Statistics for Computing MA4413

Lecture 4

Conditional Probability, Law of Total Probability and Bayes' Rule

Kevin Burke

kevin.burke@ul.ie

Conditional Probability

In the previous lecture we encountered **conditional probability**, i.e.,

$$\Pr(A \mid B)$$

which represents an *updated* probability of A given the *information* that B has occurred.

Good decisions should be based on using the information at hand.

Conditional Probability

Let's assume that we have some idea about the probability of a bug in some software, for example, $\Pr(\text{bug}) = 0.1$.

We are then given *new information*: the code was written by an inexperienced programmer.

How would we *update* our *prior* probability of $\Pr(\text{bug}) = 0.1$?

$$\Pr(\text{bug} \,|\, \text{inexperienced}) < 0.1?$$

$$\Pr(\text{bug} \,|\, \text{inexperienced}) = 0.1?$$

$$\Pr(\text{bug} \,|\, \text{inexperienced}) > 0.1?$$

Conditional Probability

It should be quite clear that "bug in software" and "inexperienced programmer" are *dependent events*.

Furthermore, it is reasonable to expect $\Pr(\text{bug} \,|\, \text{inexperienced}) > 0.1$.

What if we had been told that the programmer has brown hair instead?

We might imagine that programming ability and hair colour are *independent events* so that $\Pr(\text{bug} \,|\, \text{brown hair}) = \Pr(\text{bug}) = 0.1$.

Can you think of an event that might lead to $\Pr(\text{bug} \,|\, \text{event}) < 0.1$?

Conditional Probability Formula

Recall from the previous lecture (multiplication rule) that

$$\Pr(B)\Pr(A \mid B) = \Pr(A \cap B).$$

Dividing both sides by $\Pr(B)$ gives us a formula for calculating **conditional probability**:

$$\boxed{\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}}.$$

Conditional Probability Formula

We get an expression for $\Pr(B \mid A)$ by swapping the letters around:

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{\Pr(A \cap B)}{\Pr(A)}$$

since $\Pr(B \cap A) = \Pr(A \cap B)$.

Note that $\Pr(A \mid B)$ and $\Pr(B \mid A)$ are *not* the same thing.

Example: Flipping a Coin and Rolling a Die

In the previous lecture we dealt with the experiment of flipping a coin and rolling a die.

We had the events

- $A =$ "head & even number".
- $B =$ "head & any number" $=$ "the coin shows a head".
- $C =$ "any face & a five" $=$ "the die shows a five".

We calculated

- $\Pr(A) = \frac{1}{4}$, $\Pr(B) = \frac{1}{2}$ and $\Pr(C) = \frac{1}{6}$.
- $\Pr(A \cap B) = \frac{1}{4}$, $\Pr(A \cap C) = 0$ and $\Pr(B \cap C) = \frac{1}{12}$.

We also worked out which events were independent and dependent but now, using the formula for *conditional probability*, we can go further.

Example: Flipping a Coin and Rolling a Die

Consider

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{2}{1} \times \frac{1}{4} = \frac{2}{4} = \frac{1}{2}.$$

Thus,

- $\Pr(A) = \Pr(\text{“head \& even”}) = \frac{1}{4} = 0.25.$
- $\Pr(A \mid B) = \Pr(\text{“head \& even”} \mid \text{“coin shows head”}) = \frac{1}{2} = 0.5.$

This result makes intuitive sense: since we now know that the coin shows a head, we are more sure about the result being “head & even”.

We *update* the *prior* probability using the current *information*.

Example: Flipping a Coin and Rolling a Die

A more detailed explanation of the previous result:

Without any information, there are twelve possible outcomes $\{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$.

Three of these (i.e., 25%) fall into the category “head & even”.

Given the information that the coin shows a head, there are now only six possible outcomes $\{H1, H2, H3, H4, H5, H6\}$.

Three of these (i.e., 50%) fall into the category “head & even”.

Example: Flipping a Coin and Rolling a Die

Let's now consider

$$\Pr(A \mid C) = \frac{\Pr(A \cap C)}{\Pr(C)} = \frac{0}{\frac{1}{6}} = \frac{6}{1} \times 0 = 0.$$

Thus,

- $\Pr(A) = \Pr(\text{“head \& even”}) = \frac{1}{4} = 0.25.$
- $\Pr(A \mid C) = \Pr(\text{“head \& even”} \mid \text{“dies shows a five”}) = 0.$

Given the information that the die shows a five, we know that the result cannot be "head & even" $\Rightarrow$ our *updated probability* is $\Pr(A \mid C) = 0$.

Question 1

Continue with the previous example of flipping a coin and rolling a die.

a) Calculate $\Pr(B \mid A)$, $\Pr(C \mid A)$, $\Pr(B \mid C)$ and $\Pr(C \mid B)$.

b) Compare the above *updated* probabilities with the relevant *prior* probabilities. Give a brief explanation of the results.

c) Which events are independent?

d) Which events are mutually exclusive?

Question 2 (questions on next slide)

A software company examined blocks of code written by its employees. Each block of code was tested for bugs and, in addition, the skill level of the employee was also recorded. See table below.

		Skill Level			
		High	Average	Low	Total
Bug in Code	No	140	600	100	840
	Yes	5	70	40	115
	Total	145	670	140	955

We will let $B =$ "bug" and, hence, $B^c =$ "no bug".

Also let $S_H =$ "skill: high", $S_A =$ "skill: average" and $S_L =$ "skill: low".

Note that $S_H \cap B^c$ contains 140 cases, $S_H \cap B$ contains 5 cases, H contains 145 cases etc.

Question 2

Note: Use the appropriate probability notation.

a) Calculate the probability that the programmer has: (i) high skill, (ii) average skill and (iii) low skill.

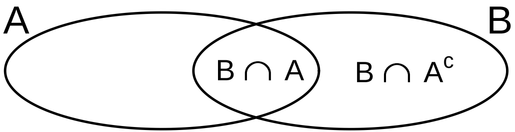
b) Calculate the probability of a bug.

c) Calculate the probability of a bug *given that* the code was written by a programmer with: (i) high skill, (ii) average skill and (iii) low skill.

d) Comment on the above conditional (i.e., updated) probabilities compared with $\Pr(B)$ calculated in part (b). Is the presence of bugs independent of the skill level?

e) Show that $\Pr(S_A \mid B) > \Pr(S_L \mid B)$. Explain the reason for this.

Law of Total Probability



From the above we can see that

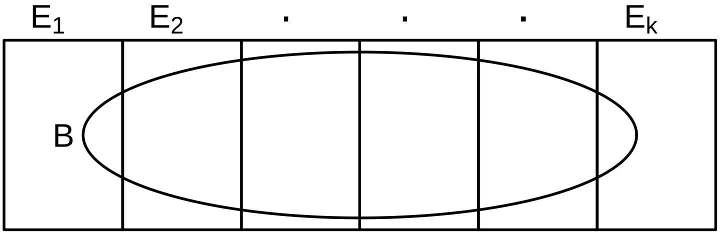
$$\Pr(B) = \Pr(B \cap A) + \Pr(B \cap A^c).$$

Also $\Pr(B \cap A) = \Pr(A)\Pr(B \mid A)$ and $\Pr(B \cap A^c) = \Pr(A^c)\Pr(B \mid A^c)$ using the *multiplication rule*.

$$\Rightarrow \boxed{\Pr(B) = \Pr(A)\Pr(B \mid A) + \Pr(A^c)\Pr(B \mid A^c)}.$$

This is the simplest example of the **law of total probability**.

Law of Total Probability



If there are k *mutually exclusive* and *exhaustive* events, $E_1, E_2, \ldots, E_k$, then the **law of total probability** is

$$\boxed{\begin{aligned} \Pr(B) &= \Pr(B \cap E_1) + \Pr(B \cap E_2) + \cdots + \Pr(B \cap E_k) \\ &= \Pr(E_1)\Pr(B \mid E_1) + \Pr(E_2)\Pr(B \mid E_2) + \cdots + \Pr(E_k)\Pr(B \mid E_k) \end{aligned}}.$$

Example: Internet Trolls

"Troll" is the slang word used to describe an internet user who aims to irritate other users - often through the medium of comment posting.

Let's assume that 10% of internet users can be classed as trolls and that 99% of their comments are irritating. Normal internet users (i.e., non-trolls) only post irritating comments 4% of the time.

If we let T = "the user is a troll" and I = "the comment is irritating" then from the information above we have:

$$\Pr(T) = 0.1 \qquad \Pr(I \mid T) = 0.99$$

$$\Pr(T^c) = 0.9 \qquad \Pr(I \mid T^c) = 0.04$$

Example: Internet Trolls

Using the *law of total probability* we can work out the probability that an irritating comment is posted:

$$
\begin{aligned}
\Pr(I) = \Pr(I \cap T) + \Pr(I \cap T^c) &= \Pr(T)\Pr(I \mid T) + \Pr(T^c)\Pr(I \mid T^c) \\
&= 0.1(0.99) + 0.9(0.04) \\
&= 0.099 + 0.036 \\
&= 0.135.
\end{aligned}
$$

Thus, 13.5% of comments are irritating.

Question 3

Return to the example used in Question 2.

a) Calculate $\Pr(B)$ using the law of total probability and previously calculated values for $\Pr(B \cap S_H)$, $\Pr(B \cap S_A)$ and $\Pr(B \cap S_L)$.

b) Calculate $\Pr(S_H)$, $\Pr(S_A)$ and $\Pr(S_L)$ using similar means.

(note: we previously calculated the *total* probabilities directly from the table)

Bayes' Rule

Bayes' Rule provides a method for assessing the likelihood of an event E_1 *given* current information B.

Bayes' Rule is derived simply by combining the *conditional probability* formula and the *multiplication rule*:

$$\boxed{\Pr(E_1 \,|\, B) = \frac{\Pr(E_1 \cap B)}{\Pr(B)} = \frac{\Pr(E_1)\Pr(B \,|\, E_1)}{\Pr(B)}},$$

where, typically, $\Pr(B)$ is calculated using the *law of total probability*, i.e., $\Pr(B) = \Pr(B \cap E_1) + \cdots + \Pr(B \cap E_k)$.

Example: Internet Troll Detection

We know from earlier that 13.5% of all comments online are irritating, i.e., $\Pr(I) = 0.135$.

We take a look at our blog and notice an irritating comment - was this the work of a troll?

In other words, *given that the message is irritating* what is the probability of a troll having posted it?

$$\begin{aligned}\Pr(T \mid I) = \frac{\Pr(T)\Pr(I \mid T)}{\Pr(I)} &= \frac{0.1(0.99)}{0.135} \\ &= \frac{0.099}{0.135} \\ &\approx 0.73.\end{aligned}$$

There is a 73% chance that this message was left by an internet troll.

Question 4

A manufacturer of laptops sources processors from three companies: A_1, A_2 and A_3. Specifically, 20% of stock comes from A_1, 55% comes from A_2 and the remainder comes from A_3. Assume that 10% of A_1's stock is defective, 4% of A_2's and 1% of A_3's.

a) What is the probability that a defective processor is installed?

b) A customer comes back with a faulty laptop - we determine that the processor is the issue. Which company is most likely to have produced this processor?

c) What is the probability that a processor from A_1 works *correctly*?

d) Given that the processor is working correctly, what is the probability it came from company A_1?

e) If *all* stock came from A_3, what would $\Pr(D)$ be? What would $\Pr(A_3 \mid D)$ be? What would $\Pr(A_1 \mid D)$ be?