"The Law of Large Numbers is the reason why statistics works."

There is a remarkable fact called the "Law of Large Numbers," which assures us that, no matter what distribution *X* has, if we could make an endless sequence of observations on independent random variables that have the same distribution as *X* (and provided that *X* has an expected value), we would find that

$$\frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \to E(X) \quad as\ n \to \infty .$$

As the sample size grows, the sequence of sample means, $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \ldots$, converges to the expected value, E(X) (the "mean of the distribution").

This means that the average of a large number of independent observations will, with high probability, be close to the expected value.
(Or when the sample is large, the probability that the sample mean will fall close to the mean of the distribution is high.)

To illustrate the Law of Large Numbers, let's generate a long sequence of independent Bernoulli random variables, and see what happens. I made my computer do this, using success probability θ = 0.75. The first 25 observations:

11010 11111 11001 01110 01010

Thus the observed sequence of sample means is

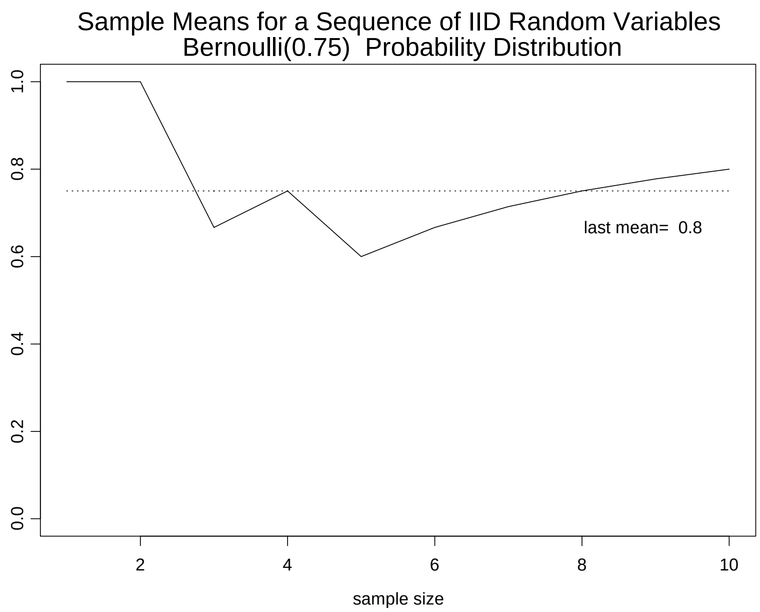
1 , 2/2, 2/3, 3/4, 3/5, 4/6, 5/7, 6/8, 7/9, 8/10, ...

$$x_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6, \cdots$$

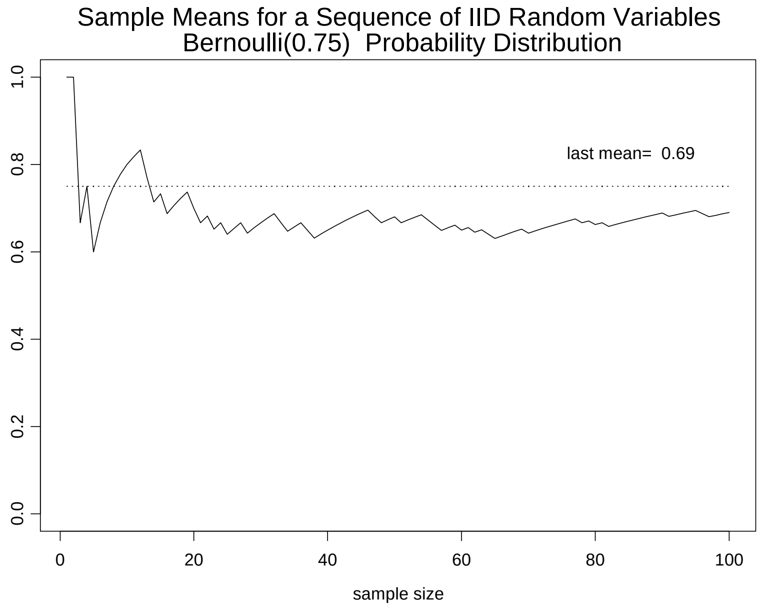
or

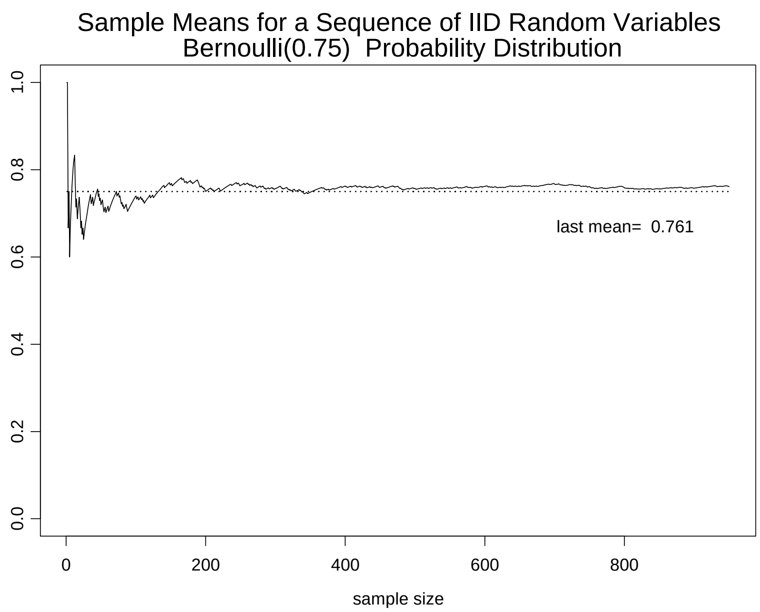
$$\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5, \hat{p}_6, \ldots$$

If we plot more of the means we see...

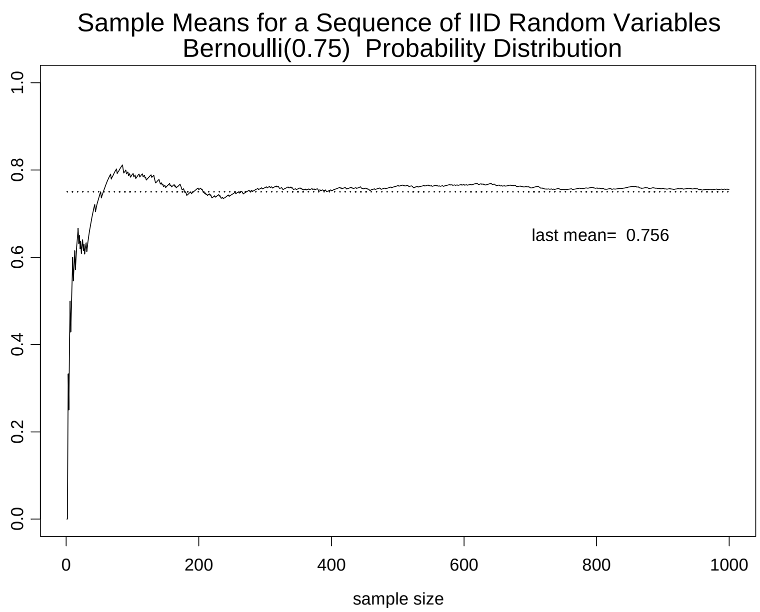

Sample Means for a Sequence of IID Random Variables
Bernoulli(0.75) Probability Distribution
last mean= 0.8
sample size
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10

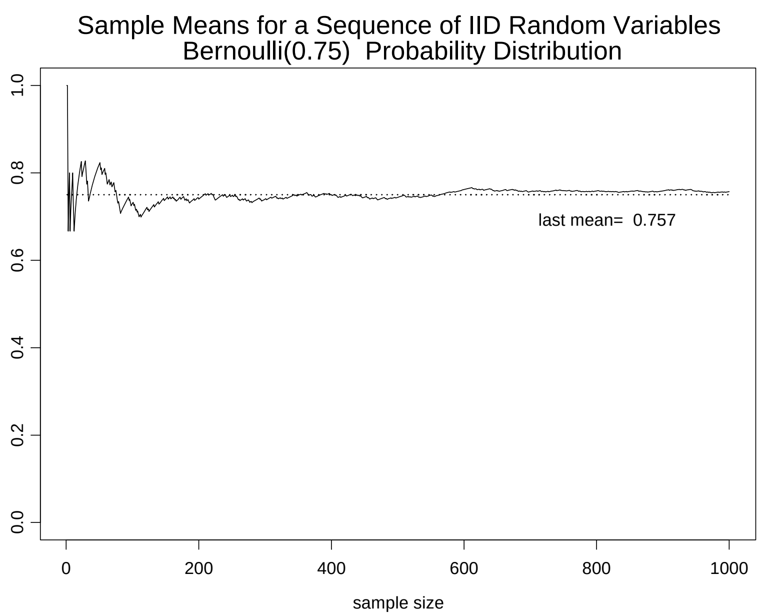
If we plot more means we see



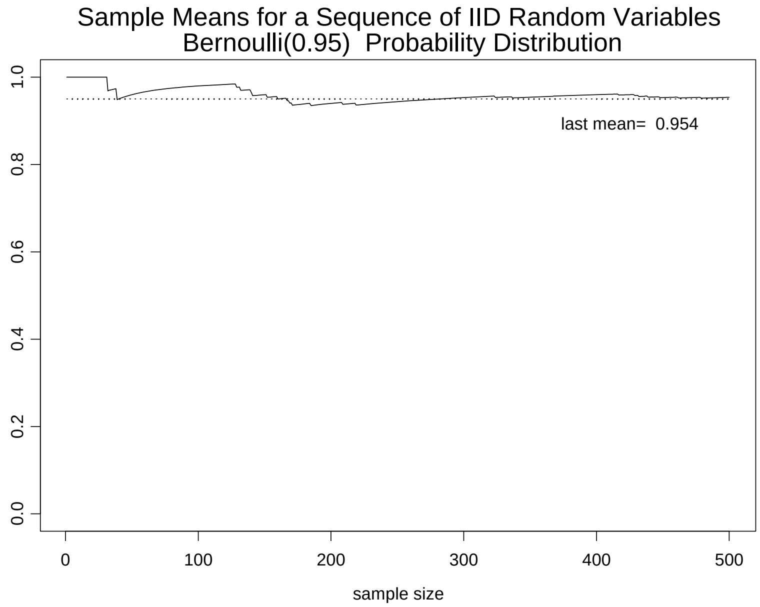


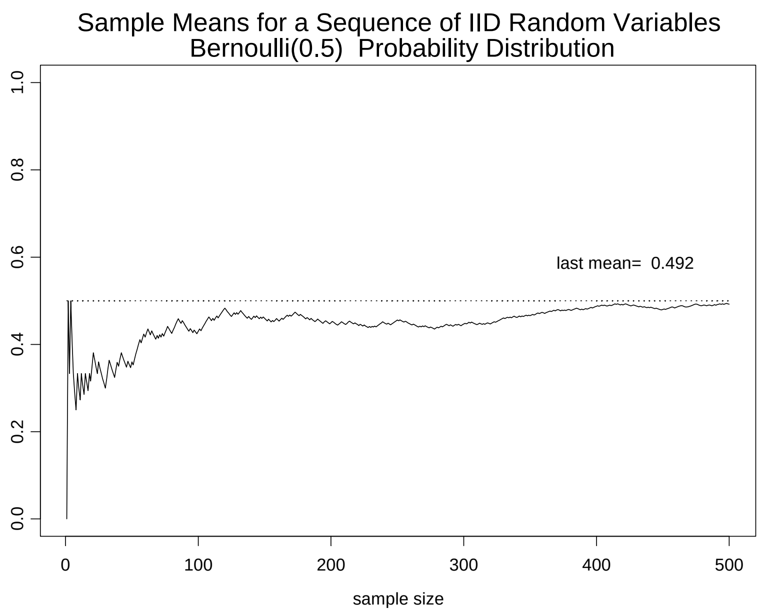
Here are the results for two more sequence of 1000 Bernoulli(0.75) trials:





What is different in the sequence of numbers between these two graphs?





$X \sim ber(0.95)$ then $Var(X)=0.95(0.05)=0.0475$

$Y \sim ber(0.5)$ then $Var(Y)=0.5(0.5)=0.25$

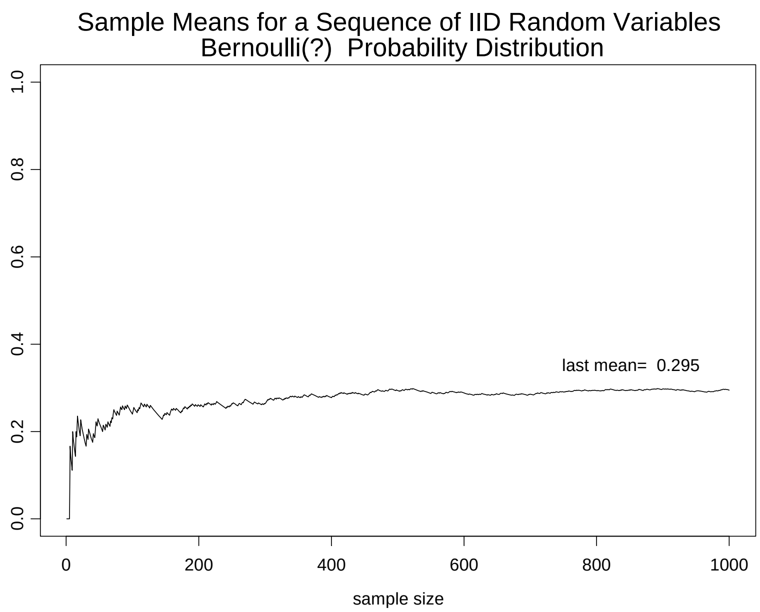
When the sample is large, the probability that the sample mean will be close to the expected value (or the "mean of the distribution") is high.

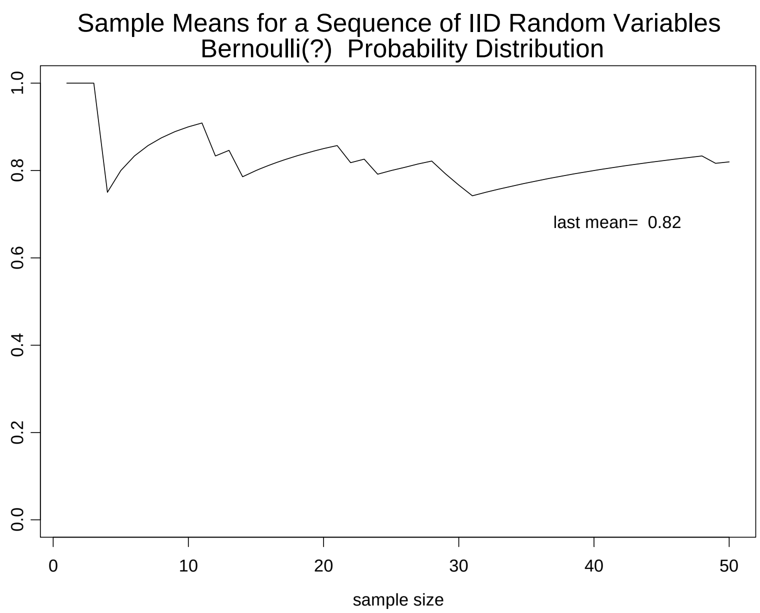
For the Bernoulli distribution, the expected value is the success probability, θ. Since the sample mean is the proportion of successes, what the Law of Large Numbers says, in this case, is

> ***the proportion of successes in the sample will be close to the probability of success, θ, with high probability.***

This is the justification for using the proportion of successes in a sample as an estimate for the true probability of success.

Here is a sequence of means of independent Bernoulli random variables. Can you tell what value of θ I used?





The Law of Large Numbers applies to all probability distributions that have expected values. This includes all of those that we have come across — Bernoulli, Binomial, Poisson, Uniform, and Normal.

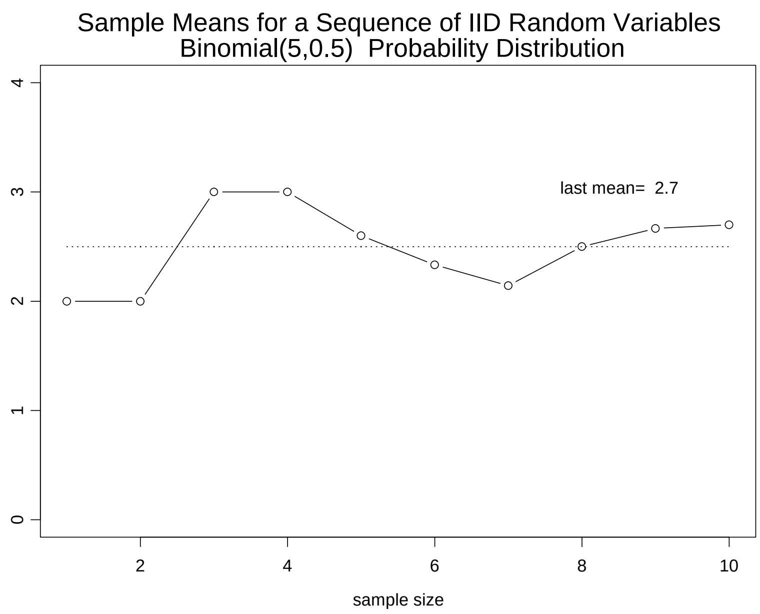
Here is a sequence of observations on independent <u>Binomial</u> random variables with n = 5 and θ = 1/2. The expected value is $E(X) = n\theta = 2.5$.

2 , 2 , 5 , 3 , 1 , 1 , 1 , 5 , 4 , 3 , ...

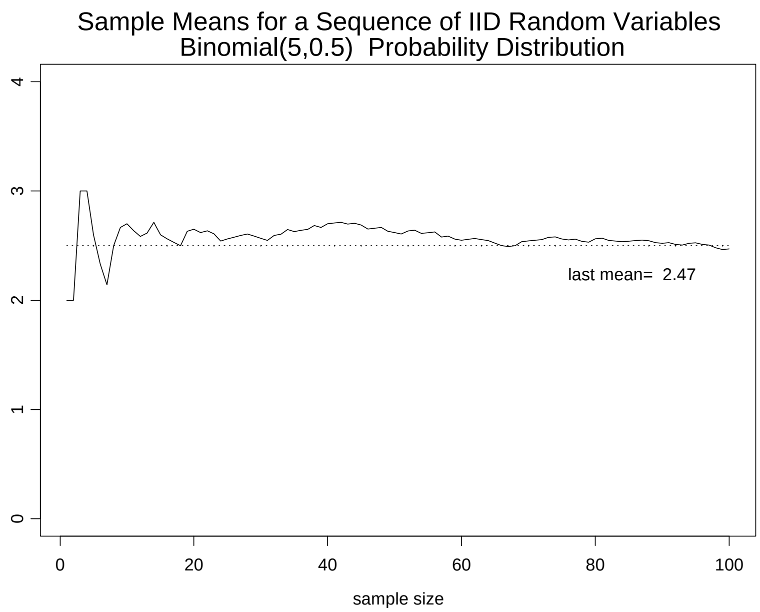
The first ten sample means are :

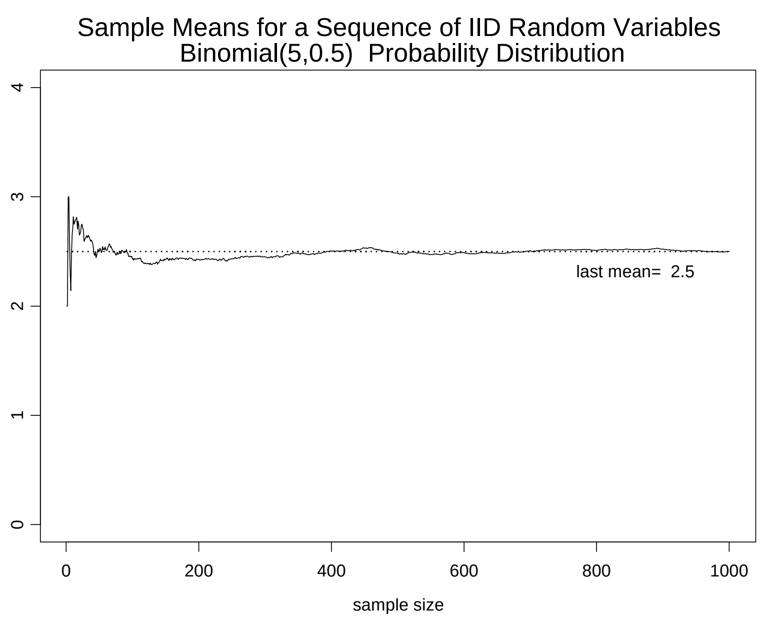
2 , 4/2, 9/3, 12/4, 13/5, 14/6, 15/7, 20/8, 24/9, 27/10, ...

2 , 2 , 3 , 3 , 2.6 , 2.33, 2.14 , 2.5 , 2.67, 2.7 , ...

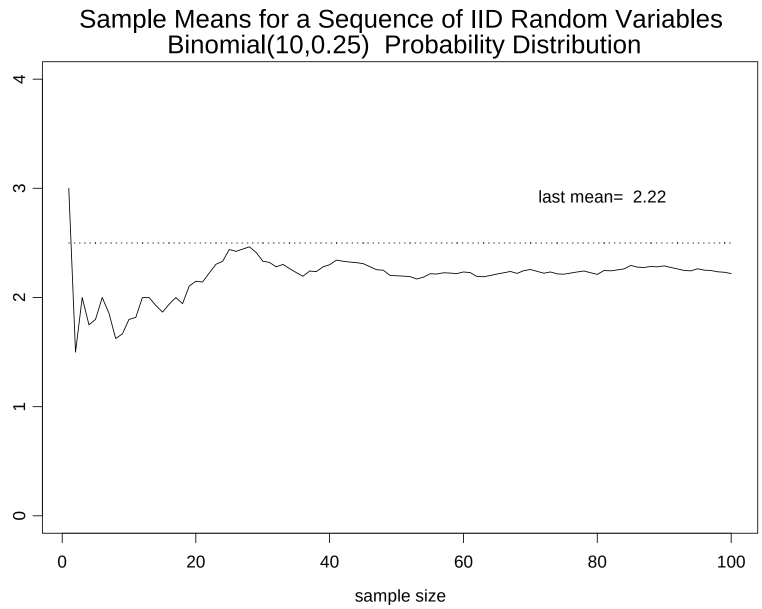


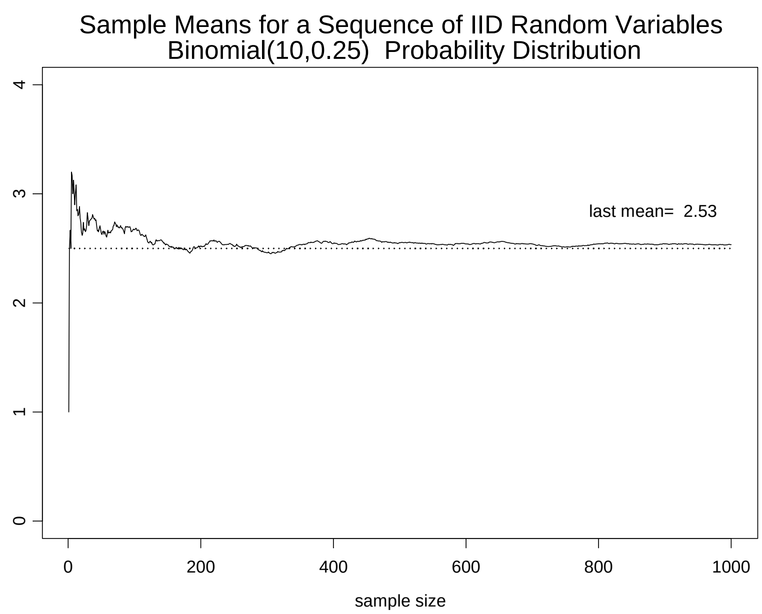
And as the sample gets larger:



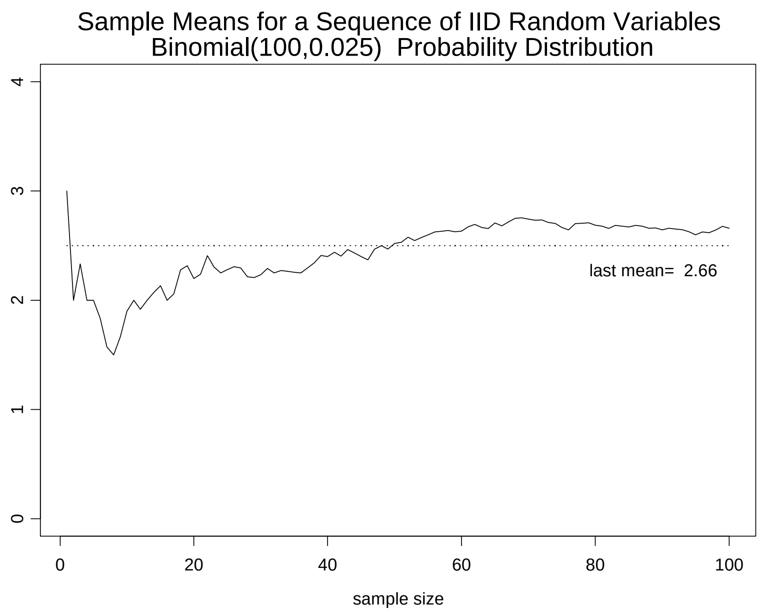


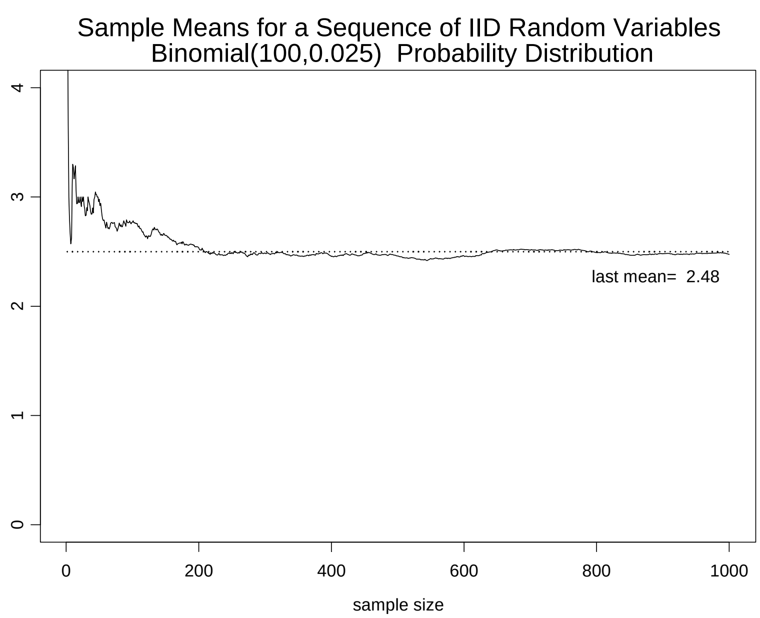
Here's what happened when I sample from a Binomial distribution with n=10 and θ=1/4. Again E(X)=2.5.





Here's what happened when I sample from a Binomial distn with n=100 and θ=0.025. Again E(X)=2.5.





The Poisson(λ) distribution has mean λ. What happens if we set λ = 2.5, and make a sequence of independent observations on this distribution?

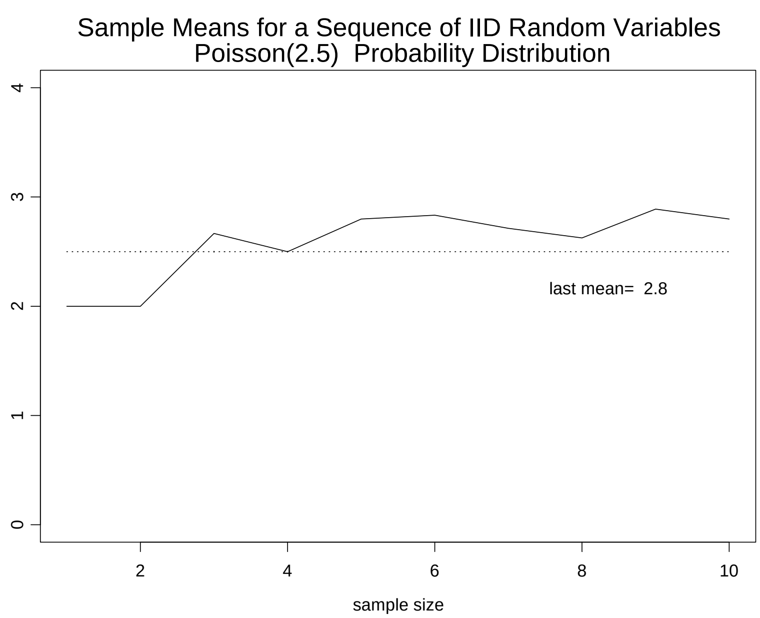
The first few observations are

2 , 2 , 4 , 2 , 4 , 3, 2, 2, 5, 2, ...

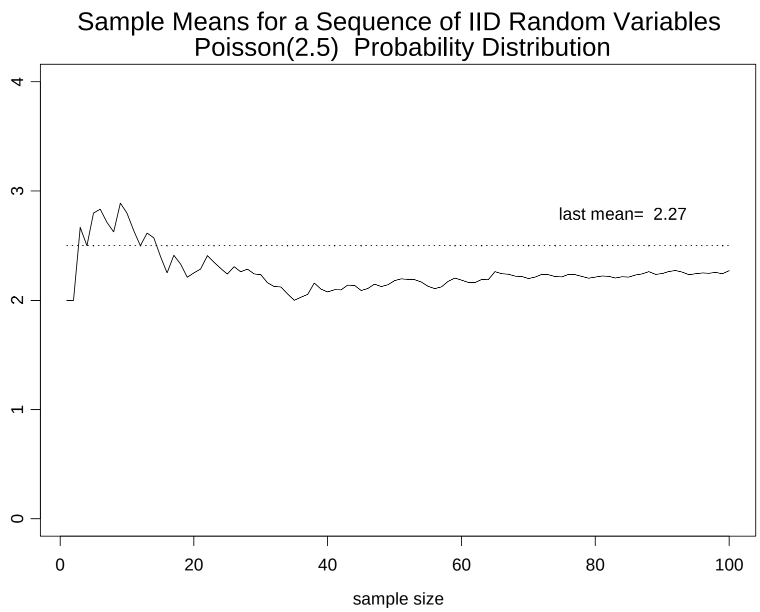
and the means are

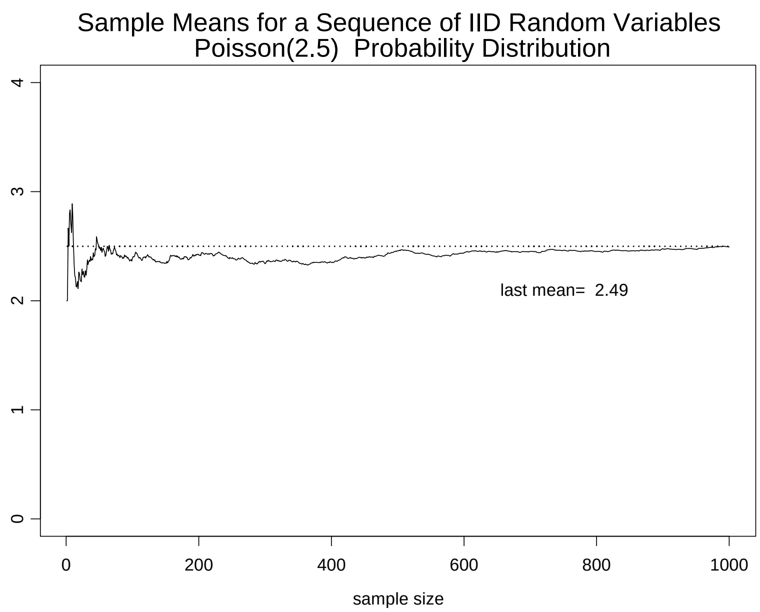
2 , 2 , 8/3, 2.5, 2.8, 2.833, ...

$x_1, \; \bar{x}_2, \; \bar{x}_3, \; \bar{x}_4, \; \bar{x}_5, \; \bar{x}_6, \ldots$



And if the sample size is increased?

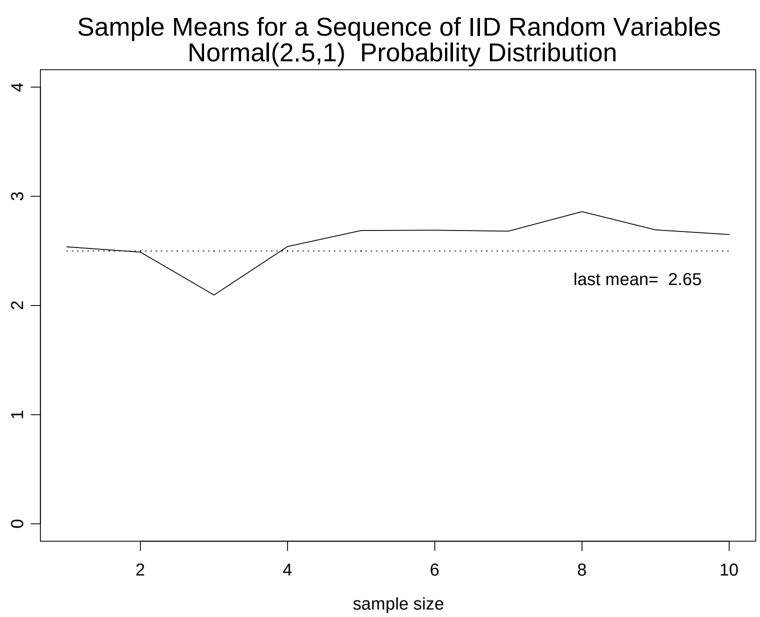




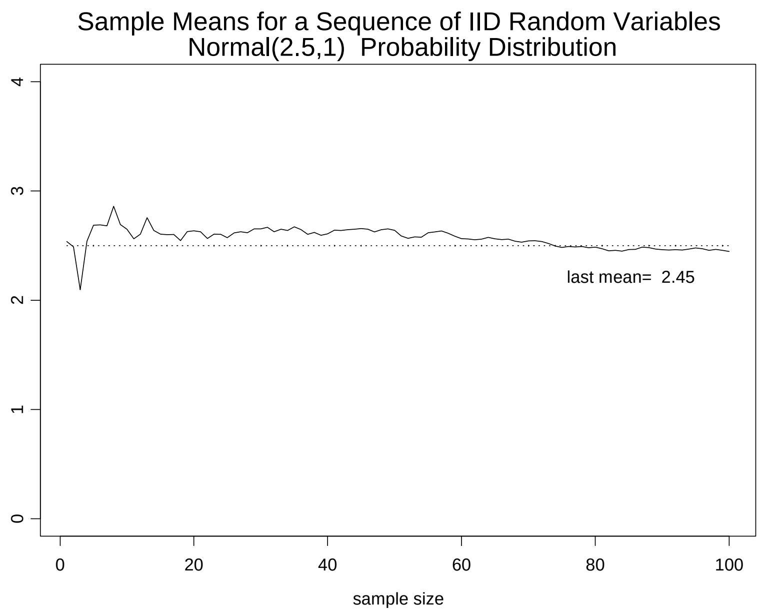
How about the normal probability distribution? Here is what happened when I observed 1000 independent normal random variables with mean $\mu = 2.5$ and standard deviation $\sigma = 1$. The first few observations were:

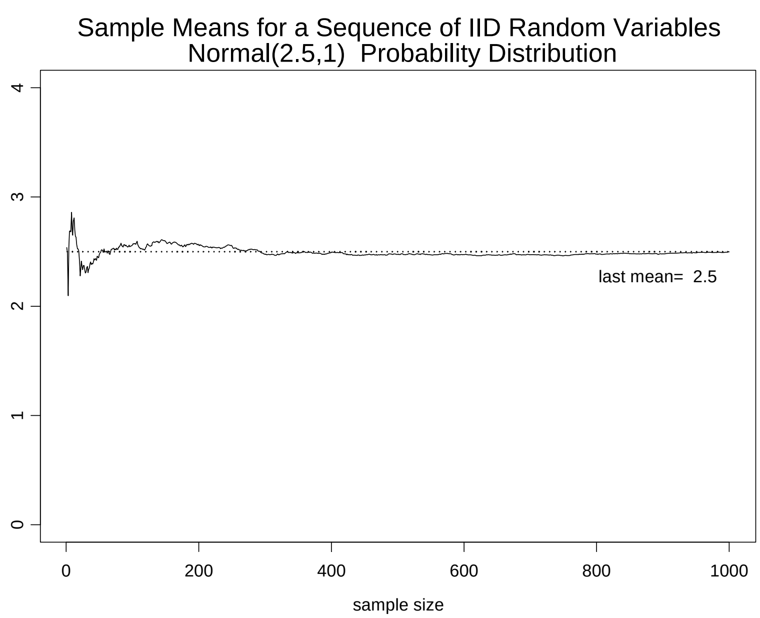
2.536691, 2.442423, 1.313549, 3.870355, 3.271964, ...

This plot shows the first ten sample means, $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{10}$



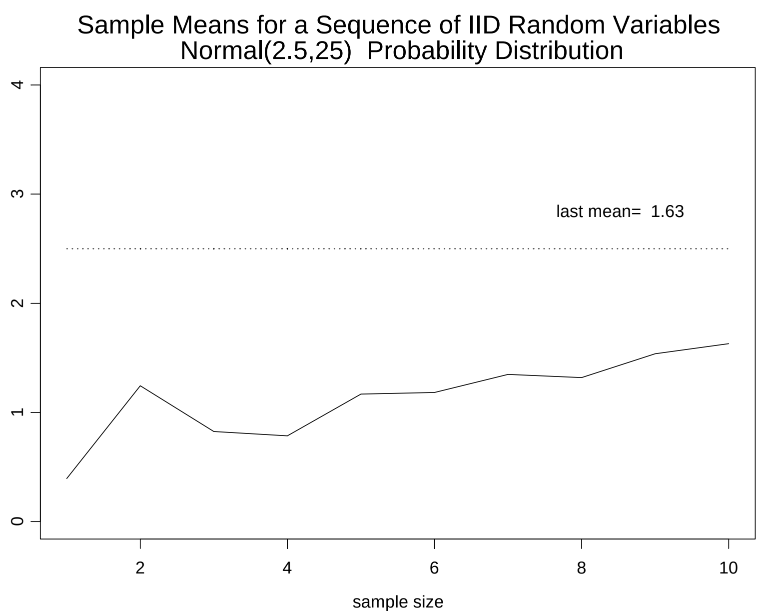
And as the sample size increases:



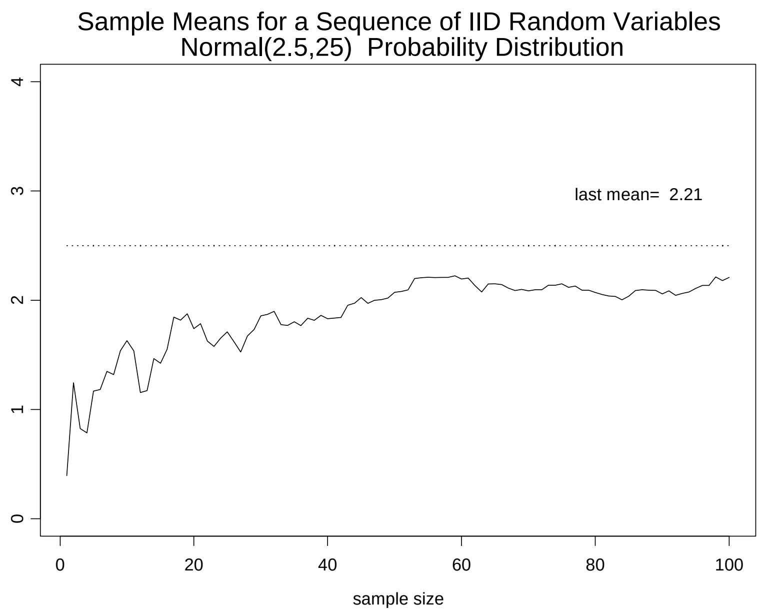


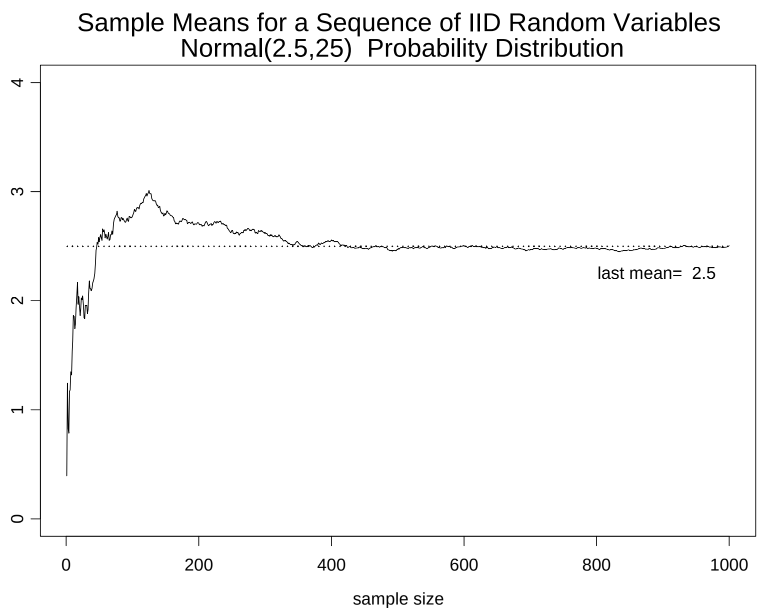
What if the standard deviation is increased? Here is what happened when I generated 1000 independent normal random variables with $\mu = 2.5$ and $\sigma = 5$.

Here's a plot showing the first ten <u>means</u>:



And as the sample size increases:





Remember that the Law of Large Numbers always applies:

if $X_1, X_2, X_3, \ldots$ are independent random variables that all have the same probability distribution as *X*, then $X_1^2, X_2^2, X_3^2, \ldots$ are independent random variables that all have the same probability distribution as X^2, so the Law of Large Numbers ensures that their average will approach the expected value of X^2 as the sample size grows:

$$\frac{\sum_{i=1}^{n} X_i^2}{n} \rightarrow E(X^2)$$

And in general

$$\frac{\sum_{i=1}^{n} g(X_i)}{n} \rightarrow E\left(g(X)\right)$$

Demonstration:
I generated 1000 independent random variables, $X_1, X_2, X_3, \ldots$ using a binomial distribution with n = 5 and θ = 0.5. We know that the sequence of averages converges to E(*X*) = 2.5.

Let's see what happens to the sequence of averages of the squared values:

$$\sum_{i=1}^{n} X_i^2 / n.$$

Also we've seen that the expected value of X^2 is

$E(X^2)=n\theta(1-\theta) + n^2\theta^2 = 5(0.5)(0.5) + 25(0.5)(0.5) = \mathbf{7.5}$.

Here are the first ten observations are :

3 , 3 , 4 , 3 , 0 , 3 , 5 , 2 , 3 , 3 , ...

And the sample means are

3 , 6/2, 10/3, 13/4, 13/5, 16/6, 21/7, 23/8, 26/9, 29/10,...

The entire sequence of sample means looks like those we saw before, converging to the expected value of *X*, 2.5.

The <u>squares</u> of the first ten observations (the observed values of X^2) are :

$$9 , 9 , 16 , 9 , 0 , 9 , 25 , 4 , 9 , 9 , \ldots$$

And *their* averages are

$$9 , 9 , 34/3 , 43/4 , 43/5 , 52/6 , 72/7 , 81/8 , 90/9 , 99/10 , \ldots$$

If we continue with this process and see what happens to <u>these</u> averages as the number of observations grows, we find