STA257

Neil Montgomery | HTML is official | PDF versions good for in-class use only
Last edited: 2016-12-04 09:45

some sums of r.v.s

If $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p), then $\sum X_i \sim \text{Binomial}(n, p)\ldots$

If $X_1, \ldots, X_n$ are i.i.d. Geometric(p), then $\sum X_i \sim \text{NegBin}(n, p)\ldots$

This is fundamentally a "lookup table" technique.

Others (exercises):

- sum of n independent Exp(λ) is Gamma(n, λ)
- sum of n independent Poisson(λ) is Poisson$(n\lambda)$
- sum of $X_i \sim \text{Binomial}(n_i, p)$ is Binomial$\left(\sum n_i, p\right)$
- distribution of sum of $X_i \sim \text{Binomial}(n_i, p_i)$ (different p_i!) cannot be determined using mgf technique.

the normal distributions

First, suppose X has mgf $m_X(t)$ and $Y = a + bX$. What is $m_Y(t)$?

So what is the mfg of a general $N(\mu, \sigma^2)$?

Finally, if $X_1, \ldots, X_n$ are independent with $X_i \sim N(\mu_i, \sigma_i^2)$?, what distribution is $X = \sum X_i$?

sequences of random variables, convergence

(optional background) sequences of functions

Depending on your background, you might have heard of:

- pointwise convergence $f_n(x) \to f(x)$ (converges for every x)
- uniform convergence (convergence happens all at the same rate

Uniform convergence is stronger and has benefits - you can pass limits, derivatives, integrals, etc. through uniform convergence with no problem.

In this course we have sometimes magically passed these things along with the $E()$ operator through *power series*, because power series converge *uniformly* inside their radius of convergence.

But don't worry if you've never heard of this or forgot it all.

sequences of random variables

Very common in probability and statistics. We have seen some already:

As a model for a variable in a dataset we have considered the "i.i.d." sequence $X_1, X_2, \ldots, X_n$.

I have introduced the notion of "Sample Average" $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

When we derived Poisson from Binomial, we (implicitly) considered a sequence $X_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$ and wondered about $n \to \infty$.

We're going to wonder again about $n \to \infty$

Again, with random variables we care most about probabilities and not their actual values.

Case 1: getting closer to a constant, probably

Consider $X_1, X_2, \ldots, X_n$ i.i.d. with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, and consider $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

From last week:

$$E\left(\overline{X}_n\right) = \mu \qquad \text{and} \qquad \mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}$$

What happens when n gets bigger?

simulation example 1 - Bernoulli(0.5)

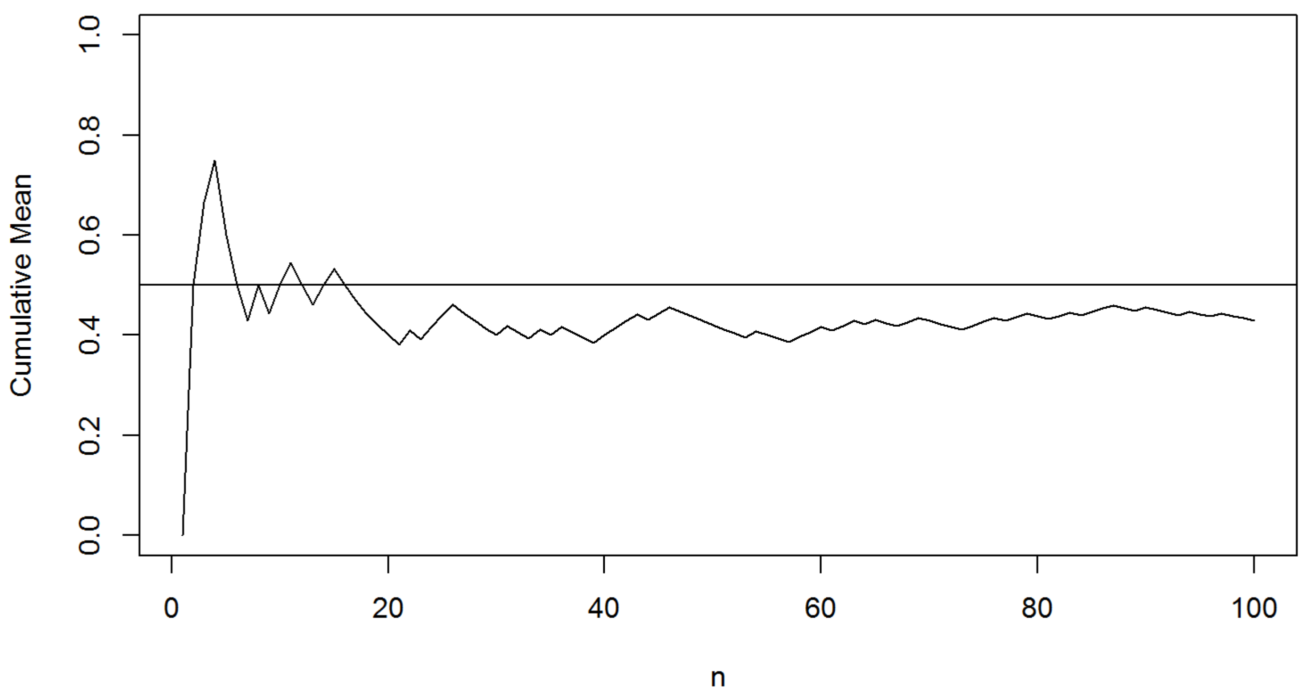
My computer can pretend to observe Bernoulli random variables. Here are $n = 30$ Binomial(1,0.5) simulations:

```
rbinom(n=30, size=1, prob=0.5)
```

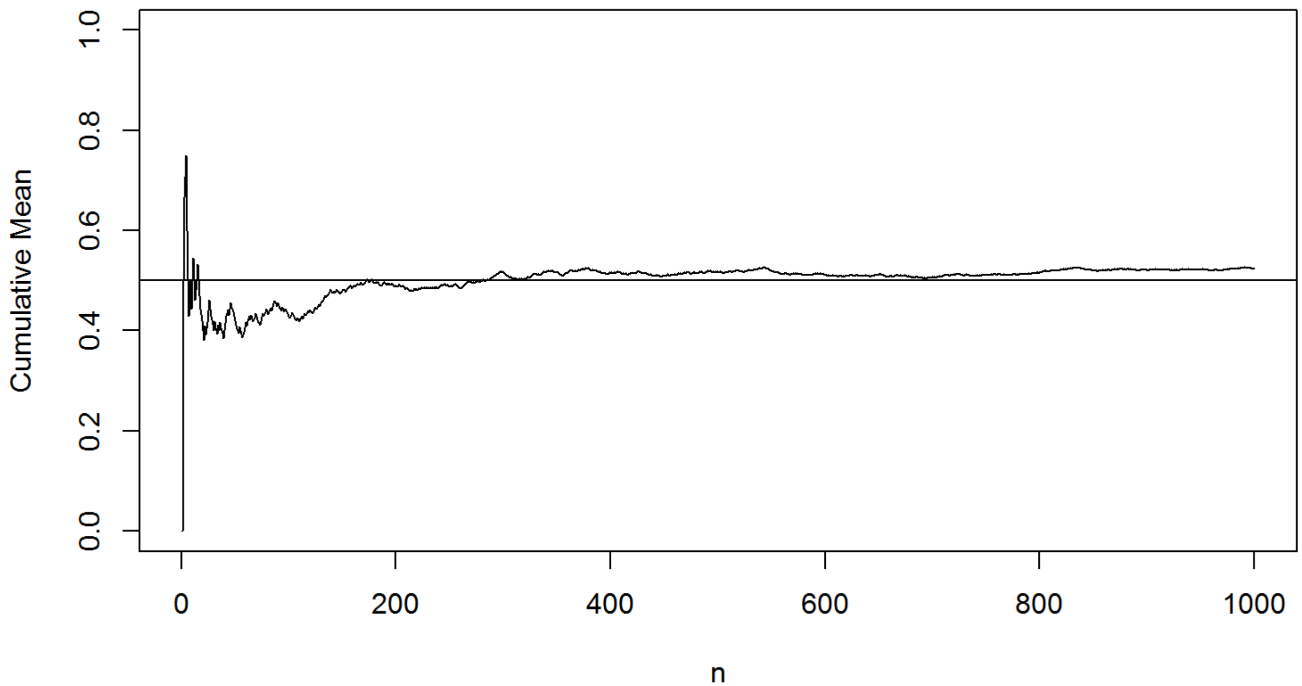
```
##  [1] 1 1 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 0 1 0 1 1 0 1 0 0 0 1 1 1
```

I am going to let n get larger and plot n versus cumulative values of $\overline{X}_n$.

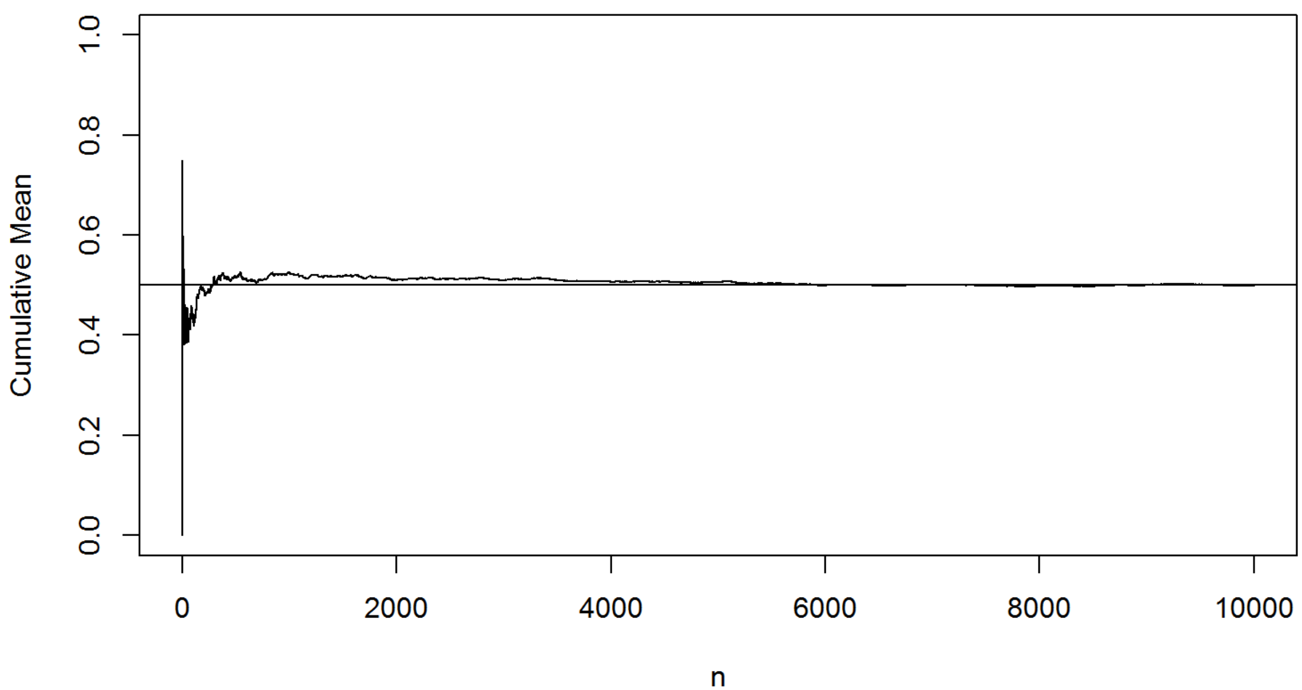
simulation example 1 - Bernoulli(0.5)



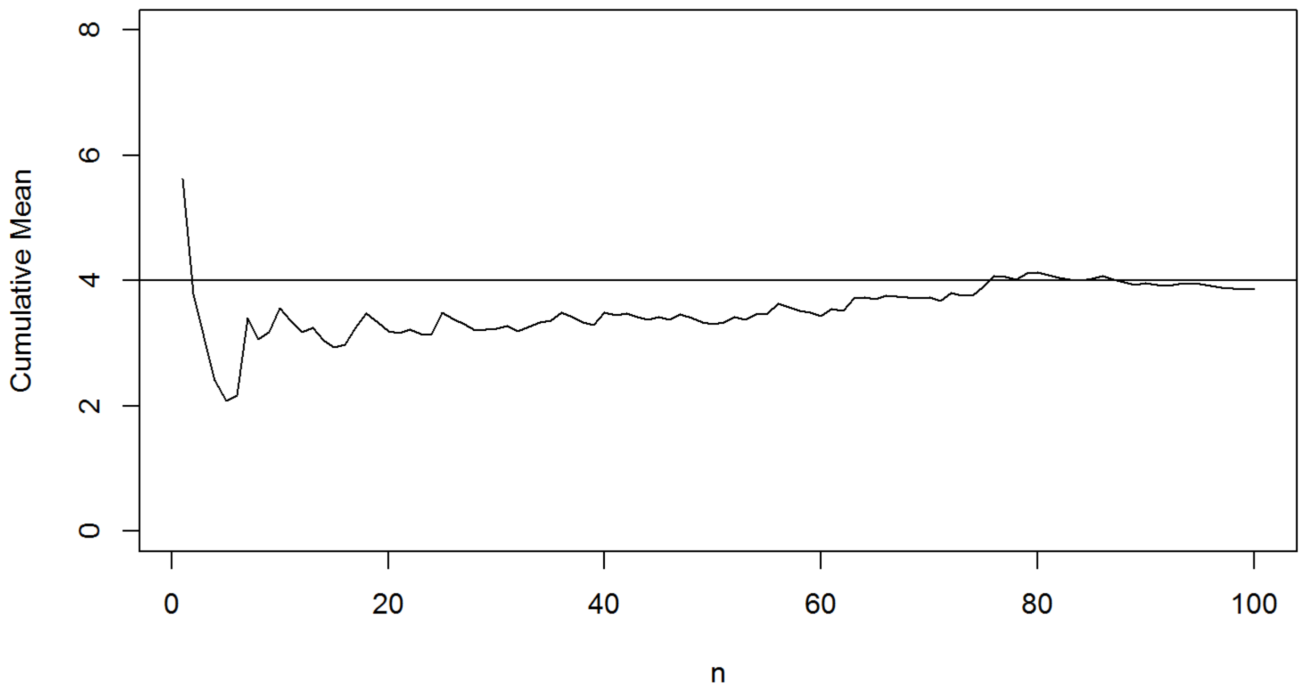
simulation example 1 - Bernoulli(0.5)



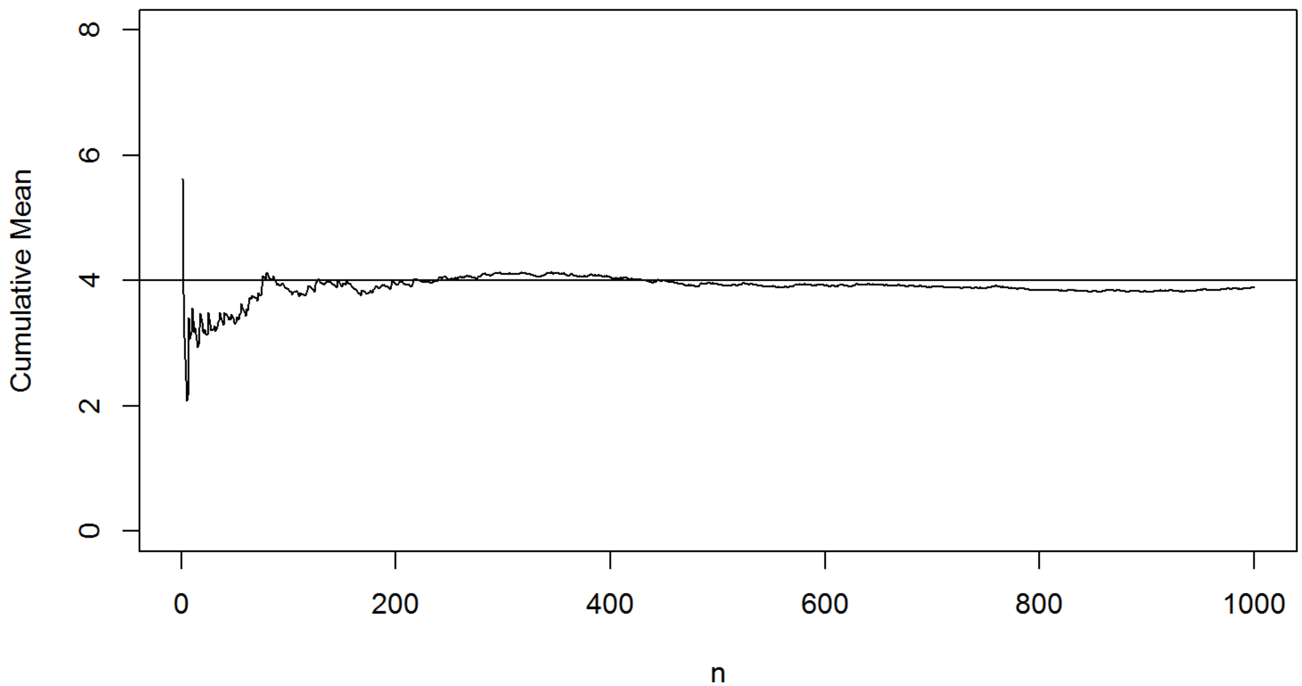
simulation example 1 - Bernoulli(0.5)



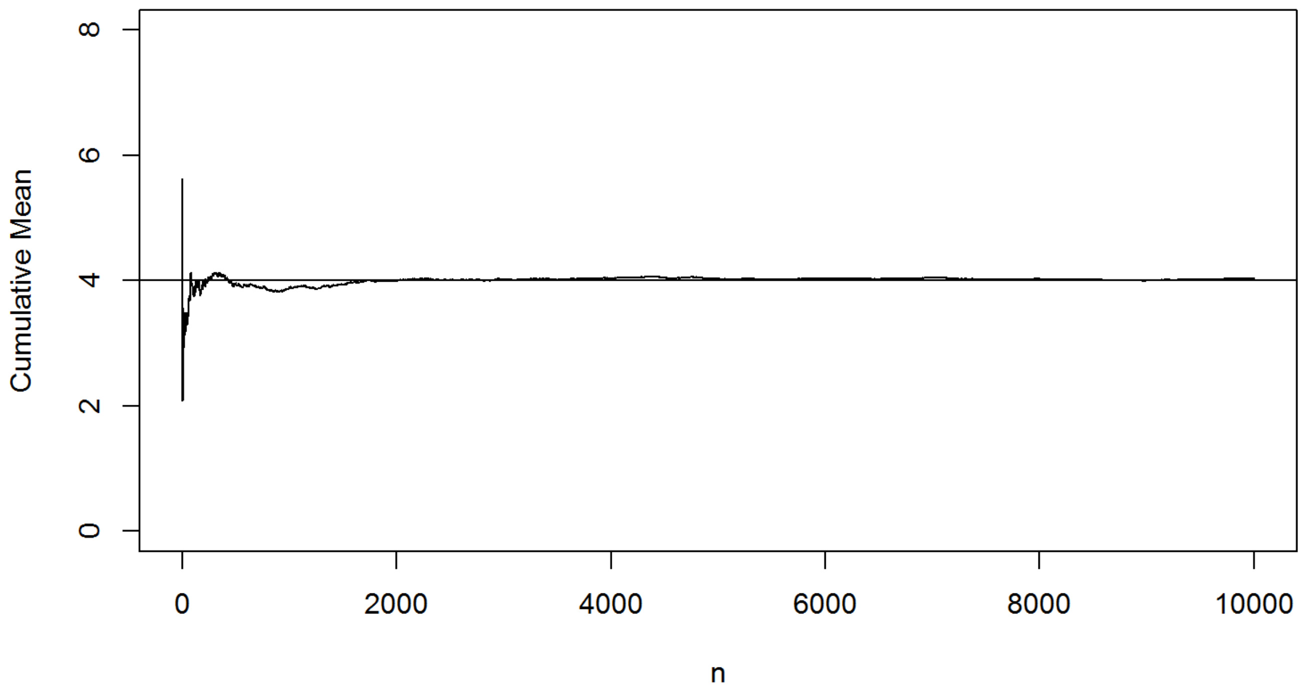
simulation example 2 - Exponential(0.25)



simulation example 2 - Exponential(0.25)



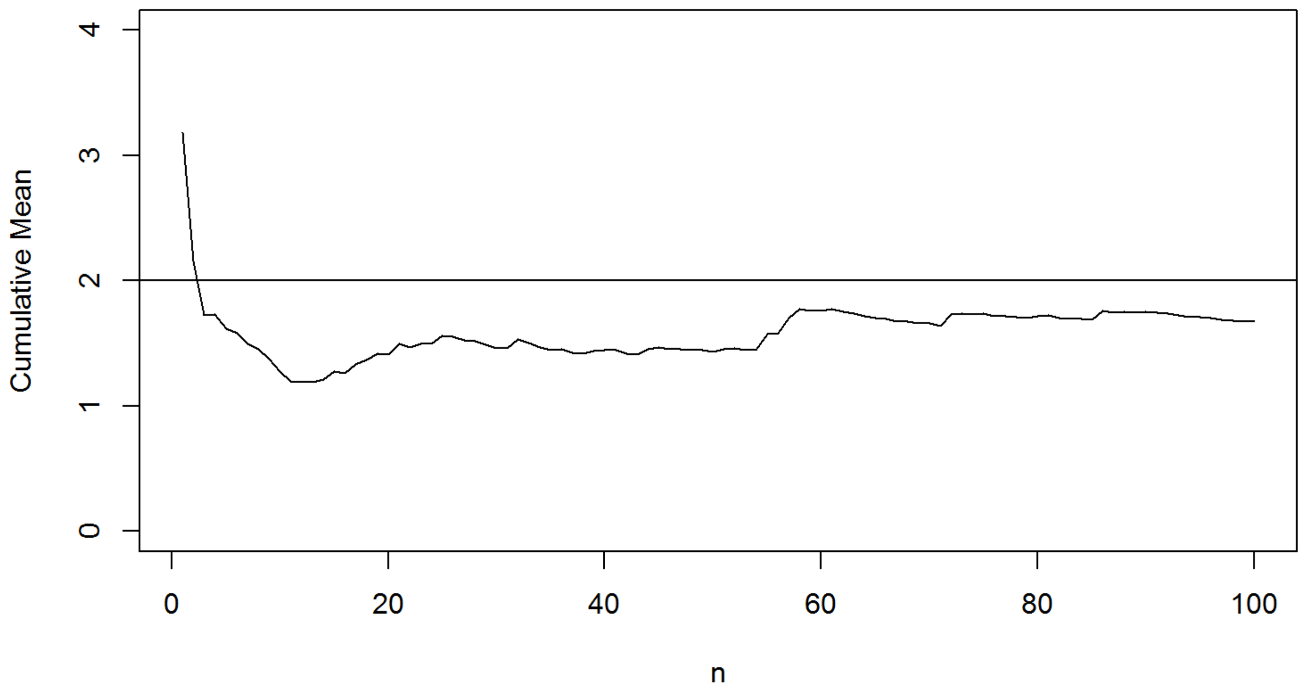
simulation example 2 - Exponential(0.25)



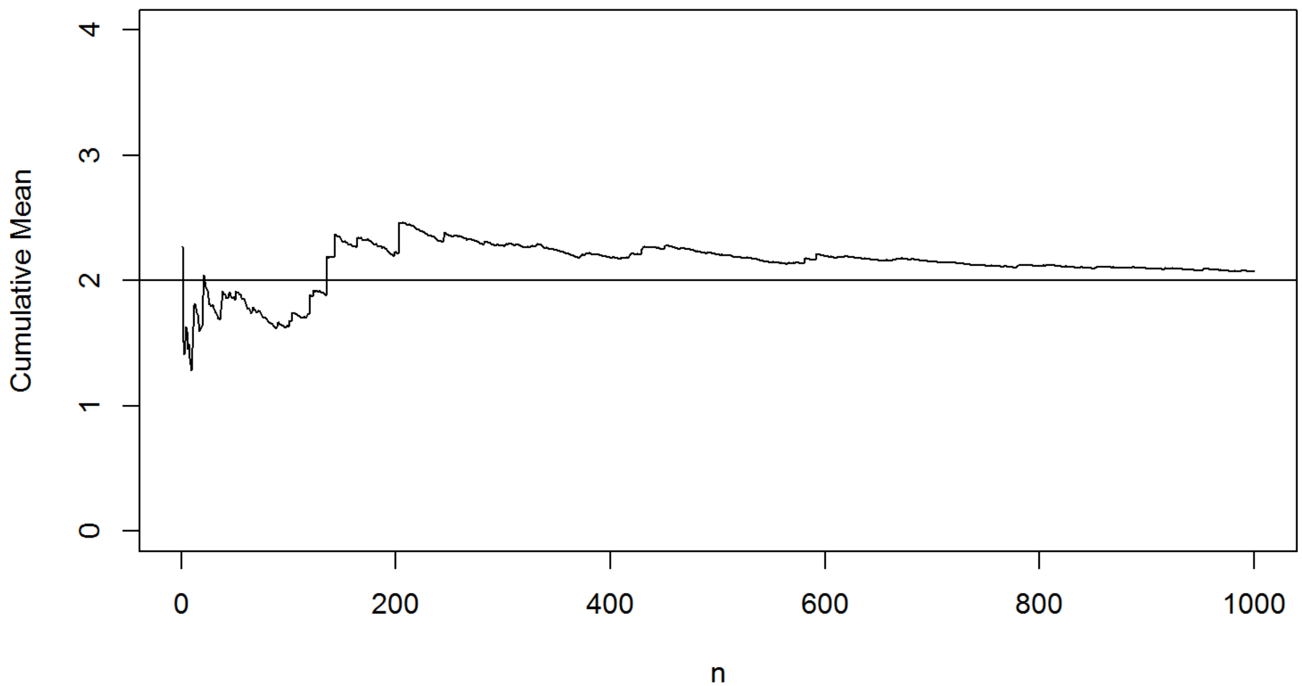
simulation example 3 - The *Neil* distribution

Positive random variable X with density $\frac{8}{\pi}\frac{x}{x^4+4}$

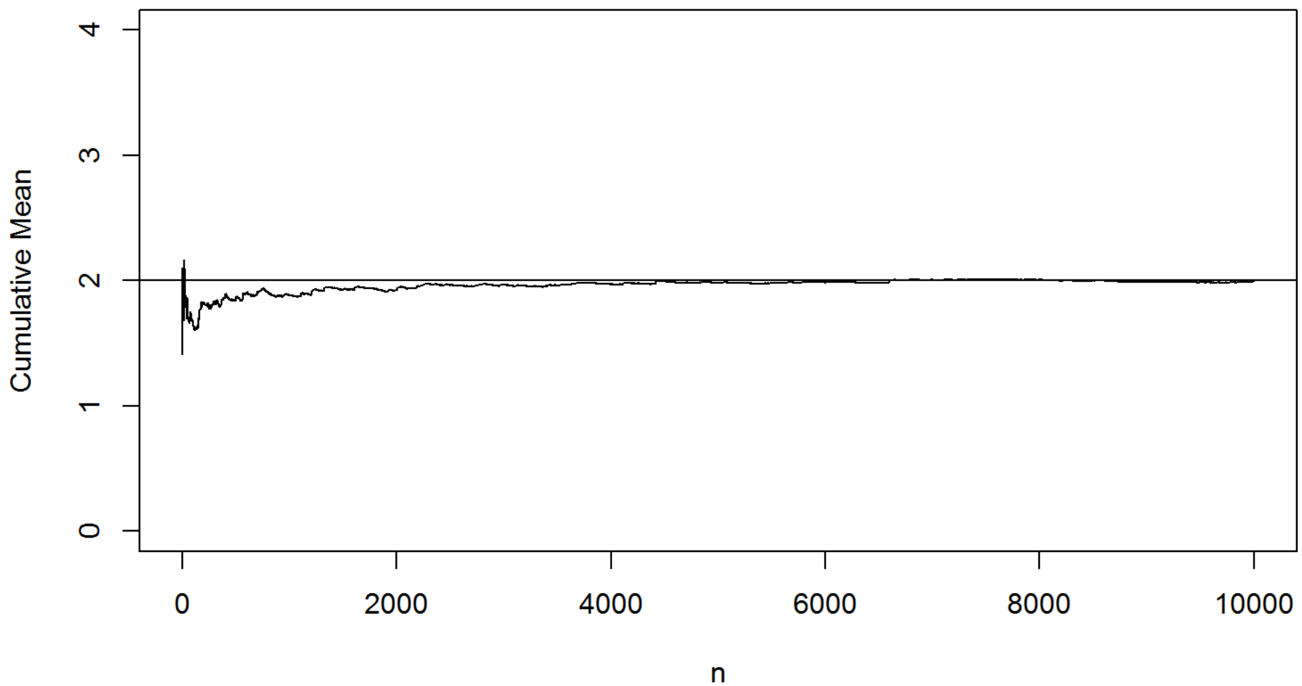
$E(X) = 2$ (hard!!) and $\mathrm{Var}(X)$ does not exist.



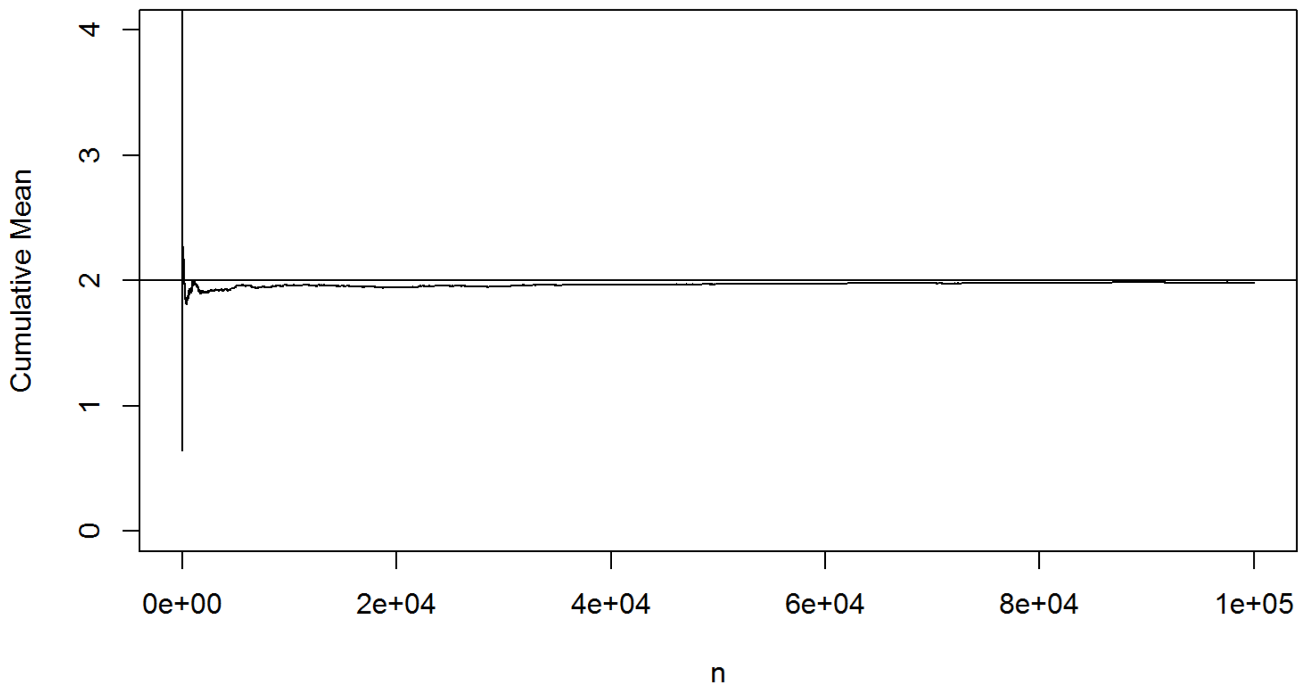
simulation example 3 - The *Neil* distribution



simulation example 3 - The *Neil* distribution

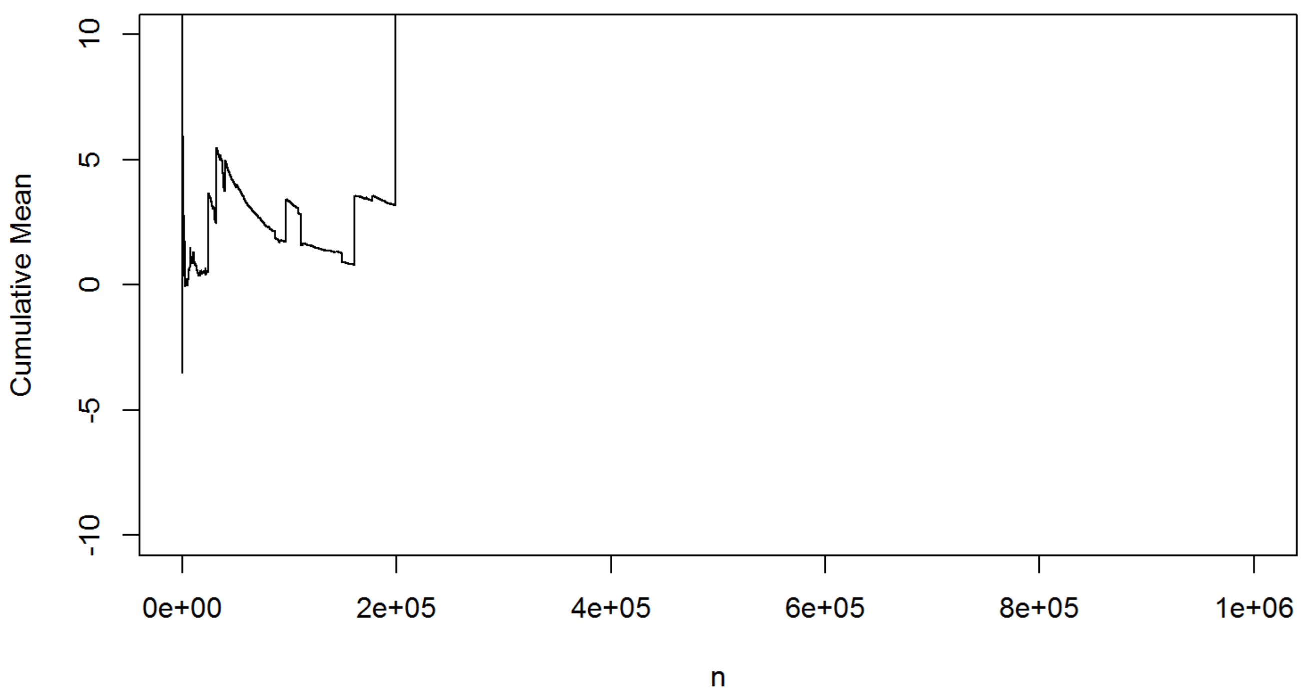


simulation example 3 - the *Neil* distribution



simulation example 4 - the Cauchy distribution

Density $\frac{1}{\pi}\frac{1}{1+x^2}$. No expected value



the (weak) Law of Large Numbers

Examples 1 and 2 had random variables with mean μ and variance σ^2. The convergence of the $\overline{X}_n$ to μ is explained by…

Theorem: If $X_1, X_2, X_3, \ldots$ are independent with the same mean μ and variance σ^2, then for all $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = 0.$$

Proof:…

This does *not* explain example 3 (no variance), and doesn't really explain 4 either.

Good for theory and philosophy. Not so good in practice (rate of convergence?)

convergence in probability

The WLLN is an example of *convergence in probability*:

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0,$$

and in the WLLN $X = \mu$ is a constant random variable.

Common notation: $X_n \xrightarrow{P} X$

Convergence in probability has to do with *values* of random variables.

The *distribution* of X (what we care about) only deals with things like $P(a < X < b)$, which is less stringent than convergence in probability.

convergence in distribution

Roughly speaking, the distribution of X_n converges to the distribution of X if their cdfs converge: $F_{X_n} \to F_X$

Formal definition: X_n *converges in distribution* to X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at every point where $F_X(x)$ is continuous.

Common notations:

$$X_n \Longrightarrow X$$
$$X_n \xrightarrow{D} X$$

Definition not so easy to use.

verifying convergence in distribution

Stated without proof; all imply $X_n \xrightarrow{D} X$.

Theorem: X_n with pmf $p_{X_n}(x)$ that converge to a pmf $p_X(x)\ldots$

Example: $X_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$ and $X \sim \text{Poisson}(\lambda)$

Theorem: X_n with density $f_{X_n}(x)$ that converge to a density $f_X(x)\ldots$

Future example for STA261 students: $X_n \sim t_n$ and $X \sim N(0, 1)$

THEOREM: X_n with m.g.f. $m_{X_n}(t)$ that converge to an m.g.f. $m_X(t)$ in a neighborhood of 0 implies $X_n \xrightarrow{D} X$.

Example: $X_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)\ldots$

the fundamental theorem of statistics

the distribution of $\overline{X}_n$

Let's consider again $X_1, X_2, \ldots$ now i.i.d. with m.g.f. $m(t)$. The common mean and variance are μ and σ^2.

$E\left(\overline{X}_n\right) = \mu$ and $\mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}$

Consider:

$$Y_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

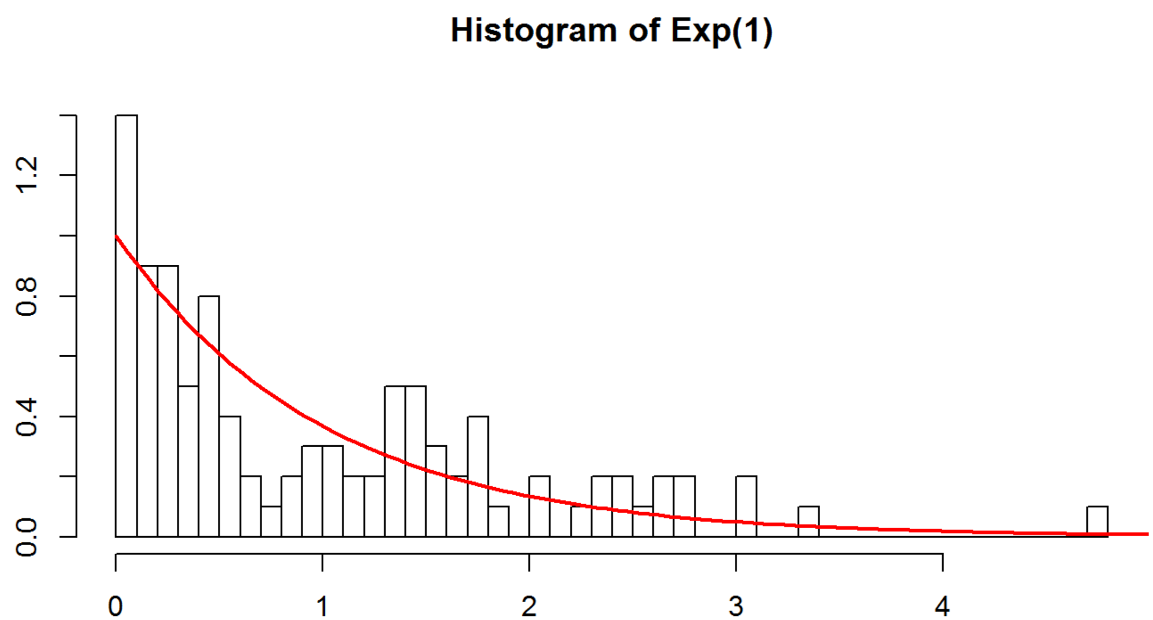
Then $E(Y_n) = 0$ and $\mathrm{Var}(Y_n) = 1$.

What could be said about the *distribution* of Y_n?

histograms (for simulating Y_n)

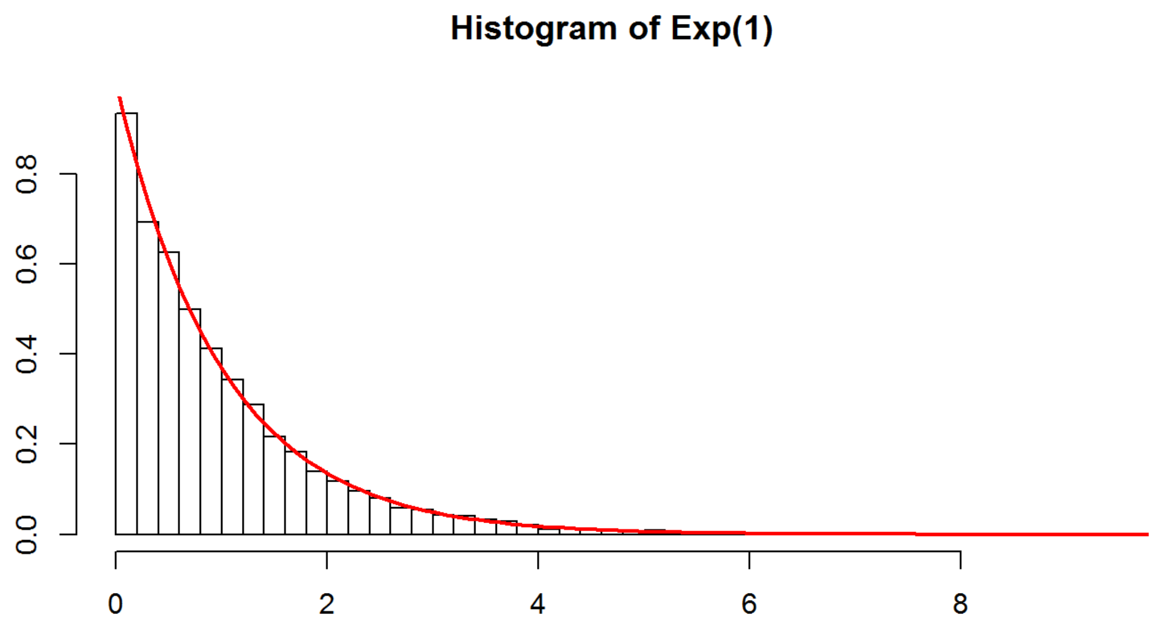
A *histogram* takes a sequence of numbers (the "data"), splits the range into "bins", and produces a bar graph of the count inside each bin.

A histogram is a "density estimator". Here's a histogram of $k = 100$ randomly samples from an $\text{Exp}(1)$ distribution, with the density in red:



more simulation = smoother histogram

$k = 10^4$



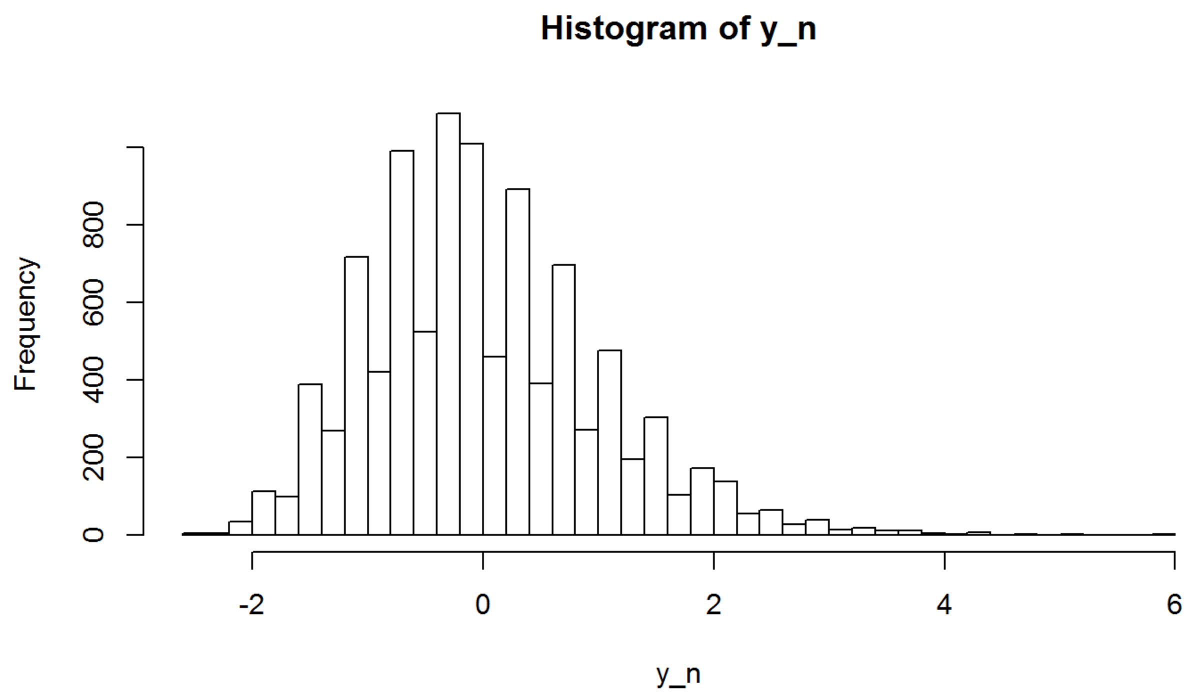
simulating Y_n

Fix n and a distribtion for X_i.

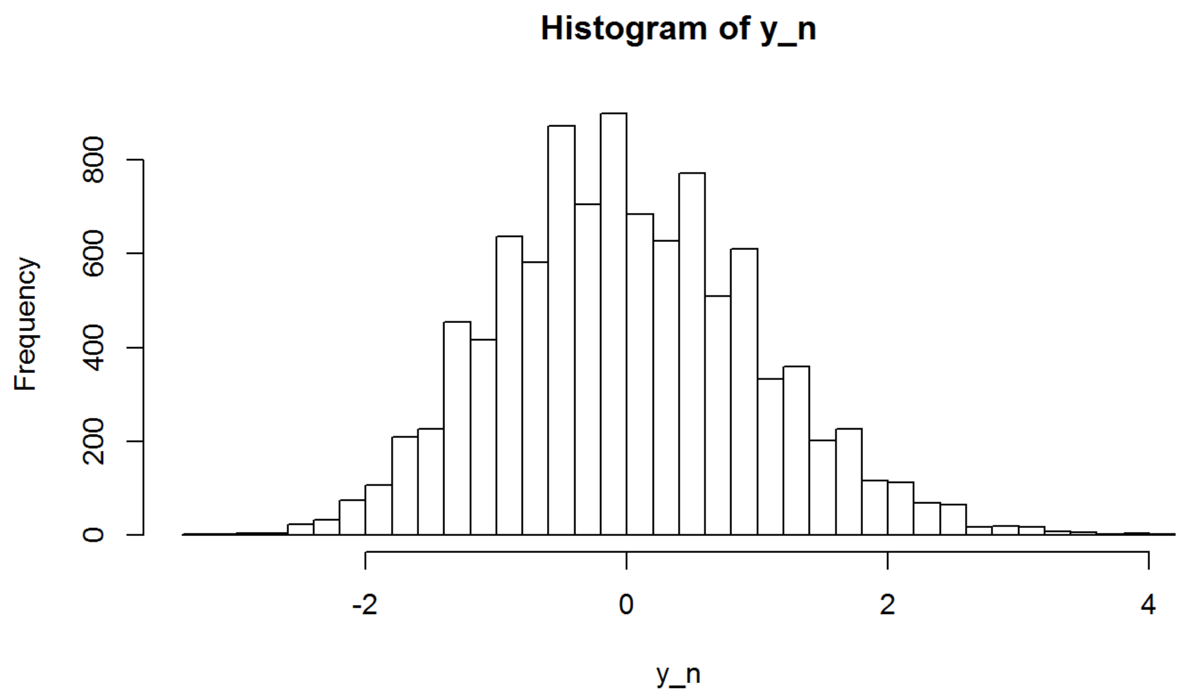
1. Simulate a sample of size n from the distribution.
2. Calculate $Y_n^{(1)}$ from this sample.
3. Repeat k times to obtain $Y_n^{(1)}, \dots, Y_n^{(k)}$
4. Make a histogram of the $Y_n^{(j)}$

Important: n is fixed and fundamental to the simulation. A larger k makes a nicer histogram and is more of a choice to make.

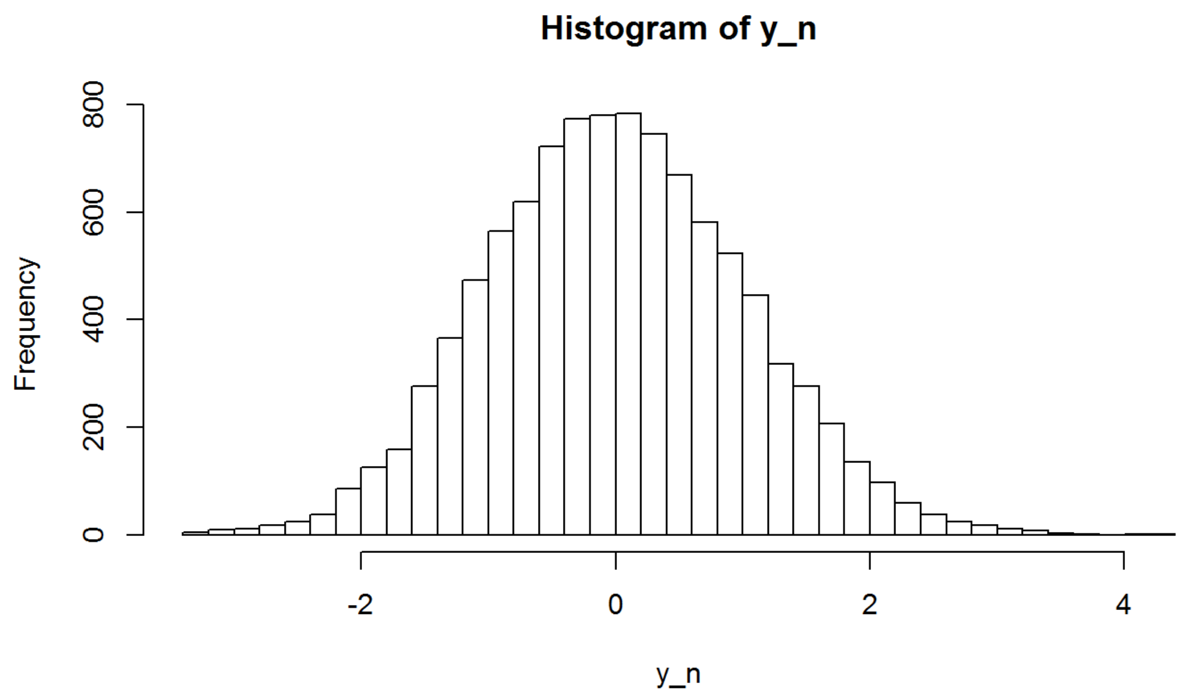
example 1 with $n = 10$ and $X_i \sim \text{Geometric}(1/3)$



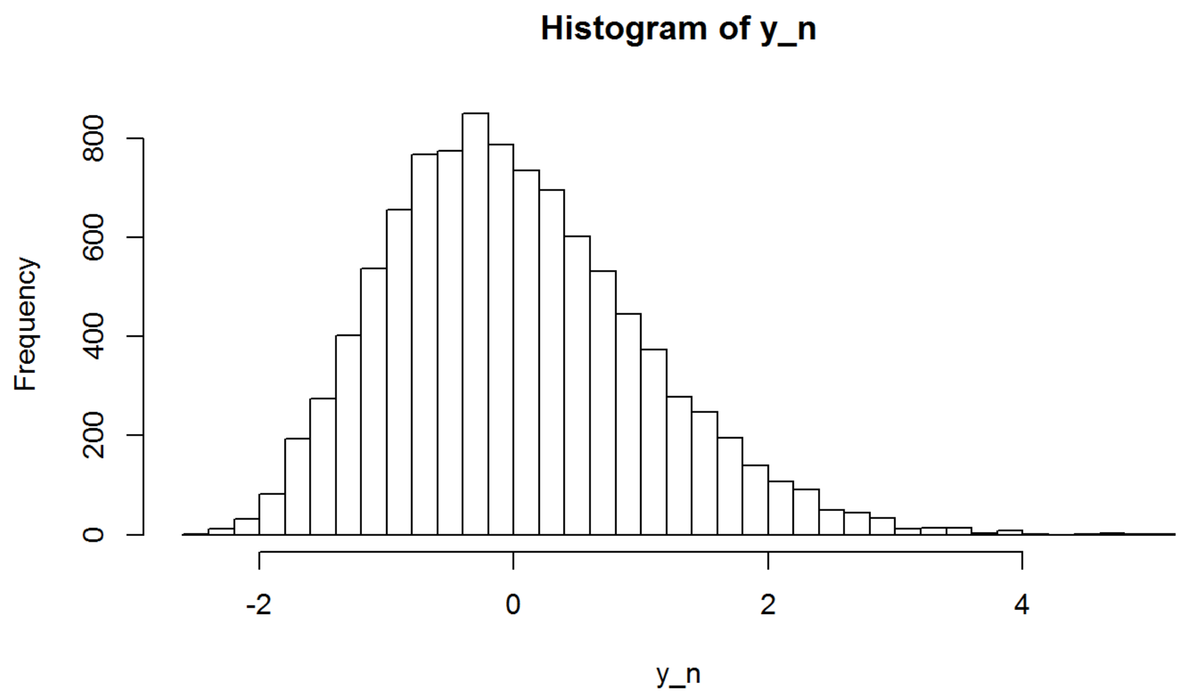
example 1 with $n = 50$ and $X_i \sim \text{Geometric}(1/3)$



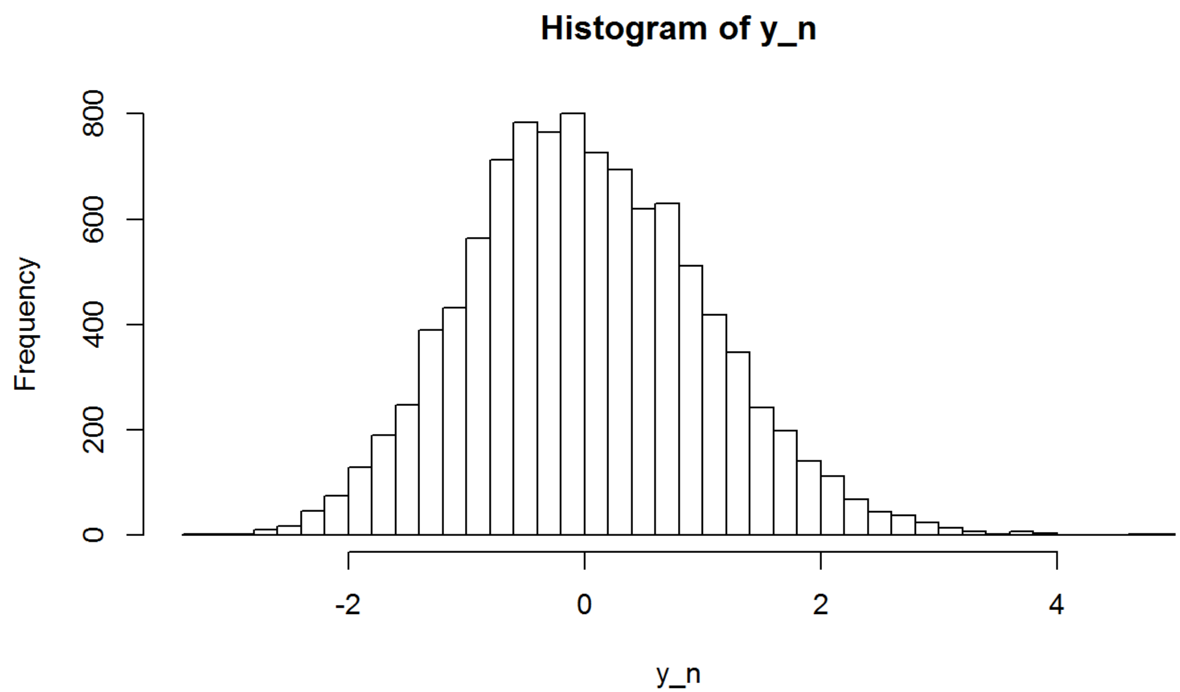
example 1 with $n = 500$ and $X_i \sim \text{Geometric}(1/3)$



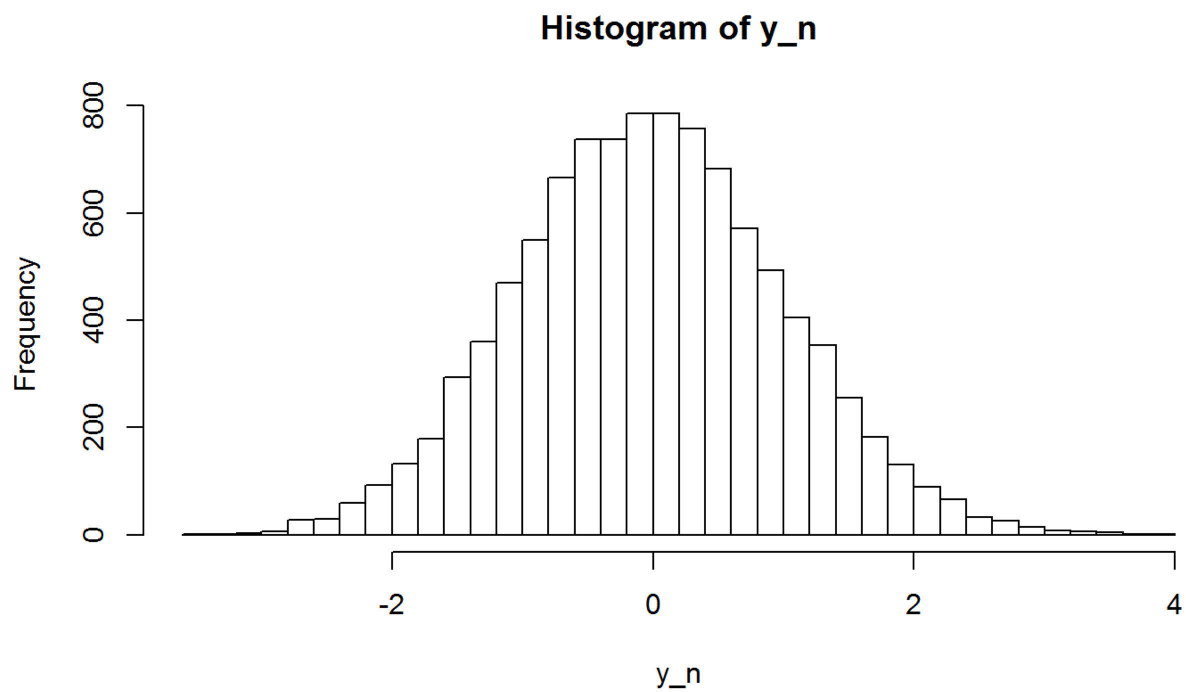
example 2 with $n = 10$ and $X_i \sim \mathrm{Exp}(0.25)$



example 2 with $n = 50$ and $X_i \sim \text{Exp}(0.25)$



example 2 with $n = 500$ and $X_i \sim \mathrm{Exp}(0.25)$



The Central Limit Theorem

Theorem: $X_1, X_2, \ldots$ are i.i.d. with m.g.f. $m(t)$, mean μ, and variance σ^2. Then:

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z$$

where $Z \sim N(0, 1)$.

Proof:…

This is a *limit theorem*. What is crucial is that the convergence can be *fast*.

normal approximations

For $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2 and n large enough*

$$\sum_{i=1}^{n} X_i \sim^{\text{approx}} N(n\mu, n\sigma^2)$$

$$\overline{X} \sim^{\text{approx}} N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim^{\text{approx}} N(0, 1)$$

* depends on the underlying distribution. The more "skewed" or "heavy-tailed", the larger the n required.

example - Uniform(0,1)

$X_1, X_2, \ldots, X_n$ are Uniform(0,1) and $n = 20$. What's the chance that $\sum X_i > 11$