Homework 3

This homework is due for 07/30/2018, 06:00pm. Please indicate your name and your netID. You can give it back to me either during a course, or during my office hours. If you want to let it in my mailbox in the lobby of WWH, please do it during the week, and write me an email to inform me.

Problem 1 - 3 points

We roll a fair die, and consider the events:

- $E_2$: “we roll a multiple of 2”,
- $E_3$: “we roll a multiple of 3”.

(a) Compute $P(E_2)$ and $P(E_3)$.

(b) Is $E_2$ independent of $E_3$? (Justify your answer).

Problem 2 - 3 points

For each question below, describe the sample space and the events you consider.

In this exercise, we consider that for a pregnant woman, it is equiprobable to deliver a girl or a boy. A mother has two children. Compute:

(a) the probability that both children are boys knowing that at least one is a boy,

(b) the probability that both children are boys knowing that the oldest is a boy.

A mother has two children, and we pick up randomly one of them. Compute:

(c) the probability that both children are boys knowing that the one picked up randomly is a boy.

Problem 3 - 3 points

An urn contains 10 balls numbered from 0 to 9. Five times, we pick a ball in the urn and note its number (we do not put it back in the urn).

We obtain a tuple $S = (s_1, s_2, s_3, s_4, s_5)$ of five numbers where $s_i$ is the number of the $i$-th ball that has been picked up.

(a) give the probability that $S$ contains an ascending or descending sequence of length 5, that is: either $\forall i < 5, s_{i+1} = s_i + 1$ or $\forall i < 5, s_{i+1} = s_i - 1$. 
(b) give the probability that $S$ contains an ascending sequence of length 4 (4 consecutive elements of the tuple $S$ are consecutive ascending integers).

(c) give the probability that $S$ contains an ascending sequence of length 5 knowing that it contains an ascending sequence of length 4.

**Problem 4 - 5 points**

When I was younger, I used to forget so often the keys of my apartment (in my apartment, so that I was locked outside) that I have been able to establish the following statistics:

- knowing that I forgot my keys the day $d$, the probability that I forgot it the day $d+1$ was $\frac{1}{10}$.
- knowing that I did not forgot my keys the day $d$, the probability that I forgot it the day $d+1$ was $\frac{4}{10}$.

Let the day 1 be the first of January, and suppose the probability that I forgot my keys on the day 1 is $x$, where $0 < x \leq 1$. We consider $d \leq 365$. Note $P_d$ the (absolute) probability that I forgot my keys the day $d$. Note $E_d$ the event “I forgot my keys the day $d$”, and $\neg E_d$ the event “I did not forgot my keys the day $d$”.

(a) Show that $E_d$ and $\neg E_d$ form a frame of discernment.

(b) Give an expression of $P_{d+1}$ in function of $P_d$.

(c) Give an expression of $P_d$ in function of $x$ and $d$.

(d) Use Matlab to fill the following table, which components are the values of $P_d$ for different $d$ and different initial value $x$.

<table>
<thead>
<tr>
<th></th>
<th>$d = 2$</th>
<th>$d = 4$</th>
<th>$d = 6$</th>
<th>$d = 8$</th>
<th>$d = 10$</th>
<th>$d = 50$</th>
<th>$d = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0.001$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x = 0.999$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(e) Would you bet that I forgot my keys the 31th of December of the same year?

*Hint:* questions (d) and (e) can be addressed independently of (c).

**Problem 5 - 6 points**

I have a list of emails that I classified (manually) in 2 categories: spam and non spam. I want to design a classification algorithm that uses machine learning for deciding if a new email arriving in my mailbox is a spam or not. This exercise is a first step through such a classifier.

I proceed as follows: first, I choose three boolean attributes as characteristics of an email that give good evidence of whether it is a spam or not. For a given email $e$, these attributes are $A_1(e)$, $A_2(e)$ and $A_3(e)$, each taking a value in \{true, false\}. These attributes can be for instance: “$e$ contains my first name”, or “$e$ contains the words amazing and sales”. For a given email, I note:

- $A_1$ the event: “$A_1(e) =$ true” and $\neg A_1$ the event “$A_1(e) =$ false”
• $A_2$ the event: “$A_2(e) = \text{true}$" and $\neg A_2$ the event “$A_2(e) = \text{false}$”

• $A_3$ the event: “$A_3(e) = \text{true}$” and $\neg A_3$ the event “$A_3(e) = \text{false}$”

• $S$ the event “$e$ is a spam” and $\neg S$ the event “$e$ is not a spam”.

I make the assumption that for a given email $e$, the events $S, \neg S, A_1, A_2, A_3, \neg A_1, \neg A_2$ and $\neg A_3$ are conditionally independents.

Now, I use the list of classified emails to determine $P(S)$, $P(A_1)$, $P(A_2)$, $P(A_3)$, $P(A_1|S)$, $P(A_2|S)$, $P(A_3|S)$ (this is the learning step) as follows: I consider that the probability of an event is the the number of emails in the classified list satisfying the event over the number of emails of the list. I obtain:

• $P(S) = 0.4$,

• $P(A_1) = 0.4$, $P(A_2) = 0.5$, $P(A_3) = 0.6$,

• $P(A_1|S) = 0.3$, $P(A_2|S) = 0.6$, $P(A_3|S) = 0.2$.

A new email $e$ arrives in my mail box: it satisfies $A_1(e) = \text{false}$, $A_2(e) = \text{true}$ and $A_3(e) = \text{true}$.

(a) Show that $S$, $\neg S$ form a frame of discernment.

(b) Use the Bayes law to give an expression of $P(S|\neg A_1 \cap A_2 \cap A_3)$ involving conditional probabilities knowing $S$ and $\neg S$.

(c) Give the probability that $e$ is a spam.