CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 5:

- Conditional Probability concluded;
- Dependence vs Causality;
- Base Rate Fallacy;
- Law of Total Probability;
- Bayes Rule

Review: Independence and Dependence

We say that two events A and B are independent if

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

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

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

or:

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

Example:

Suppose in a particular city, 40% of the population is male, and 60% female, and 20% of the population smokes. If male smokers are 8% of the population, then are smoking and gender independent? That is, are the following two events independent?

A = Smoker

B = Male

Independence and Dependence

We say that two events A and B are independent if

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

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

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

or:

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

Example:

Suppose in a particular city, 40% of the population is male, and 60% female, and 20% of the population smokes. If male smokers are 8% of the population, then are smoking and gender independent? That is, are the following two events independent?

YES. Check:

A = Smoker

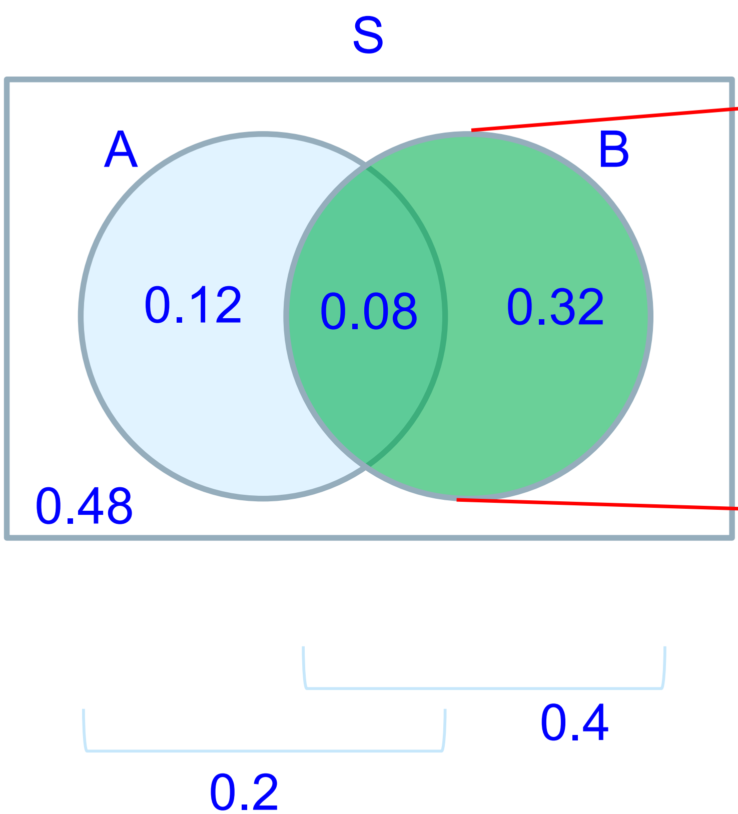
$P(A \cap B) = 0.08 = 0.4 * 0.2 = P(A) * P(B)$

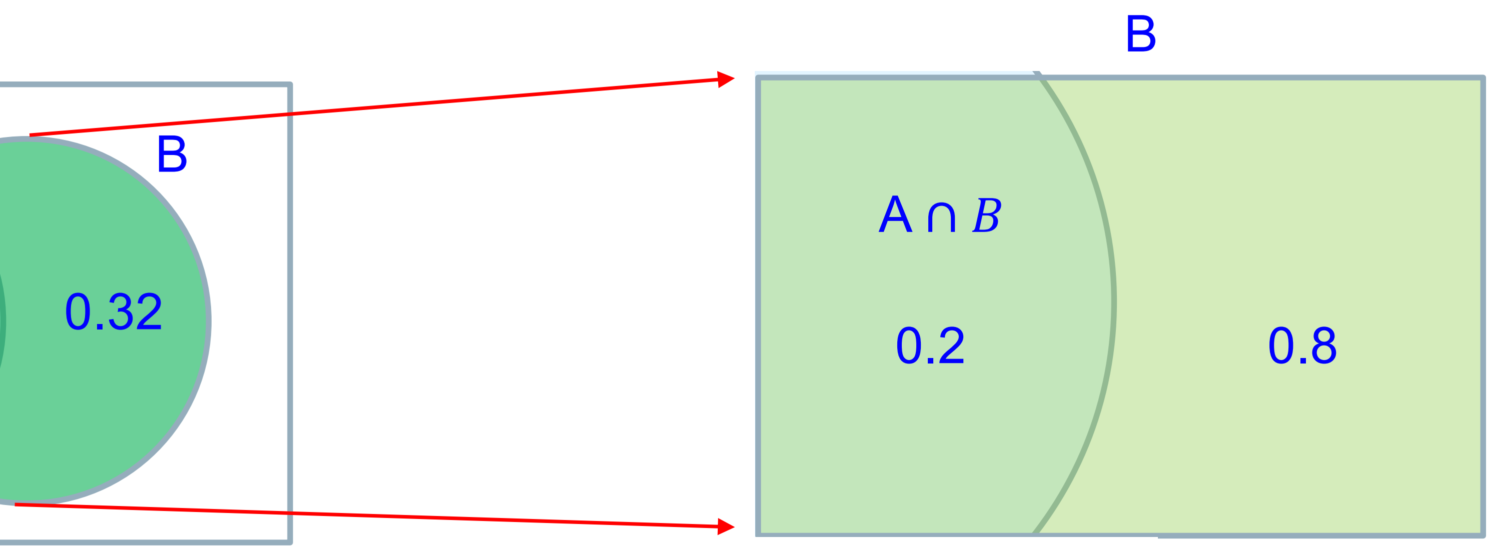
B = Male

Independence and Dependence

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

Example: Suppose in a particular city, 40% of the population is male, and 60% female, and 20% of the population smokes. If male smokers are 8% of the population, then are smoking and gender independent? That is, are the following two events independent? A = Smoker B = Male





P(A ∩ B) = 0.08 = 0.4 * 0.2 = P(A) * P(B)

Independence and Dependence

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

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

Some important results about independence....

0. Independence does NOT mean the events are disjoint: If two non-empty events are disjoint, then they are **dependent:**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\varnothing)}{P(B)} = \frac{0}{P(B)} \neq P(A)$$

1. Independence is symmetric:

$$P(A \mid B) = P(A) \quad \Longleftrightarrow \quad \frac{P(A \cap B)}{P(B)} = P(A) \quad \Longleftrightarrow \quad \frac{P(B \cap A)}{P(A)} = P(B) \quad \Longleftrightarrow \quad P(B \mid A) = P(B)$$

Independence and Dependence

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

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

Some important results about independence....

2. **Independence extends to complements:**

A and B are independent

iff A^c and B are independent

iff A and B^c are independent

iff A^c and B^c are independent

Intuitively: Independence means information about A does not give you any information about B.

Intuition: If knowing whether a person is male gives you no information about whether the person smokes, then

- Knowing if they are female gives you no information about whether they don't smoke;

- Knowing if the don't smoke gives you no information if they are male, etc., etc.

Independence and Dependence

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

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

Some important results about independence....

3. **Independence is can be generalized to more than 2 events**, for example, events **A, B, and C are mutually independent** iff

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

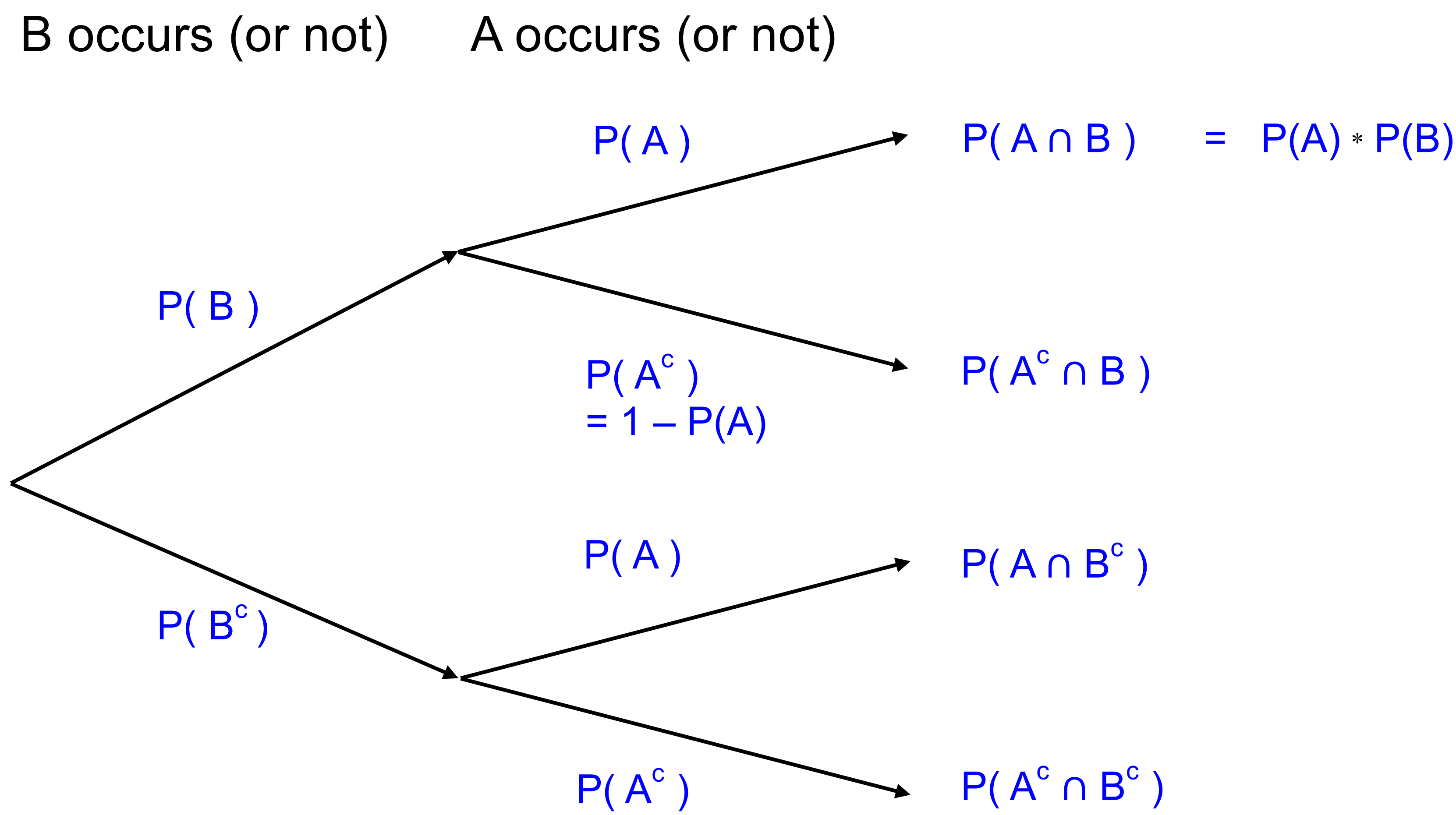
Simple Example: A = 1st coin flip is heads, B = 2nd is tails, C = 3rd is heads.

A more complex example is given in Problem 34 in the End-of-Chapter problems for Chapter 1, where it is shown that pair-wise independence does not imply mutual independence. But: mutual independence implies pair-wise independence.

Independence and Dependence

How does this relate to tree diagrams?

When the events are independent, then we have the familiar tree diagram in which we simply write the probabilities of the events on each arc:



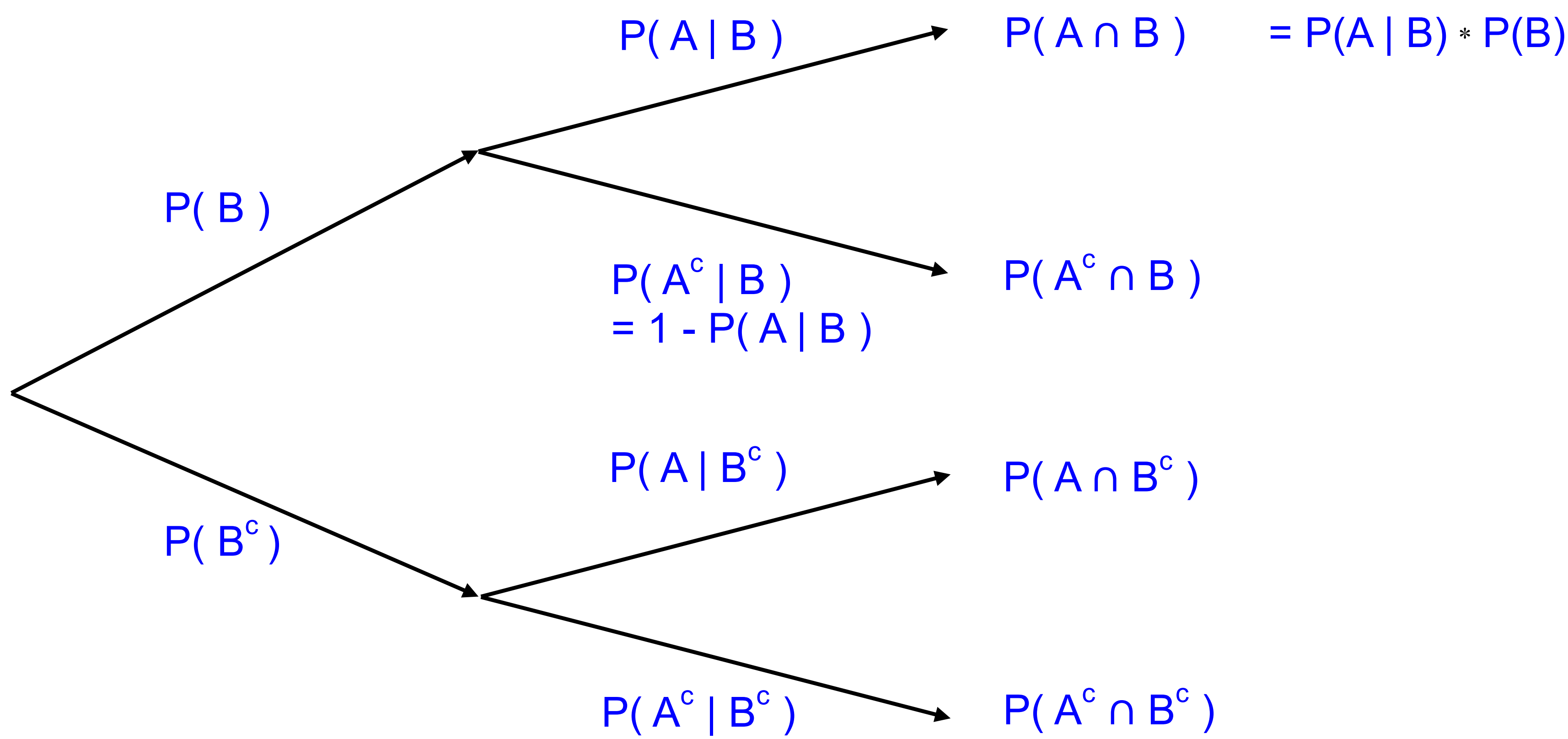
Independence and Dependence

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

When the events are NOT independent, it is more complicated…

P(A | B) considers an event B followed by an event A, and how the occurence of B affects the occurence of A. What are the labels on a tree diagram of this random experiment?

B occurs (or not) A occurs (or not)



Independence and Dependence

Digression: Dependence does not imply causality!

The NEW ENGLAND JOURNAL of MEDICINE

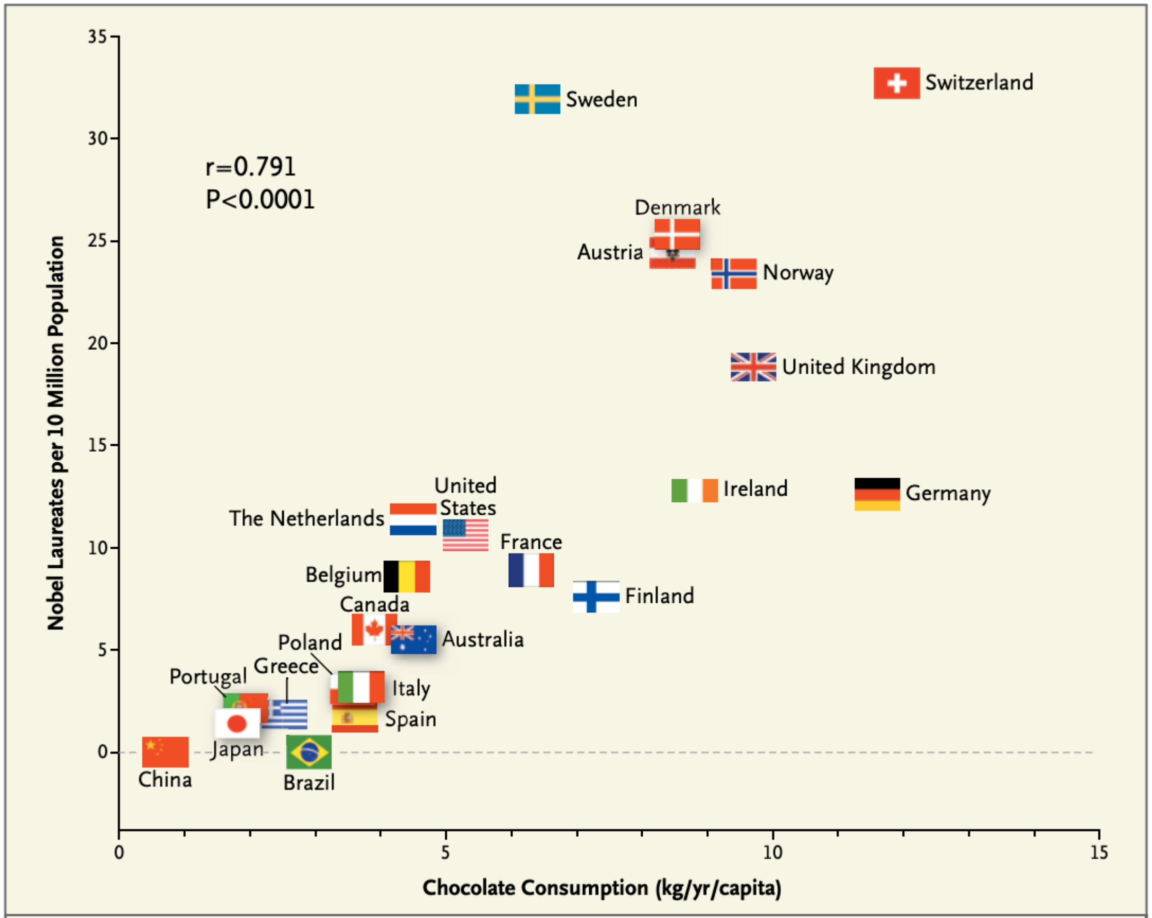


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Base Rate Fallacy

Conditional Probability is an excellent tool for evaluating what can happen under various conditions, but it is **sensitive to extreme conditions**.

Traffic police have breathalyzers to detect when people are legally drunk, but they are not perfect. In 5% of the cases, they give "false positives," meaning they say a person is drunk when they are in fact sober. However, they ALWAYS detect a truly drunk person (there are no "false negatives")

Suppose 1 in 1000 drivers is driving drunk, and the police officers stop a driver at random, and give the driver a breathalyzer test. It indicates that the driver is drunk. We assume you don't know anything else about him or her.

What is the probability he or she really IS drunk?

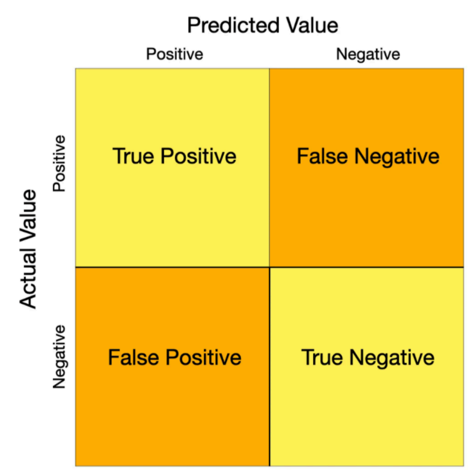
95%

50%

20%

2%

Confusion Matrix



Base Rate Fallacy

Actual Value \ Predicted Value	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Conditional Probability is an excellent tool for evaluating what can happen under various conditions, but it is sensitive to extreme conditions.

Traffic police have breathalyzers to detect when people are legally drunk, but they are not perfect. In 5% of the cases, they give "false positives," meaning they say a person is drunk when they are in fact sober. However, they ALWAYS detect a truly drunk person (there are no "false negatives")
Suppose 1 in 1000 drivers is driving drunk, and the police officers stop a driver at random, and give the driver a breathalyzer test. It indicates that the driver is drunk. We assume you don't know anything else about him or her.

What is the probability he or she really is drunk? (-Wikipedia)

2%

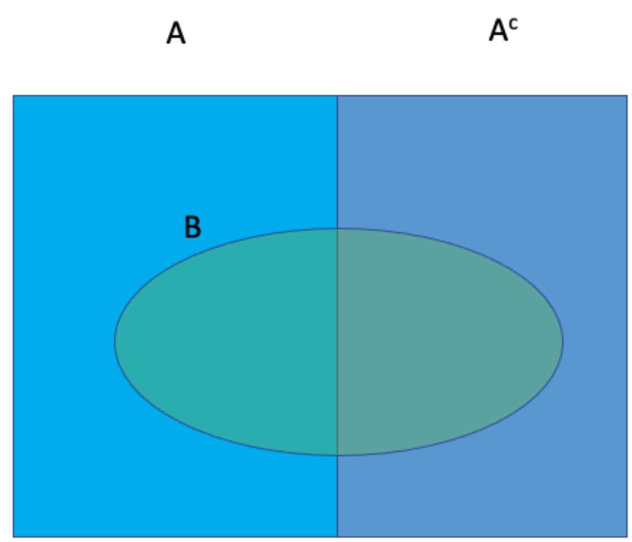
Suppose we consider 1000 "trials" of this experiment. On average, 1 driver is drunk, and it is 100% certain that for that driver there is a *true* positive test result, so there is 1 *true* positive test result.

But 999 drivers are not drunk, and among those drivers we have 5% *false* positive test results, so there are 49.95 *false* positive test results.

Therefore, the probability that one of the drivers among the 1 + 49.95 = 50.95 positive test results really is drunk is 1/50.95 = 0.0197 or about 2%.

Law of Total Probability

Conditional Probability addresses the probability of one event in the context of one other event. Essentially, it is case analysis: What is the probability of event B in two cases, A and A^c ?



Law of Total Probability

Conditional Probability addresses the probability of one event in the context of one other event. Sometimes, it gets more complicated, because there are many different events. The **Law of Total Probability** can help!

Consider this problem (from the textbook):

I have three bags that each contain 100 marbles:

Bag 1 has 75 red and 25 blue marbles;

Bag 2 has 60 red and 40 blue marbles;

Bag 3 has 45 red and 55 blue marbles;

I choose a bag at random, and then a marble at random from that bag. What is the probability that I get a red marble?

Events: A = Marble is red

B1 = Bag 1 was chosen

B2 = Bag 2 was chosen in the first step

B3 = Bag 3 was chosen

Law of Total Probability

I have three bags that each contain 100 marbles:

Bag 1 has 75 red and 25 blue marbles;

Bag 2 has 60 red and 40 blue marbles;

Bag 3 has 45 red and 55 blue marbles;

I choose a bag at random, and then a marble at random from that bag. What is the probability that I get a red marble?

Events: A = Marble is red

B1 = Bag 1 was chosen

B2 = Bag 2 was chosen in the first step

B3 = Bag 3 was chosen

$P(A \mid B1) = 75/100 = 0.75$

$P(A \mid B2) = 60/10 = 0.6$

$P(A \mid B3) = 45/100 = 0.45$

Law of Total Probability

The Law of Total Probability:

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i).$$

This is essentially conditional probability where the conditions form a partition of the sample space.

Or, simply think of it as "case analysis." (you've probably be doing this already without calling it anything special!).

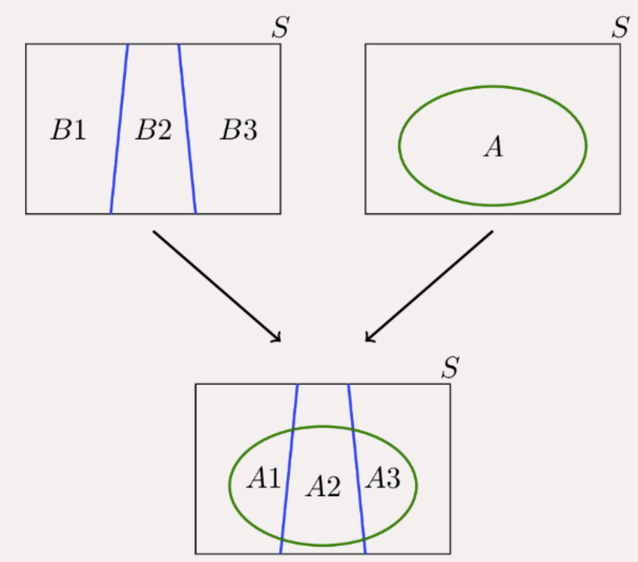


Fig.1.24 - Law of total probability.

Law of Total Probability

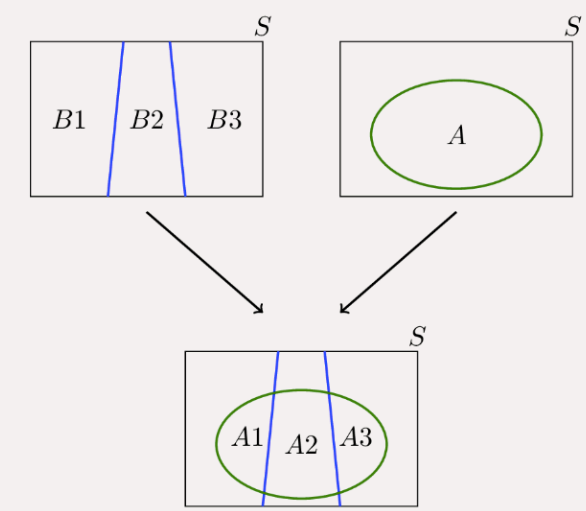


Fig.1.24 - Law of total probability.

I have three bags that each contain 100 marbles:

Bag 1 has 75 red and 25 blue marbles;

Bag 2 has 60 red and 40 blue marbles;

Bag 3 has 45 red and 55 blue marbles;

I choose a bag at random, and then a marble at random from that bag. What is the probability that I get a red marble?

Events: A = Marble is red

B1 = Bag 1 was chosen

B2 = Bag 2 was chosen in the first step

B3 = Bag 3 was chosen

P(A | B1) = 75/100 = 0.75

P(A | B2) = 60/10 = 0.6

P(A | B3) = 45/100 = 0.45

P(A) = P(A | B1) * P(B1) + P(A | B2) * P(B2) + P(A | B3) * P(B3)

= 0.74 / 3 + 0.6 / 3 + 0.45 / 3 = 0.6

Bayes' Rule

We can rearrange the conditional probability rule in a way that makes the sequence of the events irrelevant -- which happened first, A or B? Or did they happen at the same time? Does it matter?

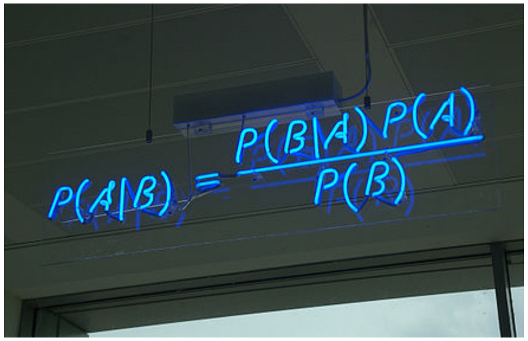
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

We can do a little algebra to define conditional probabilities in terms of each other:

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

so:

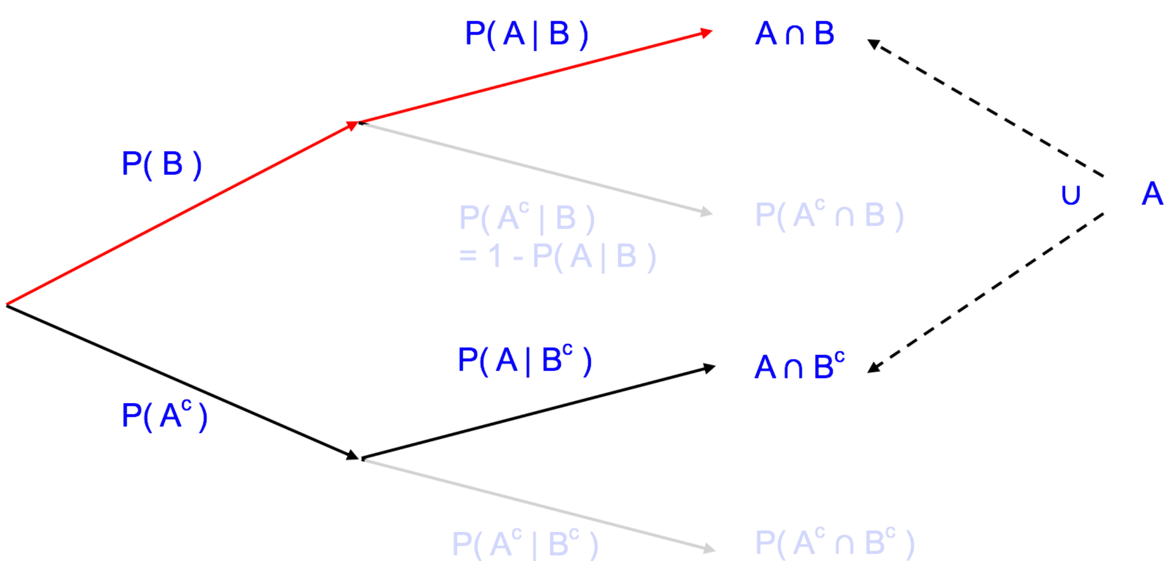
$$P(B \mid A) = \frac{P(A \mid B) * P(B)}{P(A)}$$



Bayes' Rule

The best way to understand this is to view it with a tree diagram!

P(B | A) = the probability that when A happens, it was "preceeded" by B:



If A has happened, what is the probability that it did so on the path where B also occurred?

Note:

$$A = P(A \cap B) \cup P(A \cap B^c)$$

So what percentage of A is due to $A \cap B$?

Same calculation as:

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

Bayes' Rule

This has an interesting flavor, because we can ask about causes of outcomes:

A Priori Reasoning -- "I randomly choose a person and observe that he is male; what the probability that it is a smoker?"

"The first toss of a pair of dice is a 5; what is the probability that the total is greater than 8?"

A Posteriori Reasoning -- "I find a cigarette butt on the ground, what is the probability that it was left by a man?"

"The total of a pair of thrown dice is greater than 8; what is the probability that the first toss was a 5?"

This seems odd, because instead of reasoning forward from "causes to effects" we are reasoning backwards from "effects to causes" but really it is just different ways of phrasing the mathematical formulae. Time is not really relevant!