Large Numbers and Convergence of Random Variables

Benjamin Good *

*Physics Department, Swarthmore College, 500 College Ave., Swarthmore PA, 19081

This paper aims to introduce the concept of random variables in a semi-formal manner (ie, without the formal machinery of measure theory). In addition, we explain the various notions of convergence used for random variables and state the law of large numbers.

Introduction

Although we tend to picture a random variable X as a number, the value of X depends on what happens in the world. We might try to define a mathematical notion of "randomness" to capture this behavior, but it proves almost impossible to do so without the use of a lot of circular definitions. Instead, mathematicians define the random variable X as a *function* from the space of events to $\mathbb{R}$. If we denote the set of all events by Ω, then $X : \Omega \to \mathbb{R}$ and we should technically write $X(\omega)$ instead of X. One can think of a random variable as a "black box": you tell X what happened in the real world, and it will give you a number back. This is helpful because it pushes the tricky concept of randomness (should it even exist) back into the real world and lets us concentrate on precise mathematical statements.

Convergence

With this definition in hand, we can now discuss the notion of convergence. Let X be a random variable and let $X_1, X_2, \ldots$ be a sequence of random variables. Since these are really just functions over Ω, we can simply apply the definition of convergence of a function from analysis. The sequence X_n **converges to** X (written $X_n \to X$) if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \qquad [1]$$

for every $\omega \in \Omega$. In other words, $X_n \to X$ if for every $\omega \in \Omega$ and $\epsilon > 0$, there exists an N such that $|X_n(\omega) - X(\omega)| < \epsilon$ for all $n \geq N$. A more intuitive way to view the definition is that no matter what happens in the real world, the random variables take on values such that the sequence converges. In probability theory, this is called **sure convergence**.

Almost Sure Convergence

If we weaken the requirements slightly, we arrive at a second notion of convergence. The sequence X_n **converges almost surely** to X (written $X_n \xrightarrow{\text{a.s.}} X$) if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \qquad [2]$$

or alternatively, if

$$P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}) = 0 \qquad [3]$$

In other words, $X_n \xrightarrow{\text{a.s.}} X$ if the set of events that do not lead to convergence has measure zero. Almost sure convergence is used much more often in probability theory than sure convergence. This is because we would never expect to observe those events where almost sure convergence fails (they have probability zero) so we have no need for the stronger form.[1]

Convergence in Probability

There is a still weaker form of convergence which relies even more on probabilistic ideas. The sequence X_n **converges in probability** to X (written $X_n \xrightarrow{\text{P}} X$) if

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \qquad [4]$$

for every $\epsilon > 0$. In other words, $X_n \xrightarrow{P} X$ if for every $\epsilon > 0$ and $\delta > 0$, there exists an N such that $P(|X_n - X| > \epsilon) < \delta$ for all $n \geq N$. Almost sure convergence implies convergence in probability, although the latter is used more often in introductory textbooks because it is usually easier to demonstrate for a given sequence. However, convergence in probability does not rule out the case where $P(\exists m \geq N : |X_m - X| > \epsilon) = 1$. In other words, a sequence of random variables that converges in probability can still have an infinite number of violations of the convergence inequality.

Convergence in Distribution

The final mode of convergence relies exclusively on the idea of probability. Let $F_n(a) = P(X_n < a)$ and $F(a) = P(X < a)$. The sequence X_n **converges in distribution** to X (written $X_n \xrightarrow{\text{D}} X$) if

$$\lim_{n\to\infty} F_n(a) = F(a) \qquad [5]$$

for all a such that $F(a)$ is continuous. Convergence in distribution forms the basis of the Central Limit Theorem. Convergence in probability implies convergence in distribution. Thus, we have a "hierarchy" of convergence definitions:

$$X_n \to X \Rightarrow X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{\text{P}} X \Rightarrow X_n \xrightarrow{\text{D}} X \qquad [6]$$

The Law of Large Numbers

We can now state the law of large numbers in both its strong and weak form. Let $X_1, X_2, \ldots$ be i.i.d. with $E(|X_i|) < \infty$ and $E(X_i) = \mu$ and let $S_n = X_1 + X_2 + \ldots + X_n$.

Weak Law of Large Numbers: $S_n/n \xrightarrow{\text{P}} \mu$ as $n \to \infty$.

Strong Law of Large Numbers: $S_n/n \xrightarrow{\text{a.s.}} \mu$ as $n \to \infty$.

Ergodicity. If $X_1, X_2, \ldots$ represent samples observed over time, one can interperet the law of large numbers as saying that the "time average" of these samples converges to the "state space average" (given by the expectation) when the length of time is long enough.

[1] This sounds kind of circular to me... how do we know what probability zero means in terms of "what we observe"?

Probabilities as Frequencies. Let A be some event. If we define X_i by

$$X_i = \begin{cases} 1 : \text{event } A \text{ occurs on the } i\text{th attempt} \\ 0 : \text{else} \end{cases} \qquad [\mathbf{7}]$$

then it is clear that $E(|X_i|) = E(x) = P(A) < \infty$. Let $n_A = X_1 + X_2 + \ldots + X_n$ be the number of times that event A occurs in n attempts. The law of large numbers yields

$$\frac{a_A}{n} \xrightarrow{\text{P / a.s.}} P(A) \qquad [\mathbf{8}]$$

Thus, the observed frequency of an event converges to its probability. Some even use this as the *definition* of probability.[2]

[2] This is also slightly circular because the notion of probability was used in the definition of convergence.