Decision Theory and Bayesian Analysis

Dr. Vilda Purutcuoglu

[1]

[1] Edited by Anil A. Aksu based on lecture notes of STAT 565 course by Dr. Vilda Purutcuoglu

Contents

LECTURE 1
Bayesian Paradigm

1.1. Bayes theorem for distributions

If A and B are two events,

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}. \tag{1.1}$$

This is just a direct consequence of the multiplication law of probabilities that says we can express $P(A \mid B)$ as either $P(A)P(B \mid A)$ or $P(B)P(A \mid B)$. For discrete distributions, if Z, Y are discrete random variables

$$P(Z = z \mid Y = y) = \frac{P(Z = z)P(Y = y \mid Z = z)}{P(Y = y)}. \tag{1.2}$$

- How many distributions do we deal with here?

We can express the denominator in terms of the distribution in the numerator[**1**].

$$P(Y = y) = \sum_z P(Y = y, Z = z) = \sum_z P(Z = z)P(Y = y \mid Z = z). \tag{1.3}$$

- This is sometimes called the law of total probability

In this context, it is just an expression of the fact that as z ranges over the possible values of Z , the probabilities on the left hand-side of equation 1.2 make up the distribution of Z given $Y = y$, and so they must add up to one. The extension to continuous distribution is easy. If Z, Y are continuous random variable,

$$f(Z \mid Y) = \frac{f(Z)f(Y \mid Z)}{f(Y)}. \tag{1.4}$$

where the denominator is now expressed as an integral:

$$f(Y) = \int f(Z)f(Y \mid Z)\mathrm{d}Z. \tag{1.5}$$

$$f = \begin{cases} continous \quad name? \\ discrete \quad name? \end{cases} \tag{1.6}$$

1.2. How Bayesian Statistics Uses Bayes Theorem

Theorem 1.7 (Bayes' theorem)**.**

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

$P(B)$=if we are interested in the event B, $P(B)$ is the initial or prior probability of the occurence of event B. Then we observe event A $P(B \mid A)$ = How likely B is when A is known to have occurred is the posterior probability $P(B \mid A)$.

Bayes' theorem can be understood as a formula for updating from prior to posterior probability, the updating consists of multiplying by the ratio $P(B \mid A)/P(A)$. It describes how a probability changes as we learn new information. Observing the occurrence of A will increase the probability of B if $P(B \mid A) > P(A)$. From the law of total probability,

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c) + P(A \mid B^c)P(B^c). \tag{1.8}$$

where $P(B^c) = 1 - P(B)$.

Lemma 1.9.

$$P(A \mid B) - P(A) = \frac{P(A) - P(A \mid B^c)P(B^c)}{1 - P(B^c)} - P(A)$$

Proof.

$$P(A \mid B) - P(A) = \frac{P(A) - P(A \mid B^c)P(B^c) - P(A) + P(A)P(B^c)}{P(B)}$$

$$P(A \mid B) - P(A) = \frac{P(B^c)(P(A) - P(A \mid B^c))}{P(B)}$$

$$P(A \mid B) - P(A) = P(B^c)(\frac{P(B)P(A \mid B) + P(B^c)P(A \mid B^c)}{P(B)} - \frac{P(A \mid B^c)}{P(B)})$$

$$P(A \mid B) - P(A) = P(B^c)(P(A \mid B) - \frac{P(A \mid B^c)(1 - P(B^c))}{P(B)})$$

$$P(A \mid B) - P(A) = P(B^c)(P(A \mid B) - P(A \mid B^c))$$

□

1.2.1. Generalization of the Bayes' Theorem

Let $B_1, ..., B_n$ be a set of mutually exclusive events. Then

$$P(B_r \mid A) = \frac{P(B_r)P(A \mid B_r)}{P(A)} = \frac{P(B_r)P(A \mid B_r)}{\sum_{i=1}^n P(B_r)P(A \mid B_r)}. \tag{1.10}$$

- Assuming that $P(B_r) > 0$,$P(A \mid B) > P(A)$ if and only if $P(A \mid B) > P(A \mid B^c)$.

- In Bayesian inference we use Bayes' theorem in a particular way.
- Z is the parameter (vector) θ.
- Y is the data (vector) X.

So we have

$$f(\theta \mid X) = \frac{f(\theta)f(X \mid \theta)}{f(X)} \tag{1.11}$$

$$f(X) = \int f(\theta)f(X \mid \theta)\mathrm{d}\theta. \tag{1.12}$$

$$f(\theta) = \tag{1.13}$$

$$f(\theta \mid X) = \tag{1.14}$$

$$f(X \mid \theta) = \tag{1.15}$$

1.2.2. Interpreting our sense

How do we interpret the things we see, hear, feel, taste or smell?

Example 1.2.1. I hear a song on the radio I identify the singer as Robbie Williams. Why do I think it's Robbie Williams?. Because he sounds like that. Formally, $P($ What I hear Robbie Williams $) >> P($What I hear someone else $)$

Example 1.2.2. I look out of the window and see what appears to be a tree. It has a big, dark coloured part sticking up out of the ground that branches into thinner sticks and on the ends of these are small green things. Clearly, $P($ What I hear Robbie Williams $) >> P($What I hear someone else $)$

1.3. Prior to Posterior

1.4. Triplot

[1]

[1]All plots are generated in R, relevant codes are provided in Appendix R Codes

1.4.1. Weak Prior Information

It is the case where the prior information is much weaker that the data. This will occur, for instance, if we do not have strong information about Q before seeing the data, and if there are lots of data. Then in triplot, the prior distribution will be much broader and flatter that the likelihood. So the posterior is approximately proportional to the likelihood.

Example 1.4.1. Classify the follow differential equations as ODE's or PDE's, linear or nonlinear, and determine their order. For the linear equations, determine whether or not they are homogeneous.

LECTURE 2
Some Common Probability Distributions

LECTURE 3
Inference

Basic Statistics

R Codes

Listing 1. Triplot Code in R

```
##################################################
#                                               #
#   Mean and Standard Deviation Calculation       #
#                 by Anil Aksu                    #
#   It is developed to show some basics of R      #
#                                    #
##################################################

## library required to read excel files
require(gdata)
## functions
source('getMean.R')
source('getSigma.R')
## the random data read from excel file
RandomNumbers <- read.xls("RandomNumbers.xlsx", perl ...
    = "C:\\Perl\\bin\\perl.exe")
## let's calculate the mean
Mean <- getMean(RandomNumbers)
## the standard deviation
Sigma <- getSigma(RandomNumbers)
## this function gets numbers from console

## let's output them
print("Random Numbers")
print(RandomNumbers[,1])
print("The Mean")
print(Mean)
print("The Standard Deviation")
print(Sigma)
```

BIBLIOGRAPHY

1. Allen B. Dawney. *Think Bayes: Bayesian Statistics in Python.* O'REILLY, 2013.