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:**

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

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

```
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 properties.

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

verify_non_negativity(J)
```

```
[1] TRUE
```

As we can see and was evident from the previous code block that printed the matrix the first property is satisfied. 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 $\{1,2,3\}$; we have 2 random variables with 3 possible values each, therefore $3*2=6$ conditional distributions. 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:

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

Base elements of this second formula are elements of the joint distribution, therefore we have derived all ingredients we need for the computation. In the previous definitions Y and Z can be swapped to get the formulas for Z conditioned on Y. Below we implement the functions to get the marginal and conditional distributions:

```
# support of y
Y = 1:3
# support of z
Z = 1:3

# marginal distribution of z
get_marginal_distro_Z <- function() {
    marginal = c()
    for (z in 1:3) {
        marginal = c(marginal, sum(J[, z]))
    }
    return(marginal)
}

# marginal distribution of y
get_marginal_distro_Y <- function() {
    marginal = c()
    for (y in 1:3) {
        marginal = c(marginal, sum(J[y, ]))
    }
    return(marginal)
}

# 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])
}
```

Here we call the functions to get for each possible conditioning the conditional probability distribution:

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
```

---

3. Make sure they are probability distributions.

---

3. Answer:

To make sure every conditional distribution we created is a probability distribution we can simply use the functions defined over point $A.1$ and check whether they give us a positive response. To reduce the amount of code required we create a new function that checks them both:

```
# function that verifies if a probability
# distribution satisfies both properties
is_valid_distro <- function(A) {
    return(verify_non_negativity(A) & verify_sum_to_one(A))
}
```

Pr(Y | Z = 1):

```
is_valid_distro(get_conditional_distro_Y(Y, 1))
```

[1] TRUE

Pr(Y | Z = 2):

```
is_valid_distro(get_conditional_distro_Y(Y, 2))
```

[1] TRUE

Pr(Y | Z = 1):

```
is_valid_distro(get_conditional_distro_Y(Y, 3))
```

[1] TRUE

Pr(Z | Y = 1):

```
is_valid_distro(get_conditional_distro_Z(1, Z))
```

[1] TRUE

Pr(Z | Y = 2):

```
is_valid_distro(get_conditional_distro_Z(2, Z))
```

[1] TRUE

Pr(Z | Y = 3):

```
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))
}

sample_from_J <- function() {
    indices = sample(length(J), size = 1, prob = J)
    tuples = index_to_tuple(indices)
    return(tuples)
}
```

We can think of an alternative method that exploits the previously computed marginal and conditional distributions. We first sample Z from its marginal distribution and then sample Y from the distribution conditioned on the result of the first sampling.

```r
# sampling from marginal Z
sample_from_marginal_Z <- function() {
    distro = get_marginal_distro_Z()
    sample(length(distro), size = 1, prob = distro)
}

# sampling from Y conditioned on Z
sample_from_Y_conditioned <- function(z) {
    distro = get_conditional_distro_y(Y, z)
    sample(length(distro), size = 1, prob = distro)
}

sample_marginal_conditional_1 <- function() {
    z = sample_from_marginal_Z()
    y = sample_from_Y_conditioned(z)
    return(c(y, z))
}
```

Of curse we can also do the opposite, first sampling from the marginal distribution of Y and then from the conditional distribution of Z conditioned on the result of the first sampling.

```r
# sampling from marginal Y
sample_from_marginal_Y <- function() {
    distro = get_marginal_distro_Y()
    sample(length(distro), size = 1, prob = distro)
}

# sampling from Z conditioned on Y
sample_from_Z_conditioned <- function(y) {
    distro = get_conditional_distro_z(y, Z)
    sample(length(distro), size = 1, prob = distro)
}

sample_marginal_conditional_2 <- function() {
    y = sample_from_marginal_Z()
    z = sample_from_Y_conditioned(y)
    return(c(y, z))
}
```

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**: the prior distribution $\pi(\theta)$ is a probability distribution that represents the knowledge we have about the parameter of interest θ before any evidence gathered from observing the data is taken into account.
- **Likelihood function**:

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

Since our knowledge about the problem suggests us that the problem can be modeled using an exponential distribution the choice for a conjugate prior ends on a Gamma distribution. We indicate the shape and rate parameters of the Gamma respectively with α and β, while we use μ and σ^2 for mean and variance.

Since $\mu = \frac{\alpha}{\beta}$ and $\sigma^2 = \frac{\alpha}{\beta^2}$ we can easily derive:

$\alpha = \frac{\mu^2}{\sigma^2} = \frac{0.003^2}{0.00173^2} = 3.007$

$\beta = \frac{\mu}{\sigma^2} = \frac{0.003}{0.00173^2} = 1002.372$

Therefore:

$\pi(\theta) \sim Gamma(\alpha = 3.007, \beta = 1002.372)$

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: Sat Apr 20 13:22:02 2024`