Conditional Probability

Data Science Immersive

February 27, 2019

FLATIRON SCHOOL

Conditional Probability

- Agenda today…
  - Review independent probability
  - Learn Dependent Probability
    - Theorems
    - Examples
  - Bayes’ Theorem & Bayesian Statistics
    - Derivation
    - Examples

After today, you'll be able to...

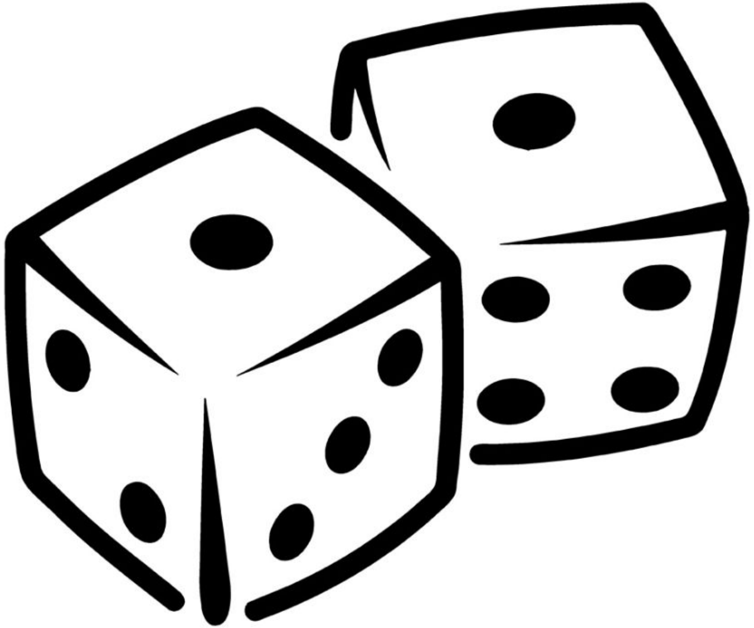
- Understand and explain the difference between independent and dependent events
- Gain an intuitive understanding of why conditional probability is necessary
- Calculate and compute conditional probabilities in given context
- Calculate conditional probabilities using Bayes' theorem
- Explain the difference between frequentist framework and Bayesian

Review

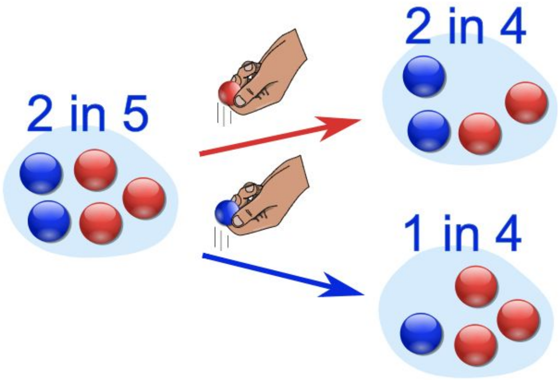
- Event
  - An **event** is the outcome of a random experiment
- Sample space
  - A **sample space** is a collection of every single possible outcome in a trial
- Independent probability is calculated by event divided by all the possible events in the sample space
  - The occurrence of one event does not affect, or dependent on the outcome of another

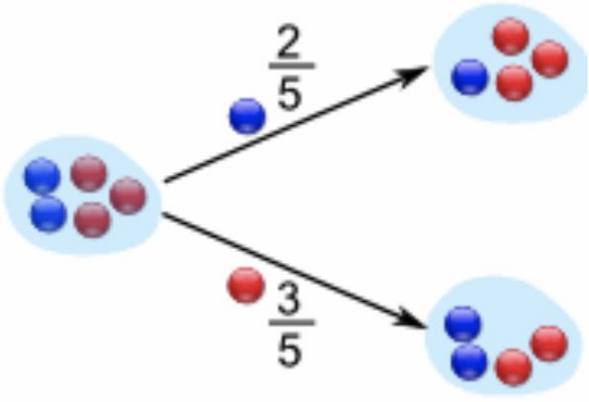
Review

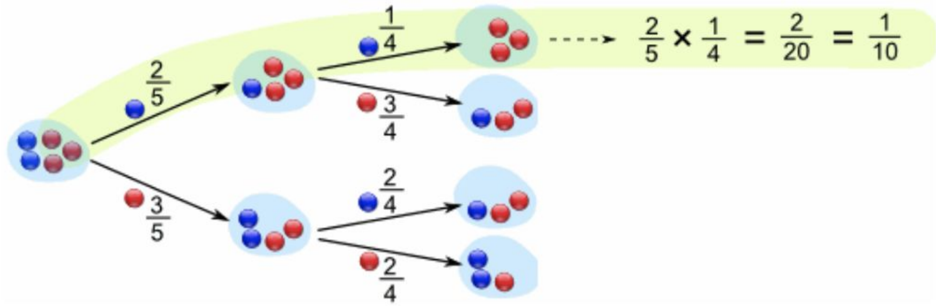
- Independent probability of an event is calculated as P(A) divided by **all** possible events.
- Some statistical distributions make such assumptions about events
  - Poisson distribution
  - Binomial distribution



Conditional Probability

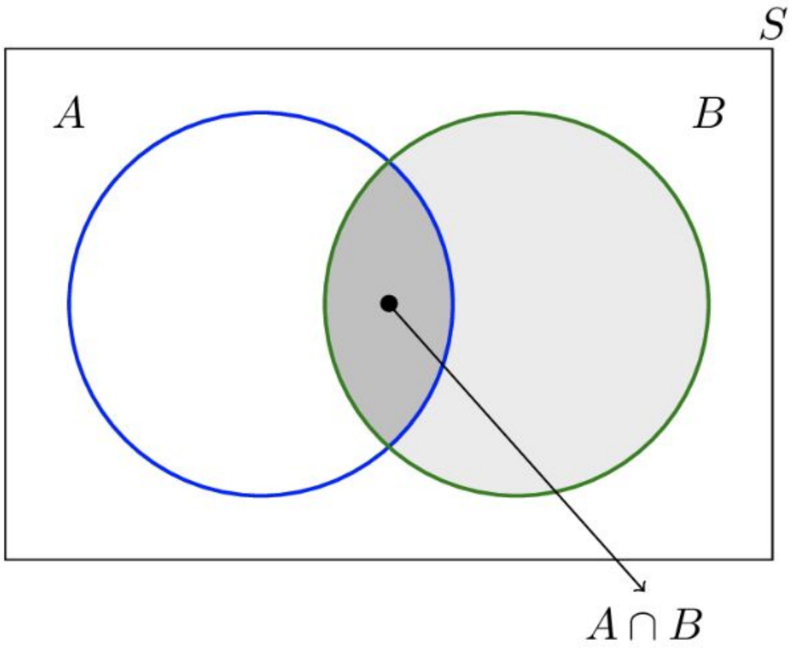






Conditional Probability

- Conditional Probability emerges in the examination of experiments where a result of a trial may influence the results of the upcoming trials. For example:
    - The probability of drawing an Ace given already drew an Ace
    - The probability of being in a good mood given the weather is nice



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

Example 1

The percentage of adults who are men and alcoholic is 2.25%. What is the probability of being an alcoholic given that someone is a male?

$P(A \mid M) = P(A \,\&\, M) / P(M) = 2.25\% / 50\% = 5\%$

Example 2

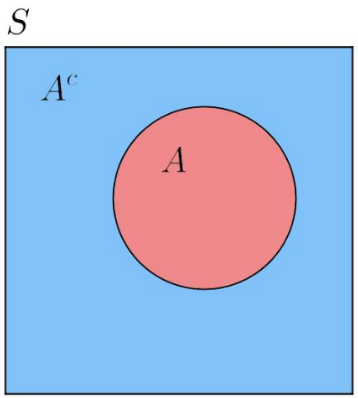
You are going to visit your distant cousins who recently had two children. You are told that at least one of them is a girl. What is the probability of both of them being girls?

Example 3

You are going to visit your distant cousins who recently had two children. You are told that the older one is a girl. What is the probability of both of them being girls?

Theorems of Conditional Probability

1. $P(A') + P(A) = 1$



Sample Space S, event A, and complement A^c

2. The Product Rule

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

The product rule is useful when the conditional prob is easy to compute but the intersection is not

Theorems of Conditional Probability

- Chain Rule
  - The chain rule (also called the general product rule) permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities.

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

- We can extend this to three groups, and get:

$$P(A \cap B \cap C) = P(A|B \cap C)\, P(B \cap C) = P(A|B \cap C)\, P(B|C)\, P(C)$$

- Which then can give us a more general format of

$$P(A1 \cap A2 \cap \ldots \cap An) = P(A1|A2 \cap \ldots \cap An)P(A2|\, A3 \cap \ldots \cap An)P(An-1|An)P(An)$$

Bayes' Theorem

- Bayes' Theorem underlies the foundation of Bayesian Inference, an incredibly powerful way in which statistics are probability are computed
- It is derived from conditional probability

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

$$P(A \cap B) = P(A|B)\, P(B) = P(B|A)\, P(A)$$

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

Bayes' Theorem

P(A) is called the **prior**; this is the probability of our hypothesis without any additional prior information. It could also be a belief we have prior to seeing the data.

P(B) is called the **marginal likelihood**; this is the total probability of observing the evidence. In many applications of Bayes Rule, this usually serves as normalization constant, which we will explain in further detail.

P(B|A) is called the **likelihood**; this is the probability of observing the new evidence, given our initial hypothesis.

P(A|B) is called the **posterior**; this is what we are trying to estimate.

Bayes’ Theorem

Why Bayes’?

- Bayes’ theorem allows us to accommodate degrees of belief into the equation, which accounts for uncertainty
- Has practical implication in natural language processing--i.g. Naive bayes for classification or topic modeling, such as Latent Dirichlet Allocation
- Stochastic methods for parameter estimation, i.g. Markov Chain Monte Carlo
- Network analysis or graph theory
- And many more applications in philosophy, cognitive sciences, even legal systems

Example 1

Assume that the probability of having lung cancer is 2%, and the probability of being a smoker is 15%. Empirically, we know that 20% of the people who have lung cancer are smokers. What is the probability of having lung cancer given you are a smoker?

$P(\text{Cancer}) = 0.02$

$P(\text{Smoker}) = 0.15$

$P(\text{Smoker} \mid \text{Cancer}) = .20$

$P(\text{Cancer} \mid \text{Smoker}) = P(\text{Cancer}) * P(\text{Smoker} \mid \text{Cancer}) \; / \; P(\text{Smoker})$

$= 0.02 * 0.20 / 0.15 = 0.0267 = 2.67\%$

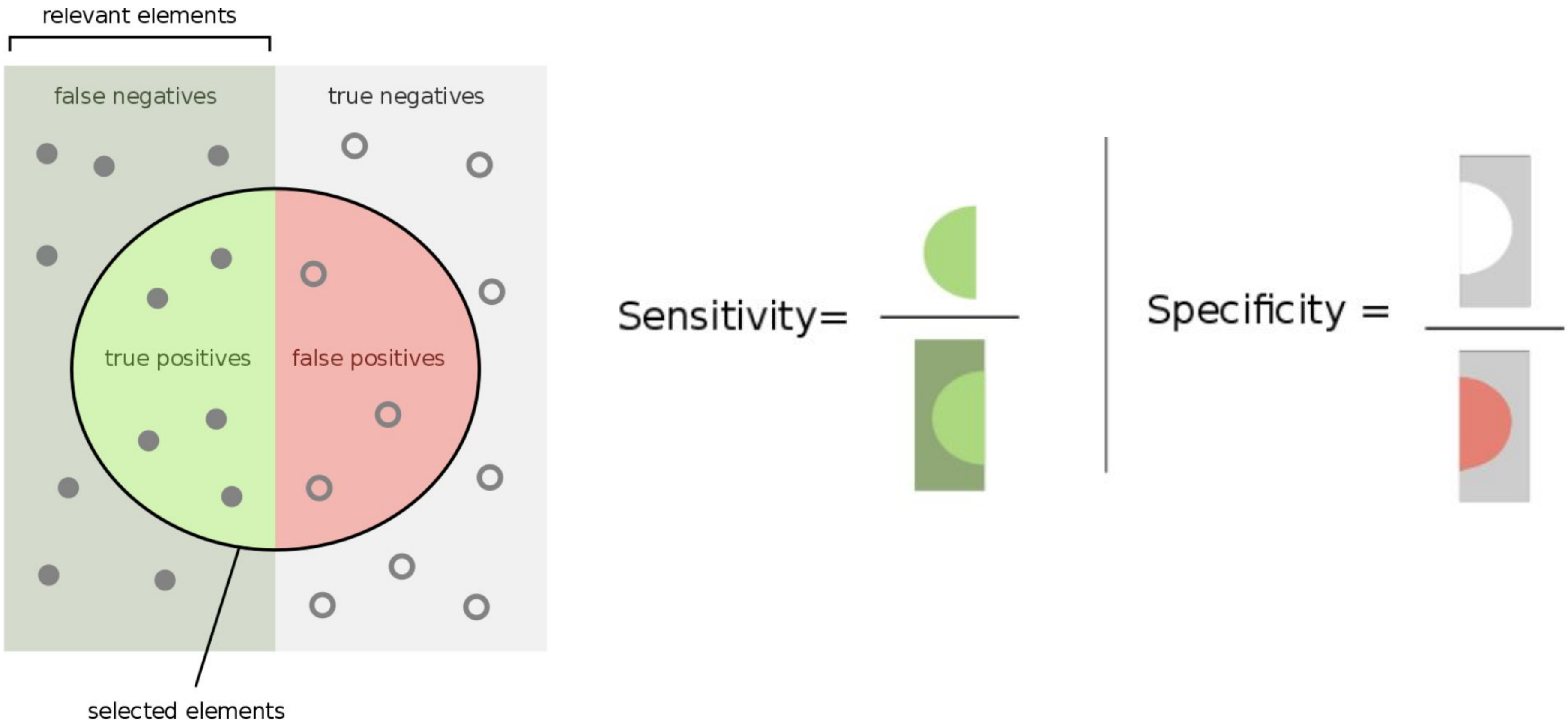
Example 2 - False Positives

- You tested positive for a disease with 5% false positive rate
- The probability of getting a positive test when you have the disease is 90%
- We know in the population 1% of people have this disease
- What is the probability of having this diseases given that you tested positive?

Example 2 Continued - Bayesian updating

What is the probability of you having the disease given that you tested positive twice?

Sensitivity vs. Specificity



Frequentist vs. Bayesian

- Frequentist statisticians rely on the imaginary sampling of an infinite population and derive a probability value that summarizes the result of the experiment
  - Inference made by a frequentist statistician only depends on the frequency of events, or samples observed
- Bayesian statisticians not only rely on actual evidence observed, but also on be