

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.

1) **Gaussian**

Let X be a Binomial random Variable.

$$X = \text{Bin}(n, p) \quad (1)$$

$$= \text{Bin}(10, 0.9) \quad (2)$$

The mean μ of X ,

$$\mu = n \times p \quad (3)$$

$$= 9 \quad (4)$$

The Variance σ^2 of X ,

$$\sigma^2 = n \times p \times (1 - p) \quad (5)$$

$$= 0.9 \quad (6)$$

Let,

$$Z = \frac{X - \mu}{\sigma} \quad (7)$$

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,

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

The Q-function from the Normal-Distribution

$$Q(x) = \Pr(Z > x) \quad (9)$$

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

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}} dt \quad (11)$$

$$X < 6.5 \quad (12)$$

$$\therefore Z < \frac{6.5 - \mu}{\sigma} \quad (13)$$

$$Z < -2.63 \quad (14)$$

Note:- The additional 0.5 correction term is present.

From (11)

$$\Pr(Z < -2.63) = 1 - \Pr(Z > -2.63) \quad (15)$$

$$= 1 - Q(-2.63) \quad (16)$$

On Computation,

$$\Pr(Z < -2.63) = 0.0043 \quad (17)$$

2) **Binomial**

$$\Pr(X = k) = {}^nC_k p^k q^{n-k} \quad (18)$$

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} \quad (19)$$

$$\Pr(X \leq k) = \sum_{t=0}^k {}^nC_t p^t q^{n-t} \quad (20)$$

For, $k = 6$

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

From (21)

$$\Pr(X \leq 6) = 0.012 \quad (22)$$

3) Conclusion

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

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

4) Proof

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,

$$\Pr(X = k) = {}^nC_k p^k q^{n-k} \quad (23)$$

$$\mu = np \quad (24)$$

$$\sigma^2 = npq \quad (25)$$

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

$$n! \approx n^n e^{-n} \sqrt{2\pi n} \quad (26)$$

Now,

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k} \quad (27)$$

$$\approx \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k} (n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} p^k q^{n-k} \quad (28)$$

$$= \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sqrt{\frac{n}{2\pi k(n-k)}} \quad (29)$$

Let,

$$\delta = k - np \quad (30)$$

$$k = np + \delta \quad (31)$$

$$n - k = nq - \delta \quad (32)$$

Now,

$$\ln\left(\frac{np}{k}\right) = \ln\left(\frac{np}{np + \delta}\right) \quad (33)$$

$$= -\ln\left(\frac{np + \delta}{np}\right) \quad (34)$$

$$= -\ln\left(1 + \frac{\delta}{np}\right) \quad (35)$$

Similarly,

$$\ln\left(\frac{nq}{n-k}\right) = \ln\left(\frac{nq}{nq - \delta}\right) \quad (36)$$

$$= -\ln\left(\frac{nq - \delta}{nq}\right) \quad (37)$$

$$= -\ln\left(1 - \frac{\delta}{nq}\right) \quad (38)$$

Using,

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

Now,

$$\ln\left(\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{(n-k)}\right) = -k \ln\left(1 + \frac{\delta}{np}\right) - (n-k) \ln\left(1 - \frac{\delta}{nq}\right) \quad (40)$$

$$= -(\delta + np) \left(\frac{\delta}{np} - \frac{\delta^2}{2n^2 p^2}\right) - (nq - \delta) \left(-\frac{\delta}{nq} - \frac{\delta^2}{2n^2 q^2}\right) \quad (41)$$

$$\approx -\delta \left[1 + \frac{\delta}{2np} - 1 + \frac{\delta}{2nq}\right] \quad (42)$$

$$= -\frac{\delta^2}{2npq} \quad (43)$$

$$\therefore \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{(n-k)} = e^{-\frac{\delta^2}{2npq}} \quad (44)$$

Moreover,

$$\sqrt{\frac{n}{2\pi k(n-k)}} = \sqrt{\frac{n}{2\pi(np + \delta)(nq - \delta)}} \quad (45)$$

$$\approx \sqrt{\frac{1}{2\pi npq}} \quad (46)$$

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

Now,

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

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

Now,

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

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,

$$\Pr(a \leq X \leq b) = \int_{a-0.5}^{b+0.5} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}} dk \quad (50)$$

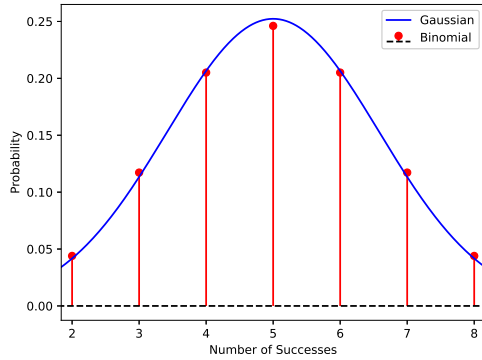
$$\Pr(a' \leq Z \leq b') = \int_{a'-0.5}^{b'+0.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (51)$$

Where,

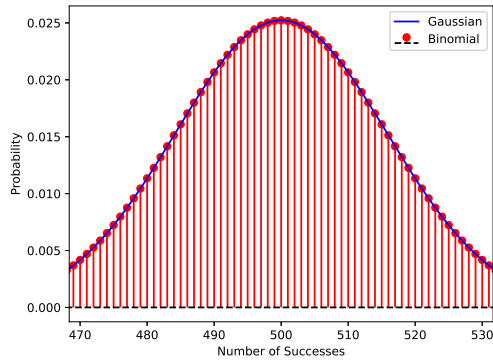
$$Z = \frac{X - \mu}{\sigma} \quad (52)$$

And a' and b' are corresponding a and b

5) Graphs



(a) 10 trials



(b) 1000 trials

Fig. 0: A comparison of Binomial and Gaussian