Conditional probability & Independence

Lecture 1b (STAT 24400 F24)

Definition

The **conditional probability of A given B** is defined as

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

provided that $\mathbb{P}(B) \neq 0$.

Remarks

- Intuitively, this is the probability that, if we know the event B occurred, the chance that the event A has also occurred.
- The relevant sample space becomes B rather than the original Ω.
- A conditional probability is a probability measure satisfying the axioms etc.

Example (deck of cards)

After shuffling a deck of cards,
what is the probability that K♣ is last, given that A♣ is first?

Solution: let A be the event that K♣ is last,
& let B be the event that A♣ is first $\rightsquigarrow$ what is $\mathbb{P}(A \mid B)$?

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\texttt{K}\clubsuit \text{ is last and } \texttt{A}\clubsuit \text{ is first})}{\mathbb{P}(\texttt{A}\clubsuit \text{ is first})}$$

$$= \frac{\frac{50!}{52!}}{\frac{51!}{52!}} = \frac{1}{51}$$

The underlying sample space:

$\Omega = \{$all possible orderings of the 52 cards$\}$, each equally likely
$\rightsquigarrow$ uniform probability measure over 52! many possibilities

Example (fire alarm)

In a university building (say, Eckhart),
on a day with no fire, there is a 0.001 chance of a fire alarm (false positive);
if there is a fire, the alarm has a 0.99 chance of going off.
On average fires happen once in three years.
If the alarm goes off, what is the probability that there is a fire?

Solution: The goal is to calculate $\mathbb{P}(F \mid A)$
where F = event that there is a fire, A = event that alarm goes off.

$$\mathbb{P}(F \mid A) = \frac{\mathbb{P}(F \cap A)}{\mathbb{P}(A)} = \frac{0.00090}{0.001899} = 0.47$$

	Fire	No fire	Total
Alarm	$0.99 \cdot 0.00091 = 0.00090$	$0.001 \cdot 0.99909 = 0.000999$	0.001899
No alarm			
Total	$\frac{1}{365 \times 3} = 0.00091$	$1 - 0.00091 = 0.99909$	1

Remarks

- Some common terms
  - Sensitivity (true positive rate) $\mathbb{P}(A \mid F) = 0.99$
  - Specificity (true negative rate) $\mathbb{P}(A^c \mid F^c) = 1 - \mathbb{P}(A \mid F^c) = 0.999$
  - False positive rate $= 1-$ Specificity
  - False negative rate $= 1-$ Sensitivity
  - PPV (positive predictive value): $\mathbb{P}(F \mid A) = 0.47$
  - FDR (false discovery rate): $\mathbb{P}(F^c \mid A) = 1 - \mathbb{P}(F \mid A) = 1 - 0.47 = 0.53$
  - FOR (false omission rate): $\mathbb{P}(F \mid A^c)$
  - NPV (negative predictive value): $\mathbb{P}(F^c \mid A^c) = 1 - \mathbb{P}(F \mid A^c)$
- Note that the conditional probab. $\mathbb{P}(A \mid F)$ & $\mathbb{P}(F \mid A)$ are very different.
- Similarly, $\mathbb{P}(A \mid F^c)$ & $\mathbb{P}(F^c \mid A)$ are very different.
  "False positives are unlikely" $\not\Leftrightarrow$ "a positive result is unlikely to be false".

Discussion: How to improve

What are our choices if we would like to change the status quo?

- If the alarm has some threshold (e.g., how much heat/smoke), we may adjust the trade-off between false positive and false negative.
  - Make the alarm less sensitive:
    Set the threshold higher $\rightsquigarrow$ fewer false positives, but miss more fires
  - Make the alarm more sensitive:
    Set the threshold lower $\rightsquigarrow$ miss fewer fires, but more false positives
- We may choose to invest in improving technology to produce a more precise alarm that will have fewer false positive and fewer false negative.

Application in clinical trials

Analogy to statistics application in clinical trials – Research studies to evaluate if a new medical treatment, such as a new drug, is safe and effective in people. In a simplified version,

fire *vs.* no fire = drug is effective *vs.* not effective
alarm *vs.* no alarm = trial is declared a success *vs.* not a success

- Set the threshold for success to be higher
  $\rightsquigarrow$ fewer false positives, but miss a larger number of effective drugs
- Set the threshold for success to be lower
  $\rightsquigarrow$ discover more effective drugs, however at the cost of more false positives (approving drugs that are actually not effective)
- Or, invest resources so that the trials are more accurate (better $\mathbb{P}(A|F)$ & $\mathbb{P}(A^c|F^c)$) with more precise estimates (mostly commonly, by using a larger sample size)

Law of total probability

For the fire example, to calculate $\mathbb{P}(A)$, the overall probability of alarm going off whether there is a fire or not, we implicitly used the law of total probability:

If events $B_1, B_2, \ldots, B_n$ partition the sample space, then

i.e., B_i's are disjoint, and $\Omega = B_1 \cup B_2 \cup \cdots \cup B_n$

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)$$

How did we use this rule?

$$\mathbb{P}(A) = \mathbb{P}(A \cap F) + \mathbb{P}(A \cap F^c) = \mathbb{P}(A \mid F) \cdot \mathbb{P}(F) + \mathbb{P}(A \mid F^c) \cdot \mathbb{P}(F^c)$$

The law also holds for a countably infinite partition $B_1, B_2, \ldots$
(i.e. B_i's are disjoint, and $\Omega = B_1 \cup B_2 \cup \cdots \cup B_n \cup \cdots = \cup_{n=1}^{\infty} B_n$)

Law of total probability

Proof: We'll prove the law of total probability for the simple case $n = 2$.

Using the axioms of probability measures and laws of set theory:

B_1 & B_2 disjoint $\rightsquigarrow$ $A \cap B_1$ & $A \cap B_2$ are disjoint
(could prove formally, using commutativity and associativity of $\cap$; omitted)

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) && \leftarrow \text{since } A = A \cap \Omega \\
&= \mathbb{P}(A \cap (B_1 \cup B_2)) && \leftarrow \text{since } \Omega = B_1 \cup B_2 \\
&= \mathbb{P}((A \cap B_1) \cup (A \cap B_2)) && \leftarrow \text{by distributive law} \\
&= \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) && \leftarrow \text{by axiom 3 } (A \cap B_1 \text{ \& } A \cap B_2 \text{ are disjoint}) \\
&= \mathbb{P}(A \mid B_1) \cdot \mathbb{P}(B_1) + \mathbb{P}(A \mid B_2) \cdot \mathbb{P}(B_2) && \leftarrow \text{by def. of conditional prob.}
\end{aligned}$$

(The construction of the proofs for larger n or countably infinite n can be analogous.)

Bayes' Rule

For any event A and any events $B_1, B_2, \ldots$ that partition the sample space,

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_j \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j)}$$

Proof:

$$\begin{aligned}
\mathbb{P}(B_i \mid A) &= \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} && \leftarrow \text{by def. of conditional prob.} \\
&= \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\mathbb{P}(A)} && \leftarrow \text{by def. of conditional prob.} \\
&= \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_j \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j)} && \leftarrow \text{by law of total prob.}
\end{aligned}$$

Redo fire alarm example:

$$\mathbb{P}(F \mid A) = \frac{\mathbb{P}(A \mid F) \cdot \mathbb{P}(F)}{\mathbb{P}(A \mid F) \cdot \mathbb{P}(F) + \mathbb{P}(A \mid F^c) \cdot \mathbb{P}(F^c)}$$

Multiplication law

For events $A_1, \ldots, A_n$,

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

For example, the multiplication law for $n = 2$ is

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1)$$

which has been used in previous examples.

Multiplication law for independence events

Two events A and B are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

If $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are nonzero, it's equivalent to say

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

or

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

Intuitively, *knowing that A occurred does not change the likelihood that B occurred*, and vice versa.

Examples: Independence

Are A & B independent in these examples?

- Flip a coin twice. $A =$ 1st coin is Heads, $B =$ 2nd coin is Tails Yes
- Draw two cards. $A =$ 1st card is `K`, $B =$ 2nd card is `Q` No
- Draw two cards. $A =$ 1st card is `K`, $B =$ 2nd card is red Yes
- A practical case: Randomly sample one patient from a hospital.
  $A =$ age 0–15, $B =$ hospitalized for an infection No
- A tricky case: Randomly sample one patient from a hospital.
  $A =$ first name starts with "A", $B =$ hospitalized for an infection No

Example: Multiplication law for independent events

Suppose you roll a fair dice repeatedly.
What is the probability that you get a 6 for the first time, on the 3rd roll?

Define events:
$A = \{\text{1st roll} \neq 6\}$, $B = \{\text{2nd roll} \neq 6\}$, $C = \{\text{3rd roll} = 6\}$

$$\mathbb{P}(A \cap B \cap C) \underset{\substack{\uparrow \\ \text{since } A,\, B,\, C \text{ are} \\ \text{mutually independent}}}{=} \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}$$

Example: Multiplication law for dependent events

Suppose you draw three cards at random from a standard deck.

If we get a number (`2`, `3`, ..., `10`), that value is the number of points earned.
If we get a `J`, `Q`, or `K`, then we earn 10 points. An `A` earns 0 points.

What is the prob. that you earn 10 points, then 5 points, then 10 points?

Define events:

$A = \{\text{1st card} = \text{10pts}\}$, $B = \{\text{2nd card} = \text{5pts}\}$, $C = \{\text{3rd card} = \text{10pts}\}$

$$\mathbb{P}(A \cap B \cap C) \underset{\substack{\uparrow \\ \text{multiplication law}}}{=} \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B) = \frac{16}{52} \cdot \frac{4}{51} \cdot \frac{15}{50}$$