Homework #01

SMDS-2023-2024

Statstical Methods in Data Science II a.y. 2022-2023

M.Sc. in Data Science

deadline: April 26th, 2024

A. Simulation

1. Consider the following joint discrete distribution of a random vector (Y, Z) taking values over the bi-variate space:

$$\begin{aligned}\mathcal{S} = \mathcal{Y} \times \mathcal{Z} \quad = \quad & \{(1,1);(1,2);(1,3); \\ & (2,1);(2,2);(2,3); \\ & (3,1);(3,2);(3,3)\}\end{aligned}$$

The joint probability distribution is provided as a matrix `J` whose generic entry `J[y,z]` $= Pr\{Y = y, Z = z\}$

```
J
```

```
     1    2    3
1 0.06 0.17 0.10
2 0.10 0.12 0.11
3 0.14 0.02 0.18
```

```
S
```

```
      row col
(1,1)   1   1
(1,2)   1   2
(1,3)   1   3
(2,1)   2   1
(2,2)   2   2
(2,3)   2   3
(3,1)   3   1
(3,2)   3   2
(3,3)   3   3
```

You can load the matrix `S` of all the couples of the states in $\mathcal{S}$ and the matrix `J` containing the corresponding bivariate probability masses from the file "Hmwk.RData". How can you check that J is a probability distribution?

2. How many *conditional distributions* can be derived from the joint distribution `J`? Please list and derive them.

Answer:

3. Make sure they are probability distributions.

4. Can you simulate from this `J` distribution? Please write down a working procedure with few lines of R code as an example. Can you conceive an alternative approach? In case write down an alternative working procedure with few lines of R

B. Bulb lifetime: a conjugate Bayesian analysis of exponential data

You work for Light Bulbs International. You have developed an innovative bulb, and you are interested in characterizing it statistically. You test 20 innovative bulbs to determine their lifetimes, and you observe the following data (in hours), which have been sorted from smallest to largest.

1, 13, 27, 43, 73, 75, 154, 196, 220, 297,
344, 610, 734, 783, 796, 845, 859, 992, 1066, 1471

Based on your experience with light bulbs, you believe that their lifetimes Y_i can be modeled using an exponential distribution conditionally on θ where $\psi = 1/\theta$ is the average bulb lifetime.

1. Write the main ingredients of the Bayesian model.

2. Choose a conjugate prior distribution $\pi(\theta)$ with mean equal to 0.003 and standard deviation 0.00173.

3. Argue why with this choice you are providing only a vague prior opinion on the average lifetime of the bulb.

4. Show that this setup fits into the framework of the conjugate Bayesian analysis.

5. Based on the information gathered on the 20 bulbs, what can you say about the main characteristics of the lifetime of your innovative bulb? Argue that we have learnt some relevant information about the θ parameter and this can be converted into relevant information about the unknown average lifetime of the innovative bulb $\psi = 1/\theta$.

6. However, your boss would be interested in the probability that the average bulb lifetime $1/\theta$ exceeds 550 hours. What can you say about that after observing the data? Provide her with a meaningful Bayesian answer.

C. Exchangeability

Let us consider an infinitely exchangeable sequence of binary random variables $X_1, ..., X_n, ...$

1. Provide the definition of the distributional properties characterizing an infinitely echangeable binary sequence of random variables $X_1, ..., X_n, ...$. Consider the De Finetti representation theorem relying on a suitable distribution $\pi(\theta)$ on $[0, 1]$ and show that

$$\begin{aligned} E[X_i] &= E_\pi[\theta] \\ E[X_i X_j] &= E_\pi[\theta^2] \\ Cov[X_i, X_j] &= Var_\pi[\theta] \end{aligned}$$

2. Prove that any couple of random variabes in that sequence must be non-negatively correlated.

3. Find what are the conditions on the distribution $\pi(\cdot)$ so that $Cor[X_i, X_j] = 1$.

4. What do these conditions imply on the type and shape of $\pi(\cdot)$? (make an example).

Last update by LT: Fri Apr 12 11:09:28 2024