Discrete Mathematics & Mathematical Reasoning
Chapter 7 (section 7.3): Conditional Probability & Bayes' Theorem

Colin Stirling

Informatics

Slides originally by Kousha Etessami

Reverend Thomas Bayes (1701-1761),
studied logic and theology as an undergraduate student
at the University of Edinburgh from 1719-1722.

Bayes' Theorem

Bayes Theorem

Let A and B be two events from a (countable) sample space Ω, and $P : \Omega \rightarrow [0, 1]$ a probability distribution on Ω, such that $0 < P(A) < 1$, and $P(B) > 0$. Then

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}$$

This may at first look like an obscure equation,
but as we shall see, it is useful....

Proof of Bayes' Theorem:

Let A and B be events such that $0 < P(A) < 1$ and $P(B) > 0$.

By definition, $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. So: $P(A \cap B) = P(A \mid B)P(B)$.

Likewise, $P(B \cap A) = P(B \mid A)P(A)$.

Likewise, $P(B \cap \overline{A}) = P(B \mid \overline{A})P(\overline{A})$. (Note that $P(\overline{A}) > 0$.)

Note that $P(A \mid B)P(B) = P(A \cap B) = P(B \mid A)P(A)$. So,

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

Furthermore,

$$\begin{aligned} P(B) &= P((B \cap A) \cup (B \cap \overline{A})) = P(B \cap A) + P(B \cap \overline{A}) \\ &= P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A}) \end{aligned}$$

So: $$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}. \quad \square$$

Using Bayes' Theorem

Problem: There are two boxes, Box B_1 and Box B_2.

Box B_1 contains 2 red balls and 8 blue balls.
Box B_2 contains 7 red balls and 3 blue balls.

Suppose Jane first randomly chooses one of two boxes B_1 and B_2, with equal probability, $1/2$, of choosing each.

Suppose Jane then randomly picks one ball out of the box she has chosen (without telling you which box she had chosen), and shows you the ball she picked.

Suppose you only see that the ball Jane picked is red.

Question: Given this information, what is the probability that Jane chose box B_1?

Using Bayes' Theorem, continued

Answer: The underlying sample space, Ω, is:

$$\Omega = \{(a, b) \mid a \in \{1, 2\}, b \in \{\text{red}, \text{blue}\}\}$$

Let $F = \{(a, b) \in \Omega \mid a = 1\}$ be the event that box B_1 was chosen. Thus, $\overline{F} = \Omega - F$ is the event that box B_2 was chosen.

Let $E = \{(a, b) \in \Omega \mid b = \text{red}\}$ be the event that a red ball was picked. Thus, $\overline{E}$ is the event that a blue ball was picked.

We are interested in computing the probability $P(F \mid E)$.

We know that $P(E \mid F) = \frac{2}{10}$ and $P(E \mid \overline{F}) = \frac{7}{10}$.
We also know that: $P(F) = 1/2$ and $P(\overline{F}) = 1/2$.
Can we compute $P(F \mid E)$ based on this? Yes, using Bayes'.

Using Bayes' Theorem, continued

Note that, $0 < P(F) < 1$, and $P(E) > 0$.

By Bayes' Theorem:

$$
\begin{aligned}
P(F \mid E) &= \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid \overline{F})P(\overline{F})} \\
&= \frac{(2/10) * (1/2)}{(2/10) * (1/2) + (7/10) * (1/2)} \\
&= \frac{2/20}{2/20 + 7/20} = \frac{2}{9}. \quad \square
\end{aligned}
$$

Note that, without the information that a red ball was picked, the probability that Jane chose Box B_1 is $P(F) = 1/2$.
But given the information, E, that a red ball was picked, the probability becomes much less, changing to $P(F \mid E) = 2/9$.

More on using Bayes' Theorem: Baysian Spam Filters

Problem: Suppose it has been observed empirically that the word "Congratulations" occurs in 1 out of 10 spam emails, but that "Congratulations" only occurs in 1 out of 1000 non-spam emails. Suppose it has also been observed empirically that about 4 out of 10 emails are spam.

In Bayesian Spam Fitering, these empirical probabilities are interpreted as genuine probabilities in order to help estimate the probability that a incoming email is spam.

Suppose we get a new email that contains "Congratulations".

Let C be the event that a new email contains "Congratulations".
Let S be the event that a new email is spam.

We have observed C. We want to know $P(S \mid C)$.

Bayesian spam filtering example, continued

Bayesian solution: By Bayes' Theorem:

$$P(S \mid C) = \frac{P(C \mid S)P(S)}{P(C \mid S)P(S) + P(C \mid \overline{S})P(\overline{S})}$$

From the "empirical probabilities", we get the estimates:

$P(C \mid S) \approx 1/10; \quad P(C \mid \overline{S}) \approx 1/1000;$

$P(S) \approx 4/10; \quad P(\overline{S}) \approx 6/10.$

So, we estimate that:

$$\begin{aligned} P(S \mid C) &\approx \frac{(1/10)(4/10)}{(1/10)(4/10) + (1/1000) * (6/10)} \\ &\approx \frac{.04}{.0406} \approx 0.985 \end{aligned}$$

So, with "high probability", such an email is spam. (However, much caution is needed when interpreting such "probabilities".)

Generalized Bayes' Theorem

Suppose that $E, F_1, \ldots, F_n$ are events from sample space Ω, and that $P : \Omega \to [0,1]$ is a probability distribution on Ω. Suppose that $\cup_{i=1}^{n} F_i = \Omega$, and that $F_i \cap F_j = \emptyset$ for all $i \neq j$.
Suppose $P(E) > 0$, and $P(F_j) > 0$ for all j. Then for all j:

$$P(F_j \mid E) = \frac{P(E \mid F_j)P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i)P(F_i)}$$

Suppose Jane first randomly chooses a box from among n different boxes, $B_1, \ldots, B_n$, and then randomly picks a coloured ball out of the box she chose. (Each Box may have different numbers of balls of each colour.)
We can use the *Generalized Bayes' Theorem* to calculate the probability that Jane chose box B_j (event F_j), given that the colour of the ball that Jane picked is red (event E).

Proof of Generalized Bayes' Theorem: Very similar to the proof of Bayes' Theorem. Observe that:

$$P(F_j \mid E) = \frac{P(F_j \cap E)}{P(E)} = \frac{P(E \mid F_j)P(F_j)}{P(E)}$$

So, we only need to show that $P(E) = \sum_{i=1}^{n} P(E \mid F_i)P(F_i)$. But since $\bigcup_i F_i = \Omega$, and since $F_i \cap F_j = \emptyset$ for all $i \neq j$:

$$\begin{aligned} P(E) &= P(\bigcup_i (E \cap F_i)) \\ &= \sum_{i=1}^{n} P(E \cap F_i) \quad \text{(because } F_i\text{'s are disjoint)} \\ &= \sum_{i=1}^{n} P(E \mid F_i)P(F_i). \quad \square \end{aligned}$$