# 7 Posterior Probability and Bayes

**Examples:**

1. In a computer installation, 60% of programs are written in $C^{++}$ and 40% in Java. $60\%$ of the programs written in $C^{++}$ compile on the first run and $80\%$ of the Java programs compile on the first run.

   (a) What is the overall proportion of programs that compile on first run?

   (b) If a randomly selected program compiles on the first run what is the probability that it was written in $C^{++}$?

2. In a certain company

   50% of documents are written in WORD; 30% in LATEX; 20% in HTML.
   From past experience it is know that :
   40% of the WORD documents exceed 10 pages
   20% of the LATEX documents exceed 10 pages
   20% of the HTML exceed 10 pages

   (a) What is the overall proportion of documents containing more than 10 pages?

   (b) A document is chosen at random and found to to have more than 10 pages. What is the probability that it has been written in LATEX?

3. Enquiries to an on-line computer system arrive on 5 communication lines. The percentage of messages received through each line are:

| Line | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| % received | 20 | 30 | 10 | 15 | 25 |

From past experience, it is known that the percentage of messages exceeding 100 characters on the different lines are:

| Line | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| % exceeding 100 characters | 40 | 60 | 20 | 80 | 90 |

(a) Calculate the overall proportion of messages exceeding 100 characters.

(b) If a message chosen at random is found to exceed 100 characters, what is the probability that it came through line 5?

4. A binary communication channel carries data as one of two sets of signals denoted by 0 and 1. Owing to noise, a transmitted 0 is sometimes received as a 1, and a transmitted 1 is sometimes received as a 0. For a given channel, it can be assumed that a transmitted 0 is correctly received with probability $0.95$ and a transmitted 1 is correctly received with probability 0.75. Also, 60$\%$ of all messages are transmitted as a 0. If a signal is sent, determine the probability that:

(a) a 1 was received;

(b) a 0 was received;

(c) an error occurred;

(d) a 1 was transmitted given than a 1 was received;

(e) a 0 was transmitted given that a 0 was received.

**Example 1**

| | Compiles on first run | Does not compile on first run | |
|---|---|---|---|
| $C^{++}$ | 72 | 48 | 120 |
| Java | 64 | 16 | 80 |
| | 136 | 64 | 200 |

The probability that a program has been written in $C^{++}$ when we know that it has compiled in the first run?

If E is the event of compiling on the first run, we seek:

$$P(C^{++}|E) =?$$

Now

$$P(C^{++} \cap E) = P(C^{++})P(E|C^{++})$$

and

$$P(E \cap C^{++}) = P(E)P(E|C^{++})$$

Since

$$P(E \cap C^{++}) = P(C^{++} \cap E)$$

Then that

$$P(E)P(C^{++}|E) = P(C^{++})P(E|C^{++})$$

So

$$P(C^{++}|E) = \frac{P(C^{++})P(E|C^{++})}{P(E)}$$

P($C^{++}|E$) is the $Posterior\ Probability$ of $C^{++}$ **after** it has been found that the program has compiled on the first run.

$$P(C^{++}|E) = \frac{P(C^{++})P(E|C^{++})}{P(E)}$$

$$\begin{aligned} P(E) &= P(C^{++})P(E|C^{++}) + P(J)P(E|J) \\ &= 120/200 * 72/120 + 80/200 * 64/80 + = 0.68 \end{aligned}$$

Then

$$P(C^{++}|E) = \frac{120/200 * 72/120}{.68}$$

In $R$ :

```
totalprob <-((80/200)*(64/80))+((120/200)*(72/120)) # total probability
conditc <-(120/200)*(72/120)/totalprob #posterior probability of C++
conditc
[1] 0.5294118
```

Analogously,

$$P(J|E) = \frac{P(J)P(E|J)}{P(C^{++})P(E|C^{++}) + P(J)P(E|J)}$$

In $R$

```
totalprob <-((80/200)*(64/80))+((120/200)*(72/120)) # total probability
conditjava <-(80/200)*(64/80)/totalprob #posterior probability of Java
conditjava
[1] 0.4705882
```

**Prior and Posterior Probabilities after a program has been found to have compiled on first run**

| | $C^{++}$ | Java |
|---|---|---|
| Prior | 0.60 | 0.40 |
| Posterior | 0.53 | 0.47 |

**Bayes' Theorem**

## Bayes' Rule

### Bayes' rule for two events

Consider two events $A$ and $B$.

$$P(A \cap B) = P(A)P(B|A)$$

and

$$P(B \cap A) = P(B)P(A|B)$$

Since

$$P(A \cap B) = P(B \cap A),$$

It follows from the multiplication law that

$$P(B)P(A|B) = P(A)P(B|A)$$

which implies that

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

$P(A|B)$ is **Posterior Probability** of A **after** B has occurred.

**Example 4**

A binary communication channel carries data as one of two sets of signals denoted by 0 and 1. Owing to noise, a transmitted 0 is sometimes received as a 1, and a transmitted 1 is sometimes received as a 0.

For a given channel, it can be assumed that a transmitted 0 is correctly received with probability $0.95$ and a transmitted 1 is correctly received with probability 0.75. It is also known that 70$\%$ of all messages are transmitted as a 0.

If a signal is sent, what is the probability that

1. a 1 was transmitted given that a 1 was received;

2. a 0 was transmitted given that a 0 was received.

**Solution**

Let:
$R_0$ is event that a zero is received.
$T_0$ is event that a zero is transmitted, $P(T_0 = .7)$

$R_1$ is event that a one is received.
$T_1$ is the event that a one is transmitted, $P(T_1 = .3)$

$$P(T_0|R_0) = \frac{P(T_0)P(R_0|T_0)}{P(R_0)}$$

Now:

$$P(T_0)P(R_0|T_0) = .7 \times .95$$

and

$$R_0 = (T_0 \cap R_0) \cup (T_1 \cap R_0)$$

So

$$P(R_0) = P(T_0)P(R_0|T_0) + P(T_1)P(R_0|T_1)$$

Therefore

$$P(R_0) = .7 \times .95 + .3 \times .25 = .74$$

$$P(T_0|R_0) = \frac{.7 \times .95}{.7 \times .95 + .3 \times .25} = \frac{.665}{.74} = .90$$

Similarly

$$P(T_1|R_0) = \frac{.3 \times .25}{.7 \times .95 + .3 \times .25} = \frac{.075}{.74} = .10$$

**Example**

From past experience it is know that :

50% of documents are written in WORD;
30% in LATEX;
20% in HTML.

These are the *prior probabilities*

- $P(Word) = .5;$
- $P(Latex) = .3;$
- $P(Html) = .2.$

40% of the WORD documents exceed 10 pages
20% of the LATEX documents exceed 10 pages
20% of the HTML exceed 10 pages
A a document chosen at random was found to exceed 10 pages, what the probability is that it has been written in Latex?

Let $E$ be the event that a document, chosen at random, contains more than 10 pages.

$$P(Latex|E) = ?$$

$$P(Latex|E) = \frac{P(Latex)P(E|Latex)}{P(E)}$$

Now

$$\begin{aligned} P(E) &= P(Word)P(E|Word) + P(Latex)P(E|Latex) + P(Html)P(E|Html) \\ &= 0.5 \times 0.4 + 0.3 \times 0.2 + 0.2 \times 0.2 = .3 \end{aligned}$$

So:

$$\begin{aligned} P(Latex|E) &= \frac{P(Latex)P(E|Latex)}{P(E)} \\ &= \frac{0.3 \times 0.2}{.3} = .2 \end{aligned}$$

Similarly

$$P(Word|E) = \frac{.5 \times .4}{.3} = .67$$

and

$$P(Html|E) = \frac{.2 \times .2}{.3} = .13$$

**Prior and Posterior Probabilities after document has been found to contain more than 10 pages**

| | Word | Latex | Html |
|---|---|---|---|
| Prior | 0.50 | 0.30 | 0.20 |
| Posterior | 0.67 | 0.20 | 0.13 |

## General Bayes Rule

If a sample space can be partitioned into $k$ mutually exclusive and exhaustive events:

$$A_1, A_2, A_3, \cdots A_k$$

$$S = A_1 \cup A_2 \cup A_3 \cdots \cup A_k$$

Then for any event E,:

$$P(E) = P(A_1)P(E|A_1) + P(A_2)P(E|A_2) \cdots P(A_k)P(E|A_k)$$

$$P(A_i|E) = \frac{P(A_i)P(E|A_i)}{P(E)}$$

**Proof:** For any $i,\ 1 \leq i \leq k$

$$\begin{aligned} E \cap A_i &= A_i \cap E \\ P(E)P(A_i|E) &= P(A_i)P(E|A_i) \\ P(A_i|E) &= \frac{P(A_i)P(E|A_i)}{P(E)} \end{aligned}$$

$P(A_i|E)$ is called the **POSTERIOR PROBABILITY**

## SOME APPLICATIONS OF BAYES

- Game Show: “Let’s make a Deal

- Hardware Fault Diagnosis

- Machine Learning

- Machine Translation

**Game Show**:

You are a contestant in a game which allows you to choose one out of three doors. One of these doors conceals a laptop computer while the other two are empty. When you have made your choice, the host of the show opens one of the remaining doors, and shows you that it is one of the empty ones. You are now given the opportunity to change your door. Should you do so, or should you stick with your original choice?

The question we address is: If you change, what are your chances of winning the laptop?

When you choose a door randomly originally, there is a probability of 1/3 that the laptop will be behind the door that you choose, and a probability of 2/3 that it will be behind one of the other two doors. These probabilities do not change when a door is opened and revealed empty; therefore if you switch you will have a probability of 2/3 of winning the laptop.

Let $D_1$, $D_2$, and $D_3$ be the events that the laptop is behind door 1, door 2 and door 3 respectively. We assume that

$$P(D_1) = P(D_2) = P(D_3) = \frac{1}{3}$$

and

$$P(\overline{D}_1) = P(\overline{D}_2) = P(\overline{D}_3) = \frac{2}{3}$$

Suppose you choose door 1.

Let $E$ be the event that the host opens door 2, and reveals it to be empty. If you change after this, you will be choosing door 3. So we need to know $P(D_3|E)$. From Bayes' rule

$$P(D_3|E) = \frac{P(D_3)P(E|D_3)}{P(E)}$$

Now, from total probability,

$$E = (D_1 \cap E) \cup (D_2 \cap E) \cup (D_3 \cap E)$$

So

$$\begin{aligned} P(E) &= P(D_1 \cap E) + P(D_2 \cap E) + P(D_3 \cap E) \\ &= P(D_1)P(E|D_1) + P(D_2)P(E|D_2) + P(D_3)P(E|D_3) \end{aligned}$$

So we can write

$$P(D_3|E) = \frac{P(D_3)P(E|D_3)}{P(D_1)P(E|D_1) + P(D_2)P(E|D_2) + P(D_3)P(E|D_3)} \tag{1}$$

We need to calculate $P(E|D_1)$, $P(E|D_2)$ and $P(E|D_3)$.

$P(E|D_1)$ is the probability that the host opens door 2 and reveals it to be empty, given that it is behind door 1. When the laptop is behind door 1, the door selected by you, the host has two doors to select from. We assume that the host selects one of these at random, i.e. $P(E|D_1) = 1/2$.

$P(E|D_2)$ is the probability that the host opens door 2 and reveals it to be empty, given that it is behind door 2. This is an impossibility, so it has a probability of zero, i.e. $P(E|D_2) = 0$.

Finally $P(E|D_3)$ is the probability that the host opens door 2 and reveals it to be empty, given that it behind door 3. When the laptop is behind door 3, there has just one way of revealing an empty door; the host must open door 2. This is a certainty, so it has a probability of one, i.e. $P(E|D_3) = 1$.

Putting these values into (1) we have

$$P(D_3|E) = \frac{\frac{1}{3} \times 1}{(\frac{1}{3} \times \frac{1}{2}) + (\frac{1}{3} \times 0) + (\frac{1}{3} \times 1)} = \frac{2}{3}$$

So changing would increase your probability of winning the laptop from $1/3$ to $2/3$.

You would double your chance by changing.

The prior probability of winning the laptop was $1/3$, while the posterior probability that the laptop is behind the door you have not chosen is $2/3$

## Hardware Fault Diagnosis

Printer failures are associated with three types of problems; hardware, software and electrical connections. A printer manufacturer obtained the following probabilities from a database of tests results.

$$P(H) = 0.1 \quad P(S) = 0.6 \quad P(E) = 0.3$$

where $H$, $S$, and $E$ denote hardware, software or electrical problems respectively:
These are the prior probabilities:

It is also known that the probability of a printer failure ($F$)

- given a hardware problem is 0.9, $\quad P(F|H) = 0.9$
- given a software problem is 0.2, $\quad P(F|S) = 0.2$
- given an electrical problem is 0.5, $\quad P(F|E) = 0.5$

These are the conditional probabilities

If a customer reports a printer failure, what is the most likely cause of the problem?

We need the posterior probabilities

$$P(H|F), \quad P(S|F), \quad P(E|F),$$

the likelihood that, given that a printer fault has been reported, it will be due to faulty hardware, software or electrical.

We calculate the posterior probabilities using Bayes' rule.

$$P(H|F), \quad P(S|F), \quad P(E|F)$$

First calculate the total probability of a failure occurring.

$$F = (H \cap F) \cup (S \cap F) \cup (E \cap F)$$

$$\begin{aligned} P(F) &= P(H)P(F|H) + P(S)P(F|S) + P(E)P(F|E) \\ &= .1 \times .9 + .6 \times .2 + .3 \times .5 = .36 \end{aligned}$$

Then the posterior probabilities are:

$$\begin{aligned} P(H|F) &= \frac{P(H)P(F|H)}{P(F)} = \frac{.1 \times .9}{.36} = .250 \\ P(S|F) &= \frac{P(S)P(F|S)}{P(F)} = \frac{.6 \times .2}{.36} = .333 \\ P(E|F) &= \frac{P(E)P(F|E)}{P(F)} = \frac{.3 \times .5}{.36} = .417 \end{aligned}$$

Because $P(E|F)$ is the largest, the most likely cause of the problem is electrical.

A help desk or website dialog to diagnose the problem should check into this problem first.

## Machine Learning

Bayes' theorem is sometimes used to improve the accuracy in supervised learning.

It is used in the classification of items where the system has already learnt the probabilities.

We need to classify $f$, a new example, into the correct class. Suppose there are only two classes, $y = 1$ and $y = 2$ into which we can classify new values of $f$.

By Bayes' rule, we can write

$$P(y = 1|f) = \frac{P(y = 1 \cap f)}{P(f)} = \frac{P(y = 1)P(f|y = 1)}{P(f)} \quad (2)$$

$$P(y = 2|f) = \frac{P(y = 2 \cap f)}{P(f)} = \frac{P(y = 2)P(f|y = 2)}{P(f)} \quad (3)$$

Dividing (2) by (3) we get

$$\frac{P(y = 1|f)}{P(y = 2|f)} = \frac{P(y = 1)P(f|y = 1)}{P(y = 2)P(f|y = 2)} \quad (4)$$

Our decision is to classify a new example into class 1 if

$$\frac{P(y = 1|f)}{P(y = 2|f)} > 1$$

or equivalently if

$$\frac{P(y = 1)P(f|y = 1)}{P(y = 2)P(f|y = 2)} > 1$$

$f$ goes into class 1 iff

$$P(y = 1)P(f|y = 1) > P(y = 2)P(f|y = 2)$$

and $f$ goes into class 2 if

$$P(y = 1)P(f|y = 1) < P(y = 2)P(f|y = 2)$$

The prior probabilities of being in either of two classes are assumed to be already learnt:

$$P(y = 1) = .4 \quad P(y = 2) = .6$$

Also the conditional probabilities for the new example $f$ are also aready learnt:

$$P(f|y = 1) = .5 \quad P(f|y = 2) = .3$$

Into what class should you classify the new example?

$$P(y = 1)P(f|y = 1)) = .4 \times .5 = .20$$

and

$$P(y = 2)P(f|y = 2) = .6 \times .3 = .18$$

and since

$$P(y = 1)P(f|y = 1) > P(y = 2)P(f|y = 2),$$

the new example goes into class 1.

## Machine Translation

A string of English words $e$ can be translated into a string of French words $f$ in many different ways. Often, knowing the broader context in which $\mathbf{e}$ occurs may narrow the set of acceptable French translations but, even so, many acceptable translations remain.

Bayes' rule can be used to make a choice between them.

To every pair $(e, f)$ a number $P(f|e)$ is assigned which is interpreted as the probability that a translator when given $e$ to translate will return $f$ as the translation.

Given a string of French words $f$, the job of the translation system is to find the string $e$ that the native speaker had in mind, when $f$ was produced. We seek $P(\mathbf{e}|\mathbf{f})$.

From Bayes' rule

$$P(e|f) = \frac{P(e)P(f|e)}{P(f)}$$

We seek that $\hat{e}$ to maximise $P(e|f)$, i.e. choose $\hat{e}$ so that

$$P(\hat{e}|f) = max_e \frac{P(e)P(f|e)}{P(f)}$$

which is equivalent to $max_e\,(P(e)P(f|e))$ since P($f$) is constant regardless of $e$. This is known as the *Fundamental Equation of Machine Translation* usually written as

$$\hat{e} = argmax P(e)P(f|e)$$

As a simple example, suppose there are 3 possibilities for $e$:

$$P(e_1) = 0.2,\ P(e_2), = 0.5,\ P(e_3) = 0.3.$$

One of these has been translated into $f$. A passage $f$ has the following conditional probabilities:

$$P(f|e_1) = 0.4,\ P(f|e_2) = 0.2, \quad P(f|\mathbf{e}_3) = 0.4.$$

We can use Bayes' theorem to decide which is the most likely $e$.

$$\begin{aligned} P(e_1)P(f|e_1) &= .2 \times .4 = .08 \\ P(e_2)P(f|e_2) &= .5 \times .2 = .1 \\ P(e_3)P(f|e_3) &= .3 \times .4 = .12 \end{aligned}$$

The best translation, in this case, is $e_3$ since it has the highest posterior probability.