Chapter 3: The basic concepts of probability

Experiment: a measurement process that produces quantifiable results
(e.g. throwing two dice, dealing cards, at poker,
measuring heights of people, recording proton-proton collisions)

Outcome: a single result from a measurement
(e.g. the numbers shown on the two dice)

Sample space: the set of **all possible** outcomes from an experiment
(e.g. the set of all possible five-card hands)

The number of all possible outcomes may be
(a) **finite** (e.g. all possible outcomes from throwing a single die;
all possible 5-card poker hands)
(b) **countably infinite** (e.g. number of proton-proton events to be made before
a Higgs boson event is observed)
or (c) **constitute a continuum** **(**e.g. heights of people)

In case (a), the sample space is said to be **finite**
in cases (a) and (b), the sample space is said to be **discrete**
in case (c), the sample space is said to be **continuous**

In this chapter we consider discrete, mainly finite, sample spaces

An **event** is any subset of a sample set (including the empty set, and the whole set)

Two events that have no outcome in common are called **mutually exclusive** events.

In discussing discrete sample spaces, it is useful to use **Venn diagrams** and basic set-theory. Therefore we will refer to the **union** (A U B)**, intersection,** (A ∩ B) and **complement** ($\bar{A}$) of events A and B. We will also use set-theory relations such as

$$\overline{\mathrm{A} \cup \mathrm{B}} = \bar{\mathrm{A}} \cap \bar{\mathrm{B}}$$ (Such relations are often proved using Venn diagrams)

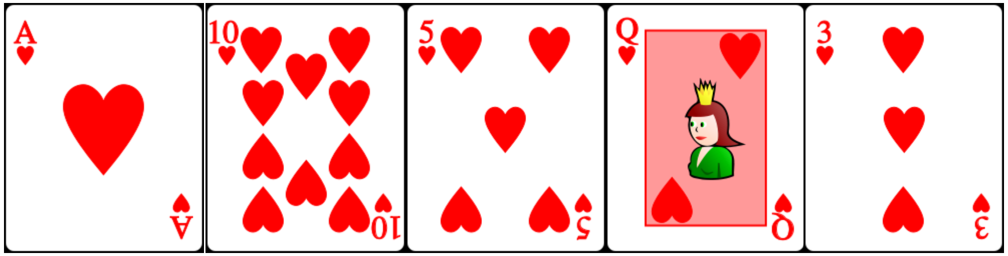
This is also called De Morgan's law, another half of De Mogan's law is:

$\overline{A \cap B} = \bar{A} \cup \bar{B}$

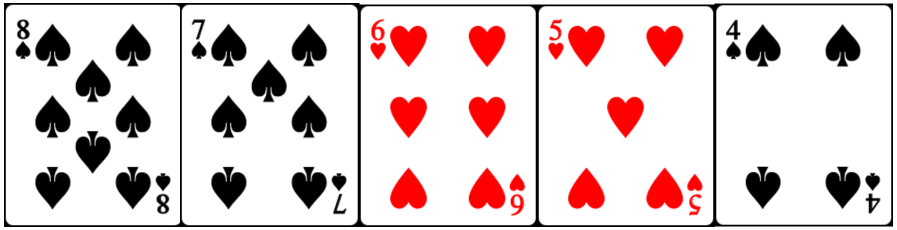
A dice has six sides, each side has a distinct number (1-6) dots

Some terminology used in card game

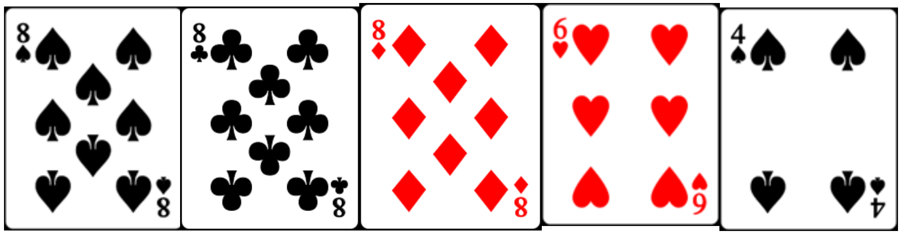
Flush: A flush is a hand of playing cards where all cards are of the same suit.



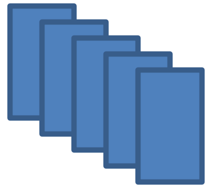
Straight:



Three of a kind:



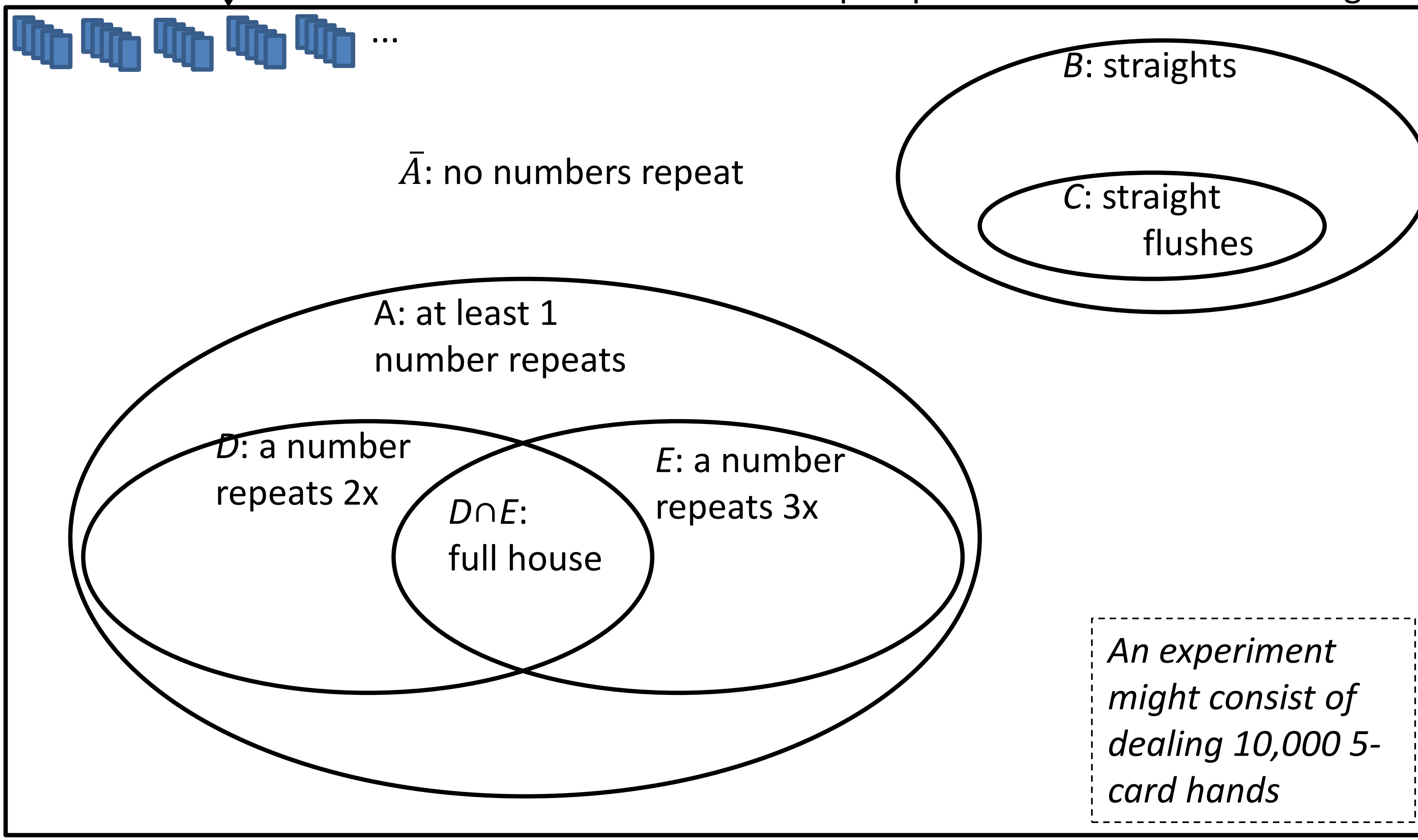
e.g.: outcome = 5-card poker hand



Event C (straight flush) has 40 outcomes

sample space S: 2,598,960 possible 5-card hands (2,598,960 outcomes)

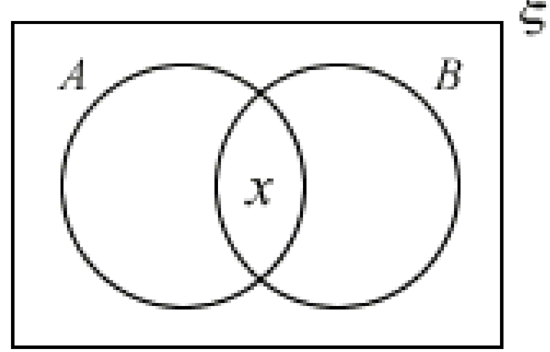
The sample space is drawn as a Venn diagram



Intersect

$A \cap B = x$

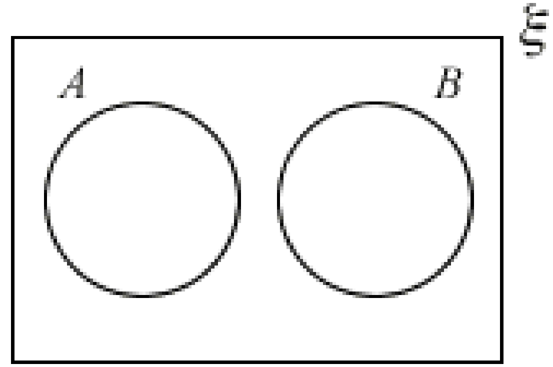
$A \cup B = n(A) + n(B) - x$



Disjoint

$A \cap B = \varnothing$

$A \cup B = n(A) + n(B)$

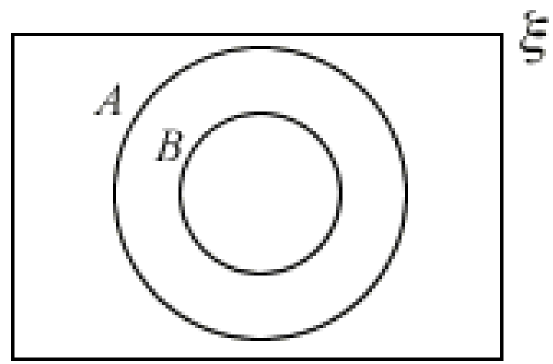


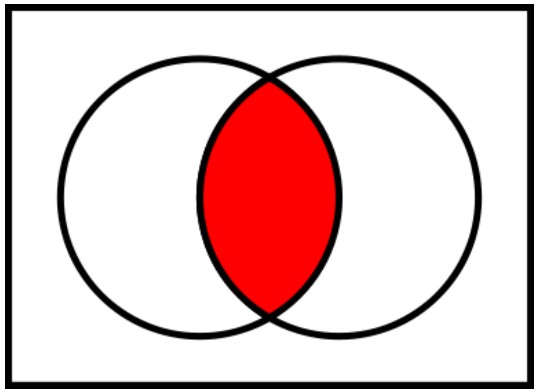
Subset

$B \subseteq A$

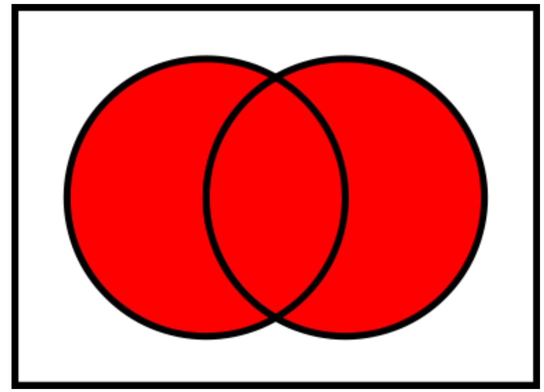
$A \cap B = B$

$A \cup B = A$

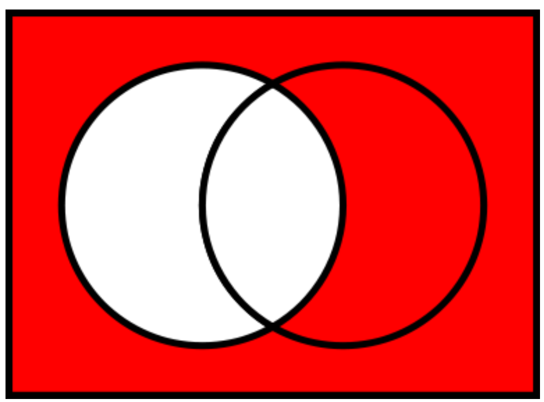




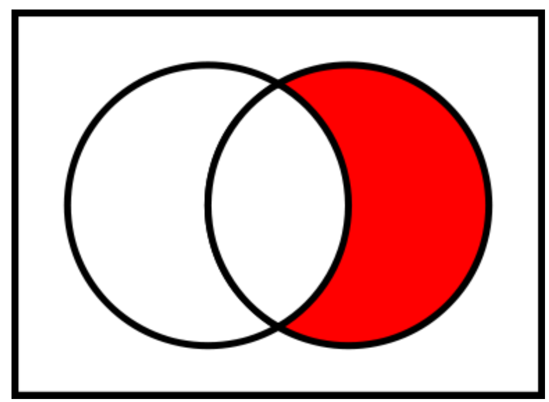
$A \cap B$



$A \cup B$

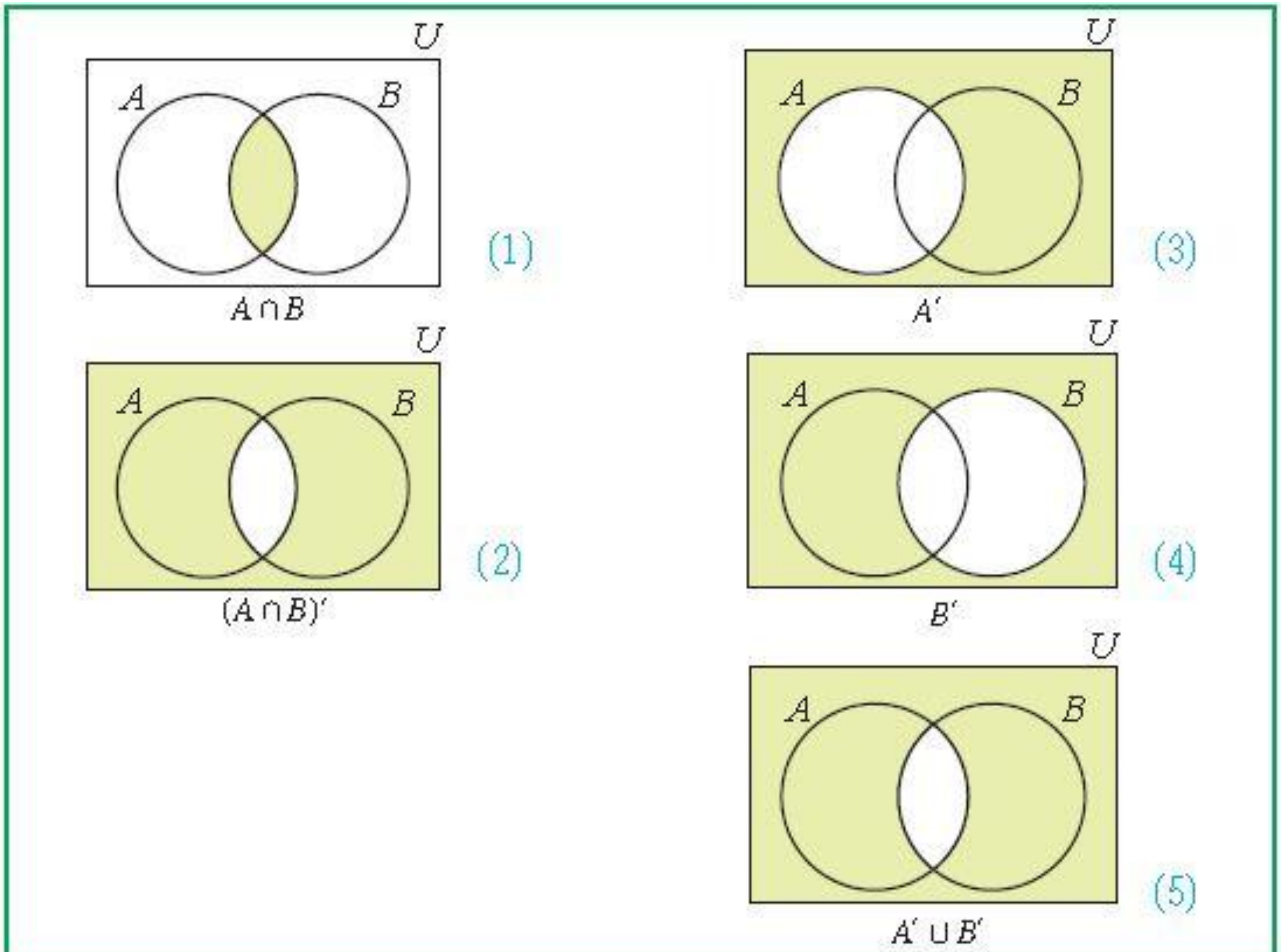


$A^c = U \setminus A$

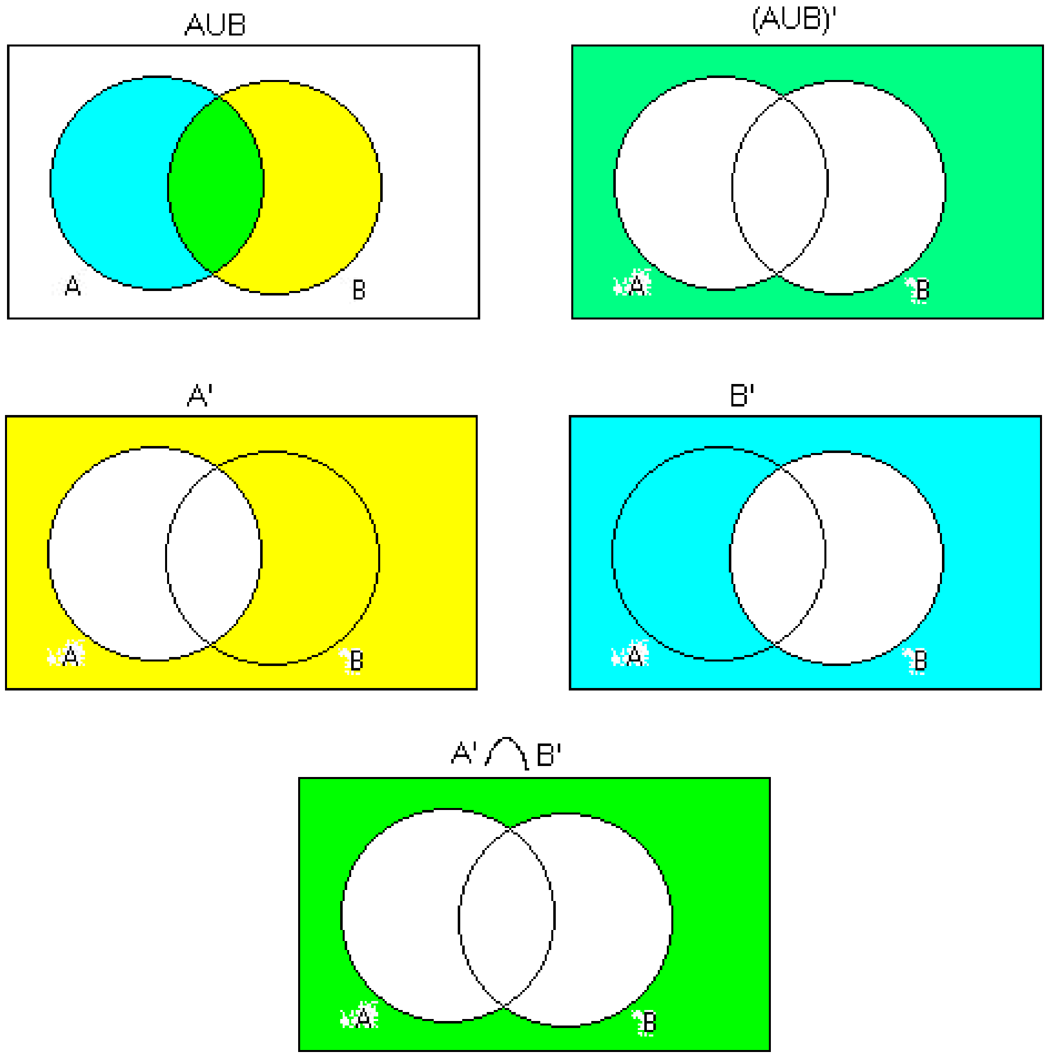


$A^c \cap B = B \setminus A$

De Morgan's Law (1): $\overline{A \cap B}=\bar{A} \cup \bar{B}$



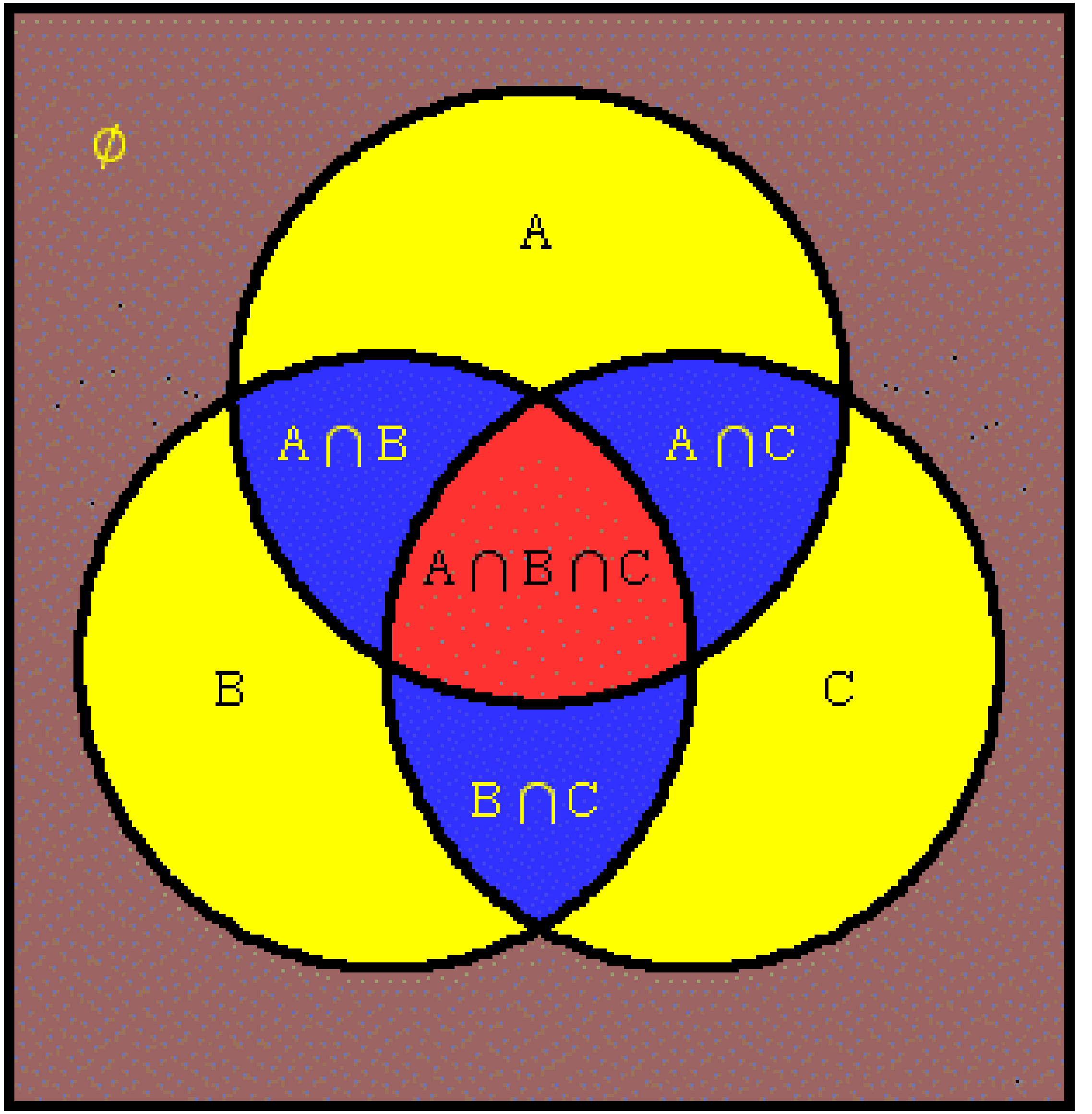
De Morgan's Law (2): $\overline{A \cup B}=\bar{A} \cap \bar{B}$



Example: U=(3,4,2,8,9,10,27,23,14)

A=(2,4,8)   B=(3,4,8,27)

$\bar{A}$=(3,9,10,27,23,14)
$\bar{B}$=(2,9,10,23,14)
$A \cup B$=(2,3,4,8,27)
$A \cap B$=(4,8)
$A - B$=(2)
$\overline{A \cup B}$=(9,10,23,14)
$\overline{A \cap B}$=(3,9,10,27,23,14,2)



∅
A
A∩B
A∩C
A∩B∩C
B
C
B∩C

The **classical definition of probability (classical probability concept)** states:

If there are m outcomes in a sample space (universal set), and all are equally likely of being the result of an experimental measurement, then the probability of observing an event (a subset) that contains s outcomes is given by $\frac{s}{m}$

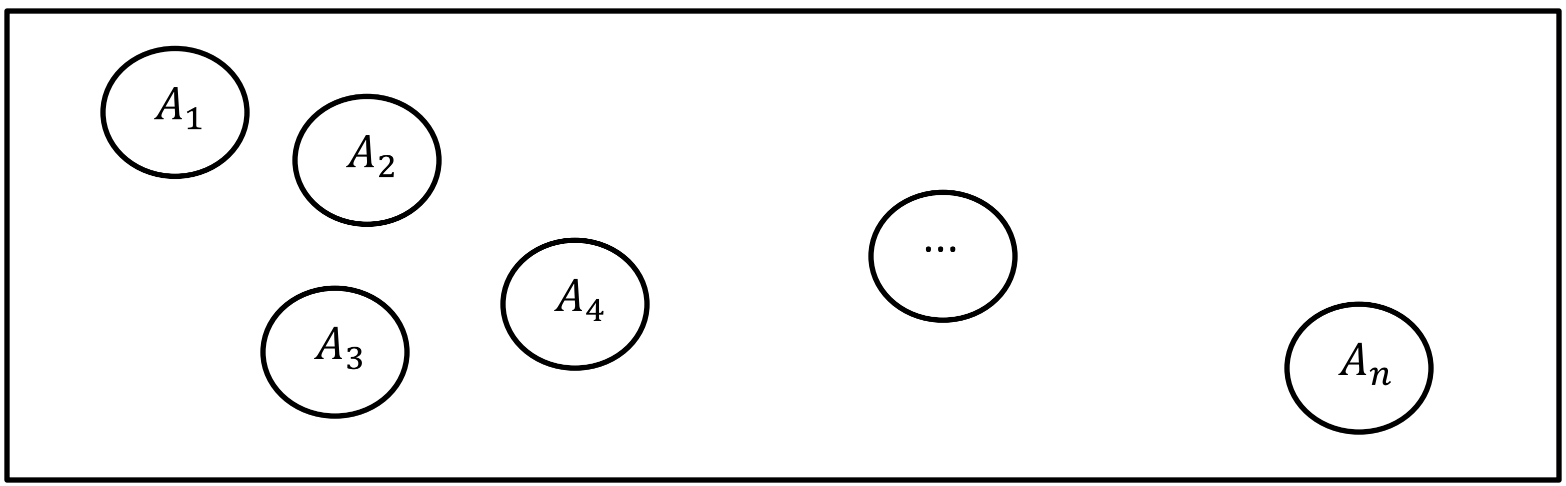
From the classical definition, we see that the ability to count the number of outcomes in an event, and the number of outcomes in the entire sample space (universal set) is of critical importance.

Counting principles:

The **addition principle:** If there are n_1 outcomes in event A_1,
n_2 outcomes in event A_2,
…
n_k outcomes in event A_k

and the events A_1, A_2, … A_k are mutually distinct (share no outcomes in common),

then the total number of outcomes in $A_1 \cup A_2 \cup \ldots \cup A_k$ is $n_1 + n_2 + \ldots + n_k$



A single outcome may consist of several distinct parts (e.g. an arrangement of 7 objects; throwing a red and a green die). Such outcomes are said to be **composite**

The **multiplication principle:** If a composite outcome can be described by a procedure that can be broken into k successive (ordered) stages such that there are

n_1 outcomes in stage 1,

n_2 outcomes in event 2,

…

n_k outcomes in event k

and if the number of outcomes in each stage is independent of the choices in previous stages and if the composite outcomes are all distinct

then the number of possible composite outcomes is $n_1 \cdot n_2 \cdot \ldots \cdot n_k$

e.g. suppose the composite outcomes of the trio (M.P,C) of class values for cars, where

M denotes the mileage class (M_1, M_2, or M_3)

P denotes the price class (P_1, or P_2)

C denotes the operating cost class (C_1, C_2, or C_3)

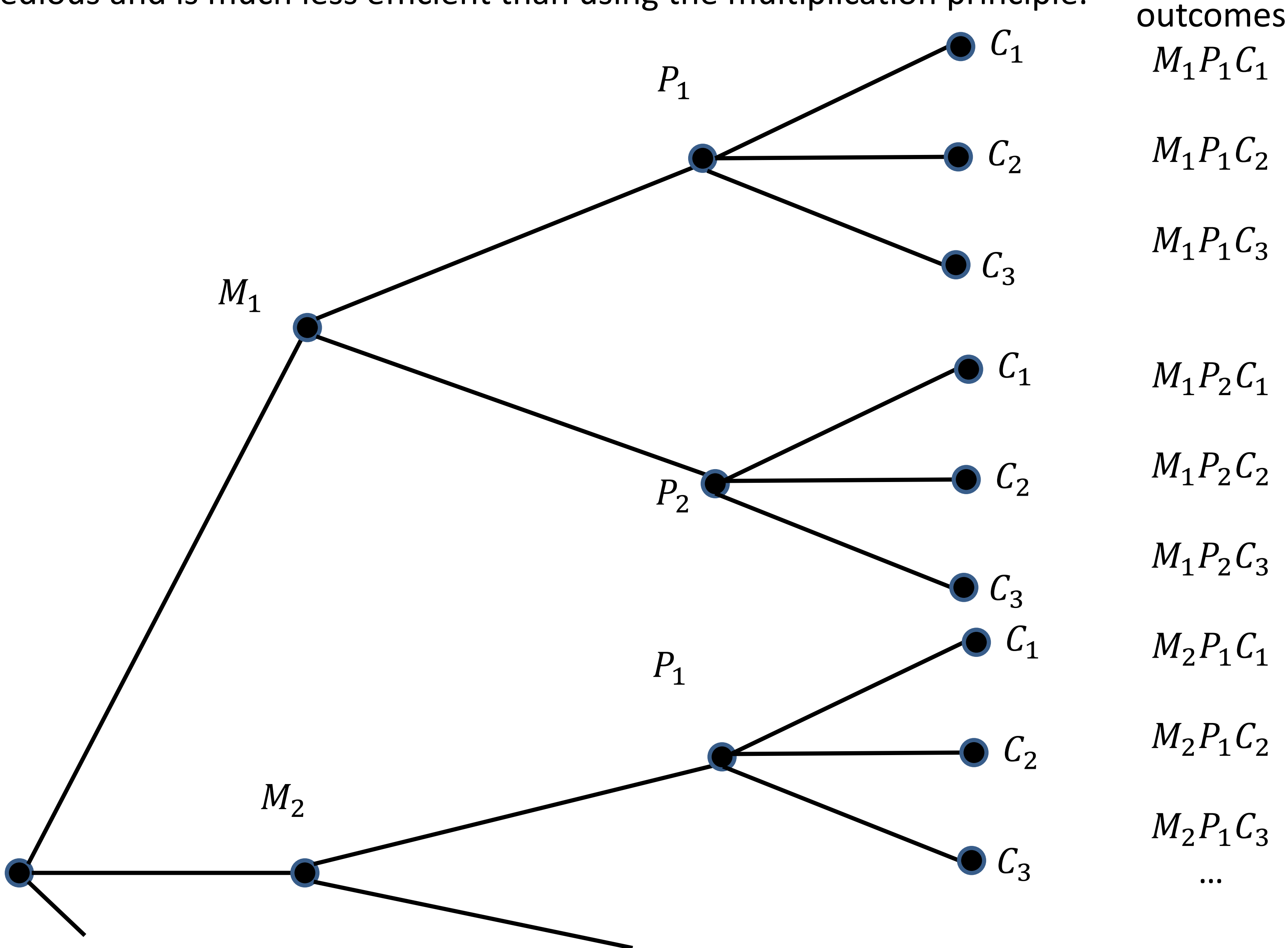
The outcome is clearly written as a 3-stage value

There are 3 outcomes in class M, 2 in class P and 3 in class C

The number of outcomes in class P does not depend on the choice made for M, etc

Then there will be $3 \cdot 2 \cdot 3 = 18$ distinct composite outcomes for car classification.

When the number of composite outcomes is relatively small, the counting can also be done via a **tree diagram** as illustrated in Fig. 3.5 (page 50) of the text. Such a method is tedious and is much less efficient than using the multiplication principle.



e.g. an outcome of an experiment consists of an operator using a machine to test a type of sample.

If there are 4 different operators, 3 different machines, and 8 different types of samples, how many experimental outcomes are possible?

Note: the conditions of the multiplication principle must be strictly adhered to for it to work.

e.g. the number of distinct outcomes obtained from throwing two identical six-sided dice cannot be obtained by considering this as a two stage process (the result from the first die and then the result from the second – since the outcomes from the two stages *are not* distinct. There *are not* 36 possible outcomes from throwing two identical six-sided dice; there are only 21 distinct outcomes.

e.g. the number of distinct outcomes obtained from throwing a red six-sided dice and a green six-sided dice can be determined by the multiplication principle. There are 36 possible outcomes in this case.

e.g. the number of 5-card poker hands comprised of a full house *can* be computed using the multiplication principle. A full house can be considered a two-stage hand, the first stage being the pair, the second stage being the three-of-a-kind. Thus the number of possible full house hands = (the number of pairs) x (the number of threes-of-a-kind)

e.g. the number of 5-card poker hands comprised of a full house that do not contain the 10's as the three-of-a-kind *cannot* be computed using the multiplication principle since the number of choices for the three-of-a-kind depends on whether-or-not the pair consists of 10's

Two common counting problems are

a) r – permutations of n distinct objects,

b) r – combinations of n distinct objects

r – permutations

Given n distinct objects, how many ways are there to arrange exactly r of the objects?

An arrangement is a sequence (ordered stages) of successive objects. An arrangement can be thought of as putting objects in slots. There is a distinct first slot, second slot, …., k'th slot, and so on. Putting object A in the first slot and B in the second is a distinct outcome from putting B in the first and A in the second. The number of possible choices for slot k+1 does not depend on *what* choice is used for slot k. Therefore the multiplication principle can be used.

Thus the number of ways to arrange r of the n distinct objects is

$$_nP_r = n\,(n-1)(n-2)\,\ldots.\,(n-r+1)$$

Multiplying and dividing by (n−r)!, this can be written

$$_nP_r = \frac{n\,(n-1)(n-2)\ldots.(n-r+1)(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

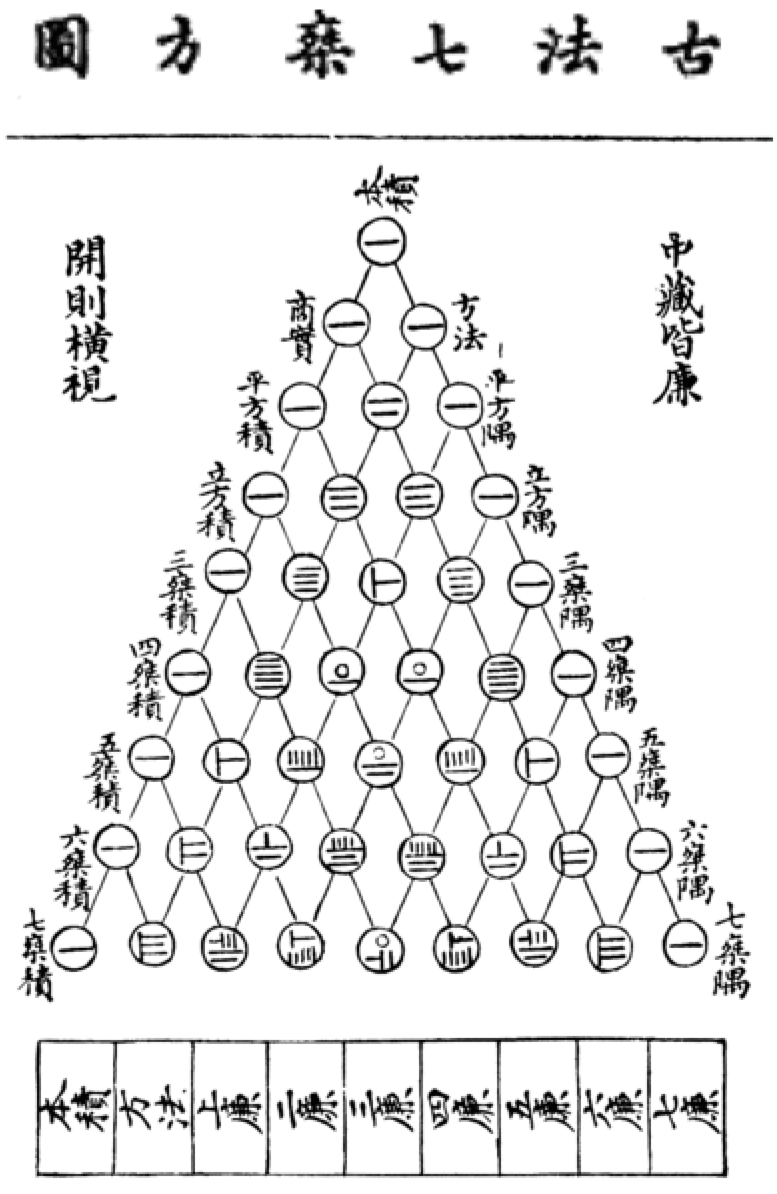
Theorem 3.2 The number of r – permutations of n distinct objects (that is the number of ways to arrange exactly r objects out of a set of n distinct objects) is

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

Note:

The number of ways to arrange *all* n objects is $\quad {}_nP_n = n!$ $\quad$ (as 0! ≡1)

The number of ways to arrange *zero* of the n objects is $\quad {}_nP_0 = \frac{n!}{n!} = 1$



Yang Hui in 1305

Pascal in 1655

Tartaglia in 1600's

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

....

Pascal triangle: the n-th row and the r-th element: $\binom{n}{r}$ with n and r start from 0 (not 1). Pascal triangle is symmetric because $\binom{n}{r}$=$\binom{n}{n-r}$. Another property is that the summation of the n-th row is 2^n

e.g. chose 3 people out of 180 to act as fire wardens. $\binom{180}{3}$ = $\frac{180!}{3!177!}$

e.g select 5 new faculty consisting of 2 chemists (from a pool of 21 applicants) and 3 physicists (from a different pool of 47 applicants)$\binom{21}{2}$x$\binom{47}{3}$

e.g. choose 5 people from a class of 143 and seat them in a row of 5 chairs at the front of the class $_{143}P_5$ = $\frac{143!}{138!}$

The **classical definition of probability**

If there are *m* outcomes in a sample space, and all are equally likely of being the result of an experimental measurement, then the probability of observing an event that contains *s* outcomes is given by $\frac{s}{m}$

e.g. Probability of drawing an ace from a deck of 52 cards.
sample space consists of 52 outcomes.
desired event (ace) is a set of 4 outcomes (number of desired outcomes is 4)
therefore the probability of getting an ace is 4/52 = 1/13 ≈ 0.0769 (7.69%)

e.g. There are 10 motors, two of which do not work and eight which do.

a) what is the probability of picking 2 working motors
   (8!/2!/6!)/(10!/2!/8!)=(8!2!8!)/(10!2!6!)=(8!8!)/(10!6!)
b) what is the probability of picking 1 working and 1 non-working motors
c) S=2*8, m=10!/2!/8! P=16*2!*8!/10!

Mathematically, in defining probabilities of events we are deriving a set function on a sample space. A set function assigns to each subset in the sample space a real number.

Example: Consider the set function that assigns to each subset (event) *A* the number *N*(A) of outcomes in the set. This set function is additive, that is, if two events *A* and *B* have no outcomes in common (are mutually exclusive), then N(A U B) = N(A) + N(B).

Counter-example: Fig. 3.7

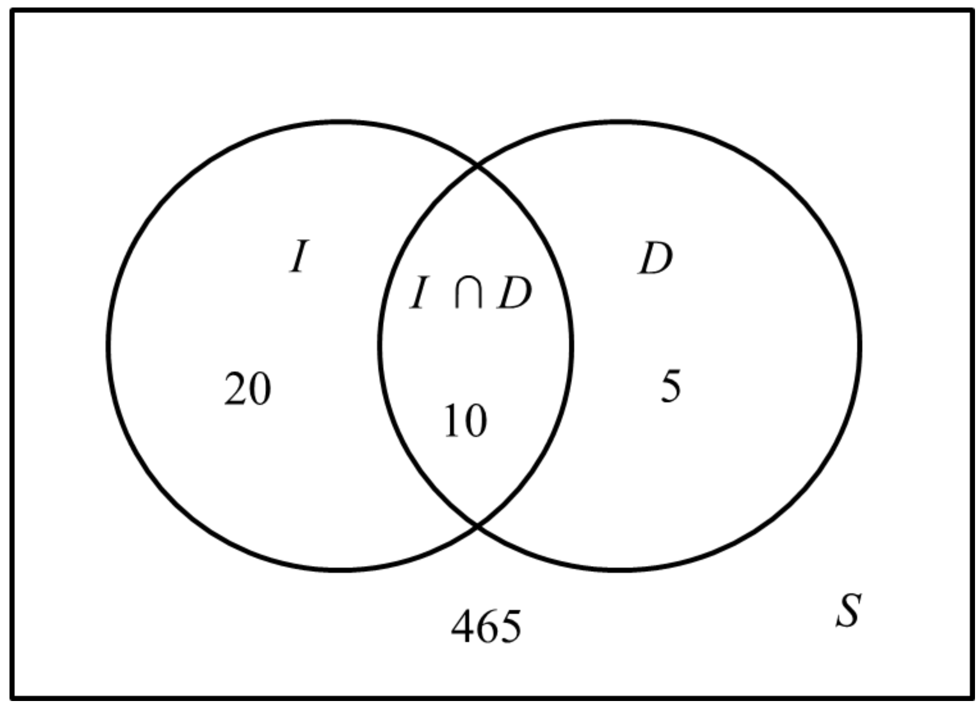


Fig. 3.7. Measurements on 500 machine parts
I = incompletely assembled, D = defective, S = satisfactory

$N(I \cup D) \neq N(I) + N(D)$ as N and D are not mutually exclusive

The axioms of probability

Let *S* be a finite sample space, *A* an event in *S*. We define *P*(*A*), the probability of *A*, to be the value of an additive set function *P*( ) that satisfies the following three conditions

Axiom 1 $0 \leq P(A) \leq 1$ for each event *A* in *S*

(probabilities are real numbers on the interval [0,1])

Axiom 2 $P(S) = 1$

(the probability of some event occurring from *S* is unity)

Axiom 3 If *A* and *B* are mutually exclusive events in *S*, then

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

(the probability function is an additive set function)

Note: these axioms do not tell us what the set function *P*(A) is, only what properties it must satisfy

The classical definition of probability defines the probability function as

$P(A) = \frac{N(A)}{N(S)}$ for any event *A* in the sample space *S*

Note that this definition satisfies all three axioms

Elementary properties of probability functions

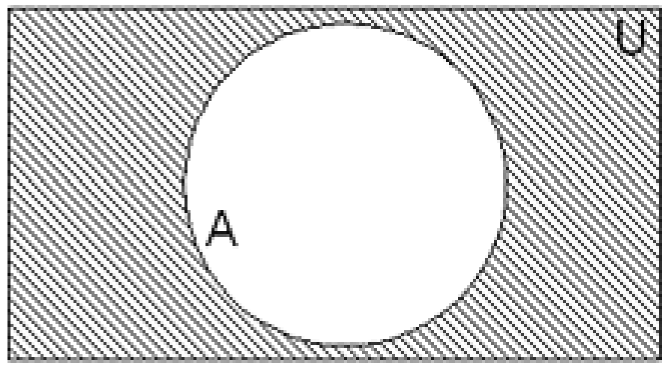
Theorem 3.4. If $A_1, A_2, \ldots, A_n$ are mutually exclusive events in a sample space S, then by induction on Axiom 3,

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

Theorem 3.7. If A is an event in S, then

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

Proof: $P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(S) = 1$



Example: drawing 5 card with at least one spade

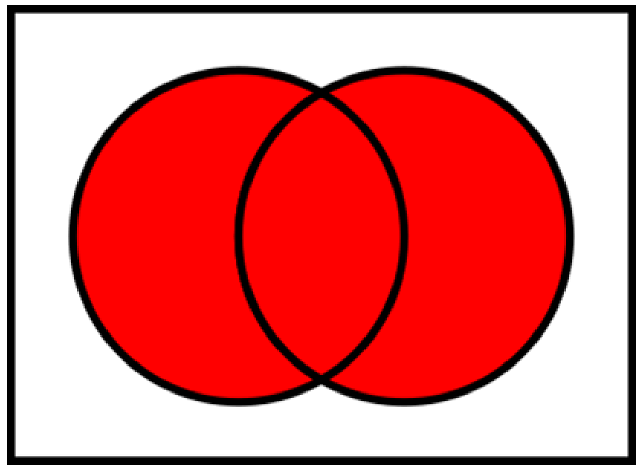
N(A) is very difficult to count, but N($\bar{A}$) is easier,

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{\binom{39}{5}}{\binom{52}{5}}$$

Elementary properties of probability functions

Theorem 3.6. If *A* and *B* are any (not necessarily mutually exclusive) events in *S*, then

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



Example: Find the probability of drawing 5 card with one spade or one club.

A: drawing 5 cards with one spade; B: drawing 5 card with one club.

N(A) = $\binom{13}{1}\binom{39}{4}$, N(B) = $\binom{13}{1}\binom{39}{4}$, N(A∪B) = $\binom{13}{1}\binom{13}{1}\binom{26}{3}$

P(A∪B) = $\dfrac{\binom{13}{1}\binom{39}{4}+\binom{13}{1}\binom{39}{4}-\binom{13}{1}\binom{13}{1}\binom{26}{3}}{\binom{52}{5}}$

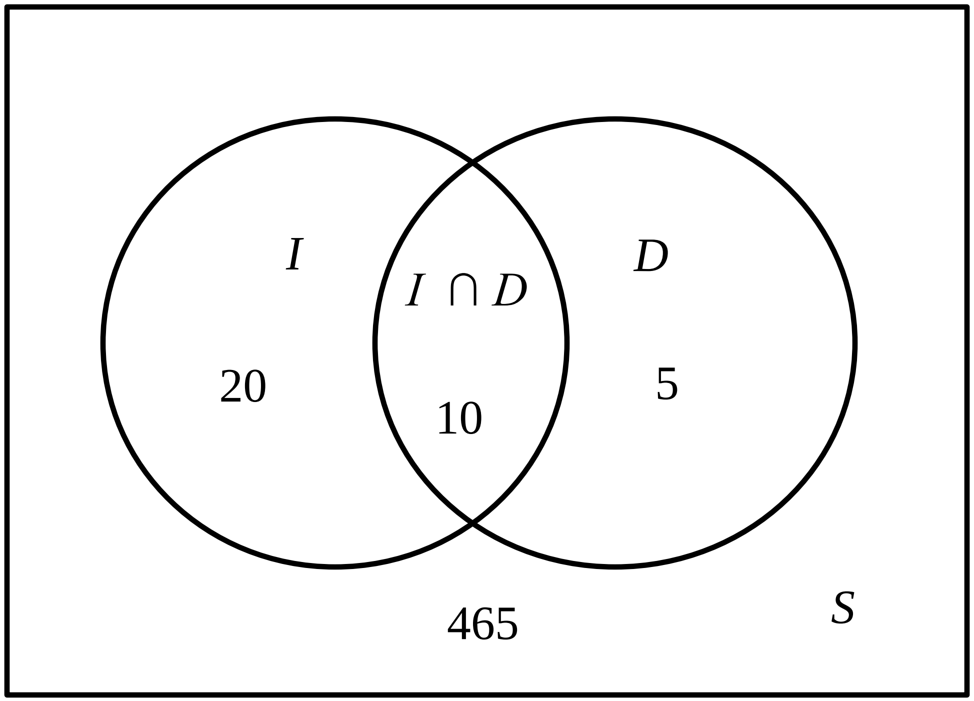


Fig. 3.7. Measurements on 500 machine parts
I = incompletely assembled, *D* = defective, *S* = satisfactory

Probability of picking up a unsatisfactory part:

I union D, P(I union D)=P(I) + P(D) – P(I intersect D)
=30/500 + 15/500 -10/500 =25/500

Conditional Probability.

The probability of an event is only meaningful if we know the sample space *S* under consideration.

The probability that you are the tallest person changes if we are discussing being the tallest person in your family, or the tallest person in this class.

This is clarified using the notation $P(A|S)$, the **conditional probability** of event *A* relative to the sample space *S*.

(When *S* is understood we simply use $P(A)$)

e.g. (using classical probability)

From Fig. 3.7,

$$P(D) = P(D|S) = \frac{N(D)}{N(S)} = \frac{N(D \cap S)}{N(S)} = \frac{10+5}{500} = \frac{3}{100} = 0.03$$

$$P(D|I) = \frac{N(D \cap I)}{N(I)} = \frac{10}{30} = \frac{1}{3} = 0.33\overline{3}$$

Note:

$$P(D|I) = \frac{\frac{N(D \cap I)}{N(S)}}{\frac{N(I)}{N(S)}} = \frac{P(D \cap I)}{P(I)}$$

Conditional Probability.

If *A* and *B* are any events in *S* and $P(B) \neq 0$, the conditional probability of *A* relative to *B* (i.e. *A* often stated 'of *A* given *B*') is

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

From the definition of conditional probability (CP) we see that

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

Since $P(B \cap A) = \mathrm{P}(\mathrm{A} \cap B)$, we have from (CP) and (*)

Theorem 3.8. If *A* and *B* are events in *S*, then

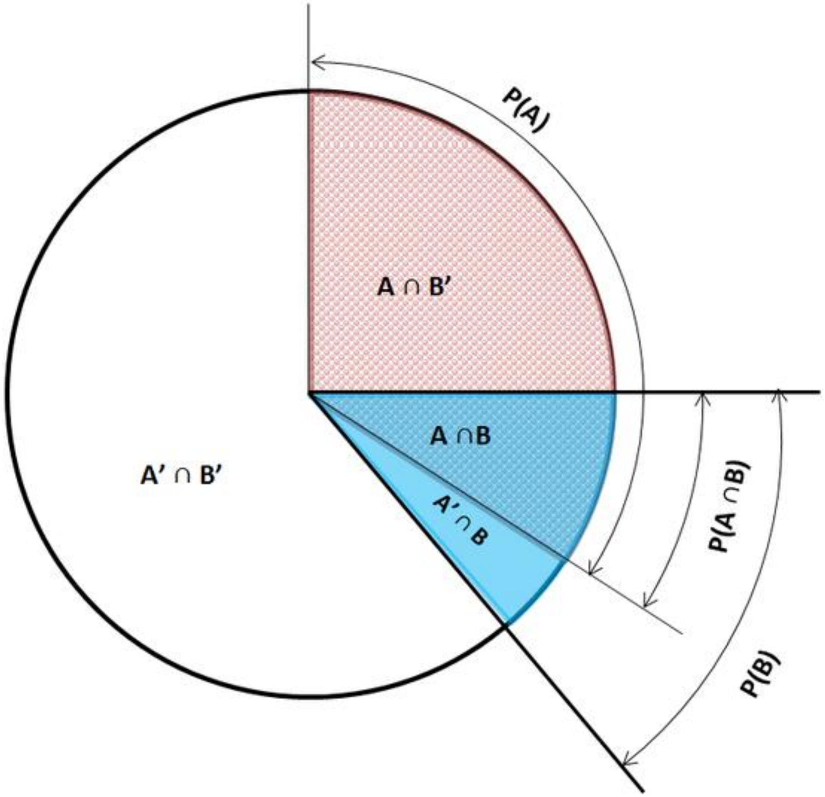
$$P(A \cap B) = \begin{cases} P(A) \cdot P(B|A) & \text{if } \mathrm{P(A)} \neq 0 \quad (\text{from } (*)) \\ P(B) \cdot P(A|B) & \text{if } \mathrm{P(B)} \neq 0 \quad (\text{from (CP)}) \end{cases}$$

Theorem 3.8 is the **general multiplication rule of probability**

Theorem 3.8 is a rather shocking statement. The definition (CP) of conditional probability implies that we compute P(A|B) by knowing P(A∩B) and P(B). However, Theorem 3.8 implies we can compute P(A∩B) by knowing P(A|B) and P(B).

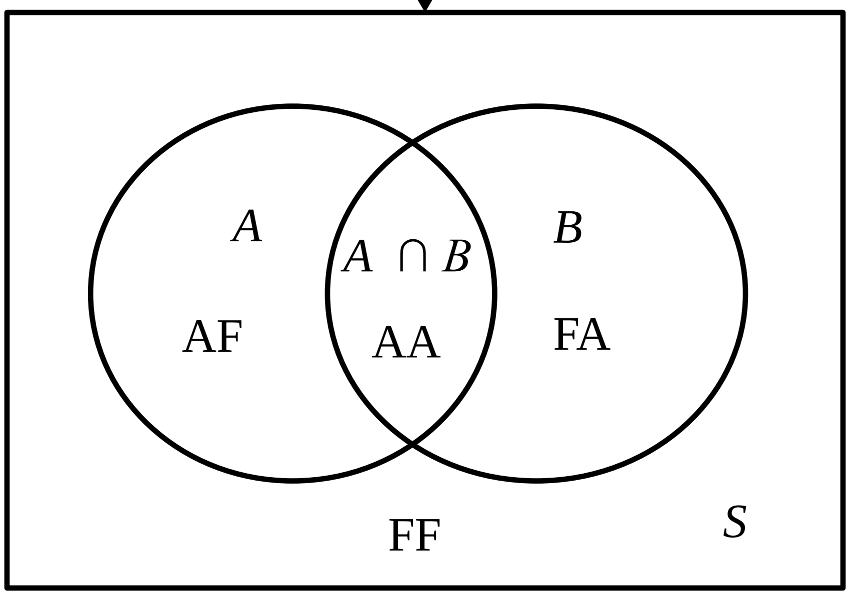
This implies that we often have another means at hand for computing P(A|B) rather than the definition (CP) !! (See next example)

The Venn Diagram on Conditional probability



e.g. Use of the general multiplication rule
20 workers, 12 are For, 8 are Against. What is the probability of randomly picking 2 workers that are Against? (Assume classical probability).

There are 4 classes out outcomes for the 2-picks: FF, FA, AF, AA
A diagram of the sample space of all 2 picks is



Set A: all outcomes where first worker is A

Set B: all outcomes where second worker is A

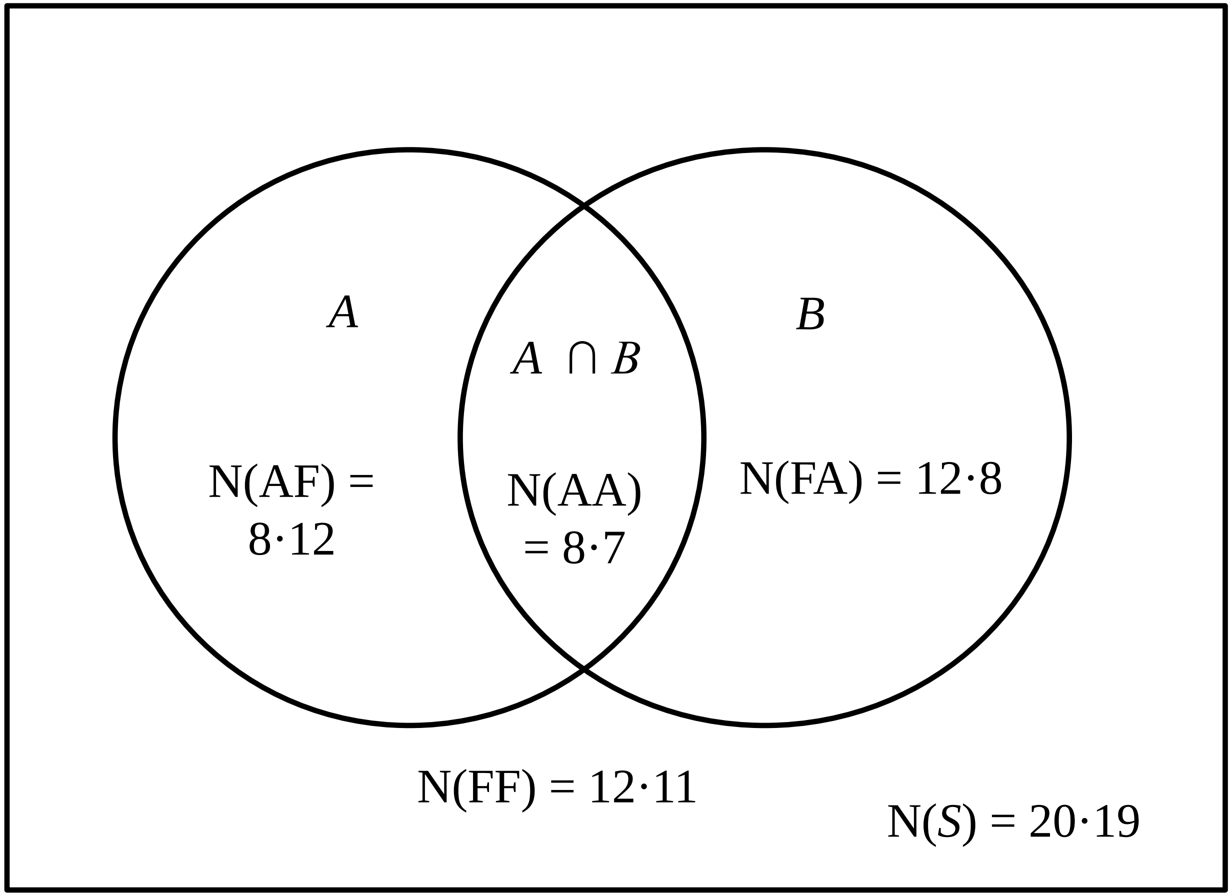
Desire $P(A \cap B) = P(A) \cdot P(B|A)$

$P(A)$ = the probability that the first is 'against'
= probability of picking one 'against' from the 20 workers
= N(against)/N(workers) = 8/20

$P(B|A)$ = the probability that the second is against given that the first pick is against
= probability of picking one 'against' from 19 workers (1 'against' removed)
= N(against)/N(workers) = 7/19

Therefore $P(A \cap B) = \frac{8}{20} \cdot \frac{7}{19} = \frac{14}{95}$

Check by classical calculation of probability



$$P(A) = \frac{8 \cdot 19}{20 \cdot 19} = \frac{8}{20}$$

$$P(B|A) = \frac{8 \cdot 7}{8 \cdot 19} = \frac{7}{19}$$

$$P(A \cap B) = \frac{N(A \cap B)}{N(S)} = \frac{8 \cdot 7}{20 \cdot 19}$$

If we find that $\boldsymbol{P(A|B) = P(A)}$**,** then we state that event *A* is **independent** of event *B*

We will see that event *A* is independent of event *B* *iff* event *B* is independent of event *A*.

It is therefore customary to state that ***A* and *B* are independent events.**

Theorem 3.9.

Two events *A* and *B* are independent events *iff* $P(A \cap B) = P(A) \cdot P(B)$

Proof:

→ If *A* and *B* are independent, that is $P(B|A) = P(B)$

Then, by Theorem 3.8,

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

← If $P(A \cap B) = P(A) \cdot P(B)$

Then, by definition of conditional probability,

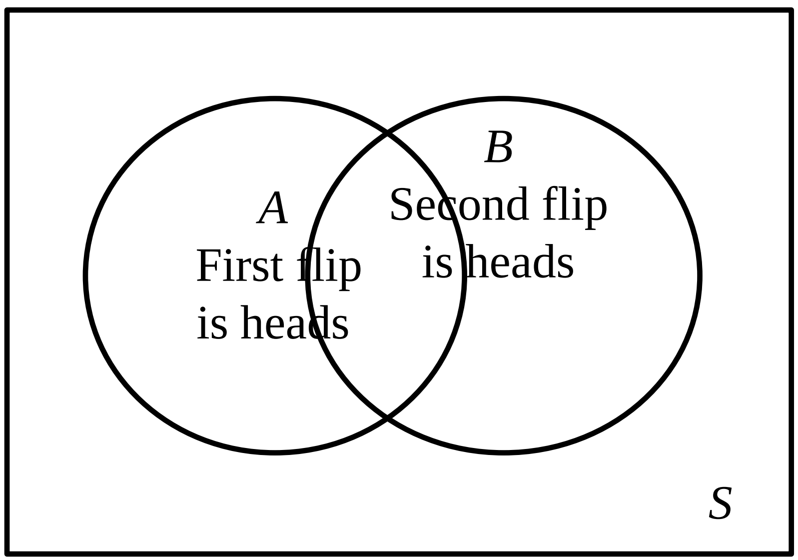
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

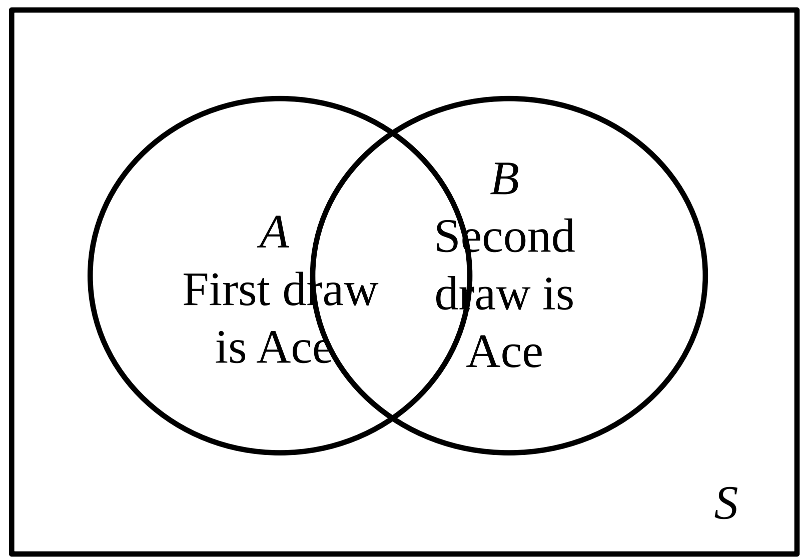
Theorem 3.9 is the **special product rule of probability** and states that **the probability that two independent events will both occur is the product of the probabilities that each alone will occur.**

Example: probability of getting two heads in two flips of a *balanced* coin

(*Assumption is* that *balance implies* that the two flips are independent)

Therefore $P(A \cap B) = P(A) \cdot P(B)$





Example: probability of selecting two aces at random from a deck of cards if first card replaced before second card drawn

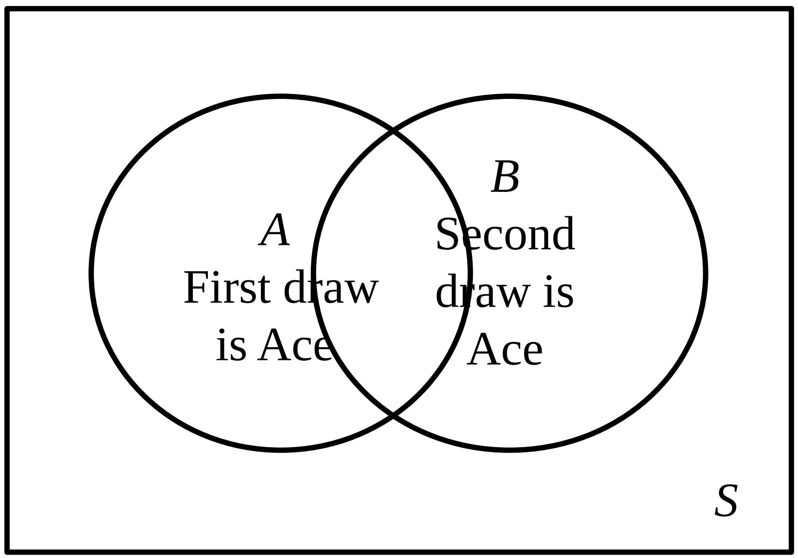
(*Assumption is* that replacing first card returns deck to original conditions making the two draws independent of each other)

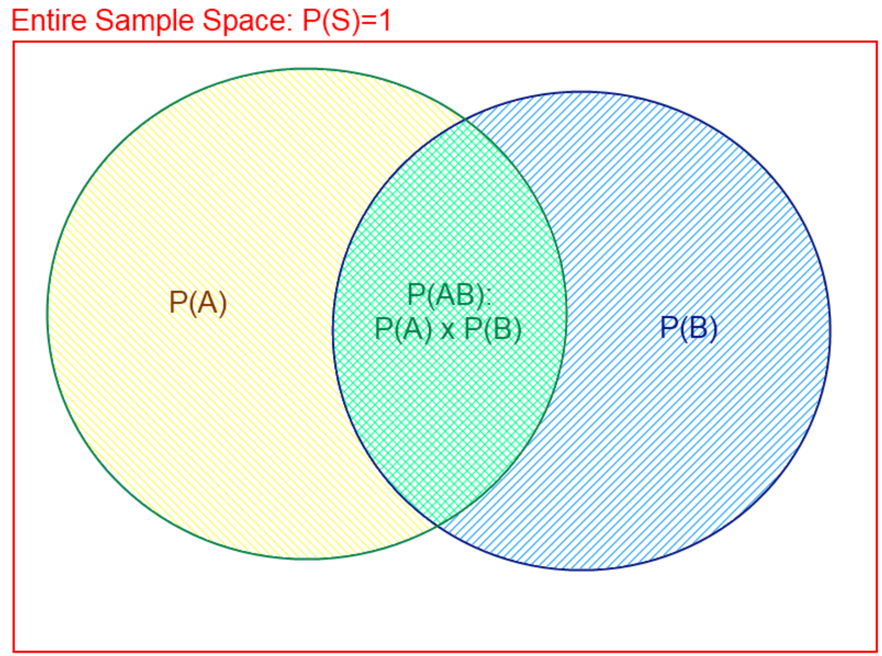
Therefore $P(A \cap B) = P(A) \cdot P(B)$

Example: probability of selecting two aces at random from a deck of cards if first card ***not*** replaced before second card drawn

(Picking the second card is now dependent on the first card choice)

Therefore $P(A \cap B) = P(A) \cdot P(B|A)$





Example: (false positives) 1% probability of getting a false reading on a test

Assuming that each test is independent of the others:

(a) probability of two tests receiving an accurate reading

$(0.99)^2$

(b) probability of 1 test per week for 2 years all receiving accurate readings

$(0.99)^{104} \approx 0.35$ (!) (65% chance that 1 or more of the 104 tests fail)

Example: redundancy in message transmission to reduce transmission errors

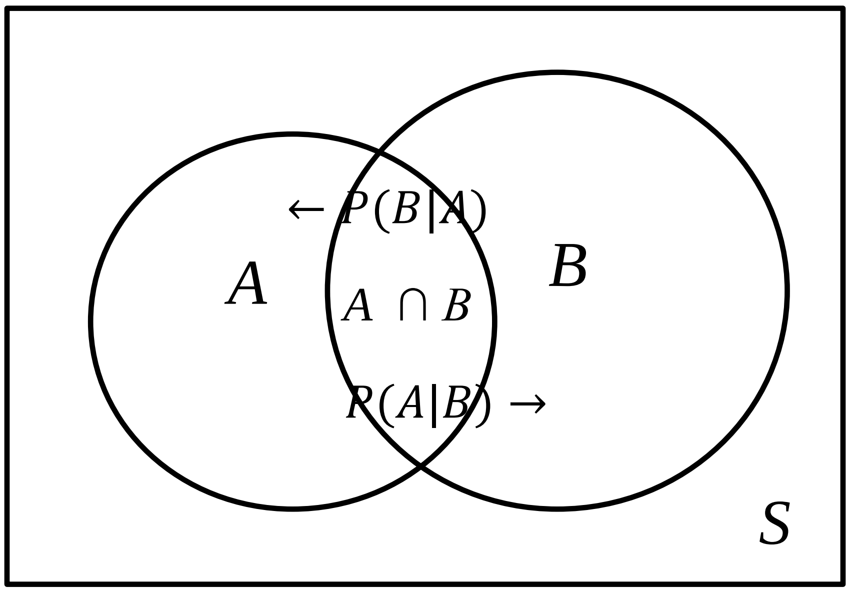
probability p that a $0 \to 1$ or $1 \to 0$ error occurs in transmission

sent	reception possibility	probability of reception	read as	Probability of reading
111	111	$(1-p)^3$	111	
	110	$p(1-p)^2$	111	
	101	$p(1-p)^2$	111	$(1-p)^2(1+2p)$
	011	$p(1-p)^2$	111	
	001	$p^2(1-p)$	000	
	010	$p^2(1-p)$	000	
	100	$p^2(1-p)$	000	$p^2(3-2p)$
	000	p^3	000	

p		0.01	0.02	0.05
Prob of reading correct	triple mode	0.9997	0.9988	0.05
	single mode	0.99	0.98	0.95

Theorem 3.8 shows that P(A|B) and P(B|A) are related, specifically:

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)} \qquad \text{(B)}$$



Remember,
P(A|B) is the ratio of the probability of event A∩B to the probability of event A
and
P(B|A) is the ratio of the probability of event A∩B to the probability of event B

Therefore to go from P(A|B) to P(B|A) one has to apply a correction, by multiplying and dividing respectively, by the probability of B and the probability of A.

In the above Figure probabilities are represented by area. P(B|A) is larger than P(A|B) by the area fraction P(B)/P(A)

Example: each year about 1/1000 people will develop lung cancer. Suppose a diagnostic method is 99.9 percent accurate (if you have cancer, it will be 99.9 Percent being diagnosed as positive, if you don’t have cancer, it will be 0.1 percent Being diagnosed as positive). If you are diagnosed as positive for lunge Cancer, what is the probability that you really have cancer?

Solution: Your being not have lung cancer is A, your being diagnosed positive is B. If you are diagnosed positive, what is the probability of being healthy? That is

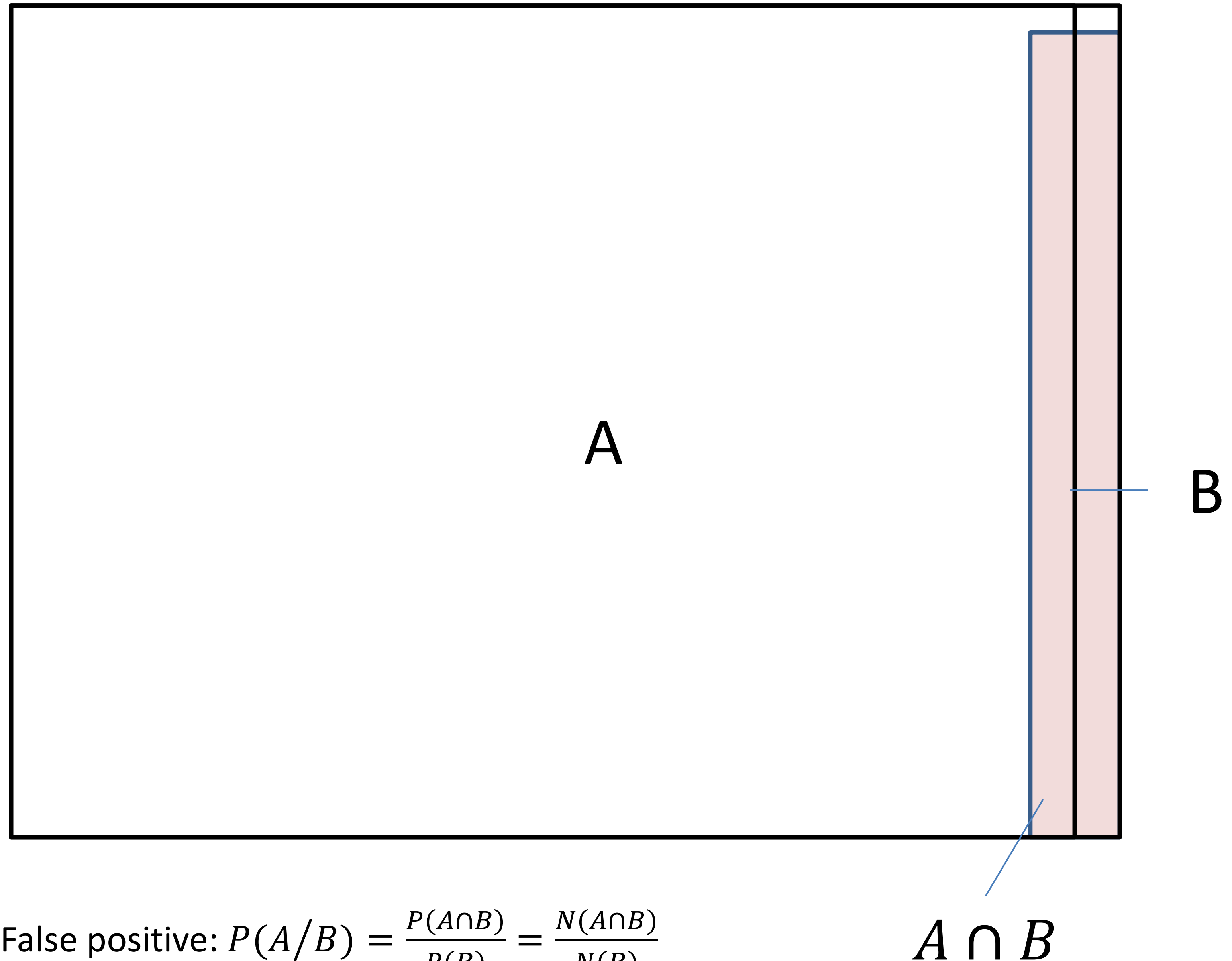
$$P(A/B) = \frac{P(A)}{P(B)} P(B/A)$$

P(A) = 0.999, P(B/A) = 0.001, P(B) = ?

P(B) = P(B/A)P(A) + P(B/A’)P(A’) = 0.001*0.999 + 0.999*0.001 = 0.001998

Substituting into the calculation:

P(A/B) = 0.999*0.001/0.0019 = 0.5

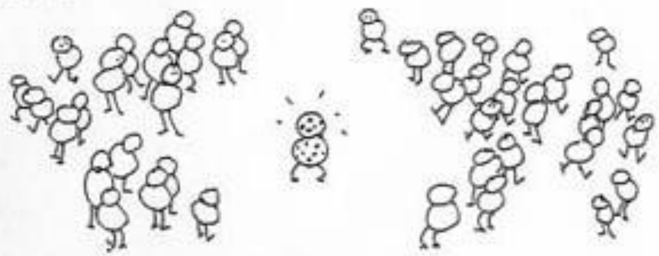


False positive: $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{N(A \cap B)}{N(B)}$

BAYES THEOREM and the case of the false positives...

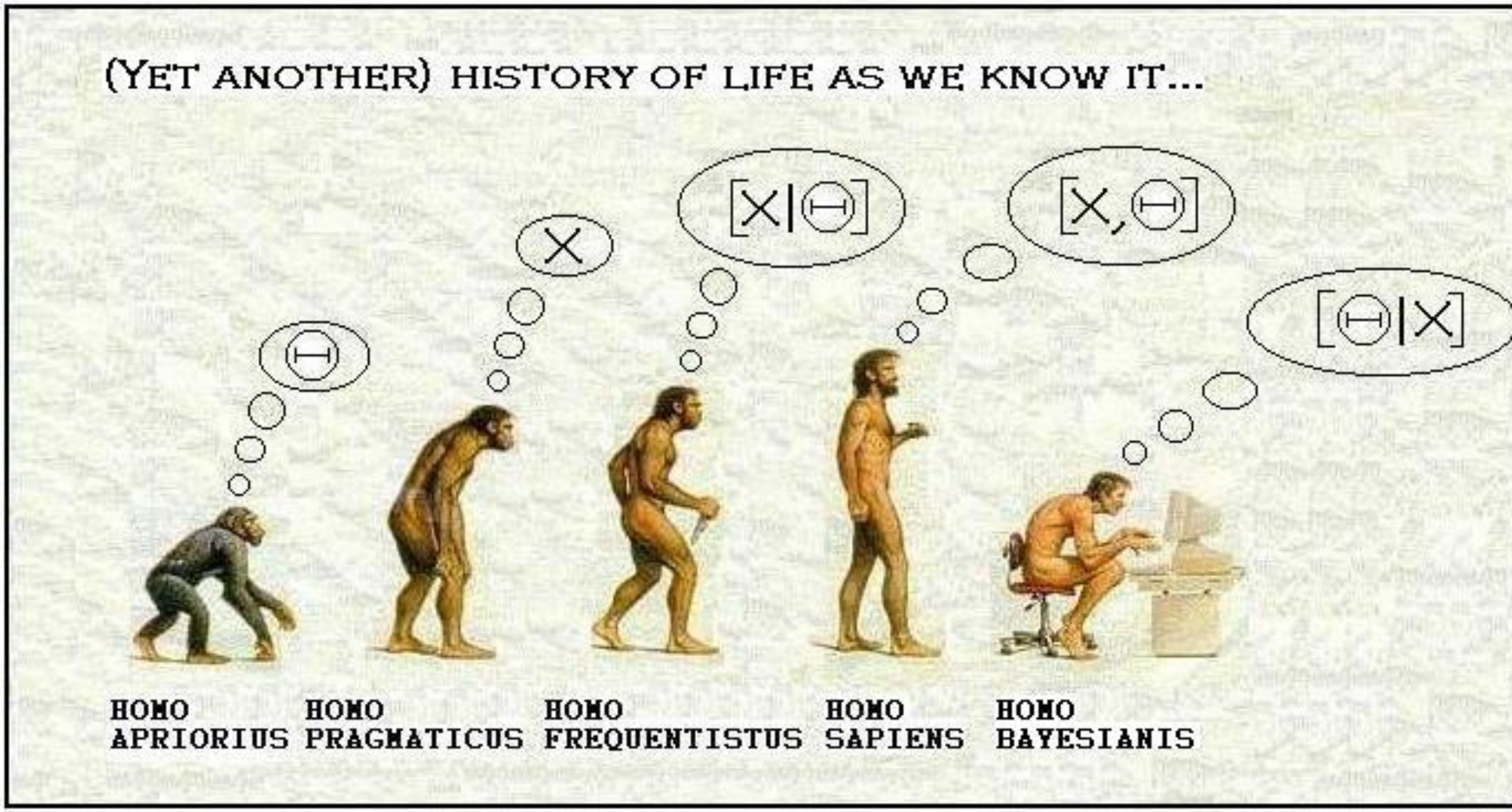
FOR A MORE SERIOUS APPLICATION OF CONDITIONAL PROBABILITY, LET'S ENTER AN ARENA OF LIFE AND DEATH...

SUPPOSE A RARE DISEASE INFECTS ONE OUT OF EVERY 1000 PEOPLE IN A POPULATION...



AND SUPPOSE THAT THERE IS A GOOD, BUT NOT PERFECT, TEST FOR THIS DISEASE: IF A PERSON HAS THE DISEASE, THE TEST COMES BACK POSITIVE *99%* OF THE TIME. ON THE OTHER HAND, THE TEST ALSO PRODUCES SOME *FALSE POSITIVES*. ABOUT *2%* OF UNINFECTED PATIENTS ALSO TEST POSITIVE. AND YOU JUST TESTED POSITIVE. WHAT ARE YOUR CHANCES OF HAVING THE DISEASE?

The evolution of thinking

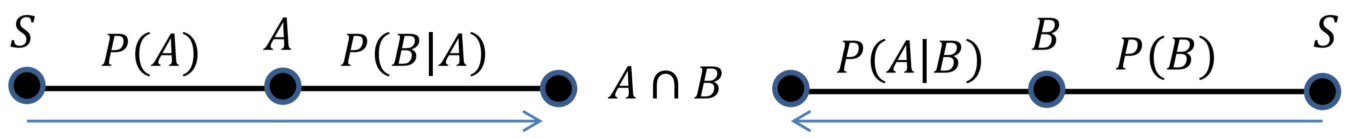


The relation

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

is a specific example of **Bayes' Theorem**. On the right hand side we have the (conditional) probability of getting outcome A considered as part of event B (having occurred). On the left hand side, we have the probability of getting outcome B considered as part of event A (having occurred).

This can be diagrammed as follows:



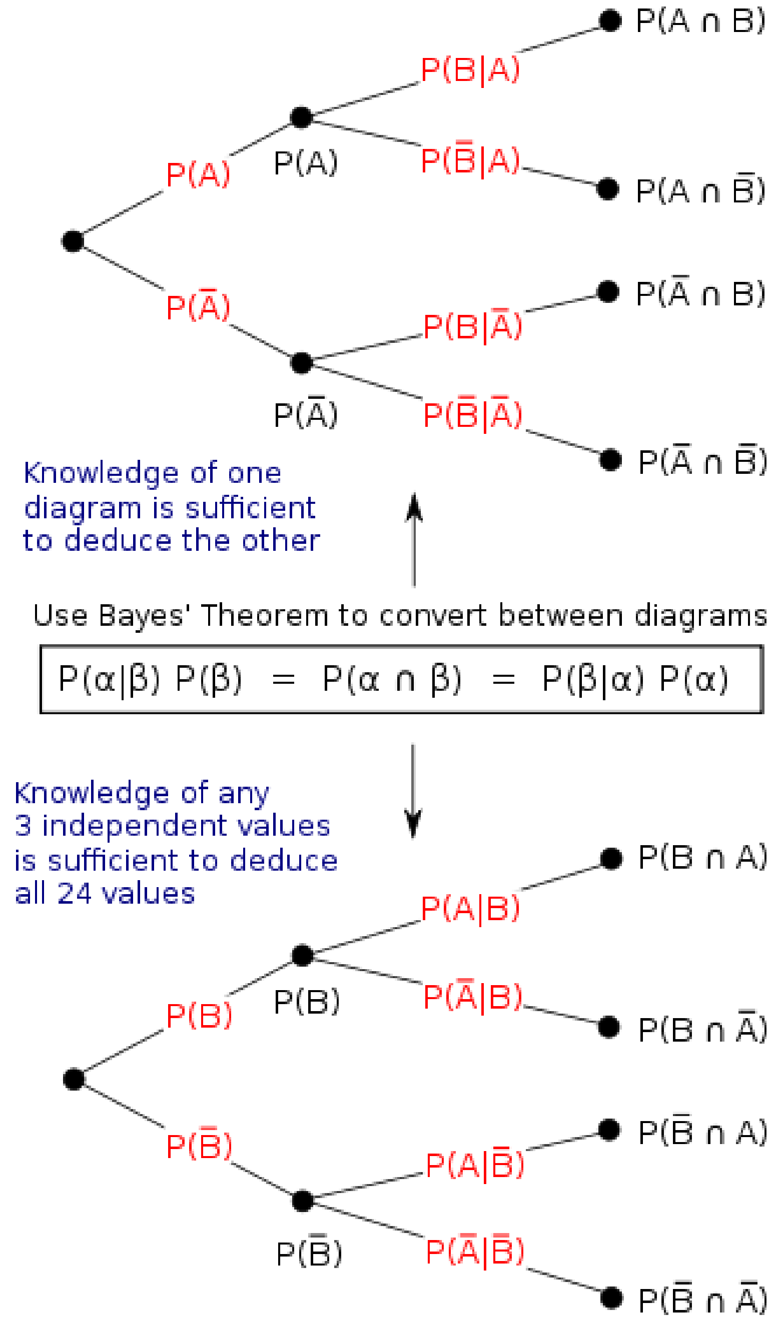
or more completely ….

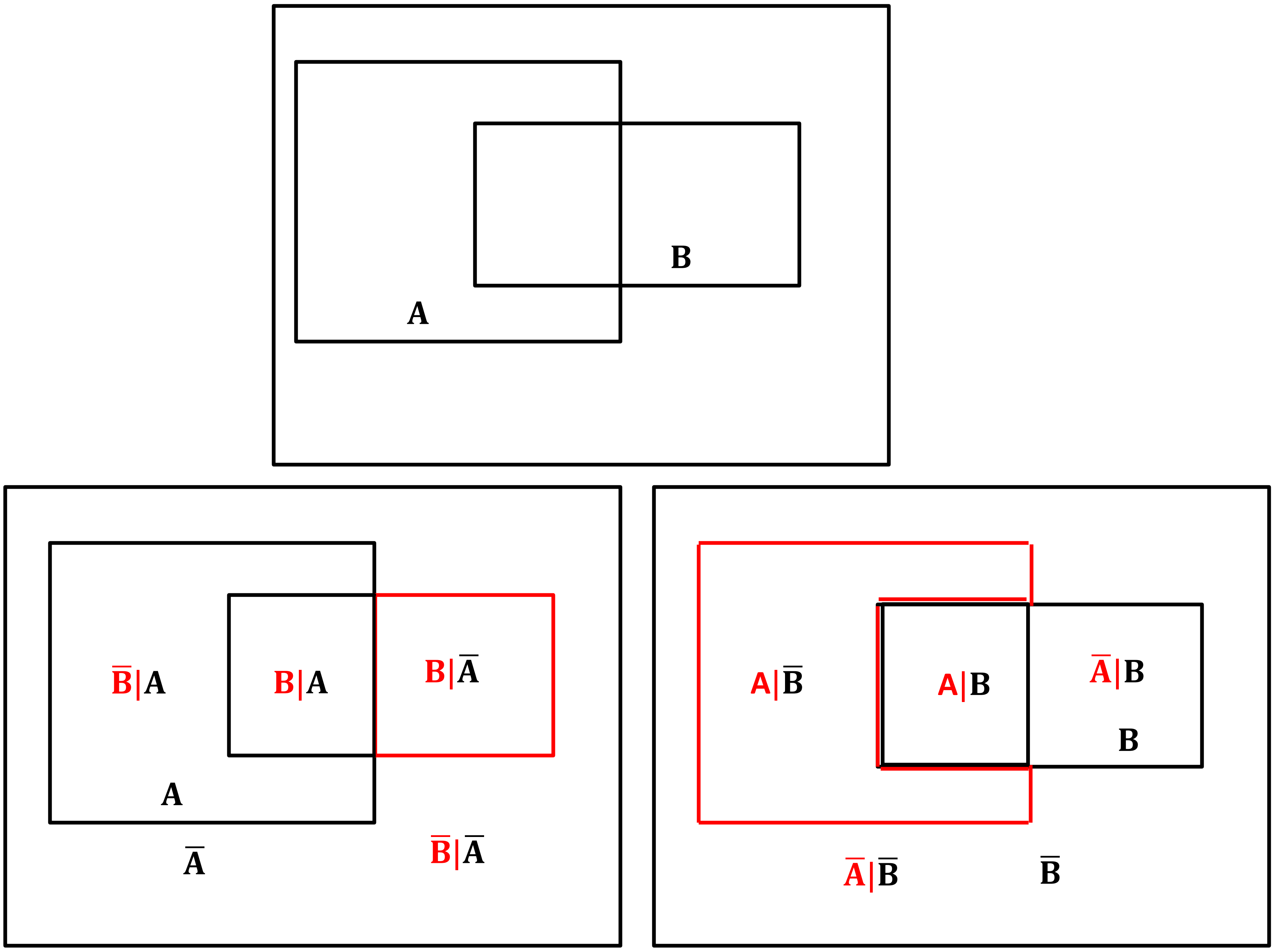
Partition all outcomes into those with and without property A and then subpartition into those with and without property B

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

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

Partition all outcomes into those with and without property B and then subpartition into those with and without property A



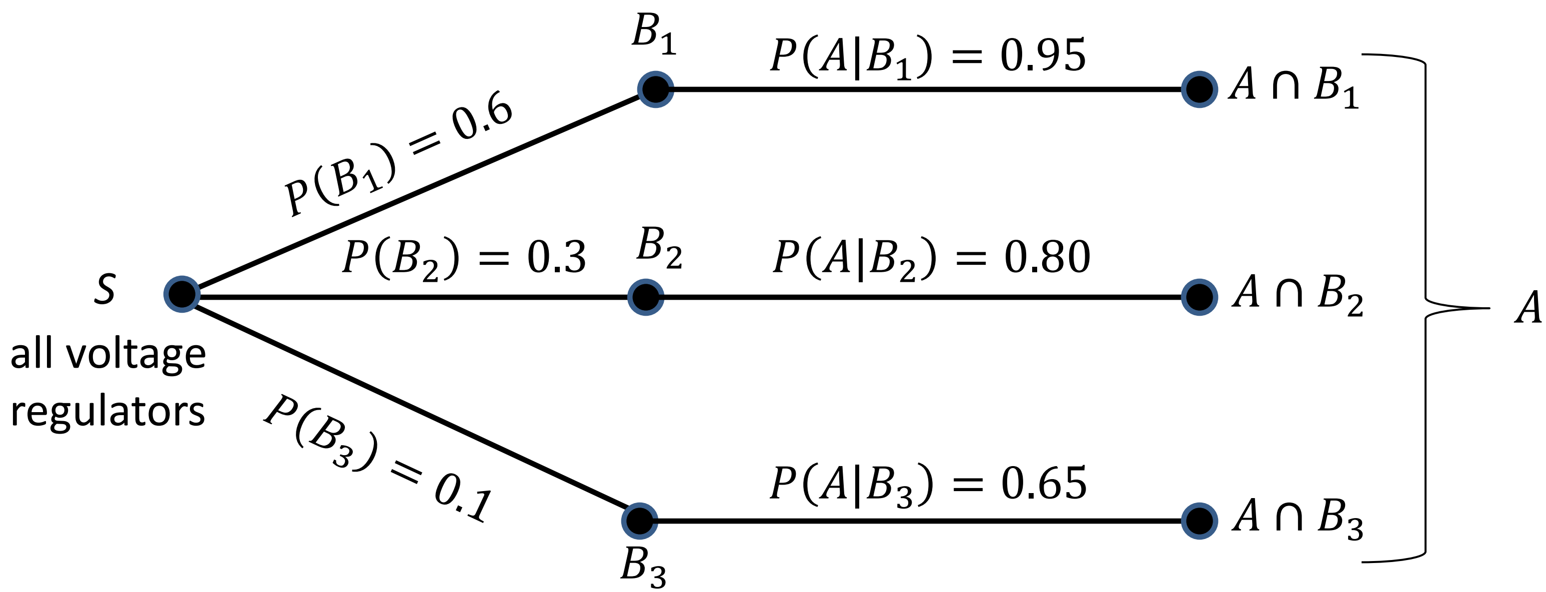


B
A
$\bar{B}|A$
$B|A$
$B|\bar{A}$
A
$\bar{A}$
$\bar{B}|\bar{A}$
$A|\bar{B}$
$A|B$
$\bar{A}|B$
B
$\bar{A}|\bar{B}$
$\bar{B}$

Bayes' result can be generalized.
Consider three mutually exclusive events, $B_1, B_2,$ and B_3, one of which must occur.

e.g. B_i are supply companies of voltage regulators to a single manufacturer.

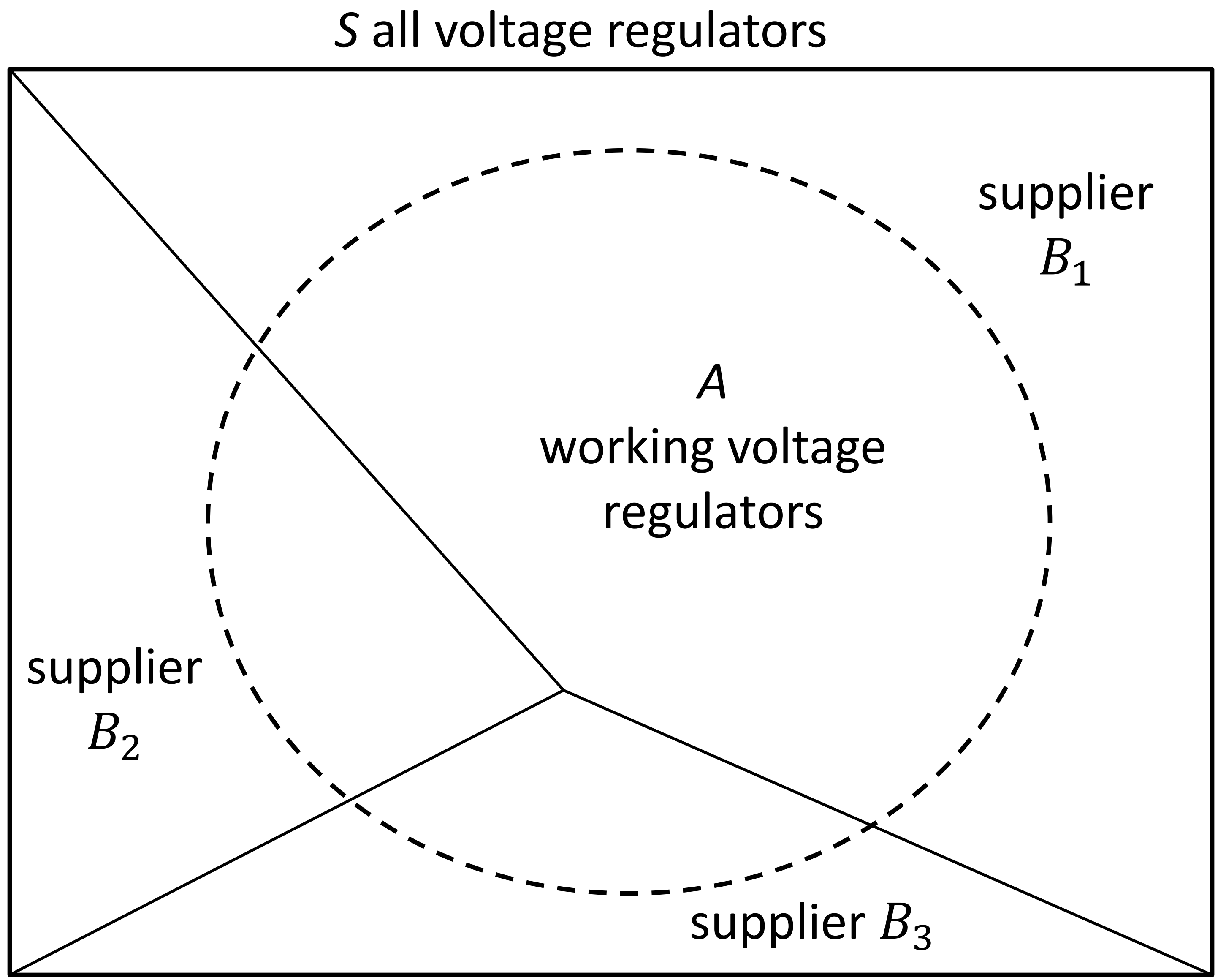
Let A be the event that a voltage regulator works satisfactorily. This might be diagrammed as follows for a hypothetical manufacturer



$P(B_1) = 0.6$ is the probability of getting a regulator from company B_1

$P(A|B_1) = 0.95$ is the probability of getting a working regulator from company B_1

Choosing at random from all regulators, what is the probability of getting a working regulator? (i.e. what is $P(A)$?)


S all voltage regulators
supplier B_1
A working voltage regulators
supplier B_2
supplier B_3

From the diagram we see

$$A = A \cap (B_1 \cup B_2 \cup B_3) = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$$

As $B_1, B_2, \text{and } B_3$ are mutually exclusive

so are $A \cap B_1, A \cap B_2, \text{and } A \cap B_3$ (see diagram)

By Theorem 3.4

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

By Theorem 3.8

$$P(A) = P(B_1) \cdot P(A|B_1) + P(B_2) \cdot P(A|B_2) + P(B_3) \cdot P(A|B_3)$$

Theorem 3.10 If $B_1, B_2, \ldots, B_n$ are mutually exclusive events of which one must occur, then for event *A*

$$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A|B_i)$$

Example: each year about 1/1000 people will develop lung cancer. Suppose a diagnostic method is 99.9 percent accurate (if you have cancer, it will be 99.9 Percent being diagnosed as positive, if you don't have cancer, it will be 0.1 percent Being diagnosed as positive). If you are diagnosed as positive for lunge Cancer, what is the probability that you really have cancer?

Now let's change the question: what is the probability that a person is tested positive?

Solution: If a person is sick, there is a probability of 0.999 to be tested positive;
If a person is healthy, there is a probability of 0.001 to be testes positive.

A: a person is healthy; $\bar{A}$: a person has lung cancer.
B: a person is tested positive.

$P(A) = 0.999;\ \ P(\bar{A}) = 0.001;\ \ P(B/A) = 0.001;\ \ P(B/\bar{A}) = 0.999.$

$$P(B) = P(A)P(B/A) + P(\bar{A})P(B/\bar{A}) = 0.999 \times 0.001 + 0.001 \times 0.999$$
$$= 0.001998$$

Theorem 3.10 expresses the probability of event *A* in terms of the probabilities that each of the constituent events B_i provided event *A*

(i.e. in our example, *P*(*A*) is expressed in terms of the probabilities that constituent B_i provided a working regulator)

Suppose, we want to know the probability that a working regulator came from a particular event B_i?

e.g. suppose we wanted to know $P(B_3|A)$

By definition of conditional probability

$$P(B_3|A) = \frac{P(B_3 \cap A)}{P(A)}$$

$$= \frac{P(A \cap B_3)}{P(A)}$$

$$= \frac{P(B_3) \cdot P(A|B_3)}{P(A)} \qquad \text{Theorem 3.8}$$

$$= \frac{P(B_3) \cdot P(A|B_3)}{\sum_{i=1}^{3} P(B_i) \cdot P(A|B_i))} \qquad \text{Theorem 3.10}$$

From the tree diagram $P(B_3|A) = \frac{0.1 \cdot 0.65}{0.6 \cdot 0.95 + 0.3 \cdot 0.80 + 0.1 \cdot 0.65} = 0.074$

The probability that a regulator comes from B_3 is 0.1

The probability that a working regulator comes from B_3 is $P(B_3|A) = 0.074$

The generalization of this three set example is **Bayes' theorem**

Theorem 3.11 If $B_1, B_2, \ldots, B_n$ are mutually exclusive events of which one must occur, then

$$P(B_r|A) = \frac{P(B_r) \cdot P(A|B_r)}{\sum_{i=1}^{n} P(B_i) \cdot P(A|B_i))}$$

for $r = 1, 2, \ldots, n$

The numerator in Bayes' theorem is the probability of achieving event *A* through the *r*'th branch of the tree.
The denominator is the sum of all probabilities of achieving event *A*.

e.g. Janet (B_1) handles 20% of the breakdowns in a computer system
Tom (B_2) handles 60%
Georgia (B_3) handles 15%
Peter (B_4) handles 5%

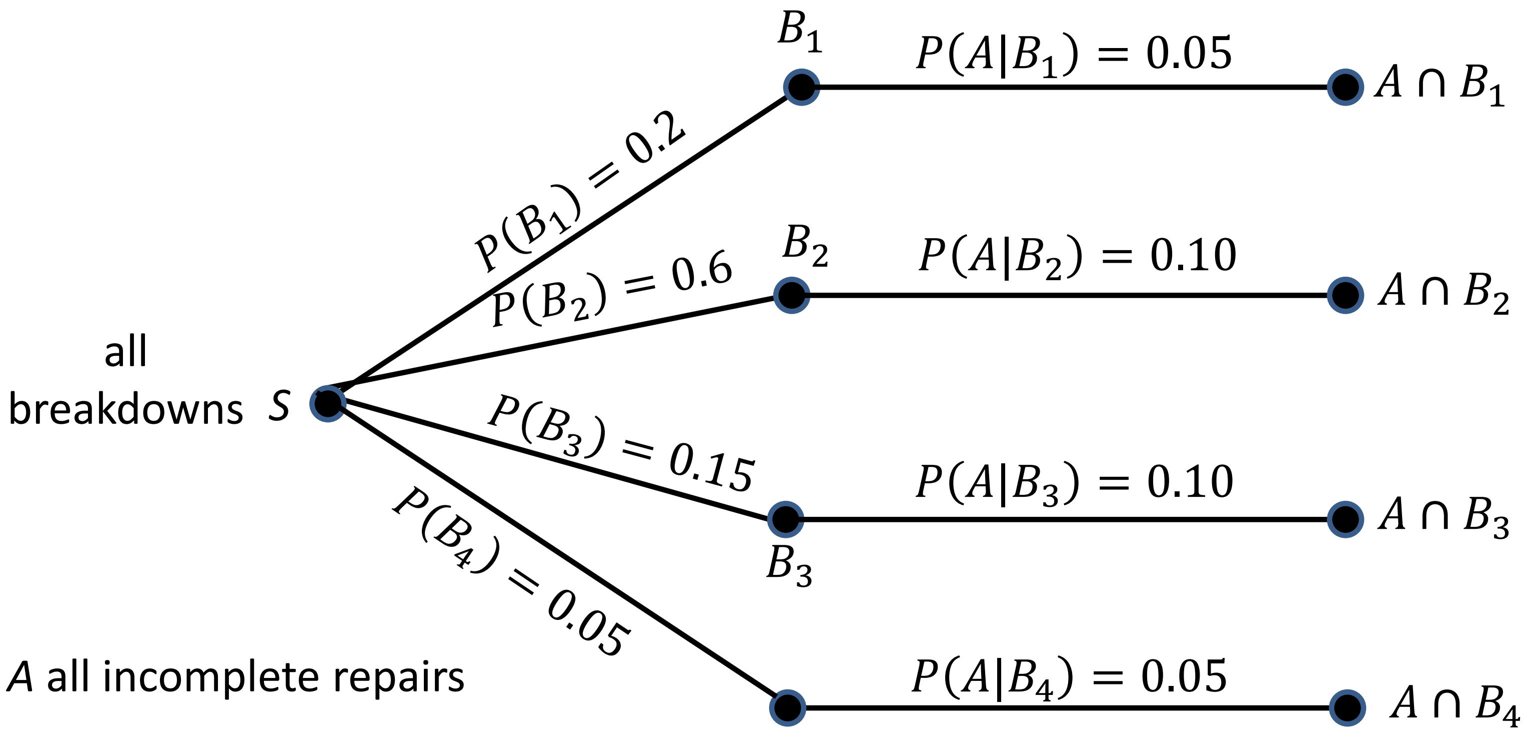
Janet makes an incomplete repair 1 time in 20 (i.e. 5% of the time)

Tom: 1 time in 10 (10%)

Georgia: 1 time in 10 (10%)

Peter: 1 time in 20 (5%)

If a system breakdown is incompletely repaired, what is the probability that Janet made the repair? (i.e. desire $P(B_1|A)$)



The probability that the incomplete repair was made by Janet is

$$P(B_1|A) = \frac{0.2 \cdot 0.05}{(0.2)(0.05)+(0.6)(0.1)+(0.15)(0.1)+(0.05)(0.05)} = 0.114$$

Therefore, although Janet makes an incomplete repair only 5% of the time, because she handles 20% of all breakdowns, she is the cause of 11.4% of all incomplete repairs

Summary of the chapter

1. Sample space, event, set operation, and Venn diagram
2. Counting principles, addition and multiplication
3. Permutation and combination
4. Classical probability, axioms
5. Independent events
6. Conditional probability
7. Bayes theorem

Formulas to remember

1. De Morgan's law

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$
$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

2. ${}_nP_r = \frac{n!}{(n-r)!}$ $\qquad \binom{n}{r} = \frac{n!}{(n-r)!r!}$
3. $P(A) = \frac{N(A)}{N(S)}$
4. $N(A \cup B) = N(A) + N(B) - N(A \cap B)$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $P(A) = 1 - P(\bar{A})$
7. For independent events A and B: $P(A \cap B) = P(A)P(B)$
8. Conditional probability: $P(A/B) = \frac{P(A \cap B)}{P(B)}$
9. If A and B are independent, $P(A/B) = P(A)$, $P(B/A) = P(B)$
10. Theorem 3.8 (bridge theorem)

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

11. Total probability

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A/B_i)$$

B_i's are mutually exclusive partition of the sample space.

12. Bayes theorem:

$$P(B/A) = \frac{P(B)P(A/B)}{P(A)} = \frac{P(B)P(A/B)}{P(B)P(A/B) + P(\bar{B})P(A/\bar{B})}$$

13. This can be generalized to give:

$$P(B_r/A) = \frac{P(B_r)P(A/B_r)}{P(A)} = \frac{P(B_r)P(A/B_r)}{\sum_{i=1}^{n} P(B_i)P(A/B_i)} \qquad r = 1,2,\cdots,n$$

14. Independent events are NOT exclusive events.