

**Topic: Basic Probability Rules,
Conditional Probabilities, and Independence**
Activity: Top 100 Films

Prerequisites: Students should be familiar with the notion of sample space and with the “equal likeliness” principle.

Materials: Top 100 Films list. See last two pages of this activity. Open the activity by asking student the percent of these films they have see.

Goals:

- To lead students to discover some basic rules of probability:
 - Complement rule
 - Addition rule for union of two events
 - Addition rule for disjoint events
 - Addition rule for finite union of events
 - Conditional probability definition
 - Multiplication rule for intersection of two events
 - Law of total probability
 - Multiplication rule for independent events
- To introduce students to common notation for events
- To provide experience with translating probability questions into event/set terminology and notation, and then solving the problems using the above rules.



Scenario:

In 1998 the American Film Institute created a list of the top 100 American films ever made. (A link to the file is on the class web page under today’s assignment.) Suppose that three people gather to watch a movie 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 probability that it has already been seen by at least one of the three people.

Notation: Let A denote the subset of these 100 films that Allan has seen, so the event $A = \{\text{films that Allan has seen}\}$. Similarly define events B and F for Beth and Frank, respectively. Recall that A' denotes the complement of A and is interpreted as “not A ”, that $A \cup B$ denotes the union of A and B and is interpreted as “ A or B or both”, and that $A \cap B$ (or simply AB) denotes the intersection of A and B and is interpreted as “ A and B .”

Notice that *events* are *sets*. [In particular, they are subsets of the sample space S .] Thus, it is legitimate to perform set operations such as complement, intersection, and union on them. On the other hand, *probabilities* are *numbers*. More specifically, they are numbers between 0 and 1 (including those extremes). Thus, it is legitimate to add, multiply, and divide probabilities but not to take complements, intersections, or unions of them.

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 following 2x2 table classifies each movie according to whether it was seen by Allan and whether it was seen by Beth. It reveals that 42 movies were seen by both Allan and Beth, so $P(A \cap B) = 42/100$.

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

(a) Translate the following events into set notation using the symbols A and B, complement, union, intersection. Also give the probability of the event as determined from the table:

Event in words	Event in set notation	Probability
Allan and Beth have both seen the film	$A \cap B$	$P(A \cap B) = .42$
Allan has seen the film and Beth has not		
Beth has seen the film and Allan has not		
Neither Allan nor Beth has seen the film		

(b) Fill in the marginal totals of the table (the row and column totals). From these totals determine the probability that Allan has seen a randomly selected film and also the probability that Beth has seen the film. (Remember that the film is chosen at random, so all 100 are equally likely.) Record these, along with the appropriate symbols, below.

$$P(\text{Allan has seen it}) = P(\quad) =$$

$$P(\text{Beth has seen it}) = P(\quad) =$$

(c) Determine the probability that Allan has not seen the film. Do the same for Beth. Record these, along with the appropriate symbols, below.

(d) If you had not been given the table, but instead had merely been told that $P(A) = .48$ and $P(B) = .58$, would you have been able to calculate $P(A')$ and $P(B')$? Explain how.

One of the most basic probability rules is the *complement rule*, which asserts that the probability of the complement of an event equals one minus the probability of the event:

$$P(A') = 1 - P(A)$$

(e) Add the counts in the appropriate cells of the table to calculate the probability that either Allan or Beth (or both) have seen the movie. Also indicate the symbols used to represent this event.

(f) If you had not been given the table but instead had merely been told that $P(A) = .48$ and $P(B) = .59$, would you have been able to calculate $P(A \cup B)$? Explain.

(g) One might naively think that $P(A \cup B) = P(A) + P(B)$. Calculate this sum, and indicate whether it is larger or smaller than $P(A \cup B)$ and by how much. Explain why this makes sense, and indicate how to adjust the right side of this expression to make the equality valid.

The *addition rule* asserts that the probability of the union of two events can be calculated by adding the individual event probabilities and then subtracting the probability of their intersection:
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- (h) Use this addition rule as a second way to calculate the probability that Allan or Beth has seen the movie, verifying your answer to (e).
- (i) As a third way to calculate this probability, first identify (in words and in symbols) the complement of the event {Allan or Beth has seen the movie}. Then find the probability of this complement from the table. Then use the complement rule to determine $P(A \cup B)$. Are your answers to (e) and (h) confirmed?
- (j) What has to be true about A and B for it to be valid to say that $P(A \cup B) = P(A) + P(B)$?

Two events A and B are said to be *disjoint* (or *mutually exclusive*) if their intersection is the empty set ϕ . In other words, two events are disjoint if they cannot both happen simultaneously. If $A \cap B = \phi$, then it follows that $P(A \cup B) = P(A) + P(B)$. This is known as the *addition rule for disjoint events*; it is a special case of the addition rule since if $A \cap B = \phi$, $P(A \cap B) = P(\phi) = 0$.

Conditional Probability:

Reconsider the original 2x2 table pertaining to Allan and Beth:

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

- (k) 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.*]
- (l) 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?
- (m) 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?

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: $P(B|A) = P(A \cap B)/P(A)$.

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

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?

$$P(D) = \qquad P(D|C) =$$

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

Now suppose you are told that Ellen has seen 80% of the films that Donna has seen.

- (d) Express this value of .8 as a conditional probability involving the events $E = \{\text{Ellen has seen it}\}$ and $D = \{\text{Donna has seen it}\}$.
- (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.

Activity for Homework if Not Finished in Class: 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.

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

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.]

AFI's Top 100 Movies: How Many Have You Seen?

The American Film Institute (AFI) tonight announced the 100 greatest American movies of all time, as selected by a blue-ribbon panel of leaders from across the film community.

(<http://www.afi.com/tv/movies.asp>)

- | | |
|--|---|
| 1. CITIZEN KANE (1941) | 25. E.T. THE EXTRA-TERRESTRIAL (1982) |
| 2. CASABLANCA (1942) | 26. DR. STRANGELOVE (1964) |
| 3. THE GODFATHER (1972) | 27. BONNIE AND CLYDE (1967) |
| 4. GONE WITH THE WIND (1939) | 28. APOCALYPSE NOW (1979) |
| 5. LAWRENCE OF ARABIA (1962) | 29. MR. SMITH GOES TO WASHINGTON (1939) |
| 6. THE WIZARD OF OZ (1939) | 30. THE TREASURE OF THE SIERRA MADRE (1948) |
| 7. THE GRADUATE (1967) | 31. ANNIE HALL (1977) |
| 8. ON THE WATERFRONT (1954) | 32. THE GODFATHER PART II (1974) |
| 9. SCHINDLER'S LIST (1993) | 33. HIGH NOON (1952) |
| 10. SINGIN' IN THE RAIN (1952) | 34. TO KILL A MOCKINGBIRD (1962) |
| 11. IT'S A WONDERFUL LIFE (1946) | 35. IT HAPPENED ONE NIGHT (1934) |
| 12. SUNSET BOULEVARD (1950) | 36. MIDNIGHT COWBOY (1969) |
| 13. THE BRIDGE ON THE RIVER KWAI (1957) | 37. THE BEST YEARS OF OUR LIVES (1946) |
| 14. SOME LIKE IT HOT (1959) | 38. DOUBLE INDEMNITY (1944) |
| 15. STAR WARS (1977) | 39. DOCTOR ZHIVAGO (1965) |
| 16. ALL ABOUT EVE (1950) | 40. NORTH BY NORTHWEST (1959) |
| 17. THE AFRICAN QUEEN (1951) | 41. WEST SIDE STORY (1961) |
| 18. PSYCHO (1960) | 42. REAR WINDOW (1954) |
| 19. CHINATOWN (1974) | 43. KING KONG (1933) |
| 20. ONE FLEW OVER THE CUCKOO'S NEST (1975) | 44. THE BIRTH OF A NATION (1915) |
| 21. THE GRAPES OF WRATH (1940) | 45. A STREETCAR NAMED DESIRE (1951) |
| 22. 2001: A SPACE ODYSSEY (1968) | 46. A CLOCKWORK ORANGE (1971) |
| 23. THE MALTESE FALCON (1941) | 47. TAXI DRIVER (1976) |
| 24. RAGING BULL (1980) | 48. JAWS (1975) |

49. SNOW WHITE AND THE SEVEN DWARFS (1937)
50. BUTCH CASSIDY AND THE SUNDANCE KID (1969)
51. THE PHILADELPHIA STORY (1940)
52. FROM HERE TO ETERNITY (1953)
53. AMADEUS (1984)
54. ALL QUIET ON THE WESTERN FRONT (1930)
55. THE SOUND OF MUSIC (1965)
56. M*A*S*H (1970)
57. THE THIRD MAN (1949)
58. FANTASIA (1940)
59. REBEL WITHOUT A CAUSE (1955)
60. RAIDERS OF THE LOST ARK (1981)
61. VERTIGO (1958)
62. TOOTSIE (1982)
63. STAGECOACH (1939)
64. CLOSE ENCOUNTERS OF THE THIRD KIND (1977)
65. THE SILENCE OF THE LAMBS (1991)
66. NETWORK (1976)
67. THE MANCHURIAN CANDIDATE (1962)
68. AN AMERICAN IN PARIS (1951)
69. SHANE (1953)
70. THE FRENCH CONNECTION (1971)
71. FORREST GUMP (1994)
72. BEN-HUR (1959)
73. WUTHERING HEIGHTS (1939)
74. THE GOLD RUSH (1925)
75. DANCES WITH WOLVES (1990)
76. CITY LIGHTS (1931)
77. AMERICAN GRAFFITI (1973)
78. ROCKY (1976)
79. THE DEER HUNTER (1978)
80. THE WILD BUNCH (1969)
81. MODERN TIMES (1936)
82. GIANT (1956)
83. PLATOON (1986)
84. FARGO (1996)
85. DUCK SOUP (1933)
86. MUTINY ON THE BOUNTY (1935)
87. FRANKENSTEIN (1931)
88. EASY RIDER (1969)
89. PATTON (1970)
90. THE JAZZ SINGER (1927)
91. MY FAIR LADY (1964)
92. A PLACE IN THE SUN (1951)
93. THE APARTMENT (1960)
94. GOODFELLAS (1990)
95. PULP FICTION (1994)
96. THE SEARCHERS (1956)
97. BRINGING UP BABY (1938)
98. UNFORGIVEN (1992)
99. GUESS WHO'S COMING TO DINNER (1967)
100. YANKEE DOODLE DANDY (1942)