Introduction to Probability

Slides are adapted from STAT414 course at PennState

https://onlinecourses.science.psu.edu/stat414/

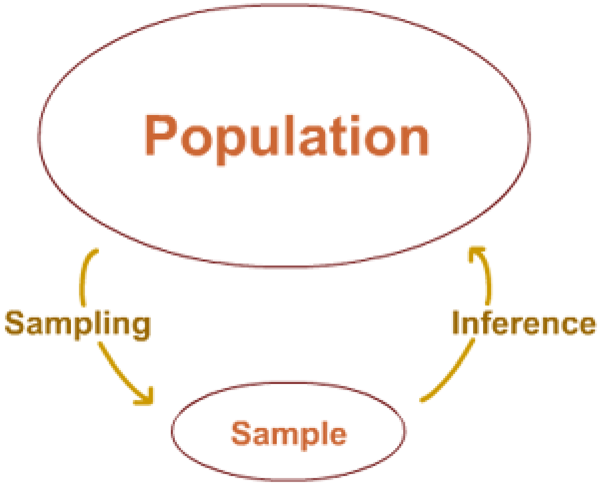
Research questions

- What is the probability of students that took probability course?
- What is the probability that a randomly selected graduate student gets more than seven hours of sleep each night?
- Do women typically cry more than men?

- **How can we answer our research questions if we can't ask every person in the population our research question?**

Population versus sample

- Take a random sample from the population, and use the resulting sample to learn something, or **make an inference**, about the population



Sample space

- The **sample space** (or **outcome space**), denoted ***S***, is the collection of all possible outcomes of a random study.
- ***S*** = {yes, no}   did you take probability course before?
- ***S*** = {h: $h \geq 0$ hours}   h: number of hours slept
- ***S*** = {0, 1, 2, ..., 31}   how many times have you cried last month?

- For each of the research questions you created:
- Formulate the question you would ask (or describe the measurement technique you would use).
- Define the resulting sample space.

Types of data

- Did you take probability course before?

yes yes yes no no no yes yes yes yes
yes no no yes yes no yes yes yes yes

- Quantitative data are called **discrete** if the sample space contains a finite or countably infinite number of values.

  S = {0, 1, 2, ...} countably infinite

- Quantitative data are called **continuous** if the sample space contains an interval or continuous span of real numbers.

  S = {h: $h \geq 0$ hours}

- Qualitative data are called **categorical** if the sample space contains objects that are grouped or categorized based on some qualitative trait. When there are only two such groups or categories, the data are considered **binary**.

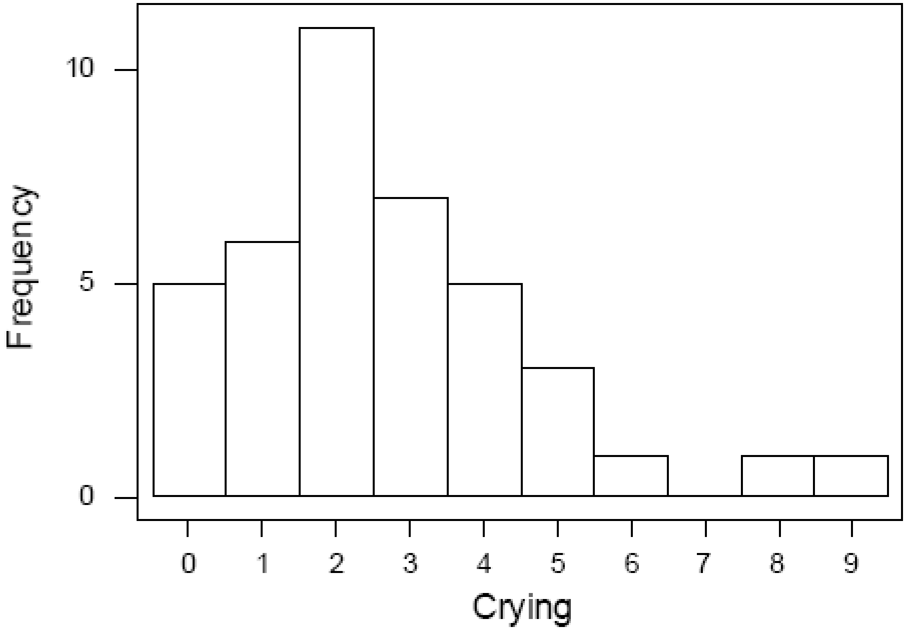
  S = {yes, no}

Frequency histogram

- How many times have you cried this month?
- a histogram gives a nice picture of the "**distribution**" of the data

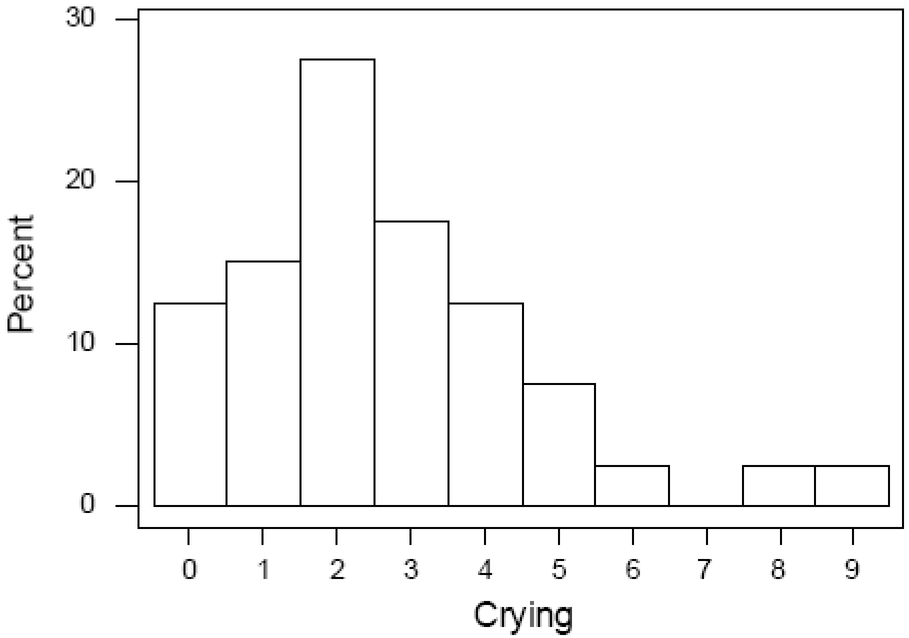
9 5 3 2 6 3 2 2 3 4 2 8 4 4
5 0 3 0 2 4 2 1 1 2 2 1 3 0
2 1 3 0 0 2 2 3 4 1 1 5

- The most common number of times that the women cried in the month was two (called the "**mode**").
- The numbers ranged from 0 to 9 (that is, the "**range**" of the data is 9).
- A majority of women (22 out of 40) cried two or fewer times, but a few cried as much as six or more times.



Relative frequency histogram

- What percentage of the surveyed women reported not crying at all in the month?
- What percentage of the surveyed women reported crying two times in the month? and three times?



Determine the number, *n*, in the sample.
Determine the frequency, f_i, of each outcome *i*.
Calculate the relative frequency (proportion) of each outcome *i* by dividing the frequency of outcome *i* by the total number in the sample *n* — that is, calculate $f_i \div n$ for each outcome *i*.

(Discrete) probability mass function

- $h(x_0) = f_0/n$ is the relative frequency (or proportion) of students, in a sample of size n, crying x_0 times
- As the sample size n increases, f_0/n tends to stabilize and approach some limiting probability $p_0 = f(x_0)$

Density histogram
Histogram of continuous data

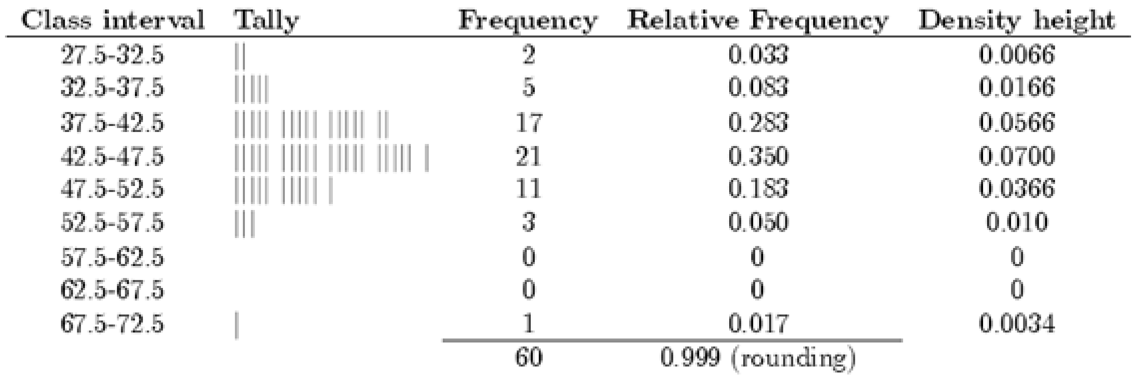
- Measured nose lengths

(error in measurement)

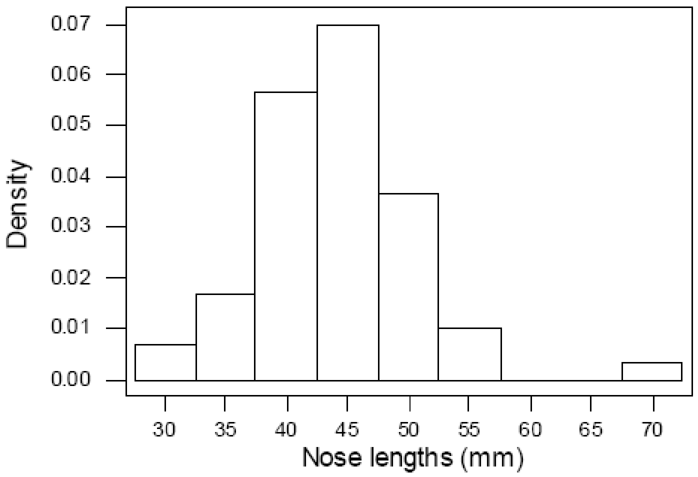
38 50 38 40 35 52 45 50 40 32 40 47 70 55 51
43 40 45 45 55 37 50 45 45 55 50 45 35 52 32
45 50 40 40 50 41 41 40 40 46 45 40 43 45 42
45 45 48 45 45 35 45 45 40 45 40 40 45 35 52

1. Determine the number, *n*, in the sample.
2. Define *k* class intervals $(c_0, c_1]$, $(c_1, c_2]$, ..., $(c_{k-1}, c_k]$.
3. Determine the frequency, f_i, of each class *i*.
4. Calculate the relative frequency (proportion) of each class by dividing the class frequency by the total number in the sample — that is, $f_i \div n$.
5. For a **frequency histogram**: draw a rectangle for each class with the class interval as the base and the height equal to the frequency of the class.
6. For a **relative frequency histogram**: draw a rectangle for each class with the class interval as the base and the height equal to the relative frequency of the class.
7. For a **density histogram**: draw a rectangle for each class with the class interval as the base and the height equal to

$h(x)=f_i / n(c_i - c_{i-1})$ for $c_{i-1}<x\le c_i$ $i = 1, 2,..., k$.



Class interval	Tally	Frequency	Relative Frequency	Density height
27.5-32.5	||	2	0.033	0.0066
32.5-37.5	|||||	5	0.083	0.0166
37.5-42.5	||||| ||||| ||||| ||	17	0.283	0.0566
42.5-47.5	||||| ||||| ||||| ||||| |	21	0.350	0.0700
47.5-52.5	||||| ||||| |	11	0.183	0.0366
52.5-57.5	|||	3	0.050	0.010
57.5-62.5		0	0	0
62.5-67.5		0	0	0
67.5-72.5	|	1	0.017	0.0034
		60	0.999 (rounding)	



the area of the entire histogram equals 1.

Event

- An **event** — denoted with capital letters *A*, *B*, *C*, ... — is just a subset of the sample space *S*. That is, for example $A \subset S$, where "⊂" denotes "is a subset of.“

- "how many pairs of jeans do you own?". Sample space *S* is *S* = {0, 1, 2, 3, ...}

- *A* is the event that a randomly selected student owns no jeans:
- *A* = student owns none = {0}

- *B* is the event that a randomly selected student owns some jeans:
- *B* = student owns some = {1, 2, 3, ...}

- *C* is the event that a randomly selected student owns no more than five pairs of jeans:
- *C* = student owns no more than five pairs = {0, 1, 2, 3, 4, 5}

- *D* is the event that a randomly selected student owns an odd number of pairs of jeans:
- *D* = student owns an odd number = {1, 3, 5, ...}

Set review

- $\emptyset$ is the "**null set**" (or "**empty set**")
- $C \cup D$ = "**union**" = the elements in C or D or both
- $A \cap B$ = "**intersection**" = the elements in A and B. If $A \cap B = \emptyset$, then A and B are called "**mutually exclusive events**" (or "**disjoint events**").
- $D' = D^c$ = "**complement"** = the elements *not* in D
- If $E \cup F \cup G \cup \ldots = S$, then E, F, G, and so on are called "**exhaustive events**."

Examples

- The union of events *C* and *D* is the event that a randomly selected student either owns no more than five pairs or owns an odd number. That is:

$C \cup D = \{0, 1, 2, 3, 4, 5, 7, 9, ...\}$

- The intersection of events *A* and *B* is the event that a randomly selected student owes no pairs and owes some pairs of jeans. That is:

$A \cap B = \{0\} \cap \{1, 2, 3, ...\}$ = the empty set Ø

- The complement of event *D* is the event that a randomly selected student owes an even number of pairs of jeans. That is:

$D^c = \{0, 2, 4, 6, ...\}$

- If $E = \{0, 1\}$, $F = \{2, 3\}$, $G = \{4, 5\}$ and so on, so that:

$E \cup F \cup G \cup ... = S$

then *E*, *F*, *G*, and so on are exhaustive events.

Probability

Probability is a number between 0 and 1, where:

- a number close to 0 means "not likely"
- a number close to 1 means "quite likely"

If the probability of an event is exactly 0, then the event can't occur. If the probability of an event is exactly 1, then the event will definitely occur.

How does an event get assigned a particular probability value?

- the personal opinion approach
  - "I think there is an 80% chance of rain today."
  - What about?
    - one day you will die?
    - you can swim around the world in 30 hours?
    - you will win the lottery some day?
    - a randomly selected student will get an A in this course?
    - *you* will get an A in this course?

- the relative frequency approach
- the classical approach

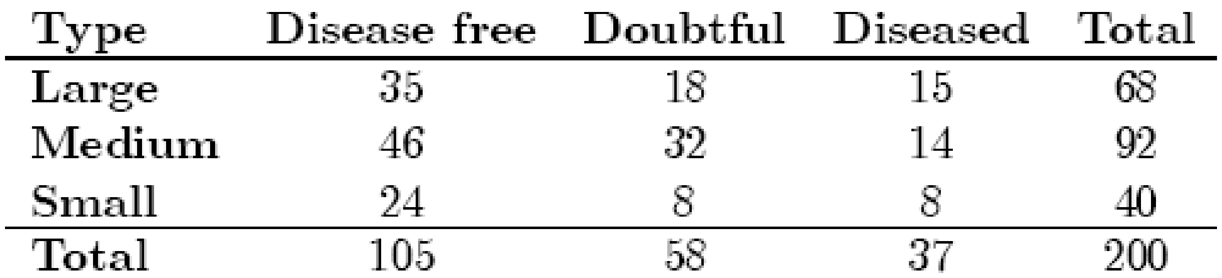
The Relative Frequency Approach

- To determine *P*(*A*), the probability of an event *A*:
- Perform an experiment a large number of times, *n*, say.
- Count the number of times the event *A* of interest occurs, call the number *N*(*A*), say.
- Then, the probability of event *A* equals: P(A) = N(A) / n

- When you toss a fair coin with one side designated as a "head" and the other side designated as a "tail", what is the probability of getting a head?

Coin Tosser	***n*, the number of tosses made**	***N*(*H*), the number of heads tossed**	***P*(*H*)**
Count Buffon	4,040	2,048	0.5069
Karl Pearson	24,000	12,012	0.5005
John Kerrich	10,000	5,067	0.5067

Examples



Type	Disease free	Doubtful	Diseased	Total
Large	35	18	15	68
Medium	46	32	14	92
Small	24	8	8	40
Total	105	58	37	200

- Some trees in a forest were showing signs of disease.
- What is the probability that one tree selected at random is large? 68/200 = 0.34.
- What is the probability that one tree selected at random is diseased?
- 37/200 = 0.185.
- What is the probability that one tree selected at random is both small and diseased?
- 8/200 = 0.04.
- What is the probability that one tree selected at random is either small or disease-free?
- (35 + 46 + 24 + 8 + 8) /200 = 121 / 200 = 0.605.
- What is the probability that one tree selected at random from the population of medium trees is doubtful of disease?
- 2/92 = 0.348.

Classical approach

- the probability of event *A* is: P(A) = N(A) / N(S)
- *N*(*A*) is the number of elements in the event *A*, and *N*(*S*) is the number of elements in the sample space *S*.

- Suppose you draw one card at random from a standard deck of 52 cards
- 13 face values (Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, and King) in 4 different suits (Clubs, Diamonds, Hearts, and Spades) for a total of 52 cards.
- *A* = {*x*: *x* is a 2, 3, or 7}
- *B* = {*x*: *x* is 2H, 3D, 8S, or KC}

Example

- What is the probability that a 2, 3, or 7 is drawn?
- What is the probability that the card is a 2 of hearts, 3 of diamonds, 8 of spades or king of clubs?
- What is the probability that the card is either a 2, 3, or 7 or a 2 of hearts, 3 of diamonds, 8 of spades or king of clubs?
- What is *P*(*A*∩*B*)?

$$P(A) = \frac{N(A)}{N(S)} = \frac{12}{52}$$

$$P(B) = \frac{N(B)}{N(S)} = \frac{4}{52}$$

$$P(A \cup B) = \frac{N(A \cup B)}{N(S)} = \frac{14}{52}$$

$$P(A \cap B) = \frac{N(A \cap B)}{N(S)} = \frac{2}{52}$$

$$A \cap B = \{2H, 3D\}$$

Axioms of probability

1. The probability of any event A must be nonnegative, that is, $P(A) \geq 0$.
2. The probability of the sample space is 1, that is, $P(S) = 1$.
3. Given mutually exclusive events $A_1, A_2, A_3, \ldots$ that is, where $A_i \cap A_j = \emptyset$, for $i \neq j$,

    a. the probability of a finite union of the events is the sum of the probabilities of the individual events, that is:

    $$P(A1 \cup A2 \cup \cdots \cup Ak)=P(A1)+P(A2)+\cdots+P(Ak)$$

    b. the probability of a countably infinite union of the events is the sum of the probabilities of the individual events, that is:

    $$P(A1 \cup A2 \cup \cdots)=P(A1)+P(A2)+\cdots$$

Example

Status	Fresh	Soph	Jun	Sen	Grad	Total
Count	1	4	20	9	9	43
Proportion	0.02	0.09	0.47	0.21	0.21	

Randomly select one student from the class.
Defining the following events:

- Fr = the event that a Freshman is selected
- So = the event that a Sophomore is selected
- Ju = the event that a Junior is selected
- Se = the event that a Senior is selected
- Gr = the event that a Graduate student is selected

The sample space is *S* = (Fr, So, Ju, Se, Gr}. Using the relative frequency approach to assigning probability to the events:

- *P*(Fr) = 0.02
- *P*(So) = 0.09
- *P*(Ju) = 0.47
- *P*(Se) = 0.21
- *P*(Gr) = 0.21

$S = \{Fr, So, Ju, Se, Gr\}$

① $P(A) \geq 0$ ✓

② $P(S) = P(Fr \cup So \cup Ju \cup Se \cup Gr)$

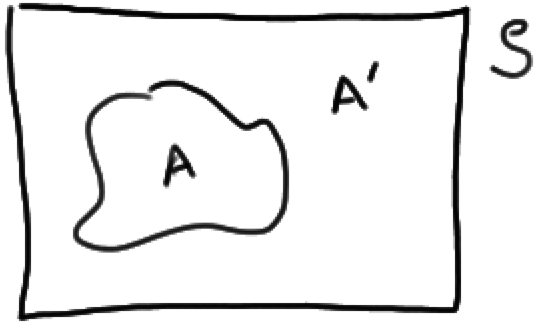
$= 43/43 = 1$ ✓

③ $P(So \cup Ju) = \frac{4+20}{43} = 0.56$

$P(So \cup Ju) = P(So) + P(Ju) = 0.09 + 0.47$

Theorems of probability

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



$$S = A \cup A'$$

$$P(S) = P(A \cup A')$$

$$\overset{③\text{ME}}{=} P(A) + P(A') \overset{②}{=} \underline{1}$$

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

✓

Theorems of probability

- $P(\emptyset) = 0$.

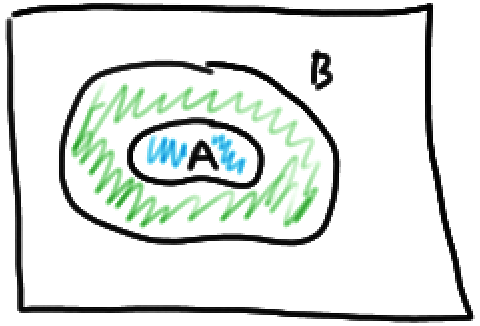
Use theorem #1 and let $A = \emptyset$ and $A' = S$

Then

$$
\begin{aligned}
P(\emptyset) = P(A) &= 1 - P(A') \\
&= 1 - P(S) \\
&\overset{\textcircled{2}}{=} 1 - 1 = 0
\end{aligned}
$$

Theorems of probability

- If events *A* and *B* are such that $A \subseteq B$, then $P(A) \leq P(B)$.



$$A \cup (A' \cap B) = B$$

$$P(A \cup (A' \cap B)) = P(B)$$

$$P(A) + P(A' \cap B) = P(B)$$

by A and $A' \cap B$ being mutually exclusive

$$\Rightarrow P(A) = P(B) - \underbrace{P(A' \cap B)}_{\geq 0 \text{ by } ①} \leq P(B)$$

$\Rightarrow P(A) \leq P(B)$ and $P(A) = P(B)$ when $P(A' \cap B) = 0$

Theorems of probability

- $P(A) \leq 1$.

By definition, event A is a subset of S.

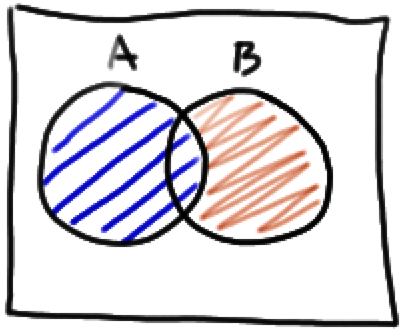
That is, $A \leq S$.

By theorem #3, $P(A) \leq P(S) \overset{②}{=} \underline{1}$

$\Rightarrow P(A) \leq \underline{1}$.

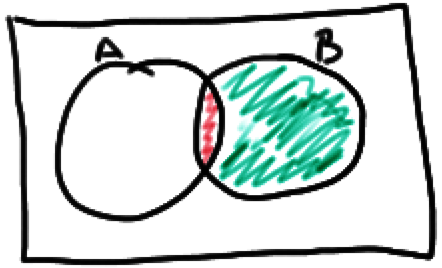
Theorems of probability

- For any two events *A* and *B*, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



$$A \cup B = A \cup (A' \cap B)$$

$$P(A \cup B) \overset{ME}{=} P(A) + \underset{(*)}{P(A' \cap B)}$$



$$B = (A \cap B) \cup (A' \cap B)$$

$$P(B) = P(A \cap B) + P(A' \cap B)$$

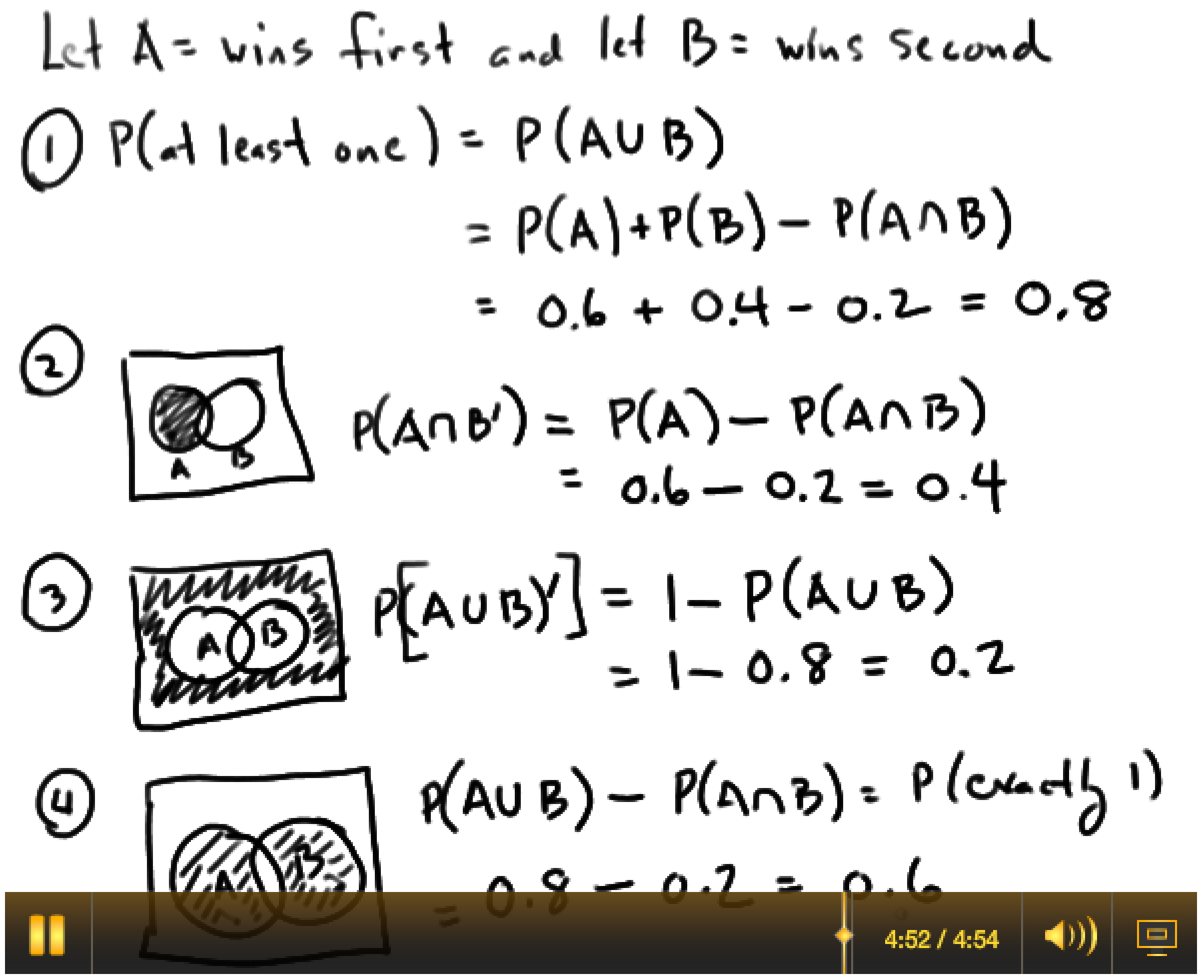
$$\Rightarrow P(A' \cap B) = P(B) - P(A \cap B)$$

Return to (*):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example

- A company has bid on two large construction projects. The company president believes that the probability of winning the first contract is 0.6, the probability of winning the second contract is 0.4, and the probability of winning both contracts is 0.2.
- What is the probability that the company wins at least one contract?
- What is the probability that the company wins the first contract but not the second contract?
- What is the probability that the company wins neither contract?
- What is the probability that the company wins exactly one contract?



Let A = wins first and let B = wins second

① $P(\text{at least one}) = P(A \cup B)$

$= P(A) + P(B) - P(A \cap B)$

$= 0.6 + 0.4 - 0.2 = 0.8$

② $P(A \cap B') = P(A) - P(A \cap B)$

$= 0.6 - 0.2 = 0.4$

③ $P[(A \cup B)'] = 1 - P(A \cup B)$

$= 1 - 0.8 = 0.2$

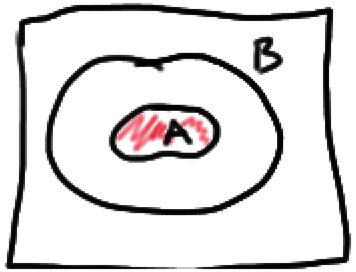
④ $P(A \cup B) - P(A \cap B) = P(\text{exactly } 1)$

$= 0.8 - 0.2 = 0.6$

Example

- f it is known that $A \subseteq B$, what can be definitively said about $P(A \cap B)$?

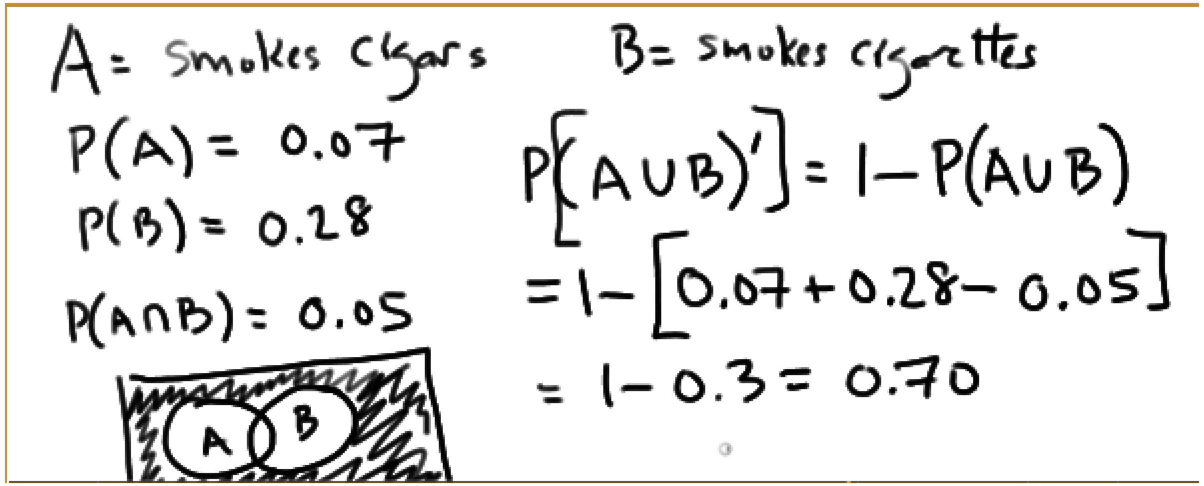
$$A \subseteq B \Rightarrow P(A) \leq P(B)$$



$$0 \leq P(A) = P(A \cap B) \leq P(B)$$

Example

- If 7% of the population smokes cigars, 28% of the population smokes cigarettes, and 5% of the population smokes both, what percentage of the population smokes neither cigars nor cigarettes?



A = smokes cigars B = smokes cigarettes

$P(A) = 0.07$

$P(B) = 0.28$

$P(A \cap B) = 0.05$

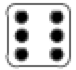
$$P[(A \cup B)'] = 1 - P(A \cup B)$$

$$= 1 - [0.07 + 0.28 - 0.05]$$

$$= 1 - 0.3 = 0.70$$

Multiplication principle

A Six and a head



S = {1H, 1T, 2H, 2T, ...}

6 x 2 = 12 possible outcomes

P(A) = N(A) / N(S)

The Multiplication Principle. If there are:
n_1 outcomes of a random experiment E_1
n_2 outcomes of a random experiment E_2
... and ...
n_m outcomes of a random experiment E_m
then there are $n_1 \times n_2 \times \ldots \times n_m$ outcomes of the composite experiment $E_1E_2 \ldots E_m$.

Permutations

- Suppose there are *n* positions to be filled with *n* different objects, in which there are:
- *n* choices for the 1st position
- $n - 1$ choices for the 2nd position
- $n - 2$ choices for the 3rd position
- ... and ...
- 1 choice for the last position
- The Multiplication Principle tells us there are then in general:
- $n \times (n - 1) \times (n - 2) \times \ldots \times 1 = n!$
- ways of filling the *n* positions.

Definition. A **permutation of *n* objects** is an ordered arrangement of the *n* objects.

Example: If there are 4 positions and 4 people to take them, for the first position there are 4 options, for the next there are 3 options since the first one is already chosen, etc.
4x3x2x1 possible ways

Permutations

- Suppose there are *r* positions to be filled with *n* different objects, in which there are:
- *n* choices for the 1st position
- *n* − 1 choices for the 2nd position
- *n* − 2 choices for the 3rd position
- ... and ...
- *n* − (r − 1) choices for the last position
- The Multiplication Principle tells us there are in general:
- $n \times (n-1) \times (n-2) \times \ldots \times [n-(r-1)]$
- ways of filling the *r* positions.

A **permutation of *n* objects taken *r* at a time**

nPr=n! / (n−r)!

$$\underline{n} \times \underline{n-1} \times \underline{n-2} \times . . \times \underline{n-(r-1)} \cdot \left[\frac{(n-r)\cdot(n-r-1)\cdots 1}{(n-r)\cdot(n-r-1)\cdots 1}\right] = \frac{n!}{(n-r)!}$$

Example:

If there are 4 people but 2 positions to take:

4x3 ways

Combinations

- Example:
- Maria has three tickets for a concert. She'd like to use one of the tickets herself. She could then offer the other two tickets to any of four friends (Ann, Beth, Chris, Dave). How many ways can 2 people be selected from 4 to go to a concert?

A B C D

ORDER MATTERS:

AB BA BC CB
AC CA BD DB
AD DA CD DC

12 ways

$$_4P_2 = \frac{4!}{(4-2)!} = \frac{24}{2} = 12$$

ORDER DOESN'T MATTER:

AB BC
AC BD
AD CD

6 ways

Combinations

Definition. The number of *unordered* subsets, called a **combination of *n* objects taken *r* at a time**, is:

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

We say “***n* choose *r***.”

The *r* represents the number of objects you'd like to select (without replacement and without regard to order) from the *n* objects you have.

Example

- Twelve (12) patients are available for use in a research study. Only seven (7) should be assigned to receive the study treatment. How many different subsets of seven patients can be selected?

$$\binom{12}{7} = \frac{12!}{7!(12-7)!} = 792$$

Example

- Let's use a standard deck of cards containing 13 face values (Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, and King) and 4 different suits (Clubs, Diamonds, Hearts, and Spades) to play five-card poker. If you are dealt five cards, what is the probability of getting a "full-house" hand containing three kings and two aces (KKKAA)?

$$P(KKKAA) = \frac{n(KKKAA)}{n(S)} = \frac{24}{2,598,960}$$

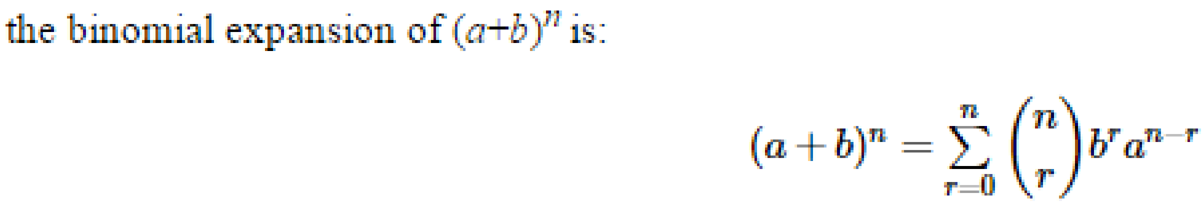
$$n(S) = n(5\text{ card hands}) = \binom{52}{5} = \frac{52!}{5!\,47!}$$

$$= 2,598,960$$

$$n(KKKAA) = \binom{4}{3} \times \binom{4}{2}$$

$$= \frac{4!}{1!\,3!} \times \frac{4!}{2!\,2!} = 4 \times 6 = 24$$

Binomial coefficients



the binomial expansion of $(a+b)^n$ is:

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} b^r a^{n-r}$$

$$(a+b)^2 = (a+b)(a+b) = a^2 + ab + ab + b^2$$
$$= a^2 + 2ab + b^2$$

$$(a+b)^3 = (a+b)(a+b)(a+b)$$
$$= a^3 + a^2b + a^2b + a^2b + b^2a + b^2a + b^2a + b^3$$
$$= a^3 + 3a^2b + 3b^2a + b^3$$

$$(a+b)^n = \underbrace{(a+b)(a+b)\cdots(a+b)}_{n \text{ factors}}$$

First, choose b from r of the n factors, and multiply them together: b^r

It leaves us with $n-r$ a's, multiply them together: a^{n-r}

Number of ways of choosing r b's from n factors: $\binom{n}{r}$

Multiplication principle: $\binom{n}{r} b^r a^{n-r}$

Now do it for $r = 0, 1, \ldots, n$ and add: $\sum_{r=0}^{n} \binom{n}{r} b^r a^{n-r}$

Distinguishable permutations

Example

Suppose we toss a gold dollar coin 8 times. What is the probability that the sequence of 8 tosses yields 3 heads (H) and 5 tails (T)?

Solution. Two such sequences, for example, might look like this:

HHHTTTTT or this **HTHTHTTT**

Assuming the coin is fair, and thus that the outcomes of tossing either a head or tail are equally likely, we can use the classical approach to assigning the probability. The Multiplication Principle tells us that there are:

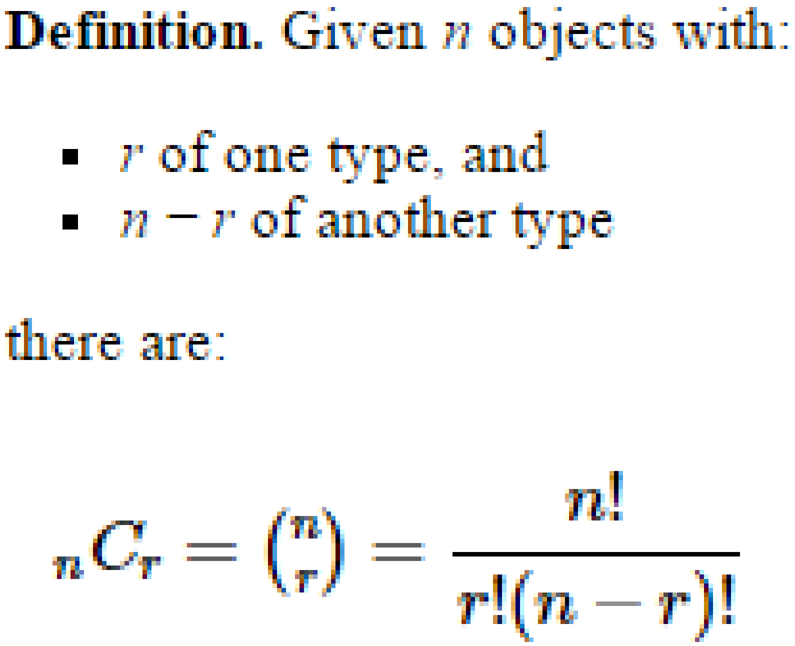
$$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$$

or 256 possible outcomes in the sample space of 8 tosses. (Can you imagine enumerating all 256 possible outcomes?) Now, when counting the number of sequences of 3 heads and 5 tosses, we need to recognize that we are dealing with arrangements or permutations of the letters, since order matters, but in this case not all of the objects are distinct. We can think of *choosing* (note that choice of word!) $r = 3$ positions for the heads (H) out of the $n = 8$ possible tosses. That would, of course, leave then $n - r = 8 - 3 = 5$ positions for the tails (T). Using the formula for a combination of n objects taken r at a time, there are therefore:

$$\binom{8}{3} = \frac{8!}{3!5!} = 56$$

distinguishable permutations of 3 heads (H) and 5 tails (T). The probability of tossing 3 heads (H) and 5 tails (T) is thus 56/256 = 0.22.

Distinguishable permutations



Definition. Given n objects with:

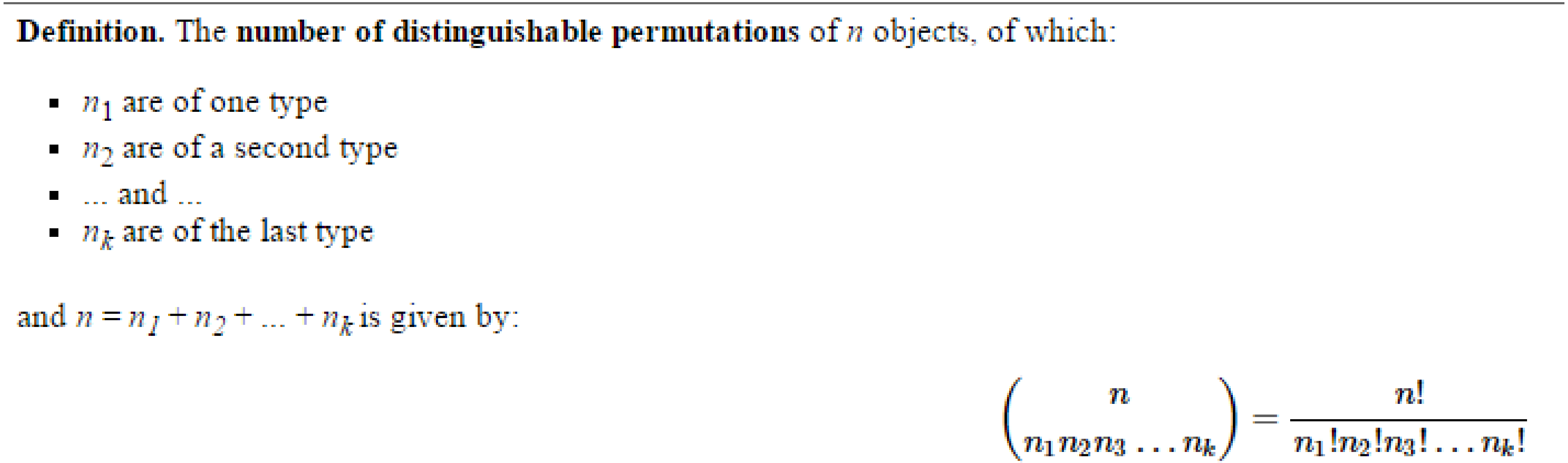
- r of one type, and
- $n-r$ of another type

there are:

$$_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

distinguishable permutations of the *n* objects.

Distinguishable permutations



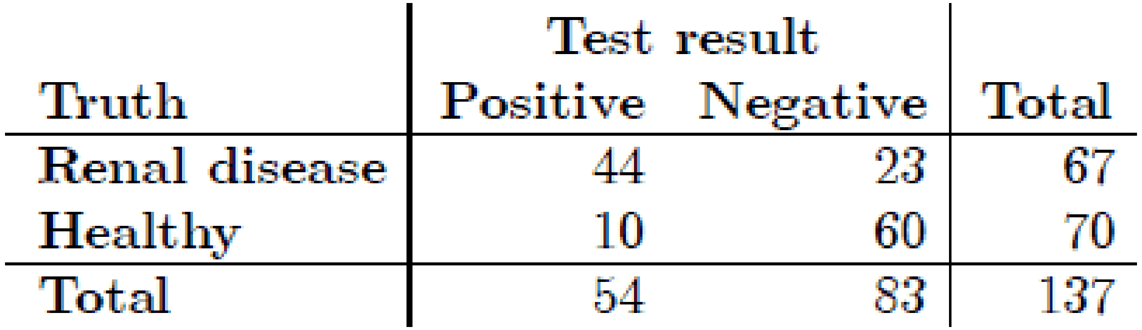
Definition. The **number of distinguishable permutations** of n objects, of which:

- n_1 are of one type
- n_2 are of a second type
- ... and ...
- n_k are of the last type

and $n = n_1 + n_2 + \ldots + n_k$ is given by:

$$\binom{n}{n_1 n_2 n_3 \ldots n_k} = \frac{n!}{n_1!n_2!n_3!\ldots n_k!}$$

Conditional probability

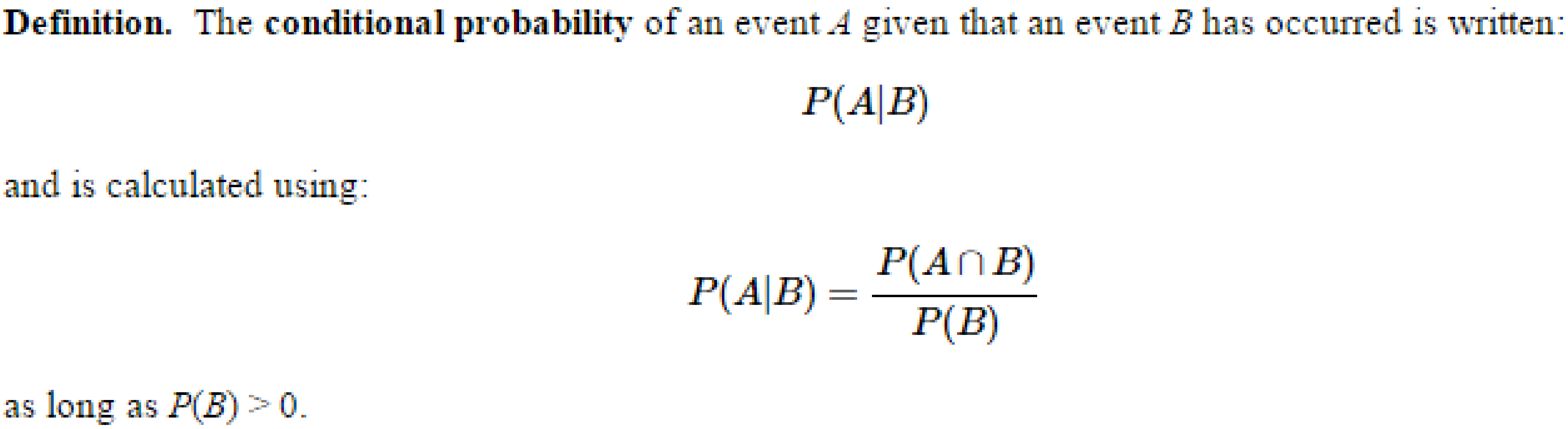


	Test result		
Truth	Positive	Negative	Total
Renal disease	44	23	67
Healthy	10	60	70
Total	54	83	137

- $P(T+) = 54/137$
- $P(D) = 67/137$
- If a person has renal disease, what is the probability that he/she tests positive for the disease?
- $P(T+ \mid D) = 44/67 = 0.65$

$$P(T+|D) = \frac{N(T+\cap D)/N(S)}{N(D)/N(S)} = \frac{P(T^+\cap D)}{P(D)}$$

Conditional Probability



Definition. The **conditional probability** of an event A given that an event B has occurred is written:

$$P(A|B)$$

and is calculated using:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

as long as $P(B) > 0$.

Conditional probability

- **Sensitivity**
- If a person has renal disease, what is the probability of testing positive?

- "**positive predictive value**" of a diagnostic test
- If I receive a positive test, what is the probability that I actually have the disease?

$$P(D|T+) = \frac{P(D \cap T+)}{P(T+)} = \frac{44/137}{54/137}$$

$$= \frac{44}{54} = 0.81$$

$$P(T+|D) = 0.65$$

NOTE: $P(D|T+) \neq P(T+|D)$

Properties of conditional probability

- Because conditional probability is just a probability, it satisfies the three axioms of probability. That is, as long as $P(B) > 0$:
- $P(A \mid B) \geq 0$
- $P(B \mid B) = 1$
- If $A_1, A_2, \ldots A_k$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \ldots \cup A_k \mid B) = P(A_1 \mid B) + P(A_2 \mid B) + \ldots + P(A_k \mid B)$ and likewise for infinite unions.

① $$P(A|B) = \frac{P(A\cap B) \geq 0}{P(B) > 0} \geq 0$$

② $$P(B|B) = \frac{P(B\cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

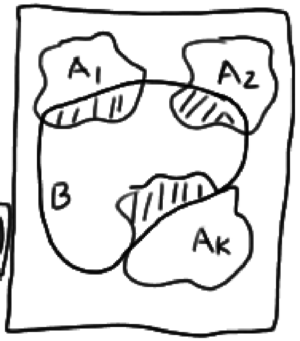
$$P[(A_1 \cup A_2 \cup \cdots \cup A_k)|B]$$

$$\overset{CP}{=} \frac{P[(A_1 \cup A_2 \cup \cdots \cup A_k) \cap B]}{P(B)}$$

$$\overset{VENN}{=} \frac{P[(A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_k \cap B)]}{P(B)}$$

$$\overset{ME}{=} \frac{P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_k \cap B)}{P(B)}$$

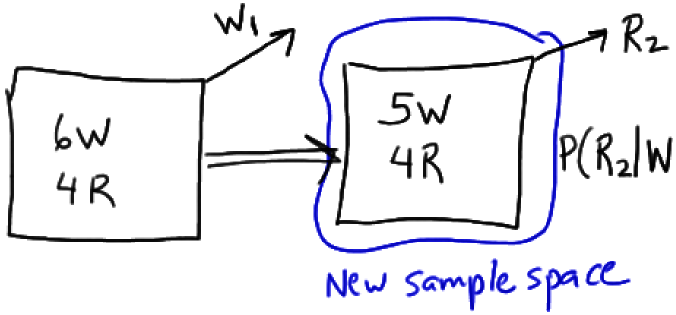
$$\overset{CP}{=} P(A_1|B) + P(A_2|B) + \cdots + P(A_k|B)$$



Examples

- A box contains 6 white balls and 4 red balls. We randomly (and without replacement) draw two balls from the box. What is the probability that the second ball selected is red, given that the first ball selected is white?

- What is the probability that both balls selected are red?

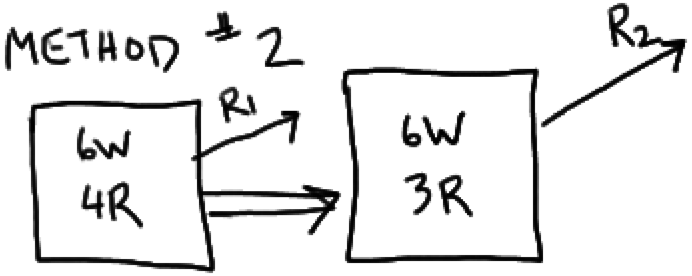


$P(R_2|W_1) = 4/9$

METHOD # 1

$$P(R_1 \cap R_2) = \frac{\binom{4}{2}\binom{6}{0}}{\binom{10}{2}} = \frac{6}{45}$$

METHOD # 2



$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1) = \frac{4}{10} \cdot \frac{3}{9} = \frac{12}{90}$$

$$= \frac{6}{45}$$

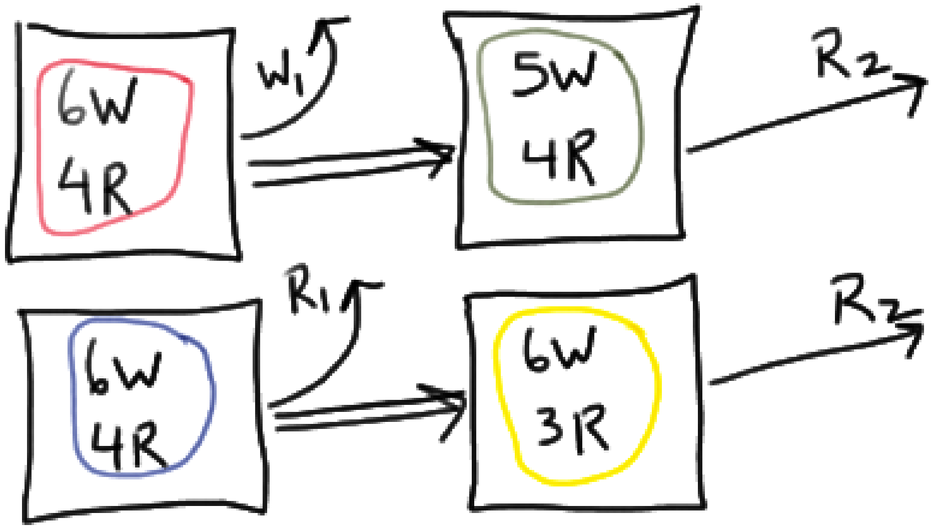
rearranging $P(R_2|R_1) = \frac{P(R_1 \cap R_2)}{P(R_1)}$

Multiplication rule

- **Definition.** The probability that two events A and B both occur is given by the **multiplication rule** as:
- $P(A \cap B) = P(A \mid B) \times P(B)$
- or by:
- $P(A \cap B) = P(B \mid A) \times P(A)$

Example

- A box contains 6 white balls and 4 red balls. We randomly (and without replacement) draw two balls from the box. What is the probability that the second ball selected is red?



$$P(R_2) = P\left[(W_1 \cap R_2) \cup (R_1 \cap R_2)\right]$$

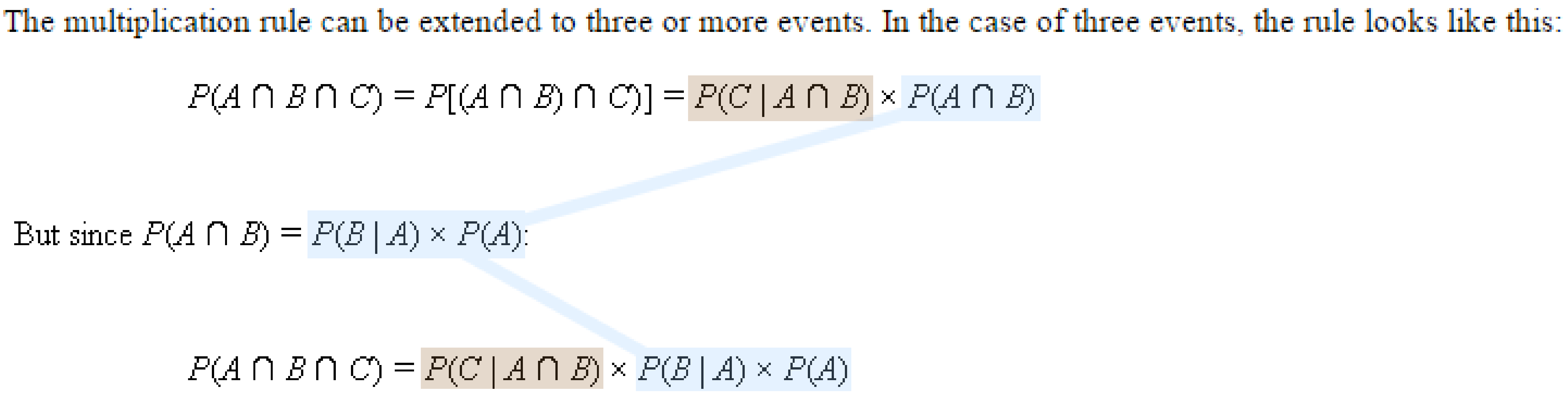
$$\overset{ME}{=} P(W_1 \cap R_2) + P(R_1 \cap R_2)$$

$$\overset{MR}{=} P(W_1) \cdot P(R_2 | W_1) + P(R_1) \cdot P(R_2 | R_1)$$

$$= \frac{6}{10} \cdot \frac{4}{9} + \frac{4}{10} \cdot \frac{3}{9}$$

$$= 24/90 + 12/90 = 36/90$$

Extended multiplication rule



The multiplication rule can be extended to three or more events. In the case of three events, the rule looks like this:

$$P(A \cap B \cap C) = P[(A \cap B) \cap C)] = P(C \mid A \cap B) \times P(A \cap B)$$

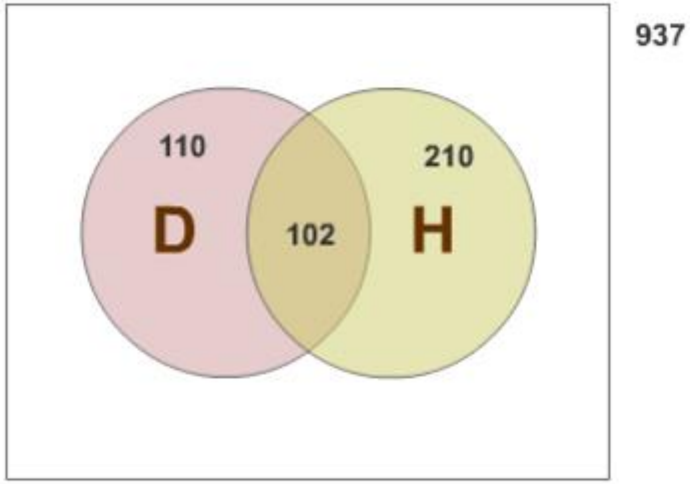
But since $P(A \cap B) = P(B \mid A) \times P(A)$:

$$P(A \cap B \cap C) = P(C \mid A \cap B) \times P(B \mid A) \times P(A)$$

Example

- Medical records reveal that of the 937 men who died in a particular region in 1999:
- 212 of the men died of causes related to heart disease,
- 312 of the men had at least one parent with heart disease
- Of the 312 men with at least one parent with heart disease, 102 died of causes related to heart disease. Using this information, if we randomly select a man from the region, what is the probability that he dies of causes related to heart disease given that neither of his parents died from heart disease? If we define two events as such:
- Let H = the event that at least one of the parents of a randomly selected man died of causes related to heart disease
- Let D = the event that a randomly selected man died of causes related to heart disease
- then we are looking for the following conditional probability:
- $P(D \mid H')$

Solution



		H ≥1 parent	H' neither	
CAUSE OF DEATH	Heart (D)	102	110	212
	Not (D')			
		312	625	937

$$P(D|H') = \frac{P(D \cap H')}{P(H')} = \frac{110/937}{625/937} = \frac{110}{625}$$

Independent Events

- A couple plans to have three children. What is the probability that the second child is a girl? And, what is the probability that the second child is a girl given that the first child is a girl?

$S = \{GGG, GBB, BGB, BBG, GGB, GBG, BGG, BBB\}$

Let A = event first is girl

$= \{GGG, GBB, GGB, GBG\}$ $\quad P(A) = \frac{4}{8} = \frac{1}{2}$

Let B = event second is girl

$= \{GGG, BGB, GGB, BGG\}$ $\quad P(B) = \frac{4}{8} = \frac{1}{2}$

$A \cap B = \{GGB, GGG\}$ $\quad P(A \cap B) = \frac{2}{8} = \frac{1}{4}$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2/8}{4/8} = \frac{2}{4} = \frac{1}{2} = P(B)$$

Independent events

- **Definition.** Events A and B are **independent events** if the occurrence of one of them does not affect the probability of the occurrence of the other. That is, two events are independent if either: $P(B|A) = P(B)$
- (provided that $P(A) > 0$) or:
- $P(A|B) = P(A)$
- (provided that $P(B) > 0$).

Independent events

Now, since independence tells us that $P(B|A) = P(B)$, we can substitute $P(B)$ in for $P(B|A)$ in the formula given to us by the multiplication rule:

$P(A \cap B) = P(A) \times P(B |A)$

yielding:

$P(A \cap B) = P(A) \times P(B)$.

- **Definition.** Events *A* and *B* are **independent events** if and only if :
- $P(A \cap B) = P(A) \times P(B)$
- Otherwise, *A* and *B* are called **dependent events.**

Theorems

- If *A* and *B* are independent events, then the events *A* and *B′* are also independent.
- If *A* and *B* are independent events, then the events *A′* and *B* are also independent.
- If *A* and *B* are independent events, then the events *A′* and *B′* are also independent.

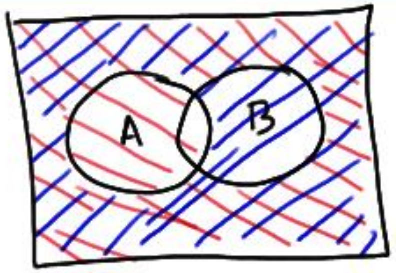
$$P(A\cap B') \overset{MR}{=} P(A)\cdot P(B'|A)$$

$$\overset{AX}{=} P(A)\cdot[1-P(B/A)]$$

$$\overset{IND}{=} P(A)\cdot[1-P(B)]$$

$$= P(A)\cdot P(B')$$

$\Rightarrow$ A and B′ are independent



$$A'\cap B' = (A\cup B)'$$

$$\boxed{P(A'\cap B')} = P[(A\cup B)']$$

$$\overset{AX}{=} 1-P(A\cup B)$$

$$= 1-[P(A)+P(B)-P(A\cap B)]$$

$$= 1-P(A)-P(B)+P(A\cap B)$$

$$= (1-P(A)\cdot(1-P(B))$$

$$\overset{AX}{=} \boxed{P(A')\cdot P(B')}$$

$\Rightarrow$ A′ and B′ are independent

Example

- A nationwide poll determines that 72% of the American population loves eating pizza. If two people are randomly selected from the population, what is the probability that the first person loves eating pizza, while the second one does not?
- $P(A \cap B') = 0.72 \times (1 - 0.72) = 0.202$

Mutual independence

- **Definition.** Three events A, B, and C are **mutually independent** if and only if the following two conditions hold:
- (1) The events are pairwise independent. That is,
- $P(A \cap B) = P(A) \times P(B)$ and...
- $P(A \cap C) = P(A) \times P(C)$ and...
- $P(B \cap C) = P(B) \times P(C)$
- (2) $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

Bayes' theorem

- A desk lamp produced by The Luminar Company was found to be defective (*D*). There are three factories (*A*, *B*, *C*) where such desk lamps are manufactured. A Quality Control Manager (QCM) is responsible for investigating the source of found defects. This is what the QCM knows about the company's desk lamp production and the possible source of defects:

Factory	**% of total production**	**Probability of defective lamps**
A	$0.35 = P(A)$	$0.015 = P(D \mid A)$
B	$0.35 = P(B)$	$0.010 = P(D \mid B)$
C	$0.30 = P(C)$	$0.020 = P(D \mid C)$

- If a randomly selected lamp is defective, what is the probability that the lamp was manufactured in factory C?

Bayes' theorem

$$P(C|D) \overset{CP}{=} \frac{P(C\cap D)}{P(D)} \overset{MR}{=} \frac{P(D|c)\cdot P(C)}{P(D)}$$

$$= \frac{P(D|c)\cdot P(c)}{P[(D\cap A)\cup(D\cap B)\cup(D\cap C)]}$$

$$\overset{\substack{ME\\MR}}{=} \frac{P(D|c)\cdot P(c)}{P(D|A)\cdot P(A)+P(D|B)\cdot P(B)+P(D|C)\cdot P(c)}$$

$$= \frac{(0.02)(0.30)}{(0.015)(0.35)+(0.01)(0.35)+(0.02)(0.30)}$$

$$= \frac{0.006}{0.01475} = 0.407$$

Bayes’ theorem

- if a randomly selected lamp is defective, what is the probability that the lamp was manufactured in factory A? And, if a randomly selected lamp is defective, what is the probability that the lamp was manufactured in factory B?

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{P(D|A) \times P(A)}{P(D)} = \frac{(0.015)(0.35)}{0.01475} = 0.356$$

$$P(B|D) = \frac{P(B \cap D)}{P(D)} = \frac{P(D|B) \times P(B)}{P(D)} = \frac{(0.01)(0.35)}{0.01475} = 0.237$$

Bayes' theorem

- The probabilities $P(A)$, $P(B)$ and $P(C)$ are often referred to as **prior probabilities**, because they are the probabilities of events A, B, and C that we know *prior* to obtaining any additional information. The conditional probabilities $P(A \mid D)$,$P(B \mid D)$, and $P(C \mid D)$ are often referred to as **posterior probabilities**, because they are the probabilities of the events *after* we have obtained additional information.
- As a result of our work, we determined:
- $P(C \mid D) = 0.407$
- $P(B \mid D) = 0.237$
- $P(A \mid D) = 0.356$

Bayes' Theorem. Let the m events $B_1, B_2, \ldots, B_m$ constitute a **partition** of the sample space S. That is, the B_i are mutually exclusive:

$$B_i \cap B_j = \emptyset \text{ for } i \neq j$$

and exhaustive:

$$S = B_1 \cup B_2 \cup \ldots \cup B_m$$

Also, suppose the prior probability of the event B_i is positive, that is, $P(B_i) > 0$ for $i = 1, \ldots, m$. Now, if A is an event, then A can be written as the union of m mutually exclusive events, namely:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_m)$$

Therefore:

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_m) \\ &= \sum_{i=1}^{m} P(A \cap B_i) \\ &= \sum_{i=1}^{m} P(B_i) \times P(A|B_i) \end{aligned}$$

And so, as long as $P(A) > 0$, the posterior probability of event B_k given event A has occurred is:

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k) \times P(A|B_k)}{\sum_{i=1}^{m} P(B_i) \times P(A|B_i)}$$

Example

- A common blood test indicates the presence of a disease 95% of the time when the disease is actually present in an individual. Joe's doctor draws some of Joe's blood, and performs the test on his drawn blood. The results indicate that the disease is present in Joe.
- One-percent (that is, 1 in 100) people have the disease. That is, if D is the event that a randomly selected individual has the disease, then $P(D) = 0.01$.
- If H is the event that a randomly selected individual is disease-free, that is, healthy, then $P(H) = 1-P(D) = 0.99$.
- The **sensitivity** of the test is 0.95. That is, if a person has the disease, then the probability that the diagnostic blood test comes back positive is 0.95. That is, $P(T+ \mid D) = 0.95$.
- The **specificity** of the test is 0.95. That is, if a person is free of the disease, then the probability that the diagnostic test comes back negative is 0.95. That is, $\mathrm{P}(T- \mid H) = 0.95$.
- If a person is free of the disease, then the probability that the diagnostic test comes back positive is $1 - \mathrm{P}(T- \mid H) = 0.05$. That is, $P(T+ \mid H) = 0.05$.
- What is the **positive predictive value** of the test? That is, given that the blood test is positive for the disease, what is the probability that Joe actually has the disease?

$$P(D|T^+) \overset{CP}{=} \frac{P(D\cap T^+)}{P(T^+)}$$

$$\overset{MR}{=} \frac{P(T^+|D)\cdot P(D)}{P\left[(T^+\cap D)\cup(T^+\cap H)\right]}$$

$$\overset{\substack{MR\\ME}}{=} \frac{P(T^+|D)\cdot P(D)}{P(T^+|D)\cdot P(D)+P(T^+|H)\cdot P(H)}$$

$$= \frac{(0.95)(0.01)}{(0.95)(0.01)+(0.05)(0.99)}$$

$$= 0.0095/0.059 = \boxed{0.16}$$

- If a second blood test on Joe comes back positive for the disease, what is the probability that Joe actually has the disease now?

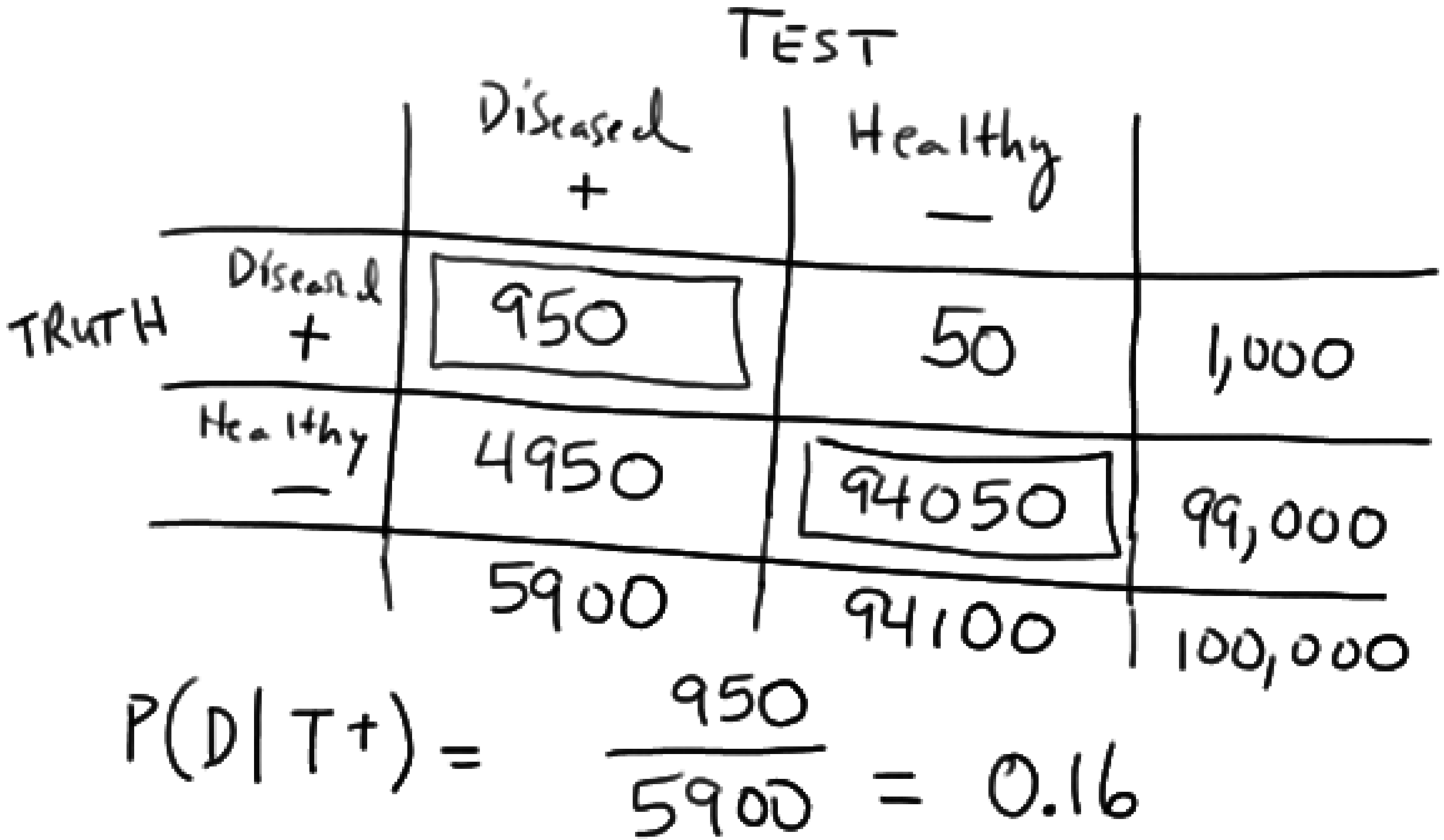
$$P(D|T^+) \overset{BR}{=} \frac{P(T^+|D)\cdot\boxed{P(D)}}{P(T^+|D)\cdot\boxed{P(D)} + P(T^+|H)\cdot\boxed{P(H)}}$$

$$= \frac{(0.95)(0.16)}{(0.95)(0.16) + (0.05)(0.84)}$$

$$= \frac{0.152}{0.194}$$

$$= \boxed{0.78}$$

Alternative way



		TEST		
		Diseased +	Healthy —	
TRUTH	Diseased +	$\boxed{950}$	50	1,000
	Healthy —	4950	$\boxed{94050}$	99,000
		5900	94100	100,000

$$P(D|T^+) = \frac{950}{5900} = 0.16$$