beta_j) \overset{a}{\sim} \mathcal{N}(0, \sigma^2/a_j^2)$ where σ^2/a_j^2 is the asymptotic variance of $\sqrt{n}(\hat{\beta}_j - \beta_j)$. Furthermore, $a_j^2 = \text{plim } \left(n^{-1}\sum_{i=1}^n \hat{r}_{ij}^2\right)$ and $\hat{r}_{ij}$ are the residuals from a regression of x_j on the other independent variables.
2. $\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^n \hat{u}_i^2$ is a consistent estimator of $\sigma^2 = var(u)$
3. For $j \geq 0$ we have

$$(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j) \overset{a}{\sim} \mathcal{N}(0,1) \tag{2}$$

where $se(\hat{\beta}_j)$ is the usual standard error of the OLS estimator $\hat{\beta}_j$. For $j \geq 1$ we have (see Lecture 4):

$$se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n \hat{r}_{ij}^2}}$$

Asymptotic normality of OLS estimator (cntd)

-Part (1) states the asymptotic normality of $\sqrt{n}(\hat{\beta}_j - \beta_j)$. The proof is sketched below for the simple linear regression case.
-Part (2) states that $\hat{\sigma}^2$ is not only an unbiased but also a consistent estimator.
-Part (3) states that the standardized OLS estimator $(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j)$ has asymptotically a $N(0,1)$ distribution.

It is important to note that this last result does not rely on MLR.6. In other words, the asymptotic standard normality of (2) holds for any distribution function of the error term (except that it must verify MLR.3 and MLR.5)!

Asymptotic normality of OLS estimator (cntd)

Proof

We give a heuristic proof of Part (1) for the case $k = 1$ ($y = \beta_0 + \beta_1 x + u$). We focus on $\hat{\beta}_1$. Since

$$a_1^2 = \text{plim}\ \left(n^{-1}\sum_{i=1}^{n}\hat{r}_{i1}^2\right) = \text{plim}\ \left(n^{-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right) = \sigma_x^2$$

we have to check that

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \overset{a}{\sim} \quad N(0, \sigma^2/\sigma_x^2)$$

We have

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\Sigma_i(x_i - \overline{x})(y_i - \overline{y})}{\Sigma_i(x_i - \overline{x})^2} \\
&= \frac{\Sigma_i(x_i - \overline{x})(\beta_1(x_i - \overline{x}) + u_i - \overline{u})}{\Sigma_i(x_i - \overline{x})^2} \\
&= \beta_1 + \frac{\Sigma_i(x_i - \overline{x})u_i}{\Sigma_i(x_i - \overline{x})^2}
\end{aligned}$$

Asymptotic normality of OLS estimator

Proof (cntd)

Therefore:

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_1 - \beta_1) &= \frac{\frac{1}{\sqrt{n}}\Sigma_i[x_i - \mu - (\overline{x} - \mu)]u_i}{\frac{1}{n}\Sigma_i(x_i - \overline{x})^2} \\
&= \frac{\frac{1}{\sqrt{n}}\left[\Sigma_i(x_i - \mu)u_i - \Sigma_i(\overline{x} - \mu)u_i\right]}{\frac{1}{n}\Sigma_i(x_i - \overline{x})^2} \\
&= \frac{\frac{1}{\sqrt{n}}\left[\Sigma_i(x_i - \mu)u_i - (\overline{x} - \mu)\Sigma_i u_i\right]}{\frac{1}{n}\Sigma_i(x_i - \overline{x})^2}
\end{aligned}
$$

where $\mu = E[x_i]$.
$\frac{1}{n}\Sigma_i(x_i - \overline{x})^2$converges in probability to $\sigma_x^2 \neq 0$ by the law of large numbers. Consequently, it is asymptotically equivalent to σ_x^2.

$\overline{x} - \mu$ converges in probability to 0 by the law of large numbers, and $\frac{1}{\sqrt{n}}\Sigma_i u_i$ converges to the distribution $N(0, \sigma^2)$ by the central limit theorem. Using a result in asymptotic theory, their product converges in probability to 0.

Asymptotic normality of OLS estimator

Proof (cntd)

Furthermore, $E\left((x_i-\mu)u_i\right)=0$, $Var\left((x_i-\mu)u_i\right)=\sigma_x^2\sigma^2$ (because of MLR.3 and MLR.5), and $(x_i-\mu)u_i$ are i.i.d. between observations. So the central limit theorem gives:

$$\frac{1}{\sqrt{n}}\Sigma_i(x_i-\mu)u_i \overset{a}{\sim} N(0,\sigma_x^2\sigma^2)$$

Since $plim[\frac{1}{\sqrt{n}}(\overline{x}-\mu)\Sigma_i u_i]=0$, it follows from a result in asymptotic theory that

$$\frac{1}{\sqrt{n}}\left[\Sigma_i(x_i-\mu)u_i-(\overline{x}-\mu)\Sigma_i u_i\right]$$

has the same asymptotic distribution as $\frac{1}{\sqrt{n}}\left[\Sigma_i(x_i-\mu)u_i\right]$. Therefore

$$\frac{1}{\sqrt{n}}\left[\Sigma_i(x_i-\mu)u_i-(\overline{x}-\mu)\Sigma_i u_i\right] \overset{a}{\sim} N(0,\sigma_x^2\sigma^2)$$

Asymptotic normality of OLS estimator

Proof (cntd)

Finally

$$\frac{\frac{1}{\sqrt{n}}\left[\Sigma_i(x_i-\mu)u_i-(\overline{x}-\mu)\Sigma_i u_i\right]}{\frac{1}{n}\Sigma_i(x_i-\overline{x})^2} \overset{a}{\sim} N\left(0,\frac{\sigma_x^2\sigma^2}{\sigma_x^4}\right)$$

$$\implies \sqrt{n}\left(\hat{\beta}_1-\beta_1\right) \overset{a}{\sim} N\left(0,\frac{\sigma^2}{\sigma_x^2}\right)$$

Asymptotic inference

Last week we showed that, under MLR.1-6, the standardized OLS estimator $(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j)$ follows a t_{n-k-1} for any sample size. Finite sample tests of hypotheses about parameters (the t tests) and confidence intervals were based on this result.

Part (3) above indicates that, under MLR.1-5, the standardized OLS estimator $(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j)$ follows asymptotically a $N(0,1)$. Large sample tests and confidence intervals can be based on this result, exactly as in the finite sample setting.

When the sample size is large, the difference is not important. This is because a t_{n-k-1} distribution approaches a $N(0,1)$ distribution for large n.

Asymptotic inference (cntd)

Therefore, in practice, when the sample size is large, large sample statistical inference (testing hypotheses about parameters, construction of confidence intervals) is approximately equivalent to finite sample statistical inference.

It can be shown that the asymptotic normality of OLS estimators also implies that the usual F statistics have approximate F distributions in large samples.

Therefore, when the sample size is large, for testing exclusion restrictions or other multiple hypotheses, one can proceed as in the theory of finite sample statistical inference.

Asymptotic efficiency of OLS estimator

Consider the SLR model

$$y = \beta_0 + \beta_1 x + u$$

the zero conditional mean $E[u|x] = 0$ can be used to construct other consistent estimators. It implies $E[u|z] = 0$ where $z = g(x)$, and thus we have $E[u] = 0$ and $E[zu] = 0$. Solving the two corresponding sample moment conditions we get (check this)

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})y_i}{\sum_{i=1}^{n}(z_i - \overline{z})x_i} = \beta_1 + \frac{\sum_{i=1}^{n}(z_i - \overline{z})u_i}{\sum_{i=1}^{n}(z_i - \overline{z})x_i}$$

$\tilde{\beta}_1$ is a consistent estimator of β_1 since the law of large numbers tells us that the numerator and denominator converge in probability to $cov(z, u)$ and $cov(z, x)$ respectively:

$$\text{plim } \tilde{\beta}_1 = \beta_1 + cov(z, u)/cov(z, x) = \beta_1$$

Asymptotic efficiency of OLS estimator

It can be shown that (like above)

$$\sqrt{n}(\tilde{\beta}_1 - \beta_1) \sim \mathcal{N}(0, \sigma^2 var(z)/cov(z,x)^2)$$

The inequality $cov(z,x)^2 \leq var(z)var(x)$ (this follows from the definition of the correlation coefficient between x and z) implies that $avar\sqrt{n}(\hat{\beta}_1 - \beta_1)$ (=$\sigma^2/var(x)$) is no larger than $avar\sqrt{n}(\tilde{\beta}_1 - \beta_1)$. This is in fact a special case of a general result:

Asymptotic Efficiency of OLS estimator in model (1)

Under assumptions MLR.1-5, $\forall j = 1, ..., k$:

$$avar(\sqrt{n}(\hat{\beta}_j - \beta_j)) \leq avar(\sqrt{n}(\tilde{\beta}_j - \beta_j))$$

where $\tilde{\beta}_j$ solve the sample counterparts of the theoretical moments $E[g_j(x_1, ..., x_k)u] = 0 \text{ for } j = 0, 1, ..., k$.

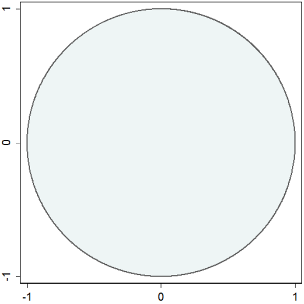
Monte Carlo Simulation

Monte Carlo methods (or Monte Carlo experiments) are a class of computational algorithms that rely on repeated random sampling to approximate formulas whenever it is not possible to have explicit expressions of the formulas (multidimensional integrals for instance).

A well known application of a Monte Carlo simulation: approximation of π.

Consider uniform rain on the square $[-1, 1] \times [-1, 1]$. Probability that a rain drop falls into the circle is

$$\Pr(\text{drop in circle}) = \frac{\pi}{4} \equiv p$$



Monte Carlo Simulation

Example

If we knew π, we could compute $Pr(\text{drop in circle}) = \pi/4$. Suppose we do not, and use a Monte Carlo simulation to approximate π.

Consider n independent raindrops, then the number of rain drops Z_n falling in the circle is a binomial random variable:

$$Z_n \sim B(n; p)$$

We can estimate p by

$$\hat{p} = \frac{Z_n}{n}$$

Thus we can estimate π by

$$\hat{\pi} = 4\hat{p} = 4\frac{Z_n}{n}$$

Monte Carlo Simulation
Example

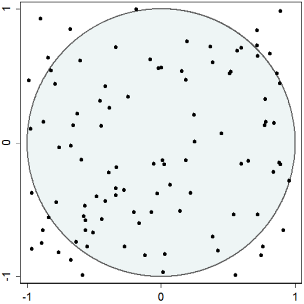
Result obtained for n = 100 raindrops: 93 points inside the dark circle

$$\hat{\pi} = \frac{4Z_n}{n} = \frac{4 \cdot 93}{100} = 3.76$$

(rather poor estimate).
However: the law or large numbers guarantees that

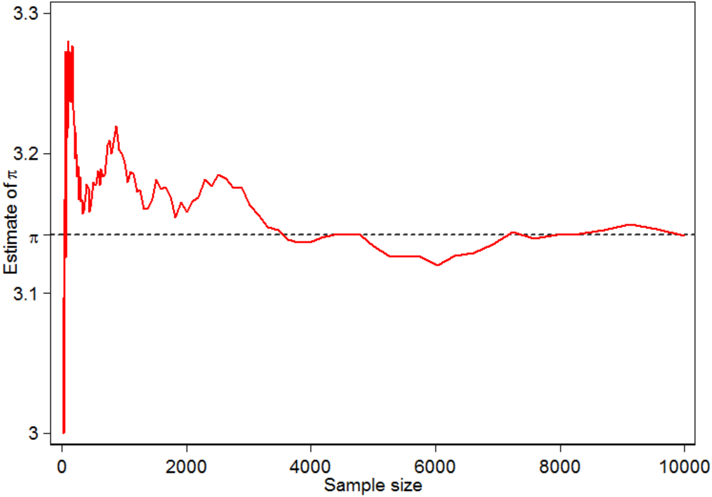
$$\hat{\pi} = 4\frac{Z_n}{n} \to \pi$$

in probability for $n \to \infty$



Monte Carlo Simulation

Example



Monte Carlo Simulation
Example

Recall the two steps used in our example

1. We have written the quantity of interest as an expectation:

$$\pi = 4 \cdot P(\text{drop in circle}) = E[4 \cdot 1_{\{\text{drop in circle}\}}]$$

2. We have replaced this algebraic representation of the quantity of interest by its sample approximation
   - Law of large numbers guarantees sample approximation converges to the algebraic representation

Monte Carlo Simulation

In econometrics, Monte Carlo simulations are used to investigate the properties of estimators, tests, and confidence intervals. The principle of the simulations is straightforward:

- Generate a sample (size N) using your population model (the data generating process)
- Calculate your estimates/statistics
- Repeat this M times

You can now use the empirical distribution of your estimates/statistics to verify questions like the following:

- are estimates consistent when $N \to \infty$?
- do the empirical standard errors of my estimates correspond to the theoretical standard errors?
- do rejecting levels of statistics match the rejection levels implied by their theoretical distributions?

I understand/can apply...

- The first four Gauss-Markov assumptions imply that OLS is consistent
- The first five Gaus-Markov assumptions imply that the OLS estimator is asymptotically normally distributed
- The methods of testing and constructing confidence intervals that we learned are approximately valid in large samples
- The basics of Monte Carlo simulations