

**LAWS OF LARGE NUMBERS**

**Contents**

1. Useful inequalities
2. The weak law of large numbers
3. The central limit theorem

**1 USEFUL INEQUALITIES**

**Markov inequality:** If  $X$  is a nonnegative random variable, then  $\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a$ .

**Proof:** Let  $I$  be the indicator function of the event  $\{X \geq a\}$ . Then,  $aI \leq X$ . Taking expectations of both sides, we obtain the claimed result.  $\square$

**Chebyshev inequality:**  $\mathbb{P}(|X - \mathbb{E}[X]| \geq \epsilon) \leq \text{var}(X)/\epsilon^2$ .

**Proof:** Apply the Markov inequality, to the random variable  $|X - \mathbb{E}[X]|^2$ , and with  $a = \epsilon^2$ .  $\square$

**2 THE WEAK LAW OF LARGE NUMBERS**

Intuitively, an expectation can be thought of as the average of the outcomes over an infinite repetition of the same experiment. If so, the observed average in a finite number of repetitions (which is called the **sample mean**) should approach the expectation, as the number of repetitions increases. This is a vague statement, which is made more precise by so-called laws of large numbers.

**Theorem 1. (Weak law of large numbers)** Let  $X_n$  be a sequence of i.i.d. random variables, and assume that  $\mathbb{E}[|X_1|] < \infty$ . Let  $S_n = X_1 + \cdots + X_n$ . Then,

$$\frac{S_n}{n} \xrightarrow{\text{i.p.}} \mathbb{E}[X_1].$$

This is called the “weak law” in order to distinguish it from the “strong law” of large numbers, which asserts, under the same assumptions, that  $X_n \xrightarrow{\text{a.s.}} \mathbb{E}[X_1]$ . Of course, since almost sure convergence implies convergence in probability, the strong law implies the weak law. On the other hand, the weak law can be easier to prove, especially in the presence of additional assumptions. Indeed, in the special case where the  $X_i$  have mean  $\mu$  and **finite variance**, Chebyshev’s inequality yields, for every  $\epsilon > 0$ ,

$$\mathbb{P}(|(S_n/n) - \mu| \geq \epsilon) \leq \frac{\text{var}(S_n/n)}{\epsilon^2} = \frac{\text{var}(X_1)}{n\epsilon^2},$$

which converges to zero, as  $n \rightarrow \infty$ , thus establishing convergence in probability.

Before we proceed to the proof for the general case, we note two important facts that we will use.

- (a) **First-order Taylor series expansion.** Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a function that has a derivative at zero, denoted by  $d$ . Let  $h$  be a function that represents the error in a first order Taylor series approximation:

$$g(\epsilon) = g(0) + d\epsilon + h(\epsilon).$$

By the definition of the derivative, we have

$$d = \lim_{\epsilon \rightarrow 0} \frac{g(\epsilon) - g(0)}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{d\epsilon + h(\epsilon)}{\epsilon} = d + \lim_{\epsilon \rightarrow 0} \frac{h(\epsilon)}{\epsilon}.$$

Thus,  $h(\epsilon)/\epsilon$  converges to zero, as  $\epsilon \rightarrow 0$ . A function  $h$  with this property is often written as  $o(\epsilon)$ . This discussion also applies to complex-valued functions, by considering separately the real and imaginary parts.

- (b) **A classical sequence.** Recall the well known fact

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a, \quad a \in \mathbb{R}. \quad (1)$$

We note (without proof) that this fact remains true even when  $a$  is a complex number. Furthermore, with little additional work, it can be shown that if

$\{a_n\}$  is a sequence of complex numbers that converges to  $a$ , then,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a.$$

**Proof of Theorem 1:** Let  $\mu = \mathbb{E}[X_1]$ . Fix some  $t \in \mathbb{R}$ . Using the assumption that the  $X_i$  are independent, and the fact that the derivative of  $\phi_{X_1}$  at  $t = 0$  equals  $i\mu$ , the characteristic function of  $S_n/n$  is of the form

$$\phi_n(t) = (\mathbb{E}[e^{itX_1/n}])^n = (\phi_{X_1}(t/n))^n = \left(1 + \frac{\mu it}{n} + o(t/n)\right)^n,$$

where the function  $o$  satisfies  $\lim_{\epsilon \rightarrow 0} o(\epsilon)/\epsilon = 0$ . Therefore,

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = e^{i\mu t}, \quad \forall t.$$

We recognize  $e^{i\mu t}$  as the characteristic function associated with a random variable which is equal to  $\mu$ , with probability one.

Applying Theorem 3 from the previous lecture (continuity of inverse transforms), we conclude that  $S_n/n$  converges to  $\mu$ , in distribution. Furthermore, as mentioned in the previous lecture, convergence in distribution to a constant implies convergence in probability.  $\square$

**Remark:** It turns out that the assumption  $\mathbb{E}[|X_1|] < \infty$  can be relaxed, although not by much. Suppose that the distribution of  $X_1$  is symmetric around zero. It is known that  $S_n/n \rightarrow 0$ , in probability, if and only if  $\lim_{n \rightarrow \infty} n\mathbb{P}(|X_1| > n) = 0$ . There exist distributions that satisfy this condition, while  $\mathbb{E}[|X_1|] < \infty$ . On the other hand, it can be shown that any such distribution satisfies  $\mathbb{E}[|X_1|^{1-\epsilon}] < \infty$ , for every  $\epsilon > 0$ , so the condition  $\lim_{n \rightarrow \infty} n\mathbb{P}(|X_1| > n) = 0$  is not much weaker than the assumption of a finite mean.

### 3 THE CENTRAL LIMIT THEOREM

Suppose that  $X_1, X_2, \dots$  are i.i.d. with common (and finite) mean  $\mu$  and variance  $\sigma^2$ . Let  $S_n = X_1 + \dots + X_n$ . The central limit theorem (CLT) asserts that

$$\frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a standard normal random variable. For a discussion of the uses of the central limit theorem, see the handout from [BT] (pages 388-394).

**Proof of the CLT:** For simplicity, suppose that the random variables  $X_i$  have zero mean and unit variance. Finiteness of the first two moments of  $X_1$  implies that  $\phi_{X_1}(t)$  is twice differentiable at zero. The first derivative is the mean (assumed zero), and the second derivative is  $-\mathbb{E}[X^2]$  (assumed equal to one), and we can write

$$\phi_X(t) = 1 - t^2/2 + o(t^2),$$

where  $o(t^2)$  indicates a function such that  $o(t^2)/t^2 \rightarrow 0$ , as  $t \rightarrow 0$ . The characteristic function of  $S_n/\sqrt{n}$  is of the form

$$(\phi_X(t/\sqrt{n}))^n = \left(1 - \frac{t^2}{2n} + o(t^2/n)\right)^n.$$

For any fixed  $t$ , the limit as  $n \rightarrow \infty$  is  $e^{-t^2/2}$ , which is the characteristic function  $\phi_Z$  of a standard normal random variable  $Z$ . Since  $\phi_{S_n/\sqrt{n}}(t) \rightarrow \phi_Z(t)$  for every  $t$ , we conclude that  $S_n/\sqrt{n}$  converges to  $Z$ , in distribution.  $\square$

The central limit theorem, as stated above, does not give any information on the PDF or PMF of  $S_n$ . However, some further refinements are possible, under some additional assumptions. We state, without proof, two such results.

- (a) Suppose that  $\int |\phi_{X_1}(t)|^r dt < \infty$ , for some positive integer  $r$ . Then,  $S_n$  is a continuous random variable for every  $n \geq r$ , and the PDF  $f_n$  of  $(S_n - \mu_n)/(\sigma\sqrt{n})$  converges pointwise to the standard normal PDF:

$$\lim_{n \rightarrow \infty} f_n(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \forall z.$$

In fact, convergence is uniform over all  $z$ :

$$\lim_{n \rightarrow \infty} \sup_z \left| f_n(z) - \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \right| = 0.$$

- (b) Suppose that  $X_i$  is a discrete random variable that takes values of the form  $a + kh$ , where  $a$  and  $h$  are constants, and  $k$  ranges over the integers. Suppose furthermore that  $X$  has zero mean and unit variance. Then, for any  $z$  of the form  $z = (na + kh)/\sqrt{n}$  (these are the possible values of  $S_n/\sqrt{n}$ ), we have

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n}}{h} \mathbb{P}(S_n = z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

MIT OpenCourseWare  
<http://ocw.mit.edu>

6.436J / 15.085J Fundamentals of Probability  
Fall 2008

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.