Applications of Bayes’ Theorem

**Example:** There are 10 urns, 9 of which are of type I and 1 of type II. Urn of type I carries 2 white balls and 2 black balls. Urn of type II carries 5 white balls and 1 black ball.

If a ball drawn randomly from a randomly chosen urn turns out to be white, then what is the probability that the chosen urn is of type II? This is a model of an inference problem.

**Solution**

$A :=$ ball drawn is white

$B_1 :=$ urn is of type I

$B_2 :=$ urn is of type II

$B_1$ and $B_2$ are disjoint events and define a partition $\Omega = B_1 \cup B_2$.

\[
P(B_2|A) = \frac{P(A|B_2) \cdot P(B_2)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)} = \frac{5/6 \cdot 1/10}{1/2 \cdot 9/10 + 5/6 \cdot 1/10}
\]

\[
= \frac{5}{27 + 5} = \frac{5}{32}
\]
**Statistical Independence:** The idea that two phenomena have nothing to do with each other has a key role in probability theory.

**Definition** We say that in an experiment $E$, two events $A$ and $B$ are *statistically independent* if,

\[ P(A \cap B) = P(A) \cdot P(B) \]

Imagine a long series of trials, each of which involves carrying out two experiments $E_1$ and $E_2$, where only $E_1$ leads to $A_1$ and only $E_2$ leads to $A_2$.

If $n = \text{total number of trials}$, $n(A_1 \cap A_2) = \text{number of trials leading to occurrence of } A_1 \text{ and } A_2$, then

\[
P(A_1 \cap A_2) \sim \frac{n(A_1 \cap A_2)}{n}
\]
\[
P(A_2) \sim \frac{n(A_2)}{n}
\]
\[
P(A_1) \sim \frac{n(A_1)}{n}.
\]

On the other hand

\[
P(A_1 \cap A_2) \sim \frac{n(A_1 \cap A_2)}{n} = \frac{n(A_1 \cap A_2)}{n} \cdot \frac{n(A_2)}{n} \sim P(A_1) \cdot P(A_2)
\]

The following example illustrates statistical independence and related subtleties. Throw two dice resulting in the outcomes $(X,Y)$.

Let $A_1 : \text{event that } X \text{ is odd}$

$A_2 : \text{event that } Y \text{ is odd}$

$A_3 : \text{event that } X + Y \text{ is odd}$. 

2
Clearly, $A_1$ and $A_2$ are independent.

\[ P(A_1) = \frac{1}{2} = P(A_2) \]

\[ P(A_3) = \frac{1}{2} \cdot \text{Prob} \{ X \text{ odd and } Y \text{ even} \} + \frac{1}{2} \cdot \text{Prob} \{ X \text{ even and } Y \text{ odd} \} \]

\[ = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \]

\[ = \frac{1}{2} \]

\[ P(A_3 | A_1) = \text{Prob} \{ Y \text{ even} \} = \frac{1}{2} \]

\[ P(A_3 | A_2) = \text{Prob} \{ X \text{ even} \} = \frac{1}{2} \]

\[ \Rightarrow P(A_3 | A_1) = P(A_3) = P(A_3 | A_2) \]

Thus $A_3$ and $A_1$ are independent and $A_3$ and $A_2$ are independent. \qed

**Definition:** Given events $A_1, A_2, \ldots, A_n$, we say these are **mutually independent** if:

\[ P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \]

\[ P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \]

\[ \vdots \]

\[ P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n). \]

In the previous example, the events $A_1, A_2, A_3$ are *not* mutually independent, even though they are pairwise independent, because

\[ P(A_1 \cap A_2 \cap A_3) = 0 \]

but

\[ P(A_1)P(A_2)P(A_3) = \left(\frac{1}{2}\right)^3. \]
Probability Trees: When an experiment is of a sequential nature, it is often convenient, especially for purposes of calculation, to represent the experiment graphically by a probability tree. It is a rooted tree and the vertices represent outcomes/events of the experiment. The edges are labelled by the conditional probabilities required to descend from a given vertex to an adjacent one. The probability associated with the event corresponding to a vertex is obtained by taking under consideration the product of the probabilities labelling the edges forming the unique path between the vertex, and the root of the tree.

Example: Flipping a coin three times:
Probability trees may also be infinite. We give an example below.

**Example:** Player A flips a fair coin. If the outcome is a head, he wins; if the outcome is a tail, player B flips. If B’s flip is a head, he wins; if not, player A flips the coin again. This process is repeated (*ad infinitum*, if necessary) until somebody wins. What is the probability that A wins?
For the probability tree above, the darkened vertices correspond to the \textit{elementary events} for which $A$ wins. Since the probability represented by each branch of the tree is 1/2, we have:

$$P\{A \text{ wins}\} \quad \text{calculated via sampling with replacement}$$
\[ P\{A_H\} + P\{A_{TH}\} + P\{A_{TTH}\} + \cdots \]
\[ = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots \]
\[ = \frac{1}{2} \left[ 1 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^6 + \cdots \right] \]
\[ = \frac{1}{2} \frac{1}{1 - \left(\frac{1}{2}\right)^2} \]
\[ = \frac{1}{2} \frac{1}{1 - 1/4} \]
\[ = \frac{2}{3} \]

There is a *big advantage for A* to flip first.

**Gambler’s Ruin** − (Application of Total Probability Law)

**Example:** (1) Toss coin. Call correctly, win 1 dollar. Call wrongly, loose 1 dollar.

**Payoff Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Toss Call</th>
<th>Head</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Tail</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 4
Initial Capital = $x$ dollars and $x$ is a positive integer.

**STRATEGY**

PLAY UNTIL EITHER:

- Win $m$ Dollars
- Lose Shirt

(i.e. has a total (RUIN) of $m$ dollars)

**Question:** What is the probability $p(x)$ of ruin?

- $A = RUIN$
- $B_1 = \text{Win first call } = p$
- $B_2 = \text{Lose first call } = (1 - p)$

$$
P(A) = P(A|B_1) P(B_1) + P(A|B_2) \cdot P(B_2)$$

$$
p(x) = p(x + 1) \cdot \frac{1}{2} + p(x - 1) \cdot \frac{1}{2} \quad 1 \leq x \leq m - 1$$

$$
= p(x + 1) \cdot p + p(x - 1) \cdot (1 - p)
$$

B.C.

$$
\begin{cases}
  p(0) &= 1 \\
  p(m) &= 0
\end{cases}
$$

$$
p(x) = C_1 + C_2x \quad \text{is the solution}
$$

$$
C_1 = 1 \quad C_1 + C_2m = 0
$$

Hence:

$$
p(x) = 1 - \frac{x}{m} \quad 0 \leq x \leq m
$$

**If** $p \neq 1/2$ **the solution is not linear**

**Example [Matching]:**

$n$ distinct items to be matched against $n$ distinct cells. What is the probability of at least 1 match?

**Solution:**

$A_k := \text{event that } k^{th} \text{ item is matched (we don’t care about the rest)}$

$$P^{(n)} = \text{Probability of at least 1 match}$$
\[
P(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j=2}^{n} P(A_i \cap A_j) + \sum_{i<j<k=3}^{n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)
\]

\[
P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_m}) = \frac{(n-m)!}{n!}
\]

\[
P_m = \sum_{a_{i_1} < i_2 < \cdots < i_m \leq n} P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_m}) = \binom{n}{m} \frac{(n-m)!}{n!}
\]

\[
P^{(n)} = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} \cdots + (-1)^{n+1} \frac{1}{n!}
\]

**Special Cases:** Number of permutations of \(n\) things in which there is at least 1 match = \(P^{(n)} \cdot n!\).

\[
n = 3 \quad P^{(n)}n! = 6 \times \left(1 - \frac{1}{2} + \frac{1}{6}\right) = 4
\]

\[
n = 4 \quad P^{(n)}n! = 24 \left(1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24}\right) = 15
\]

**Problem:** Given any \(n\) events, \(A_1, A_2, \cdots A_n\) prove that the probability of exactly \(m \leq n\) events occurring is

\[
P = P_m - \binom{m+1}{m} P_{m+1} + \binom{m+2}{m} P_{m+2} \cdots \pm \binom{n}{m} P_n
\]

where

\[
P_k = \sum_{1 \leq i_1 < i_2 \cdots \cdots i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})
\]
Good Example of Bayesian Inference

Sometimes the application of Bayes’ theorem may yield results that appear counter-intuitive.

**Example:** A laboratory test is developed to detect *mononucleosis* (mono, for short). The probability that a person selected at random has mono is 0.005. If a person has mono, 95% of the time he test will be positive. If a person does not have mono, the test will be positive only 4% of the time. These circumstances are described by the *binary channel* shown in **Figure 5**.

![Binary Channel Diagram](image)

What is the probability that a person has mono conditioned on the fact that his test came out positive?

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person has mono</td>
<td>0.005</td>
</tr>
<tr>
<td>Person does not have mono</td>
<td>0.995</td>
</tr>
<tr>
<td>Test is positive</td>
<td>0.95</td>
</tr>
<tr>
<td>Test is negative</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Fig. 5

\[
\begin{align*}
M &= \text{person has mono} \\
T &= \text{positive mono test} \\
prior \text{probabilities} &= \begin{cases} 
    P(M) = 0.005 \\
    P(\bar{M}) = 0.995 
\end{cases} \\
\text{conditional prob.} &= \begin{cases} 
    P(T|M) = 0.95 \\
    P(T|\bar{M}) = 0.04 
\end{cases}
\]
Then, by Bayes’ theorems,

\[ \{P(M|T) = \frac{P(T|M)P(M)}{P(T|M)P(M) + P(T|\bar{M})P(\bar{M})} \]

\[ = \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.04 \times 0.95} \]

\[ = \frac{0.00475}{0.00475 + 0.0398} \]

\[ = \frac{0.04455}{0.107} \]

Thus the test might give rise to too many false alarms. How to improve? Bring down the probability \( P(T|\bar{M}) \) from 0.04. Improve the test.

A useful form of Bayes’ theorem is obtained by conditioning in more than one event.

Let \( H := \) hypothesis (e.g. a disease event),
Let \( E := \) evidence of data (e.g. image data event), and
Let \( C := \) context (e.g. age group). Then,

\[ P(H|E \cap C) = \frac{P(E|H \cap C) \cdot P(H|C)}{P(E|C)} \]

To see this, observe that the r.h.s. above

\[ = \frac{P(E \cap H \cap C)}{P(E \cap C)} \cdot \frac{P(H \cap C)}{P(C)} \cdot \frac{P(C)}{P(E \cap C)} \]

\[ = \frac{P(E \cap H \cap C)}{P(E \cap C)} \]

\[ = \frac{P(H \cap (E \cap C))}{P(E \cap C)} \]

11
\[ = P(H|E \cap C) \]
\[ = l.h.s \]