ECS315 2018/1 Part VI Dr.Prapun

12 Limiting Theorems

12.1 Law of Large Numbers (LLN)

Definition 12.1. Let $X_1, X_2, \ldots, X_n$ 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 $X_1, X_2, \ldots$. Then the sample means $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ will converge to the actual mean as $n \to \infty$.

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

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

and

$$\text{Var}[M_n] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}\, X_i = \frac{1}{n}\sigma^2, \tag{34}$$

Remarks:

(a) For (34) to hold, it is sufficient to have uncorrelated X_i's.

(b) From (34), we also have

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}\sigma. \tag{35}$$

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 $X_1, X_2, \ldots, X_{10}$. 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 $\mathbb{E}[M]$.

(b) Find $\text{Var}[M]$.

(c) Find $\mathbb{E}[V_1]$.

(d*) Find $\mathbb{E}[V_2]$.

12.5. In 1.21 and 1.23, we stated an application of LLN. Back then, we have a sequence of independent repeated trials of an experiment.

Let A be the event of interest. Let a Bernoulli RV X_k indicate whether the event A happens in the kth trial. Then, the X_k are i.i.d. with

$$\mathbb{E}X_k = 1 \times P(A) + 0 \times (1 - P(A)) = P(A).$$

Note also that $\sum_{k=1}^{n} X_k$ is the same as $N(A, n)$ defined in Definition 1.22. Both of them count the number of trials in which A occurs. Therefore, the sample mean

$$M_n = \frac{1}{n} \sum_{k=1}^{n} X_k = \frac{N(A, n)}{n}$$

is the same as the *relative frequency* of event A.

By LLN, we can now conclude that M_n will converge to $\mathbb{E}X_k = P(A)$ as $n \to \infty$. The same result was stated without proof in 1.23.

Example 12.6. Back to Example 1.19.

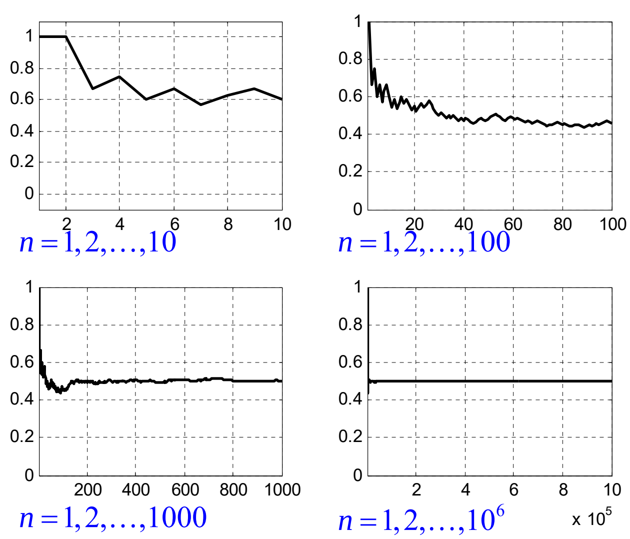

Figure 36: If a fair coin is flipped a large number of times, the proportion of heads will tend to get closer to 1/2 as the number of tosses increases.

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

$$S_n = \sum_{i=1}^{n} X_i \tag{36}$$

where the X_i are i.i.d. with common mean m and common variance σ^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

  $$X_i \overset{\text{i.i.d.}}{\sim} X, \tag{37}$$

  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 σ_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 ∞. Here, we don't want to rescale the sum S_n by the factor $\frac{1}{n}$.

12.7 (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.

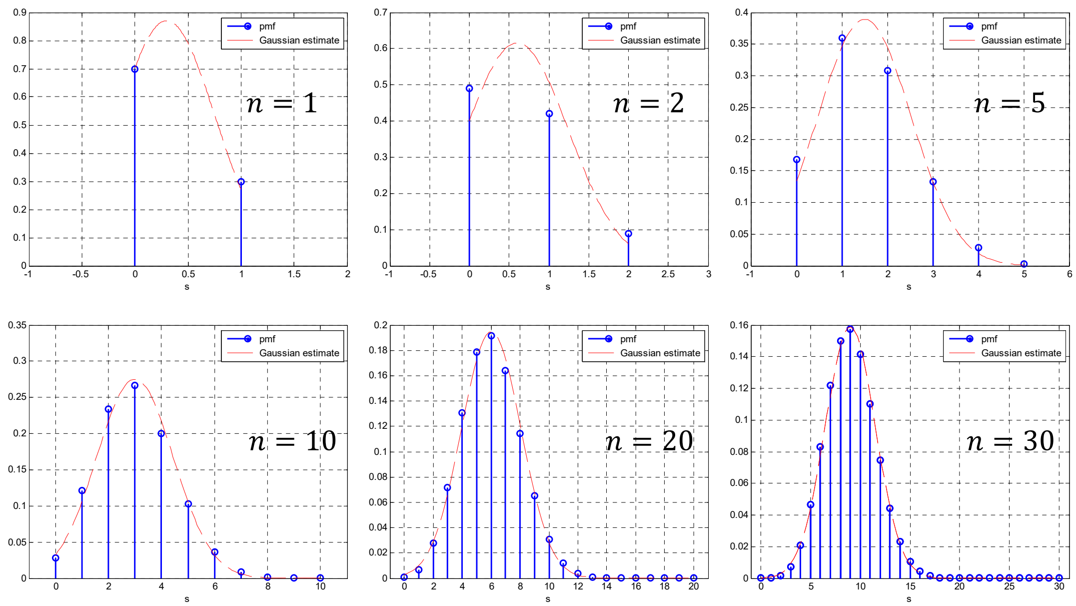

Figure 37: Gaussian approximation of the sum of i.i.d. Bernoulli random variables. The stem plots show the pmf of the sum $S_n = \sum_{k=1}^{n} X_k$ where $X_1, X_2, \ldots$ are i.i.d. Bernoulli(0.3) random variables.

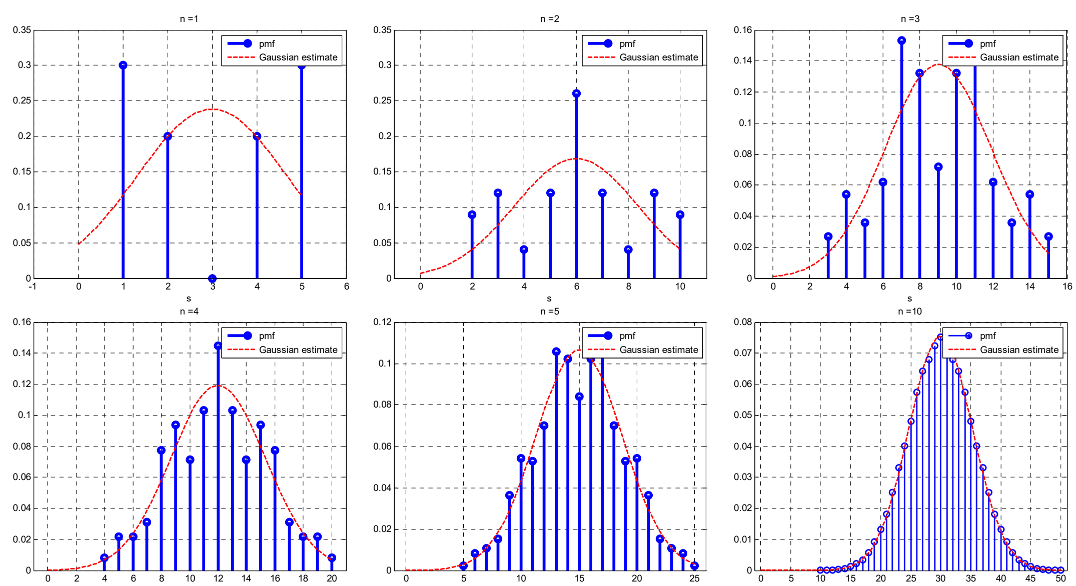

Figure 38: Gaussian approximation of the sum of i.i.d. discrete random variables. The stem plots show the pmf of the sum $S_n = \sum_{k=1}^{n} X_k$.

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 $\mathcal{N}\left(nm, n\sigma^2\right)$. 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 12.8. 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 approxmiate

$$P[X = k] \approx \frac{1}{\sqrt{2\pi \operatorname{Var} X}} e^{-\frac{(k-\mathbb{E}X)^2}{2 \operatorname{Var} X}} \tag{38}$$

$$= \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}. \tag{39}$$

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 8.55.

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