Kuan-Ting Chen
Victoria Delk
Aaron Lee
Charles Zhang

Stats 100A Homework #1

Part I

a) Law (Victoria Delk)

This article focuses on a political exchange between two people. The first person claims that "Every racist and anti-Semite in the country pretty much probably voted for Brexit", while the second paraphrases this sentiment as meaning, "You've basically tried to slur anybody who voted Leave as a bigot." For a casual listener, these statements may sound equivalent in meaning, however, there is much more nuance to the situation. In reality, the first statement is better understood as meaning that "all racists and anti-Semites voted to leave", while the second statement is closer in meaning to "all people who voted to leave are racists and anti-Semites." The tendency to equate these two statements to one another is known as the prosecutor's fallacy, and this article focuses on its capability to alter societal perceptions within the realm of politics.

Mathematically, the prosecutor's fallacy is tied to the relationship between the conditional probabilities of two events. In the context of the article, these two events can be described as `RaS` (the person is a racist and anti-Semite) and `L` (the person voted to leave), within the sample space `S` (the voting-eligible British population). This means that the point of concern is the difference between $P(L \mid RaS)$ (the first statement) and $P(RaS \mid L)$ (the second statement). Using Bayes' Theorem, we can see that these two conditional probabilities are related by $P(RaS \mid L) = [P(L \mid RaS)P(RaS)] / P(L)$. The first statement claims that all racists and anti-Semites voted to leave, meaning that the conditional probability $P(L \mid RaS) \approx 1$. On the other hand, the second statement claims that all Leavers are racists and anti-Semites, or $P(RaS \mid L) \approx 1$. However, the problem is that the prior probabilities are left unaccounted for. The article goes on to show that, assuming $P(L \mid RaS) \approx 1$ and that $P(L) = 0.51$ (based on referendum data), the only way for $P(RaS \mid L) \approx 1$ is for the prior probability $P(RaS) \approx 0.5$. In other words, statement 2 can only be true given statement 1 if about half of Britain's population was indeed racist and anti-Semetic, and therefore, the equating of these two statements is clearly misleading and an excellent demonstration of the misuse of statistics in the case of the prosecutor's fallacy.

b) Representativeness Heuristic (Charles Zhang)

This video discusses how statistics plays a role in our use of the representativeness heuristic. As humans, we have a limit to our rationality, bounded by resources and computing power, and therefore, we must resort to tactics such as the representativeness heuristic in our decision-making. When using this heuristic, we subconsciously replace the question, "how likely is it for `X` to be `Y`?" with, "how much does `X` represent a stereotypical `Y`?" The issue with this replacement is that it results in base rate neglect, where we tend to ignore the likelihood of Y occurring in the first place.

From a statistics perspective, base rate neglect can be equated to a tendency to ignore the prior probabilities of an event occurring. In the scenario described in the lecture, we want to know how likely it is that Spongebob is tough. The relevant events to this problem are the event that Spongebob is tough (B) and the event that a random resident of Bikini Bottom is tough (T). By the definition of a joint probability, we can see that $P(B, T) = P(B \mid T)P(T) = P(T \mid B)P(B)$. Accounting for base rate neglect, the prior probabilities would be removed from the equation and we'd end up with $P(B \mid T) = P(T \mid B)$. In other words, the conditional probability that Spongebob is tough is being incorrectly equated with the conditional probability that a person resembles Spongebob, and is therefore tough. As discussed in part a), this result is mathematically equivalent to the issue described by the prosecutor's fallacy, where, in reality, we must account for the prior probabilities of the events occurring, as the events are not independent. This misconception shows us how the representativeness heuristic can fail us, and how it is important to keep prior probabilities in mind in the act of decision-making.

c) Artificial Intelligence (Kuan-Ting Chen)

This article by Professor Piech talks about the concept of fairness in machine learning. He provides three different definitions of fairness using the example of loan repayment and demographics. D denotes the demographic indicator, G stands for the prediction the algorithm made, and T stands for the actual outcome about whether or not the loan was repaid. The first is parity, or whether or not the algorithm will make the same prediction regardless of the person's demographics. The author found that $P(G = 1 \mid D = 1) \neq P(G = 1 \mid D = 0)$ which means that parity is not satisfied, meaning that the algorithm was not able to make correct predictions regardless of demographic.

The next definition of fairness is calibration, or whether or not the predicted outcome is the same as the actual outcome under a condition. For example, the author found that $P(G = T \mid D = 1) \neq P(G = T \mid D = 0)$. In this case, the probability of getting a correct match between the guess and actual outcome were different depending upon the different conditions so the calibration definition is not satisfied.

The last definition of fairness that the author provides is called the equality of odds, which is that given the actual outcome and demographic, what is the probability that the algorithm will make a correct prediction. In this case, the result was that $P(G = 1 \mid D = 1, T = 1) = P(G = 1 \mid D = 0, T = 1)$, meaning that equality of odds is satisfied and that the algorithm was correct regardless of demographic. The author places extra emphasis on the fact that not all three of these definitions can be simultaneously satisfied, so it is important to determine which one is the best use for any case.

Piech uses another example of using machine learning to predict gender by image. The algorithm was much more likely to correctly guess the gender of lighter men than darker women: $P(G = T \mid D = \text{darker woman}) \neq P(G = T \mid D = \text{lighter man})$. The probability of the algorithm making a correct guess is different depending on the actual gender and the skin color of that person. Therefore, the calibration definition is not satisfied. The author exposed a deficit in training models because the majority demographic in a training set is more heavily weighted and can skew the accuracy of the model towards a certain demographic.

In regards to the misinterpretations presented in (a) and (b) in the form of the representative heuristic and prosecutor's fallacy, they can also appear in AI. In the case of machine learning, if a small, heavily biased dataset is used, then the results would be inaccurate as in the gender prediction example. This, like the misrepresentations, represents poor judgment made based on limited information.

d) Probability Models in AI (Aaron Lee)

Generative models create new instances of data, while discriminative models compare different instances of data. Hence, generative models evaluate the joint probability $P(x, y)$, where x represents the set of data instances and y, the labels. In contrast, discriminative models evaluate conditional probabilities: $P(y \mid x)$. Generative models inherently take the distribution of data into consideration when they assess the likelihood of a given example, whereas the latter assigns probabilities for each label. In generative adversarial models as in Ruiqi Gao's video, biased data sets can cause the model to generate or discriminate unfairly upon training. In the real world, bias exists in data, which is why the unsupervised training of generative models can inherently be biased as well. When $P(X \mid Y)$ does not equal $P(Y \mid X)$ misinterpretations arise.

Part II

a) Synthesis

- (a) and (b) are both discussing the prosecutor's fallacy, in which $P(X \mid Y)$ is misconstrued as being equivalent to $P(Y \mid X)$, essentially ignoring prior probabilities
    - (a) discusses this effect in the context of the manipulation of language in politics, resulting in the equating of $P(RaS \mid L)$ with $P(L \mid RaS)$, when they have entirely different interpretations
    - (b) discusses this effect through the lens of psychology, and how we subconsciously take shortcuts by ignoring the base rate of an event occurring in the first place (mathematically represented by prior probabilities $P(B)$ and $P(T)$)
    - By Bayes' Theorem ($P(X \mid Y) = P(Y \mid X)P(X) / P(Y)$), we can see how both of these misconstructions are fundamentally incorrect, as the prior probabilities $P(X)$ and $P(Y)$ are incorrectly asserted as $P(X) = P(Y)$, simplifying the equation to $P(X \mid Y) = P(Y \mid X)$

- Sources (c) and (d) discuss the use of probability and its applications to machine learning and artificial intelligence. In neural networks, which pertains to the last two sources, probability distributions are used to assess the likelihood of outcomes, as in image classification and labeling with discriminatory models
  - (c) emphasizes some of the ways that machine learning algorithms can be biased. For an algorithm to not be biased, it must be fair, which has three definitions: parity, calibration, and equality of odds. All three of these definitions boil down to the idea that regardless of the condition, the output of the algorithm should be the same. For example, `P(X | Y = 1)` should be the same as `P(X | Y = 2)` which is an example of parity. Machine learning models often fall into the trap of prosecutor's fallacy in that `P(X | Y) ≠ P(Y | X)` especially when the training data is biased; for example, a model might find the probability of an individual having a certain characteristic dependent on their race, when that is not the case
  - (d) Bayes' Theorem in the context of machine learning discriminatory models states that `P(X | Y) = P(XY) / P(Y)`, where `X` describes the event of an image being labeled with label `A`, while event `Y` describes the event that the image is actually object `A`
  - (d) In contrast, generative algorithms are governed by the concept of joint probability `P(XY)` since they capture correlations between data in a multidimensional space (associations such as trees usually have clusters of green and a brown trunk nearby)

b) Use in Statistics (Kuan-Ting Chen)

Source: https://machinelearningmastery.com/naive-bayes-for-machine-learning/

As a statistics major, conditional probability and Bayes' Theorem are important in that they provide the necessary foundation to not only calculate many different probabilities across different industries, but also to prevent misunderstandings from happening and being misused. One example is using conditional probability in conjunction with machine learning to classify binary classed items using a simplified version of Bayes' Theorem called Naive Bayes. This algorithm helps to find which category an item would belong to. But as powerful as machine learning is, and as with fairness in part (c), feeding the algorithm a good dataset is just as important, and the various definitions of fairness that involve conditional probability can help to ensure that the results are fair.

c) Use in Computer Science (Victoria Delk)

Source:

1. https://medium.com/appengine-ai/bayes-theorem-and-its-concepts-in-ai-409ec697b2d7
2. https://www.jigsawacademy.com/blogs/ai-ml/bayes-theorem-in-machine-learning

Within the field of computer science, conditional probability and Bayes' Theorem is extremely relevant in artificial intelligence and machine learning. Since machine learning deals with making decisions based on a growing dataset, Bayes' Theorem is used in order to more actually predict chances of an event happening, and is generally considered the most reliable, simple, and precise method of prediction available to us today. Using Bayes' Theorem, computer scientists have developed classifier models, such as Bayes optimal classifier and native Bayes classifier. These classifier models are used to predict a class based on the input data. With this technology we are able to perform image recognition, targeted adding, sentiment analysis, and more.

d) Use in Chemical Engineering (Aaron Lee)

Source:

1. https://pubmed.ncbi.nlm.nih.gov/18807909/#:~:text=The%20statistical%20methods%20are%20applied%20in%20chemical%20studies%20for%20data,biological%20properties%20of%20chemical%20compounds.
2. https://dspace.mit.edu/handle/1721.1/111286

Statistical analysis in the field of chemical engineering is crucial when it comes to quality assurance (QA) and quality control (QC). When working with large batches of products, chemical engineers need to assess the probabilities of a sample not meeting the passing acceptable qualifications outlined by the FDA. Additionally, statistical computation is employed during the maintenance routines of factories, equipment, and processes. Machine learning algorithms can take measurements of equipment functionality and assess the probabilities of machine failure given deviations in readings compared to an acceptable range. Using these techniques, a self-maintaining facility can be designed. Using sensor feedback as a means of regulating maintenance protocols is a form of Baysian probability: $P(X \mid Y)$ is a simple representation of complex maintenance models, where event X represents the probability of failure and event Y signifies the probability of a reading falling in a set range.

e) Use in Computer Science (Charles Zhang)

Source: https://machinelearningmastery.com/bayes-theorem-for-machine-learning/

In computer science, the primary use of conditional probabilities and Bayes' Theorem is in the realm of artificial intelligence and machine learning. One of the most common uses is as a Bayes classifier in machine learning, which labels objects in the agent's view with a probability that they are a specified object, like a dog or chair. Given a dataset and various heuristics to simplify computation, machine learning algorithms can implement Bayes' Theorem to accomplish this goal. Conditional probabilities can also be used to model casual relationships in belief networks, which can be used to streamline logical agents in cases where it is impossible to find a provably correct solution to a problem. Overall, probability provides computer scientists with a rational way to model decision-making under uncertainty.