**Exercise 2: Theory of probabilities**

Given the two following tables, **T1** and **T2**.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Unknown** | **SensitiveWord** | **Spam** | **Probability** |  |  | **Spam = T** | | **Spam = F** | |
| T | T | T | 0.108 |  |  | **SensitiveWord = T** | **SensitiveWord = F** | **SensitiveWord = T** | **SensitiveWord = F** |
| T | T | F | 0.016 |  | Unknown = T | 0.108 | 0.012 | 0.016 | 0.064 |
| T | F | T | 0.012 |  | Unknown = F | 0.072 | 0.008 | 0.144 | 0.576 |
| T | F | F | 0.064 |  |  | **T2** | | | |
| F | T | T | 0.072 |  |  |
| F | T | F | 0.144 |  |  |
| F | F | T | 0.008 |  |  |
| F | F | F | 0.576 |  |  |

**T1**

1. Fill in the missing information in **T2** to be identical to **T1**
2. What do **T1** and **T2** represent?

Joint probability distributions

1. Compute (show all possible formulas and calculations):
   1. P(Unknown=T, SensitiveWord=T, Spam=T) = 0.108 (10.8%)
   2. P(Unknown=F, SensitiveWord=F, Spam=T) = 0.008 (0.8%)
   3. P(Unknown=T, Spam=T) =P(Unknown=T, SensitiveWord=x, Spam=T)

= P(Uknown=T, SensitiveWord=T, Spam=T)+

P(Uknown=T, SensitiveWord=F, Spam=T)

= 0.108 + 0.012 = 0.12

|  |  |  |  |
| --- | --- | --- | --- |
| **Unkown** | **SensitiveWord** | **Spam** | **Probability** |
| T | T | T | 0.108 |
| T | T | F | 0.016 |
| T | F | T | 0.012 |
| T | F | F | 0.064 |
| F | T | T | 0.072 |
| F | T | F | 0.144 |
| F | F | T | 0.008 |
| F | F | F | 0.576 |

* 1. P(Spam = T) =P(Unknown= y, SensitiveWord=x, Spam=T)

**= 0.108 + 0.012+0.072+0.008 = 0.2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Unkown** | **SensitiveWord** | **Spam** | **Probability** |
| T | T | T | 0.108 |
| T | T | F | 0.016 |
| T | F | T | 0.012 |
| T | F | F | 0.064 |
| F | T | T | 0.072 |
| F | T | F | 0.144 |
| F | F | T | 0.008 |
| F | F | F | 0.576 |

* 1. P(Spam = T OR Unknown = F) = P(Spam = T) + P(Unknown = F)-P(Spam = T, Unknown = F)

**=** 1- P(Spam=F, Unknown=T)= 1-0.016-0.064 = 0.92

|  |  |  |  |
| --- | --- | --- | --- |
| **Unkown** | **SensitiveWord** | **Spam** | **Probability** |
| T | T | T | 0.108 |
| T | T | F | 0.016 |
| T | F | T | 0.012 |
| T | F | F | 0.064 |
| F | T | T | 0.072 |
| F | T | F | 0.144 |
| F | F | T | 0.008 |
| F | F | F | 0.576 |

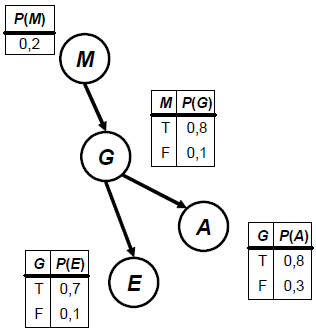
* 1. P(Spam = F | Unknown = T) = P(Spam = F, Unknown = T) / P(Unknown = T)

= (0.016+0.064)/(0.016+0.064+0.108+0.012) = 0.4

**Exercise 3: Bayesian Networks**

Considering the following Bayesian network. It consists of one graph of four nodes M, G, E and A; and four tables, T1, T2, T3 and T4:

|  |  |
| --- | --- |
| M | |
| +m | -.m |
| 0.2 | 0..8 |



|  |  |  |  |
| --- | --- | --- | --- |
|  | | G | |
| +g | .-g |
| M | +m | 0.8 | .0.2 |
| -m | 0.1 | .0.9 |

**T1**

**T2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | E | |
| +e | .-e |
| G | +g | 0.7 | .0.3 |
| -g | 0.1 | .0.9 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | A | |
| +a | .-a |
| G | +g | 0.8 | .0.2 |
| -g | 0.3 | .0.7 |

**T3**

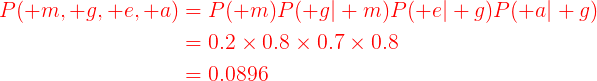
**T4**

**Note:** the signs ‘+’ and ‘-‘ stand for True and False respectively.

1. Complete the missing information in each table.

See tables above.

1. Compute P(+m, +g, +e, +a )



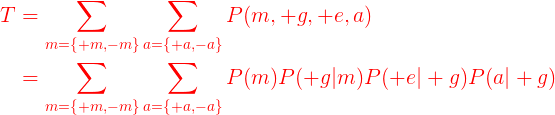
1. Given:



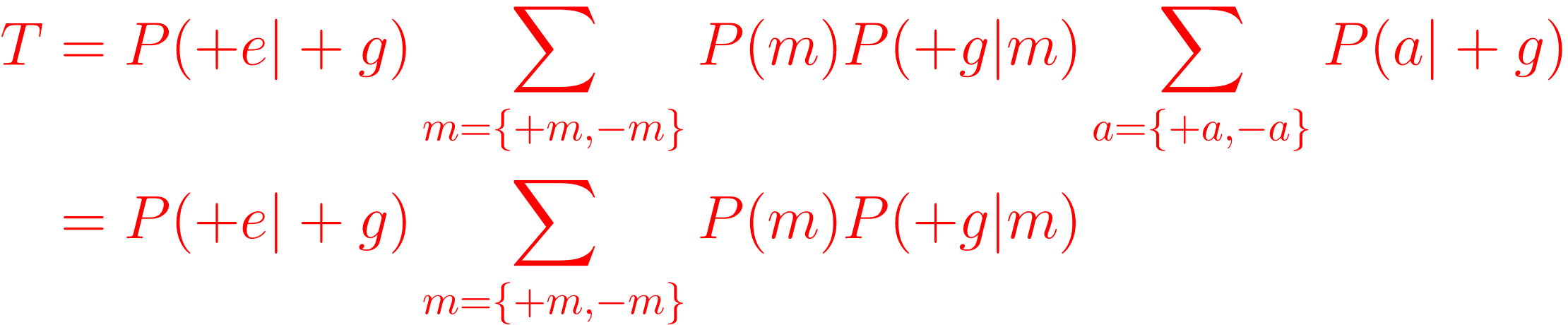
* 1. What does the factor α represent?
  + It is a constant of normalization.
  1. Give the formula to compute the probability P(+e)



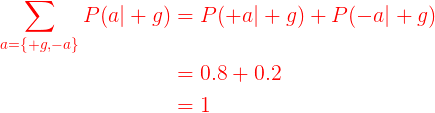
1. Give the formula of the factor T



1. Give the simplified formula of the factor T using the variable elimination technique or any other technique you know.



because



1. Compute the factor T. 