Topic 10
The Law of Large Numbers

Outline

Introduction

Public health officials want to ascertain the mean weight of healthy newborn babies in a their region of study.

- They randomly choose babies and weigh them, keeping a running average.
- At the beginning we might see some larger fluctuations in our average.
- As they continue to make measurements, we expect to see this running average settle and converge to the true mean weight of newborn babies.

This phenomena is informally known as the law of averages. In probability theory, we call this the law of large numbers.

Introduction

Exercise. Entering the following R commands to simulate and plot the running average of newborn birth weights.

```
> n<-c(1:100)                 #create a vector of integers from 1 to 100
> weight<-rnorm(100,3,0.5) #simulate weight of 100 newborns
> s<-cumsum(weight)           #keeping a running sum of total birthweights
> plot(s/n,xlab="n",ylim=c(2,4),type="l") #plot this running average
```

Describe what you see. Repeat this simulation several times and note the differences and similarities among the plots.

Distribution of Sample Mean

We begin with a sequence $X_1, X_2, \ldots$ of random variables having a common distribution. Their average, the sample mean,

$$\bar{X} = \frac{1}{n} S_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

is itself a random variable.

If the common mean for the X_i's is μ, then by the linearity property of expectation, the mean of the average,

$$E[\frac{1}{n} S_n] = \frac{1}{n}(EX_1 + EX_2 + \cdots + EX_n) = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \frac{1}{n} n\mu = \mu.$$

is also μ.

Distribution of Sample Mean

If, in addition, the X_i's are independent with common variance σ^2, then first by the quadratic identity and then the Pythagorean identity for the variance of independent random variables, we find that the variance of $\bar{X}$,

$$
\begin{aligned}
\sigma^2_{\bar{X}} &= \mathrm{Var}(\frac{1}{n}S_n) = \frac{1}{n^2}\mathrm{Var}(S_n) \\
&= \frac{1}{n^2}(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)) \\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2.
\end{aligned}
$$

- The mean of these running averages remains at μ, but
- The variance is decreasing to 0 at a rate inversely proportional to n, the number of terms in the sum.

Law of Large Numbers

Theorem. For a sequence of independent random variables $X_1, X_2, \ldots$ having a common distribution, their running average

$$\frac{1}{n} S_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

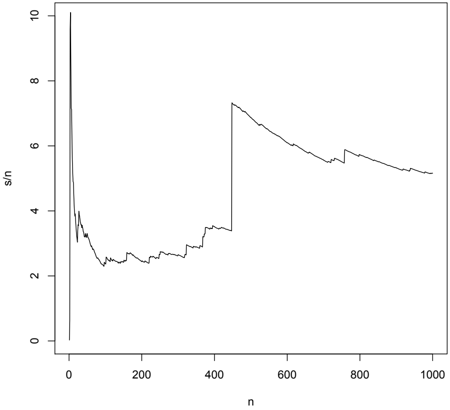
has a limit as $n \to \infty$ if and only if this sequence of random variables has a common mean μ. In this case the limit is μ.

The theorem also states that if the random variables do not have a mean, the limit will *fail* to exist.

Law of Large Numbers

Exercise. We will simulate Cauchy random variables to examine the case when the mean does not exist. Repeat the simulation below and compare your plot to the one displayed.

```
> n<-c(1:1000)
> y<-abs(rcauchy(1000))
> s<-cumsum(y)
> plot(s/n,xlab="n",
   ylim=c(0,10),type="l")
```



Law of Large Numbers

The simulations are based on standard Cauchy random variables, X. X has density function

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Let $Y = |X|$. In an attempt to compute the improper integral for $EY = E|X|$, note that

$$\int_{-b}^{b} |x| f_X(x)\ dx = 2\int_0^b \frac{1}{\pi}\frac{x}{1+x^2}\ dx = \frac{1}{\pi}\ln(1+x^2)\Big|_0^b = \frac{1}{\pi}\ln(1+b^2) \to \infty$$

as $b \to \infty$. Thus, Y has *infinite* mean. So the law of large numbers states that the running average will not have a limit.