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 CDF from the Normal-Distribution

$$F_Z(x) = \int_{-\infty}^{x+0.5} \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}}\, dt \tag{9}$$

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

$$X \leq 6 \tag{10}$$
$$\therefore Z \leq \frac{6 - \mu}{\sigma} \tag{11}$$
$$Z \leq -3.16 \tag{12}$$

$$F_Z(-3.16) = \int_{-\infty}^{-2.66} \frac{1}{\sqrt{2\pi}} \times e^{-\frac{t^2}{2}}\, dt \tag{13}$$

On Computation,

$$F_Z(-3.16) = 0.0042 \tag{14}$$

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,

$$\Pr(X = x) = {}^{n}C_{x} p^{x} q^{n-x} \tag{15}$$

$$\mu = np \tag{16}$$

$$\sigma^2 = npq \tag{17}$$

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

$$n! \approx n^n e^{-n} \sqrt{2\pi n} \tag{18}$$

Now,

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

$$\approx \frac{n^n e^{-n} \sqrt{2\pi n}}{x^x e^{-x} \sqrt{2\pi x}(n-x)^{n-x} e^{-(n-x)} \sqrt{2\pi(n-x)}} p^x q^{n-x} \tag{20}$$

$$= \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{n-x} \sqrt{\frac{n}{2\pi x(n-x)}} \tag{21}$$

Let,

$$\delta = x - np \tag{22}$$

$$x = np + \delta \tag{23}$$

$$n - x = nq - \delta \tag{24}$$

Now,

$$\ln\left(\frac{np}{x}\right) = \ln\left(\frac{np}{np+\delta}\right) \tag{25}$$

$$= -\ln\left(\frac{np+\delta}{np}\right) \tag{26}$$

$$= -\ln\left(1 + \frac{\delta}{np}\right) \tag{27}$$

Similarly,

$$\ln\left(\frac{nq}{n-x}\right) = \ln\left(\frac{nq}{nq-\delta}\right) \tag{28}$$

$$= -\ln\left(\frac{nq-\delta}{nq}\right) \tag{29}$$

$$= -\ln\left(1 - \frac{\delta}{nq}\right) \tag{30}$$

Using,

$$\ln(1+x) \approx x - \frac{x^2}{2} \tag{31}$$

Now,

$$\ln\left(\left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{(n-x)}\right) = -x\ln\left(1 + \frac{\delta}{np}\right) - (n-x)\ln\left(1 - \frac{\delta}{nq}\right) \tag{32}$$

$$= -(\delta + np)\left(\frac{\delta}{np} - \frac{\delta^2}{2n^2p^2}\right) - (nq - \delta)\left(-\frac{\delta}{nq} - \frac{\delta^2}{2n^2q^2}\right) \tag{33}$$

$$\approx -\delta\left[1 + \frac{\delta}{2np} - 1 + \frac{\delta}{2nq}\right] \tag{34}$$

$$= -\frac{\delta^2}{2npq} \tag{35}$$

$$\therefore \left(\frac{np}{x}\right)^x \left(\frac{nq}{n-x}\right)^{(n-x)} = e^{-\frac{\delta^2}{2npq}} \tag{36}$$

Moreover,

$$\sqrt{\frac{n}{2\pi x(n-x)}} = \sqrt{\frac{n}{2\pi(np+\delta)(nq-\delta)}} \tag{37}$$

$$\approx \sqrt{\frac{1}{2\pi npq}} \tag{38}$$

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

Now,

$$\Pr(X = x) \approx \sqrt{\frac{1}{2\pi npq}} e^{-\frac{(x-np)^2}{2npq}} \tag{39}$$

This is the Normal-Distribution for data with $\mu = np$ and $\sigma^2 = npq$

For data with $\mu = 0$ and $\sigma^2 = 1$, that is $Z = \frac{X-\mu}{\sigma}$

$$\Pr(Z = z) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{40}$$

Now,

$$\Pr(a \le Z \le b) = \sum_{t=a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{41}$$

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,

$$\Pr(a \le Z \le b) = \int_{a-0.5}^{b+0.5} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt \tag{42}$$