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:

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

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?

---

1. Answer:

We can check whether J is a *valid probability distribution* by verifying if it satisfies the two properties of *non-negativity* and *summation to 1*. Since J is discrete we will refer to the discrete versions of the formulas.

The first states that all elements of the probability distribution must be non-negative:

$Pr\{Y = y, Z = z\} \geq 0 \; \forall y, z \in \{1, 2, 3\}$

We can write a simple conditional statement to check if this condition is satisfied:

```
# function to verify is a probability distribution is non-negative
verify_non_negativity <- function(A){
  return(sum(A >= 0) == length(A))
}

verify_non_negativity(J)
```

```
[1] TRUE
```

The property of summation to 1 instead states that the elements of the probability distribution must sum up to 1:

$\sum_{y,z \in \{1,2,3\}} Pr\{Y = y, Z = z\} = 1$

This also can be checked with a simple conditional statement:

```
# function to verify is a probability distribution has sum equal to 1
verify_sum_to_one <- function(A){
  return(all.equal(sum(A), 1))
}
```

It's important to notice that, to account for *machine precision*, we used the method `all.equal` instead of the `==` operator.

```
verify_sum_to_one(J)
```

```
[1] TRUE
```

Also the second condition is satisfied, therefore we can say that J is indeed a valid probability distribution.

---

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

---

2. **Answer:**

We can derive 6 conditional probability distributions by fixing on either Y or Z a value in ${1,2,3}. To derive them we have just to implement the definition of conditional probability:

$Pr(Y = y|Z = z) = \frac{Pr(Y=y,Z=z)}{Pr(Z=z)}$

Where at the numerator we have the joint distribution provided by the matrix J and at the denominator the marginal distribution of the value we condition on, defined as:

$Pr(Z = z) = \sum_y Pr(Y = y, Z = z)$

Y and Z can be swapped in the previous definitions to get the formulas for the conditioning on Y. Below we implement functions to derive the needed marginal and conditional distributions;

```
Y = 1:3
Z = 1:3

# marginal distribution of z
get_marginal_distro_z <- function(z){
  return(sum(J[,z]))
}

# marginal distribution of y
get_marginal_distro_y <- function(y){
  return(sum(J[y,]))
}

# marginal distribution of y given z
get_conditional_distro_y <- function(y, z){
  return(J[y,z] / get_marginal_distro_z(z))
}

# marginal distribution of z given y
get_conditional_distro_z <- function(y, z){
  return(J[y,z] / get_marginal_distro_y(y))
}
```

$Pr(Y|Z = 1)$

```
get_conditional_distro_y(Y, 1)
```

```
        1         2         3
0.2000000 0.3333333 0.4666667
```

$Pr(Y|Z = 2)$

```
get_conditional_distro_y(Y, 2)
```

```
         1          2          3
0.54838710 0.38709677 0.06451613
```

$Pr(Y|Z = 3)$

```
get_conditional_distro_y(Y, 3)
```

```
        1         2         3 
0.2564103 0.2820513 0.4615385 
```

$Pr(Z|Y = 1)$

```
get_conditional_distro_z(1, Z)
```

```
        1         2         3 
0.1818182 0.5151515 0.3030303 
```

$Pr(Z|Y = 2)$

```
get_conditional_distro_z(2, Z)
```

```
        1         2         3 
0.3030303 0.3636364 0.3333333 
```

$Pr(Z|Y = 3)$

```
get_conditional_distro_z(3, Z)
```

```
         1          2          3 
0.41176471 0.05882353 0.52941176 
```

We can derive infinite more conditional probability distributions by fixing values outside the support, but every distribution we would get this way would have probability equal to zero as it is conditioned on something impossible.

$Pr(Y|Z = z) = 0 \; \forall z \notin \{1, 2, 3\}$

$Pr(Z|Y = y) = 0 \; \forall y \notin \{1, 2, 3\}$

---

3. Make sure they are probability distributions.

---

3. Answer:

To make sure every conditional distribution we created is a probability distribution we can simply apply the previously defined functions over them:

```
# function that verifies both properties
is_valid_distro <- function(A){
  return(verify_non_negativity(A) & verify_sum_to_one(A))
}
```

$Pr(Y|Z = 1)$

```r
is_valid_distro(get_conditional_distro_y(Y, 1))
```

```
[1] TRUE
```

$Pr(Y|Z = 2)$

```r
is_valid_distro(get_conditional_distro_y(Y, 2))
```

```
[1] TRUE
```

$Pr(Y|Z = 3)$

```r
is_valid_distro(get_conditional_distro_y(Y, 3))
```

```
[1] TRUE
```

$Pr(Z|Y = 1)$

```r
is_valid_distro(get_conditional_distro_z(1, Z))
```

```
[1] TRUE
```

$Pr(Z|Y = 2)$

```r
is_valid_distro(get_conditional_distro_z(2, Z))
```

```
[1] TRUE
```

$Pr(Z|Y = 3)$

```r
is_valid_distro(get_conditional_distro_z(3, Z))
```

```
[1] TRUE
```

---

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

4. Answer:

We can simulate from the `J` by using the `sample` method. We will have to treat the matrix as a vector and turn back the indices of the vector into tuples.

```r
index_to_tuple <- function(n){
  row = ((n - 1) %% 3) + 1
  col = ((n - 1) %/% 3) + 1

  return(c(row, col))
}
```

```r
sample_from_J <- function(n){
  indices = sample(length(J), n, replace = TRUE, prob = J)
  tuples = index_to_tuple(indices)
  return(tuples)
}
```

To check the correctness of the sampling we can confront the empirical average of the sampling with the true distribution:

```r
sample_from_J_vector <- function(n){
  indices = sample(length(J), n, replace = TRUE, prob = J)
  return(indices)
}

n_samples = 10000000

samples = sample_from_J_vector(n_samples)

emp_mean = hist(samples, breaks = seq(0, 9), plot = FALSE)$counts / n_samples

norm(J - emp_mean, type = "2")
```

```
[1] 0.0001350278
```

An alternative method would be exploiting the marginal and conditional distributions of sample separately Y and Z. We present the case where we first sample Z using the marginal distribution and then the Y conditioned on the Z obtained:

```r
# code TODO
```

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.

1. Answer:

The main ingredients of the Bayesian model are * Prior distribution * Likelihood function

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 19 09:23:46 2024