

# ECS315 2014/1    Part VI    Dr.Prapun

## 12 Limiting Theorems

### 12.1 Law of Large Numbers (LLN)

**Definition 12.1.** Let  $X_1, X_2, \dots, X_n$  be a collection of random variables with a common mean  $\mathbb{E}[X_i] = m$  for all  $i$ . In practice, since we do not know  $m$ , we use the numerical average, or **sample mean**,

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

in place of the true, but unknown value,  $m$ .

Q: Can this procedure of using  $M_n$  as an estimate of  $m$  be justified in some sense?

A: This can be done via the law of large number.

**12.2.** The law of large number basically says that if you have a sequence of **i.i.d** random variables  $X_1, X_2, \dots$ . Then the sample means  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$  will **converge to the actual mean as  $n \rightarrow \infty$** .

**12.3.** LLN is easy to see via the property of variance. Note that

$$\mathbb{E}[M_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \overbrace{\mathbb{E}X_i}^m = m$$

and

$$\text{Var}[M_n] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var} X_i = \frac{1}{n} \sigma^2, \quad (33)$$

$$= \left(\frac{1}{n}\right)^2 \text{Var}\left[\sum_{i=1}^n X_i\right]^{176}$$

$$\text{Var}[X+Y] = \text{Var} X + \text{Var} Y + 2 \text{Cov}[X, Y]$$

$$\text{Var}[X_1 + X_2 + X_3 + \dots + X_n] = \text{Var}\left[\sum_{i=1}^n X_i\right]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \mathbb{E}\left(\sum_{i=1}^n X_i\right)\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^n \underbrace{(X_i - \mathbb{E}X_i)}_{A_i}\right)^2\right]$$

$$= \mathbb{E}\left[(A_1 + A_2 + \dots + A_n)^2\right]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^n A_i\right)\left(\sum_{k=1}^n A_k\right)\right] = \sum_{i=1}^n \sum_{k=1}^n \mathbb{E}[A_i A_k]$$

$$= \sum_{i=1}^n \sum_{k=1}^n \text{cov}[X_i, X_k] = \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \text{cov}[X_i, X_k] + \sum_{i=1}^n \text{Var} X_i$$

Need to  
assume  
the  $X_i$ 's  
are

$$\downarrow \sum_{i=1}^n \text{Var} X_i$$

uncorrelated.

Remarks:

- (a) For (33) to hold, it is sufficient to have uncorrelated  $X_i$ 's.
- (b) From (33), we also have

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}\sigma. \quad (34)$$

In words, “when uncorrelated (or independent) random variables each having the same distribution are averaged together, the standard deviation is reduced according to the square root law.” [21, p 142].

**Exercise 12.4** (F2011). Consider i.i.d. random variables  $X_1, X_2, \dots, X_{10}$ . Define the sample mean  $M$  by

$$M = \frac{1}{10} \sum_{k=1}^{10} X_k.$$

Let

$$V_1 = \frac{1}{10} \sum_{k=1}^{10} (X_k - \mathbb{E}[X_k])^2.$$

and

$$V_2 = \frac{1}{10} \sum_{j=1}^{10} (X_j - M)^2.$$

Suppose  $\mathbb{E}[X_k] = 1$  and  $\text{Var}[X_k] = 2$ .

- (a) Find  $\mathbb{E}[M]$ . = 1
- (b) Find  $\text{Var}[M]$ .
- (c) Find  $\mathbb{E}[V_1]$ .
- (d) Find  $\mathbb{E}[V_2]$ .

## 12.2 Central Limit Theorem (CLT)

In practice, there are many random variables that arise as a sum of many other random variables. In this section, we consider the sum

$$S_n = \sum_{i=1}^n X_i \quad \begin{cases} \mathbb{E}[S_n] = nm \\ \text{Var}[S_n] = n\sigma^2 \end{cases} \quad (35)$$

where the  $X_i$  are i.i.d. with common mean  $m$  and common variance  $\sigma^2$ .

- Note that when we talk about  $X_i$  being i.i.d., the definition is that they are independent and identically distributed. It is then convenient to talk about a random variable  $X$  which shares the same distribution (pdf/pmf) with these  $X_i$ . This allow us to write

$$X_i \stackrel{\text{i.i.d.}}{\sim} X, \quad (36)$$

which is much more compact than saying that the  $X_i$  are i.i.d. with the same distribution (pdf/pmf) as  $X$ . Moreover, we can also use  $\mathbb{E}X$  and  $\sigma_X^2$  for the common expected value and variance of the  $X_i$ .

Q: How does  $S_n$  behave?

In the previous section, we consider the sample mean of identically distributed random variables. More specifically, we consider the random variable  $M_n = \frac{1}{n}S_n$ . We found that  $M_n$  will converge to  $m$  as  $n$  increases to  $\infty$ . Here, we don't want to rescale the sum  $S_n$  by the factor  $\frac{1}{n}$ .

**12.5** (Approximation of densities and pmfs using the CLT). The actual statement of the CLT is a bit difficult to state. So, we first give you the interpretation/insight from CLT which is very easy to remember and use:

**For  $n$  large enough, we can approximate  $S_n$  by a Gaussian random variable with the same mean and variance as  $S_n$ .**

Note that the mean and variance of  $S_n$  is  $nm$  and  $n\sigma^2$ , respectively. Hence, for  $n$  large enough we can approximate  $S_n$  by  $\mathcal{N}(nm, n\sigma^2)$ . In particular,

(a)  $F_{S_n}(s) \approx \Phi\left(\frac{s-nm}{\sigma\sqrt{n}}\right)$ .

(b) If the  $X_i$  are continuous random variable, then

$$f_{S_n}(s) \approx \frac{1}{\sqrt{2\pi}\sigma\sqrt{n}} e^{-\frac{1}{2}\left(\frac{s-nm}{\sigma\sqrt{n}}\right)^2}.$$

(c) If the  $X_i$  are integer-valued, then

$$P[S_n = k] = P\left[k - \frac{1}{2} < S_n \leq k + \frac{1}{2}\right] \approx \frac{1}{\sqrt{2\pi}\sigma\sqrt{n}} e^{-\frac{1}{2}\left(\frac{k-nm}{\sigma\sqrt{n}}\right)^2}.$$

[9, eq (5.14), p. 213].

The approximation is best for  $k$  near  $nm$  [9, p. 211].

**Example 12.6.** Approximation for Binomial Distribution: For  $X \sim \mathcal{B}(n, p)$ , when  $n$  is large, binomial distribution becomes difficult to compute directly because of the need to calculate factorial terms.

(a) When  $p$  is not close to either 0 or 1 so that the variance is also large, we can use CLT to approximate

$$P[X = k] \approx \frac{1}{\sqrt{2\pi \text{Var } X}} e^{-\frac{(k-\mathbb{E}X)^2}{2 \text{Var } X}} \quad (37)$$

$$= \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}. \quad (38)$$

This is called Laplace approximation to the Binomial distribution [25, p. 282].

(b) When  $p$  is small, the binomial distribution can be approximated by  $\mathcal{P}(np)$  as discussed in 8.45.

(c) If  $p$  is very close to 1, then  $n - X$  will behave approximately Poisson.

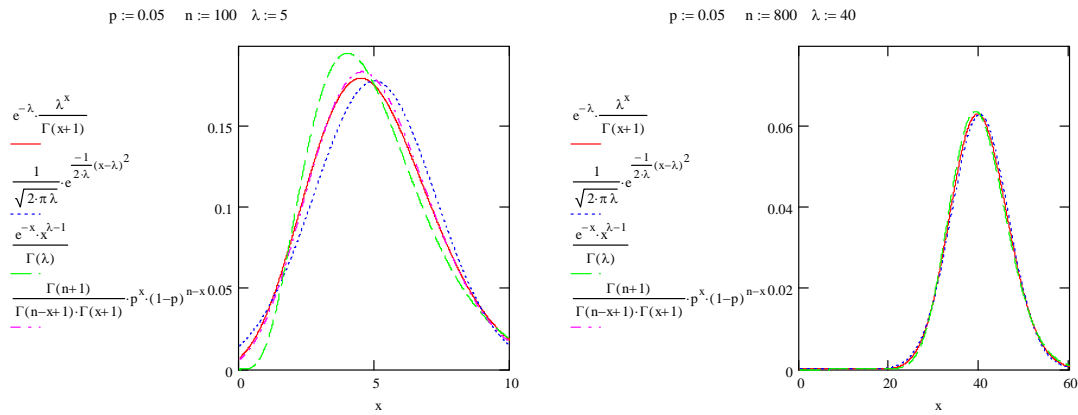


Figure 26: Gaussian approximation to Binomial, Poisson distribution, and Gamma distribution.

**Exercise 12.7** (F2011). Continue from Exercise 6.53. The stronger person (Kakashi) should win the competition if  $n$  is very large. (By the law of large numbers, the proportion of fights that Kakashi wins should be close to 55%.) However, because the results are random and  $n$  can not be very large, we can not guarantee that Kakashi will win. However, it may be good enough if the probability that Kakashi wins the competition is greater than 0.85.

We want to find the minimal value of  $n$  such that the probability that Kakashi wins the competition is greater than 0.85.

Let  $N$  be the number of fights that Kakashi wins among the  $n$  fights. Then, we need

$$P \left[ N > \frac{n}{2} \right] \geq 0.85. \tag{39}$$

Use the central limit theorem and Table 3.1 or Table 3.2 from [Yates and Goodman] to approximate the minimal value of  $n$  such that (39) is satisfied.