Rating
Mathematically Mature: may contain mathematics beyond calculus with proofs.
Section Starter Question

Consider a fair \( p = 1/2 = q \) coin tossing game carried out for 1000 tosses. Explain in a sentence what the “law of averages” says about the outcomes of this game.

Key Concepts

1. The precise statement, meaning and proof of the Weak Law of Large Numbers.

2. The precise statement and meaning of the Strong Law of Large Numbers.

Vocabulary

1. The **Weak Law of Large Numbers** is a precise mathematical statement of what is usually loosely referred to as the “law of averages”. Precisely, let \( X_1, \ldots, X_n \) be independent, identically distributed random variables each with mean \( \mu \) and variance \( \sigma^2 \). Let \( S_n = X_1 + \cdots + X_n \) and consider the **sample mean** or more loosely, the “average” \( S_n/n \). Then the Weak Law of Large Numbers says that the sample mean \( S_n/n \) converges in probability to the population mean \( \mu \). That is:

\[
\lim_{n \to \infty} \mathbb{P}_n [ |S_n/n - \mu| > \epsilon ] = 0.
\]
In words, the proportion of those samples whose sample mean differs significantly from the population mean diminishes to zero as the sample size increases.

2. The **Strong Law of Large Numbers** says that $S_n/n$ converges to $\mu$ with probability 1. That is:

$$\mathbb{P}\left[\lim_{n \to \infty} S_n/n = \mu\right] = 1.$$ 

In words, the Strong Law of Large Numbers “almost every” sample mean approaches the population mean as the sample size increases.

---

**Mathematical Ideas**

**The Weak Law of Large Numbers**

**Lemma 1** (Markov’s Inequality). If $X$ is a random variable that takes only nonnegative values, then for any $a > 0$:

$$\mathbb{P}[X \geq a] \leq \mathbb{E}[X]/a$$

*Proof.* Here is a proof for the case where $X$ is a continuous random variable
with probability density $f$:

$$
E[X] = \int_0^\infty xf(x) \, dx = \int_0^a xf(x) \, dx + \int_a^\infty xf(x) \, dx \\
\geq \int_a^\infty xf(x) \, dx \\
\geq \int_a^\infty af(x) \, dx \\
= a \int_a^\infty f(x) \, dx \\
= a \Pr[X \geq a].
$$

(The proof for the case where $X$ is a purely discrete random variable is similar with summations replacing integrals. The proof for the general case is exactly as given with $dF(x)$ replacing $f(x) \, dx$ and interpreting the integrals as Riemann-Stieltjes integrals.)

**Lemma 2** (Chebyshev’s Inequality). *If $X$ is a random variable with finite mean $\mu$ and variance $\sigma^2$, then for any value $k > 0$:

$$
\Pr[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}.
$$

*Proof.* Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov’s inequality (with $a = k^2$) to obtain

$$
\Pr[(X - \mu)^2 \geq k^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2}.
$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, the inequality above is equivalent to:

$$
\Pr[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}
$$

and the proof is complete.  

**Theorem 3** (Weak Law of Large Numbers). *Let $X_1, X_2, X_3, \ldots$, be independent, identically distributed random variables each with mean $\mu$ and variance $\sigma^2$. Let $S_n = X_1 + \cdots + X_n$. Then $S_n/n$ converges in probability to $\mu$. That is:

$$
\lim_{n \to \infty} \Pr_n[|S_n/n - \mu| > \epsilon] = 0.
$$

4
Proof. Since the mean of a sum of random variables is the sum of the means, and scalars factor out of expectations:

\[ \mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n}(n\mu) = \mu. \]

Since the variance of a sum of independent random variables is the sum of the variances, and scalars factor out of variances as squares:

\[ \text{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}[X_i] = \frac{1}{n^2}(n\sigma^2) = \sigma^2/n. \]

Fix a value \( \epsilon > 0 \). Then using elementary definitions for probability measure and Chebyshev’s Inequality:

\[ 0 \leq \mathbb{P}_n \left[ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right] \leq \mathbb{P}_n \left[ \left| S_n/n - \mu \right| \geq \epsilon \right] \leq \frac{\sigma^2}{(n\epsilon^2)}. \]

Then by the squeeze theorem for limits

\[ \lim_{n \to \infty} \mathbb{P}_n \left[ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right] = 0. \]

Jacob Bernoulli originally proved the Weak Law of Large Numbers in 1713 for the special case when the \( X_i \) are binomial random variables. Bernoulli had to create an ingenious proof to establish the result, since Chebyshev’s inequality was not known at the time. The theorem then became known as Bernoulli’s Theorem. Simeon Poisson proved a generalization of Bernoulli’s binomial Weak Law and first called it the Law of Large Numbers. In 1929 the Russian mathematician Aleksandr Khinchin proved the general form of the Weak Law of Large Numbers presented here. Many other versions of the Weak Law are known, with hypotheses that do not require such stringent requirements as being identically distributed, and having finite variance.

The Strong Law of Large Numbers

Theorem 4 (Strong Law of Large Numbers). Let \( X_1, X_2, X_3, \ldots \), be independent, identically distributed random variables each with mean \( \mu \) and variance \( \mathbb{E}\left[X^2_j\right] < \infty \). Let \( S_n = X_1 + \cdots + X_n \). Then \( S_n/n \) converges with probability 1 to \( \mu \),

\[ \mathbb{P} \left[ \lim_{n \to \infty} \frac{S_n}{n} = \mu \right] = 1. \]
The proof of this theorem is beautiful and deep, but would take us too far afield to prove it. The Russian mathematician Andrey Kolmogorov proved the Strong Law in the generality stated here, culminating a long series of investigations through the first half of the 20th century.

**Discussion of the Weak and Strong Laws of Large Numbers**

In probability theory a theorem that tells us how a sequence of probabilities converges is called a **weak law**. For coin tossing, the sequence of probabilities is the sequence of binomial probabilities associated with the first $n$ tosses. The Weak Law of Large Numbers says that if we take $n$ large enough, then the binomial probability of the mean over the first $n$ tosses differing “much” from the theoretical mean should be small. This is what is usually popularly referred to as the law of averages. However, this is a limit statement and the Weak law of Large Numbers above does not indicate the rate of convergence, nor the dependence of the rate of convergence on the difference $\epsilon$. Note furthermore that the Weak Law of Large Numbers in no way justifies the false notion called the “Gambler’s Fallacy”, namely that a long string of successive Heads indicates a Tail “is due to occur soon”. The independence of the random variables completely eliminates that sort of prescience.

A **strong law** tells how the sequence of random variables as a sample path behaves in the limit. That is, among the infinitely many sequences (or paths) of coin tosses we select one “at random” and then evaluate the sequence of means along that path. The Strong Law of Large Numbers says that with probability 1 that sequence of means along that path will converge to the theoretical mean. The formulation of the notion of probability on an infinite (in fact an uncountably infinite) sample space requires mathematics beyond the scope of the course, partially accounting for the lack of a proof for the Strong Law here.

Note carefully the difference between the Weak Law of Large Numbers and the Strong Law. We do not simply move the limit inside the probability. These two results express different limits. The Weak Law is a statement that the group of finite-length experiments whose sample mean is close to the population mean approaches all of the possible experiments as the length increases. The Strong Law is an experiment-by-experiment statement, it says (almost every) sequence has a sample mean that approaches the population
mean. This is reflected in the subtle difference in notation here. In the Weak Law the probabilities are written with a subscript: $P_n[\cdot]$ indicating this is a binomial probability distribution with parameter $n$ (and $p$). In the Strong Law, the probability is written without a subscript, indicating this is a probability measure on a sample space. Weak laws are usually much easier to prove than strong laws.

**Sources**

This section is adapted from Chapter 8, “Limit Theorems”, *A First Course in Probability*, by Sheldon Ross, Macmillan, 1976.

**Algorithms, Scripts, Simulations**

**Algorithm**

The experiment is flipping a coin $n$ times, and repeat the experiment $k$ times. Then compute the proportion for which the sample mean deviates from $p$ by more than $\epsilon$.

**Scripts**

Geogebra  

R script for the Law of Large Numbers

```r
p <- 0.5
n <- 10000
k <- 1000
coinFlips <- array ( 0+( runif (n*k) <= p), dim =c(n,k))
# 0+ coerces Boolean to numeric
headsTotal <- colSums(coinFlips)
# 0..n binomial rv sample, size k
epsilon <- 0.01
mu <- p
prob <- sum ( 0+( abs( headsTotal/n - mu ) > epsilon ) )/k
```
Octave

Octave script for the Law of Large Numbers

```octave
p = 0.5;
n = 10000;
k = 1000;

coinFlips = rand(n,k) <= p;
headsTotal = sum(coinFlips);
# 0..n binomial rv sample, size k

epsilon = 0.01
mu = p;
prob = sum(abs(headsTotal/n - mu) > epsilon)/k;
disp("Empirical probability:"), disp(prob)
```

Perl

Perl PDL script for the Law of Large Numbers

```perl
use PDL::NiceSlice;

$p = 0.5;
$n = 10000;
$k = 1000;

$coinFlips = random( $k, $n ) <= $p;
# note order of dims!!
$headsTotal = $coinFlips->transpose->sumover;
# 0..n binomial r.v. sample, size k
# note transpose, PDL likes x (row) direction for
# implicitly threaded operations

$epsilon = 0.01;
$mu = $p;

$prob = ( ( abs( ( $headsTotal / $n ) - $mu ) > $epsilon )
    -> sumover ) / $k;
print "Empirical probability: ", $prob, "\n";
```
SciPy [Scientific Python script for the Law of Large Numbers]

```python
import scipy

p = 0.5
n = 10000
k = 1000

coinFlips = scipy.random.random((n,k)) <= p
# Note Booleans True for Heads and False for Tails
headsTotal = scipy.sum(coinFlips, axis = 0)
# 0..n binomial r.v. sample, size k
# Note how Booleans act as 0 (False) and 1 (True)

epsilon = 0.01
mu = p

prob = (scipy.sum(abs(headsTotal.astype('float')/n - mu) > epsilon)).astype('float')/k
# Note the casting of integer types to float to get floats

print "Empirical probability: ", prob
```

Problems to Work for Understanding

1. Suppose $X$ is a continuous random variable with mean and variance both equal to 20. What can be said about $\mathbb{P}[0 \leq X \leq 40]$?

2. Suppose $X$ is an exponentially distributed random variable with mean $\mathbb{E}[X] = 1$. For $x = 0.5, 1, and 2$, compare $\mathbb{P}[X \geq x]$ with the Markov Inequality bound.
3. Suppose $X$ is a Bernoulli random variable with $\mathbb{P}[X = 1] = p$ and $\mathbb{P}[X = 0] = 1 - p = q$. Compare $\mathbb{P}[X \geq 1]$ with the Markov Inequality bound.

4. Make a sequence of 100 coin tosses and keep a record as in Experiment. How “typical” was your coin flip sequence? All $2^{100}$ coin flip sequences are equally likely of course, so yours is neither more nor less typical than any other in that way. However, some sets or events of coin flip sequences are more or less “typical” when measured by the probability of a corresponding event. What is your value of $S_{100}$, and the number of heads and the number of tails in your record? Using your value of $S_{100}$ let $\epsilon = |S_{100}|$ and use Chebyshev’s Inequality as in the proof of the Weak Law of Large Numbers to provide an upper bound on the probability that for all possible records $|S_{100}| > \epsilon$.

5. Let $X_1, X_2, \ldots, X_{10}$ be independent Poisson random variables with mean 1. First use the Markov Inequality to get a bound on $\mathbb{P}[X_1 + \cdots + X_{10} > 15]$. Next find the exact probability that $\mathbb{P}[X_1 + \cdots + X_{10} > 15]$ using that the fact that the sum of independent Poisson random variables with parameters $\lambda_1, \lambda_2$ is again Poisson with parameter $\lambda_1 + \lambda_2$.

6. Write a proof of Markov’s Inequality for a random variable taking positive integer values.

7. Modify the scripts to compute the proportion of sample means which deviate from $p$ by more than $\epsilon$ for increasing values of $n$. Do the proportions decrease with increasing values of $n$? If so, at what rate do the proportions decrease?
Reading Suggestion:

References


Outside Readings and Links:

1. [Virtual Laboratories in Probability and Statistics](#) Search the page for “Simulation Exercises and then run the Binomial Coin Experiment and the Matching Experiment.

2.

I check all the information on each page for correctness and typographical errors. Nevertheless, some errors may occur and I would be grateful if you would alert me to such errors. I make every reasonable effort to present current and accurate information for public use, however I do not guarantee the accuracy or timeliness of information on this website. Your use of the information from this website is strictly voluntary and at your risk.

I have checked the links to external sites for usefulness. Links to external websites are provided as a convenience. I do not endorse, control, monitor, or guarantee the information contained in any external website. I don’t guarantee that the links are active at all times. Use the links here with the same caution as
you would all information on the Internet. This website reflects the thoughts, interests and opinions of its author. They do not explicitly represent official positions or policies of my employer.

Information on this website is subject to change without notice.
Steve Dunbar’s Home Page, http://www.math.unl.edu/~sdunbar1
Email to Steve Dunbar, sdunbar1 at unl dot edu
Last modified: Processed from \LaTeX{} source on July 21, 2016