Data Mining

Lecture 4

Ananya Jana
CS360

Fall 2024

Revising the Fundamental Concepts of Probability

Probability is the measure of the likelihood or chance that a particular event will occur. Probability is defined as a the numerical value that represents how likely an event is to occur, typically ranging from 0 (impossible) to 1(certain).

Random Experiment : Experiment (action) whose result is uncertain (cannot be predicted with certainty) before it is performed.

Trial : Single performance of the random experiment.

Outcome : Result of a trial.

Sample Space S : The set of all possible outcomes of a random experiment

Event : The subset of the sample space S (to which a probability can be assigned)

Revising the Fundamental Concepts of Probability

Traditional definitions of the probability of an event A

Classical: P(A) $= \dfrac{\text{\# of possible outcomes for event } A}{\text{\# of possible outcomes for space } S}$

Relative Frequency: P(A) $= \lim_{N \to \infty} \dfrac{\text{\# of occurrences of event } A}{N \text{ (total \# of trials)}}$

Revising the Fundamental Concepts of Probability

Example: Die-rolling events

Rolling a die is a random experiment. An outcome can be any number from 1 to 6.
Sample space = {1, 2, 3, 4, 5, 6}.
Some possible examples of events are

1. A = { an even number showing up} = {2, 4, 6} (3 outcomes)
2. B = { a number greater than 5 shows up} = {6} (single outcome)
3. C = { 2 shows up} = {2} (single outcome)
4. D = {a number greater than 7 shows up} = { } (no outcome)
5. E = {2 and 4 shows up} = { } (no outcome)
6. F = {2 or 4 shows up} = { 2, 4} (2 outcomes)
7. G = { a number from 1 to 6 shows up} = {1, 2, 3, 4, 5, 6} (all outcomes)

Thus if 2 showed up as a result of rolling a die, we say that the events A, C, F and G have all occurred.

Revising the Fundamental Concepts of Probability

Let's work out their probabilities

1. A = { an even number showing up} = {2, 4, 6} (3 outcomes)
   P (an even number showing up) = 3 / 6 = 1 /2

2. B = { a number greater than 5 shows up} = {6} (single outcome)
   P(a number greater than 5 shows up) = 1 / 6

3. C = { 2 shows up} = {2} (single outcome)
   P (2 shows up) = 1 / 6

4. D = {a number greater than 7 shows up} = { } (no outcome)
   P (a number greater than 7 shows up) = 0 / 6 = 0
   Impossible events have 0 probability.

Revising the Fundamental Concepts of Probability

Let's work out their probabilities

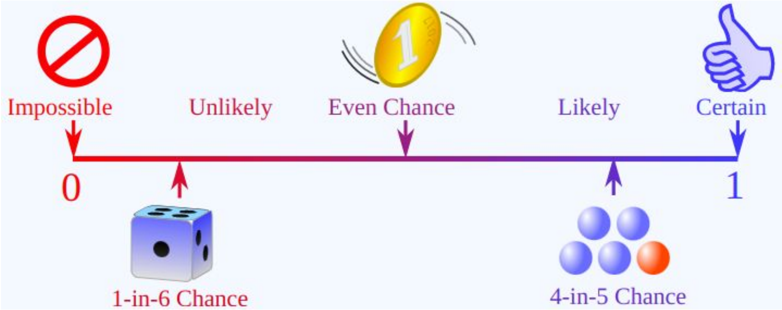
5. E = {2 and 4 shows up} = { } (no outcome)
   P (2 and 4 shows up) = 0 / 6 = 0
   Impossible events have 0 probability

6. F = {2 or 4 shows up} = { 2, 4} (2 outcomes)
   P (2 or 4 shows up} 2 / 6 = 1 / 3

7. G = { a number from 1 to 6 shows up} = {1, 2, 3, 4, 5, 6} (all outcomes)
   P (a number from 1 to 6 shows up) = 6 / 6 = 1
   Sure events have the probability 1.

Revising the Fundamental Concepts of Probability

We can show probability on a probability line.



(image source: math is fun website)

Revising the Fundamental Concepts of Probability

Example 2: Let’s consider another example. Two coins are tossed.
Let A be the event “two heads are obtained” and
B be the event “one head and one tail is obtained”.
Find P(A), P(B).

The sample space S = {HH, HT, TH, TT}

A = {HH}, hence P(A) = 1/4

B = {HT, TH}, hence P(B) = 2/4 = ½
Please notice that the outcomes HT and TH are different.

Revising the Fundamental Concepts of Probability

Example 3: A marble is drawn at random from a bag containing 3 red, 2 blue, 5 green and 1 yellow marble. What is the probability that it is not green?

We can work out the probability that the marble is green: P (marble is green) = 5 / 11
P (marble is NOT green) = 1 - P (marble is green) = 1 - (5/11) = 6 / 11.

Revising the Fundamental Concepts of Probability

Probability of the complement of an event. Let A be an event. The complement of the event is denoted as $\overline{A}$. The probability of the complement of an event A can be calculated in the following way

$$P(\overline{A}) = 1 - P(A)$$

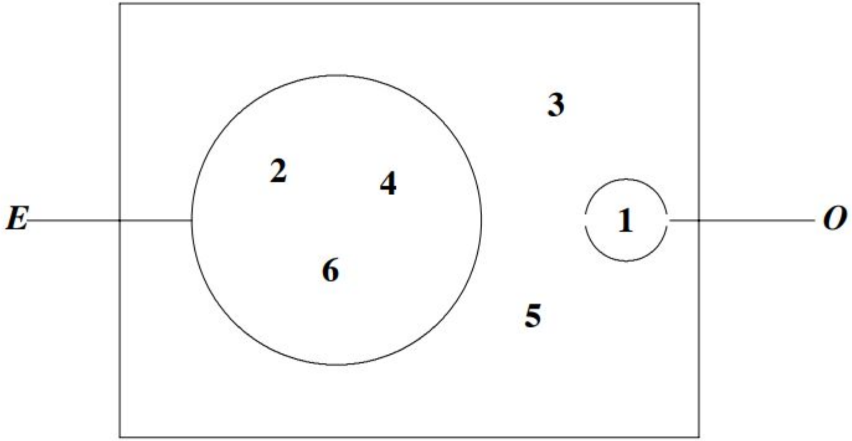
Two events are complementary if they cannot occur at the same time and they make up the whole sample space.

Revising the Fundamental Concepts of Probability

Mutually exclusive or **disjoint** events.

A set of events $A_1, A_2, .. A_n$ is said to be mutually exclusive or disjoint if the intersection of A_i and A_j is empty for all $i \neq j$ i.e. at most one event can occur (if one occurs, any other event cannot occur)

Let's assume a die is rolled.
The event E = "obtaining an even number"
The event O = "obtaining a 1"



As we can see the events E and O are mutually Exclusive as they do not intersect.

Revising the Fundamental Concepts of Probability

For any two events A and B,
$P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$

If the events A and B are **mutually exclusive**, then
$P(A \cup B) = P(A) + P(B)$

Let's assume we rolled a die. We want to calculate the probability of the event "die shows an even face or a die shows a number greater than 3".

Let us denote A as the event "die shows even face". A = {2, 4, 6}. $P(A) = 3/6 = 1/2$
Let us denote B as the event "die shows a number greater than 3". B = {4, 5, 6}. $P(B) = 3/6 = 1/2$

The intersection of A and B is not empty. A and B are not mutually exclusive events.
A intersection B = {4, 6}. $P(A \text{ and } B) = 2/6 = 1/3$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = (1/2) + (1/2) - (1/3)$

Revising the Fundamental Concepts of Probability

Conditional probability: The probability of an event A under the condition that the event B has already occurred. It is denoted as P(A|B).

How to calculate conditional probability?

Let's look at an example - A lecture on a topic of public health is held and 300 people attend. They are classified in the following way:

Gender	Doctors	Nurses	Total
Female	90	90	180
Male	100	20	120
Total	190	110	300

Revising the Fundamental Concepts of Probability

If one person is selected at random, find the following probabilities:

(a) P(a doctor is chosen);
(b) P(a female is chosen);
(c) P(a nurse is chosen);
(d) P(a male is chosen);
(e) P(a female nurse is chosen);
(f) P(a male doctor is chosen).

Gender	Doctors	Nurses	Total
Female	90	90	180
Male	100	20	120
Total	190	110	300

Revising the Fundamental Concepts of Probability

Solution:

a) The number of doctors is 190 and the total number of people is 300, so P(doctor) = 190 / 300

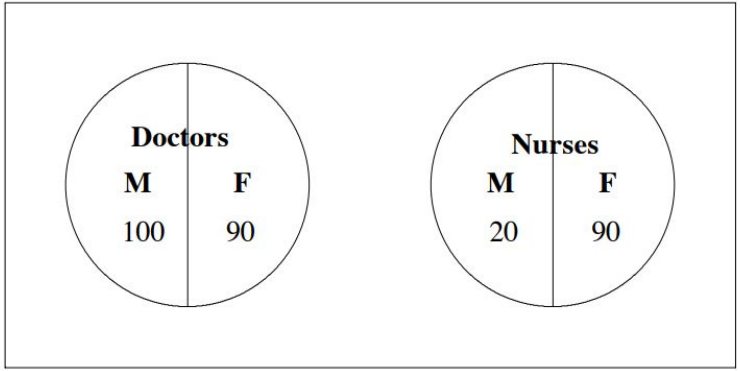
b) P(female) = 180 / 300

c) P (male) = 120 / 300

d) P(nurse) = 110 / 300

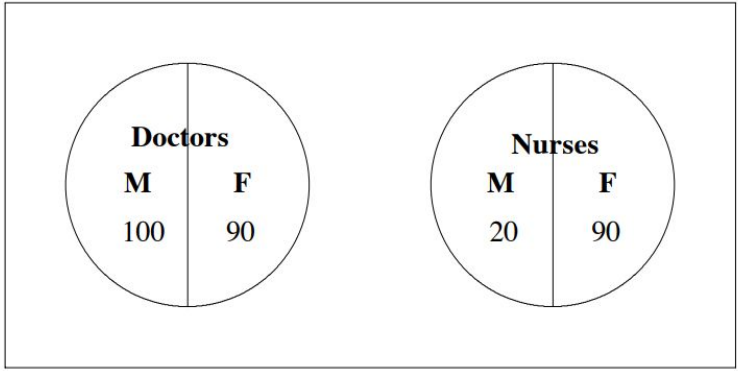
e) There are 90 female nurses, therefore P(female ∩ nurse) = 90 / 300

f) P(male doctor) = P(male ∩ doctor) = 100 / 300 .



Revising the Fundamental Concepts of Probability

Now suppose you are given the information that a female is chosen and you wish to find the probability that she is a nurse. This is what we call **conditional probability**.



We want the probability that the person chosen is a nurse, subject to the condition that we know she is female.

The notation used for this is:

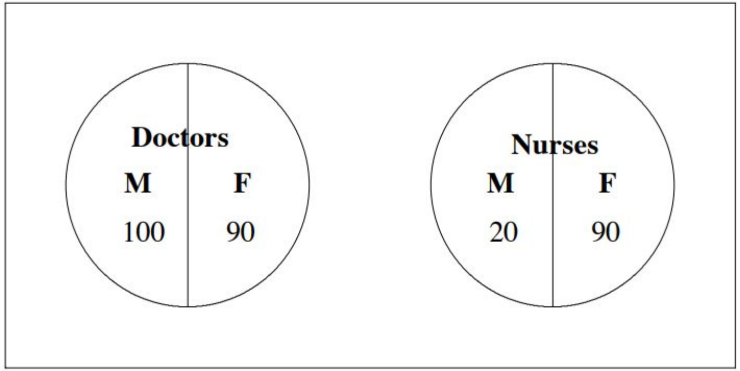
P(nurse | female)

and the notation is read as '*the probability of the person chosen being a nurse, given that she is female*'.

Since there are 180 females and of these 90 are nurses, the required probability is 90 / 180 = 1 / 2

Revising the Fundamental Concepts of Probability

$P(\text{nurse} \mid \text{female}) = 90 / 180$
$= (90 / 300) / (180 / 300)$
$= P(\text{nurse} \cap \text{female}) / P(\text{female})$



The conditional probability of A given the event B has occurred is defined by

$$P(A \mid B) = P(A \cap B) / P(B) \text{ , provided that } P(B) \neq 0.$$

Revising the Fundamental Concepts of Probability

Hence if A and B are any two events with probabilities greater than 0, then

$$P(A \cap B) = P(A \mid B).P(B)$$
$$= P(B \mid A).P(A),$$

as $P(B \cap A) = P(A \cap B)$

In conditional probability the order matters

$P(A \mid B)$ is not the same as $P(B \mid A)$.

Revising the Fundamental Concepts of Probability

Independence: Two events A and B are said to be independent if and only if $P(A \mid B) = P(A)$, that is, when the conditional probability of A given B is the same as the probability of A.

In other words if the chance of a given outcome of the event A remains the same, irrespective of whether or not another event B has occurred, then the events A and B are said to be independent.

Tasks

A. For the previous example, calculate the conditional probabilities:

1. P(female | nurse),
2. P(doctor | male),
3. P(male | doctor).

B. Consider the experiment: tossing a coin thrice. Assume A is the event that “at least two heads show up”. Write down the Sample Space, Event, and the probability of the event A i.e. P(A).