Assignment:- 4

AI1110: Probability and Random Variables
Indian Institute of Technology, Hyderabad

CS22BTECH11001
Aayush Adlakha
29 April, 2023

12.13.6.4 Suppose that 90 % of people are right handed. What is the probability that at most 6 of a random sample of 10 people are right- handed.

Solution. Let X be a Binomial random Variable.

$$X = Bin(n, p) \tag{1}$$
$$= Bin(10, 0.9) \tag{2}$$

The mean μ of X,

$$\mu = n \times p \tag{3}$$
$$= 9 \tag{4}$$

The Variance σ^2 of X,

$$\sigma^2 = n \times p \times (1 - p) \tag{5}$$
$$= 0.9 \tag{6}$$

Let,

$$Z = \frac{X - \mu}{\sigma} \tag{7}$$

Now, Z is a random variable with $\mu = 0$ and $\sigma^2 = 1$.

We can calculate the distribution of Z by assuming it be a set of discrete points on the Normal-Distribution.

Note:-The CDF of Z will converge to the normal distribution for large values of n.

[Proof on next page]

The Normal-Distribution,

$$f(x) = \frac{1}{\sqrt{2\pi}} \times e^{-\frac{x^2}{2}} \tag{8}$$

The Q-function from the Normal-Distribution

$$Q(x) = \Pr(Z > x) \tag{9}$$
$$\Pr(Z > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}}\, dt \tag{10}$$
$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}}\, dt \tag{11}$$

$$X < 6.5 \tag{12}$$
$$\therefore Z < \frac{6.5 - \mu}{\sigma} \tag{13}$$
$$Z < -2.63 \tag{14}$$

Note:- The additional 0.5 correction term is present. We want

$$\Pr(Z < 2.63) = 1 - \Pr(Z > 2.63) \tag{15}$$
$$= 1 - Q(-2.63) \tag{16}$$

On Computation,

$$\Pr(Z < 2.63) = 0.0043 \tag{17}$$

For exact answer,

$$\Pr(X = k) = {}^nC_k p^k q^{n-k} \tag{18}$$
$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} \tag{19}$$
$$\Pr(X \le k) = \sum_{t=0}^{k} {}^nC_t p^t q^{n-t} \tag{20}$$

For, $k = 6$

$$\Pr(X \le 6) = \sum_{t=0}^{6} {}^nC_t p^t q^{n-t} \tag{21}$$

On Computation,

$$\Pr(X \le 6) = 0.012 \tag{22}$$

As we can see, our approximation has an absolute error of 0.008 but a relative error of 67.14%

Conclusion:- The approximation is not very effective when n is small and p is far off from 0.5

Given that, X is a Binomial Random Variable where n is number of trials, p is probability of success and q is probability of failure.

Let μ be the mean and σ^2 be the variance. We know that,

$$\Pr(X = k) = {}^nC_k p^k q^{n-k} \tag{23}$$
$$\mu = np \tag{24}$$
$$\sigma^2 = npq \tag{25}$$

Also for large values of n, by Stirling's Approximation.

$$n! \approx n^n e^{-n} \sqrt{2\pi n} \tag{26}$$

Now,

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} \tag{27}$$
$$\approx \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k}(n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} p^k q^{n-k} \tag{28}$$
$$= \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sqrt{\frac{n}{2\pi k(n-k)}} \tag{29}$$

Let,

$$\delta = k - np \tag{30}$$
$$k = np + \delta \tag{31}$$
$$n - k = nq - \delta \tag{32}$$

Now,

$$\ln\left(\frac{np}{k}\right) = \ln\left(\frac{np}{np+\delta}\right) \tag{33}$$
$$= -\ln\left(\frac{np+\delta}{np}\right) \tag{34}$$
$$= -\ln\left(1 + \frac{\delta}{np}\right) \tag{35}$$

Similarly,

$$\ln\left(\frac{nq}{n-k}\right) = \ln\left(\frac{nq}{nq-\delta}\right) \tag{36}$$
$$= -\ln\left(\frac{nq-\delta}{nq}\right) \tag{37}$$
$$= -\ln\left(1 - \frac{\delta}{nq}\right) \tag{38}$$

Using,

$$\ln(1+x) \approx x - \frac{x^2}{2} \tag{39}$$

Now,

$$\ln\left(\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{(n-k)}\right) = -k\ln\left(1 + \frac{\delta}{np}\right) - (n-k)\ln\left(1 - \frac{\delta}{nq}\right) \tag{40}$$
$$= -(\delta + np)\left(\frac{\delta}{np} - \frac{\delta^2}{2n^2p^2}\right) - (nq - \delta)\left(-\frac{\delta}{nq} - \frac{\delta^2}{2n^2q^2}\right) \tag{41}$$
$$\approx -\delta\left[1 + \frac{\delta}{2np} - 1 + \frac{\delta}{2nq}\right] \tag{42}$$
$$= -\frac{\delta^2}{2npq} \tag{43}$$
$$\therefore \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{(n-k)} = e^{-\frac{\delta^2}{2npq}} \tag{44}$$

Moreover,

$$\sqrt{\frac{n}{2\pi k(n-k)}} = \sqrt{\frac{n}{2\pi(np+\delta)(nq-\delta)}} \tag{45}$$
$$\approx \sqrt{\frac{1}{2\pi npq}} \tag{46}$$

This holds only when k differs from the mean by a few standard deviations.

Now,

$$\Pr(X = k) \approx \sqrt{\frac{1}{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \tag{47}$$
$$\Pr(X = k) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}} \tag{48}$$

Now,

$$\Pr(a \le X \le b) = \sum_{t=a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}} \tag{49}$$

Modeling the terms of the summation as area of rectangles of height 1, in the region $(t - 0.5, t + 0.5)$

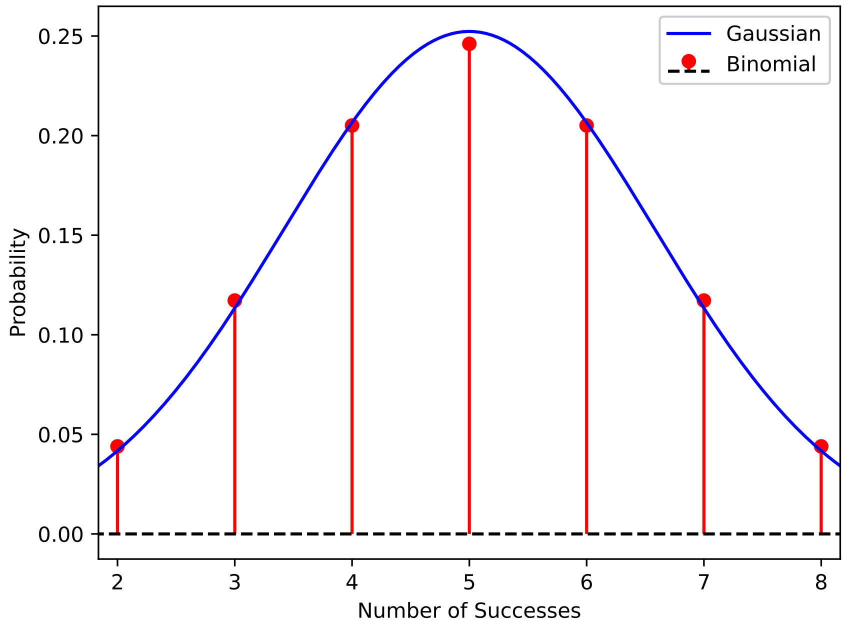
We can further approximate the Sum to be the integral,

$$\Pr(a \le X \le b) = \int_{a-0.5}^{b+0.5} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}}\, dk \tag{50}$$
$$\Pr(a' \le Z \le b') = \int_{a'-0.5}^{b'+0.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz \tag{51}$$

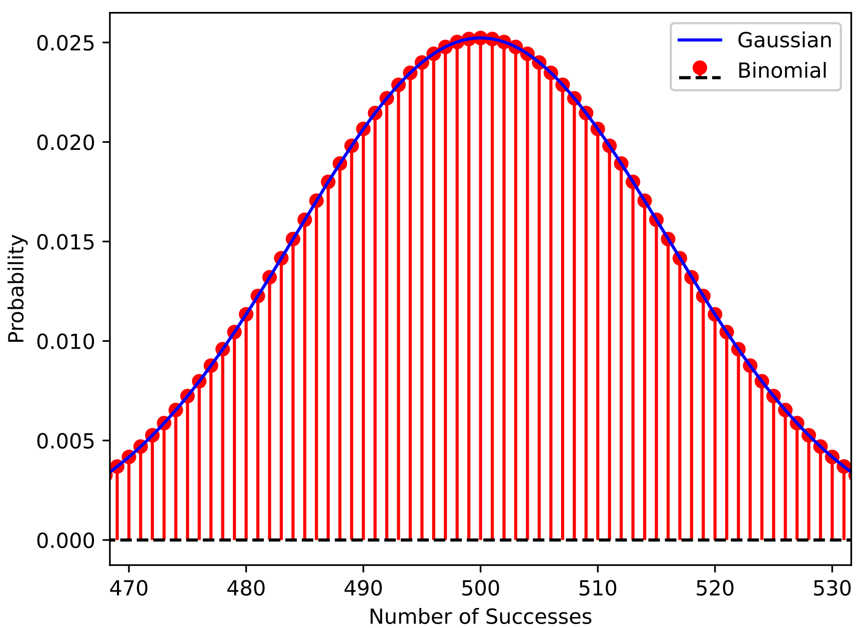
Where,

$$Z = \frac{X - \mu}{\sigma} \tag{52}$$

And a' and b' are corresponding a and b



(a) 10 trials



(b) 1000 trails

Fig. 0: A comparison of Binomial and Gaussian