INDEPENDENCE CONDITIONING

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

MATH 5670-6670 FALL 2019

PROBABILITY I

September 6, 2019

Outline

Example

Sample a card from a standard deck. Assign values:

$$A \mapsto 1 \text{ and } J \mapsto 11,\ Q \mapsto 12,\ K \mapsto 13.$$

The following pairs of events A and B satisfy the equation.

A	B
spade	ace
spade	face
black	even
black	prime
no face	multiple of 4
etc.	(pattern?)

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B).$$

Definition

1. Events A_1, A_2 are called **independent** if
   $P(A_1A_2) = P(A_1)P(A_2)$.

2. In general, events $A_1, \cdots, A_n$ are called **independent** if
   1. $P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n)$ and
   2. every selection consists of independent events.

At the first glance the definition seems to be circular, the term is explained by itself.

It is not so.

The definition is rigorous because it refers to shorter sequences.

Example

A_1, A_2, A_3 are independent if

- The first condition (2a) is satisfied:

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3),$$

- and the second condition (2b) involving selections of two sets is satisfied in view of the condition 1 for pairs:

$$P(A_1A_2) = P(A_1)P(A_2),\ P(A_1A_3) = P(A_1)P(A_3),$$
$$P(A_2A_3) = P(A_2)P(A_3).$$

Exercise. Show that the $2^n - n - 1$ is the number of equations needed to establish independence of n events.

Independent stay independent...

Divide independent $A_1, \ldots, A_n$ into two groups, say $i = 1, \ldots, m$ and $i = m+1, \ldots n$ $(1 \leq m < n)$.

Using set operations, create B out of the first group and C out of the second group.

Theorem

The new events B and C are independent.

Proof. Duh, of course! ∎

A rigorous proof is rather tedious. See how it works for examples.

Say, $n = 3$, $B = A_1'$ and $C = A_2 \cup A_3$.

Or, $n = 4$, $B = A_1 A_2$ and $C = A_3 \oplus A_4$. Etc.

Definition

Let S be the sample space, with a collection of events, and equipped with probability P. Suppose that the even B of positive probability has occurred. This decreases uncertainty. A modification is needed:

- The event B becomes the new sample space.
- All new events are now of the form $A \cap B$,
- The new probability comes from normalization:

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

"**the conditional probability of A given B equals...**".

What was first - a chicken or an egg?

In the textbook the conditional probability is introduced first, from which the notion of independence is derived.

However, the independence

- was crucial earlier in the Multiplication Rule;
- is conceptually simpler;
- is more intuitive.

Therefore I and many others consider it as a primary notion, "independent" of conditional probability.

The latter provides just an additional insight.

CP vs. independence

Clearly, A and B with positive probability are independent iff

$$\mathrm{P}(A|B) = \mathrm{P}(A) \quad \text{or/and} \quad \mathrm{P}(B|A) = \mathrm{P}(B)$$

To tie independence with CP for more events is much more tedious.
E.g., A_1, A_2, A_3 are independent iff

$$\mathrm{P}(A_1|A_2) = \mathrm{P}(A_1),\ \mathrm{P}(A_1|A_3) = \mathrm{P}(A_1),\ \mathrm{P}(A_2|A_3) = \mathrm{P}(A_2), \\ \text{and } \mathrm{P}(A_1|A_2A_3) = \mathrm{P}(A_1).$$

Now, try more events: four or five... Or say "No, thank you."

Why “conditional”

Duh! Every probability is “conditional” because it relies on the probability space Ω and the defined family of events.

For example, we assume the probability of an ace as $\frac{1}{13}$ while drawing from the standard deck of cards. It's a simple urn model with 4 aces out of total 52 items, each item has the same chance of being drawn, and thus we take the ratio $\dfrac{4}{52} = \dfrac{1}{13}$.

When the deck is reduced to, say, 24 cards (9 or higher like in the game of Thousand), the probability changes to $\dfrac{4}{24} = \dfrac{1}{6}$.

Different conditions yield possibly different probabilities.

Grave error!

The additivity fails in regard to conditions. For disjoint B_1 and B_2

$$\mathrm{P}(A|B_1 \sqcup B_2) \neq \mathrm{P}(A|B_1) + \mathrm{P}(A|B_2), \quad \text{in general.}$$

Claiming the truth is an **error**. Of what kind? Let's see. Suppose that the equality $=$ holds. By definition of CP

$$\frac{\mathrm{P}(AB_1) + \mathrm{P}(AB_2)}{\mathrm{P}(B_1) + \mathrm{P}(B_2)} = \frac{\mathrm{P}(A(B_1 \cup B_2))}{\mathrm{P}(B_1 \cup B_2)} = \frac{\mathrm{P}(AB_1)}{\mathrm{P}(B_1)} + \frac{\mathrm{P}(AB_2)}{\mathrm{P}(B_2)}.$$

That is, denoting terms by single variables, $\dfrac{a+b}{c+d} = \dfrac{a}{c} + \dfrac{b}{d}$ (**ouch!**)

Summary - basic properties

- $\mathrm{P}(A|B) = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$, where $\mathrm{P}(B) > 0$ (undefined otherwise).

- Again, given B with $\mathrm{P}(B) > 0$, the CP $\mathrm{P}(\cdot|B)$ is a probability:

  $\mathrm{P}(\mathbf{A_1} \sqcup \mathbf{A_2} \sqcup \cdots |\mathbf{B}) = \mathrm{P}(\mathbf{A_1}|\mathbf{B}) + \mathrm{P}(\mathbf{A_2}|\mathbf{B}) + \cdots;$

  In particular, $\mathrm{P}(A'|B) = 1 - \mathrm{P}(A|B)$.

- However, given A, the CP $\mathrm{P}(A|\cdot)$ is not a probability:

  $\mathrm{P}(\mathbf{A}|\mathbf{B_1} \sqcup \mathbf{B_2}) \neq \mathrm{P}(\mathbf{A}|\mathbf{B_1}) + \mathrm{P}(\mathbf{A}|\mathbf{B_2})$, in general.

More...

- Conditioning **does not imply a causal relation** (that B is a cause and A is an effect).

  Nevertheless, as a metaphor the cause-effect jargon is too tempting to pass.

- Uncertainty may decrease or increase upon conditioning.

- The defining formula can be rewritten symmetrically $\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A)$ (if $\mathrm{P}(A) > 0$).

Multiplication Rule

Given a sequence of events $B_1, B_2, ..., B_n$

$$\mathsf{P}(B_1 \cdots B_n) = \mathsf{P}(B_1) \cdot \mathsf{P}(B_2|B_1) \cdot \mathsf{P}(B_3|B_1 B_2) \cdots \mathsf{P}(B_n|B_1 \cdots B_{n-1}).$$

Proof. Definition and straightforward algebra.

Conditioning

Suppose that the sample space Ω is partitioned as the union of two or more disjoint events $B_1,\ B_2,\ldots$. Then an event A is the union of the suitable portions:

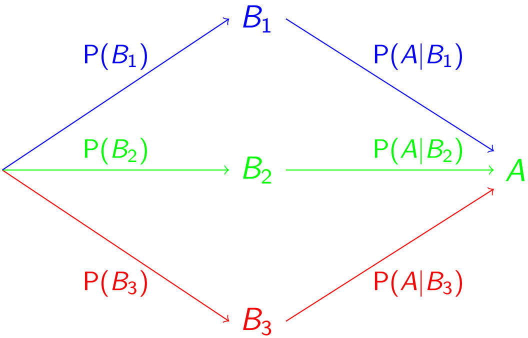
$$A = A\Omega = A(B_1 \sqcup B_2 \sqcup \cdots) = \Big(AB_1\Big) \sqcup \Big(AB_2\Big) \sqcup \cdots$$

So, this is so called "**conditioning**":

$$\begin{aligned} P(A) &= \mathrm{P}\Big((AB_1) \sqcup (AB_2) \sqcup \cdots \Big) = \mathrm{P}(AB_1) + \mathrm{P}(AB_2) + \cdots \\ &= \mathrm{P}(A|B_1)\,\mathrm{P}(B_1) + \mathrm{P}(A|B_2)\,\mathrm{P}(B_2) + \cdots . \end{aligned}$$

Visualization by the tree

Connsider three conditions B_1, B_2, B_3.



$\mathrm{P}(A)$: Add path products $\mathrm{P}(A|B_i)\mathrm{P}(B_i)$, $i = 1, 2, 3$
(such product is a probability mass of a path)

Three types of scientific reasoning

- **deduction** - from general to particular,
- **induction** - from details, particular cases, to a general pattern,
- **abduction** - from the effect to a cause
  (or, more correctly, to a factor)

(of course, there are more...)

Examples of abduction

- AI - Turing test
- CSI - whodunit?
- coral reefs are dying - why?
- a restaurant chain is performing poorly - what is the reason?
- trees turn yellow before their time - what is the culprit?
- a patient has rash, fever, and cough - what is the diagnosis?
- a spouse starts working long into nights - hmm... really?
- this soup tastes funny - what did you put in it?
- so called History - why this or that happened?
- reverse engineering..., etc.

Bayes' Theorem or Reversal of Conditioning

Consider $n = 3$ disjoint conditions $B_1, \ldots, B_n$. Say, A has occurred.

Q: What are the chances that, say, B_i was involved?

A: Formalizing the question, by simple algebra we obtain

Theorem (Bayes')

$$\mathrm{P}(B_i|A) = \frac{\mathrm{P}(A|B_i)\,\mathrm{P}(B_i)}{\mathrm{P}(A)}, \; i = 1, \ldots, n.$$

Or, referring to the tree,

$$\frac{\text{mass of the path through } B_i}{\text{total mass } \mathrm{P}(A)}$$

Example: chances and fractions

Consider disjoint conditions B_1, B_2, B_3 with probabilities $\frac{3}{12}, \frac{7}{12}, \frac{2}{12}$ and the CPs $\mathrm{P}(A|B_i)$: 50%, 30%, 70%.

Rewrite the data: $\mathrm{P}(A|B_i) \sim 5, 3, 7; \quad \mathrm{P}(B_i) \sim 3 : 7 : 2$. Then

$$\begin{aligned}
\mathrm{P}(B_1|A) &= \frac{5 \cdot 3}{5 \cdot 3 + 3 \cdot 7 + 7 \cdot 2} &= \frac{15}{50} &= 30\%, \\
\mathrm{P}(B_2|A) &= \frac{3 \cdot 7}{5 \cdot 3 + 3 \cdot 7 + 7 \cdot 2} &= \frac{21}{50} &= 42\%, \\
\mathrm{P}(B_3|A) &= \frac{7 \cdot 2}{5 \cdot 3 + 3 \cdot 7 + 7 \cdot 2} &= \frac{14}{50} &= 28\%,
\end{aligned}$$

or $\mathrm{P}(B_i|A) \sim 15 : 21 : 14$. **Do not simplify the fractions!**

False positive and false negative

Consider a test for a disease D. The outcomes T_+ or T_- indicate the presence or absence of the disease. No test is 100% error-proof. Therefore two errors are possible:

false negative with the rate $\alpha = \mathsf{P}(T_- | D)$

the test says "NO" in spite of having the disease

false positive with the rate $\beta = \mathsf{P}(T_+ | D')$

the test says "YES" in spite of being disease-free

Example

A test indicated the presence of a relatively rare disease.

What are the chances of factually having the disease?

Let D have prevalence $p = 0.001$ (one in thousand).

Let the test be 99% accurate, i.e., $\mathrm{P}(T_+|D) = 0.99$.

Let the FPR $\beta = \mathrm{P}(T_+|D')$ be between 0.01 and 0.1. Then

$$\begin{aligned} \mathrm{P}(D|T_+) &= \frac{\mathrm{P}(T_+|D)\mathrm{P}(D)}{\mathrm{P}(T_+)} = \frac{0.99 * 0.001}{0.99 * 0.001 + \beta * 0.999} \\ &= \frac{99}{99 + 99,900\beta} \approx \frac{1}{1 + 10,000\beta} \in [0.1\%, 1\%] \end{aligned}$$