NUMB3RS Activity: Filtering Suspects
Episode: "Judgment Day"

Topic: Conditional Probability Leading to Bayesian Filtering   **Grade Level**: 6-12
Objective: Introduce conditional probability in Bayesian filtering
Time: about 10 minutes

Introduction

Thomas Bayes (1702-1761) was a mathematician and minister who published little in mathematics, but what he wrote has been very significant. His ideas of conditional probability (*the chances of A given B*) have influenced modern theories of decision making and are the basis for many filtering programs, most notably those that filter spam from our e-mail inboxes.

Example When rolling a fair number cube, we can calculate the probability of rolling a 3 to be $\frac{1}{6}$. However, if we are given information about the outcome, the probability of rolling a 3 changes. For example, if we are told that an odd number was rolled, the probability that we rolled a 3 is now $\frac{1}{3}$.

Discuss with Students

In **"Judgment Day*"*,** Charlie helps the FBI sort through their case files for relevant suspects. He applies a logical filter to find the files most relevant to a specific crime.

NUMB3RS **Example** By the type of crime and where it occurred, Don has narrowed his investigation down to four suspects. Resources are scarce, and time is of the essence. Don uses Bayesian filtering logic to choose which suspect to go after first.

Here are the four suspects and their characteristics.

Suspect 1	Suspect 2	Suspect 3	Suspect 4
6 ft 2 in.	5 ft 3 in.	5 ft 5 in.	6 ft 1 in.
235 lbs	165 lbs	145 lbs	200 lbs
Brown hair	Red hair	Black hair	Blonde hair

If each suspect had an equal chance of committing the crime, the probability for any of the suspects is $\frac{1}{4}$. So, for example, the probability that it was Suspect 4 is $\frac{1}{4}$.

The probability of each suspect changes, given certain information. For example, if witnesses say that the suspect was over 6 feet tall, the probability of Suspect 4 given that the person was over 6 feet tall is now $\frac{1}{2}$. This is an example of Bayesian filtering, or the process of recalculating the probability of something, given new information.

Student Page Answers: **Case 1**: a. *1/6* b. *1/2* **Case 2**: *1/3* **Case 3**: a. *1/3* b. *1; Suspect 5 must be the criminal.* **Case 4**: *0; Suspect 3 could not be responsible for the fire.* **Your Turn**: *Answers will vary; check student's work.*

Name: _____   Date: _____

NUMB3RS Activity: Filtering Suspects

The six suspects below are being investigated for their involvement in four independent cases. Use the list of suspects to answer the questions below.

Suspect 1	Suspect 2	Suspect 3	Suspect 4	Suspect 5	Suspect 6
6 ft 3 in.	5 ft 8 in.	5 ft 4 in.	6 ft 1 in.	6 ft 1 in.	5 ft 10 in.
220 lbs	205 lbs	160 lbs	190 lbs	215 lbs	175 lbs
Blonde hair	Black hair	Blonde hair	Brown hair	Red hair	Brown hair
Beard	Beard	Goatee	Beard	Goatee	No Beard

Case 1: Agent Terry Lake is investigating a robbery and knows that it was
committed by one of the suspects above.
   a. What are the chances that Suspect 2 committed the robbery?_____
   b. Terry receives a tip that the robbery was committed by a dark-haired
   (brown or black hair) man with facial hair. Now what are the chances
   that the robbery was committed by Suspect 2? _____

Case 2: A tall (over 6 ft) man was seen leaving the scene of an assault. What
are the chances that the man was Suspect 1? _____

Case 3: Based on the footprints found in the mud at the scene of a kidnapping,
Agent David Sinclair has concluded that a certain criminal must have
weighed more than 200 pounds.
   a. What are the chances that Suspect 5 is the criminal?_____
   b. An eye-witness comes forward and says that the person seen fleeing
   the scene had a goatee. Given this new information, what are the
   chances that Suspect 5 is the criminal? Explain your answer.
   _____
   _____

Case 4: Agent Eppes is following up on a lead that a man with dark hair under 6
feet tall was seen near a warehouse just before it was set on fire. What
are the chances that Suspect 3 is responsible for the fire? Explain your
answer. _____
   _____

Your turn: Create a scenario such that Suspect 6 has a $\frac{1}{2}$ chance of
committing a crime. _____
   _____

Extensions

Activity: Bayesian Filter's Role with Respect to Spam E-mails

Introduction

The way a Bayesian filter works to prevent spam is by analyzing the content of spam messages and non-spam messages. By seeing which words and combinations of words appear most often in spam, but rarely in non-spam, the filter can determine which e-mails have a higher probability of being spam than others. That is, it can "learn" which e-mail to eliminate and which to let through.

Additional Resources

http://www.paulgraham.com/antispam.html

This site has links to essays, some of which contain advanced mathematics, describing spam filters in detail.

http://www.process.com/precisemail/bayesian_filtering.htm

The essay "Introduction to Bayesian Filtering: Using Bayes' Formula to Keep Spam Out of Your Inbox" may be appropriate for higher-level classes. Bayes Theorem is introduced and the application of it in spam filtering is discussed at length.

For the Student

Describe a process you would use to create your own personal spam filter using Bayesian filtering.

- Brainstorm words and word combinations that identify spam.
- No spam filter will be perfect. Consider whether it is better to occasionally let some spam through or to occasionally eliminate regular mail.
- Discuss ways that spammers adapt to spam filters. How do they alter their e-mails to get through the filters?

Related Topic: More on Probability

Explore probability theory with the Probability Simulation Application on your TI-84 Plus graphing calculator. This application simulates the rolling of dice, tossing of coins, picking marbles from a bag, spinning a spinner, and generating random numbers on your calculator. Settings allow you to adjust weighting for fair or unfair outcomes.

In addition, students can export data for further exploration.