7.6: Bayes' Theorem and Applications

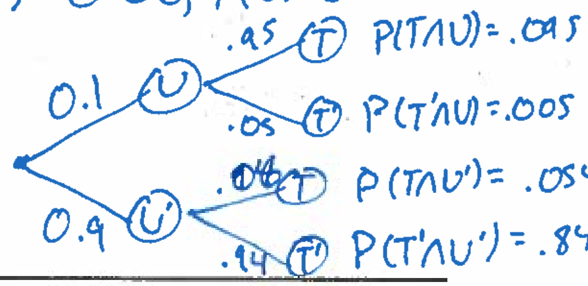
~~7.5: Conditional Probability and Dependence~~

Example 1(a). (Steroids Testing) Gamma Chemicals advertises its anabolic steroid detection test as being 95% effective at detecting steroid use, meaning that it will show a positive result on 95% of all anabolic steroid users. This also it implies that the probability of a false negative is .05. This means that there is a 5% chance that a user will test negative. It also states that its test has a false positive rate of 6%. This means that that the probability of a nonuser testing positive is .06. Estimating that about 10% of all of its athletes are using, Enormous State University (ESU) begins testing its football players. The quarterback, Hugo V. Huge, tests positive and is promptly dropped from the team. Hugo claims that he is not using steroids. How confident can we be that he is not telling the truth?

(1) Name Events $\begin{cases} T: \text{test positive for steroids} \\ U: \text{uses steroids} \end{cases}$

(2) Given Info: $P(T|U) = 0.95$, $P(T|U') = 0.06$, $P(U) = 0.10$

(3) What we want: $P(U|T)$



So $P(U|T) = \dfrac{\cancel{.054}\ .095}{.095 + .054}$

$\approx .64$

Not very confident

Bayes' Theorem (Simple Version) Let A and B be events.

1. $P(B) = P(B|A)P(A) + P(B|A')P(A')$
2. $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A')P(A')}$
3. $P(A|B) = \frac{P(\text{Using } A \text{ and } B \text{ branches})}{\text{Sum of } P(\text{Using branches ending in } B)}$

Example 1(b). Redo Example 1(a) using a table instead of a tree.

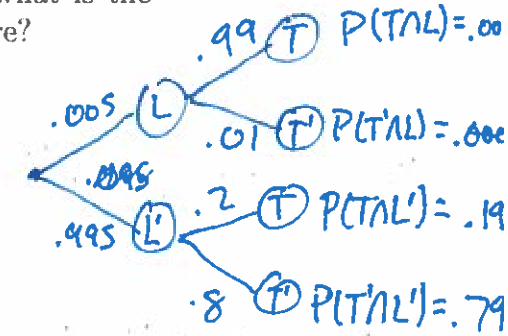
	T	T'	Total
U	$P(T|U)\cdot P(U) = .095$	.005	0.1
U'	.054	.846	0.9
Total	.149	.851	1

$P(U|T) = \dfrac{P(U \cap T)}{P(T)}$

$= \dfrac{.095}{.149}$

$\approx .64$

Example 2. (Lie Detectors) The Watson and Holmes Lie Detector Company manufactures the latest in lie detectors, and the Count-Your-Pennies (CYP) store chain is eager to use them to screen its employees for theft (they were counting there pennies and noticed some missing). Watson and Holmes' advertising claims that the test misses a lie only once in every 100 instances. On the other hand, an analysis by an independent consumer group reveals 20% of people who are telling the truth fail the test anyways. Furthermore, the local police department estimates that 1 out of every 200 employees has engaged in theft. When the CYP store first screened its employees, the test indicated that Ms. Prudence B. Good was lying when she claimed she had never stolen from CYP. Assuming that people were only lying if they stole, what is the probability that she was lying and had in fact stolen from the store?



(1) Events: L: subject is lying/subject stole

T: Tests positive

(2) Given: $P(T|L)=.99, P(T|L')=.2, P(L)=.005$

(3) Want: $P(L|T)=\dfrac{.00495}{.00495+.199}\approx .024$. Only a 2.4% chance she actually stole from the store.

Definition 1. A partition of the sample space S is a collection of events, $(A_1, A_2, \ldots, A_n)$, in S such that $A_1 \cup A_2 \cup \cdots \cup A_n = S$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$ (with $1 \leq i, j \leq n$). Essentially, a partition is a collection of mutually exclusive events that cover the sample space.

Example 3.

(a) Given any event A the collection (A, A') forms a partition of S.

(b) Suppose you roll a die twice. For $1 \leq i \leq 6$, let A_i be the event that the first roll results in an i being face up. Then the collection $(A_1, A_2, A_3, A_4, A_5, A_6)$ forms a partition of the sample space.

Remark. When doing any problems in probability, particularly those involving conditional probabilities and/or Bayes' Theorem, you should always name all of your events, represent the information given using the terminology of probability theory and determine (if applicable) what is the partition of your sample space.

Bayes' Theorem (Full Version) Let $(A_1, A_2, \ldots, A_n)$ be a partition of S and let B be any event.

1. $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$
2. $P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+\cdots+P(B|A_n)P(A_n)}$

Example 3. A survey conducted by the Bureau of Labor Statistics found that approximately 27% of the high school graduating class of 2010 went on to a 2-year college, 41% went on to a 4-year college, and the remaining 32% did not go on to college. Of those who went on to a 2-year college, 52% worked at the same time, 32% of those going on to a 4-year college worked, and 78% of those who did not go on to college worked. What percentage of those working had not gone on to college and what percentage of the graduates went to work?

(1) Events: C_1: 2-year college, C_2: 4-year college, C_3: No college, W: Went to work

(2) Given: $P(C_1)=.27$, $P(C_2)=.41$, $P(C_3)=.32$, $P(W|C_1)=.52$, $P(W|C_2)=.32$, $P(W|C_3)=.78$

(3) Want: ~~$P(W|C_3)$~~ $P(C_3|W)$ and $P(W)$

Using Bayes' Theorem

$P(W) = (.27)(.52) + (.41)(.32) + (.32)(.78) = .5212$

$P(C_3|W) = \frac{(.32)(.78)}{P(W)} \approx .48$

	W	W'	Total
C_1	.1404	.1296	.27
C_2	.1312	.2788	.41
C_3	.2496	.0704	.32
Total	.5212	.4788	1

Therefore $P(W) = .5212$

and

$P(C_3|W) = \frac{.1404}{.5212} \approx .48$

Example 4. Suppose a die is rolled three times. Let A_{oo} be the event that the first two rolls result in odds, A_{ee} be the event that the first two rolls result in even, A_{oe}: the first roll is odd and the second is even, and A_{eo}: the first roll is even and the second roll is odd. The E be the event that the sum of the three rolls is at least 16. Suppose your super nice teacher has already calculated the conditional probabilities

$$P(E|A_{oo}) = \tfrac{1}{54}, \quad P(E|A_{ee}) = \tfrac{5}{54}, \text{ and } P(E|A_{oe}) = P(E|A_{eo}) = \tfrac{1}{27}.$$

Calculate the probability of E using Bayes' formula and verify that your calculations are correct using a direct method (such as a decision algorithm).

To use Bayes' formula, you must first calculate

$P(A_{oo})$, $P(A_{ee})$, $P(A_{oe})$ and $P(A_{eo})$.

$$P(A_{oo}) = P(A_{ee}) = P(A_{oe}) = P(A_{eo}) = \frac{C(3,1)\cdot C(3,1)}{6^2} = \frac{1}{4}$$

So $$P(E) = \frac{1}{54}\cdot\frac{1}{4} + \frac{5}{54}\cdot\frac{1}{4} + \frac{1}{27}\cdot\frac{1}{4} + \frac{1}{27}\cdot\frac{1}{4} = \frac{5}{108} \approx .0463$$

<u>Directly:</u> $P(\text{at least } 16) = P(16) + P(17) + P(18)$

$$P(18) = \frac{1}{6^3}, \quad P(17) = \frac{C(3,2)\cdot C(1,1)}{6^3} = \frac{3}{6^3},$$

$$P(16) = \frac{C(3,2)\cdot C(1,1) + C(3,2)\cdot C(1,1) \cancel{+ C(3,1)\cdot C(2,1)\cdot C(1,1)}}{6^3} = \frac{6}{6^3}$$

So $$P(\text{at least } 16) = \frac{1+3+6}{6^3} = \frac{10}{216} = \frac{5}{108} \approx .0463$$