Tutorial 11: Limit Theorems

Baoxiang Wang & Yihan Zhang
bxwang, yhzhang@cse.cuhk.edu.hk
April 10, 2017

Outline

- The Central Limit Theorem (CLT)
- Normal Approximation Based on CLT
- De Moivre-Laplace Approximation to the Binomial
- Problems and solutions

Formally

- Let $S_n = X_1 + \cdots + X_n$, where the X_i are i.i.d. random variables with mean μ, variance σ^2.

- Define

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Zero mean and unit variance

- An easy calculation yields

$$\mathbf{E}[Z_n] = \frac{\mathbf{E}[X_1 + \cdots + X_n] - n\mu}{\sigma\sqrt{n}} = 0$$

- For variance, we have

$$\mathrm{var}(Z_n) = \frac{\mathrm{var}(X_1 + \cdots + X_n)}{\left(\sigma\sqrt{n}\right)^2} = \frac{n\sigma^2}{n\sigma^2} = 1$$

The Central Limit Theorem (CLT)

- *The Central Limit Theorem* The CDF of $Z_n = \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ converges to standard normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}dx$$

in the sense that

$$\lim_{n\to\infty} \mathbf{P}(Z_n \le z) = \Phi(z)$$

The distribution of the r.v. Z_n approaches a normal distribution.

Normal Approximation Based on CLT

- Given $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables with mean μ and variance σ^2. If n is large, the probability $\mathrm{P}(S_n \leq c)$ can be approximated by treating S_n as if it were normal, according to the following procedure.

I. Calculate the mean $n\mu$ and variance $n\sigma^2$.

II. Calculate $z = (c - n\mu)/\sigma\sqrt{n}$.

III. Use the approximation

$$\mathbf{P}(S_n \leq c) \approx \Phi(z).$$

De Moivre-Laplace Approximation to the Binomial

- Plugging $\mu = p, \sigma = \sqrt{p(1-p)}$, we get the following *de Moivre-Laplace Approximation to the Binomial.*
- If S_n is a binomial random variable with parameters n and p, n is large, and k, l are nonnegative integers, then

$$\mathbf{P}(k \leq S_n \leq l) \approx \Phi\left(\frac{l + 1/2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

Problem 8.

- Before starting to play the roulette in a casino, you want to look for biases that you can exploit.
- You therefore watch 100 rounds that result in a number between 1 and 36, and count the number of rounds for which the result is odd.
- If the count exceeds 55, you decide that the roulette is not fair.

Question

- Assuming that the roulette is fair, find an approximation for the probability that you will make the wrong decision.

Solution

- Let S be the number of times that the result was odd, which is a binomial random variable, with $n = 100$ and $p = 0.5$,
- so that

$$\mathbf{E}[S] = 100 \cdot 0.5 = 50$$

- and

$$\sigma_s = \sqrt{100 \cdot 0.5 \cdot 0.5} = 5$$

- Using the normal approximation to the binomial, we find

$$\mathbf{P}(S > 55) = \mathbf{P}\left(\frac{S-50}{5} > \frac{55-50}{5}\right)$$
$$= 1 - \mathbf{P}(z \leq 1)$$
$$\approx 1 - \Phi(1)$$
$$= 1 - 0.8413$$
$$= 0.1587$$

Alternative solution

- A better approximation can be obtained by using the de Moivre-Laplace approximation, which yields

$$\mathbf{P}(S > 55) = \mathbf{P}(S \geq 55.5)$$

$$= \mathbf{P}\left(\frac{S-50}{5} > \frac{55.5-50}{5}\right)$$

$$\approx \ 1 - \Phi(1.1)$$

$$= 1 - 0.8643$$

$$= 0.1357.$$

Problem 9

- During each day, the probability that your computer's operating system crashes at least once is 5%, independent of every other day.
- You are interested in the probability of at least 45 crash-free days out of the next 50 days.

Question

- (a) Find the probability of interest by using the normal approximation to the binomial.

- (b) Repeat part (a), this time using the Poisson approximation to the binomial.

Solution (a)

- Let S be the number of crash-free days, which is a binomial random variable with $n = 50$ and $p = 0.95$,
- so that

$$\mathbf{E}[S] = 50 \cdot 0.95 = 47.5$$

- And

$$\sigma_s = \sqrt{50 \cdot 0.95 \cdot 0.05} = 1.54$$

- Using the normal approximation to the binomial, we find

$$\mathbf{P}(S \geq 45) = \mathbf{P}\left(\frac{S - 47.5}{1.54} \geq \frac{45 - 47.5}{1.54}\right)$$

$$\approx 1 - \Phi(-1.62)$$

$$= \Phi(1.62)$$

$$= 0.9474.$$

Using the de Moivre-Laplace approximation, we yields

$$\mathbf{P}(S \geq 45) = \mathbf{P}(S > 44.5)$$

$$= \mathbf{P}\left(\frac{S-47.5}{1.54} \geq \frac{44.5-47.5}{1.54}\right)$$

$$= 1 - \mathbf{P}(z \leq -1.95)$$

$$\approx 1 - \Phi(-1.95)$$

$$= \Phi(1.95)$$

$$= 0.9744.$$

Solution (b)

- The random variable S is binomial with parameter $p = 0.95$.
- However, the random variable $50 - S$ (the number of crashes) is also binomial with parameter $p = 0.05$.
- Since the Poisson approximation is exact in the limit of small p and large n, it will give more accurate results if applied to $50 - S$.

- We will therefore approximate $50 - S$ by a Poisson random variable with parameter $\lambda = 50 \cdot 0.05 = 2.5$. Thus,

$$\mathbf{P}(S \geq 45) = \mathbf{P}(50 - S \leq 5)$$

$$= \sum_{k=0}^{5} \mathbf{P}(n - S = k)$$

$$= \sum_{k=0}^{5} e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= 0.958.$$

- It is instructive to compare with the exact probability which is

$$\sum_{k=0}^{5} \binom{50}{k} 0.05^k 0.95^{50-k} = 0.962$$

Interpretation

- The Poisson approximation is closer.
- This is consistent with the intuition that the normal approximation to the binomial works well when p is close to 0.5 or n is very large, which is not the case here.
- On the other hand, the calculations based on the normal approximation are generally less tedious.

Problem 11

- Let $X_1, Y_1, X_2, Y_2, \cdots$ be independent random variables, uniformly distributed in the unit interval $[0, 1]$, and let

$$W = \frac{(X_1 + \cdots + X_{16}) - (Y_1 + \cdots + Y_{16})}{16}$$

Question

- Find a numerical approximation to the quantity

$$\mathbf{P}(|W - \mathbf{E}[W]| < 0.001)$$

Solution

- Note that W is the sample mean of 16 independent identically distributed random variables of the form $X_i - Y_i$, and a normal approximation is appropriate.
- The random variables $X_i - Y_i$ have zero mean, and variance equal to $2/12$.
- Therefore, the mean of W is zero, and its variance is $(2/12)/16 = 1/96$.

- Thus,

$$\mathbf{P}(|W| < 0.001) = \mathbf{P}\left(\frac{|W|}{\sqrt{\frac{1}{96}}} < \frac{0.001}{\sqrt{\frac{1}{96}}}\right)$$

$$\approx \Phi\left(0.001\sqrt{96}\right) - \Phi\left(-0.001\sqrt{96}\right)$$

$$= 2\Phi\left(0.001\sqrt{96}\right) - 1$$

$$= 2\Phi(0.0098) - 1$$

$$\approx 2 \cdot 0.504 - 1 = 0.008$$

Alternative solution

- Let us also point out a different approach that bypasses the need for the normal table.
- Let Z be a normal random variable with zero mean and standard deviation equal to $1/\sqrt{96}$.
- The standard deviation of Z, which is about 0.1, is much larger than 0.001.
- Thus, within the interval $[-0.001, 0.001]$, the PDF of Z is approximately constant.

- $\mathbf{P}(z - \delta \leq Z \leq z + \delta) \approx f_Z(z) \cdot 2\delta,$
- $z = 0, \delta = 0.001,$

$$\mathbf{P}(|W| < 0.001) = \mathbf{P}(-0.001 \leq z \leq 0.001)$$
$$\approx f_Z(0) \cdot 0.002$$
$$= \frac{0.002}{\sqrt{2\pi}(\frac{1}{\sqrt{96}})}$$
$$= 0.0078$$