Topic 13: Conditional Probability and Bayes' Rule

02-680: Essentials of Mathematics and Statistics

December 9, 2024

1 Basics

We think of conditional probability as a restriction to some subset of our sample space. If we know one thing happened already whats the probability of another conditioned on the thing we know.

The basic rule of conditional probability is the following: For some probability space $(\Omega, \mathcal{A}, P)$, and for events $A, B \in \mathcal{A}$

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

Using the axiom of total probability we can show that

$$p(A \mid \Omega) = \frac{p(A \cap \Omega)}{p(\Omega)} = \frac{p(A)}{1}$$

Going back to the example from last topic if $A_{firstheads}$ is defined as before and $A_{allheads} = \langle \texttt{Heads}, \texttt{Heads} \rangle$, we can say that

$$p(A_{allheads} \mid A_{firstheads}) = \frac{p(A_{allheads} \cap A_{firstheads})}{p(A_{firstheads})} = \frac{1/4}{1/2} = \frac{1}{2}$$

This is based on the fact that

$$A_{allheads} \subseteq A_{firstheads} \implies A_{allheads} \cap A_{firstheads} = A_{allheads}.$$

This can be interpreted as "the probability you end up with both coin flips being heads, given that the first coin was heads."

If we examine the opposite: "the probability the first coin was heads, given that you end up with both coin flips being heads" the logic here still makes sense

$$p(A_{firstheads} \mid A_{allheads}) = \frac{1/4}{1/4} = 1.$$

Another example. Consider the following scenario for people's illnesses:

Flu? (F)	Cough? (C)	$p(C/\overline{C} \cap F/\overline{F})$
F	C	0.3
F	$\overline{C}$	0.2
$\overline{F}$	C	0.15
$\overline{F}$	$\overline{C}$	0.25

How do we calculate the probability you have flu given that you have a cough?

$$p(F \mid C) = \frac{p(C \cap F)}{p(C)}$$

But how do we get $p(C)$?

1.1 Sum Rule.

If we have some partition of the space $A_1, A_2, ..., A_n$ (remember a partition is a set of sets that all are disjoint, but $\bigcup_{i=1}^{n} A_i = \Omega$), then we can say the following:

$$p(B) = \sum_{i=1}^{n} p(B \cap A_i).$$

So, going back to our original question above, we can say that

$$p(C) = p(C \cap F) + p(C \cap \overline{F}) = 0.3 + 0.15$$

Therefore

$$p(F \mid C) = \frac{p(C \cap F)}{p(C)} = \frac{0.3}{0.45} = \frac{2}{3}$$

1.2 Product and Chain Rules.

The product rule, is an algebraic manipulation of the conditional probability definition, but comes in handy quite often (if $p(B) > 0$):

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

The chain rule is an expansion of this to multiple events $D_1, D_2, ... D_n$ (notice there is no condition on these events)

$$p(D_1 \cap D_2 \cap ... \cap D_n)$$
$$= p(D_n \mid D_1 \cap D_2 \cap ... \cap D_{n-1})p(D_1 \cap D_2 \cap ... \cap D_{n-1})$$
$$= p(D_n \mid D_1 \cap D_2 \cap ... \cap D_{n-1})p(D_{n-1} \mid D_1 \cap D_2 \cap ... \cap D_{n-2})p(D_1 \cap D_2 \cap ... \cap D_{n-2})$$
$$= ...$$
$$= p(D_n | D_1 \cap D_2 \cap ... \cap D_{n-1})...p(D_3|D_1 \cap D_2)p(D_2|D_1)p(D_1)$$
$$= p(D_1)p(D_2|D_1)p(D_3|D_1 \cap D_2)...p(D_n|D_1 \cap D_2 \cap ... \cap D_{n-1})$$

1.3 Independence

Events A and B are **independent** if

$$p(A \cap B) = p(A) \cdot p(B).$$

They are **conditionally independent** if

$$p(A \cap B|C) = p(A|C) \cdot p(B|C).$$

Example.   A box contains two coins: a regular coin ($p(\texttt{Heads}) = p(\texttt{Tails}) = \frac{1}{2}$) and one fake two-headed coin ($p(\texttt{Heads}) = 1$), we choose a coin and flip it twice.

Define the following events.

A — First coin toss results in a $\texttt{Heads}$ (let $\overline{A}$ be $\texttt{Tails}$)

B — Second coin toss results in a $\texttt{Heads}$ (let $\overline{B}$ be $\texttt{Tails}$)

C — the regular coin has been selected (let $\overline{C}$ be the biased coin)

So, the conditional probabilities are as follows:

$$
\begin{array}{llll}
p(A|C) & = \frac{1}{2} & p(A|\overline{C}) & = 1 \\
p(B|C) & = \frac{1}{2} & p(B|\overline{C}) & = 1 \\
p(A \cap B|C) & = \frac{1}{4} & p(A \cap B|\overline{C}) & = 1 \\
p(C) = p(\overline{C}) & = \frac{1}{2} &
\end{array}
$$

Since

$$p(A \cap B|C) = \frac{1}{4} = p(A|C) \cdot (B|C)$$

and

$$p(A \cap B|\overline{C}) = 1 = p(A|\overline{C}) \cdot (B|\overline{C})$$

we can see A and B are *conditionally* independent, but are they (regular) independent?

3

We know from the sum rule that if we have a partition (C and $\overline{C}$) we can say that

$$p(A) = p(A \cap C) + p(A \cap \overline{C})$$

then using the product rule,

$$= p(A|C)p(C) + p(A|\overline{C})p(\overline{C}) = \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{3}{4}$$

$p(B) = \frac{3}{4}$ using the same logic.

But to determine if A and B are independent we need to know $p(A \cap B)$, we can use the same trick:

$$
\begin{aligned}
p(A \cap B) &= p(A \cap B|C)p(C) + p(A \cap B|\overline{C})p(\overline{C}) \\
&= p(A|C)p(B|C)p(C) + p(A|\overline{C})p(B|\overline{C})p(\overline{C}) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot 1 \cdot \frac{1}{2} \\
&= \frac{1}{8} + \frac{1}{2} = \frac{5}{8} = \frac{10}{16}.
\end{aligned}
$$

But since $p(A) \cdot p(B) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$, we can see the coin flips are not independent.

Example. Consider again rolling a (6-sided) die and let

$$
\begin{aligned}
A &= \{1, 2\} \\
B &= \{2, 4, 6\} \\
C &= \{1, 4\}
\end{aligned}
$$

$$p(A) = \frac{1}{3} \quad p(B) = \frac{1}{2}$$

$$p(A \cap B) = \frac{1}{6} = p(A) \cdot p(B)$$

But,

$$p(A \mid C) = \frac{1}{2} \quad p(B \mid C) = \frac{1}{2}$$

$$p(A \cap B \mid C) = p(\{2\} \mid C) = 0$$

thus

$$p(A \cap B \mid C) = 0 \neq \frac{1}{4} = p(A \mid C) \cdot p(B \mid C).$$

4

2 Bayes' Rule

For some probability space $(\Omega, \mathcal{A}, P)$, and for some event that already happened $B \in \mathcal{A}$. Then the probability that some event $A \in \mathcal{A}$ also happens:

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

(where $p(B) > 0$).

Bayes' Rule * tells us

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

we call

    $p(A)$ — the prior

    $p(B \mid A)$ — the likelihood

    $p(A \mid B)$ — the posterior

Another form:

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

Example. Let A be event where you have the flu, and B be the event you cough. Given the following information:

    Prior: $p(A) = 0.05$ (background probability that you have a cough for no reason)

    Likelihood:

        − $p(B \mid A) = 0.8$ (probability you have a cough given you have the flu),

        − $p(B \mid \overline{A}) = 0.2$ (probability you have a cough if you don't have the flu)

Whats the probability you have the flu if you have a cough ($p(A \mid B)$)?

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B \mid A)p(A) + p(B \mid \overline{A})p(\overline{A})} = \frac{(0.8)(0.05)}{(0.8)(0.05) + (0.2)(0.95)} = \frac{0.04}{0.04 + 0.19} \approx 0.174$$

What if you're really susceptible to a cough, so $p(A) = 0.5$:

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B \mid A)p(A) + p(B \mid \overline{A})p(\overline{A})} = \frac{(0.8)(0.5)}{(0.8)(0.5) + (0.2)(0.5)} = \frac{0.4}{0.4 + 0.1} = 0.8$$

*originally published in 1763 by Thomas Bayes

Useful References

Wasserman. "All of Statistics: A Concise Course in Statistical Inference" §1 Degroot and Schervish. "Probability and Statistics" §2