

Statistical Methods for Data Analytics

Assignment 1

1. Probability calculation: As of Jan 18, 332,210,522 people in the world got COVID, and 5,494,965 died from it. The world population is 7,794,795,127.
 - a) What is the probability of getting COVID?
 $332,210,522 / 7,794,795,127 = 4.26\%$
 - b) What is the probability of dying from COVID?
 $5,494,965 / 7,794,795,127 = 0.07\%$
 - c) What is the probability of dying conditional that one has COVID?
 $5,494,965 / 332,210,522 = 1.65\%$
 - d) See the chart on the infection fatality rate by age group in slide 6 in Topic 1. What is the probability of dying conditional that you have COVID and you're a 21-year-old man? Give a rough number based on the chart.

Infection fatality rate = 0.0055% approx.

4. MLE for normal distribution

- b) Given this synthetic data in (a), estimate μ and σ using the MLE. Use `optim()` or `nlm()` function. Check whether the estimated parameters are close to the actual ones.

Ans: - The actual ones are $\mu = -0.1611651$ $\sigma = 1.5765628$ whereas the estimated ones are $\mu = -0.1612627$ $\sigma = 1.5606212$, they have minimal difference.

- c) Repeat this exercise with 5000 generated numbers instead of 50 numbers with `set.seed(200)`. What difference can you tell?

Ans: - The actual ones are $\mu = 0.01837722$ $\sigma = 1.96820547$ whereas the estimated ones are $\mu = 0.01850456$ $\sigma = 1.96821879$, they have minimal difference.

By increasing the generated numbers, the values of actual and estimated ones increase significantly.