Week1

Abdallah Daaboul

11/2/2020

Exercise 1

$P(Spam) = 0.65$

$P(ClassifiedAsSpam|Spam) = 0.75$

$P(ClassifiedAsSpam|NotSpam) = 0.06$

By applying the Bayes Rule we get:

$$P(NotSpam|ClassifiedAsSpam) = \frac{P(ClassifiedAsSpam|NotSpam) * P(NotSpam)}{P(ClassifiedAsSpam)}$$

$P(NotSpam) = 1 – P(Spam) = 0.35$

$P(ClassifiedAsSpam|NotSpam) = 0.06$

$P(ClassifiedAsSpam) = P(ClassifiedAsSpam|Spam) * P(Spam) + P(ClassifiedAsSpam|NotSpam) * P(NotSpam) = 0.75 * 0.65 + 0.06 * 0.35 = 0.5085$

Therefore:

$$P(NotSpam|ClassifiedAsSpam) = \frac{0.06 * 0.35}{0.5085} = 0.04 = 4\%$$

Exercise 2

b)

Reflect your thoughts about the above texts. For example, does the aleatory and epistemic uncertainties make sense? Describe how you understand the term uncertainty. Where does uncertainty arise from, how does it relate to data analysis? :

I totally agree with this article. The way it divides aleatory and epimestic uncertainties is very convincing. Yes, there is times where we are uncertain about something just because we don't have the expertise in its subject or at least we lack in knowledge towards it. While uncertainty could be independant from our knowledge and isn't subjective at all and this is the case of the aleatory uncertainty.

Uncertainty is simply when we can't be sure 100 % about a subject. Uncertainty has a big range and could go from lacking a little bit of the total truth to knowing just few information about what could happen. Uncertainty is related to something that will happen in the near futur. And we will try to predict the outcome of what will happen or model this outcome by creating a probabilistic model. At this point, we could talk about epistemic and aleatory uncertainties. Eather the outcome of the event could be known by improving our knowledge or is is just an aleatory event that we can't modelise perfectly.

Uncertainty is a major key to data analysis. because being uncertain about something lead us to be more curious and find relation between this feature and the rest of the existing features in order to get wiser and try to understand more our initial variable. Uncertainty leads to knowledge. Everything we know, was neglected at some point before becoming an acquired notion.

c)

– The probability that a 6 appears when a fair die is rolled, where A observes the outcome of the die roll and B does not:

In this case, A will know if we got a 6 or not. Thus, his answer will be eather yes or no. If A observed a 6, the answer will be yes (P(6)=1 in this case). If not, the answer is no (P(6)=0 in this case).

However, B doesn't know the answer yet. Therefore, the probability of getting a 6 is the probability that the outcome of the dice is a 6 which is 1/6.

– The probability that Brazil wins the next World Cup, where A is ignorant of soccer and B is a knowledgeable sports fan:

The person without knowledge about football doesn't have any idea about it. He doesn't know which team is better actually, nor the team with the best history, nor the actual FIFA rakings of the countries. Therefore his answer could be: the probability that Brazil wins the world cup is the same as for any other country. Thus, P(Brazil wins the worldcup)) = 1/32.

For the person with soccer knowledge, the answer depends on the type of knowledge he has. For example if he only knows the history of football and neglect recent performances, he could say: brazil won 5 out of 21 world cups thererfore P(Brazil wins the worldcup)= 5/21. If this person knows the raking of Brazil today which is 3, he could say that brazil has a big probability of winning the world cup while this probability should be less than the one for france (ranked 2) and belgium (ranked 1) and greater than the other countries.

Exercise 3

a)

The formula for the marginal distribution is : $P(y = \alpha) = \sum P(y = \alpha|\theta i) * P(\theta i)$

$P(y=0) = \sum P(y=0|\theta i) * P(\theta i) = 0.5 * (Binomial(0, 10, p = 0.2) + Binomial(0, 10, p = 0.6)) = 0.5 * (0.1073741824 + 0.0001048576)$
$P(y=1) = \sum P(y=1|\theta i) * P(\theta i) = 0.5 * (Binomial(1, 10, p = 0.2) + Binomial(1, 10, p = 0.6)) = 0.5 * (0.268435456 + 0.0001048576)$
$P(y=2) = \sum P(y=2|\theta i) * P(\theta i) = 0.5 * (Binomial(2, 10, p = 0.2) + Binomial(2, 10, p = 0.6)) = 0.5 * (0.301989888 + 0.010616832)$
$P(y=3) = \sum P(y=3|\theta i) * P(\theta i) = 0.5 * (Binomial(3, 10, p = 0.2) + Binomial(3, 10, p = 0.6)) = 0.5 * (0.201326592 + 0.042467328)$
$P(y=4) = \sum P(y=4|\theta i) * P(\theta i) = 0.5 * (Binomial(4, 10, p = 0.2) + Binomial(4, 10, p = 0.6)) = 0.5 * (0.088080384 + 0.111476736)$
$P(y=5) = \sum P(y=5|\theta i) * P(\theta i) = 0.5 * (Binomial(5, 10, p = 0.2) + Binomial(5, 10, p = 0.6)) = 0.5 * (0.0264241152 + 0.2006581248)$
$P(y=6) = \sum P(y=6|\theta i) * P(\theta i) = 0.5 * (Binomial(6, 10, p = 0.2) + Binomial(6, 10, p = 0.6)) = 0.5 * (0.005505024 + 0.250822656)$
$P(y=7) = \sum P(y=7|\theta i) * P(\theta i) = 0.5 * (Binomial(7, 10, p = 0.2) + Binomial(7, 10, p = 0.6)) = 0.5 * (0.000786432 + 0.214990848)$
$P(y=8) = \sum P(y=8|\theta i) * P(\theta i) = 0.5 * (Binomial(8, 10, p = 0.2) + Binomial(8, 10, p = 0.6)) = 0.5 * (0.000073728 + 0.120932352)$
$P(y=9) = \sum P(y=9|\theta i) * P(\theta i) = 0.5 * (Binomial(9, 10, p = 0.2) + Binomial(9, 10, p = 0.6)) = 0.5 * (0.000004096 + 0.040310784)$
$P(y=10) = \sum P(y=10|\theta i) * P(\theta i) = 0.5 * (Binomial(10, 10, p = 0.2) + Binomial(10, 10, p = 0.6)) = 0.5 * (0.000001 + 0.0060466176)$

I will use R in order to do these caculations using a simple function and a loop

```
# This function represents the formula above
MarginalDistrib =  function(x) {
  temp = 0.5 * dbinom(x,10,0.2)  + 0.5 * dbinom(x,10,0.6)
  return(temp)
}

x=0:10

for(i in 1:11) {
  x[i]=MarginalDistrib(i-1)
  cat("y=",i-1, "=> P=",x[i],"\n")
}
```

```
## y= 0 => P= 0.05373952
## y= 1 => P= 0.1350042
## y= 2 => P= 0.1563034
## y= 3 => P= 0.121897
## y= 4 => P= 0.09977856
## y= 5 => P= 0.1135411
## y= 6 => P= 0.1281638
## y= 7 => P= 0.1078886
## y= 8 => P= 0.06050304
## y= 9 => P= 0.02015744
## y= 10 => P= 0.00302336
```

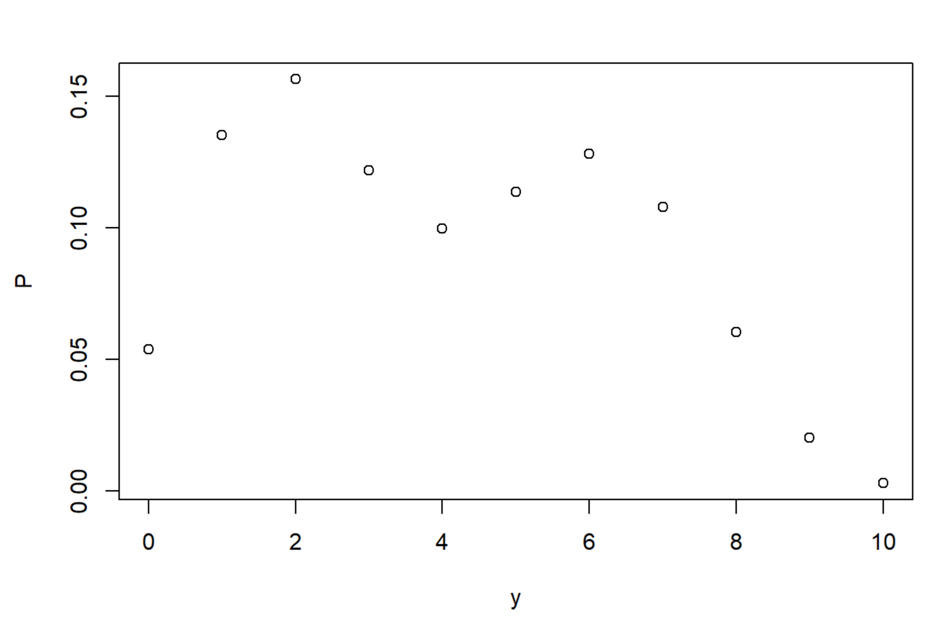
Let's check if the sum of our probabilities is 1

```
print(sum(x))
```

```
## [1] 1
```

Density

```
plot(0:10 , x , xlab='y',ylab='P')
```



b)

$$Pr(\theta = 1|y = 3) = \frac{Pr(y = 3|\theta = 1) * Pr(\theta = 1)}{Pr(y = 3)}$$

$Pr(y = 3|\theta = 1) = dbinom(3, 10, 0.2) = 0.201326592$

$Pr(\theta = 1) = 0.5$

$Pr(y = 3) = \sum P(y = 3|\theta i) * P(\theta i) = 0.5 * (dbinom(3, 10, p = 0.2) + dbinom(3, 10, p = 0.6)) = 0.5 * (0.201326592 + 0.042467328) = 0.5 * 0.2437 = 0.12185$

$$Pr(\theta = 1|y = 3) = \frac{0.201326592 * 0.5}{0.12185} = 0.8261247 = 82\%$$

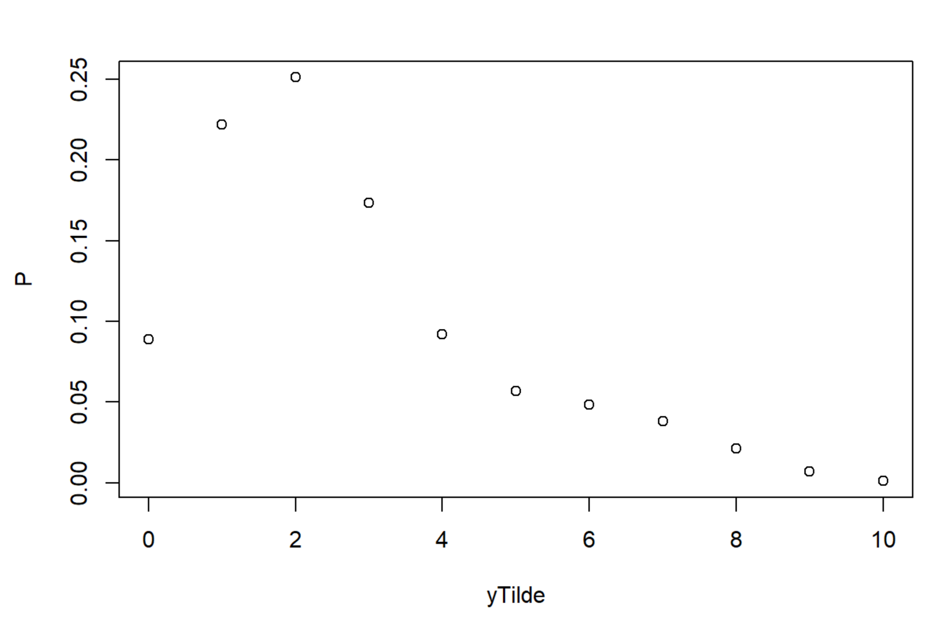
c)

$P(\tilde{y}|y = 3) = \sum P(\tilde{y}|\theta i) * P(\theta i|y = 3)$

```
# P(y|θi)
likelihood =  function(x,theta) {

  if (theta== 1) {
    return( dbinom(x,10,0.2) )
  }
  else
    return(dbinom(x,10,0.6))
}


#P (θi | y=3 ) =  Pr(y = 3 | θ = i) * Pr(θ = i) / Pr(y = 3)
posterior = function (theta){
  return (likelihood(3,theta) * 0.5 / MarginalDistrib(3) )
}

Pred = function (x) {
  likelihood(x,1) * posterior(1) +   likelihood(x,2) * posterior(2)
}



x=0:10

for(i in 1:11) {
  x[i]=Pred(i-1)
  cat("y_tilde=",i-1, "=> P=",x[i],"\n")
}
```

```
## y_tilde= 0 => P= 0.08868856
## y_tilde= 1 => P= 0.2219497
## y_tilde= 2 => P= 0.2512346
## y_tilde= 3 => P= 0.1736543
## y_tilde= 4 => P= 0.09215588
## y_tilde= 5 => P= 0.05677456
## y_tilde= 6 => P= 0.04823777
## y_tilde= 7 => P= 0.03809946
## y_tilde= 8 => P= 0.02112652
## y_tilde= 9 => P= 0.007025261
## y_tilde= 10 => P= 0.001053366
```

```
plot(0:10 , x, xlab='yTilde', ylab ='P')
```



Check the sum

```
print(sum(x))
```

```
## [1] 1
```

the posterior predictive probability that $\tilde{y}$ = 3 is : 0.1736543

```
x[4]
```

```
## [1] 0.1736543
```