PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit - 3 - Probability Distributions

QB SOLVED

Central Limit Theorem

Exercises for Section 4.11

1. A 500-page book contains 250 sheets of paper. The thickness of the paper used to manufacture the book has mean 0.08 mm and standard deviation 0.01 mm.
   a) What is the probability that a randomly chosen book is more than 20.2 mm thick (not including the covers)?
   b) What is the 10th percentile of book thicknesses?
   c) Someone wants to know the probability that a randomly chosen page is more than 0.1 mm thick. Is enough information given to compute this probability? If so, compute the probability. If not, explain why not.

[Text Book Exercise – Section 4.11 – Q. No. 2 – Pg. No. 300]
Solution:

a) What is the probability that a randomly chosen book is more than 20.2 mm thick (not including the covers)?

Let $X_1 \ldots X_{250}$ denotes thickness of the 250 sheets of paper.

Let $S = X_1 + \cdots + X_{250}$ be the total thickness of the 250 sheets of paper.

Given that $n = 250, \mu = 0.08, \sigma = 0.01.$

So, S is approximately normally distributed. From the Central Limit theorem we get $S_n \sim N\,(n\mu, n\sigma^2)$

$$\mu_S = (250)(0.08) = 20$$

$$\sigma_{S_n}^2 = n\sigma^2 = 250(0.01)^2 = 0.025$$

$$S_n \sim N\ (20, 0.025)$$

The z-score corresponding to 20.02 is,

$$z = \frac{20.2 - 20}{\sqrt{0.025}} = 1.26$$

The area to the right of z = 1.26 is (1 – 0.8962) =0.1038.

$$P(S > 20.2) = 0.1038$$

b) What is the 10th percentile of book thicknesses?

Let S_{10} be the 10th percentile.

The z-score for 10th percentile is approximately z = -1.28

Therefore,

$$-1.28 = \frac{(S_{10} - 20)}{\sqrt{0.025}}$$

$$S_{10} = 20 + (-1.28)\sqrt{0.025} = 19.797$$

c) Someone wants to know the probability that a randomly chosen page is more than 0.1 mm thick. Is enough information given to compute this probability? If so, compute the probability. If not, explain why not.

No, because we don't know the distribution of page thicknesses. We only have the mean and standard deviation of it. In particular, we do not know whether the page thicknesses are normally distributed. So, there is no enough information to compute the required probability.

2. The concentration of particles in a suspension is 30 per mL.
   a) What is the probability that a 2 mL sample will contain more than 50 particles?
   b) Ten 2 mL samples are drawn. What is the probability that at least 9 of them contain more than 50 particles?
   c) One hundred 2 mL samples are drawn. What is the probability that at least 90 of them contain more than 50 particles?

[Text Book Exercise – Section 4.11 – Q. No. 14 – Pg. No. 301]
Solution:

a) What is the probability that a 2 mL sample will contain more than 50 particles?

Let X be the number of particles contained in a 2 ml sample. Then,

$X \sim Poisson\ (60)$.

So, X is approximately normal with mean $\mu_X = 60$ and standard deviation $\sigma_X = \sqrt{60} = 7.7459$.

Clearly, $X \sim Poisson\ (60)$. We can use Normal Approximation to the Poisson.

Thus $X \sim N(\mu = 60, \sigma^2 = 60)$.

To find $P\ (X > 50)$.

$$z = \frac{50 - 60}{7.7459} = -1.29$$

The area to the right of z = -1.29 is (1 – 0.0985) = 0.9015.

$P\ (X > 50) = 0.9015$.

b) Ten 2 mL samples are drawn. What is the probability that at least 9 of them contain more than 50 particles?

Let Y be the number of samples that contain more than 50 particles. From part (a), the probability that a sample contains more than 50 particles is 0.9015.

Therefore, $Y \sim Bin(10{,}0.9015)$.

$$P\ (Y \geq 9) = P(Y = 9) + P(Y = 10)$$

$$= \frac{10!}{9!\,(10-9)!}\,(0.9015)^9\,(1 - 0.9015)^{10-9} - \frac{10!}{10!\,(10-10)!}\,(0.9015)^{10}\,(1 - 0.9015)^{10-10}$$

$$= 0.7419$$

c) One hundred 2 mL samples are drawn. What is the probability that at least 90 of them contain more than 50 particles?

Here we draw 100 2ml samples.

From part (a), the probability that a sample contains more than 50 particles is 0.9015.

Therefore, $X \sim Bin\ (100, 0.9015)$.

So, X is approximately normal with mean $\mu_X = 100(0.9015) = 90.15$ and standard deviation $\sqrt{100(0.9015)(0.0985)} = 2.9798$.

To find, $P\ (X > 90)$, we use continuity correction and find the z-score of 89.5.

$$z = \frac{89.5 - 90.15}{2.9798} = -0.22$$

The area to the right of $z = -0.22$ is $(1-0.4129) = 0.5871$.

$P\ (X > 90) = 0.5871$