Lecture 5 : Weak Law of Large Numbers

STAT205 *Lecturer: Jim Pitman* Scribe: *Moorea Brega* <brega@stat.berkeley.edu>

References: Durrett [Sections 1.4, 1.5]

5.1 Independence

Denote by $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. Consider a sequence of random variables $X_1, X_2, \ldots$ with some distributions on $(\mathbb{R}, \mathcal{B})$,

$$P(X_i \leq x) = F_i(x)$$

where F_i is the cumulative distribution function of X_i. If $F_i(x) \equiv F(x)$, we say the X_i are *identically distributed.* If

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \leq x_i\}\right) = \prod_{i=1}^{n} F_i(x_i) \tag{5.1}$$

for all choices of $x_i \in \mathbb{R}$, $X_1, X_2, \ldots, X_n$ are *independent.*

Remark 5.1 *For any choice of cumulative distribution functions F_i there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables X_i which are independent.*

The remark follows from our discussion of product spaces. Consider only two random variables, X_1 and X_2. As seen in Lecture 4, there is a product measure on $\Omega = \mathbb{R} \times \mathbb{R}$ with projection maps X_i, $X_i(\omega) = x_i$ where $\omega = (x_1, x_2)$. This idea extends easily to finite n, with $\Omega = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} = \mathbb{R}^n$ and projection maps $X_i(\omega) = x_i$ for $i = 1, 2, \ldots, n$. The following is a simple example of the finite n case where the random variables have densities.

Example 5.2 *Consider*

$$F_i(x) = \int_{-\infty}^{x} f_i(y)dy$$

where f_i is the density of X_i. Then, the joint law of $(X_1, \ldots, X_n)$ on $\mathbb{R}^n$ has density

$$h(x) = \prod_{i=1}^{n} f_i(x_i)$$

with respect to Lebesgue measure $dx_1, dx_2, \ldots dx_n$.

For an infinite sequence of random variables, $X_1, X_2, \ldots$, we must consider the infinite product space, $\Omega = \mathbb{R} \times \mathbb{R} \times \cdots = \mathbb{R}^\infty$ equipped with the projections $X_i(\omega) = x_i$ where $\omega = (x_1, x_2, \ldots)$.

Remark 5.3 *It is a nontrivial fact of measure theory that there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is the product σ-field generated by X_i, so that for every n, $X_1, X_2, \ldots X_n$ are independent as in equation (5.1). The proof uses Kolmogorov's extension theorem.*

5.1.1 Construction of Independent Random Variables

As a small digression, we consider a method of constructing independent random variables X_i on $([0,1], Leb)$. As discussed previously (Lecture 2), we can create a random variable X_1 with distribution F_1 by using the inverse of the distribution, $X_1 = F_1^{-1}(U_1)$, where U_1 is a uniform$[0,1]$ random variable.

Using this method, to generate n independent random variables X_i with distributions F_i, we will need to start with n independent uniform random variables. The following is a useful method for generating any number of independent uniforms from a single uniform$[0,1]$ random variable. First, we consider the simple case of generating two i.i.d. uniform random variables from a single uniform U. We begin by considering the binary expansion of U,

$$U = \frac{D_1}{2} + \frac{D_2}{2^2} + \frac{D_3}{2^3} + \ldots$$

where D_i is the i^{th} digit in the binary expansion. Each D_i takes on the value 0 or 1 with equal probability on subintervals of [0,1]. Thus, if we then let

$$U_1 = \frac{D_1}{2} + \frac{D_3}{2^2} + \frac{D_5}{2^3} + \ldots$$
$$U_2 = \frac{D_2}{2} + \frac{D_4}{2^2} + \frac{D_6}{2^3} + \ldots,$$

the random variables U_1 and U_2 are uniform$[0,1]$ and independent (a result of the fact that functions of disjoint collections of independent random variables are independent). This method can be used to generate a finite or an infinite sequence of independent uniform random variables. For an infinite sequence of random variables, we consider

$$\mathbb{N} = \bigcup_{i=1}^{\infty} N_i$$

where $|N_i| = \infty$ and the N_i are disjoint. The construction above is repeated with U_i defined using the digits D_j where $j \in N_i$.

5.2 Weak Law of Large Numbers

The Weak Law of Large Numbers is a statement about sums of independent random variables. Before we state the WLLN, it is necessary to define convergence in probability. We say Y_n *converges in probability* to Y and write $Y_n \xrightarrow{\mathbb{P}} Y$ if, $\forall \epsilon > 0$,

$$P(\omega : |Y_n(\omega) - Y(\omega)| > \epsilon) \to 0, \quad n \to \infty.$$

Theorem 5.4 (Weak Law of Large Numbers) *Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $E|X| < \infty$ and define $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} EX.$$

Proof: In this proof, we employ the common strategy of first proving the result under an L^2 condition (i.e. assuming that the second moment is finite), and then using truncation to get rid of the extraneous moment condition.

First, we assume $EX^2 < \infty$. Because the X_i are iid,

$$\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\mathrm{Var}(X)}{n}.$$

By Chebychev's inequality, $\forall\ \epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - EX\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\mathrm{Var}(X)}{n\epsilon^2} \to 0.$$

Thus, $\frac{S_n}{n} \xrightarrow{\mathbb{P}} EX$ under the finite second moment condition. To transition from L^2 to L^1, we use truncation. For $0 < x < \infty$ let

$$X_{xk} = X_k \mathbf{1}_{(|X_k| \leq x)}$$
$$Y_{xk} = X_k \mathbf{1}_{(|X_k| > x)}$$

Then, we have $X_k = X_{xk} + Y_{xk}$ and

$$\begin{aligned}\frac{S_n}{n} &= \frac{1}{n}\sum_{k=1}^{n} X_{xk} + \frac{1}{n}\sum_{k=1}^{n} Y_{xk} \\ &= U_{xn} + V_{xn}\end{aligned}$$

Applying Jensen's inequality, we have

$$E\left|\frac{1}{n}\sum_{k=1}^{n} Y_{xk}\right| \leq \frac{1}{n}\sum_{k=1}^{n} E|Y_{xk}| = E(|X|\mathbf{1}_{(|X|>x)})$$

and by DCT,

$$E(|X|\mathbf{1}_{(|X|>x)}) \to 0, \qquad x \to \infty.$$

Fix $1 > \epsilon > 0$ and choose x such that

$$E\left(|X|\mathbf{1}_{(|X|>x)}\right) = E|Y_{x1}| < \epsilon^2.$$

Let $\mu_x = E(X_{x1})$ and $\mu = E(X)$. Then, we also have

$$|\mu_x - \mu| \leq |E(Y_{x1})| < \epsilon^2 < \epsilon.$$

Let $B_n = \{|U_{xn} - \mu_x| > \epsilon\}$ and $C_n = \{|V_{xn}| > \epsilon\}$. Noting that $E(X_{xk}^2) \leq x^2 < \infty$, we can apply the Weak Law of Large Numbers to U_{xn}. Thus, we choose $N > 0$ such that $\forall\ n > N$,

$$\mathbb{P}(B_n) = \mathbb{P}(|U_{nx} - \mu_x| > \epsilon) < \epsilon.$$

Now, by Chebyshev's inequality, we also have

$$\mathbb{P}(C_n) = \mathbb{P}(|V_{xn}| > \epsilon) \leq \frac{E|V_{xn}|}{\epsilon} \leq \frac{E|Y_{x1}|}{\epsilon} \leq \epsilon$$

But on $B_n^c \cap C_n^c = (B_n \cup C_n)^c$, we have $|U_{xn} - \mu_x| \leq \epsilon$ and $|V_{xn}| \leq \epsilon$, and therefore

$$\left|\frac{S_n}{n} - \mu\right| \le |U_{xn} - \mu_x| + |V_{xn}| + |\mu_x - \mu| \le 2\epsilon + \epsilon^2 \le 3\epsilon.$$

Thus, $\forall\ n > N$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - EX\right| > 3\epsilon\right) \le \mathbb{P}(B_n \cup C_n) \le 2\epsilon.$$

$\blacksquare$