**Activity: Conditional Probability and The Law of Total Probability**

**Concepts:** Conditional Probability, Independent Events, the Multiplication Rule, the Law of Total Probability

**Prerequisites:** Student should be able to use the basic laws of probability, know how to read 2x2 data tables, and how to translate probability questions into event/set terminology and notation.

**Conditional Probability:**
**Scenario: Top 100 Films II**
In 1998 the American Film Institute created a list of the top 100 American films ever made (www.afi.com/tv/movies.asp). Suppose that two people gather to watch one of these movies and, to avoid potentially endless debates about a selection, decide to choose a movie at random from the "top 100" list. You will investigate the various probabilities associated with this scenario.

Let A denote the subset of these 100 films that Allan has seen, so the event A = {films that Allan has seen}. Similarly define events B for Beth. The following 2x2 table classifies each movie according to whether it was seen by Allan and whether it was seen by Beth. The "at random" selection implies that each of the 100 films is equally likely to be chosen; i.e., each has probability 1/100. Thus, the probabilities of these various events can be calculated by counting how many of the 100 films comprise the event of interest. For example, the table reveals that 42 movies were seen by both Allan and Beth, so
$P(A \cap B)=42/100=.42$ (read "the probability of A and B").

| | Beth yes | Beth no | Total |
|---|---|---|---|
| Allan yes | 42 | 6 | |
| Allan no | 17 | 35 | |
| Total | | | 100 |

(a) Fill in the totals on the table. To make sure you are reading the table correctly, we will quickly compute a few probabilities in class…

(b) Given the knowledge that Allan has seen a film, what is the *conditional probability* that Beth has seen it? [*Hint*: Restrict your consideration to films that Allan has seen, and ask yourself what fraction *of them* has Beth seen.]

(c) How does this conditional probability of Beth having seen the film given that Allan has seen it compare with the (unconditional) probability of Beth having seen the film in the first place? Does the knowledge that Allan has seen the film make it more or less likely (or neither) that Beth has seen it?

(d) Suggest how this conditional probability could have been calculated from $P(A \cap B)$, $P(A)$, and $P(B)$. Which of these three is not needed?

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

We denote the *conditional probability* of an event B given that the event A has occurred by P(B|A). It can be calculated as: $P(B|A) = P(A \cap B)/P(A)$.

(e) Use this definition of conditional probability to calculate P(A'|B') in this context, and explain in words what the resulting probability means

(f) Determine the (unconditional) probability that Allan has seen the film.

(g) Determine the conditional probability that Allan has seen the film given that Beth has.

(h) Does the knowledge that Beth has seen the film make it more or less likely that Allan has seen it, or does that probability not change given the knowledge that Beth has seen the film?

**Independence and the Multiplication Rule:**

Now consider hypothetical data representing the number of these films seen by Chuck and by Donna:

| | Donna yes | Donna no | Total |
|---|---|---|---|
| Chuck yes | 15 | 10 | |
| Chuck no | 45 | 30 | |
| Total | | | 100 |

(a) Compare Donna's (unconditional) probability of having seen the film with the conditional probability that she has seen it given that Chuck has. Does the knowledge that Chuck has seen the film change the probability that Donna has seen it?

P(D) = P(D|C) =

Two events A and B are said to be *independent* if P(A|B)=P(A); otherwise they are *dependent*.

(b) Are the events {Allan has seen the film} and {Beth has seen it} independent? How about {Chuck has seen the film} and {Donna has seen it}? Explain.

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

(c) Algebraically derive an equivalent expression for independence that involves $P(A \cap B)$, $P(A)$, and $P(B)$.

(d) Now suppose you are told that Ellen has seen 80% of the films that Donna has seen. Express this value of .8 as a conditional probability involving the events E = {Ellen has seen it} and D = {Donna has seen it}.

(e) Can you use the information given about Donna and Ellen to determine the proportion of films that have been seen by both Donna and Ellen? If so, please do. [*Hint*: Solve for $P(D \cap E)$ from the expression for $P(E|D)$.]

> The *multiplication rule*, which follows immediately from the definition of conditional probability, asserts that: $P(A \cap B) = P(A)\,P(B|A)$. This can equivalently be written as: $P(A \cap B) = P(B)\,P(A|B)$. When the events are *independent*, this becomes $P(A \cap B) = P(A)\,P(B)$.

(f) Explain how the multiplication rule for independent events follows from the more general multiplication rule.

**<u>Scenario:</u> Graduate School Admissions**

Suppose that you apply to two graduate schools A and B, and that you believe your probability of acceptance by A to be .7, your probability of acceptance by B to be .6, and your probability of acceptance by both to be .5.

(g) Are the events {acceptance by A} and {acceptance by B} independent? Explain. [*Hint*: Use the alternative definition that you derived in (f).]

(h) Determine the conditional probability of acceptance by B given acceptance by A? How does it compare to the (unconditional) probability of acceptance by B?

(i) What is the probability that you are accepted by at least one of the two schools?

Now suppose that the events {acceptance by A} and {acceptance by B} are independent, with probability of acceptance by A equal to .7 and probability of acceptance by B equal to .6.

(j) Determine the probability of acceptance by both schools. Then determine the probability of acceptance by at least one school. Also indicate appropriate symbols and set operations to describe these events.

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

The multiplication rule for a series of *independent* events $A_1, A_2, \ldots, A_k$ asserts that $P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2)\ldots P(A_k)$.

Suppose that you also apply to graduate schools C and D, that you consider all acceptances to be independent of each other, and that you believe the probabilities of acceptance to be .8 and .5, respectively.

(k) What is the probability of acceptance by all four schools?

(l) What is the probability of acceptance by at least one of the four schools? [*Hint*: First find the probability of the complement of this event.]

**Law of Total Probability:**

**Scenario: Presidential Election Votes**

Exit polls conducted by CNN during the 2000 Presidential election recorded the race of voters (and many other variables) as well as for whom they voted. Results available at http://www.cnn.com/ELECTION/2000/results/index.epolls.html reveal that:

- 54% of white voters (W) voted for George W. Bush (B)
- 9% of African-American (A) voters voted for Bush
- 35% of Hispanic voters (H) voted for Bush
- 41% of Asian voters (I) voted for Bush
- 39% of other races (O) voted for Bush

(m) Is it correct to find the (unconditional) percentage who voted for Bush by taking the average of these five percentages? Explain why or why not. If not, indicate under what conditions that would be correct.

(n) Translate the percentages into probability statements using the symbols given in parentheses.

.54 = .09 = .35 = .41 = .39 =

The CNN exit poll results further revealed that 80% of those interviewed were white, 10% were African-American, 7% Hispanic, 2% Asian, and 1% other races.

Record these as (unconditional) probabilities in the bottom row of the *probability table* below.

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

| | White (W) | African-American (A) | Hispanic (H) | Asian (I) | Others (O) | Total |
|---|---|---|---|---|---|---|
| Bush (B) | | | | | | |
| Gore (G) | | | | | | |
| Others (E) | | | | | | |
| Total | | | | | | 1.00 |

(o) Use the multiplication rule to find P(W∩B). Also express this event in words. Then record this probability in the upper left cell of the probability table.

(p) Similarly find the other intersection probabilities that comprise the first row of the probability table.

(q) Sum those five probabilities in the first row of the table to determine the (unconditional) probability of voting for Bush.

You have derived the *Law of Total Probability (LTP)*, which enables you to calculate an unconditional probability when one knows conditional probabilities. It says that if the events $A_1$, $A_2, \ldots, A_k$ are mutually exclusive and exhaustive (i.e., they form the entire sample space with their union), then for any event *B*: $P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$. In other words, the probability of B is a weighted average of the conditional probabilities given a certain group designation, with weights equal to the group probabilities.

(r) Use the CNN information that Al Gore received 42% of the white vote, 90% of the African-American vote, 62% of the Hispanic vote, 55% of the Asian vote, and 55% of the other races' vote to fill in the second row of the table and to calculate Gore's overall vote percentage.

**<u>Scenario:</u> Weekend Weather**

Suppose that a weather forecast says that there is a 30% chance of rain on Saturday and a 40% chance of rain on Sunday.

(s) What is the maximum possible value for the probability of rain on *both* days? Fill in the following probability table to summarize this situation:

| | rain on Sunday | no rain on Sunday | total |
|---|---|---|---|
| rain on Saturday | | | .3 |
| no rain on Saturday | | | |
| total | .4 | | |

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

(t) What is the maximum possible value for the probability of rain on *at least one* of the two days? Fill in the following probability table to summarize this situation:

| | rain on Sunday | no rain on Sunday | total |
|---|---|---|---|
| rain on Saturday | | | .3 |
| no rain on Saturday | | | |
| total | .4 | | |

(u) In either of the cases above (i.e. (s) or (t)), are the events "rain on Saturday" and "rain on Sunday" independent? Why or why not?

(v) Fill in the following the table to summarize the situation that the events "rain on Saturday" and "rain on Sunday" are independent. Show below using the definition and the multiplication rule that these two events are independent.

| | rain on Sunday | no rain on Sunday | total |
|---|---|---|---|
| rain on Saturday | | | .3 |
| no rain on Saturday | | | |
| total | .4 | | |

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF