1. Data on leaving certificate subject choices showed that in a certain year, 21% of students took Chemistry but not French, 40% took French but not Chemistry, and 9% took both French and Chemistry. Let C be the event that a randomly chosen student took Chemistry, and F be the event the student took French.

(a) Construct an appropriate 2 × 2 table in percentages. Include margin totals.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **F** | **Fc** | **Sum** |
| **C** | 0.09 | 0.21 | **0.30** |
| **Cc** | 0.40 | 0.30 | **0.70** |
| **Sum** | **0.49** | **0.51** | **1.00** |

(b) Compute P(C ∩ F) and

0.09.

(c) P(C ∪ F) .

0.09 + 0.21 + 0.40 = 0.70.

(d) Are C and F disjoint (mutually exclusive)? How do you know?

No, because C ∩ F is not the null (or empty) set – because its probability value is not zero.

(e) Are C and F independent? How do you know?

No, because P(C ∩ F) = 0.09 P(C) X P(F) = 0.30 X 0.49 = 0.147.

(f) For two randomly chosen students, calculate the probability both take French. (Assume students choose courses independently).

For two independent events, such as these two random ones, P(A ∩ B) = P(A) X P(B).

Therefore, for those 2 students, A and B:

* P(FA ∩ FB) = P(FA) X P(FB) = 0.49 X 0.49 = 0.2401.

(g) For two randomly chosen students, calculate the probability neither takes French

For two independent events, such as these two random ones, P(A ∩ B) = P(A) X P(B).

Therefore, for those 2 students, A and B:

* P(FAc ∩ FBc) = P(FAc) X P(FBc) = (1 - 0.49) X (1- 0.49) = 0.51 X 0.51 = 0.2601.

2. Suppose that A and B are disjoint events for which P(A) = 0.1 and P(B) = 0.2. What is the probability that

Let’s do a 2 x 2 table for the events first:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **B** | **Bc** | **Sum** |
| **A** | 0 | 0.1 | **0.1** |
| **Ac** | 0.2 | 0.7 | **0.9** |
| **Sum** | **0.2** | **0.8** | **1.0** |

(a) either A or B occurs

P(A ∪ B) = 0.1 + 0 + 0.2 = 0.3.

This is the sum of the values in the A row and the B column in the 2 x 2 table above (i.e. in the pink, lavender and blue cells).

(b) A occurs but B does not

P(A ∩ Bc) = 0.1

This is the value in the cell where A intersects with Bc (i.e. in the pink cell).

(c) both A and B occur?

P(A ∩ B) = 0.

This is the value in the cell where the A row intersects with the B column (i.e. in the lavender cell).

In any case, as we’re told that A and B are disjoint events, P(A ∩ B) = 0.

3. 60 percent of drivers have had driver education. During their first year of driving, drivers without driver education have a 0.08 probability of having an accident, but new drivers with driver education have a 0.03 probability of having an accident.

Let’s try to complete some of a 2 x 2 table (its lower sum row only) using the immediately useful information that we are given above, and then use the General Multiplication Rule (P(X∩Y) = P(Y)\*P(X|Y) to complete the remaining cells:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **E** | **Ec** | **Sum** |
| **A** | 0.018  (0.6 x 0.03) | 0.032  (0.4 x 0.08) | **0.05** |
| **Ac** | 0.582  (0.6 x 0.97) | 0.368  (0.4 x 0.92) | **0.95** |
| **Sum** | **0.6** | **0.4** | **1.0** |

* A = Accident;
* E = Education.

What is the probability a new driver with no accident in the first year had driver education?

We are now being asked for P(E|Ac) = P(E∩Ac) / P(Ac) = 0.582/0.95 = 0.613.

4. A classification algorithm classifies 1000 objects in to one of two classes. It incorrectly classifies 15 out of 100 class 1 objects and 51 class 2 objects.

**C1**

**S**

**C2**

**A1**

**E1**

**E2**

**A2**

P(C1)

P(C2)

P(A1|C1)

P(A1C1)

P(E1|C1)

P(E2|C2)

P(A2|C2)

P(E1C1)

P(E2C2)

P(A2C2)

In this tree diagram:

* S=Sample Space
* C = Class
* A = Accurate
* E = Error

We will now use the information provided in the problem text to solve the expression in this diagram.

|  |  |  |
| --- | --- | --- |
| **Count/Probability** | **Derivation** | **Value** |
| S | Given in problem text | 1000 |
| **C1** | **Given in problem text** | **100** |
| C2 | S - C1 | 900 |
| P(C1) | C1/S | 0.1 |
| P(C2) | C2/S | 0.9 |
| E1 | Given in problem text | 15 |
| **A1** | **100 - E1** | **85** |
| E2 | Given in problem text | 51 |
| A2 | 900 - E2 | 849 |
| P(A1|C1) | A1/C1 | 0.85 |
| P(E1|C1) | E1/C1 | 0.15 |
| P(E2|C2) | E2/C2 | 0.0567 |
| P(A2|C2) | A2/C2 | 0.9433 |
| P(A1C1) | P(A1|C1)\*P(C1) | 0.085 |
| **P(E1C1)** | **P(E1|C1)\*P(C1)** | **0.015** |
| **P(E2C2)** | **P(E2|C2)\*P(C2)** | **0.051** |
| P(A2C2) | P(A2|C2)\*P(C2) | 0.849 |

(a) What is the overall error rate?

The error rates for both C1 and C2 are highlighted in the 2 bold red rows in the table above. The overall error rate is their sum: 0.015 + 0.051 = 0.066.

(b) What proportion of those predicted to belong to class 1 are correctly classified?

As highlighted in the 2 bold blue rows in the table above, as 85 of the 100 objects that were classified in C1, the proportion of correctly classified objects is 85%.

(c) A competing algorithm run on the same data, classifies 200 objects as class 1, 50% of which are correct. Which algorithm is better?

If we recalculate the error probability for C1 using the new data provided above, we get:

* P(E1C1) = P(E1|C1)\*P(C1) = 0.5\*(200/1000) = 0.5\*0.2 = 0.1.

As the competing algorithm’s error rate for only C1 (0.1) is much higher than the overall error rate (0.066) for the original algorithm, we can conclude that the original algorithm is better (more accurate).

5. Events A, B and C have probabilities 0.1, 0.3 and 0.2 respectively. Also A and B are mutually exclusive (disjoint), A and C are independent, and B and C are independent. Find the probability that exactly one of the events A, B, and C occur.

P() = P(A)P(C) = 0.1 X 0.2 = 0.02 (because A and C are independent).

P() = P(B)P(C) = 0.3 X 0.2 = 0.06 (because B and C are independent).

This enables us to complete the following Venn diagram:

S

0.06

0.02

0.40

A

B

C

From this diagram we can see that the probability of only event A occurring is 0.08, event B only is 0.24, and event C only is 0.12.

Therefore, the total probability of only one of events A, B or C occurring is 0.08 + 0.24 + 0.12 = 0.44.