Assignment Question 2

AI1110: Probability and Random Variables

Rishitha Surineni
CS22BTECH11050

Question:12.13.5.13
It is known that 10% of certain articles manufactured are defective.What is the probability that in a random sample space of 12 such articles, 9 are defective?
Solution:

(I) **Calculation of Probability Using Gaussian Distribution approach**
Let X_i be a random variable corresponding to i^{th} article such that

$$X_i = \begin{cases} 1 & \text{if the article is defective} \\ 0 & \text{if the article is not defective} \end{cases} \tag{1}$$

$X_1, X_2, ..., X_{12}$ is a sequence of independent and identically distributed random variables.This sequence forms a Binomial Distribution with mean μ and variance σ^2
For this Binomial Distribution, n = 12 and p = 0.1.
The mean and standard deviation of the Binomial distribution are

$$\mu = np \tag{2}$$
$$\mu = 12 \times 0.1$$
$$\mu = 1.2 \tag{3}$$
$$\sigma = \sqrt{np(1-p)} \tag{4}$$
$$\sigma = \sqrt{12 \times 0.1 \times 0.9}$$
$$\sigma = 1.04 \tag{5}$$

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

$$\text{Standardized sample mean} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
$$= \frac{S_n - \mu n}{\sigma\sqrt{n}} \tag{6}$$

According to Central Limit Theorem
as $n \to \infty$,$\frac{S_n - \mu n}{\sigma\sqrt{n}} \to N(0,1)$
here N(0,1) denotes a standard normal distribution with mean 0 and variance 1.

(II) **Proof of Central Limit Theorem**
Let $X_1, X_2, .., X_n$ be independent and identically distributed Random Variables with mean μ and variance σ^2.
Let

$$S_n = \sum_{i=1}^{n} X_i \tag{7}$$
$$Z_i = \frac{X_i - \mu}{\sigma} \tag{8}$$
$$E(Z_i) = E\left(\frac{X_i - \mu}{\sigma}\right) \tag{9}$$
$$= \frac{1}{\sigma} E(X_i - \mu)$$
$$= \frac{1}{\sigma}(E(X_i) - E(\mu))$$
$$= \frac{1}{\sigma}(\mu - \mu)$$
$$E(Z_i) = 0 \tag{10}$$
$$Var(Z_i) = Var\left(\frac{X_i - \mu}{\sigma}\right) \tag{11}$$
$$= \frac{1}{\sigma^2} Var(X_i - \mu)$$
$$= \frac{1}{\sigma^2} Var(X_i)$$
$$= \frac{1}{\sigma^2}\sigma^2$$
$$Var(Z_i) = 1 \tag{12}$$

Let M_{Z_i} be Moment Generating Function of Z_i.

$$M_{Z_i}(t) = E\left(e^{tZ_i}\right) \tag{13}$$
$$M_{Z_i}(0) = E(1) = 1 \tag{14}$$
$$E(Z_i) = M'_{Z_i}(0) \tag{15}$$

$$M'_{Z_i}(0) = 0 \text{ From eq(10)} \tag{16}$$

$$Var(Z_i) = E\left(Z_i^2\right) - E(Z_i)^2 \tag{17}$$

$$\begin{aligned} E\left(Z_i^2\right) &= Var(Z_i) + E(Z_i)^2 \\ &= 1 - 0 = 1 \end{aligned}$$

$$M''_{Z_i}(0) = E\left(Z_i^2\right) = 1 \tag{18}$$

Consider $Y \sim N(0,1)$, Then

$$M_Y(t) = e^{\frac{t^2}{2}} \tag{19}$$

Let $T_n = \sum_{i=1}^{n} Z_i$

$$\begin{aligned} M_{\frac{T_n}{\sqrt{n}}}(t) &= E\left(e^{t\frac{T_n}{\sqrt{n}}}\right) && (20) \\ &= E\left(e^{\frac{t}{\sqrt{n}}\sum_{i=1}^{n} Z_i}\right) \\ &= E\left(e^{\frac{t}{\sqrt{n}}Z_1}.e^{\frac{t}{\sqrt{n}}Z_2}...e^{\frac{t}{\sqrt{n}}Z_n}\right) \\ &= E\left(e^{\frac{t}{\sqrt{n}}Z_1}\right).E\left(e^{\frac{t}{\sqrt{n}}Z_2}\right)...E\left(e^{\frac{t}{\sqrt{n}}Z_n}\right) \\ &= M_{Z_1}\left(\frac{t}{\sqrt{n}}\right).M_{Z_2}\left(\frac{t}{\sqrt{n}}\right)...M_{Z_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left[M_{Z_i}\left(\frac{t}{\sqrt{n}}\right)\right]^n && (21) \end{aligned}$$

As $n \to \infty$, $\lim_{n\to\infty} M_{\frac{T_n}{\sqrt{n}}}(t)$ takes the form 1^∞
Applying natural logarithm to the eq(21)

$$\ln\left(M_{\frac{T_n}{\sqrt{n}}}(t)\right) = n \ln\left(M_{Z_i}\left(\frac{t}{\sqrt{n}}\right)\right)$$

$$\lim_{n\to\infty} \ln\left(M_{\frac{T_n}{\sqrt{n}}}(t)\right) = \lim_{n\to\infty} n \ln\left(M_{Z_i}\left(\frac{t}{\sqrt{n}}\right)\right) \tag{22}$$

Let $u = \frac{1}{\sqrt{n}}$
Then as $n \to \infty$, $u \to 0$

Substituting in eq(22) we get

$$\begin{aligned} \lim_{n\to\infty} \ln\left(M_{\frac{T_n}{\sqrt{n}}}(t)\right) &= \lim_{u\to 0} \frac{1}{u^2} \ln\left(M_{Z_i}(tu)\right) \\ &= \lim_{u\to 0} \frac{1}{2u} \frac{tM'_{Z_i}(tu)}{M_{Z_i}(tu)} \\ &= \frac{t}{2} \frac{1}{M_{Z_i}(0)} \lim_{u\to 0} \frac{M'_{Z_i}(tu)}{u} \end{aligned}$$

[From eq(16)]

$$\begin{aligned} &= \frac{t}{2} \lim_{u\to 0} \frac{tM''_{Z_i}(tu)}{1} \\ &= \frac{t^2}{2} M''_{Z_i}(0) \end{aligned}$$

[From eq(18)]

$$M_{\frac{T_n}{\sqrt{n}}}(t) = e^{\frac{t^2}{2}} \tag{23}$$

Therefore

$$M_{\frac{T_n}{\sqrt{n}}}(t) = M_Y(t) \tag{24}$$

According to Uniqueness Theorem of Moment Generating Function,
If two random variables have the same MGF then they have the same distribution.
Therefore,
as $n \to \infty$ $\frac{T_n}{\sqrt{n}}$ and Y have the same distribution.
From eq(7) and eq(8)

$$\frac{T_n}{\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \tag{25}$$

Therefore
as $n \to \infty$,$\frac{S_n - \mu n}{\sigma\sqrt{n}} \to N(0,1)$
Hence Proved.

(III) **Continuation of the Gaussian Distribution approach**
Let, E be the event that exactly 9 articles are defective.
For the event E

$$S_n = 9 \text{[as exactly 9 articles are defective]}$$

Z-score for a Gaussian distribution is given as

$$Z = \frac{X - \mu}{\sigma} \tag{26}$$

where x is the observed value We can calculate $\Pr(E)$ by using Q-function.
Q-function $Q(x)$is defined as the complement of the Cumulative Distribution Function(CDF) of the standard normal distribution up to the

point x. Let Z_1 and Z_2 are the z-scores for X_1 and X_2 respectively

To calculate the probability of having exactly 9 defective articles, we can calculate the probability that the number of defective articles is greater than 8.5 and subtract the probability that it is greater than 9.5.

Since the Gaussian distribution is continuous, we consider the probability of the number of defective articles falling within the range of 8.5 to 9.5 (inclusive) as the approximation for having exactly 9 defective articles.

$$X_1 = 9.5$$
$$X_2 = 8.5$$

As S_n=9

$$Z_1 = \frac{x_1 - np}{\sqrt{np(1-p)}}$$
$$Z_1 = \frac{9.5 - 1.2}{1.04}$$
$$Z_1 = 7.98 \tag{27}$$
$$Z_2 = \frac{x_2 - np}{\sqrt{np(1-p)}}$$
$$Z_2 = \frac{8.5 - 1.2}{1.04}$$
$$Z_2 = 7.02 \tag{28}$$
$$\Pr(E) = Q(Z_2) - Q(Z_1) \tag{29}$$
$$\Pr(E) = Q(7.02) - Q(7.98) \tag{30}$$

Q-function can be calculated from the error function $[erf(Z)]$, where Z is the Z- score.

$$Q(Z) = 1 - \text{erf}(Z) \tag{31}$$
$$\text{erf}(Z) = \frac{2}{\sqrt{\pi}} \int_0^Z e^{-t^2}\, dt \tag{32}$$

Therefore (33)

$$Q(Z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^Z e^{-t^2}\, dt \tag{34}$$

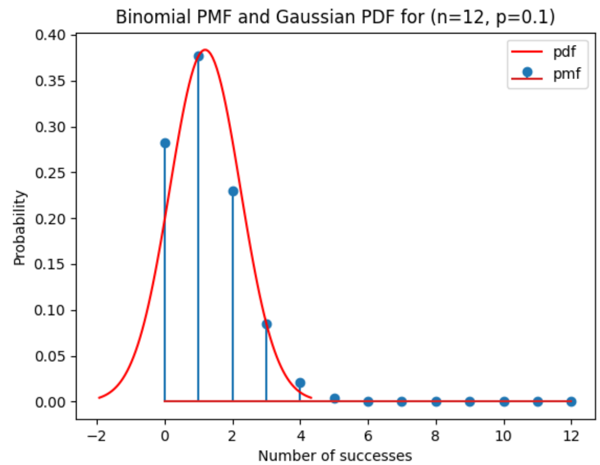
$Q(x)$ can be calculated by using python code. Use the link below to find the code to calculate the value of Q-function.
https://github.com/Rishitha745/Bonus_2/blob/main/Codes/Q_func.py

Based on the output of the above code we get the value of Q-functions as

$$Q(7.98) = 7.316664535724922 \times 10^{-16} \tag{35}$$
$$Q(7.02) = 1.1093412281591434 \times 10^{-12} \tag{36}$$

From eq(30)

$$\begin{aligned}\Pr(E) &= 1.1093412281591434 \times 10^{-12} \\ &\quad - 7.316664535724922 \times 10^{-16} \\ \Pr(E) &= 1.108609561705571 \times 10^{-12}\end{aligned} \tag{37}$$



(IV) **Calculation of Probability using Binomial Distribution approach**

If we consider the distribution as a Binomial Distribution then the $\Pr(E)$ can be calculated as

$$\Pr(E) = {}^{n}C_r p^r (1-p)^{n-r} \tag{38}$$
$$r = 9 \tag{39}$$

As exactly 9 articles are defective

$$\Pr(E) = {}^{12}C_9 (0.1)^9\, 0.9^3 \tag{40}$$
$$\Pr(E) = 1.6038 \times 10^{-07} \tag{41}$$

(V) **Comparision of Answers obtained from Binomial Distribution and Gaussian Distribution**

Binomial Distribution converges to Gaussian Distribution when the size of sample space is large.

If the following condition is satisfied then the Binomial distribution converges to Gaussian Distribution

$$np(1-p) \geqslant 10 \tag{42}$$

From Given

$$n = 12$$
$$p = 0.1$$

Therefore LHS of the above condition becomes

$$np\,(1-p) = 12 \times 0.1 \times 0.9 = 1.08$$
$$1.08 < 10$$

Therefore the above condition is not satisfied and there would be significant difference in the probabilities calculated by using Binomial Distribution approach and Gaussian Distribution approach.(here, as the size of sample space is low the Binomial approach gives more accurate answer and the probability from Gaussian Distribution approach is significantly lower than the actual probability)
Let ’N’ be the size of sample space for the Binomial Distribution to converge to Gaussian Distribution, then

$$\begin{aligned} N\,(0.1)\,(0.9) &\geqslant 10 \\ N &\geqslant \frac{10}{0.09} \\ N &\geqslant 111.11 \end{aligned} \tag{43}$$

Therefore the size of Sample Space should be atleast 112.