

Law of Large Numbers

The sample mean is itself a random variable

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

If $E[X_i] = \mu$, $E[\bar{X}_n] = \mu$

and if $V[X_i] = \sigma_i^2$

$$P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$$

LLN Explained

$$P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$$

Pick any number $\epsilon > 0$

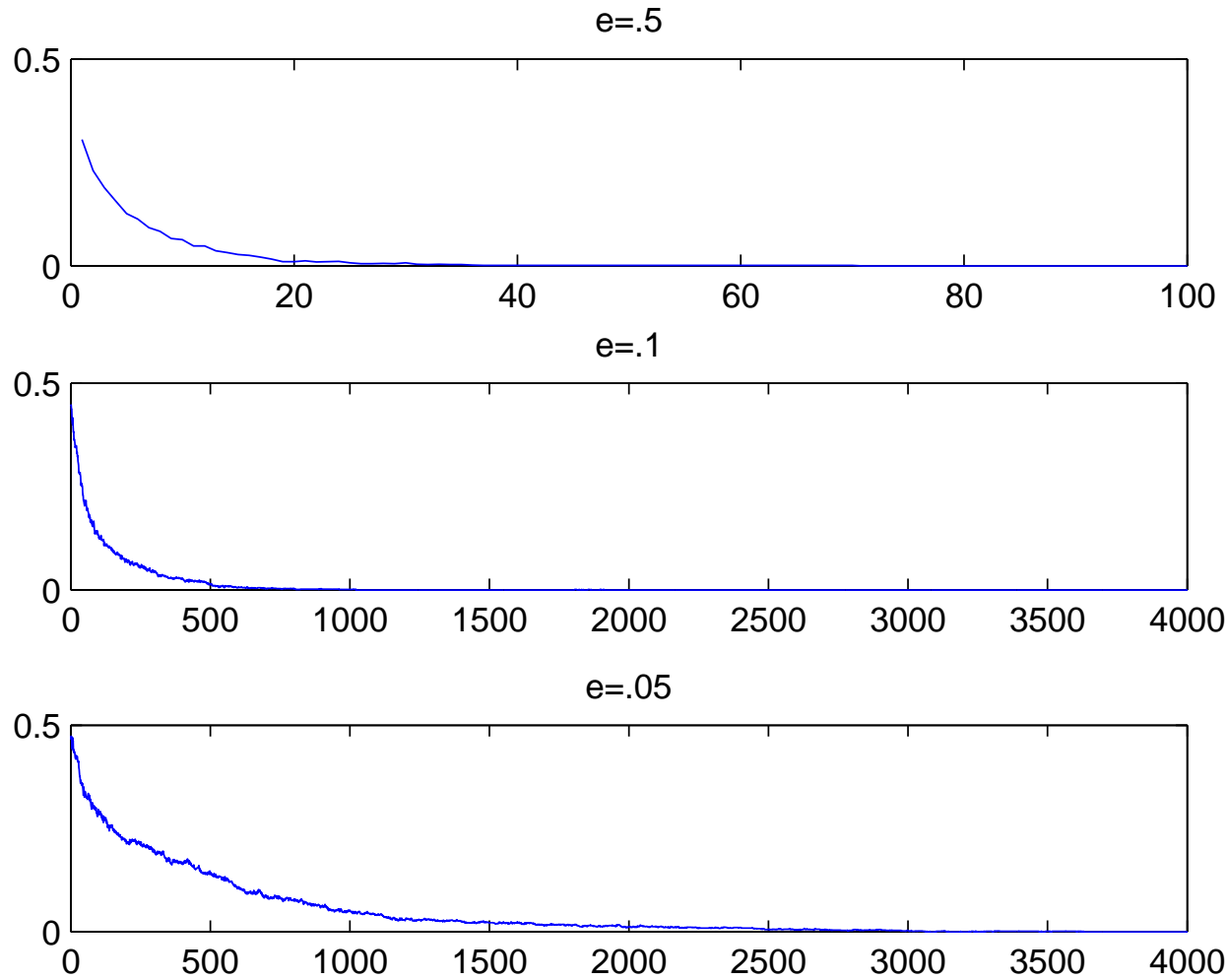
As we increase the sample, the probability that the sample mean is at most ϵ in error goes to zero.

Choose any desired tolerance ϵ

Choose any desired probability $\delta > 0$ for failure to meet the tolerance

LLN says we can find a sample size so that the probability of failure is less than δ

LLN Illustrated



What the LLN is for

The LLN isn't a practical result

The LLN is a comforting result

The LLN allows us to hope

It allows us to approximate the theoretical mean with the sample mean

The approximation gets better as the sample size increases

More generally, we can approximate all distribution moments by their sample moments

The LLN does not tell us how good our sample mean is.

Generalization of means

Almost any quantity of interest can be made into a mean
For a sample X_1, \dots, X_n , consider estimating $P[X \leq x]$

Define a new random variable to suit our purpose
 $Z_i = 1$ if $X_i \leq x$, 0 otherwise

As $n \rightarrow \infty$, $\bar{Z}_n \rightarrow P[X \leq x]$

Repeat for all x , and we can estimate the cdf everywhere
This is the empirical cdf.

Central Limit Theorem

Let X_1, X_2, \dots, X_n be a sequence of independent random variables with mean $E[X_i] = \mu$ and variance $V[X_i] = \sigma^2$.

$$\text{Let } Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/N}}$$

As n increases, the pdf of Z approaches the standard normal

Use of CLT

In practice, we only have one sample, one mean

Our one sample mean will not be the true mean

How can we say something about the true mean?

The LLN allows us to hope that the sample mean approximates μ

How do we know how well we did?

The CLT allows us to make a guess through approximation

It does not matter what distribution we draw from, if the sample is large enough, the sample mean will have a normal distribution

Measurement Error Example

$N = 25$, $E[X_i] = \mu$, $V[X_i] = 1$

Suppose we want to know the probability that \bar{X}_n is at most .25 from μ

$$\begin{aligned} P(|\bar{X} - \mu| < .25) &= P(-.25 < \bar{X} - \mu < .25) \\ &= P\left(-\frac{.25}{\sqrt{1/25}} < \frac{\bar{X} - \mu}{\sqrt{1/25}} < \frac{.25}{\sqrt{1/25}}\right) \\ &\approx \Phi\left(\frac{.25}{\sqrt{1/25}}\right) - \Phi\left(-\frac{.25}{\sqrt{1/25}}\right) \\ &= \Phi(.25 * 5) - \Phi(-.25 * 5) \\ &= .8944 - .1056 = .7888 \end{aligned}$$