Covid 19

The present assignment requires you to analyse Coronavirus disease 2019 (COVID-19) data from various data sources including data from India and around the world. You will see yourself unearthing very interesting information about COVID-19 "hidden" in the data.

Problem 1. The data file *Covid19IndiaData 30032020.xlsx* presents the Indian patientlevel data until 30th March 2020. Source for latest Indian COVID-19 data: https://api. rootnet.in/.

(i) Calculate and plot the probability mass function (pmf) of the age of infected patients — this includes *Hospitalized*, *Recovered* and *Dead*. Evaluate the expected age of an infected patient from this pmf and the variance of the pmf. Does the variance seem *high* or *low*? What can you say about the expectation calculated in light of the variance of the pmf?

(ii) Calculate and plot the pmfs of the age of *Recovered* and *Dead* patients. Calculate the expectation and variance of the pmfs. Are the expectations the same as in Case (i)? What can you say about COVID-19 by comparing the expectation values?

(iii) Find the conditional pmf of the age of all infected patients conditional to the *gender* of the patient. Are the pmfs identically distributed? Compare the expectations and comment of the possible reason(s) for any difference.

Problem 2. The data file *linton supp tableS1 S2 8Feb2020.xlsx* presents patient-level case data from China and other parts of the world. This includes the following information — *Exposure date (E)*, *Symptoms onset date (O)*, *Hospitalisation date (H)*, and *Date of death (X)* in case of deceased patients. The data also includes whether the patient is/was a resident of Wuhan (China). Please note that details of surviving and deceased patients (until 31st January 2020) are included as different Sheets. The data can be downloaded from http://www.mdpi.com/2077-0383/9/2/538/s1 and may require some *cleaning* before being used. You can compare your results to the results published in the Journal of Clinical Magazine in February 2020 (https://www.mdpi.com/2077-0383/9/2/538/pdf).

(i) Calculate and plot the pmf of the incubation period, which is defined as the duration between the date of infection exposure (E) and the date of onset of symptoms (O). Calculate the mean incubation period and the variance of the distribution. Please use the left exposure data as the date of infection.

(ii) Now calculate the expected incubation period by excluding Wuhan residents and compare the values with part (i). What can you comment based on the comparison?

(iii) Calculate the pmfs of the onset to hospitalization ($H-O$) for dead patients, onset to death ($X-O$) and hospitalization to death ($X-H$). Do you see a similarity in the distribution? Comment. Also, compare the $H-O$ pmf for surviving and dead patients; comment on the difference.