IEOR 6711: Stochastic Models I

Fall 2013, Professor Whitt

Lecture Notes, Tuesday, September 3

Laws of Large Numbers

1 Overview

We start by stating the two principal laws of large numbers: the strong and weak forms, denoted by SLLN and WLLN. We want to be clear in our understanding of the statements; that leads us to a careful definition of a **random variable** and an examination of the basic **modes of convergence for a sequence of random variables**. We also want to focus on the proofs, but in this course (as in the course textbook) we consider only relatively simple **proofs that apply under extra moment conditions**. Even with these extra conditions, important proof techniques appear, which relate to the basic axioms of probability, in particular, to **countable additivity**, which plays a role in understanding and proving the **Borel-Cantelli lemma** (p. 4) . We think that it is helpful to fucus on these more elementary cases before considering the most general conditions.

Key reading for this first week: §§1.1-1.3, 1.7-1.8, the Appendix, pp. 56-58.

2 The Classical Laws of Large Numbers

Theorem 2.1 *Let $X_1, X_2, \ldots$ be IID random variables. Let $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$, $S_0 \equiv 0$, be the associated partial sums. If $E[|X_1|] < \infty$, then*

(a) ***SLLN***

$$\frac{S_n}{n} \to E[X_1] \quad \textit{as} \quad n \to \infty \quad \textit{w. p. 1} .$$

(b) ***WLLN***

$$\frac{S_n}{n} \Rightarrow E[X_1] \quad \textit{as} \quad n \to \infty ,$$

where $\equiv$ denotes equality by definition, w.p.1 is convergence with probability 1 (almost sure convergence) and $\Rightarrow$ denotes convergence in distribution.

Definition 2.1 (*convergence in distribution*) *There is convergence $Y_n \Rightarrow Y$ if the associated probability distributions of these random variables converge, i.e., if $P_{Y_n} \to P_Y$ as $n \to \infty$ or, equivalently, if the associated cumulative distribution functions (cdf's) converge, i.e., if $F_{Y_n}(x) \to F_Y(x)$ as $n \to \infty$ for all x that are continuity points of the function $F_Y(x)$, where $F_Y(x) \equiv P(Y \leq x)$. (More on this later.)*

Proof of (a) under extra condition $E[X_1^4] < \infty$. p. 56 of Ross. Draws heavily on Borel-Cantelli and thus §1.1.

Proof of (b) under extra condition $E[X_1^2] < \infty$. Two parts: (i)

$$E\left[\left(\frac{S_n}{n} - E[X_1]\right)^2\right] = Var\left(\frac{S_n}{n}\right) = \frac{Var(X_1)}{n} \to 0 \quad \text{as} \quad n \to \infty .$$

(ii) Convergence in mean squared (L_2) implies convergence in probability, which in turn implies convergence in distribution. In fact, convergence in probability is equivalent to convergence in distribution when the limit is constant (deterministic or non-random). Chebychev's inequality (which is a special case of Markov's inequality, Lemma 1.7.1 in the book) shows convergence in mean squared implies convergence in probability. ■

We remark that the IID condition can also be relaxed in a variety of ways; there is a large literature.

3 Random Variables and Functions of Random Variables

To properly understand why there are two versions of the LLN, it is necessary to understand what is a random variable and the possible modes of convergence for a sequence of random variables.

(i) What is a **random variable**?

A (real-valued) random variable, often denoted by X (or some other capital letter), is a **function** mapping a probability space (S, P) into the real line $\mathbb{R}$. This is shown in Figure 1. (It is also common to write Ω for the sample space and ω for an element in that set.) Associated with each point s in the domain S the function X assigns one and only one value $X(s)$ in the range $\mathbb{R}$. (The set of possible values of $X(s)$ is usually a proper subset of the real line; i.e., not all real numbers need occur. If S is a finite set with m elements, then $X(s)$ can assume at most m different values as s varies in S.)

As such, a random variable has a probability distribution. We usually do not care about the underlying probability space, and just talk about the random variable itself, but it is good to know the full formalism. The distribution of a random variable is defined formally in the obvious way

$$F(t) \equiv F_X(t) \equiv P(X \le t) \equiv P(\{s \in S : X(s) \le t\}) \ ,$$

where again $\equiv$ means "equality by definition," P is the probability measure on the underlying sample space S and $\{s \in S : X(s) \le t\}$ is a subset of S, and thus an *event* in the underlying sample space S. See Section 1.1 of Ross; he puts this out very quickly. (Key point: recall that P attaches probabilities to events, which are subsets of S.)

If the underlying probability space is discrete, so that for any event E in the sample space S we have

$$P(E) = \sum_{s \in E} p(s),$$

where $p(s) \equiv P(\{s\})$ is the *probability mass function* (pmf), then X also has a pmf p_X on a new sample space, say S_1, defined by

$$p_X(r) \equiv P(X = r) \equiv P(\{s \in S : X(s) = r\}) = \sum_{s \in \{s \in S : X(s) = r\}} p(s) \quad \text{for} \quad r \in S_1. \qquad (1)$$

Example 3.1 (*roll of two dice*) Consider a random roll of two dice. The natural sample space is

$$S \equiv \{(i, j) : 1 \le i \le 6, 1 \le j \le 6\},$$

where each of the 36 points in S is assigned equal probability $p(s) = 1/36$. The random variable X might record the sum of the values on the two dice, i.e., $X(s) \equiv X((i, j)) = i + j$. Then the new sample space is

$$S_1 = \{2, 3, 4, \ldots, 12\}.$$

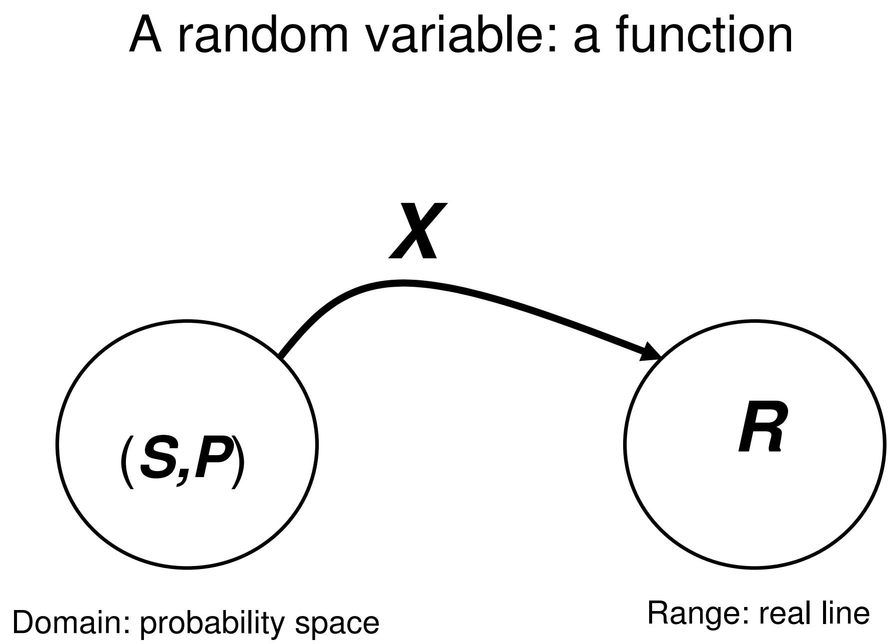

Figure 1: A (real-valued) random variable is a function mapping a probability space into the real line.

In this case, using formula (1), we get the pmf of X being $p_X(r) \equiv P(X = r)$ for $r \in S_1$, where

$$\begin{aligned}
p_X(2) &= p_X(12) = 1/36, \\
p_X(3) &= p_X(11) = 2/36, \\
p_X(4) &= p_X(10) = 3/36, \\
p_X(5) &= p_X(9) = 4/36, \\
p_X(6) &= p_X(8) = 5/36, \\
p_X(7) &= 6/36.
\end{aligned}$$

(ii) What is a **function of a random variable**?

Given that we understand what is a random variable, we are prepared to understand what is a function of a random variable. Suppose that we are given a random variable X mapping the probability space (S, P) into the real line $\mathbb{R}$ and we are given a function h mapping $\mathbb{R}$ into $\mathbb{R}$. Then $h(X)$ is a function mapping the probability space (S, P) into $\mathbb{R}$. As a consequence, $h(X)$ is itself a new random variable, i.e., a new function mapping (S, P) into $\mathbb{R}$, as depicted in Figure 2.

As a consequence, the distribution of the new random variable $h(X)$ can be expressed in different (equivalent) ways:

$$F_{h(X)}(t) \equiv P(h(X) \leq t) \equiv P(\{s \in S : h(X(s)) \leq t\}),$$

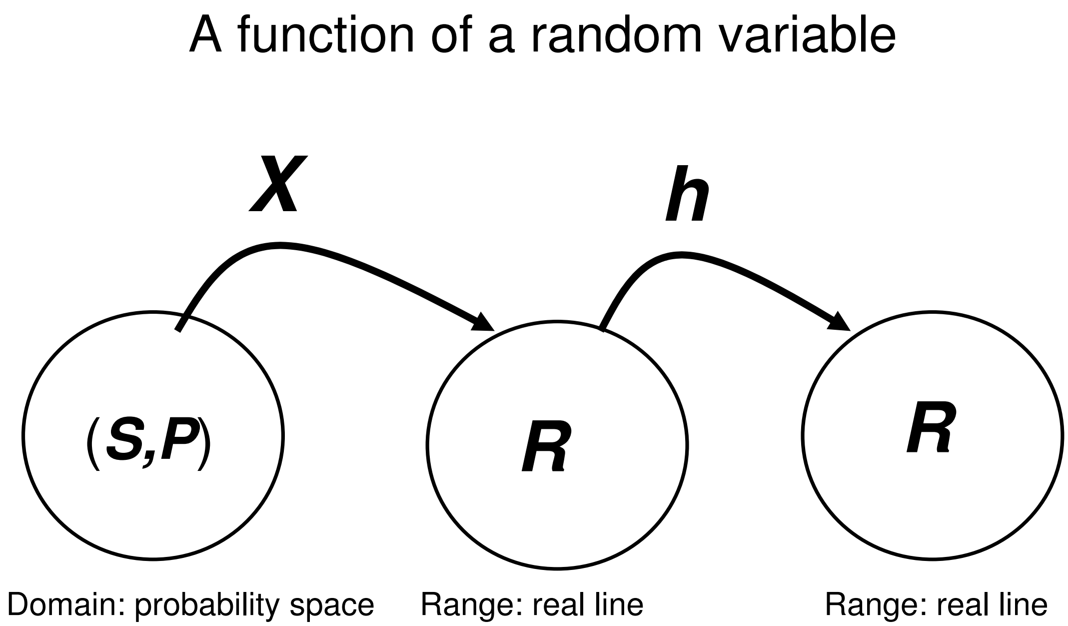

Figure 2: A (real-valued) function of a random variable is itself a random variable, i.e., a function mapping a probability space into the real line.

$$
\begin{aligned}
&\equiv P_X(\{r \in \mathbb{R} : h(r) \le t\}), \\
&\equiv P_{h(X)}(\{k \in \mathbb{R} : k \le t\}),
\end{aligned}
$$

where P is the probability measure on S in the first line, P_X is the probability measure on $\mathbb{R}$ (the distribution of X) in the second line and $P_{h(X)}$ is the probability measure on $\mathbb{R}$ (the distribution of the random variable $h(X)$ in the third line.

Example 3.2 (*more on the roll of two dice*) As in Example 3.1, consider a random roll of two dice. There we defined the random variable X to represent the sum of the values on the two rolls. Now let

$$h(x) = |x - 7|,$$

so that $h(X) \equiv |X - 7|$ represents the absolute difference between the observed sum of the two rolls and the average value 7. Then $h(X)$ has a pmf on a new probability space $S_2 \equiv \{0, 1, 2, 3, 4, 5\}$. In this case, using formula (1) yet again, we get the pmf of $h(X)$ being $p_{h(X)}(k) \equiv P(h(X) = k) \equiv P(\{s \in S : h(X(s)) = k\})$ for $k \in S_2$, where

$$
\begin{aligned}
p_{h(X)}(5) &= P(h(X) = 5) \equiv P(|X - 7| = 5) = 2/36 = 1/18, \\
p_{h(X)}(4) &= P(h(X) = 4) \equiv P(|X - 7| = 4) = 4/36 = 2/18, \\
p_{h(X)}(3) &= P(h(X) = 3) \equiv P(|X - 7| = 3) = 6/36 = 3/18, \\
p_{h(X)}(2) &= P(h(X) = 2) \equiv P(|X - 7| = 2) = 8/36 = 4/18,
\end{aligned}
$$

$$
\begin{aligned}
p_{h(X)}(1) &= P(h(X)=1) \equiv P(|X-7|=1) = 10/36 = 5/18, \\
p_{h(X)}(0) &= P(h(X)=0) \equiv P(|X-7|=0) = 6/36 = 3/18.
\end{aligned}
$$

In this setting we can compute probabilities for events associated with $h(X) \equiv |X-7|$ in three ways: using each of the pmf's p, p_X and $p_{h(X)}$.

(iii) How do we compute the **expectation** (or expected value) of a (probability distribution) or a random variable?

See Section 1.3. The expected value of a discrete probability distribution P is

$$
\text{expected value} \quad = \quad \text{mean} \quad = \sum_k kP(\{k\}) = \sum_k kp(k) \ ,
$$

where P is the probability measure on S and p is the associated pmf, with $p(k) \equiv P(\{k\})$. The expected value of a discrete random variable X can be written in two ways, as shown in the two lines below:

$$
\begin{aligned}
E[X] &= \sum_k kP(X=k) = \sum_k kp_X(k) \\
&= \sum_{s \in S} X(s)P(\{s\}) = \sum_{s \in S} X(s)p(s) \ .
\end{aligned}
$$

In the continuous case, with probability density functions (pdf's), we have corresponding formulas, but the story gets more complicated, involving calculus for computations. The expected value of a continuous probability distribution P with density f is

$$
\text{expected value} \quad = \quad \text{mean} \quad = \int_{s \in S} xf(x)\,dx \ .
$$

The expected value of a continuous random variable X with pdf f_X is

$$
E[X] = \int_{-\infty}^{\infty} xf_X(x)\,dx = \int X(s)f(s)\,ds \ ,
$$

where f is the pdf on S and f_X is the pdf "induced" by X on $\mathbb{R}$.

(iv) How do we compute the **expectation of a function of a random variable**?

Now we need to put everything above together. For simplicity, suppose S is a finite set, so that X and $h(X)$ are necessarily finite-valued random variables. Then we can compute the expected value $E[h(X)]$ in **three different ways**:

$$
\begin{aligned}
E[h(X)] &= \sum_{s \in S} h(X(s))P(\{s\}) = \sum_{s \in S} h(X(s))p(s) \\
&= \sum_{r \in \mathbb{R}} h(r)P(X=r) = \sum_{r \in \mathbb{R}} h(r)p_X(r) \\
&= \sum_{t \in \mathbb{R}} tP(h(X)=t) = \sum_{t \in \mathbb{R}} tp_{h(X)}(t) \ ,
\end{aligned}
$$

where $p(s) \equiv P(\{s\})$ is the pmf associated with P on S, while $p_X(r)$ is the pmf of X and $p_{h(X)}(t)$ is the pmf of $h(X)$.

Similarly, we have the following expressions when all these probability distributions have probability density functions (the continuous case). First, suppose that the underlying probability distribution (measure) P on the sample space S has a probability density function (pdf) f. Then, under regularity conditions, the random variables X and $h(X)$ have probability density functions f_X and $f_{h(X)}$. Then we have:

$$
\begin{aligned}
E[h(X)] &= \int_{s \in S} h(X(s)) f(s)\, ds \\
&= \int_{-\infty}^{\infty} h(r) f_X(r)\, dr \\
&= \int_{-\infty}^{\infty} t f_{h(X)}(t)\, dt \ .
\end{aligned}
$$

4 Implications for the LLN

So what does our study of random variables imply for the LLN? We see that the SLLN states that $S_n/n \to E[X_1]$ as $n \to \infty$ w.p.1. We have almost sure convergence of the sequence of functions $\{S_n/n\}$, all defined on the underlying sample space S. That is, for each s in a subset of S having probability 1, we have convergence of the sequence of numbers $S_n(s)/n \to E[X_1]$ as $n \to \infty$. All this is defined precisely in terms of the basic notion of the convergence of a sequence of numbers.

On the other hand, the WLLN states that $S_n/n \Rightarrow E[X_1]$ as $n \to \infty$. That means that the associated probability distributions converge. The statement concerns the probability distributions on the range of the functions (random variables). That is, $F_{S_n/n}(x) \to F_{E[X_1]}(x)$ and $n \to \infty$ for all x that are continuity points of the limiting cdf $F_{E[X_1]}(x)$. The only value that is *not* a continuity point is the mean itself $E[X_1]$, where the cdf has a unit jump.

We often are satisfied with the WLLN, because we often only care about the distributions of the random variables.