VE414 Lecture 1

Jing Liu

UM-SJTU Joint Institute

May 14, 2019

Theorem (Bayes' Theorem)

Let $\mathcal{A}$ and $\mathcal{B}$ be two events, and $\Pr(\mathcal{B}) \neq 0$, then

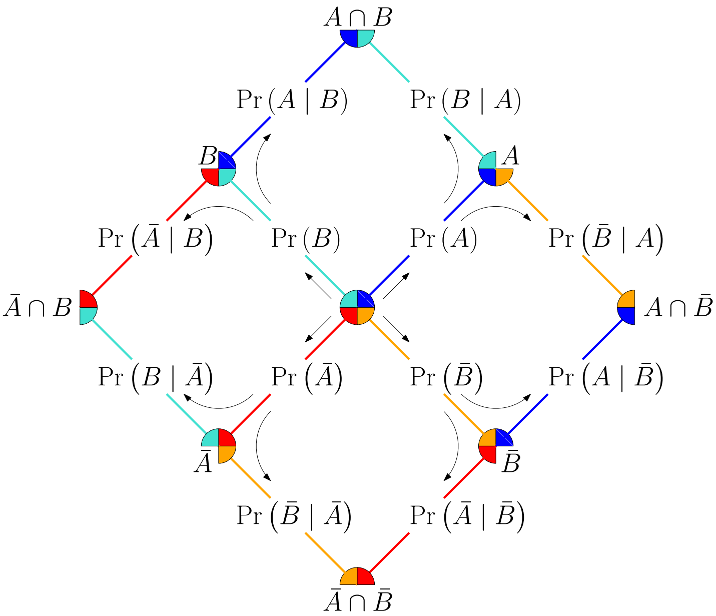
$$\Pr(\mathcal{A} \mid \mathcal{B}) = \frac{\Pr(\mathcal{B} \mid \mathcal{A}) \Pr(\mathcal{A})}{\Pr(\mathcal{B})}$$

In general, for some partition $\{\mathcal{A}_j\}$ of the sample sample, then

$$\Pr(\mathcal{A}_i \mid \mathcal{B}) = \frac{\Pr(\mathcal{B} \mid \mathcal{A}_i) \Pr(\mathcal{A}_i)}{\sum_j \Pr(\mathcal{B} \mid \mathcal{A}_j) \Pr(\mathcal{A}_j)}$$

- It is named after Thomas Bayes, however, he actually has very little to do with this formula, it was Laplace that worked out the mathematical form.
- It is not difficult to understand using the concept of conditional probability.

Q: Can you give a tree diagram of the probability space for the simple form?


$A \cap B$
$\Pr(A \mid B)$
$\Pr(B \mid A)$
B
A
$\Pr(\bar{A} \mid B)$
$\Pr(B)$
$\Pr(A)$
$\Pr(\bar{B} \mid A)$
$\bar{A} \cap B$
$A \cap \bar{B}$
$\Pr(B \mid \bar{A})$
$\Pr(\bar{A})$
$\Pr(\bar{B})$
$\Pr(A \mid \bar{B})$
$\bar{A}$
$\bar{B}$
$\Pr(\bar{B} \mid \bar{A})$
$\Pr(\bar{A} \mid \bar{B})$
$\bar{A} \cap \bar{B}$

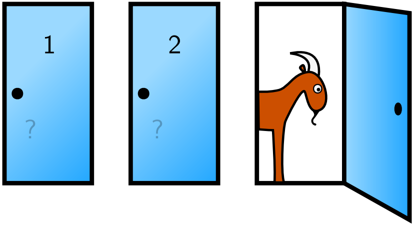
- The famous Monty Hall problem arises from a popular television game show

Let's Make a Deal

is often used to illustrate the Bayes' theorem and Bayesian inference.

Monty Hall Problem

Suppose you are on a game show, and you are given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 2, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to switch to door No. 1?"



Is it to your advantage to switch your choice?

- A typical solution:

  Let $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ denote the events in which the car is placed behind door No.1, No.2 and No.3, respectively. Without loss of generality, suppose we select door No.2 and Monty opens No.3.

$$\Pr(\mathcal{C}_2 \mid \mathcal{O}_3) = \frac{\Pr(\mathcal{O}_3 \mid \mathcal{C}_2)\Pr(\mathcal{C}_2)}{\sum_{j=1}^{3}\Pr(\mathcal{O}_3 \mid \mathcal{C}_j)\Pr(\mathcal{C}_j)} = \frac{1/2 \cdot 1/3}{1 \cdot 1/3 + 1/6 + 0 \cdot 1/3} = \frac{1}{3}$$

$$\Pr(\mathcal{C}_1 \mid \mathcal{O}_3) = \frac{\Pr(\mathcal{O}_3 \mid \mathcal{C}_1)\Pr(\mathcal{C}_1)}{\sum_{j=1}^{3}\Pr(\mathcal{O}_3 \mid \mathcal{C}_j)\Pr(\mathcal{C}_j)} = \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1/6 + 0 \cdot 1/3} = \frac{2}{3}$$

  where $\mathcal{O}_1$, $\mathcal{O}_2$ and $\mathcal{O}_3$ denote the events in which Monty opens door No.1, No.2 and No.3, respectively.

  uses Bayes' theorem to see switching is the right choice. But it often fails to attract one's attention to the reasoning behind this solution thus this choice.

- Before Monty opens door No.3 and reveals the goat behind it,

$$\Pr\left(\mathcal{C}_2\right) = \frac{1}{3} = \Pr\left(\mathcal{C}_1\right)$$

we have no reason to switch, however, the information provided by Monty,

$$\Pr\left(\mathcal{C}_2 \mid \mathcal{O}_3\right) = \frac{1}{3} \qquad \text{and} \qquad \Pr\left(\mathcal{C}_1 \mid \mathcal{O}_3\right) = \frac{2}{3}$$

allows us to improve our chance of winning the car.

- Using Bayes' theorem, we essentially updated our understanding regarding the likelihood of where the car is from the data (not behind door No. 3).
- We could introduce some random variables to make this more concise:

$$Y = \begin{cases} -1, & \text{if } \mathcal{C}_1 \text{ happens}, \\ 0, & \text{if } \mathcal{C}_2 \text{ happens}, \\ 1, & \text{if } \mathcal{C}_3 \text{ happens}, \end{cases} \qquad \text{and} \qquad X = \begin{cases} -1, & \text{if } \mathcal{O}_1 \text{ happens}, \\ 0, & \text{if } \mathcal{O}_2 \text{ happens}, \\ 1, & \text{if } \mathcal{O}_3 \text{ happens}. \end{cases}$$

- Notice how the information provided by Monty updated our understanding

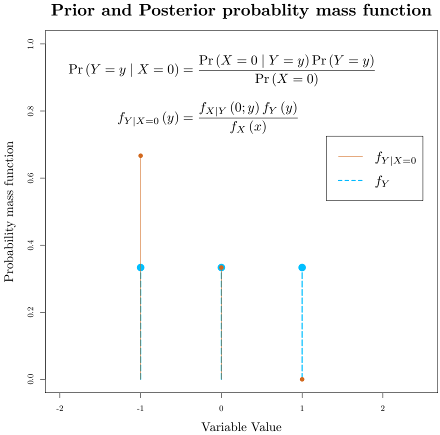


Figure: R Code: `monty_hall_prior_posterior_414.R`

Definition

In Bayesian analysis, without any data, the unobserved parameter or variable is modelled as a random variable Y having a marginal distribution

$$f_Y$$

This marginal distribution is known as the prior. The data is modelled as a random variable X given Y having a conditional distribution

$$f_{X|Y}$$

Given a particular realisation of the data $X = x$ is observed, the function

$$\mathcal{L}(y; x) = f_{X|Y}(x \mid y)$$

as a function of y only is called the likelihood. The conditional distribution of Y,

$$f_{Y|X=x}$$

given the observed $X = x$, is known as the posterior.

- This tiny problem gives a taste of what Bayesian inference is about, namely, updating one's understanding when data become available, i.e.

$$f_Y \longrightarrow f_{Y|X=x}$$

using Bayes' theorem

$$\Pr(Y = y \mid X = x) = \frac{\Pr(X = x \mid Y = y)\Pr(Y = y)}{\Pr(X = x)}$$

$$\Longrightarrow f_{Y|X=x}(y) = \frac{f_{X|Y}(x \mid y)\, f_Y(y)}{f_X(x)}$$

- However, using Bayes' theorem does not make someone a frequentist, using it for everything does! That is, Bayesianism is about creating models that are very much like the mechanics behind learning from experience, being a Bayesian means the learning rule is always some form of Bayes' theorem.