GR5203 - 09/12/2019

Thursday, September 12, 2019 8:30 AM

<u>example</u> Suppose a blood test is 100% effective in detecting (certain) disease when it is present. Also, the test yields "false positives", i.e. for 1% of healthy population it will yield a positive test. If .1% of the population has the disease, what's the probability that a person who tests positive actually has the disease?

D = person has the disease

$\oplus$ = person tests positive

$$P(D \mid \oplus) = ?$$

<u>Know</u>:

$$P(\oplus \mid D) = 1$$

$$P(\oplus \mid D^c) = .01$$

$$P(D) = .001$$

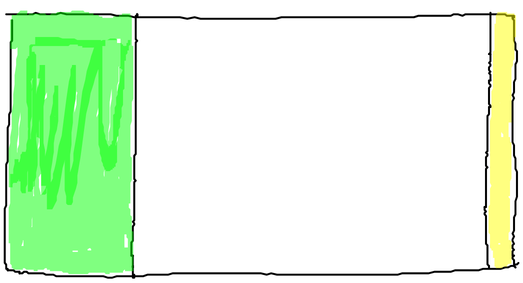
$$P(D \mid \oplus) = \frac{P(D \cap \oplus)}{P(\oplus)} =$$

$$= \frac{P(\oplus \mid D)\, P(D)}{P(\oplus \mid D)\, P(D) + P(\oplus \mid D^c)\, P(D^c)}$$

$$= \frac{1 \cdot .001}{1 \cdot .001 + .01 \cdot .999}$$

$$= \frac{100}{1099} \approx \frac{1}{11}$$

Venn Diagram



$D =$ (yellow) $\qquad P(D) = .001$

$\oplus \cap D^c =$ (green)

$$P(\oplus \cap D^c) = P(\oplus \mid D^c)\, P(D^c)$$

$$= .01 \times .999 = .00999 \approx .01$$

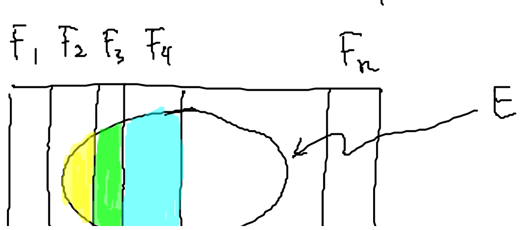
$$\oplus = (\oplus \cap D^c) \cup (\oplus \cap D)$$

$$= (\oplus \cap D^c) \cup D$$

The prevelance & "false positive" rate need to always be considered together. //

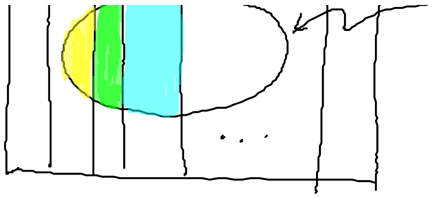
Suppose $F_1, F_2, \ldots, F_n$ are events such that

$$S = \bigcup_{k=1}^{n} F_k \quad \text{and} \quad F_j \cap F_i = \emptyset, \; i \neq j.$$



$$E = E \cap S =$$

$$= E \cap \left(\bigcup_{k=1}^{n} F_k \right)$$



$$= E \cap \left(\bigcup_{k=1}^{n} F_k \right)$$

$$= \bigcup_{k=1}^{n} \underbrace{(E \cap F_k)}_{\text{disjoint events}}$$

$$P(E) = P\left(\bigcup_{k=1}^{n} (E \cap F_k) \right) =$$

$$= \sum_{k=1}^{n} P(E \cap F_k) = \sum_{k=1}^{n} P(E \mid F_k) P(F_k)$$

$$= P(E)$$

The Law of Total Probability.

Bayes Formula

Think of E = outcome of an experiment (data)

$F_1, F_2, \dots, F_n$ = hypothesis $\left(\bigcup_{k=1}^{n} F_k = S \right.$

$\left. i \neq j,\ F_i \cap F_j = \emptyset \right)$

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

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

$$\sum_{i=1} P(E \mid F_i)\, P(F_i)$$

Bayes Formula.

example: A plane is missing and it is equally likely to have gone down in on of 3 regions. Let $1-\beta_i$ be the probability that the plane would found upon a search of the i^{th} region, $i=1,2,3$. What's the probability that the plane is in region $i \in \{1,2,3\}$, given that the search of region 1 was unsuccessful?

F_1, F_2, F_3 = events that the plane is in region 1, 2, or 3.

$$\bigcup_{i=1}^{3} F_i = S$$

E = search of region 1 was unsuccessful.

prior: $P(F_1) = P(F_2) = P(F_3) = \frac{1}{3}$

model: $P(E \mid F_1) = \beta_1$, $P(E \mid F_2) = 1 = P(E \mid F_3)$

$$P(F_1 \mid E) = \frac{P(E \mid F_1)\, P(F_1)}{\sum_{i=1}^{3} P(E \mid F_i)\, P(F_i)}$$

$$= \frac{\beta_1 \cdot \frac{1}{3}}{\beta_1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{\beta_1}{\beta_1 + 2}$$

$$P(F_2 \mid E) = \ldots\ldots = \frac{1}{\beta_1 + 2} = P(F_3 \mid E)$$

As long as $\beta_1 < 1$, $P(F_1 \mid E) < P(F_2 \mid E)$

$\Rightarrow$ plane is now more likely to be in region 2 or 3. //

In general, $P(E \mid F)$ is not equal to $P(E)$.

Independence

Sometimes new information does not change your beliefs.

$$P(E \mid F) = P(E)$$

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

$\iff \underbrace{P(E\cap F) = P(E)P(F)}$.

<u>Def</u> Two events E and F are <u>independent</u> if $P(E\cap F) = P(E)P(F)$

(dependent if not independent).

<u>example</u> Draw a card from a standard 52 card deck.

E = "ace", F = "spade"

$P(E\cap F) = \frac{1}{52}$; $P(E) = \frac{4}{52} = \frac{1}{13}$

$P(F) = 13/52 = 1/4$

$\Rightarrow P(E\cap F) = P(E)P(F)$

$\Rightarrow$ E and F are independent.

We will write $E \perp\!\!\!\perp F$ ← "independent" //

<u>example</u> 2 coins, "fair"

$\{HH, HT, TH, TT\}$ equally likely.

H_1 = 1st coin comes up heads,

H_2 = 2nd coin comes up heads.

$P(H_1) = 2/4 = 1/2$; $P(H_2) = 1/2$

$P(HH) = P(H_1 \cap H_2) = 1/4$

$\Rightarrow \quad P(H_1 \cap H_2) = P(H_1)\, P(H_2)$

$\Rightarrow \quad H_1 \perp\!\!\!\perp H_2$. //

<u>example</u> ① Toss 2 fair dice

E_1 = sum of upturned faces is 6.

F = 1st die comes up 4.

$P(E_1 \cap F) = 1/36$, $P(F) = 1/6$

$P(E_1) = 5/36$

$\Rightarrow \quad P(E_1 \cap F) \neq P(E_1)\, P(F)$

$\Rightarrow \quad E_1 \not\!\perp\!\!\!\perp F$

---

② 2 fair dice, F = same

E_2 = sum of upturned faces is 7.

$P(E_2 \cap F) = 1/36$, $P(F) = 1/6$

$P(E_2) = 6/36 = 1/6$

$\Rightarrow \quad P(E_2 \cap F) = P(E_2) \cdot P(F)$

$\Rightarrow \quad E_2 \perp\!\!\!\perp F$.

Always check your intuition!

<u>Note</u>: If $E \perp\!\!\!\perp F$ then so are

$$E \perp\!\!\!\perp F^c$$

$$P(E \cap F^c) = P(E) - P(E \cap F)$$

$$= P(E) - P(E)P(F)$$

$$= P(E)\left[1 - P(F)\right] =$$

$$= P(E)\,P(F^c)$$

$$\Rightarrow \quad E \perp\!\!\!\perp F^c. \qquad //$$

<u>example</u> (motivating independence of more than 2 events)

Suppose E, F, G are s.t.

$$E \perp\!\!\!\perp F, \quad F \perp\!\!\!\perp G, \quad E \perp\!\!\!\perp G$$

"pairwise independent"

$$\text{Is } E \perp\!\!\!\perp (F \cap G)\text{?}$$

2 fair dice

E = sum is 7

F = 1^{st} die is 4

G = 2^{nd} die is 3

$E \perp\!\!\!\perp F \quad E \perp\!\!\!\perp G \quad F \perp\!\!\!\perp G$

G = 2nd die is 5

Then $E \perp\!\!\!\perp F$, $E \perp\!\!\!\perp G$, $F \perp\!\!\!\perp G$

checked before — similarly — throws are independent

$\Rightarrow$ pairwise independence.

$$P(E) = \frac{1}{6} \neq P(E \mid F \cap G) = 1$$

$$\Rightarrow \quad E \not\!\perp\!\!\!\perp F \cap G. \qquad //$$

So pairwise independence is <u>not</u> enough!

<u>Def</u> Three events E, F, G are <u>independent</u> if $P(E \cap F \cap G) = P(E)P(F)P(G)$

<u>and</u>

$$\left.\begin{aligned} P(E \cap F) &= P(E)P(F) \\ P(E \cap G) &= P(E)P(G) \\ P(F \cap G) &= P(F)P(G) \end{aligned}\right\} \text{pairwise independence}$$

<u>example</u> (illustration that this definition is better)

Assume E, F, G independent,

? $E \perp\!\!\!\perp (F \cup G)$?

$$P(E \cap (F \cup G)) = P((E \cap F) \cup (E \cap G))$$

$$= P(E\cap F) + P(E\cap G) - P(E\cap F\cap G)$$

$$= P(E)P(F) + P(E)P(G) - P(E)P(F)P(G)$$

$$= P(E)\{\underbrace{P(F) + P(G) - P(F)P(G)}_{P(F\cup G)}\}$$

$$= P(E)P(F\cup G)$$

$$\Rightarrow E \perp\!\!\!\perp (F\cup G)\,!$$

And this will hold for any events formed from different subcollections among E, F, G.

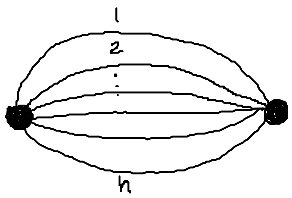
Def For any finite collection of events $F_1, F_2, \dots, F_n$ we say that they are independent if for every subset

$$\{i_1, i_2, \dots, i_k\} \subset \{1, \dots, n\} \quad (1 \le k \le n)$$

$$P(F_{i_1}\cap F_{i_2}\cap \dots \cap F_{i_k}) = P(F_{i_1})P(F_{i_2})\dots P(F_{i_k})$$

---

example 1) A system of n components is said to be parallel if it functions as long as at least one component is functional.



Assume that component i functions with probability p_i, $1 \le i \le n$.

What's the chance that this system is functional?

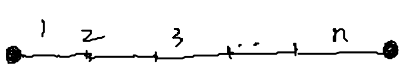
A_i = component i is functional.

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - P\left(\left(\bigcup_{i=1}^{n} A_i\right)^c\right)$$

$$\overset{\text{de Morgan}}{=} 1 - P\left(\bigcap_{i=1}^{n} A_i^c\right)$$

$$= 1 - \prod_{i=1}^{n} P(A_i^c)$$

$$= 1 - \prod_{i=1}^{n} (1 - p_i).$$



Consequitive system

this system functions only if all components function.

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} p_i.$$

//

2) Roll 2 fair dice, independently, many times.

$\left\{ \begin{pmatrix} 1 \\ 6 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \ldots \right\}$

$\downarrow$

$\{ 7, 5, 8, \ldots \}$

What is the probability that sum of 5 appears before sum of 7?

E = sum of 5 before sum of 7

E_n = event that neither 5 nor 7 appear in the first $(n-1)$ rolls <u>and</u> 5 appears on the n^{th} trial.

(e.g. $\{ 6, 8, 12, 5, \ldots \}$)

$\uparrow$

E_4

$$E = \bigcup_{n=1}^{\infty} E_n \,,$$

E_i's are disjoint (disjoint events are dependent!)

$$P(E) = \sum_{n=1}^{\infty} P(E_n)$$

$$P(E) = \sum_{n=1} P(E_n)$$

(1, 4) (2, 3) (3, 2) (4, 1)

$$P(E_n) = \left(\frac{26}{36}\right)^{n-1} \cdot \frac{4}{36} = \left(\frac{13}{18}\right)^{n-1} \cdot \frac{1}{9}$$

geometric series:

$$\boxed{\text{If } |a| < 1 \text{ then } \sum_{n=0}^{\infty} a^n = \frac{1}{1-a}}$$

$$P(E) = \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} \cdot \frac{1}{9}$$

$$= \frac{1}{9} \cdot \sum_{k=0}^{\infty} \left(\frac{13}{18}\right)^{k}$$

$$= \frac{1}{9} \cdot \frac{1}{1 - \frac{13}{18}} = 2/5 \quad //$$

Another approach

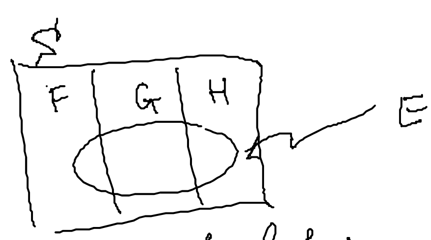
Let E = sum of 5 occurs before sum of 7

F = 1^{st} trial yields a sum of 5

G = 1^{st} trial yields a sum of 7

H = 1^{st} trial yields neither 5 nor 7

$$F \cup G \cup H = S$$



according to the law of total probability

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

$$= 1 \cdot \frac{1}{9} + 0 \cdot P(G) + P(E) \cdot \frac{13}{18}$$

$$P(E) \cdot \frac{5}{18} = \frac{1}{9} \Rightarrow P(E) = \frac{2}{5} \quad //$$

---

example: You are playing a game to win a big prize. You have a choice

1) win the prise by shooting basket ball once & making it;

or

2) shoot 3 times & win the prize if you make it at least 2 times.

Which option should you choose?

"Make" any given shot w/ prob. p, $0 \le p \le 1$.

① $P(\text{win}) = p$

② $P(\text{win}) = P(SSS \cup SSF \cup SFS \cup FSS)$

$$= P(SSS) + P(SSF) + P(SFS) + P(FSS)$$

$$= p^3 + 3p^2(1-p)$$

Compare $\quad p^3 + 3p^2(1-p) \quad ? \quad p$

$$p^2 + 3p - 3p^2 \quad ? \quad 1$$

$$0 \quad ? \quad 2p^2 - 3p + 1$$ (compare @ your leisure)

Calculus .... shows if $p \in [0, \frac{1}{2})$

$\Rightarrow \quad p > p^3 + 3p^2(1-p)$

if $\quad p \in (\frac{1}{2}, 1]$

$\Rightarrow \quad p < p^3 + 3p^2(1-p)$

$\Rightarrow$ option is for weaker player //