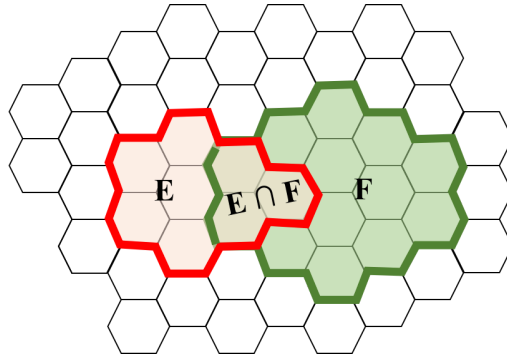


# Law of Total Probability

An astute person once observed that when looking at a picture, like the one we saw for conditional probability:



that event  $E$  can be thought of as having two parts, the part that is in  $F$ , ( $E$  and  $F$ ), and the part that isn't, ( $E$  and  $F^C$ ). This is true because  $F$  and  $F^C$  are (a) mutually exclusive sets of outcomes which (b) together cover the entire sample space. After further investigation this proved to be mathematically true, and there was much rejoicing:

$$P(E) = P(E \text{ and } F) + P(E \text{ and } F^C)$$

This observation proved to be particularly useful when it was combined with the chain rule and gave rise to a tool so useful, it was given the big name, law of total probability:

$$\begin{aligned} P(E) &= P(E \text{ and } F) + P(E \text{ and } F^C) \\ &= P(E|F)P(F) + P(E|F^C)P(F^C) \end{aligned}$$

## The Law of Total Probability (LOTP)

If we combine our above observation with the chain rule, we get a very useful formula the Law of Total Probability of LOTP for short:

$$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

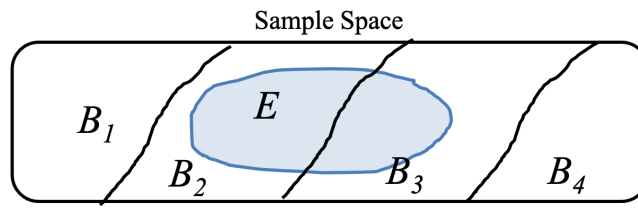
There is a more general version of the rule. If you can divide your sample space into any number of [mutually exclusive](#) events:  $B_1, B_2, \dots, B_n$  such that every outcome in the sample space falls into one of those events, then:

$$\begin{aligned} P(E) &= \sum_{i=1}^n P(E \text{ and } B_i) && \text{Extension of our observation} \\ &= \sum_{i=1}^n P(E|B_i)P(B_i) && \text{Using chain rule on each term} \end{aligned}$$

## Generalization to Many Background Events

The events  $F$  and  $F^C$  are always mutually exclusive and they always cover the entire sample space, no matter what  $F$  represents! If you can find more than two background events that are also mutually exclusive, and their union covers the entire sample space (the universe of outcomes) then you can use the generalized version of the law of total probability.

To generalize the law of total probability, imagine we can divide the sample space into several mutually exclusive background events ( $B_1, B_2, \dots, B_n$ ), where these sets cover the entire sample space. In this case, any event  $E$  can be decomposed by considering the likelihood of  $E$  within each of these disjoint sets.



In the image above, you could compute  $P(E)$  to be equal to

$$P \left[ (E \text{ and } B_1) \text{ or } (E \text{ and } B_2) \text{ or } \dots \text{ or } (E \text{ and } B_n) \right]$$

There are many real world cases where (a) it is much easier to think of the probability of an event  $E$  in the context of a background event  $B_i$  and the sample space can be discretized into several mutual exclusive background events  $B_i$ . Lets start with an example with three events  $B_1, B_2, B_3$ . Suppose you are trying to determine the likelihood that a randomly selected individual will test positive for a certain disease,  $P(E)$ . The population can be divided into three mutually exclusive groups :

1.  $B_1$ : Individuals who are high-risk (e.g., individuals with a known exposure to the disease),
2.  $B_2$ : Individuals who are medium-risk (e.g., individuals with a family history of the disease but no direct exposure),
3.  $B_3$ : Individuals who are low-risk (e.g., the general population without known risk factors).

Each of these groups has a different probability of testing positive for the disease, and the total probability of a random individual testing positive can be broken down as follows:

$$\begin{aligned}
 P(E) &= P(E \text{ and } B_1) + P(E \text{ and } B_2) + P(E \text{ and } B_3) && \text{LOTP} \\
 &= P(E|B_1)P(B_1) + P(E|B_2)P(B_2) + P(E|B_3)P(B_3) && \text{Chain Rule} \\
 &= \sum_{i=1}^3 P(E | B_i)P(B_i) && \text{Sum Notation}
 \end{aligned}$$

Where:

- $P(E|B_1)$  is the probability of testing positive given someone in the high-risk group
- $P(E|B_2)$  is the probability of testing positive given someone in the medium-risk groups,
- $P(E|B_3)$  is the probability of testing positive given a person is in the low-risk group.
- $P(B_1)$ ,  $P(B_2)$  and  $P(B_3)$  are the probabilities of a person being in the high-risk, medium-risk and low-risk groups.

This works because everyone can belongs to one of the background events ( $B_1, B_2, B_3$ ), in other words the sets span the sample space. Moreover each person is in only one of the sets, and as such they are mutually exclusive. It is helpful because its easier to think of the probability of  $E$ , testing positive, in the context of the background events, where you know how at risk the patient is.