

Basic probability and statistics via R

Contents

1	The law of large numbers	1
1.1	Theory	1
1.2	Application	2
2	Summary	4
3	Version: 27 April 2021	4

1 The law of large numbers

We will give a quick review of the law of large numbers and illustrate applications of this theorem via R.

1.1 Theory

You may recall that in probability theory, we are given the distribution of a non-negative integer-valued random variable X and we *define* its **mean** to be

$$\mathbb{E}X = \sum_{x \in \mathbb{N}} x \mathbb{P}(X = x).$$

The justification of the use of the word *mean* or *average* is given by the following theorem. Recall that a countable collection of random variables are **independent** if every finite subset of them is independent, and they are **identical** if they have the same distribution. We say that a sequence of random variables $(X_n)_{n \in \mathbb{Z}^+}$ converges **almost-surely** to a random variable X if there exists an event Ω' with $\mathbb{P}(\Omega') = 1$ such that for every $\omega \in \Omega'$, we have convergence in the usual pointwise sense: $X_n(\omega) \rightarrow X(\omega)$, as $n \rightarrow \infty$.

Theorem 1.1 (Law of large numbers (almost-sure version)). *Let $(X_n)_{n \in \mathbb{Z}^+}$ be a sequence of independent and identically distributed (i.i.d.) random variables. If $\mathbb{E}|X_1| < \infty$, then*

$$n^{-1}(X_1 + \cdots + X_n) \rightarrow \mathbb{E}X_1$$

almost-surely.

The almost-sure version of the law of large numbers has a somewhat difficult proof, but to prove convergence in the mean-square and in probability is easy. We say that X_n converges to X **in the mean-squared** if

$$\mathbb{E}|X_n - X|^2 \rightarrow 0$$

and we say that X_n converges to X **in probability** if for every $\epsilon > 0$, we have

$$\mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0.$$

Theorem 1.2 (Law of large numbers (convergence in mean-squared)). *Let $(X_n)_{n \in \mathbb{Z}^+}$ be a sequence of i.i.d. random variables. If $\mathbb{E}|X_1|^2 < \infty$, then*

$$n^{-1}(X_1 + \cdots + X_n) \rightarrow \mathbb{E}X_1$$

in the mean-squared.

Proof. Let $S_n = X_1 + \cdots + X_n$. Note that $\mathbb{E}S_n = n\mathbb{E}X_1$. We have

$$\begin{aligned} \mathbb{E}|n^{-1}S_n - \mathbb{E}X_1|^2 &= n^{-2}\mathbb{E}|S_n - n\mathbb{E}X_1|^2 \\ &= n^{-2}\text{var}(S_n) \\ &= n^{-2}\sum_{i=1}^n \text{var}(X_i) \\ &= n^{-1}\text{var}(X_1) \rightarrow 0, \end{aligned}$$

where we need an independence assumption to commute the variance as a sum of variances. □

Convergence in probability is an easy consequence of Markov's inequality.

Lemma 1.1 (Markov's inequality). *Let $X \geq 0$ be a random variable, then for all $a \geq 0$, we have*

$$a^2\mathbb{P}(X > a) \leq \mathbb{E}X^2.$$

Lemma 1.2. *If X_n converges to X in the mean-square, then it converges in probability.*

Proof. Let $\epsilon > 0$. Markov's inequality gives

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \epsilon^{-2}\mathbb{E}|X_n - X|^2 \rightarrow 0.$$

□

Exercise 1.1. Let $g : [0, 1] \rightarrow \mathbb{R}$ be a continuous function. Let $(X_i)_{i \in \mathbb{Z}^+}$ be independent random variables that are uniformly distributed on the interval $[0, 1]$. Show that

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \rightarrow \int_0^1 g(x) dx.$$

Recall that if T_n is a sequence of estimators for a parameter θ , then we say that T_n is **consistent** if $T_n \rightarrow \theta$ in probability.

Exercise 1.2. Show that the usual sample variance is a consistent estimator for the *true* variance.

1.2 Application

The law of large numbers gives us another way to compute probabilities. If we want to compute say $p = \mathbb{P}(Z > 1)$, one way to is to consider i.i.d. random variables with the same law as Z , and consider the count

$$S_n = \mathbf{1}[X_1 > 1] + \cdots + \mathbf{1}[X_n > 1],$$

where $\mathbf{1}[X_i > 1] = 1$ if $X_i > 1$ and zero otherwise. Since the X_i are i.i.d. the Bernoulli random variables $\mathbf{1}[X_i > 1]$ are also i.i.d. and thus the law of large numbers applies

$$n^{-1}S_n \rightarrow p,$$

and thus the average for large values of n approximates p .

In R it is possible to simulate random variables, and thus it is possible to approximate probabilities. In short, we can do most of the exercises of elementary probability by simulation.

Exercise 1.3 (The hat check problem). A professor completely forgets the names and faces of all $n = 20$ students in his class. He randomly hands back their midterms. Let p_n be the probability that no one receives their own test back. It is known that as $n \rightarrow \infty$ we have $p_n \rightarrow e^{-1}$, and the approximation is quite good for even $n = 20$. Demonstrate this fact with R.

Solution. Note that with R we can easily generate a random permutation of the 20 papers, as a permutation of 20 numbers. Each random permutation generated by R is considered to be independent. We can check whether a permutation is *in-order* by defining a function *Hand*; we repeat this experiment (10000 times) with the replicate function and keep score.

```
inorder = seq(1, 20,1)
Hand <-function(){
x <- sample(20,20, replace=F);
is.element(TRUE, x==inorder)
}
r <-replicate(10000, Hand())
sum(r==FALSE)/10000
```

```
## [1] 0.3602
```

```
1/exp(1)
```

```
## [1] 0.3678794
```

Exercise 1.4 (Boxes and balls). A box initially contains 8 blue balls and 4 red balls. Terry closes his eyes and randomly picks a ball and then follows the following procedure: if it is a blue ball, he puts the blue ball back in, and also adds another blue ball; if it is a red ball, then he removes the red ball from the bin and in addition, removes another red ball. Suppose you did not see what Terry did and close your eyes and pick a ball at random and it is red; then what is the probability that Terry picked a blue ball? Use R.

Solution. First, we code one iteration of this procedure, which reports the first ball, and the second ball. We will use Bernoulli random variables, whereby *zero* will mean a blue ball, and *one* will mean a red ball, has been drawn. Next, we play the *game* a large number of (independent) times.

We also, store as two separate vectors, the outcome of the first pick, and the second pick. We want to count the number of times the second pick resulted in a red ball, and out of those times, the first pick was a blue ball, and then take the ratio. We write a simple loop to accomplish this.

```
game <-function() {
x = rbinom(1,1, 4/12);
y = rbinom(1,1, 2/10);
if (x==0)
y = rbinom(1,1, 4/13);
c(x,y)
}

z = replicate(10000, game());
F = z[1,]
S = z[2,]

br=0
for (i in 1:10000)
```

```
if (S[i]==1 && F[i]==0)
br <- br+1
br
```

```
## [1] 2116
```

```
r=sum(S==1)
r
```

```
## [1] 2782
```

```
br/r
```

```
## [1] 0.7606039
```

Exercise 1.5 (Uniform random variables). Let $(U_i)_{i \in \mathbb{Z}^+}$ be a sequence of independent random variables that are uniformly distributed in $[0, 1]$. Let

$$S_n = X_1 + \cdots + X_n.$$

Let

$$T = \inf\{n \geq 1 : S_n > 1\}$$

so that T is the first time the sum is greater than 1. Use R to compute $\mathbb{E}T$. You will not be disappointed.

Exercise 1.6 (Histograms). Carefully explain why plotting a (probability) histogram of some sample data may give an approximation of the probability density function of some random variable.

Exercise 1.7 (Integration). Let $f(x) = x^2$. Compute

$$\int_0^2 f(x) dx$$

by appealing to the law of large numbers and running R simulations.

2 Summary

We reviewed the law of large numbers and gave some proofs of some basic results. We showed how the law of large numbers can be used in conjugation with R to compute or estimate various probabilities arising from elementary probability theory.

3 Version: 27 April 2021

- Rmd Source