STA286 Lecture 04

Neil Montgomery

Last edited: 2017-01-17 10:50

the axiomatic approach

Some of the basic rules can be formally proven, which is great fun!

Theorem 1 $P(\emptyset) = 0$

Theorem 2 $P(A') = 1 - P(A)$

Theorem 3 If $A \subset B$ then $P(A) \leq P(B)$

Theorem 4 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Corollaries to Theorem 4: $P(A \cup B) = P(A) + P(B)$ when A and B are disjoint, and $P(A \cup B) \leq P(A) + P(B)$ (always).

conditional probability

partial information

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$\begin{aligned} A &= \{2, 5\}, \\ B &= \{2, 4, 6\}, \\ C &= \{1, 2\}. \end{aligned}$$

partial information

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$A = \{2, 5\},$$
$$B = \{2, 4, 6\},$$
$$C = \{1, 2\}.$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

partial information

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$
\begin{aligned}
A &= \{2, 5\}, \\
B &= \{2, 4, 6\}, \\
C &= \{1, 2\}.
\end{aligned}
$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

Let's use a "personal probability" philosophy for the momemnt.

partial information

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$A = \{2, 5\},$$
$$B = \{2, 4, 6\},$$
$$C = \{1, 2\}.$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

Let's use a "personal probability" philosophy for the momemnt.

What if I peek and tell you "Actually, B occurred". What is the (*your?*) probabality of A given this partial information? It is $\frac{1}{3}$.

partial information

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$\begin{aligned} A &= \{2, 5\}, \\ B &= \{2, 4, 6\}, \\ C &= \{1, 2\}. \end{aligned}$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

Let's use a "personal probability" philosophy for the momemnt.

What if I peek and tell you "Actually, B occurred". What is the (*your?*) probability of A given this partial information? It is $\frac{1}{3}$.

I roll the die again, peek, and tell you "Actually, C occurred". Now the probability of A is $\frac{1}{2}$.

partial information

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$A = \{2, 5\},$$
$$B = \{2, 4, 6\},$$
$$C = \{1, 2\}.$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

Let's use a "personal probability" philosophy for the momemnt.

What if I peek and tell you "Actually, B occurred". What is the (*your?*) probability of A given this partial information? It is $\frac{1}{3}$.

I roll the die again, peek, and tell you "Actually, C occurred". Now the probability of A is $\frac{1}{2}$.

Intuitively people use a "sample space restriction" approach in these simple cases.

elementary definition of conditional probability

Given B with $P(B) > 0$,

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

"The conditional probability of A given B"

The answers for the previous example coincide with the intuitive approach.

elementary definition of conditional probability

Given B with $P(B) > 0$,

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

"The conditional probability of A given B"

The answers for the previous example coincide with the intuitive approach.

Fun fact: For a fixed B with $P(B) > 0$, the function $P_B(A) = P(A|B)$ is a probability function. (You can prove this.)

useful expressions for calculation - I

$P(A \cap B) = P(A|B)P(B)$ often comes in handy.

Consider the testing for, and prevalence of, a viral infection such as HIV.

Denote by A the event "tests positive for HIV", and by B the event "is HIV positive."

useful expressions for calculation - I

$P(A \cap B) = P(A|B)P(B)$ often comes in handy.

Consider the testing for, and prevalence of, a viral infection such as HIV.

Denote by A the event "tests positive for HIV", and by B the event "is HIV positive."

For the ELISA screening test, $P(A|B)$ is about 0.995. The prevalence of HIV in Canada is about $P(B) = 0.00212$.

useful expressions for calculation - I

$P(A \cap B) = P(A|B)P(B)$ often comes in handy.

Consider the testing for, and prevalence of, a viral infection such as HIV.

Denote by A the event "tests positive for HIV", and by B the event "is HIV positive."

For the ELISA screening test, $P(A|B)$ is about 0.995. The prevalence of HIV in Canada is about $P(B) = 0.00212$.

The probability of a randomly selected Canadian being HIV positive and testing positive is:

$$P(A \cap B) = P(A|B)P(B) = 0.0021094$$

useful expressions for calculation - II

If $B_1, B_2, \ldots$ is a partition of S with all $P(B_i) > 0$, then:

$$\begin{aligned} P(A) &= P\left(\bigcup_i (A \cap B_i)\right) \\ &= \sum_i P(A \cap B_i) \end{aligned}$$

useful expressions for calculation - II

If $B_1, B_2, \ldots$ is a partition of S with all $P(B_i) > 0$, then:

$$\begin{aligned} P(A) &= P\left(\bigcup_i (A \cap B_i)\right) \\ &= \sum_i P(A \cap B_i) \\ &= \sum_i P(A|B_i)P(B_i) \end{aligned}$$

Common simple version: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

useful expressions for calculation - II

If $B_1, B_2, \ldots$ is a partition of S with all $P(B_i) > 0$, then:

$$\begin{aligned} P(A) &= P\left(\bigcup_i (A \cap B_i)\right) \\ &= \sum_i P(A \cap B_i) \\ &= \sum_i P(A|B_i)P(B_i) \end{aligned}$$

Common simple version: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Continuing with the HIV example, suppose we also know $P(A|B^c) = 0.005$ ("false positive").

useful expressions for calculation - III

We can now calculate $P(A)$, the probability of a randomly selected Canadian testing positive.

$$
\begin{aligned}
P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\
&= 0.995 \cdot 0.00212 + 0.005 \cdot (1 - 0.00212)
\end{aligned}
$$

useful expressions for calculation - III

We can now calculate $P(A)$, the probability of a randomly selected Canadian testing positive.

$$
\begin{aligned}
P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\
&= 0.995 \cdot 0.00212 + 0.005 \cdot (1 - 0.00212) \\
&= 0.0070988
\end{aligned}
$$

The simple formula gets a grandiose title: ***"THE! LAW! OF! TOTAL! PROBABILITY!!!!!"***

Now, in the HIV example, we also might be interested in $P(B|A)$, the probability that someone is HIV+ given that they test positive.

$P(B|A)$

A little algebra:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

In our example this is $\frac{0.0021094}{0.0070988} = 0.2971$.

Bayes' rule in general

If $B_1, B_2, \ldots$ is a partition of S with all $P(B_i) > 0$, then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_i P(A|B_i)P(B_i)}$$

independence

motivation - revisit the die toss example

I'll roll a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$A = \{2, 5\},$$
$$B = \{2, 4, 6\}$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

What if I peek and tell you "Actually, B occurred". What is the probabality of A given this partial information? It is $\frac{1}{3}$.

The probability of A didn't change after the new information:

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

definition(s) of independence

A and B are (pairwise) *independent* (notation $A \perp B$) if:

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

No requirement for $P(A)$ or $P(B)$ to be positive.

$A_1, A_2, A_3, \ldots$ (possibly infinite) are (mutually) *independent* if for any finite subcollection of indices $I = \{i_1, \ldots, i_n\}$:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

independence of two classes of events

Note that if $A \perp B$, then also $A \perp B^c$ and so on. Consider:

$$\mathcal{A} = \{\emptyset, A, A^c, S\}$$
$$\mathcal{B} = \{\emptyset, B, B^c, S\}$$

Classes of events $\mathcal{A}$ and $\mathcal{B}$ are *independent* all pairs of events with one chosen from each class are independent.

The suggests a concept of "independent experiments", which will be revisited.

the “any” and “all” style of examples

(Note: in probability modeling, independence is usually *assumed*.)

A subway train is removed from service if *any* of its doors are stuck open. There is a probability p of a door getting stuck open on one day of operations. A train has n doors.

Example question: what is the chance a train is removed from service due to stuck doors on one day of operations?

p^n “all doors fail”

$1 - p^n$ “not all doors fail”

$(1 - p)^n$ “no doors fail”

$1 - (1 - p)^n$ “not *no doors fail*, in other words *any doors fail*”