ECS315 2013/1 Part VII Dr.Prapun

14 Limiting Theorems

14.1 Law of Large Numbers (LLN)

Definition 14.1. 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,

$\bar{X}$

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

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.

14.2. 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 $n \to \infty$.

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

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

and

$$\operatorname{Var}[M_n] = \operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \overset{\sigma^2}{\operatorname{Var} X_i} = \frac{1}{n}\sigma^2, \tag{39}$$

$\operatorname{Var}[aX] = a^2 \operatorname{Var} X$

$\operatorname{Var}[X+Y+Z] = ?$

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

Remarks:

(a) For (39) to hold, it is sufficient to have uncorrelated X_i's.

(b) From (39), we also have

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}\sigma. \tag{40}$$

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 14.4 (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

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

and

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

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

(a) Find $\mathbb{E}\left[M\right]$.

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

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

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

14.2 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

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

$$S_n = \sum_{i=1}^{n} X_i \tag{41}$$

$$\mathrm{Var}[S_n] = \sum_{i=1}^{n} \overbrace{\mathrm{Var}\, X_i}^{\sigma^2} = n\sigma^2$$

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

$$X_i \overset{\text{i.i.d.}}{\sim} X, \tag{42}$$

  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?

For the S_n defined above, there are many cases for which we know the pmf/pdf of S_n.

Example 14.5. When the X_i are i.i.d. Bernoulli(p),

Example 14.6. When the X_i are i.i.d. $\mathcal{N}(m, \sigma^2)$,

Note that it is not difficult to find the characteristic function of S_n if we know the common characteristic function $\varphi_X(v)$ of the

X_i:

$$\varphi_{S_n}(v) = (\varphi_X(v))^n .$$

If we are lucky, as in the case for the sum of Gaussian random variables in Example 14.6 above, we get $\varphi_{S_n}(v)$ that is of the form that we know. However, $\varphi_{S_n}(v)$ will usually be something we haven't seen before or difficult to find the inverse transform. This is one of the reason why having a way to approximate the sum S_n would be very useful.

There are also some situations where the distribution of the X_i is unknown or difficult to find. In which case, it would be amazing if we can say something about the distribution of S_n.

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 $\frac{1}{n}$.

14.7 (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.

$S_n \approx \sim \mathcal{N}(nm, n\sigma^2)$

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

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

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

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

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

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

Example 14.8. 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

$$P[X=k] \approx \frac{1}{\sqrt{2\pi \operatorname{Var} X}} e^{-\frac{(k-\mathbb{E}X)^2}{2 \operatorname{Var} X}} \tag{43}$$

$$= \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}. \tag{44}$$

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.

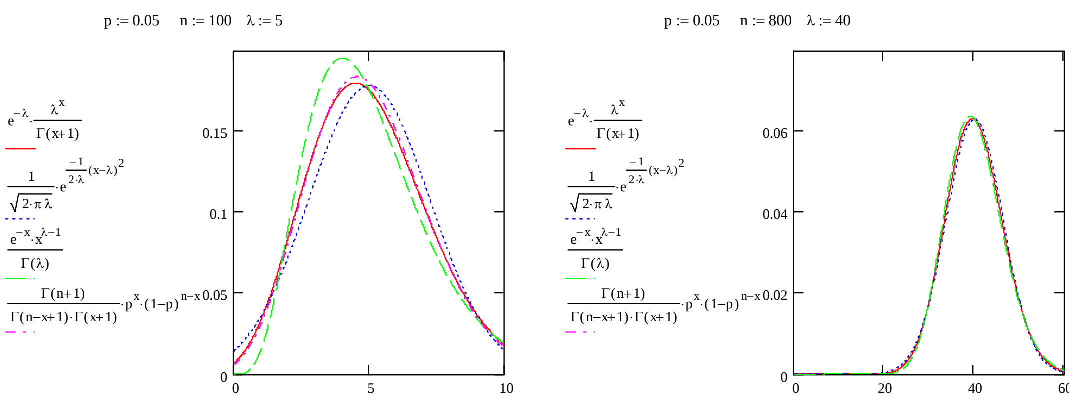

Figure 29: Gaussian approximation to Binomial, Poisson distribution, and Gamma distribution.

Exercise 14.9 (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

$$P\left[N > \frac{n}{2}\right] \geq 0.85. \tag{45}$$

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 (45) is satisfied.

14.10. A more precise statement for CLT can be expressed via the convergence of the characteristic function. In particular, suppose that $(X_k)_{k\geq 1}$ is a sequence of i.i.d. random variables with mean m and variance $0 < \sigma^2 < \infty$. Let $S_n = \sum_{k=1}^{n} X_k$. It can be shown that

(a) the characteristic function of $\frac{S_n - mn}{\sigma\sqrt{n}}$ converges pointwise to the characteristic function of $\mathcal{N}(0,1)$ and that

(b) the characteristic function of $\frac{S_n - mn}{\sqrt{n}}$ converges pointwise to the characteristic function of $\mathcal{N}(0,\sigma)$.

To see this, let $Z_k = \frac{X_k - m}{\sigma} \overset{\text{iid}}{\sim} Z$ and $Y_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Z_k$. Then, $\mathbb{E}Z = 0$, $\operatorname{Var} Z = 1$, and $\varphi_{Y_n}(t) = \left(\varphi_Z(\frac{t}{\sqrt{n}})\right)^n$. By approximating $e^x \approx 1 + x + \frac{1}{2}x^2$. We have $\varphi_X(t) \approx 1 + jt\mathbb{E}X - \frac{1}{2}t^2\mathbb{E}\left[X^2\right]$ and

$$\varphi_{Y_n}(t) = \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n \to e^{-\frac{t^2}{2}},$$

which is the characteristic function of $\mathcal{N}(0,1)$.

- The case of Bernoulli(1/2) was derived by Abraham de Moivre around 1733. The case of Bernoulli(p) for $0 < p < 1$ was considered by Pierre-Simon Laplace [9, p. 208].