Lecture 4: Independent Events and Bayes' Theorem

- Independent Events
- Law of Total Probability
- Bayes' Theorem
- Case Study: Prosecutor's Fallacy

Introduction

– In the last lecture we discussed basic counting methods, conditional probability, and independence of two events

– Today we will introduce independence of several events, and

– how to chose a partition of the sample space so that an important source of uncertainty is reduced if we learn which one of the partition events occurs

– We will look at one of the core theorems of probability, Bayes' Theorem

Independence

– The events A and B are said to be independent if $P(A|B) = P(A)$. That is, if knowing B doesn't change the probability of A. By the definition of conditional probability, this implies that if A and B are independent, we must have $P(A \cap B) = P(A)P(B)$

– **Gambler's Fallacy:**

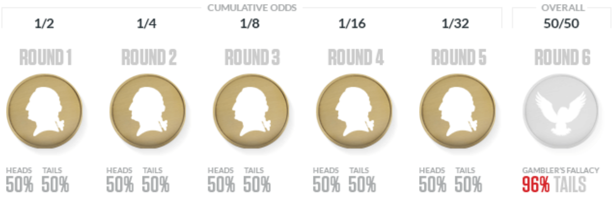

Figure from
https://www.sportsbettingdime.com/guides/betting-psychology/avoiding-gamblers-fallacy/

Past independent events do not have an effect on future outcomes.
Relationship to hot-hand fallacy

Independence of Several Events

– **(Mutually) Independent Events:** The k events $A_1, \ldots, A_k$ are independent (or mutually independent) if, for every subset $A_{i_1}, \ldots, A_{i_j}$ of j of these events $(j = 2, 3, \ldots, k)$

$$P(A_{i_1} \cap \ldots \cap A_{i_j}) = P(A_{i_1}) \cdots P(A_{i_j})$$

– In order for three events A, B and C to be independent, the following four relations must ALL be satisfied:

1. $P(A \cap B) = P(A)P(B)$
2. $P(A \cap C) = P(A)P(C)$
3. $P(B \cap C) = P(B)P(C)$
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$

Pairwise Independence

D.S. Example 2.2.4: Suppose that a fair coin is tossed twice so that the sample space $S = \{HH, HT, TH, TT\}$ is simple. Define the following three events:

- A = {H on first toss}={HH, HT}
- B = {H on second toss}={HH, TH}, and
- C = {Both tosses the same}={HH,TT}.

Then $A \cap B = A \cap C = B \cap C = A \cap B \cap C = \{HH\}$. Hence, $P(A) = P(B) = P(C) = 1/2$, and

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C) = 1/4$$

So $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$, but $P(A \cap B \cap C) \neq P(A)P(B)P(C)$. The results can be summarized by saying that A, B and C are pairwise independent, but all three events are not independent.

Partitions and Law of Total Probability

– **Partition:** Let S denote the sample space of some experiment, and consider k events $B_1, \ldots, B_k$ in S such that $B_1, \ldots, B_k$ are disjoint and $\bigcup_{i=1}^{k} B_i = S$. It is said that these events form a **partition** of S.

– Note that $\sum_{j=1}^{k} P(B_j) = 1$ and $\sum_{j=1}^{k} P(B_j|A) = 1$ for any other event A.

– **Law of Total Probability:** Suppose the events $B_1, \ldots, B_k$ form a finite partition of S and $P(B_j) > 0$ for all $j = 1, \ldots, k$. Then for every event A in S,

$$P(A) = \sum_{j=1}^{k} P(B_j)P(A|B_j)$$ **Draw a Venn Diagram!**

Selecting Bolts

– Tow boxes contain long bolts and short bolts. Suppose that one box contains 60 long bolts and 40 short bolts, and that the other box contains 10 long bolts and 20 short bolts. Suppose also that one box is selected at random and a bolt is then selected at random from that box. We would like to determine the probability that this bolt is long.

– Let B_1 be the event that the first box is selected, let B_2 be the event that the second box is selected, and let A be the event that a long bolt is selected. Then

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2)$$

Since a box is selected at random, we know that $P(B_1) = P(B_2) = 1/2$. Furthermore, $P(A|B_1) = 60/100 = 3/5$, $P(A|B_2) = 10/30 = 1/3$. Hence,

$$P(A) = \frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{1}{3} = \frac{7}{15}$$

Party and Policy

Suppose that in country S, 40% of the people support party A, 30% of the people support party B, 20% support party C, and 10% support party D. Let Q be a certain policy. We're given that 50% of the supporters of party A are in favor of Q, 40% of the supporters of party B are in favor of Q, 30% of the supporters of party C are in favor of Q, and 100% of the supporters of party D are in favor of Q. If we draw a citizen from this imaginary country at random, what is the probability that the citizen supports Q? Let Q denotes the event "is in favor of policy Q", A be the event "support party A" (and so on for the rest of the parties). We can find $P(Q)$ using the law of total probability:

$$\begin{aligned} P(Q) &= P(A)P(Q|A) + P(B)P(Q|B) + P(C)P(Q|C) + P(D)P(Q|D) \\ &= 0.4 \times 0.5 + 0.3 \times 0.4 + 0.2 \times 0.3 + 0.1 \times 1 = 0.48 \end{aligned}$$

Draw a tree diagram helps.

Conditional Versions

– **Conditional Version of Law of Total Probability:** The law of total probability has an analog conditional on another event C, namely,

$$P(A|C) = \sum_{j=1}^{k} P(B_j|C)P(A|B_j \cap C)$$

– **Conditional Indepedence:** We say that events $A_1, \ldots, A_k$ are conditionally independent given B if, for every subcollection $A_{i_1}, \ldots, A_{i_j}$ of j of these events $(j = 2, 3, \ldots, k)$,

$$P(A_{i_1} \cap \cdots \cap A_{i_j}|B) = P(A_{i_1}|B) \cdots P(A_{i_j}|B)$$

Bayes' Rule/Theorem

– **Bayes' Rule/Theorem:** Suppose the events $A_1, \ldots, A_k$ form a finite partition of S and $P(A_j) > 0$ for all $j = 1, \ldots, k$, and let A be an event such that $P(A) > 0$. Then for $i = 1, \ldots, k$,

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)}$$

– **Interpretation:** Inverting Probabilities

– **Prior Probability:** $P(A_i)$

– **Posterior Probability:** $P(A_i|B)$

Examples

Example 11: Suppose Duke has only three possible majors: economics, statistics, and biology. Also, suppose 25% of the students are in economics, 45% are in stats, and the rest are in bio (with no double majors).

Campus folklore says that 90% of econ majors want to go on dates, compared with 10% of stats majors and 50% of bio majors.

Your room-mate meets someone at Duke Gardens and gets a date. What is the probability that he/she is dating an economist?

Note that the majors define a finite partition, and the campus folklore gives us the conditional probabilities $P(B|A_i)$

The point of Bayes' rule is to reverse the conditioning to get $P(A_i|B)$.

Examples

Let A_1 = being an econ major, A_2 = being a biology major, A_3 = being a stats major, and B = going on a date. Then we want to find $P(A_1|B)$. From the question, we have:

$$P(A_1) = 0.25,\ P(A_2) = 0.45,\ P(A_3) = 0.30.$$

$$P(B|A_1) = 0.90,\ P(B|A_2) = 0.10,\ P(B|A_3) = 0.50.$$

Using Bayes' rule,

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{j=1}^{3} P(A_j)P(B|A_j)}$$

$$= \frac{0.25 \times 0.90}{(0.25 \times 0.90) + (0.45 \times 0.10) + (0.30 \times 0.50)} = \frac{0.225}{0.42} = 0.5357$$

Party and Policy (Cont'd)

Supposed that our friend Robbie is a citizen of the imaginary country S we introduced before. We know he doesn't support policy Q because his Facebook status is "I really dislike Q :(". Which party does he support?

$$P(A|Q^c) = \frac{P(A)P(Q^c|A)}{P(Q^c)} = \frac{0.4 \times 0.5}{0.52} \approx 0.385$$

$$P(B|Q^c) = \frac{P(B)P(Q^c|B)}{P(Q^c)} = \frac{0.3 \times 0.6}{0.52} \approx 0.346$$

$$P(C|Q^c) = \frac{P(C)P(Q^c|C)}{P(Q^c)} = \frac{0.2 \times 0.7}{0.52} \approx 0.269$$

$$P(D|Q^c) = \frac{P(D)P(Q^c|D)}{P(Q^c)} = 0$$

So it's pretty hard to tell! We know for sure that he doesn't support party D, which makes sense: 100% of the supporters of D are in favor of Q, so he couldn't be supporting D.

Spam Filters

Some spam filters use Bayes' rule to compute the probability that a message is spam given the words it contains. Let S be the event "the message is spam" and C be the event "the message contains the string of words "You won a prize!"". Then, the Bayes filter would compute

$$P(S|C) = \frac{P(S)P(C|S)}{P(C)}$$

where $P(S)$ is the probability that a "random" message is spam, $P(C)$ is the probability that a message contains the string "You won a prize!", and $P(C|S)$ is the probability that give a message is a spam, it contains the string "You won a prize!". Clearly, $P(C|S)$ is way greater than $P(C)$, so $P(S|C)$ will be pretty close to 1. There are many spam filters that essentially do this. If you're interested, you can read more at:
`https://en.wikipedia.org/wiki/Naive_Bayes_spam_filtering`

In-class Exercise

To be done in-class with one teammate: ELISA is a test for HIV. Like all technology, it is not completely reliable, but the Food And Drug Administration (FDA) has collected extensive information on its error rates.

- If a person has HIV, ELISA has probability 0.997 of signaling.
- If a person does not have HIV, then ELISA does not signal with probability 0.985.
- About 0.32% of the U.S. population has HIV.

Suppose someone walks into a clinic to get an HIV test at random and the test comes back positive. What is the chance that the person has HIV?

Answer: approximately 17.59%

Two Preparatory Steps

Bayes' rule is one of the core formulas in the course, so make sure you're familiar with it and know how to apply it.

1. First identify the marginal probabilities of each possible outcome of the first event $P(A_1), P(A_2), \ldots, P(A_k)$
2. Then identify the probability of the outcome B, conditioned on each possible scenario for the first event, $P(B|A_1), P(B|A_2), \ldots, P(B|A_k)$. Once each of these probabilities are identified, they can be applied directly within the formula.

– Drawing a tree diagram makes it easier to understand how two events are connected

– If there are so many scenarios that drawing a tree diagram would be complex

Island Problem (Eggleston 1983, Appendix 3)

A murder has been committed on an island, cut off from the outside world, on which $1001 (= N + 1)$ inhabitants remain. The forensic evidence at the scene consists of a measurement x, on a "crime trace" characteristic, which can be assumed to come from the criminal,

1. The mainland police arrive and arrest a random islander, Jack. It is found that Jack matches the trait with the criminal. There is no other relevant evidence.
2. The probability of a random member of the population having characteristic x is $p = 0.004$
3. Prosecuting counsel asserts that the probability that Jack is guilty is $1 - p$, or 0.996, (because the probability that Jack would show characteristic x if he were not guilty would be 0.4%) and that this proves guilt "beyond a reasonable doubt".

Can you spot what's wrong with this logic?

Bayesian Argument

1. Let I denote innocence and G guilt
2. $P(\text{I}) = \frac{N}{N+1}$, $P(\text{G}) = \frac{1}{N+1}$
3. $P(\text{Jack has x }|\text{G}) = 1$, $P(\text{Jack has x }|\text{I}) = p$
4. We want $P(\text{G}|\text{Jack has x })$

$$
\begin{aligned}
P(\text{G}|\text{Jack has x }) &= \frac{P(\text{G and Jack has x })}{P(\text{Jack has x })} = \frac{P(\text{Jack has x}\,|\,\text{G})P(G)}{P(\text{Jack has x })} \\
&= \frac{P(\text{Jack has x}\,|\,\text{G})P(G)}{P(\text{Jack has x}\,|\,\text{G})P(G) + P(\text{Jack has x}\,|\,\text{I})P(I)} \\
&= \frac{1 \times 1/(N+1)}{1 \times 1/(N+1) + p \times N/(N+1)} = \frac{1}{1 + p \times N}
\end{aligned}
$$

$N = 1001$, $p = 0.004$, $P(\text{G}|\text{Jack has x }) \approx 0.20$

– **Prosecutor's Fallacy:** $P(I|\textbf{Evidence}) \neq P(\textbf{Evidence}|I) = p$

The Sally Clark Case

1. Sally Clark, a British woman, was accused in 1998 of having killed her first child at 11 weeks of age and then her second child at 8 weeks of age.
2. The prosecution had expert witness Sir Roy Meadow, a professor and consultant paediatrician, testify that the probability of two children in the same family dying from SIDS is about 1 in 73 million.
3. The chance of a mother killing her two children is around 1 in 1, 000, 000.
4. Meadow acknowledged that 1-in-73 million is not an impossibility, but argued that such accidents would happen "once every hundred years"

The Independence Assumption

1. The chance of a family which are non-smokers and over 25 having a SIDS death is around 1 in 8,500.
2. Meadow estimated it from single-SIDS death data, and the assumption that the probability of such deaths should be independent between infants

$$P(\text{both SIDS}) = (1/8500) \times (1/8500) = (1/73,000,000)$$

3. If this is true, then by Bayes' rule,

$$\begin{aligned} P(G|E) &= \frac{P(E|G)P(G)}{P(E)} = \frac{P(E|G)P(G)}{P(E|G)P(G) + P(E|I)P(I)} \\ &= \frac{1 \times (1/1,000,000)}{1 \times (1/1,000,000) + (1/73,000,000)(999,999/1,000,000)} \\ &\approx 0.986 \end{aligned}$$

The Underestimated Chance

1. The chance of a family which are non-smokers and over 25 having a SIDS death is around 1 in $8,500$.
2. The chance of a family which has already had a SIDS death having a second is around 1 in 100.
3. So by the multiplication rule,

$$\begin{aligned} P(\text{both SIDS}) &= P(\text{first SIDS}) \times P(\text{second SIDS}|\text{first SIDS}) \\ &= (1/8500) \times (1/100) = (1/850,000) \end{aligned}$$

4. If this is true, then by Bayes' rule,

$$\begin{aligned} P(G|E) &= \frac{1 \times (1/1,000,000)}{1 \times (1/1,000,000) + (1/850,000)(999,999/1,000,000)} \\ &\approx 0.4594 \end{aligned}$$

The Sally Clark Case (Cont'd)

1. 1-in-73 million greatly underestimated the chance of two successive accidents, but, even if that assessment were accurate, the court seems to have missed the fact that the 1-in-73 million number meant nothing on its own.
2. Mrs. Clark was convicted in 1999, resulting in a press release by the Royal Statistical Society which pointed out the mistakes
3. In 2002, Ray Hill (Mathematics professor at Salford) attempted to accurately compare the chances of these two possible explanations
4. After it was found that the forensic pathologist who had examined both babies had withheld exculpatory evidence, a higher court later quashed Sally Clark's conviction, on 29 January 2003
5. Sally Clark died unintentionally on 16 March 2007 from acute alcohol intoxication. She never recovered from the serious psychological trauma resulting from the experience of the deaths of two children, then being unjustly convicted of their murder with subsequent imprisonment leading to her being separated from her third baby.

Other Examples & Legal Impact

1. D.S. Example 2.2.9: People v. Collins, 68 Cal.2d 319, 438 P.2d 33 (1968)
2. O. J. Simpson murder trial, 1994
3. R v Adams [1996] 2 Cr App R 467, [1996] Crim LR 898, CA and R v Adams [1998]
4. The Lucia de Berk Case, Netherlands, 2003

– **Legal impact:** Though the prosecutor's fallacy typically happens by mistake, in the adversarial system lawyers are usually free to present statistical evidence as best suits their case; retrials are more commonly the result of the prosecutor's fallacy in expert witness testimony or in the judge's summation.

Misunderstandings of p-values in the scientific world

Recap

Today, we talked about

- Independence of several events
- Partition and Law of total probability
- How to apply Bayes' theorem
- Prosecutor's Fallacy

Suggested reading:

- D.S. Sec. 2.1, 2.2, 2.3
- OpenIntro3: Sec. 2.2.5, 2.2.6, 2.2.7