Bayesian 1

Jan Rundt

November 7, 2022

1 Exercise 1

First of all the prior P(not spam) = 0,35

Likelihood P(classified as spam | not spam) = 0,06

Marginal probability for classified as spam = P(classified as spam | not spam)P(not spam) + P(classified as spam | spam)P(spam) = $0,06 * 0,35 + 0,75 * 0,65 = 0,021 + 0,4875 \approx 0,51$

So using the bayesian formula:

$$P(\text{not spam}|\text{class spam}) = \frac{P(\text{class spam}|\text{not spam})P(\text{not spam})}{P(\text{class spam}|\text{not spam})P(\text{not spam}) + P(\text{class spam}|\text{spam})P(\text{spam})}$$

$$\frac{0,06 * 0,35}{0,51} = \frac{0,21}{0,51} \approx 0,04$$

So the posterior probability that the message is not spam after being marked as spam is 4%.

2 Exercise 2

2.1 Task a

Interestingly, while reading the first chapter of BDA and the text "Dicing with uncertainty" (O'Hagan, 2004), I myself tend to fall back into just the kind of incorrect thinking about frequencies versus probabilities that the authors describe. A frequentist probability still mostly feels like a description of certainty, although that's not the way it should be interpreted. And this although I just finished a course that described frequentist statistics and also pointed this fact out quite clearly. So no wonder that anyone consuming statistics that is not a statistician him- or herself is prone to make that same kind of mistake.

The conceptual split into aleatory and epistemic uncertainty is an interesting one, and one that I have not encountered before, at least with this description. In a way, taking perhaps the most extreme example, one could picture a toddler experiencing the world as a collection of aleatory uncertainties, as a lot of the consequences of its actions must seem fairly random to it - e.g. a chair tripping over when you lift its leg up too high. Similar to how a toddler experiences the world, through scientific investigation, a scientific process enters an area where we before only could experience randomness and tries to find the hidden causes of events, sorting out the epistemic uncertainties, diminishing the amount of randomness, finding the "truly" aleatory uncertainties within the field.

I guess this is what uncertainty to me does seem like, in relation to data analysis. Data analysis seems like a way of reducing the seeming randomness of the data, by finding those characteristics that can be epistemically explained, and pinpointing those uncertainties that can be described and quantified as an aleatory uncertainty. Data analysis therefore becomes a way of understanding the world by reducing the amount and classifying the nature of our uncertainty in our description of it.

2.2 Task b

Statement: "The probability of event E is considered subjective if two rational persons A and B can assign unequal probabilities to E, PA(E) and PB(E). These probabilities can also be interpreted as conditional: PA(E) = P(E|IA) and PB(E) = P(E|IB), where IA and IB represent the knowledge available to persons A and B, respectively."

2.2.1 1. A roll of a die

In this case the epistemic uncertainty of persons A and B is so big, it could almost be called categorical. A is in full knowing, there is no uncertainty in his/her case, and he/she does only have to trust his/her memory, if any considerable time has passed from seeing the result of the die rolling. We could of course, for the sake of staying true to the statement, call this a case of very unequal subjective probabilities where person A has a very high probability - only limited by his/her eyesight and memory - of knowing, while B can only do a basic probabilistic guess of 1/6. We could also view it as conditional probabilities, where the conditions in one case is quite superior to the other, in rendering realistic probabilities.

2.2.2 2. Soccer match

In this example the subjective probability or conditional probability makes more sense. Here especially the concept of conditional probability becomes interesting, as clearly it is not only a question of subjective views, randomly assigned to the different persons, which lies behind the difference in their opinion. Establishing how big the difference between the knowledge of the two persons really is would be a method to weigh their different opinions, or the probabilities that they assign to the event of Brazil winning. Lets say both of them are still gamblers, clearly the difference chance they have of winning a bet on the World Cup would be different considering their knowledge, and this difference could probably be measured in some way.

3 Exercise 3

3.1 Task a

The marginal distribution is given by the formula:

$$\begin{aligned} p(y) &= p(\theta = 1)p(y|\theta = 1) + p(\theta = 2)p(y|\theta = 2) \\ &= 0.5 * Bin(n, y, p_1) + 0.2 * Bin(n, y, p_2) \end{aligned}$$

Where n=10, $p_1 = 0.2$, $p_2 = 0.6$ and Bin is the binary distribution:

$$\binom{n}{y}(1-p)^{n-y}p^y$$

In R with code:

```
p_y <- function(y){
  0.5*dbinom(y, 10, 0.2) + 0.5*dbinom(y, 10, 0.6)
}
y <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
plot(y, p_y(y), "h", ylab="Probabilities")
```

See below Figure 1 for the plot.

3.2 Task b

We get the probability $Pr(\theta = 1|y = 3)$ with the bayesian formula, using prior $Pr(\theta = 1) = 0.5$, likelihood $P(y = 3|\theta = 1) \approx 0.2$ and marginal distribution from the formula in a) above.

$$\begin{aligned} Pr(\theta = 1|y = 3) &= \frac{Prior * Likelihood}{Marginaldistribution} \\ &= \frac{0.5 * 0.2}{0.12} \approx 0.83 \end{aligned}$$

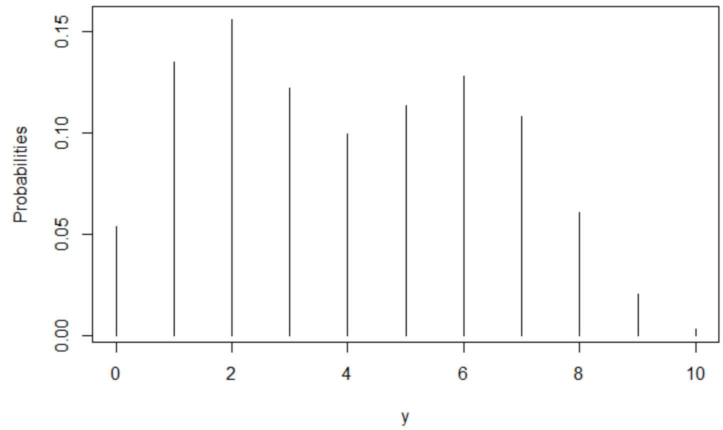

Figure 1: Marginal probability density for y plotted.

3.3 Task c

With the posterior distribution for θ, we get a new formula for the marginal distribution:

$$\begin{aligned} posterior_p(y) &= p(\theta = 1)p(y|\theta = 1) + p(\theta = 2)p(y|\theta = 2) \\ &= 0.83 * Bin(n, y, p_1) + 0.17 * Bin(n, y, p_2) \end{aligned}$$

Coded in R:

```
posterior_y <- function(y){
  0.83*dbinom(y, 10, 0.2) + 0.17*dbinom(y, 10, 0.6)
}
```

And as a plotted density:

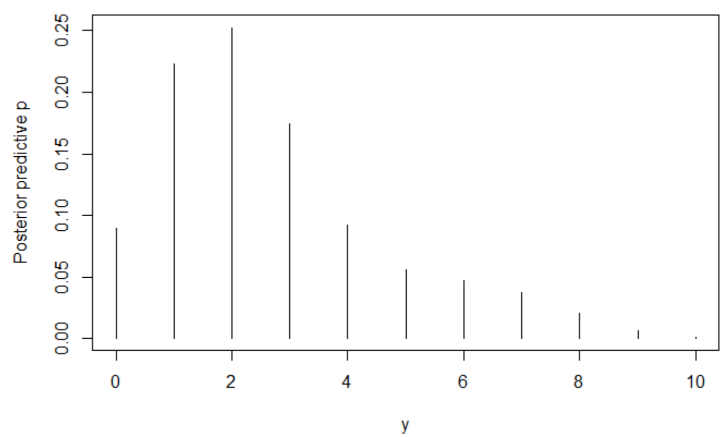

Figure 2: Posterior predictive distribution for y plotted.

The posterior predictive probability from the formula above is $posterior_p(3) \approx 0.17$